

Structural and functional characterization of kunitz inhibitor from *Cicer arietinum L.*

Thesis Submitted to AcSIR
For the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGICAL SCIENCES



by

Ameya Dipak Bendre

Registration Number: 10BB12A26061

Under the guidance of

Dr. Dhanasekaran Shanmugam

(Research Supervisor)

Dr. C. G. Suresh

(Research Co-Supervisor)

DIVISION OF BIOCHEMICAL SCIENCES
CSIR-NATIONAL CHEMICAL LABORATORY
PUNE – 411008, INDIA

July 2018



*Dedicated to
Caffeine, Sugar and the Man at the backstage!*

सीएसआईआर - राष्ट्रीय रासायनिक प्रयोगशाला

(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)

डॉ. होमी भाभा मार्ग, पुणे - 411 008, भारत



CSIR - NATIONAL CHEMICAL LABORATORY

(Council of Scientific & Industrial Research)

Dr. Homi Bhabha Road, Pune - 411 008, India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, **“Structural and functional characterization of kunitz inhibitor from *Cicer arietinum* L.”** submitted by **Mr. Ameya Dipak Bendre** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, embodies original research work under my/our supervision/guidance.

We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any Degree or Diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Date: 20th July 2018

Place: Pune

Mr. Ameya Dipak Bendre

(Student)

Dr. Dhanasekaran Shanmugam

(Research Supervisor)

Dr. C. G. Suresh

(Research Co-Supervisor)

Communication Channels

NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
: +91-20-25893400



FAX

Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
SPO's Office : +91-20-25902664

WEBSITE

www.ncl-india.org

DECLARATION

I, **AMEYA DIPAK BENDRE**, hereby declare that the work incorporated in the thesis and entitled “**Structural and functional characterization of kunitz inhibitor from *Cicer arietinum L.***” submitted for the award of the Degree of **Doctor of Philosophy in Biological Sciences** to the **Academy of Scientific & Innovative Research (AcSIR)**, New Delhi, has been carried out by me at Division of Biological Sciences, CSIR-National Chemical Laboratory, Pune-411008, India, under the supervision of Dr. Dhanasekaran Shanmugam. The work is original and has not been submitted as a part or full by me for any degree or diploma to this or any other university. I further declare that the material obtained from other resources has been duly acknowledged in this thesis.

Date: 20th July 2018



Ameya Dipak Bendre

(AcSIR Enrollment No. 10BB12A26061)

Acknowledgements

An acknowledgement of my appreciation

A bow card of an honest confession

A letter to show a physical symbol of recognition

A tribute to you

A written commendation

I could use all the words under the sun until my face turns blue

A very simple document just to say thank you...Just to say thank you...!

(Adopted from 'Thank you' by Michael Robinson)

- ✚ **Supervisors:** Dr. C. G. Suresh, Dr. Dhanasekaran Shanmugam, Dr. Sureshkumar Ramasamy
- ✚ **DAC members:** Dr. Ashok Giri, Dr. Sushama Gaikwad, Dr. A. Sudalai, Dr. G. Suryavanshi, Dr. Anjan Banerjee
- ✚ **CSIR-NCL:** Director(s) CSIR-NCL, SAO-NCL, AcSIR at CSIR-NCL, Head(s) of Biochemical Sciences Division and staff of Biochemical Sciences Division
- ✚ University Grant Commission, India for doctoral fellowship
- ✚ **RRCAT Indore (India):** Dr. Ashwani Kumar, Dr. Ravindra Makde, Dr. Biplab Ghosh
- ✚ **NanoTemper Technologies, Bangalore (India):** Dr. Sivramaiyah Nallapeta, Dr. Saji Menon
- ✚ **BAIF, Pune:** Dr. Santosh Jadhav, Dr. Suresh Gokhale
- ✚ **Lab 1866 group:** Dr. Manas Sule, Dr. Nishant Varshney, Dr. Tulika Javkar, Dr. Priyabrata Panigrahi, Dr. Ruby Singh
- ✚ **CD Spectroscopy:** Dr. Ekta Shukla, Sanskruthi Agrawal, Dr. Shakeel Abbasi
- ✚ **Plant Molecular biology group:** Dr. Yashvant Kumar, Dr. Gayatri Gurjar, Dr. Amey Bhide, Dr. Sheon Mary
- ✚ **DS Lab**

Special thanks to...

Shridhar

Deepanjan

Deepak Manu



Vijay

Debjyoti

Yashpal

Shiva

Dudes' Lab

CONTENTS

		Page no.
A	<i>List of abbreviations</i>	1
B	<i>Abstract</i>	3
Chapter 1	Understanding the Kunitz legume family	6
Section 1	Introduction	6
1.1	Protein-protein interactions	7
1.2	Plant protease inhibitors	9
1.3	Classification of protease inhibitors	10
1.3.1	<i>Based on their chemical nature</i>	10
1.3.2	<i>Based on their specificity</i>	10
1.3.3	<i>Based on sequence and structure features</i>	11
1.4	Kunitz type trypsin inhibitors (KTIs)	13
1.4.1	<i>Kunitz type inhibitors in animals</i>	13
1.5	Mechanism of protease inhibition	14
1.6	Applications of inhibitors	15
1.7	Need for detailed and systematic study	17
1.8	About protein crystallography	18
Section 2	Comprehensive analysis of KTI structures	19
2.1	Materials and methods	19
2.2	Amino acid distribution	23
2.3	Size and shape	23
2.4	Fold and secondary structure elements	24
2.5	Disulphide bond	26
2.6	Interactions within amino acid side chains	27
2.7	Aromatic-aromatic interactions	27
2.8	Cation– π interaction	28
2.9	Active site loop	29
	References	32
Chapter 2	Cloning and characterization of recombinant CaTI2	41
Section 1	Molecular cloning, expression and purification of recombinant CaTI2	41
1.1	Introduction	41

1.2	Materials	42
1.3	Methods	42
1.3.1	<i>Sequence analysis</i>	42
1.3.2	<i>cDNA synthesis, molecular cloning and heterologous expression in E. coli</i>	43
1.3.3	<i>Purification of 6×His-tagged CaTI2</i>	44
1.3.4	<i>Cloning, heterologous expression and purification of recombinant CaTI2 from P. pastoris</i>	44
1.4	Results and discussion	45
1.4.1	<i>Sequence analysis of chickpea KTIs</i>	45
1.4.2	<i>Cloning, expression and purification of CaTI2 from E. coli and P. pastoris</i>	47
Section 2	Biochemical and biophysical features of CaTI2	50
2.1	Introduction	51
2.2	Materials	52
2.3	Methods	52
2.3.1	<i>CaTI2 activity assays</i>	52
2.3.2	<i>Determination of K_d by MST</i>	53
2.3.3	<i>Far UV-circular dichroism spectroscopic studies</i>	53
2.3.4	<i>Differential Scanning Fluorimetry (DSF)</i>	54
2.3.5	<i>Thermal denaturation</i>	54
2.3.6	<i>Effect of pH on CaTI2</i>	54
2.3.7	<i>Treatment of the CaTI2 with urea</i>	55
2.3.8	<i>Effect of solvent polarity on CaTI2</i>	55
2.4	Results and discussion	55
2.4.1	<i>Inhibition of BPT by CaTI2</i>	55
2.4.2	<i>Dissociation constant determination</i>	56
2.4.3	<i>Circular Dichroism Spectroscopy</i>	57
2.4.4	<i>Thermal denaturation</i>	57
2.4.5	<i>pH-induced denaturation</i>	59
2.4.6	<i>CaTI2 stability towards Urea</i>	62
2.4.7	<i>Stability of CaTI2 in organic solvents</i>	64
2.5	Graphical Summary	66
	References	68

Chapter 3	Structural features of CaTI2	71
Section 1	Crystal structure determination and analysis of CaTI2	71
1.1	Materials	73
1.2	Methodology	74
1.2.1	<i>Heterologous expression and purification</i>	74
1.2.2	<i>Protein crystallization and crystal growth</i>	74
1.2.3	<i>Data collection and processing</i>	74
1.2.4	<i>Matthew's coefficient</i>	76
1.2.5	<i>Structure determination</i>	76
1.2.6	<i>Structure validation</i>	78
1.3	Results and discussion	79
1.3.1	<i>Overall structure of CaTI2</i>	82
1.3.2	<i>Comparison with homologous structures</i>	85
Section 2	Assessment of interactions between CaTI2 and Trypsin	89
2.1	Methodology	90
2.1.1	<i>Molecular docking of CaTI2 with BPT</i>	90
2.1.2	<i>Crystallization of CaTI2 and BPT complex</i>	91
2.2	Results and Discussion	91
	Schematic summary	97
	References	98
Chapter 4	Summary and conclusion	102
Section 1	Summary of the thesis	102
Section 2	Conclusion	104
	References	106
C	<i>Annexures</i>	
C1	Symposia and workshops attended	107
C2	Poster presentations	108
C3	List of publications	109
C4	Bio data	

List of Abbreviations

ACN	Acetonitrile
Aka	Also known as
APS	Ammonium persulphate
BAPNA	N α -Benzoyl-L-arginine 4-nitroanilide
bp	Basepair
BPT	Bovine pancreatic trypsin
CaTI2	<i>Cicer arietinum L</i> trypsin inhibitor 2
CaTI3	<i>Cicer arietinum L</i> trypsin inhibitor 3
CCP4	Collaborative Computational Project No. 4
CD	Circular Dichroism
cDNA	Complimentary deoxyribonucleic acid
DMSO	Dimethyl sulphoxide
dNTP	Deoxy-nucleotide triphosphate
DrTI	<i>Delonix regia</i> Trypsin inhibitor
DSF	Differential Scanning Fluorimetry
EcTI	<i>Enterolobium contortisiliquum</i> Trypsin inhibitor
EPR	Electron Spin Resonance
ETI	<i>Erythrina caffra</i> Trypsin inhibitor
Eth	Ethanol
Fnorm	Frobenius norm of two matrices
IPTG	Isopropyl β -D-1-thiogalactopyranoside
iRDP	<i>in silico</i> Rational Design of Proteins
kDa	Kilo Dalton
KTI	Kunitz type inhibitor
LB	Luria-Bertani media
LLG	Log Likelihood Gain
LITI	<i>Leucaena leucocephala</i> trypsin inhibitor
MEGA	Molecular Evolutionary Genetics Analysis

Meth	Methanol
mM	Milli molar
Mol. wt.	Molecular weight
MRE	Mean residue Ellipticity
mRNA	Messenger RNA
MST	Microscale thermophoresis
NCBI	National Centre for Biotechnology Information
Ni-NTA	Ni ²⁺ -nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
NRMSD	Normalized root mean square deviation
ns	Nanoseconds
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PI	Protease inhibitor
PPI	Protein-protein interaction
Prop	Propanol
RMSD	Root mean square deviation
Rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
STI	Soybean trypsin inhibitor
TKI	<i>Tamarindus indica</i> Kunitz inhibitor
T_m	Melting temperature

Abstract of the thesis

Enzyme-inhibitor complexes have been studied as a model system to understand protein-protein interactions. Mostly, inhibitors interact with enzymes the way the substrate binds to the enzyme. Thus studying such complexes gives valuable information about various types of interactions and mechanisms by which a molecule binds to the protein and brings about any particular biochemical reaction/change.

Kunitz inhibitors (KTI) are the inhibitors of the various serine proteases, a widest class of proteases found in nature. The KTI function during growth, development, and stress conditions in plants. It is also believed that they are involved in defense from pathogenic proteases. These findings lead to experiments where scientists attempted to use KTI as pest control agents. On ingestion of leaves or artificial diet containing KTI, the insect pest could not digest the food taken in. As a result insect pest died in larval stage due to malnourishment and starvation.

We report in this thesis, the extensive studies carried out on one of the ten KTI identified from commercially important pulse crop Chickpea, CaTI2. The CaTI2 inhibitor belongs to Kunitz legume (PF00197) superfamily. First identified in Soybean (named STI), the family characterizes with single peptide inhibitors having about 20kDa molecular weight and single reactive site. The inhibitors inhibit several members of serine protease family including plasmin, human plasma kallikrein, trypsin, chymotrypsin and Factor XIIa etc. The inhibitory loop lies between 4th and 5th β sheet. It consists of an Arg residue that brings about the inhibition of trypsin with the help of distant Asn residue (Asn13 in STI).

Three dimensional structure determination, along with biochemical and biophysical characterization, have been carried out on CaTI2 in this thesis. The objective of this study is to achieve a better understanding of the mechanism of action of these inhibitors, which in turn is expected to help in protein engineering them for further applications. Results from the studies reported here have helped us to compare the structural, mechanistic and evolutionary relationships of industrially and therapeutically important inhibitor.

Chapter 1: Understanding the Kunitz legume family

This chapter presents the literature survey of the research reported on mainly plant KTIs. The knowledge about the classes, distribution, general mechanism and applications of various inhibitor families have been discussed briefly. We have tried to comprehend and give detailed account of structural aspects of available plant KTIs. The known KTI structures were analyzed for various parameters like amino acid composition, crystal parameters, and various interactions among the side chains of residues. The inhibitory loop regions of available structures are compared to deduce the common mechanism of inhibition. The characteristic fold, different applications of KTIs has also been discussed.

Chapter 2: Cloning and characterization of recombinant CaTI2

CaTI2 is one of the KTIs identified from Chickpea. It was cloned in two expression systems, first in *E. coli*. Due to low expression levels, it was then cloned in *P. pastoris* system. The recombinant protein was purified and its biochemical activity was checked against Bovine pancreatic trypsin (BPT), both qualitatively and quantitatively. Active inhibitor was then further assessed for its biophysical characteristics such as pH and temperature stability, stability under varying polarity of alcohols and in presence of urea. The biophysical characteristics were correlated with biochemical function under similar conditions. The melting temperature and dissociation constant were determined using Differential Scanning Fluorimetry (DSF) and Micro Scale Thermophoresis, respectively (MST). Secondary structure transitions were monitored using Circular Dichroism spectroscopy (CD).

Although the protein was expressed in *E. coli* system, for better yield, we switched to *P. pastoris*. The native CD spectra profile matched with other β -trefoil proteins suggesting that the purified inhibitor was folded in its native conformation. Biophysical characterization revealed that the protein maintained its maximum activity under basic pH and up to 45 °C. It was almost inactive in extreme acidic conditions and higher temperatures. It had smaller K_d value compared to that of Soybean trypsin inhibitor suggesting that it could form stronger complex with trypsin. The stability of the inhibitor was found to be directly proportional to the solvent polarity. As urea is known to destabilize the structure, with increasing concentrations of urea, inhibitory potential was lost which also reflected in CD spectra suggesting loss of native conformation.

Chapter 3: Structural features of CaTI2

The CaTI2 was purified to homogeneity. After several crystallization trials, the data was collected for one of the CaTI2 crystals which diffracted at 2.9Å resolution. The inhibitor like other Kunitz family proteins folded in a β trefoil fold, a fold with twelve antiparallel β sheets connected by loops with negligible helical components. The inhibitory loop was present between 4th and 5th β sheets. The loop lacked all conserved residues that were observed in available Kunitz structures. The loop contains unique sequence arrangement and geometry unusual to plant Kunitz inhibitors. ConSurf analysis confirmed the above observation of variations in inhibitory loop while like all other Kunitz members; core β structures were highly conserved. It was also observed that the P1 position in loop was occupied by Ile77.

In absence of CaTI2-BPT complex structure, to confirm the P1 residue, molecular docking was performed with BPT. Docking results showed that Ile77 binds to trypsin active pocket in non-substrate fashion. The interfacial area of docked complex is found to be larger than that of STI-BPT complex. It supported findings of MST experiments that CaTI2-BPT complex might be stronger than STI-BPT complex. It was also observed that the inhibition of trypsin must be occurring due to entire inhibitor molecule and not due to just inhibitory loop element. To validate this finding, CaTI2-BPT complex was purified and crystallized. The complex structure determination is underway.

Chapter 4: Summary and conclusion

In this chapter we have tried to summarize the findings of our work. Our observations illustrate how variations at sequence level are translated through higher structural components of a protein. These sequence variations not only reflect in quaternary structure but also define the mode of action of the inhibitor. This chapter discusses the major highlights of the thesis and summarizes the characteristics of the inhibitor under study.

The *in vitro* and *in silico* studies described in this thesis contribute to our knowledge of the interplay between the stability, structure and function of the enzyme-inhibitor system at molecular level, which can prove instrumental to improve their efficiency in future.

Chapter 1



Understanding the Kunitz legume family

Section 1: Introduction

The central dogma of biology can be summed up in three basic processes: the recognition of molecules, the crosstalk between these biomolecules and their interactions through binding. When the molecules interacting with each other are connected, a network of biological interactions is established. Based on the type of biomolecules involved, the interactions are classified in different groups. These interactions could be mainly between proteins and carbohydrates, protein and nucleic acids, or protein and protein. These interactions when considered together forms an ‘Interactome’ of a system (1).

Two decades ago the scientific community was introduced to a terminology called ‘Interactome’ by a French scientist Bernard Jacq (2). The term interactome literally means the set of overall physical interactions between the molecular members of any biological system say, a cell. The term is indirectly also implied to genetic interactions. Every protein that is synthesized in the cell is interrelated with other cellular proteins (3). In general, the interactome term is vaguely applied to protein-protein interactions (PPI) network because if we consider the overall networks of interactions within the interactome, the number of PPI dominates over other types of interactions networks (4).

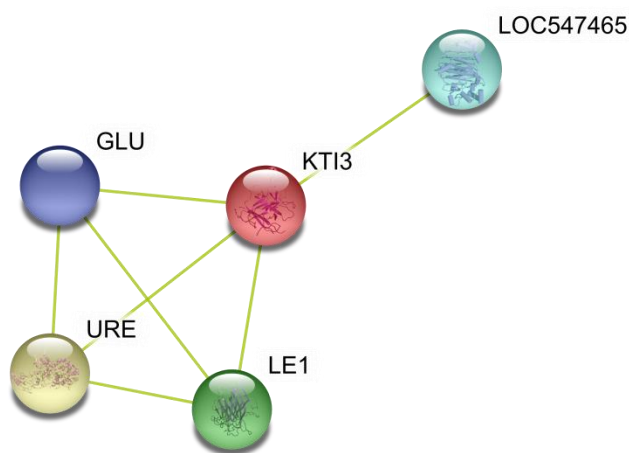


Figure 1: Schematic representation of set of interactions of Soybean trypsin inhibitor (KTI3) within Soybean interactome

Consider an interactome data of Soybean plant (5). Figure 1 shows that a Kunitz inhibitor from Soybean, KTI3 (aka STI) has interactions with four other proteins from Soybean

proteome. The proteins may not be always physically binding to each other but share same function. These proteins are Urease (Ure), Lectin (LE1), β -conglycinin (LOC547465) and an uncharacterized protein GLU. The structures inside the sphere are the crystal structures deposited for respective proteins. If the structure is not available, then the sphere appears empty. This is the way in which interactome network related to any protein is represented.

1.1 Protein-protein interactions

Proteins are the most diverse and versatile biomolecules in living systems. These not only act as *building blocks of the cell* but are also significant *regulators of various biological processes*. Often, the protein molecule does not act alone but is assisted by other proteins or biomolecules. Most of the biological processes in nature are governed by such PPIs. Some proteins display a transient effect i.e. their interactions are short lived, as seen in the case of enzymes; while some protein interactions are longer than transient and thus, form stable complexes, for example, as in case of the cytoskeleton (6). There are many ways in which PPIs are classified. PPIs can be homo-oligomeric or hetero-oligomeric (7). Homo-oligomeric interactions occur between identical chains while hetero-oligomeric as the name indicates, occur within chains of different types. The interactions may include multi-protein complexes of hetero-oligomeric nature. These also could be formed of homo-oligomeric proteins.

Another way of classifying PPIs is obligate and non-obligate interactions. In obligate PPI the individual chains are not stable structures on their own but the overall complex is stable *in vivo* (8). For example, a protein called Arc repressor protein that is involved in DNA binding. The protein exists in homodimer form to perform its function of DNA binding while monomer is highly unstable (9). The members of non-obligate PPIs exist independently and are stable as individual moiety like enzyme-inhibitor complexes. Whether the enzyme is present or not, the inhibitor peptide is stable on its own; similarly, in absence of the inhibitor molecule, the enzyme exists as a stable functional unit. On complexation, the enzyme and the inhibitor form a non-obligate transient interaction complex. In most cases, the members of non-obligate complexes are initially not localized together and thus exist as individual stable proteins (10). Usually, the non-obligate interactions form induced-fit complexes (11).

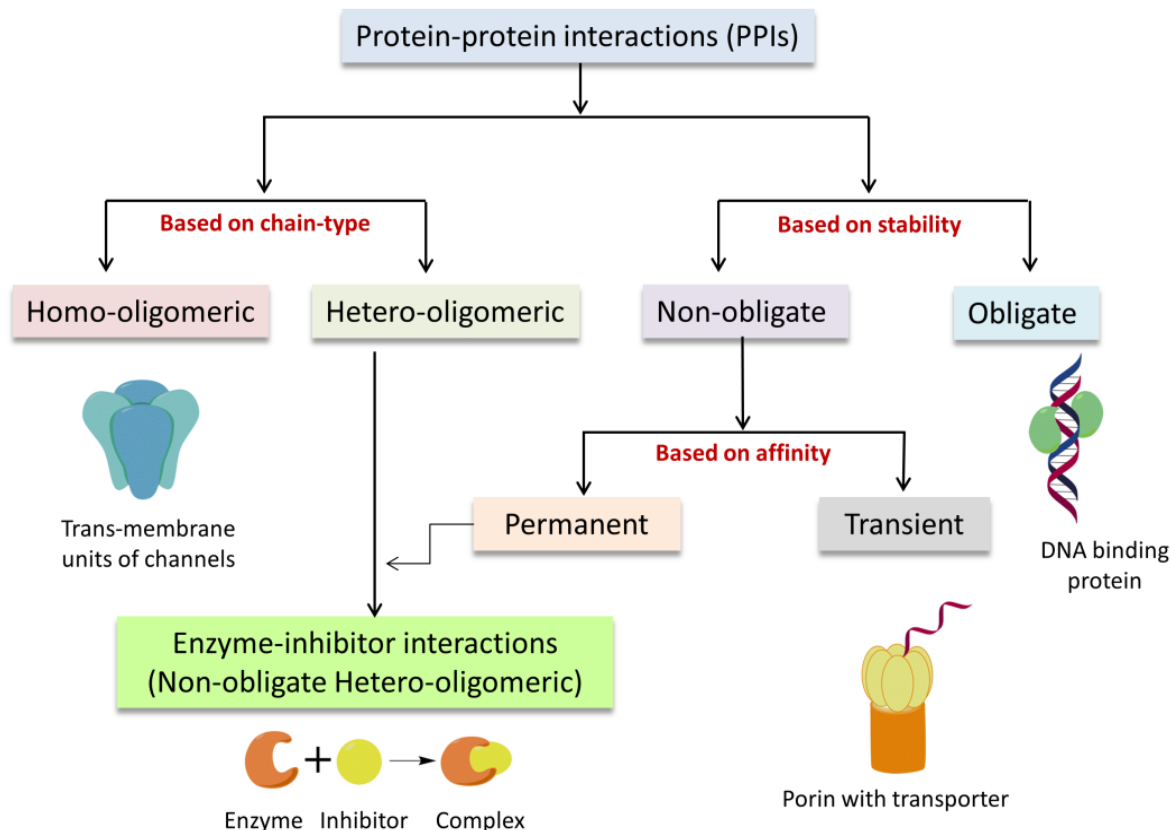


Figure 2: Types of protein-protein interactions

Interactomic studies with help of mainly yeast two-hybrid system and mass-spectrometry help in identifying the PPIs (12). The high resolution crystal structure data enable us to validate the findings of interactomics predictions. It also helps us to understand wide spectrum of individual interactions at atomic level occurring at different amino acids present between interfaces of interacting proteins. To study transient, homo-oligomeric obligate interactions, membrane proteins especially channels or porins form the ideal systems. On the other hand enzyme-inhibitor complexes such as protease-inhibitors prove to be a model system to learn and understand the permanent, non-obligate hetero-oligomeric interactions (13). The regulation of the activity of proteolytic enzymes by protein protease inhibitors is through PPIs. The complexes of protease inhibitors with their cognate proteases facilitated the study of PPIs.

Protein protease inhibitors (PIs) are ubiquitous in nature (14). Animals, plants and micro-organisms contain a number of protease inhibitors which form stoichiometric protein-protein complexes with several proteases (15). Under particular set of condition these complexes are reversible. PIs are present in multiple forms in different tissues of organisms.

1.2 Plant protease inhibitors

Plants form the rich source of proteins of different kinds. Modern classification of plant proteins is based on two aspects, protein function and molecular/ biochemical relationships. The functional proteins present in the seeds are further classified into three main classes (16). They are

1. Storage proteins which store nitrogen, carbon and sulfur
2. Structural and metabolic proteins that are essential for growth and structure of the seed
3. Defense proteins that are involved in plant defense

It was believed that the plant PIs act as the storage proteins due to their presence mainly in seeds. But there are evidences showing their involvement in defense from proteases from pathogens and pests (17). Along with PIs a wide variety of proteins are considered as defense proteins. These include structural proteins such as hydroxyl-proline rich glycoproteins, glycine rich proteins, amylases, thionins, chitinases, phytoalexins and pathogenesis related proteins. Defense proteins play a key role in protecting plants against microbial pathogens, invertebrate pests, environmental stress etc (18). Plants generate various inter and intra cellular signals to activate genes for the induction of various substances in response to an external attack (19). They include antibiotics, alkaloids, terpenes and defense proteins such as protease inhibitors (20).

Plant PIs have been investigated for various aspects including their utility as agents that prevent unwanted proteolysis thereby controlling physiological processes such as growth regulation, stress management, the protein turnover and metabolism etc. PIs are also studied as a model system for understanding PPIs (21). The most abundant source of PIs is plants, and these proteins are widely distributed throughout the plant kingdom especially in the

fabaceae family. Among the members of fabaceae, PIs are usually present in the storage tissues such as seeds, and while in case of solanaceae inhibitors are observed in tubers (22). They are usually localized inside the vacuolar protein bodies of the cell. A large number of PIs have been isolated and characterized which are mostly low molecular weight proteins about 4-20kDa, soluble and their polypeptide chains are non-glycosylated (23).

A vast body of research has been undertaken to study protease inhibitors in the context of plant defense. The expression of genes of protease inhibitors in important crop plants to protect them against pests and pathogens is gaining much attention in recent times (24).

1.3 Classification of protease inhibitors

1.3.1 Based on their chemical nature: PIs are broadly grouped into two categories viz: chemical protease inhibitors and protein protease inhibitors. Chemical inhibitor molecules mostly are synthetic small molecules which are highly specific and usually form covalent contacts with protease active pocket e.g. PMSF. There are exceptions where chemical inhibitors bind to proteases reversibly like Benzamidine hydrochloride.

Table 1: Chemical inhibitors with their substrate proteases

Chemical Inhibitor	Substrate	Ref.
Phenyl-methyl-sulfonyl Fluoride (PMSF)	Serine proteases, Metalloproteases, Aspartic proteases	25
Benzamidine Hydrochloride	Trypsin, Thrombin	26
Glu-Gly-Arg-Chloromethyl ketone (GGACK)	Amylase, Factor Xa, VIIa, Urokinase, Trypsin	27
Waldelactone	Caspase 11	28
Vinyl Sulfones	Cysteine Protease	29

1.3.2 Based on their specificity: Protease inhibitors have been classified into four families based on their specificity for the four mechanistic classes of proteolytic enzymes (19). A

particular inhibitory reactive site usually inhibits the protease belonging to one of the four mechanistic classes. Proteases belonging to the category of serine and cysteine type display strong covalent interactions with the substrate proteins they attack, whereas aspartic and metalloproteases do not show such strong affinity towards substrates. Serine and cysteine proteases have strong nucleophile in their reactive site. The nucleophile attacks the peptide bond in such a way that the carboxyl end of respective amino acid from peptide backbone of the substrate protein. Aspartic proteases, lack any such nucleophile (30). By the principle of acid/base catalysis and with the assistance of water molecule in reactive hole, they perform the function of proteolysis. Metalloproteases mimic the mechanism of Aspartic proteases but include the metal ion in the reaction (31). Majority of PIs are specific for serine proteases, the most widespread and most extensively studied enzymes. Many serpins are multi-headed type as a result of internal gene duplication and fusion (32). The known serpins are not all homologous suggesting that they must have arisen by convergent rather than divergent evolution. The classification based on sequence and the structure includes more than 15 families (14).

1.3.3 Based on sequence and structural features: InterPro database listed about 10 families of PIs. Classification is based on structural differences observed in the different inhibitors (33). Serpin is the most versatile family of inhibitors found in nature (34). The peptides of Squash type trypsin inhibitors are usually of about 30 amino acids long. It has a peculiar structure called “Knottin” or a cysteine knot made of three conserved disulphide bonds (35). The inhibitors of Mustard family contain 60-63 amino acid residues. Apparently the protein family is restricted to cruciferae family. Mustard PIs are rich in glycine and cysteine (36). Pin I and Pin II inhibitors are found in solanaceae family. The inhibitor has about 3-4 small stretches within peptide that are responsible for inhibition of the serine protease known as IRDs (Inhibitory repeat domains) (37). Bowman Birk inhibitors are small peptides of about 6kDa molecular weight with two inhibitory loops (38). The cysteine inhibitors are 20-22kDa proteins are rich in glycine and serine amino acids and are found in gramineae family. They are known to be potent antifungal agents (39). Carboxy peptidase inhibitors from tomato and potato are known to accumulate in leaves on wounding. The inhibitors were localized in spaces of vacuolar compartments (40). Out of ten inhibitor families from InterPro, except metalloprotease inhibitor and cysteine proteinase inhibitor families, all other

families represent mainly serine protease inhibitors. Considering the volume of work done on plant PIs, Bowman Birk type and the Kunitz type inhibitors are the families of extensively studied for their structural, functional and applied aspects compared to other families.

Table 2: PI families listed in InterPro database

PI family	Plant PI code	IntePro Id
Bowman Birk serine proteinase inhibitor	BBI	IPR000877
Cereal trypsin/ amylase inhibitor	BRI	IPR001768
Cysteine proteinase inhibitor	CYS	IPR000737
Metallocoxy peptidase inhibitor	MCI	IPR000737
Mustard trypsin inhibitor	MSI	IPR000010
Potato type I inhibitor	PI1	Not available
Potato type II protease inhibitor	PI2	Not available
Serpin	SPI	IPR000215
Soybean trypsin inhibitors (Kunitz)	KNI	IPR002160
Squash inhibitors	SQI	IPR000737

Among the proteases found in nature serine proteases dominate over all other classes of proteases. Majority of the microbial proteases identified are serine proteases (41). Serine proteases are classified by their substrate specificity, particularly by the type of residue found at P1 position of their substrate, as trypsin-like (positively charged residues Lys/Arg), elastase-like (small hydrophobic residues Ala/Val), and chymotrypsin-like (large hydrophobic residues Phe/Tyr/Leu) (42). The active site of trypsin like enzymes consist of a catalytic triad of Ser195, His57, and Asp102 residues (chymotrypsin numbering system) and a pocket in an enzyme that can stabilize an alkoxide i.e., an oxyanion hole (43). Serine proteases fall into two broad categories based on their structure: chymotrypsin-like (trypsin-like) or subtilisin-like. In humans, they are responsible for coordinating various physiological functions, including digestion, immune response, blood coagulation and reproduction (44).

1.4 Kunitz type trypsin inhibitors (KTIs)

In 1945, the first KTI was reported from Soybean by M. Kunitz. It was about 20kDa protein with trypsin inhibitory potential (45). The inhibitor gained popularity due to its unique properties. Soon various similar inhibitor sequences were identified from various plant and animal sources. These inhibitors are also termed as STI type since several isoforms were identified from Soybean. Prior to the discovery of STI, most of the inhibitors known from plant sources were small peptides of 20-30 amino acids and were rich in glycine while the KTIs contained about 170 to 200 amino acids. But STI type inhibitors gained special attention because they were abundant in legumes with multiple variants and peculiar characteristics listed below (46):

- Molecular weight about 18–24kDa
- Single polypeptide chain: single inhibitory loop
- 4-5 conserved cysteine with 2 disulphide bonds
- Type of proteases inhibited: serine, cysteine and aspartic proteinases
- No inhibition of metalloproteinase
- Quaternary structure containing a β -trefoil fold

These protein families are described in the Pfam database (47). It includes the annotations of sequences listed different protein families based on the domains it contained. The Kunitz inhibitors in plants are clubbed under Kunitz legume family in Pfam database with Pfam ID: PF00197. About 756 sequences in Pfam are distributed on 53 plant species. Plant KTIs are clubbed together under I3 families in the MEROPS database (48).

1.4.1 Kunitz type inhibitors in animals: KTIs found in animals are quite different from that of their counterparts in plants, though they can inhibit trypsin and/or chymotrypsin. Considering the length of polypeptide, the animal KTIs are usually much smaller than that from plants. These KTIs are about 50-60 long amino acid chains with inhibitory potential towards trypsin and chymotrypsin (49). In MEROPS database family I2 includes all animal KTIs. Like STI in plants, BPTI in animals represents canonical animal KTIs. In Pfam this family is known as Kunitz BPTI family (PF00014). A Kunitz inhibitor from a spider *Araneus ventricosus* contains 51 amino acids (50). But there are some exceptions like a KTI from

Eastern brown snake contains 83 amino acids (51). Apart from inhibition of serine proteases, animal KTIs are known to perform several functions which are linked to various physiological processes, such as ion channel blocking, blood coagulation, fibrinolysis, and inflammation. *Araneus ventricosus* Kunitz inhibitor is involved in K⁺ channel blocking (52).

1.5 Mechanism of protease inhibition

In order to bring about the inhibition of any protease, the inhibitor must bind to the protease in such a way that the substrate protein cannot bind the reaction center. The inhibitor could directly bind the active site of protease or it could bind somewhere else to the protease but it could bring about the conformational changes in a protease that the substrate protein may not bind to protease. It has been observed that in case of almost all protein inhibitors, the inhibitor binds directly to reactive pocket of a protease in competing with the substrate proteins for same binding site (53). It suggests that in order to compete with substrate, the inhibitor molecule must mimic the substrate and form a strong binding with the protease, probably; stronger than the substrate. There were two mechanisms proposed for binding of the inhibitor to the protease active site. One hypothesis assumed that the when an inhibitor binds to the protease, it is partially cleaved by the protease. The inhibitor is cleaved into two parts generating new N and C termini. Since, a part of the inhibitor remains bound to the active pocket of the enzyme, further proteolysis of the substrate proteins is prevented. This mechanism was proposed to be operational in *Cucurbita maxima* trypsin inhibitor (54)

According to another hypothesis, the inhibitor binds to the protease but is not cleaved into two parts. The inhibitor remains bound in active pocket protecting other proteins from getting cleaved. The second mechanism was observed in most of the inhibitors. Usually the inhibitor has same residue in its inhibitory loop as the residue in substrate proteins on which the protease attacks. This explains the binding of the inhibitor to the protease in the same way as the substrate (55).

Inhibitors belonging to different families may have mechanistic differences involving different residues to achieve inhibition in a protease specific manner. Usually the underlying mechanism of proteases inhibition remains the same. Consider the mechanism of inhibition of trypsin. The inhibitory loop from the inhibitor binds to catalytic pocket just like the

substrate protein binds to the trypsin during cleavage. Trypsin cleaves at carboxyl end of Lys/Arg residue in substrate protein unless followed by Pro (56). In case of Bowman Birk inhibitor (e.g. PsTI from *Pisum sativum*), Lys residue neutralizes the Ser195 from the trypsin (57). If Pin II inhibitors are considered, (e.g. *Solanum lycopersicum* inhibitor II), Arg residue at P1 position in loop 1 blocks nucleophilic attack of Ser195 and the network of hydrogen bonds with trypsin active pocket and within the loop achieve the overall inhibition (58). In a canonical KTI like STI, Ser195 of trypsin is tackled by Arg63 residue in the inhibitory loop (59).

1.5.1 Mode of action of inhibitor: *If inhibitor binds in substrate like manner, then, why it is not digested by the protease?* It has long been believed that inhibitors bind to enzymes in the manner of a good substrate, the current problem concerning protease inhibitor-protease inhibition is thus not to explain why turnover of the inhibitors is so slow (60).

Radisky E. S. et al. proposed the explanation to this question with structural analysis of inhibitors (61). It was proposed that the catalytic triad of trypsin hydrolyzes the peptide backbone of inhibitor at the Arg/Lys residue in the inhibitory loop just like it does in case of substrate protein. A Michaelis complex is formed which is converted to an acyl-enzyme complex. At this stage enzyme hydrolyzes the inhibitor at P1 residue; but the rate of hydrolysis is extremely slow. The reaction appears to be static. The inhibitor remains attached to trypsin active pocket and the resultant acylenzyme complex appears to be stable. The cleaved backbone is eventually re-ligated. Thus, trypsin cannot attack other proteins.

1.6 Applications of inhibitors

As the family name suggests, KTIs are mostly and abundantly found in legumes; mainly in seed tissue (62). Thus it was initially believed that KTIs must be storage proteins. Later studies highlighted that soybean trypsin inhibitor genes were expressed in seeds, still the inhibitor was seen to be viable even after seed germination and during embryonic development; questioning its function as a storage protein (63). After the discovery of their ability to inhibit serine proteases, it was postulated that the KTIs must be regulating the plant endogenous proteases (64). But there are very scarce reports where direct relation between a

KTI and an endogenous plant protease is described. Thus exact biological role of these KTIs still remain under the wrap.

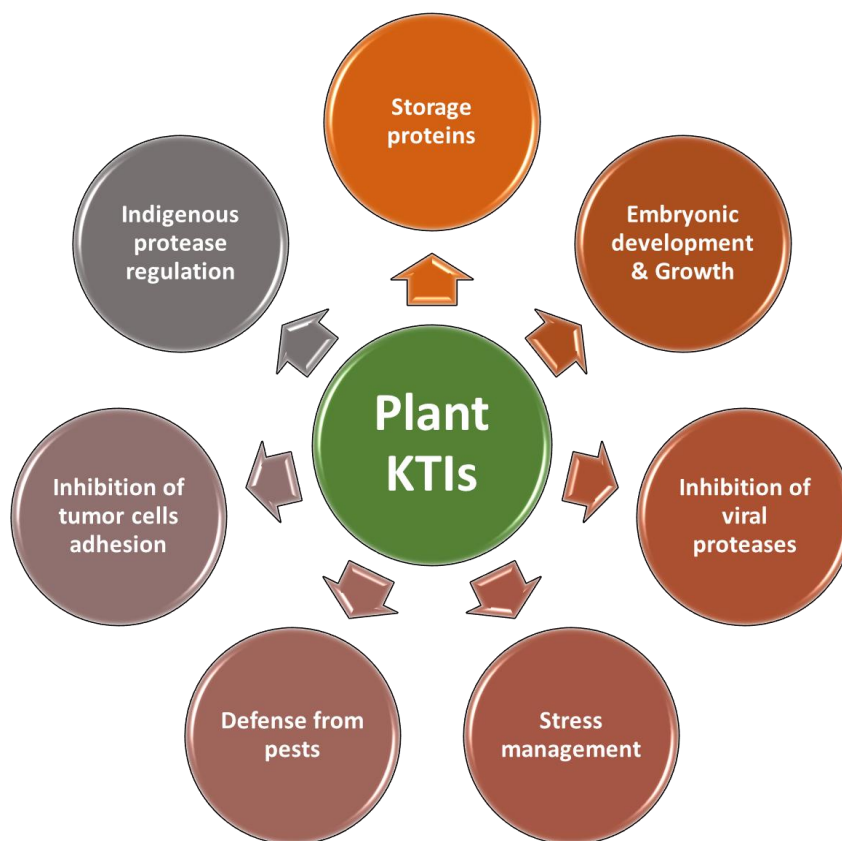


Figure 3: Diverse applications and functions of KTIs: Based on these basic postulated functions, novel applications could be discovered.

However, the role of KTIs in their assistance in defense against invertebrate pests and microbial pathogens has been very well established. The insect gut proteases are rich in serine proteases. About 95% of the total proteases of mid-gut comprises of trypsin and trypsin like proteases (65). The protein content in diet directly controls the secretion of proteases in insect gut. On ingestion of effective concentrations of PIs, the insects cannot digest the dietary proteins. It causes serious malnourishment to pest insect and results in larval death. This fact formed the basis of several studies where insects were fed with artificial diet containing one or cocktail of PIs. It resulted in reduced number of insect larvae attacking the plant. Bringing in further advancement in this research, transgenic crops were

raised which would express recombinant inhibitors in their leaves. In one such study tobacco plants were made express cowpea trypsin inhibitor (66). It was observed that due to expression of foreign inhibitor gene, the crop developed resistance against *Heliothis zea* larvae. Unfortunately in certain cases, it was noticed that the larvae developed resistance against PIs and modulated their own proteases for sustainable survival (67).

To overcome this hurdle, one can engineer the inhibitor and design more potent, versatile inhibitors. To perform such experiments, one should have structural details of the inhibitors at hand. Some reports hint towards the involvement of KTIs in stress management. In white clover plants, it was observed that the Kunitz inhibitor gene expression was differentially regulated during biotic stress (68). During water stress, intrinsic proteolytic activity increases. The expression of two KTIs from white clover; Tr-KPI1 and Tr-KPI2 was increased when exposed to water stress (69). Plant KTIs were also explored for their putative applications in medicine. A KTI from Chinese black soybean was reported to be effective against certain types of cancers. It was shown that it could prevent the cell proliferation in MCF-7 breast cancer and HepG2 hepatoma (70). BrTI, an inhibitor from *B. rufa* was reported to inhibit tumor cells adhesion to extracellular matrix glycoproteins (71). Serine proteases of viruses have been target for drug design and discovery. Thiazolidine derivatives are known to inhibit NS3/4A serine protease, an integral protease from Hepatitis C virus. Plant KTIs could be used instead to inhibit viral proteases and thus could be used to develop a new remedy for viruses like Hepatitis C, HIV, Herpes, and CMV (72).

1.7 Need for detailed and systematic study

The speculated function of KTIs is defense from foreign proteases and growth regulation during wounding stress. But their exact physiological role remains covered. Several questions remain unanswered even after discovery and identification of hundreds of KTIs. How these KTIs respond to plant's own proteases is not yet explored. Most of the studies concentrate on the activity determination against different serine proteases. For some KTIs structural information is available but gene level expression and biophysics of the protein is not studied. For some KTIs claimed to be more potent than canonical KTIs but are not further investigated. STI is the only KTI which is characterized for structural, biochemical, biophysical aspects. But it was observed that it contained allergenic domain due to which

limited it for further clinical studies (73). Under such scenario it becomes necessary that known KTIs should be thoroughly assessed for their stability, gene expression level under different conditions in different plant organs, systematic interactions with other proteins, and biophysical characters under different sets of conditions. In order to develop any molecule into novel candidate for commercial product whether it is a pest control agent or a drug, above mentioned parameters must be well studied.

1.8 About protein crystallography

As mentioned earlier, plant protease inhibitors are studied as model systems to understand various types of PPIs. Omics studies help to deduce probable candidates involved in validating and establishing actual and exact type of interactions involved in binding partners of any interactome.

Today there are about 86 structures of KTIs are available, including structures solved at different pH conditions, mutants, different resolutions. However, among these there are only three complex structures available. More and more structural information is required to validate proposed mechanisms and hypotheses since such information serves as direct evidence.

A protein structure can be deduced by NMR spectroscopy, Cryo electron microscopy or macromolecular crystallography (74). NMR spectroscopy though provides in solution structure, it cannot be used for structure determination of about 20kDa proteins. Cryo-electron microscopy is still in its infancy. It is used for high molecular weight protein complexes and it still needs to reach the high resolution capabilities like X-ray crystallography. Under such conditions, X-ray crystallography proves to best approach for structural investigations of KTIs.

In present thesis we have explored one of the KTIs from the economically important crop Chickpea (*Cicer arietinum L.*). The inhibitor was reported to be efficient in inhibition of bovine trypsin. Despite being unique in amino acid sequence; it while retaining its inhibitory capabilities not only arrests the growth of insect pest but also proven to be antifungal in nature. We have determined its crystal structure in this study. We have also tried to associate its biophysical properties with its biochemical activity. Finally, attempts were made to

understand its interactions with commercially available bovine pancreatic trypsin using available structural information and bioinformatics techniques.

Section 2: Comprehensive analysis of KTI structures

To understand the PPIs, one must first understand the basic structural properties of individual protein members of the PPI interaction system. There are several fundamental parameters that could be used to understand, analyze and evaluate the protein structures. Several features like amino acid composition, secondary structure elements, and a range of attractive/repulsive forces influence the overall 3D architecture of any protein. Some of these factors are discussed in the sections below, with respect to Kunitz legume family.

In RCSB PDB about 86 different structures of plant proteins categorized as a Kunitz type inhibitor are available (Table 3). Among these are few unique structural entities, and also several mutants, non-inhibitor proteins, enzyme-inhibitor complexes and structures at different pH conditions. Some of the structures have been selected to understand the peculiar features that are responsible for the characteristic properties of KTIs.^{#1} (Ref for selected PDBs: 75-86)

2.1 Materials and methods: The crystal structures were retrieved from RCSB-PDB in ‘.pdb’ format. They were analyzed for different structural parameters like inter-chain/intra-chain interactions using iRDP server (integrated Rational Design of Proteins) (87). The structures were visualized using PyMOL 2.0 and CCP4mg (88, 89).

Table 3: List of structures available in RCSB-PDB[§] categorized under plant Kunitz type inhibitor

PDB	Protein	Source	Resolution (Å)	Space group	Year
5DSS	MP4	<i>Mucuna pruriens</i>	2.8	P 2 ₁ 2 ₁ 2	2016
1WBA	WINGED BEAN ALBUMIN 1	<i>Psophocarpus tetragonolobus</i>	1.8	I 4 ₁ 2 2	1987

^{#1} Due to poor refinement parameters, some structures could not be used for certain analyses

[§] Data available up to December 2017

5DZU	cathepsin D inhibitor	<i>Solanum tuberosum</i>	2.12	C 1 2 1	2015
2GZB	BbCI	<i>Bauhinia bauhinioides</i>	1.7	P 1 2 ₁ 1	2007
4ZOT	BbKI	<i>Bauhinia bauhinioides</i>	1.4	P 2 ₁ 2 ₁ 2 ₁	2015
3IIR	Miraculin like protein	<i>Murraya koenigii</i>	2.9	P 4 ₃ 2 ₁ 2	2009
3ZC8	Miraculin like protein	<i>Murraya koenigii</i>	2.24	C 1 2 1	2012
3ZC9	Miraculin like protein	<i>Murraya koenigii</i>	2.2	C 1 2 1	2013
1R8N	DrTI	<i>Delonix regia</i>	1.75	P 2 ₁ 2 ₁ 2 ₁	2003
1R8O	CTI	<i>Copaifera langsdorffii</i>	1.83	P 4 ₃ 2 ₁ 2	2004
4AN6	TKI	<i>Tamarindus indica</i>	1.94	P 2 ₁ 2 ₁ 2 ₁	2012
4AN7	TKI porcine trypsin complex	<i>Tamarindus indica, Sus scrofa</i>	2.23	P 2 ₁ 2 2 ₁	2012
1AVA	AMY2/BASI complex	<i>Hordeum vulgare</i>	1.9	P 2 ₁ 2 ₁ 2 ₁	1997
2IWT	Thioredoxin h2-BASI complex	<i>Hordeum vulgare subsp. vulgare, Hordeum vulgare</i>	2.3	P 4 ₁ 2 ₁ 2	2006
3BX1	Barley alpha-Amylase/Subtilisin Inhibitor (BASI)	<i>Hordeum vulgare, Bacillus lentus</i>	1.85	P 4 ₁ 2 ₁ 2	2008
2QN4	BASI	<i>Oryza sativa subsp. japonica</i>	1.8	P 2 ₁ 2 ₁ 2	2007
1EYL	WCI	<i>Psophocarpus tetragonolobus</i>	1.9	P 6 ₁ 2 2	2000

1FMZ	WCI mutant N14K	<i>Psophocarpus tetragonolobus</i>	2.05	P 6 ₁ 2 2	2000
1FN0	WCI mutant N14D	<i>Psophocarpus tetragonolobus</i>	2	P 6 ₁ 2 2	2000
1WBC	WCI	<i>Psophocarpus tetragonolobus</i>	2.95	P 6 ₁ 2 2	1995
1XG6	WCI P1 mutant	<i>Psophocarpus tetragonolobus</i>	2.15	P 6 ₅	2004
2BEA	WCI mutant N14G	<i>Psophocarpus tetragonolobus</i>	2.35	P 1	2005
2BEB	WCI mutant N14T	<i>Psophocarpus tetragonolobus</i>	2.81	P 2 ₁ 2 ₁ 2	2005
2ESU	WCI mutant N14Q	<i>Psophocarpus tetragonolobus</i>	1.94	P 6 ₁ 2 2	2005
2ET2	WCI mutant N14A	<i>Psophocarpus tetragonolobus</i>	2.1	P 6 ₁ 2 2	2006
2QYI	Engineered inhibitor and bovine trypsin binary complex	<i>Psophocarpus tetragonolobus</i> , <i>Bos taurus</i>	2.6	C 1 2 1	2007
2WBC	WCI	<i>Psophocarpus tetragonolobus</i>	2.3	P 6 ₁ 2 2	1997
3I29	Mutant WCI and trypsin	<i>Psophocarpus tetragonolobus</i> , <i>Bos taurus</i>	2.4	C 1 2 1	2009
3I2A	Chimeric STI- WCI	<i>Psophocarpus tetragonolobus</i>	2.3	I 4	2010
3I2X	Chimeric ETI- WCI	<i>Psophocarpus tetragonolobus</i>	2.85	C 1 2 1	2010
3QYD	Chimeric trypsin inhibitor	<i>Psophocarpus tetragonolobus</i>	2.97	P 1 2 ₁ 1	2011

4H9W	M mutant of WCI	<i>Psophocarpus tetragonolobus</i>	2.5	P 6 ₁ 2 2	2014
4HA2	A91 mutant of WCI	<i>Psophocarpus tetragonolobus</i>	2.9	P 1 2 ₁ 1	2014
4TLP	A91 mutant of WCI	<i>Psophocarpus tetragonolobus</i>	1.9	P 2 ₁ 2 ₁ 2 ₁	2014
4WBC	WCI	<i>Psophocarpus tetragonolobus</i>	2.13	P 6 ₁ 2 2	1999
1TIE	ETI	<i>Erythrina caffra</i>	2.5	P 6 ₅ 2 2	1992
1AVU	STI	<i>Glycine max</i>	2.3	P 2 ₁ 2 ₁ 2 ₁	1998
1AVW	STI and porcine trypsin complex	<i>Glycine max, Sus scrofa</i>	1.75	P 2 ₁ 2 ₁ 2 ₁	1998
1AVX	STI and porcine trypsin complex	<i>Glycine max, Sus scrofa</i>	1.9	P 4 ₁ 2 ₁ 2	1998
1BA7	STI	<i>Glycine max</i>	2.5	P 1	1998
4J2K	EcTI	<i>Enterolobium contortisiliquum</i>	1.75	C 1 2 1	2013
4J2Y	EcTI bovine trypsin complex	<i>Bos taurus, Enterolobium contortisiliquum</i>	2	C 1 2 1	2013
4IHZ	CrataBL	<i>Crateva tapia</i>	1.5	C 1 2 1	2013
4II0	CrataBL	<i>Crateva tapia</i>	1.75	C 1 2 1	2013
3S8J	PPI	<i>Carica papaya</i>	2.6	P 3 ₁	2011
3S8K	PPI	<i>Carica papaya</i>	1.7	P 3 ₁	2011
5FNW	Potato STI	<i>Solanum tuberosum</i>	2.45	P 4 ₃ 2 2	2016
5FNX	Potato STI	<i>Solanum tuberosum</i>	2.65	P 4 ₃ 2 2	2016
5FZU	N19D POTATO ST	<i>Solanum tuberosum</i>	2.43	P 4 ₃ 2 2	2016
5FZY	Potato STI	<i>Solanum tuberosum</i>	2.47	C 2 2 2 ₁	2016
5FZZ	Potato STI	<i>Solanum tuberosum</i>	2.55	P 2 2 ₁ 2 ₁	2016
5G00	Potato STI	<i>Solanum tuberosum</i>	2.5	P 4 ₃ 2 2	2016

2DRE	Water-soluble chlorophyll protein	<i>Lepidium virginicum</i>	2	P 2 ₁ 2 ₁ 2 ₁	2006
2GO2	BbKI	<i>Bauhinia bauhinioides</i>	1.87	P 2 ₁ 2 ₁ 2 ₁	2007
3E8L	API-A bovine trypsin complex	<i>Bos taurus</i> , <i>Sagittaria sagittifolia</i>	2.48	C 2 2 2 ₁	2009
5HPZ	Water-soluble chlorophyll protein	<i>Brassica oleracea var. viridis</i>	1.96	P 2 2 ₁ 2 ₁	2016
3TC2	Trypsin inhibitor	<i>Solanum tuberosum</i>	1.6	P 1 2 ₁ 1	2012

2.2 Amino acid distribution: Only twenty amino acids (with certain modified amino acids) are available in the nature but with their different combinations, innumerable different proteins are formed. Thus, the placement of amino acids in the peptide as well as overall amino acid composition of the protein is responsible for the global and local structure of any protein. This in turn can control the function and properties of protein molecules. Hence, it becomes inevitable to understand the amino acid composition of the protein. Certain family of proteins called glycine rich proteins is the family of defense proteins in plants. It holds its characteristic name because the candidates of this family are rich in amino acid glycine. From the following chart (Figure 4) we can conclude that polar amino acids contribute more to the overall composition as compared to aromatic and charged amino acids. However, this trend is not observed in *Enterolobium contortisiliquum* Trypsin inhibitor (EcTI). In EcTI, negatively charged amino acids dominate over polar amino acid composition.

2.3 Size and shape: In case of protease inhibitors, the inhibitory regions complement the enzyme's reaction center. In simple words, they complement each other in a lock and key mechanism (90). So the interface area, shape and dimensions of the cavity or channel of a protein structure can be easily calculated in terms of simple dimension unit 'Å' and can be used for structural comparisons. Solvent accessible area or solvent accessible surface area (ASA) and the gap index are some of these parameters that could be used to assess different structures for their size and shape (91).

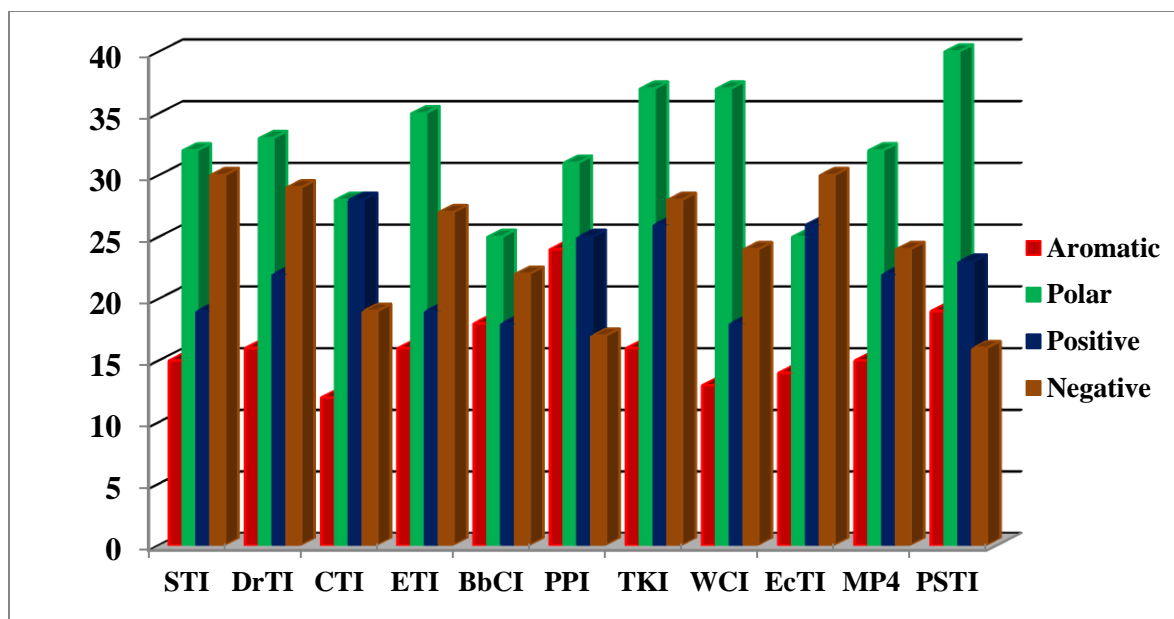


Figure 4: Composition of different types of amino acids in selected KTI structures

When a protein in its native 3D conformation structure is dipped in water, the sum of the areas where water molecules form contact with the structure can be termed crudely as solvent accessible area. This imagination was mathematically brought into reality by two scientists Shrake and Rupley in 1973. It was called ‘rolling ball algorithm’ (92). This algorithm uses a sphere (of solvent) of a particular radius to ‘probe’ the surface of the molecule. A probe of given radius is rolled around the surface of the protein, and the path traced out by its center is the accessible surface. Typically, the probe has the same radius as water (1.4Å) and hence the surface described is often referred to as the solvent accessible surface. The KTI structures contain β -strands that are surrounded by loops from all sides making the loops highly solvent accessible. About 70% of total solvent accessible area comes from looped structures. Residues present in active site are exposed to solvent. In all experimentally determined structures, the inhibitory loop appears to be solvent accessible.

2.4 Fold and secondary structure elements: The amino acid sequence in a polypeptide gives us the primary structure of a protein. The primary structure organizes itself into different patterns giving rise to secondary structure elements. A list of different type of

secondary structure elements present in a protein, and which dictate formation of the tertiary structure, is given below.

1. 3_{10} helix (G): [i+3] hydrogen bond
2. α helix (H): [i+4] hydrogen bond
3. π helix (I): [i+5] hydrogen bond
4. beta bridge (B)
5. beta bulges (E)
6. Turn (T)
7. Bend (a region of high curvature) (S)
8. Random coil (C)

These eight types are clubbed into three classes, helix (G, H and I), strand (E and B) and loop (C, T and S). This segregation of secondary structure components into three classes was based on an algorithm ‘Define Secondary Structure of protein’- abbreviated as DSSP-, originally proposed by Kabsch and Sander in 1983 for a program called ‘Pasc1’. DSSP was based on intra-backbone hydrogen bonds of a protein (93). Table 4 describes the percent secondary structure character in overall secondary structure composition of a given KTI. The values in the table clearly indicate that a dominant part of a KTI structure is composed of β bridges, bulges and loops. The structure comprises of about an average of 43% β structures, 55% loops and only 2% helices. More than 50% of the structural part is made up of loops or unordered structures. This indicates the dynamic nature of these proteins.

Table 4: Percent secondary structure composition of KTIs as calculated by iRDP server

Secondary structure	β structure		Helix			Loop		
	B	E	G	H	I	S	C	T
STI	3.31	37.02	1.66	0	0	9.39	34.25	10.5
DrTI	5.95	31.35	1.62	0	0	21.62	29.19	10.27
CTI	1.8	35.93	0	0	0	17.96	32.34	11.98
ETI	2.33	37.79	4.07	0	0	13.37	27.33	12.21

BbCI	5.42	42.77	0	0	0	17.47	21.69	11.45
PPI	4.89	33.7	2.72	0	0	19.02	30.98	7.61
TKI	4.32	38.38	0	0	0	14.05	27.03	9.73
WCI	4.37	34.97	3.28	0	0	15.3	32.79	9.29
EcTI	3.41	35.8	1.7	0	0	10.8	29.55	14.77

2.5 Disulphide bond: Disulphide bonds are formed between cysteine residues present in peptide chains. Cysteine (Cys) residues are structurally, functionally and evolutionarily very important (94). Disulphide bonds could be either inter-chain or intra-chain. These are helpful in maintaining the 3D structure of a protein and protein folding. Cys being hydrophilic in nature, it is usually found surrounded by hydrophobic residues. It helps in locating the hydrophobic patches in protein.

Usually KTIs have 4 to 5 conserved Cys residues. These are responsible for formation of two disulfide bonds within a single peptide. Table shows the inhibitors with respective positions of Cys residues. However, exceptions are known. Some inhibitors do not contain any Cys residue but still can maintain their Kunitz architecture. BbCI does not contain any Cys residue on the other hand CTI has only two Cys residues (77, 79). It is believed that the diversion of proteins from higher number of Cys residues to lower is a sign of evolutionary diversion from ancestral Kunitz type precursors (95). Structural integrity and biological activity of BbKI and BbCI are maintained by H-bonds since Cys-Cys bridges are missing.

Table 5: Cysteine content in respective KTI molecules

Protein	PDB Id	Cysteine content	Disulphide Bonds
STI	1AVU	4	2
DrTI	1R8N	4	2
CTI	1R8O	2	1
ETI	1TIE	4	2
BbCI	2GZB	0	0

PPI	3S8J	4	2
TKI	4AN6	4	2
WCI	2WBC	4	2
EcTI	4J2K	4	2
PSTI	3TC2	4	2
MP4	5DSS	4	2

2.6 Interactions within amino acid side chains: The tertiary and the quaternary structure of the protein arise due to folding of a primary structure into secondary structural elements and their particular orientation in 3D space. What is responsible for the folding of a primary structure into a well-defined characteristic 3D functional structure of a protein? There are several non-covalent but prominent and highly effective interactions occurring within the residues of the polypeptide. These interactions may bring distant amino acid residues closer and may repel certain side chains, finally defining the position and orientation of each amino acid in the peptide forming an absolute protein structure. Following are the different types of interactions observed in a protein structure:

1. Hydrogen bond network
2. Aromatic-aromatic interactions
3. Cation- π interaction
4. Aromatic-sulfur interactions
5. Salt bridges

All the above interactions are not necessarily seen in every protein. However, in most cases, hydrogen bond networks are the most common and influential interaction. These are amongst the predominant interactions observed within same or between different peptide chains (96).

2.7 Aromatic-aromatic interactions: Networks formed by aromatic amino acids interacting with each other in a peptide chain provide additional stability to the structure. Such interactions are also seen in KTIs. Aromatic interactions are not strong but effective enough to restrict the movement of the peptide backbone by forming a network. In STI, Phe2 residue interacts with the Phe66 across a distance of 5.7Å and in case of BbCI, active site residue

Phe58 shows aromatic interactions with a distant Tyr18. The distance between their centroids is 6.16Å. These individual aromatic interactions can prove crucial in stabilizing the active site loop in STI and BbCI. DrTI, CTI, PPI and TKI also exhibit network of aromatic interactions. In DrTI, a network of aromatic interactions is observed between Phe63, Tyr134, Phe174 and Tyr20. Phe63 interacts with Phe174 (5.81Å) and Tyr20 (6.54Å), while Phe174 interacts with Tyr134 (5.44Å), as depicted in figure 7. In case of CTI Phe59 interacts with residue Tyr17 (5.93Å) of chain A and Phe163 (5.79Å) of chain B. PPI has network in which there is interaction between Phe62 and 3 aromatic amino acids Phe139, Phe179, Tyr20 with centroid distances 6.75Å, 5.51Å, 6.3Å respectively. Similar to PPI, CTI and DrTI, Phe61 of TKI shows interactions with Phe171 (5.49Å) and Tyr19 (6.06Å). Interestingly, in all these networks Phe, Tyr residues are conserved. This might be an indication that DrTI, CTI, PPI and TKI could have common structural ancestors.

2.8 Cation- π interaction: The interactions between a cation and a plane of electron rich system like a cyclic carbon ring, such as a benzene ring, are known as cation- π interactions. These interactions are strong but non-covalent in nature. In proteins, side chains of the aromatic amino acids provide a surface of negative electrostatic potential that can bind to a wide range of cations through a predominantly electrostatic interaction. Usually Lys or Arg act as the cation, but the frequency of Arg as a cation is higher compared to Lys. The findings of Dennis A. Dougherty et al. suggest that there is at least one Cation- π interaction for every 77 amino acids in a protein (97). Over 25% Trp residues are involved in Cation- π interactions. In enzyme catalyzed reactions, these interactions play crucial roles. In the transition state, these interactions stabilize the buildup of positive charges.

STI, DrTI, CTI, ETI, BbCI, WCI and PPI exhibit cation- π interactions in or near (within range of 2 residues) the active site. In STI, Arg63 and Arg65 act as a cation while side chains of Tyr62 and Phe66 act as a π system respectively. DrTI and TKI mimic this system of STI. Arg60-Phe61 forms the cation- π system in TKI. In DrTI cationic Arg62 interacts with Phe63. Tyr74 serves as a system for both Lys5 and Lys72. The active site of CTI has got 2 cationic residues, Arg64 and Lys60, which are both served by the side chain of Phe112. Lys66 of PPI has interactions with the aromatic ring outside the active chain (Tyr145).

EcTI, BbCI and WCI do not exhibit any cation- π interactions within active site loop. It would be interesting to note that active site of WCI has got neither cation- π interaction nor aromatic-aromatic interactions within or nearby its active site.

2.9 Active site loop: The trypsin active site, due to its unique narrow and deep pocket, interacts strongly with basic Lys and Arg side chains of substrates or inhibitors. The contribution of P1 residue is important for enzyme-inhibitor interaction in the region involved in the binding. The amino acid composition of inhibitory loops of respective KTIs has been summarized in Table 6.

Table 6: Amino acid composition of inhibitory loops of KTIs with their respective P1 residues

Protein	PDB ID	P1 residue	Reactive loop composition
STI	1AVU	Arg-65	Ser60-Pro61-Tyr62-Arg63-Ile64-Arg65-Phe66-Ile67-Ala68
DrTI	1R8N	Lys-69	Ser65-Pro66-Glu67-Glu68-Lys69-Gln70-Gly71-Lys72-Ile73
CTI	1R8O	Arg64	Ala61-Ser62-Pro63-Arg64-Ser65-Lys66-Tyr67-Ile68-Ser69
ETI	1TIE	Arg63	Ser60-Arg61-Leu62-Arg63-Ser64-Ala65-Phe66-Ile67-Pro68
BbCI	2GZB	Ala63	Thr60-Pro61-Leu62-Ala63-Ile64-Ala65-Ile66-Ile67
PPI	3S8J	Val68	Ala64-Ile65-Lys66-Asn67-Val68-Lys69-Asp70-Asn71
TKI	4AN6	Arg66	Ser63-Arg64-Ala65-Arg66-Ile67-Ser68-His69
WCI	2WBC	Loop1: Leu65	Loop1: Gln63-Phe64-Leu-65-Ser66-Leu67-Phe68-Ile69 Loop2: Asn38-Glu39-Pro40-Cys41-Pro42-Leu43
EcTI	4J2K	Arg64	Thr61-Pro62-pro63-Arg64-Ile65-Ala66-Ile67-Leu68
PSTI	3TC2	Loop1: Phe75	Se71-Ser72-Ser73-His74-Phe75-Gly76-Gln77-Ile78

The KTIs can be grouped in two categories; KTIs with Arg at P1 position and KTIs without Arg at P1 position. When these inhibitors are aligned (Figure 5) it can be noticed that the KTIs with P1-Arg have similar topology and geometry of the inhibitory loop. On the other hand other KTIs exhibit variations in loop geometry. The reactive site of Kunitz trypsin inhibitors usually has a Lys or Arg, just like the substrate proteins. The canonical KTIs from *Acacia confusa*, *Cassia obtusifolia*, *Enterolobium* and *Leucaena leucocephala* follow the same trend (98). In the case of elastase inhibitors, this position is occupied by an Ala residue (99). Some inhibitors bear a second independent reactive site, while others have distinct alterations in this region; such as the inhibitor isolated from *Cicer arietinum L.* (later described in this work) does not contain Lys/Arg but can still act as a potent inhibitor (100). KTIs with P1-Arg/Lys share similar geometry of the inhibitory loop. On the other hand other KTIs exhibit variations in loop geometry.

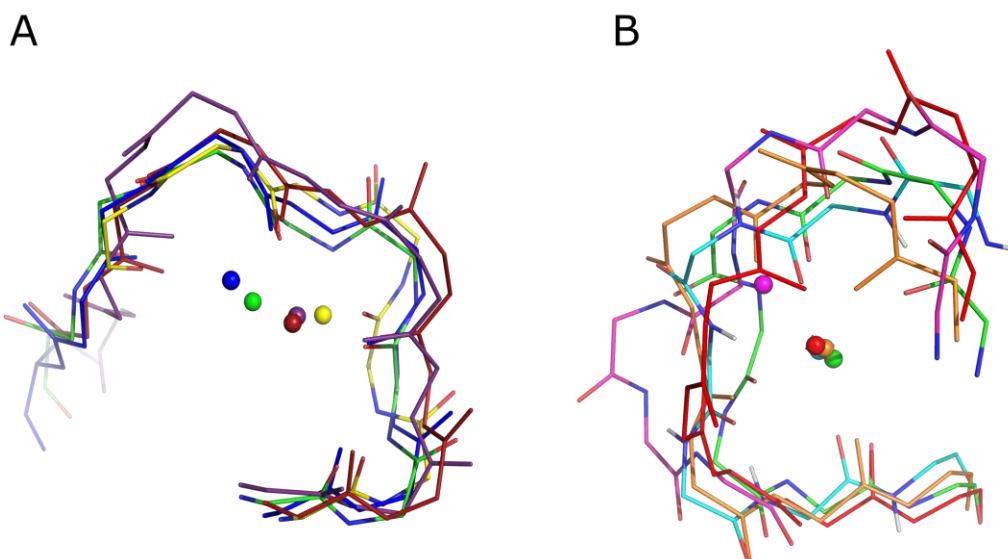


Figure 5: Line representation of backbones of inhibitory loops of KTIs aligned with loop of STI, along with respective centroids: A- KTIs with Arg residue at P1 position. Color indicators: Green- STI, Blue- CTI, Violet- ETI, Yellow- TKI, Red- EcTI. B- KTIs with residue other than Arg at P1 position. Color indicators: Green- STI, Cyan- WCI, Pink- PPI, Yellow- BbCI, Red- DrTI.

In a complex of TKI with trypsin [PDB: 4AN7], within the radius of 4Å, a strong network of hydrogen bonds can be seen between Arg66 of TKI and Ser195 of trypsin. Similar interactions were expected in case of STI complex [PDB: 1AVW]. However, it was observed that in addition to Ser195 other residues from catalytic triad of trypsin i.e., His57 and Asp102 also interact with Arg63 (Figure 6).

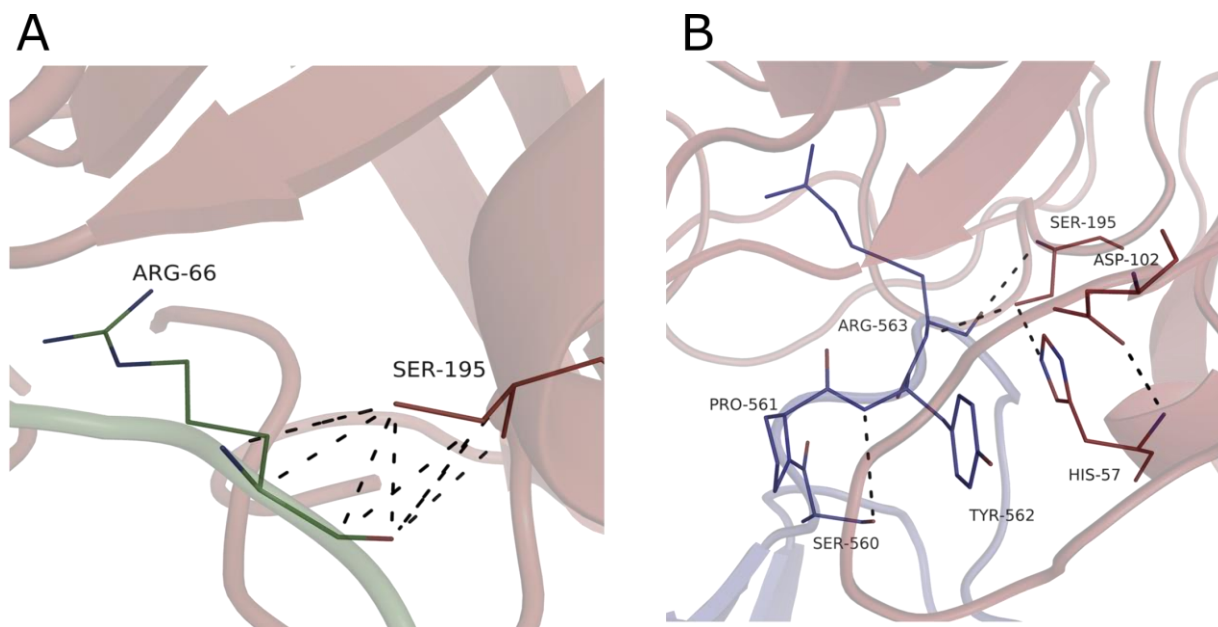


Figure 6: Polar interactions between KTI and trypsin active pocket: A- Arg66 of TKI (green) interacts with Ser195 of trypsin (red) [PDB: 4AN7]. No other polar contacts are seen within 3Å. **B-** Series of polar interactions is seen in catalytic triad of trypsin (red) and Arg63 of STI (blue) [PDB: 1AVW].

In this thesis, we have analyzed the crystal structure of CaTI2, a KTI from *Cicer arietinum L.* for its structural properties. We have compared the structure of CaTI2 to that of STI structure in order to understand the similarities and the differences with respect to canonical KTIs and correlated the structural variations to biochemical activity.

Note: A part of this chapter has been published in the following review article-

Ameya D. Bendre, Sureshkumar Ramasamy and C. G. Suresh. "Analysis of Kunitz inhibitors from plants for comprehensive structural and functional insights", *International Journal of Biological Macromolecules* (2018)113: 933-943. DOI: 10.1016/j.ijbiomac.2018.02.148

References

1. Vidal, Marc, Michael E. Cusick, and Albert-László Barabási. "Interactome networks and human disease." *Cell* 144.6 (2011): 986-998.
2. Sanchez, Catherine, et al. "Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database." *Nucleic acids research* 27.1 (1999): 89-94.
3. De Las Rivas, Javier, and Celia Fontanillo. "Protein–protein interactions essentials: key concepts to building and analyzing interactome networks." *PLoS computational biology* 6.6 (2010): e1000807.
4. Bonetta, Laura. "Protein–protein interactions: interactome under construction." *Nature* 468.7325 (2010): 851.
5. Tripathi, Prateek, et al. "The interactome of soybean GmWRKY53 using yeast 2-hybrid library screening to saturation." *Plant signaling & behavior* 10.7 (2015): e1028705.
6. Schmidt, Mirko HH, and Ivan Dikic. "The Cbl interactome and its functions." *Nature reviews Molecular cell biology* 6.12 (2005): 907.
7. Mintseris, Julian, and Zhiping Weng. "Structure, function, and evolution of transient and obligate protein–protein interactions." *Proceedings of the National Academy of Sciences* 102.31 (2005): 10930-10935.
8. Zhu, Hongbo, et al. "NOXclass: prediction of protein-protein interaction types." *BMC bioinformatics* 7.1 (2006): 27.
9. Best, Robert B., Yng-Gwei Chen, and Gerhard Hummer. "Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor." *Structure* 13.12 (2005): 1755-1763.
10. Fernandez-Recio, Juan, et al. "Optimal docking area: a new method for predicting protein–protein interaction sites." *PROTEINS: Structure, Function, and bioinformatics* 58.1 (2005): 134-143.
11. Nooren, Irene MA, and Janet M. Thornton. "Diversity of protein–protein interactions." *The EMBO journal* 22.14 (2003): 3486-3492.
12. Causier, Barry. "Studying the interactome with the yeast two-hybrid system and mass spectrometry." *Mass spectrometry reviews* 23.5 (2004): 350-367.

13. Jones, Susan, and Janet M. Thornton. "Principles of protein-protein interactions." *Proceedings of the National Academy of Sciences* 93.1 (1996): 13-20.
14. Bode, Wolfram, and Robert Huber. "Natural protein proteinase inhibitors and their interaction with proteinases." *European Journal of Biochemistry* 204.2 (1992): 433-451.
15. Laskowski Jr, Michael, and M. A. Qasim. "What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?." *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1477.1-2 (2000): 324-337.
16. Shewry, Peter R., and Rod Casey. "Seed proteins." *Seed proteins*. Springer, Dordrecht, 1999. 1-10.
17. Broadway, Roxanne M., et al. "Plant proteinase inhibitors: a defense against herbivorous insects?." *Entomologia experimentalis et applicata* 41.1 (1986): 33-38.
18. Stintzi, A., et al. "Plant 'pathogenesis-related' proteins and their role in defense against pathogens." *Biochimie* 75.8 (1993): 687-706.
19. Bowles, Dianna J. "Defense-related proteins in higher plants." *Annual review of biochemistry* 59.1 (1990): 873-907.
20. Ryan, Clarence A. "Protease inhibitors in plants: genes for improving defenses against insects and pathogens." *Annual review of phytopathology* 28.1 (1990): 425-449.
21. Helland, Ronny, et al. "The crystal structures of the complexes between bovine β -trypsin and ten P1 variants of BPTI." *Journal of molecular biology* 287.5 (1999): 923-942.
22. Bewley, J. Derek, and Michael Black. "Seeds." *Seeds*. Springer, Boston, MA, 1994. 1-33.
23. Valueva, T. A., and V. V. Mosolov. "Protein inhibitors of proteinases in seeds: 1. Classification, distribution, structure, and properties." *Russian journal of plant physiology* 46.3 (1999): 307-321.
24. Estruch, Juan J., et al. "Transgenic plants: an emerging approach to pest control." *Nature biotechnology* 15.2 (1997): 137.
25. James, Gordon T. "Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers." *Analytical biochemistry* 86.2 (1978): 574-579.

26. Koeppe, Roger E., and Robert M. Stroud. "Mechanism of hydrolysis by serine proteases: direct determination of the pKa's of aspartyl-102 and aspartyl-194 in bovine trypsin using difference infrared spectroscopy." *Biochemistry* 15.16 (1976): 3450-3458.
27. Fuentes-Prior, Pablo, et al. "Structural basis for the anticoagulant activity of the thrombin–thrombomodulin complex." *Nature* 404.6777 (2000): 518.
28. Kobori, M., et al. "Wedelolactone suppresses LPS-induced caspase-11 expression by directly inhibiting the IKK complex." *Cell death and differentiation* 11.1 (2004): 123.
29. Palmer, James T., et al. "Vinyl sulfones as mechanism-based cysteine protease inhibitors." *Journal of medicinal chemistry* 38.17 (1995): 3193-3196.
30. Davies, David R. "The structure and function of the aspartic proteinases." *Annual review of biophysics and biophysical chemistry* 19.1 (1990): 189-215.
31. Jiang, Baohong, et al. "Salvianolic acid B functioned as a competitive inhibitor of matrix metalloproteinase-9 and efficiently prevented cardiac remodeling." *BMC pharmacology* 10.1 (2010): 10.
32. Laskowski Jr, Michael, and Ikunoshin Kato. "Protein inhibitors of proteinases." *Annual review of biochemistry* 49.1 (1980): 593-626.
33. Apweiler, Rolf, et al. "The InterPro database, an integrated documentation resource for protein families, domains and functional sites." *Nucleic acids research* 29.1 (2001): 37-40.
34. Carrell, R. W., P. A. Pemberton, and D. R. Boswell. "The serpins: evolution and adaptation in a family of protease inhibitors." *Cold Spring Harbor symposia on quantitative biology*. Vol. 52. Cold Spring Harbor Laboratory Press, 1987.
35. Chiche, Laurent, et al. "Squash inhibitors: from structural motifs to macrocyclic knottins." *Current Protein and Peptide Science* 5.5 (2004): 341-349.
36. Volpicella, M., et al. "Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in *Pichia pastoris*." *FEBS letters* 468.2-3 (2000): 137-141.
37. Mishra, Manasi, et al. "Stress inducible proteinase inhibitor diversity in *Capsicum annum*." *BMC plant biology* 12.1 (2012): 217.
38. Werner, Milton H., and David E. Wemmer. "Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution." *Biochemistry* 31.4 (1992): 999-1010.

39. Joshi, Bimba N., et al. "Cysteine protease inhibitor from pearl millet: a new class of antifungal protein." *Biochemical and biophysical research communications* 246.2 (1998): 382-387.
40. Ryan, C. A. "Proteinase inhibitors in plant leaves: a biochemical model for pest-induced natural plant protection." *Trends in biochemical sciences* 3.3 (1978): 148-150.
41. Walsh, K. A., and P. E. Wilcox. "[3] Serine proteases." *Methods in enzymology*. Vol. 19. Academic Press, 1970. 31-41.
42. Kraut, Joseph. "Serine proteases: structure and mechanism of catalysis." *Annual review of biochemistry* 46.1 (1977): 331-358.
43. Bier, M., and F. F. Nord. "On the mechanism of enzyme action. XLVI. The effect of certain ions on crystalline trypsin and reinvestigation of its isoelectric point." *Archives of biochemistry and biophysics* 33.2 (1951): 320-332.
44. Karran, Eric Howard, et al. "Human serine protease." U.S. Patent No. 6,004,794. 21 Dec. 1999.
45. Kunitz, M. "Crystallization of a trypsin inhibitor from soybean." *Science* 101.2635 (1945): 668-669.
46. Oliva, Maria Luiza V., et al. "A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds." *Biochimie* 92.11 (2010): 1667-1673.
47. Bateman, Alex, et al. "The Pfam protein families database." *Nucleic acids research* 32.suppl_1 (2004): D138-D141.
48. Rawlings, Neil D., Fraser R. Morton, and Alan J. Barrett. "MEROPS: the peptidase database." *Nucleic acids research* 34.suppl_1 (2006): D270-D272.
49. MacLean, James A., et al. "Family of Kunitz proteins from trophoblast: expression of the trophoblast Kunitz domain proteins (TKDP) in cattle and sheep." *Molecular Reproduction and Development: Incorporating Gamete Research* 65.1 (2003): 30-40.
50. Wan, Hu, et al. "A spider-derived Kunitz-type serine protease inhibitor that acts as a plasmin inhibitor and an elastase inhibitor." *PLoS One* 8.1 (2013): e53343.
51. Millers, Emma-Karin I., et al. "Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the venom of the Australian common brown snake (*Pseudonaja textilis*)." *The FEBS journal* 276.11 (2009): 3163-3175.

52. Millers, E. K., et al. "Crystallization and preliminary X-ray analysis of a Kunitz-type inhibitor, textilinin-1 from *Pseudonaja textilis textilis*." *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 62.7 (2006): 642-645.
53. Richardson, M. "The proteinase inhibitors of plants and micro-organisms." *Phytochemistry* 16.2 (1977): 159-169.
54. Liu, Jianhua, et al. "NMR studies of internal dynamics of serine proteinase protein inhibitors: Binding region mobilities of intact and reactive-site hydrolyzed *Cucurbita maxima* trypsin inhibitor (CMTI)-III of the squash family and comparison with those of counterparts of CMTI-V of the potato I family." *Protein science* 7.1 (1998): 132-141.
55. Jackson, Richard M., and Robert B. Russell. "The serine protease inhibitor canonical loop conformation: examples found in extracellular hydrolases, toxins, cytokines and viral proteins1." *Journal of molecular biology* 296.2 (2000): 325-334.
56. Dodson, Guy, and Alexander Wlodawer. "Catalytic triads and their relatives." *Trends in biochemical sciences* 23.9 (1998): 347-352.
57. Clemente, Alfonso, et al. "Pea (*Pisum sativum* L.) Protease Inhibitors from the Bowman– Birk Class Influence the Growth of Human Colorectal Adenocarcinoma HT29 Cells in Vitro." *Journal of agricultural and food chemistry* 53.23 (2005): 8979-8986.
58. Tamhane, Vaijayanti A., et al. "Plant Pin-II family proteinase inhibitors: structural and functional diversity." *Funct Plant Sci Biotechnol* 6 (2012): 42-58.
59. Migliolo, Ludovico, et al. "Structural and mechanistic insights into a novel non-competitive Kunitz trypsin inhibitor from *Adenantha pavonina* L. seeds with double activity toward serine-and cysteine-proteinases." *Journal of Molecular Graphics and Modelling* 29.2 (2010): 148-156.
60. Peräkylä, Mikael, and Peter A. Kollman. "Why does trypsin cleave BPTI so slowly?." *Journal of the American Chemical Society* 122.14 (2000): 3436-3444.
61. Radisky, Evette S., and Daniel E. Koshland. "A clogged gutter mechanism for protease inhibitors." *Proceedings of the National Academy of Sciences* 99.16 (2002): 10316-10321.
62. Orf, J. H., and T. H. Hymowitz. "Inheritance of the Absence of the Kunitz Trypsin Inhibitor in Seed Protein of Soybeans 1." *Crop Science* 19.1 (1979): 107-109.

63. Jofuku, K. Diane, and Robert B. Goldberg. "Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants." *The Plant Cell* 1.11 (1989): 1079-1093.
64. Luiza Vilela Oliva, Maria, et al. "Structural and functional properties of Kunitz proteinase inhibitors from leguminosae: a mini review." *Current Protein and Peptide Science* 12.5 (2011): 348-357.
65. Telang, Manasi A., et al. "Characterization of two midgut proteinases of *Helicoverpa armigera* and their interaction with proteinase inhibitors." *Journal of Insect Physiology* 51.5 (2005): 513-522.
66. Hoffmann, Michael P., et al. "Field evaluation of transgenic tobacco containing genes encoding *Bacillus thuringiensis* δ -endotoxin or cowpea trypsin inhibitor: efficacy against *Helicoverpa zea* (Lepidoptera: Noctuidae)." *Journal of economic entomology* 85.6 (1992): 2516-2522.
67. Brunelle, France, Conrad Cloutier, and Dominique Michaud. "Colorado potato beetles compensate for tomato cathepsin D inhibitor expressed in transgenic potato." *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America* 55.3 (2004): 103-113.
68. McManus, Michael T., et al. "Expression of the soybean (Kunitz) trypsin inhibitor in leaves of white clover (*Trifolium repens* L.)." *Plant Science* 168.5 (2005): 1211-1220.
69. Islam, Afsana, et al. "Transcription of biotic stress associated genes in white clover (*Trifolium repens* L.) differs in response to cyst and root-knot nematode infection." *PloS one* 10.9 (2015): e0137981.
70. Ye, Xiujuan, and Tzi Bun Ng. "A trypsin-chymotrypsin inhibitor with antiproliferative activity from small glossy black soybeans." *Planta medica* 75.05 (2009): 550-556.
71. Nakahata, Adriana M., et al. "Structural and inhibitory properties of a plant proteinase inhibitor containing the RGD motif." *International journal of biological macromolecules* 40.1 (2006): 22-29.
72. Krummenacher, Claude, et al. "Structure of unliganded HSV gD reveals a mechanism for receptor-mediated activation of virus entry." *The EMBO journal* 24.23 (2005): 4144-4153.

73. Moroz, Leonard A., and William H. Yang. "Kunitz soybean trypsin inhibitor: a specific allergen in food anaphylaxis." *New England Journal of Medicine* 302.20 (1980): 1126-1128.
74. Schulz, Georg E., and R. Heiner Schirmer. *Principles of protein structure*. Springer Science & Business Media, 2013.
75. Song, Hyun Kyu, and Se Won Suh. "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." *Journal of molecular biology* 275.2 (1998): 347-363.
76. Krauchenco, Sandra, et al. "Crystal structure of the Kunitz (STI)-type inhibitor from *Delonix regia* seeds." *Biochemical and biophysical research communications* 312.4 (2003): 1303-1308.
77. Krauchenco, Sandra, et al. "Three-dimensional structure of an unusual Kunitz (STI) type trypsin inhibitor from *Copaifera langsdorffii*." *Biochimie* 86.3 (2004): 167-172.
78. Onesti, Silvia, Peter Brick, and David M. Blow. "Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina coffra* seeds." *Journal of molecular biology* 217.1 (1991): 153-176.
79. Dattagupta, Jiban K., et al. "Refined crystal structure (2.3 Å) of a double-headed winged bean α -chymotrypsin inhibitor and location of its second reactive site." *Proteins: Structure, Function, and Bioinformatics* 35.3 (1999): 321-331.
80. Azarkan, Mohamed, et al. "The plasticity of the β -trefoil fold constitutes an evolutionary platform for protease inhibition." *Journal of Biological Chemistry* (2011): jbc-M111.
81. Patil, Dipak N., et al. "Structural basis for dual inhibitory role of tamarind K unitz inhibitor (TKI) against factor X a and trypsin." *The FEBS journal* 279.24 (2012): 4547-4564.
82. Hansen, Daiane, et al. "Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor." *Biochemical and biophysical research communications* 360.4 (2007): 735-740.
83. Zhou, Dongwen, et al. "Crystal structures of a plant trypsin inhibitor from *Enterolobium contortisiliquum* (EcTI) and of its complex with bovine trypsin." *PloS one* 8.4 (2013): e62252.

84. Kumar, Ashish, et al. "MP-4 contributes to snake venom neutralization by *Mucuna pruriens* seeds through an indirect antibody-mediated mechanism." *Journal of Biological Chemistry* (2016): jbc-M115.
85. Guerra, Yasel, et al. "Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: Could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop?." *Journal of structural biology* 195.2 (2016): 259-271.
86. Meulenbroek, Elisabeth M., et al. "Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato." *Acta Crystallographica Section D* 68.7 (2012): 794-799.
87. Panigrahi, Priyabrata, et al. "Engineering proteins for thermostability with iRDP web server." *PloS one* 10.10 (2015): e0139486.
88. DeLano, Warren L. "The PyMOL molecular graphics system." <http://www.pymol.org> (2002).
89. McNicholas, S., et al. "Presenting your structures: the CCP4mg molecular-graphics software." *Acta Crystallographica Section D: Biological Crystallography* 67.4 (2011): 386-394.
90. Gáspári, Zoltán, et al. "Reconciling the lock-and-key and dynamic views of canonical serine protease inhibitor action." *FEBS letters* 584.1 (2010): 203-206.
91. Chothia, Cyrus. "The nature of the accessible and buried surfaces in proteins." *Journal of molecular biology* 105.1 (1976): 1-12.
92. Shrake, A., and J. A. Rupley. "Environment and exposure to solvent of protein atoms. Lysozyme and insulin." *Journal of molecular biology* 79.2 (1973): 351-371.
93. Kabsch, W., and C. Sander. "DSSP: definition of secondary structure of proteins given a set of 3D coordinates." *Biopolymers* 22 (1983): 2577-2637.
94. Mariotti, Marco, et al. "Composition and evolution of the vertebrate and mammalian selenoproteomes." *PLoS One* 7.3 (2012): e33066.
95. Miseta, Attila, and Peter Csutora. "Relationship between the occurrence of cysteine in proteins and the complexity of organisms." *Molecular biology and evolution* 17.8 (2000): 1232-1239.

96. Desiraju, Gautam R., and Thomas Steiner. The weak hydrogen bond: in structural chemistry and biology. Vol. 9. International Union of Crystal, 2001.
97. Dougherty, Dennis A. "Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp." *Science* 271.5246 (1996): 163-168.
98. Hung, Chih-Hung, Ming-Chih Lee, and Jung-Yaw Lin. "Inactivation of *Acacia confusa* trypsin inhibitor by site-specific mutagenesis." *FEBS letters* 353.3 (1994): 312-314.
99. Yoo, Byung-Chun, et al. "Characterization of *Cucurbita maxima* phloem serpin-1 (CmPS-1): a developmentally regulated elastase inhibitor." *Journal of Biological Chemistry* (2000).
100. Bendre, Ameya Dipak, Sureshkumar Ramasamy, and C. G. Suresh. "MS009. P47." *Foundations of Crystallography* 70 (2017): C271.

Chapter 2



Cloning and characterization of recombinant
CaTl2

In the present chapter, we have cloned the CaTI2 gene from cDNA in two expression systems, *E. coli* and *P. pastoris*. Recombinant CaTI2 expression was confirmed, and the protein was purified to homogeneity. The purified protein was then assayed for its biochemical activity both qualitatively and quantitatively followed by its characterization using various biophysical techniques.

Section 1: Cloning, heterologous expression and purification of recombinant CaTI2

1.1 Introduction

Chickpea is world's 3rd legume crop with India as the top producer (1). In past 30 years, chickpea production in India has increased almost 28% (2). Noticeably, the number of pathogens and pests attacking chickpea has simultaneously increased from 49 to 175, of which 35 are from India (3). The losses due to pests/pathogens reported are as high as 100% (4). Serine proteases account for as much as 95% of total gut proteases in insect pests (5). However, when insects consume a diet that is rich in protease inhibitors, they cannot assimilate the ingested food and eventually starve to death. This strategy has been exploited by researchers to control the pests attacking crop plants (6). Studies on Kunitz inhibitors from chickpea have gained attention, in order to evaluate their inhibitory potential against insect proteases. Reports suggest that when *Helicoverpa armigera* larvae are fed with artificial diet containing a Kunitz inhibitor from chickpea, CaTI2, the larvae showed decreased growth and eventual death (7). Although about 10 KTI sequences have been predicted from Chickpea, there is only meager information available about their structural and functional aspects (8).

Josefina Hernández-Nistal et al with their immunolocalization experiments proved that unlike classical KTIs, the CaTI2 is expressed in embryonic axis post seed germination and proposed that it might be engaged in embryo axis development (9). Also, the expression levels of CaTI2 derived from CTDB (Table 1) indicate that the inhibitor is expressed more in roots than in other parts of Chickpea (10).

As mentioned above, a KTI, CaTI2 from chickpea was capable of inducing anti-metabolic effects of *H. armigera* larvae while Nair et al in their work studied the antimicrobial

properties of CaTI2 on fungal pathogen, *Fusarium oxysporum f. s. p. ciceris* along with inhibition of trypsin (11).

Table 1: CaTI2 gene expression levels in RPM (Reads per million) in various tissues of Chickpea. Data reproduced from CTDB database (10)

Gene	Shoot	Root	Mature leaf	Flower bud	Young pod
CaTI2	8.9	129.8	0.0	21.6	3.7

Due to such varied functions and properties of the CaTI2 and the scope for their potential applications, detailed biochemical and structural characterization of the inhibitor is imperative.

1.2 Materials

Tris base and other salts were procured from Sigma (USA). PrintEnrich TMSEC 650 column and molecular weight markers were from Bio-Rad. All DNA manipulation enzymes were procured from New England Biolabs (NEB). The RNA isolation kit, Plant Spectrum was from Sigma Aldrich India. cDNA synthesis Kit, High-Capacity cDNA Reverse Transcription Kit was from Life technologies Pvt. Ltd. DNA isolation and purification kits, cloning plasmids and *E. coli* strains and Ni-NTA Superflow resin were from Qiagen. pGEMT, pPICZαA vectors were from Invitrogen (Thermo Fisher Scientific USA) while pET series vectors were from Novagen (Merck, USA).

1.3 Methods

1.3.1 Sequence analysis

The CaTI2 gene (NCBI reference no.: NM_001279121.1), did not contain any introns. The amino acid sequence of CaTI2 (NCBI reference no.: NP_001266050.1) was checked for presence of signal peptide using SignalP 3.0 server (12). The consensus sequence of Kunitz legume family and other representative and reviewed sequences of KTIs were retrieved from NCBI database. The sequences were aligned using Clustal omega and ESript 3.0 (13) and a

neighbor joining tree was constructed with the help of Mega 7 (14) with 10000 iterative cycles. The tree was visualized using iTOL server (15).

1.3.2 cDNA synthesis, molecular cloning and heterologous expression in E. coli

Total RNA was extracted from sprouted chickpea seeds. The CaTI2 gene was amplified from cDNA which was synthesized from total RNA. PCR was performed in a gradient thermal cycler using conditions: 94°C/5 min, 30 cycles of [94°C/30s, 55°C/30s, 68°C/60s] and final extension at 68°C/10 min. The PCR product was eluted from 1% (w/v) agarose gel and ligated at 16°C for 12hr with the pGEMT vector. The ligation mixture was transformed into *E. coli* DH5 α cells and selected on LB agar containing 100 μ g/ml Ampicillin. Colonies were screened for cells transformed successfully with plasmids using colony PCR and restriction enzyme digestion of plasmid. The gene was sequenced using T7 promoter specific primers.

Table 2: Primers used for gene amplification

Primer	Sequence	T_m
Forward	CGC CGC ATA TGT TCT CAA ACG AAG ATG TTG AAC	62.4°C
Reverse	AAC GCG CTC GAG TTA AAC AAC AGA TTT AAC AAT TCC ATC	62.7°C

After the confirmation of successful amplification and cloning in pGEM-T vector, the gene was sub cloned into pET30a vector. The CaTI2-pET30a construct was transformed into BL21 (DE3) strain of *E. coli* expression system by heat-shock treatment. The transformed cells were grown at 37°C in LB medium containing 34 μ g/ml of kanamycin. The positively transformed colonies were re-streaked on a LB agar kanamycin+ plate. From the positive colonies, the cells were grown at 37°C in LB medium containing 34 μ g/ml kanamycin at 180rpm until OD₆₀₀ of 0.6 - 0.8 is reached. Protein overexpression was induced by adding IPTG to a final concentration of 1mM and incubating the culture for 16h at 37°C. Cells were collected by centrifugation at 8000rpm for 10min followed by re-suspension of cell pellet in 25 μ l/ml culture, of lysis buffer (100mM Tris, 100mM NaCl, 0.5% (v/v) Triton X-100, pH 7.8). The cells were further lysed and homogenized by sonication and crude extract was obtained by centrifugation at 12,000rpm for 20min at 4°C.

1.3.3 Purification of 6×His-tagged CaTI2

The supernatant of crude extract was subjected to affinity chromatography using Ni-NTA Superflow resin. The Ni-NTA resin was equilibrated with 10 volumes of 100mM Tris, pH 7.8. The cleared lysate was added to Ni-NTA resin and mixed gently by rotating (20rpm on a rotary shaker) at 4°C for 1h. The lysate-Ni-NTA mixture was loaded onto the column and then washed with 10 volumes of equilibration buffer containing 30mM imidazole (wash 1). The recombinant protein was eluted with 100mM Tris, containing 250mM imidazole [pH 7.8] (elution). This partially purified CaTI2 was then further purified by size exclusion chromatography. The fractions collected during the purification stages were analyzed by SDS-PAGE and fractions that were positive for recombinant CaTI2 were pooled and used subsequently. The purified CaTI2 inhibitor was confirmed by western blot using anti-6xHis tag primary antibodies, HRP coupled secondary antibodies and SIGMAFAST™ 3, 3'-Diaminobenzidine developer tablets.

1.3.4 Cloning, heterologous expression and purification of recombinant CaTI2 from P. pastoris:

To sub-clone the CaTI2 cDNA in the *P. pastoris* expression vector pPICZαA, in frame with the α-factor secretion signal, the cDNA fragment from the pGEMT clone, whose sequence was verified to be correct, was re-amplified using primers with restriction sites for EcoRI and XbaI. The PCR product was purified and ligated into linearized vector pPICZαA. The CaTI2-pPICZαA construct was transformed into *E. coli* DH5α competent cells for maintaining the clone. The clones were sequenced to verify the absence of any mutations. The plasmid construct was linearized with the restriction enzyme SacI and transformed into the GS115 strain of *P. pastoris* by electroporation using a 2kV single pulse program (MicroPulser™ Electroporator Bio-rad).

A single positive colony for CaTI2 expression was inoculated and pre-cultured in a conical flask containing 100mL of medium containing 15ml 10% (v/v) glycerol, 15ml of 10x yeast nitrogen base and 300μl biotin (0.2mg/ml) at 28°C with 180rpm shaking for 48hr. Cells were harvested by centrifugation at 3000×g for 10min at RT and washed with sterile water followed by inoculation in a 1000ml medium containing 2% (w/v) tryptone, 1% (w/v) yeast extract, 0.1M potassium phosphate, pH 6.5, 1.34% (w/v) yeast nitrogen base, 1.5% (v/v)

methanol and 0.2gm biotin. The culture was incubated at 28°C with 180rpm shaking for 72hr. Every 24hrs, the culture was induced by adding absolute methanol equal to the 1.5% of total media volume. The cells were harvested by centrifugation at 3000×g for 20min and discarded. The supernatant was lyophilized to reduce the volume and extensively dialyzed against 100mM Tris-HCl [pH 7.8]. The dialyzed liquid was subjected to affinity chromatography using Ni-NTA Superflow resin. The Ni-NTA resin was equilibrated with 10 volumes of 100mM Tris-HCl [pH 7.8]. The dialyzed solution was added to 1ml Ni-NTA resin and mixed gently by rotating (20rpm on a rotary shaker) at 4°C for 1hr. The mixture was loaded onto the column and washed with 10 column volumes Tris-Cl buffer. The bound recombinant protein was eluted with 100mM Tris-HCl [pH 7.8] containing 300mM imidazole. The eluted protein was concentrated using a 3kDa cutoff concentrator (Millipore) and imidazole was removed by buffer exchange using 100mM Tris-Cl buffer [pH 7.8]. The concentrated protein was purified to homogeneity by performing size exclusion chromatography with FPLC Biologic Duoflow™ (Bio-Rad) using 30ml PrintEnrich TMSEC 650. The protein bands were visualized using SDS-PAGE followed by silver staining. The recombinant protein was confirmed by western blot using anti-6x-His tag primary antibodies, HRP coupled secondary antibodies and SIGMAFAST™ 3, 3'-Diaminobenzidine developer tablets.

1.4 Results and discussion

1.4.1 Sequence analysis of chickpea KTIs

Ten genes with KTI domain have been previously reported from chickpea, but only two of these (CaTI2 and CaTI3) have been previously characterized at the protein level and so these were selected for alignment with other representative KTIs (16). It is evident from the sequence alignment shown in figure 1 that there is significant variation in the sequence of the inhibitor loop region (residues 60 to 69 of consensus sequence highlighted with light green shade), which is responsible for the trypsin inhibitory activity, in case of both CaTI2 and CaTI3. Also the brown box denotes the P1 residue, i.e., the residue that is responsible for trypsin inhibition by directly interacting with the Ser195 nucleophile in the protease. It should be noted that, despite the remarkable sequence variations in inhibitor loop region, CaTI2 was previously reported to have robust trypsin inhibition activity (17).

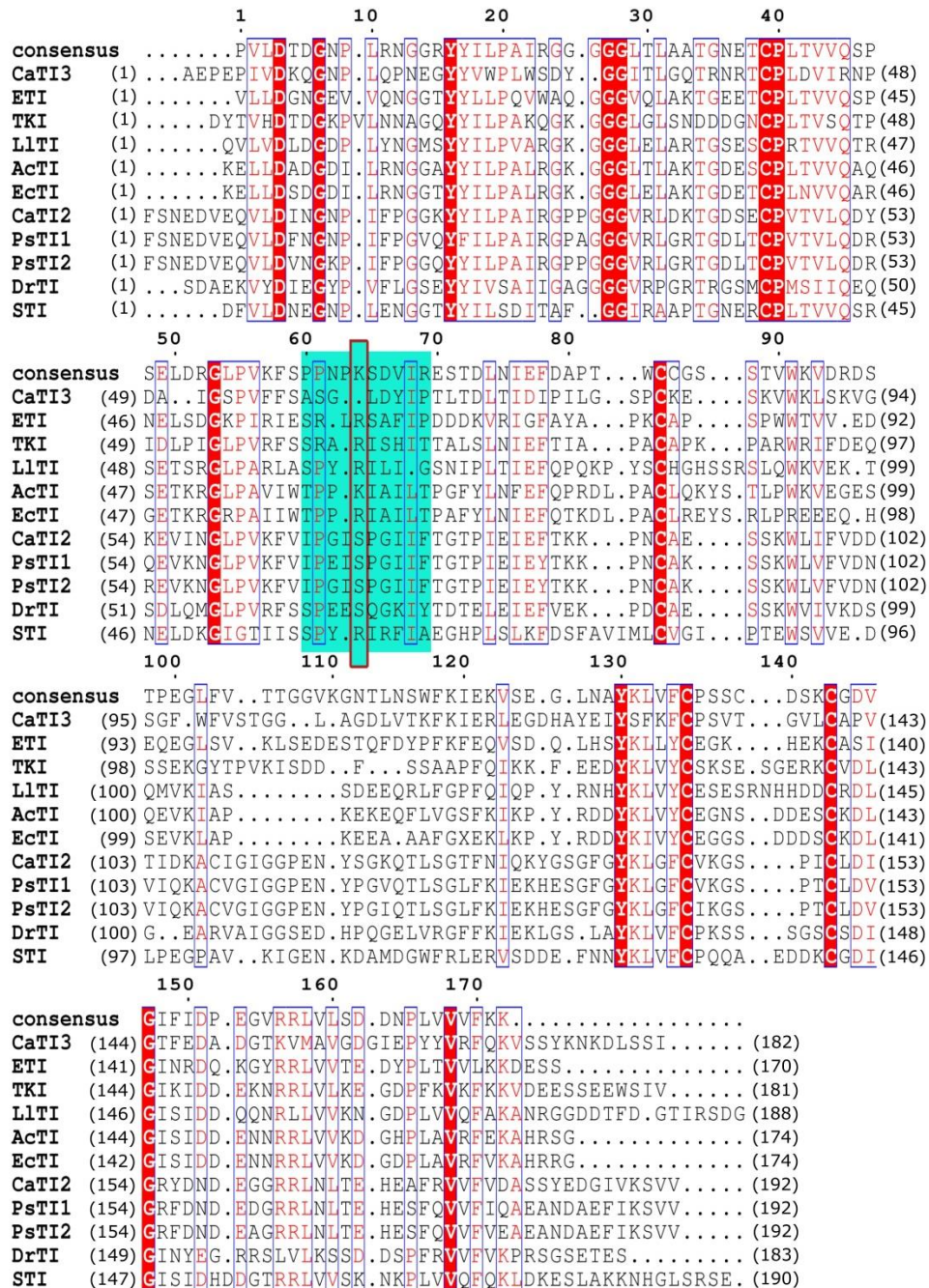


Figure 1: Multiple sequence alignment of CaTI2 with representative KTIs from Kunitz legume family. Light green box indicates the inhibitory loop region with conserved Arg residue indicated by brown box in it.

Like other canonical KTIs, CaTI2 consists of 5 conserved cysteine residues, 4 of which are involved in two disulphide linkages. In the well-characterized STI, Asn13 and Arg65 are crucial for inhibitory action (18). Surprisingly, these residues are not conserved in CaTI2, as

is shown in its alignment (Figure 1). Moreover, since the amino acid composition of the predicted reactive loop of CaTI2 is quite different when compared to other KTIs, it is important to further decipher the mechanism by which this protein is capable of inhibiting trypsin. In phylogenetic analysis (Figure 2), sequences belonging to the same or closely related subfamilies of the KTI were clustered together, i.e., all canonical KTIs have been grouped into the same clade colored dark blue, while CaTI2 and CaTI3 have diverged into different clades. CaTI2 appears to be very similar to the KTIs from *P. sativum*. The KTIs shown in the purple clade do not contain Arg residue at P1 position. Although the DrTI does not contain Arg at P1 position, it still mimics the substrate when bound to trypsin, since it contains Lys at the P1 position. Pea and Chickpea KTIs do not contain Arg or Lys at P1 position and its structural significance is yet to be understood.

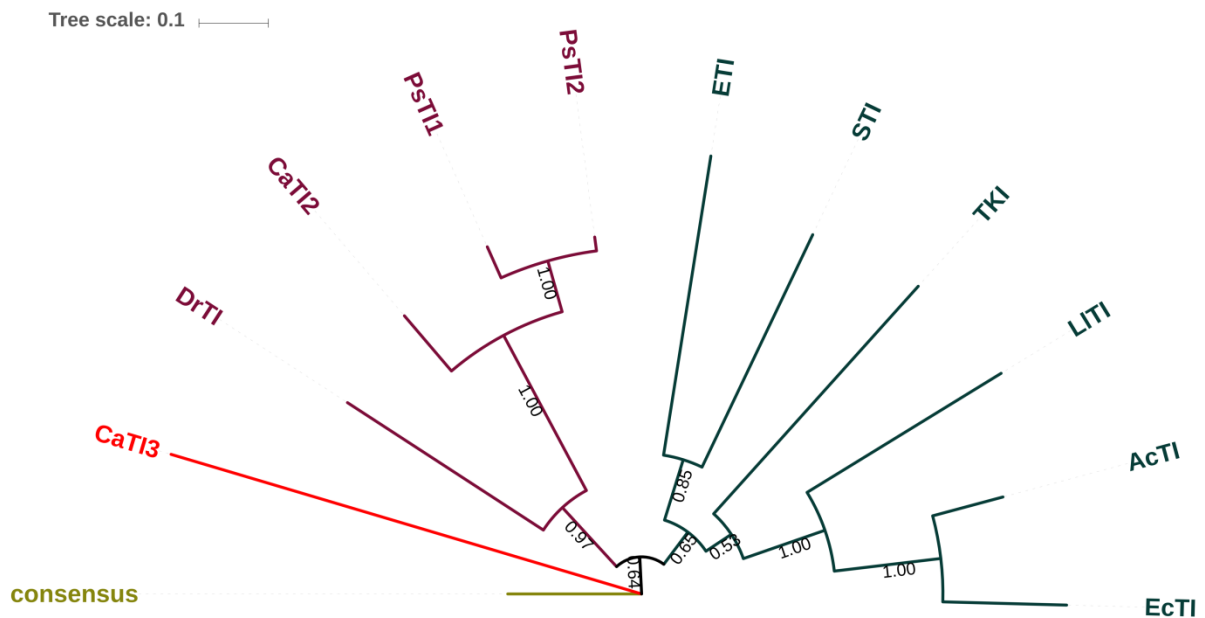


Figure 2: A neighbor-joining tree of selective KTIs indicates that CaTI3 branch out from other sequences. KTIs with Arg at P1 position clustered together in one clade (dark green) while rest other yet unique KTIs including CaTI2 form one clade (purple). CaTI3 (Red) formed clade forming a new group.

1.4.2 Cloning, expression and purification of CaTI2 from *E. coli* and *P. pastoris*

The 585bp long amplicon synthesized from cDNA was confirmed by sequencing to have no mutations resulting from PCR amplification.

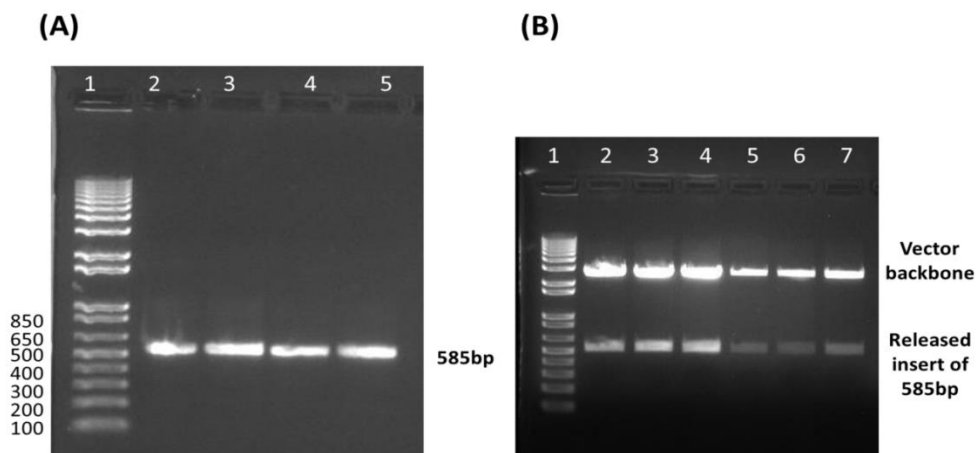


Figure 3: (A) Amplification of 585bp total cDNA. (B) Restriction enzyme digestion of CaTI2-pGEM-T construct for confirmation.

The *E. coli* strain BL21 (DE3) cells were transformed with plasmid pET30a/CaTI2 to allow the expression of a 6×His-tagged recombinant protein. Optimization of recombinant protein overexpression at different post induction temperatures (37°C, 28°C and 16°C) was carried out and eventually 37°C was selected as the optimum temperature for further experiments because there was no expression observed at 16°C and there was no considerable difference in expression levels at 28 and 37°C. The molecular mass of the recombinant CaTI2 expressed from the pET30a vector was estimated to be ~23kDa by SDS-PAGE,

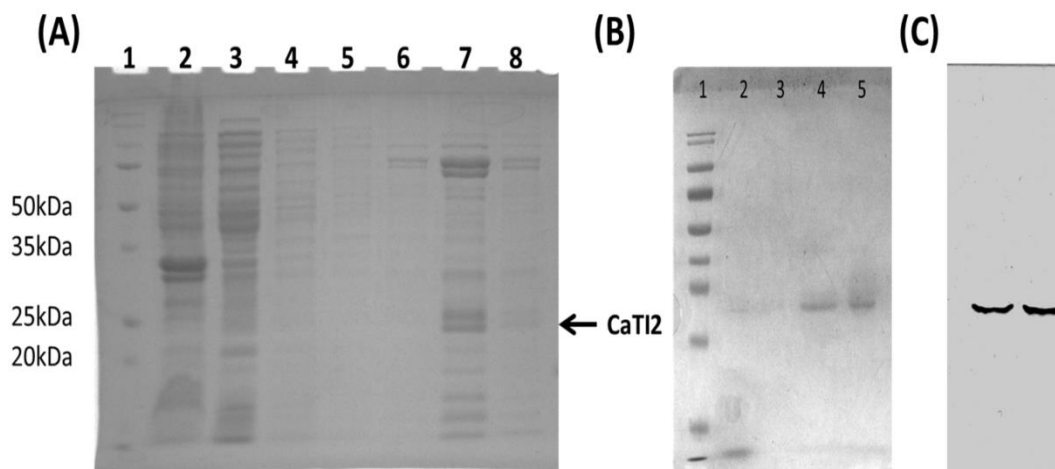


Figure 4: Overexpression of CaTI2 in *E. coli*. (A) Expression profile of CaTI2. Lane 1: Bio-Rad Dual color marker, 2: Cell pellet, 3: Fraction unbound to Ni-NTA, 4-6: Wash fractions, 7 and 8: Elution fractions. (B) Purification of Ni-NTA elution fractions by gel-

filtration. Lane 1: Bio-Rad Dual color marker, Lane 2 to 5 fractions with pure CaTI2 (C): Western blot confirmation of protein bands in lane 4 and 5 of gel in B.

Although the expression was successful in *E. coli* system, the expression levels were not enough to carry out further experiments. About 200 μ g of CaTI2 could be purified from 6ltr of LB broth. Attempts were made to increase the yield of the recombinant protein by variations in different culture parameters such as variations in expression vectors (pET30a, pET22b, pColdTF, pACYCduet, pET100 Directional TOPO, pET33b), temperature (16°C, 28°C, 37°C), IPTG concentration (0.1-1mM) and different combinations of these parameters in different *E. coli* strains. But there was no significant improvement in expression levels. To overcome this challenge we switched to the *P. pastoris* expression system.

P. pastoris (GS115) transformed with the pPICZ α A vector containing the CaTI2 expression cassette were grown on YPDZ agar (Yeast Peptone Dextrose Zeocin) at 28°C. In the cells with stable transformation, the expression plasmid gets integrated into the genome and confers resistance to Zeocin, which can be used to isolate clonal expression lines. Protein expression was induced using methanol and the expressed protein was purified using Ni-NTA affinity chromatography. Following purification, a prominent band near ~23-24kDa corresponding to the recombinant inhibitor was detected by SDS-PAGE analysis (Figure 5A), and the identity of the protein and molecular size were confirmed by Western blotting (Figure 5B).

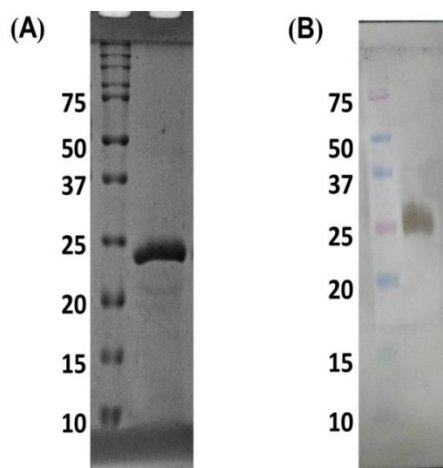


Figure 5: (A): Purified CaTI2 in lane 2 along with protein standard markers in lane 1; (B): A western blot analysis confirming expression of recombinant CaTI2 on 10% SDS-PAGE.

After a final step of purification by size exclusion chromatography the recombinant CaTI2 was used for further biochemical and structural characterization (Figure 6). By switching to the *Pichia* expression system, the final quantity of protein purified increased to ~5mg/L culture.

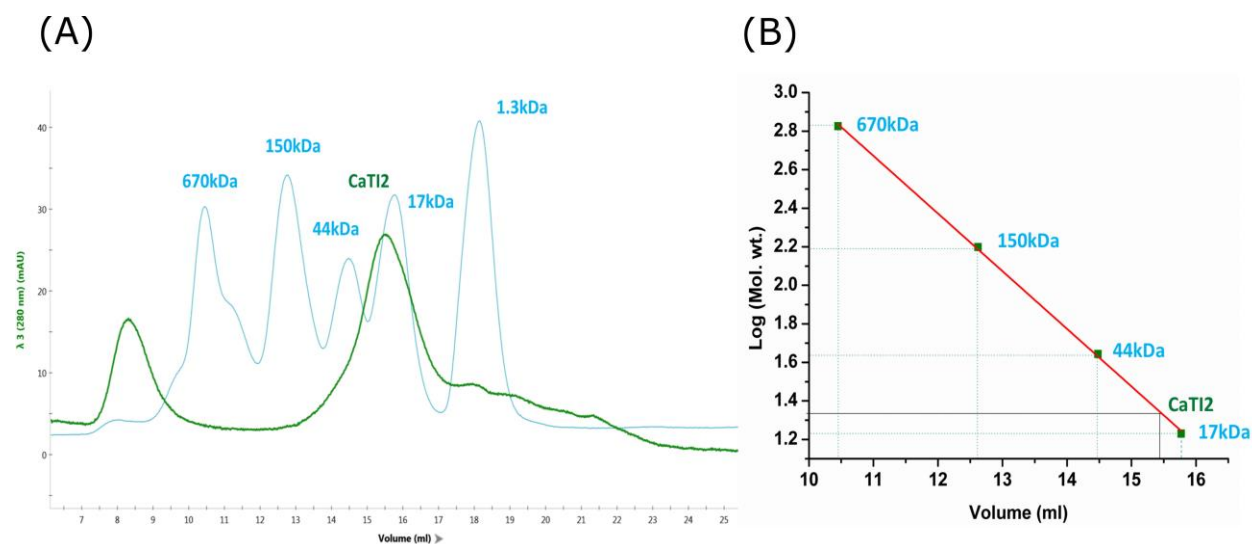


Figure 6: (A) Gel filtration profile of concentrated Ni-NTA elution fractions of CaTI2 in purple color. Sky blue colored profile indicates the peaks of proteins present in molecular weight marker. (B) Plot of logarithm of molecular weight against volume eluted. The molecular weight of recombinant CaTI2 was found to be about 23kDa.

Section 2: Biochemical and biophysical features of CaTI2

In this section of the chapter studies on the structural and functional transitions of CaTI2 under various non-optimal conditions is discussed. The folding and stability of a protein were assessed under varying pH, temperature and chaotropic agents, either individually or in combination, in order to track the conformational changes and subsequent alterations in the function of the protein. Spectroscopic techniques like far-UV circular dichroism (CD) and differential scanning fluorimetry (DSF) along with the protease-inhibition activity assay were employed to assess the differential mode of structure-function transition under different conditions.

2.1 Introduction

In modern day physics, understanding the structure and dynamics of proteins play indispensable part when studying protein folding. Insights into biochemical and biophysical behavior of a biomolecule are essential for practical applications in the areas of biotechnology and bioengineering. Advancements in tools and technologies required to study these aspects of protein biophysics enhances the progress of research tremendously.

Advances in NMR spectroscopy, Cryo electron microscopy and Electron paramagnetic resonance EPR which in earlier days were limited to studies of small molecules now provide structural information about proteins. DSF is a high throughput technique mainly used to study protein stability. It uses intrinsic tryptophan fluorescence to calculate melting temperature of a protein when subjected to linear temperature ramp (19). A relatively new alternative to Surface Plasmon Resonance (SPR) or Isothermal Titration Calorimetry (ITC) is MicroScale Thermophoresis (MST). MST allows the qualitative and quantitative study of biomolecular interactions. It detects the directed movement of a biomolecule across the microscale temperature gradient in free and bound (to ligand) state (20).

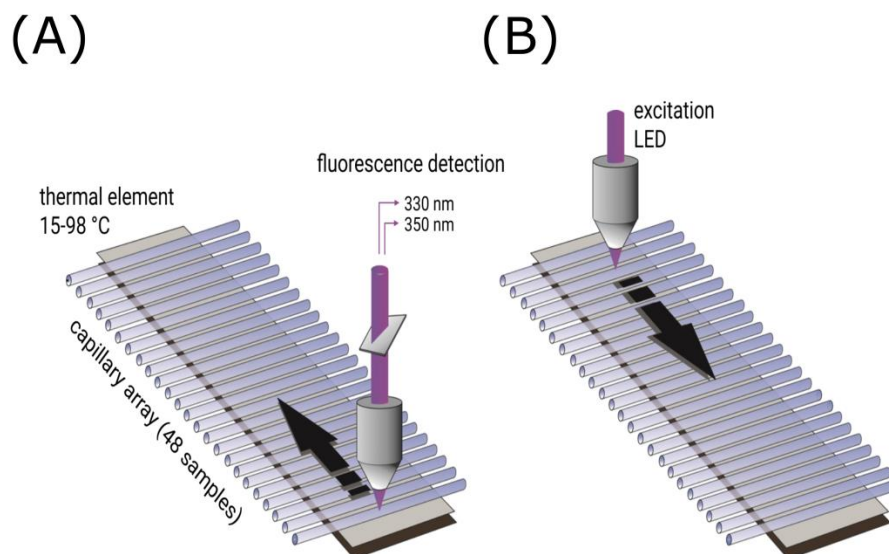


Figure 7: Basic instrumentation of (A) DSF and (B) MST. The overall instrumentation remains same. The difference lies in the output signal recorded. (Source: www.2bind.com). In DSF an intrinsic fluorescence is recorded at 330nm and 350nm when sample is heated at

various temperatures. While in MST, LED is used to excite the sample molecules and their relative movement is recorded.

CD spectroscopy is popular for its effectiveness in studying protein folding, its secondary and tertiary structure features, and structural transitions when a protein is subjected to variable environmental conditions. It uses the differences in the absorption of left-handed polarized light versus right-handed polarized light which arise due to structural asymmetry (21).

This study will focus on use of MST, nanoDSF, CD spectroscopy and biochemical assays to shed light on some of the biophysical and biochemical characteristics of the Chickpea Kunitz inhibitor CaTI2. These techniques will allow in solution studies of the behavior of CaTI2 while X-ray crystallography has been used study static 3D structure, as described in next chapter.

2.2 Materials

All chemicals including BApNA and bovine pancreatic trypsin were purchased from Sigma Aldrich (India). Prometheus NT.48 Series nanoDSF Grade Standard Capillaries (Germany) were used for DSF experiments while K005 Monolith NT.115 Capillaries (Nanotemper technologies, Germany) were used for MST experiments.

2.3 Methods

2.3.1 CaTI2 activity assays

The inhibitory activity of CaTI2 was checked qualitatively by Dot-Blot assay. In this assay a blue colored film is coated with gelatin. On this film, if protease solution is added, the gelatin layer is digested which on washing results in exposing blue color of the film. If the solution contains inhibitor along with the trypsin, the enzyme is inactivated by the inhibitor and the gelatin remains intact and original blue color of the film is not exposed (22). To check if the recombinant protein is active, this assay was used at various stages since quantitative assessment is not feasible at every step along the protein purification pipeline.

The protease inhibitor activity of purified recombinant CaTI2 was assayed on commercially available bovine pancreatic trypsin. To check the activity, a previously reported BApNA assay was used with certain modifications (23). BApNA is an artificial but specific substrate for trypsin containing Arg amino acid bound to Para-nitroaniline. Briefly, we tracked the inhibition of trypsin activity by CaTI2 by monitoring BApNA cleavage by the protease in the presence and absence of the inhibitor. We first determined that the detection of pNA (410nm) was linear up to 30µg/ml concentration, and estimated the specific activity of trypsin to be 1.75µg pNA produced/20µg of protein/min under optimal assay conditions. Inhibition assays were carried out by pre-incubating 20µg trypsin with different amounts of CaTI2 (0µg to 24µg) for 5min in assay buffer, before the addition of excess BApNA. Percentage inhibition was calculated based on reduction in the unit activity of trypsin in the presence of the inhibitor. Further we measured the activity of the inhibitor at various temperatures and pH conditions to study the effect of these parameters on its inhibitory capabilities. The inhibitor was incubated at temperatures ranging from 20°C to 90°C for 1hr, 2hr and 4hr. To understand the effect of pH on CaTI2 activity, the inhibitor was incubated in different buffers at different pH values (pH range 2-12) for 1hr, 4hr and 24hr, before checking the inhibitor activity.

2.3.2 Determination of K_d by MST

Interaction between bovine pancreatic trypsin and CaTI2 was studied by MST technique using the Monolith NT.115 instrument (Nano Temper, Munich, Germany). Purified CaTI2 was labeled using the Monolith His-Tag Labeling Kit RED-tris-NTA dye, and 20nM of labeled inhibitor protein was titrated against increasing concentrations (from 3.265nM to 53.3µM) of un-labeled bovine pancreatic trypsin after loading into hydrophilic silicon capillaries. Each measurement was performed three times. Experiments were carried out in proprietary MST buffer provided by Nano Temper technologies. Data was analyzed using the Monolith software.

2.3.3 Far UV-circular dichroism spectroscopic studies

CD spectra of the purified CaTI2 were recorded using a Jasco J-815-150S spectropolarimeter (Jasco, Tokyo, Japan) connected to a PTC343 Peltier unit circulating water bath at 25°C. Far-UV CD spectra was recorded in a rectangular quartz cell of 1mm path length in the range of

190–250nm at a scan speed of 100nm/min with a response time of 1sec and a slit width of 1nm. Purified CaTI2 at a concentration of 0.70mg/ml was used to monitor the secondary structure. Each spectrum was recorded as an average of three scanned spectra. Spectra were corrected for buffer contribution and observed values were converted to mean residue ellipticity (MRE) in $\text{deg.cm}^2/\text{dmol}$ defined as $\text{MRE} = M\theta\lambda/10dcr$, where, M is the molecular weight of the protein, $\theta\lambda$ is CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ml, and r is the average number of amino acid residues in the protein. The relative content of various secondary structure elements was calculated by using CDPro software (24).

2.3.4 Differential Scanning Fluorimetry (DSF)

Measurements on chemical and thermal unfolding were performed with the Prometheus NT.48 nanoDSF (NanoTemper Technologies, Germany). DSF experiments were based on changes in intrinsic fluorescence of the protein where the shift in fluorescence emission is plotted as the ratio between 350 and 330nm. The capillaries were filled with mixture of 20 μl of protein placed onto the capillary tray of the Prometheus NT.48. Start to end temperatures were set as 20°C to 95°C and the heating rate was defined as 1°C/min. The inflection point of the resulting sigmoidal curve provided the melting temperature (T_m) of the protein. The T_m of the inhibitor was studied in native conditions, and in the presence of urea and organic solvents. Effect of pH on the T_m of the protein was studied by recording DSF of the inhibitor in buffers of varying pH.

2.3.5 Thermal denaturation

The inhibitor (0.7mg/ml) was incubated at temperature ranging from 25°C to 95°C with interval of 5°C for 5min each. DSF and CD measurements were recorded as described earlier. In another set of experiment, activity of each sample was checked at different temperatures and time interval to determine its thermal stability.

2.3.6 Effect of pH on CaTI2

Purified CaTI2 (0.56mg/ml) incubated in buffers of different pH at 25°C were used for DSF and CD measurements as described above. The following buffers (25mM each) were used for varying the assay pH - Glycine–HCl (pH 1.0 and 3.0), citrate–phosphate (pH 5.0), potassium phosphate (pH 7.0), Tris–HCl (pH 9.0) and glycine–NaOH (pH 11.0). To study the pH

stability of the CaTI2 inhibitor, it was incubated in buffers with different pH at 28°C. Suitable aliquots were removed at regular time intervals and assayed for inhibitor activity. The readings were corrected for blank samples containing only buffers of respective pH without the enzyme and the inhibitor.

2.3.7 Treatment of the CaTI2 with urea

For urea mediated denaturation, the denaturant was freshly prepared; pH was adjusted to 8.0 and filtered through a 0.22µm syringe filter before use. The protein samples (0.50mg/ml) were incubated in 0–5M urea solution at pH 8.0 and DSF and CD measurements were recorded as mentioned above. All the CD spectra were corrected for respective blanks of urea buffer contributions. Due to instrumental limitation, the CD spectra of urea treated samples were recorded in the 250nm to 200nm range. The CaTI2 protein was treated with increasing concentrations of urea at room temperature and percent residual activity was measured using the BApNA assay.

2.3.8 Effect of solvent polarity on CaTI2

The stability of CaTI2 in presence of organic solvents [25% (v/v)] of different polarity was studied using far-UV CD, DSF measurements. Organic solvents used were methanol (METH), ethanol (ETH), Isopropanol (PROP). All the CD spectra were corrected for respective blanks of urea buffer contributions.

2.4 Results and discussion

2.4.1 Inhibition of BPT by CaTI2

CaTI2 has been previously reported to efficiently inhibit trypsin and in this study the activity of the recombinant CaTI2 was checked by Dot-Blot assay. It was observed that when the inhibitor mixed trypsin solution was loaded on gelatin coated film, the gelatin layer remained intact which was observed to be digested in case of pure trypsin solution. This suggested that that recombinant inhibitor added to the trypsin solution inactivated the trypsin thus protecting the gelatin layer. Thus we could conclude that the purified recombinant inhibitor was active (Figure 8A).

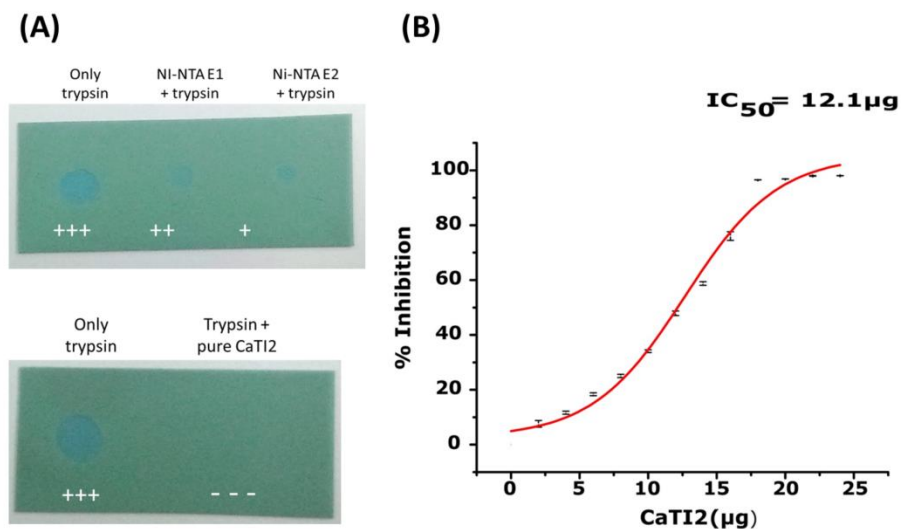


Figure 8: Assessment of inhibitory activity of CaTI2. (A) Dot-Blot assay showing when trypsin is active, blue color of the film is exposed. Pure CaTI2 inhibits trypsin keeping gelatin film intact. (B) Percent inhibition of trypsin was determined by the BApNA assay and plotted to determine the IC₅₀ value.

After qualitative confirmation of the activity, the inhibitor was assayed with BApNA for quantitative measurement of trypsin inhibition. BApNA is an artificial substrate for trypsin. When trypsin is treated with BApNA, BApNA mimics the substrate protein and is cleaved by the trypsin to produce para-nitroaniline (pNA) which a yellow colored product. The absorbance of the pNA produced by trypsin is measured at 410nm. When inhibitor binds to the trypsin, the BApNA is not digested to pNA and thus the reaction mixture does not turn yellow indicating that the inhibitor has actively inhibited the trypsin. In this study, bovine pancreatic trypsin was used to study the inhibitory activity of CaTI2, using a standard assay where trypsin activity is measured by the cleavage of the artificial substrate BApNA. From these studies we determined the IC₅₀ value for trypsin inhibition by CaTI2 as ~12 μg (Figure 8B).

2.4.2 Dissociation constant determination

When MST analysis is done without immobilizing the biomolecules, it provides direct *in solution* measurement of interactions between the biomolecules. It detects the change in the relative movement of a biomolecule before and after the binding of a ligand and thus helps us

in determining the affinity of the ligand towards the receptor molecule. It is highly sensitive technique and can detect the interactions at pico molar concentrations. We determined the K_d value for trypsin binding by CaTI2 to be around $0.631\mu\text{M}$ (Figure 9A), which is $\sim 18.7\mu\text{M}$ in case of STI as reported previously (25).

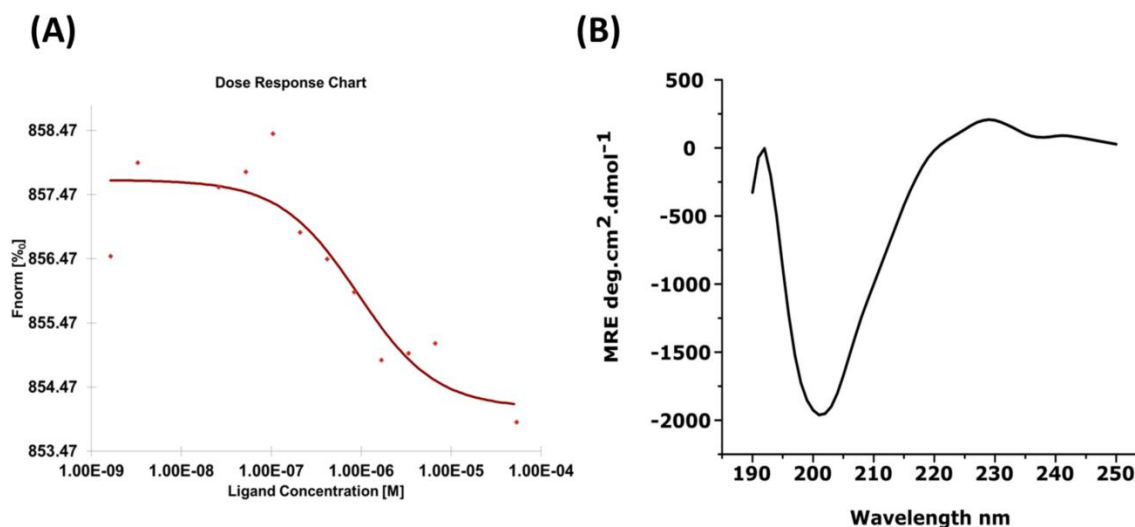


Figure 9: (A) Titration curve for determination of dissociation constant in terms of Fnorm [%] (B) Far-UV CD spectrum of CaTI2 (32nM)

2.4.3 Circular Dichroism Spectroscopy

Far-UV CD spectrum of CaTI2 showed negative ellipticity minima at 202nm and positive maxima at 229nm (Figure 9B). The spectrum was analyzed with CDPro to estimate the secondary structure contents as 5.7% α -helix and 39.4% β -sheet, suggesting that CaTI2 is a β -sheet rich protein. The ordered CD profile also indicated that the recombinant protein was in properly folded/ native form.

2.4.4 Thermal denaturation

Thermal stability of the protein provides information about the covalent and non-covalent forces involved in stabilizing the protein structure. The protein was functionally stable up to 45°C for 1hr; lost 40% of inhibitory activity upon incubation for 2hr while, 80% when incubated for 4hr (Figure 10a). Thermal denaturation of CaTI2 is an irreversible process as the activity was not recovered upon cooling the samples.

CD spectroscopic analysis: Monitoring the secondary structure during thermal denaturation by far-UV CD spectroscopy revealed a gradual decrease in the negative ellipticity when heated up to 85°C (Figure 6B). Slight structural alterations occurred up to 55°C. Major change leading to loss of the structure was observed at and above 60°C. Analysis of far-UV CD spectra in CONTINLL indicated that β -sheet content of the protein decreased significantly above 55°C with a simultaneous increase in helices and unordered structure. This marks the major change occurring in secondary structure conformation of the protein upon thermal denaturation. Further, thermal denaturation of the inhibitor is an irreversible process as the conformation was not completely regained upon cooling the sample.

Table 3: Secondary structure analysis of CaTI2 by CDPro (CONTINLL) program

Sample	α -helix	β -sheet	Turns	Unordered	NRMSD
Control	5.6	39.7	21.7	33	0.070
Thermal unfolding					
25°C	5.6	39.7	21.7	33	0.070
35°C	6.4	35.3	22.8	35.4	0.039
45°C	6.8	35.8	23.8	34.8	0.058
55°C	6.8	35	23.2	35.1	0.055
65°C	8.2	29.7	24.2	37.9	0.031
75°C	8.8	27.2	23.7	40.2	0.059
85°C	11.6	24.3	23.9	40.0	0.060
Re-cooled to 25°C	7.9	34.5	22.6	35.0	0.023
pH-induced denaturation at 24hr incubation					
2.0	7.8	39.5	25.1	27.6	0.050
4.0	4.3	41.7	21.8	32.2	0.094
6.0	5.4	40.3	22.0	32.2	0.043

8.0	5.2	40.6	22.3	32.6	0.070
10.0	5.9	40.8	21.7	31.6	0.062
12.0	4.8	43.2	21.4	30.6	0.093
Urea at 24hr incubation					
1M	3.3	38.8	23.2	35.6	0.095
2M	3.3	39.4	22	37.6	0.098
3M	3.3	40.6	22.2	33.9	0.095
4M	1	42.9	21.2	34.9	0.094
Organic solvents (25% v/v) at 24hr incubation					
Methanol	15	28.4	22.4	34.1	0.096
Ethanol	16.9	26.2	22	34.9	0.093
Isopropanol	17.1	23.2	22.5	37.2	0.051

DSF studies: Temperature induced protein denaturation was also studied using DSF or thermofluor. In MST of proteins, the directed movement of protein molecules along micro-scale temperature gradients is measured by intrinsic fluorescence emitted by tryptophan residues (26). From the thermophoretogram, the T_m of CaTI2 was determined as 61.3°C, correlating the unfolding of the protein with the loss of conformation.

2.4.5 pH-induced denaturation

Ionization status is one of the most important factor affecting the biological function and conformation of a protein. CaTI2 was active over a wide pH range; however its activity was most stable in the pH range of 8.0-10.0 for a period of up to 24hr. The activity of protein incubated at pH 8.0 was assumed to be 100%, as the protein is most stable at this pH for long term storage. A drastic loss in activity was observed at highly acidic pH (Figure 7a).

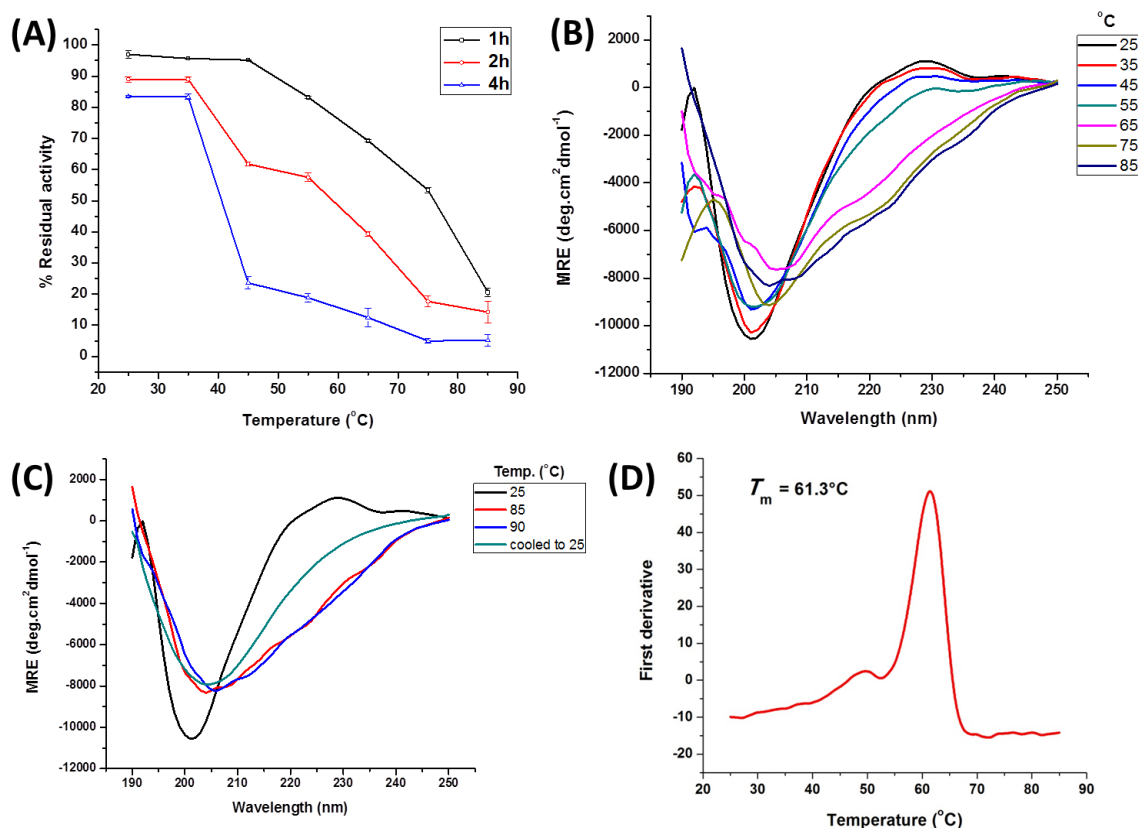


Figure 10: (A): Activity profile after 1hr, 2hr and 4hr of incubation at different temperatures (B) Far-UV CD spectra of CaTI2 incubated at respective temperatures for 5min. (C) Far-UV CD spectra of CaTI2 when re-cooled to 25°C in light blue color indicating irreversible denaturation. (D) Thermal unfolding curve, indicating the denaturation midpoints (T_m : 61.3°C) of CaTI2 using DSF.

CD analysis of CaTI2 at different pH : As indicated by far-UV CD spectra, the secondary structure of inhibitor following 12hr incubation in different pH conditions was observed to be stable in the pH range of 6.0-10.0, which correlates with its activity profile. The decrease in the negative and positive ellipticity at extreme pH values indicated changes in ordered-ness of the protein structure (Figure 11c). In spite of the loss in activity of the inhibitor at acidic pH, substantial secondary structure stability was observed for the protein (Table 3). Thus, the activity loss might be attributed to change in local conformation of trypsin-binding site or due to ionization of amino acids near the binding site rendering it unfit for proper binding.

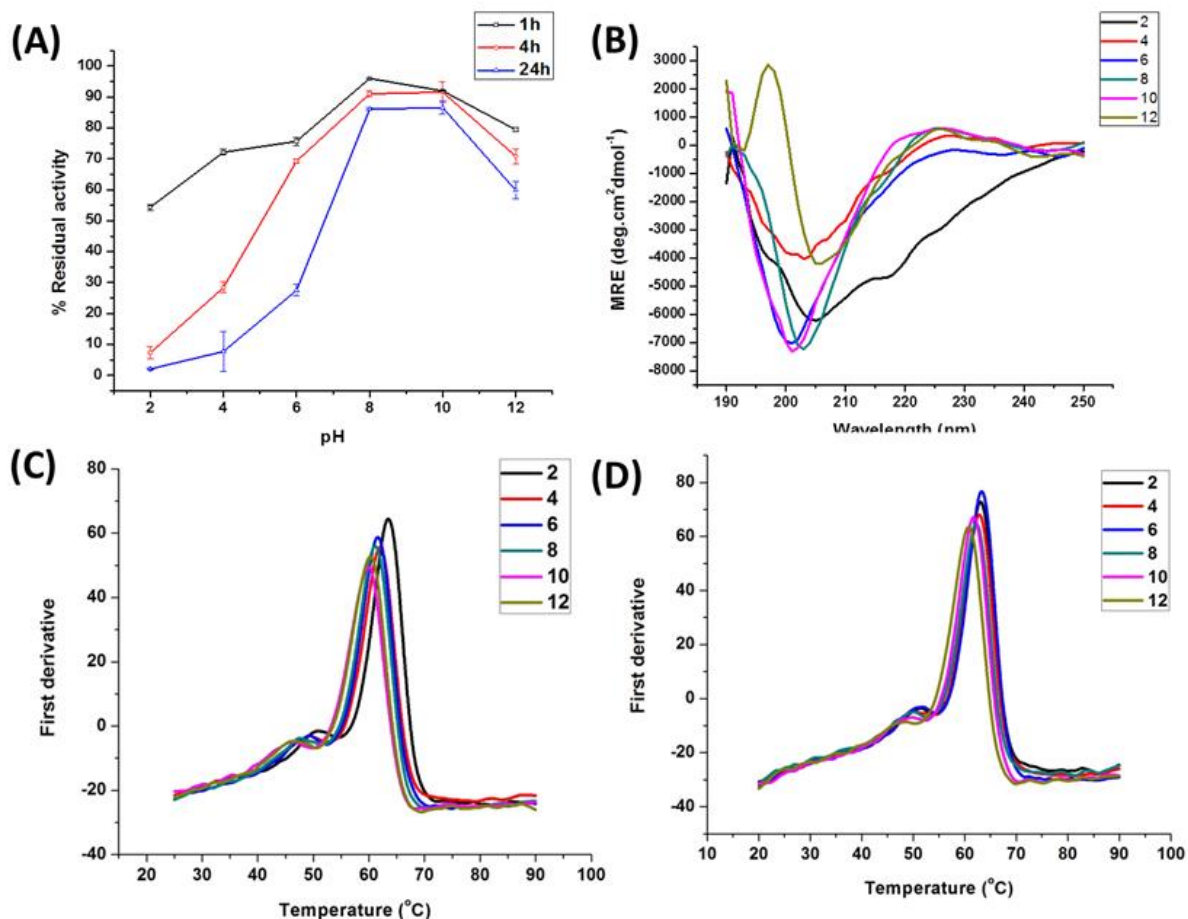


Figure 11: (A): Activity profile after 1hr, 4hr and 24hr of incubation at different pH (B) Far-UV CD spectra of CaTI2 at 24hr incubation. Thermal unfolding curve indicating the denaturation midpoints of CaTI2 using DSF at different pH for 1hr incubation (C) and 24hr incubation (D).

pH dependent thermal unfolding of CaTI2: The transitions in CaTI2 unfolding at different pH values were also studied using nanoDSF which revealed different T_m values for different pH buffers. There was negligible change observed in the melting temperatures of CaTI2 at different pH conditions and no change in T_m over increased incubation time. This supports the above observation that the protein is quite stable structurally at wide pH range, though, it loses its activity at extreme pH conditions, which could be due to ionization of amino acids near the binding site.

Table 4: Melting temperature T_m ($^{\circ}\text{C}$) of CaTI2 after 12hr incubation at respective pH

pH	T_m ($^{\circ}\text{C}$)
2	63.7 ± 0.1
4	64.5 ± 0.3
6	64.5 ± 0.5
8	63.2 ± 0.2
10	62.0 ± 0.3
12	61.1 ± 0.6

2.4.6 CaTI2 stability towards Urea

Chaotropic agents like GdnHCl or urea are widely used to study gradual unfolding and refolding of the protein. They destabilize the protein structure either by forming hydrogen bonds to the protein backbone or by decreasing the hydrophobic effect (27). Remarkably, CaTI2 was quite stable up to 2M concentration of urea and retained more than 50% of the activity even after 24hr. The inhibitor lost its activity gradually in the vicinity of 3M over a period of time, while it immediately lost activity in 4M and 5M urea (Figure 12A).

CD analysis: The far-UV CD spectra of CaTI2 showed that the protein conformation in the vicinity of >2M urea was altered. There was a drastic shift in minima from 202nm to 212nm indicating change in protein conformation. Although, the structure of the protein was not completely lost but there was a decrease in helical content of the protein, as evident from the CDPro analysis (table 3). Thus, CaTI2 was found to be fairly stable in presence of urea. Retention of the activity of CaTI2 up to 3M urea could be correlated with the retention of secondary structure (Figure 12B).

Urea dependent thermal unfolding via DSF: The protein unfolded at lower temperatures as the concentration of denaturant increased (Table 5), pointing towards the destabilisation of CaTI2 under higher concentrations of urea. DSF analysis also marks the conformational

changes in the protein with increasing concentrations of urea. Moreover, the urea based thermal unfolding (T_m) of the CaTI2 differs with incubation time suggesting alterations in the protein (Figure 12C, D). With increasing concentration of urea, the T_m CaTI2 continuously dropped. Also, a decrease in fluorescence intensity was observed.

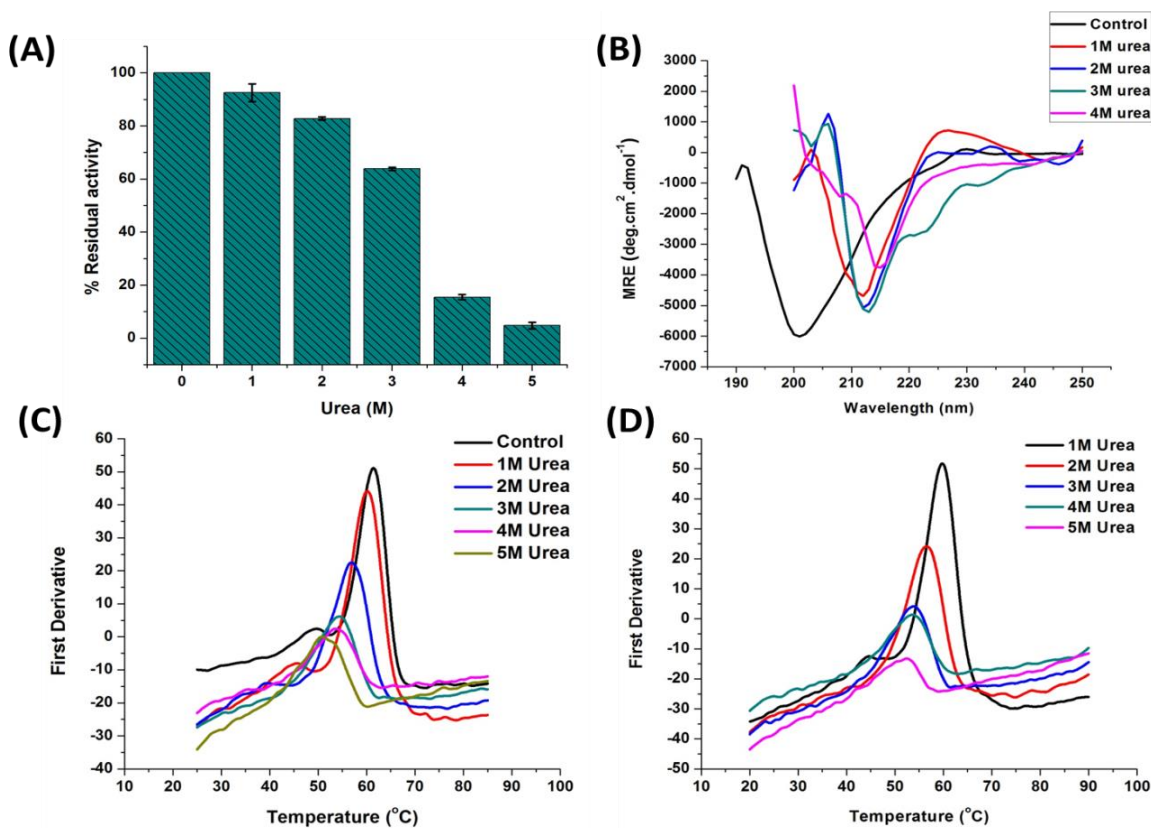


Figure 12: (A): Activity profile after 24hr of incubation at different urea concentrations (B) Far-UV CD spectra of CaTI2 at 24hr incubation in urea. (C) Thermal unfolding curve, indicating the denaturation midpoints of CaTI2 using DSF at different urea concentrations for 12hr incubation and (D) 24hr incubation.

Table 5: Thermal unfolding of CaTI2 in presence of urea

Urea (M)	T_m (°C, 1hr)	T_m (°C, 12hr)
1	59.1 ± 0.2	58.1 ± 0.2
2	56.2 ± 0.3	56.4 ± 0.3

3	54.2 ± 0.4	53.2 ± 0.4
4	53.4 ± 0.8	53.4 ± 0.7
5	51.0 ± 0.6	52.0 ± 0.4

2.4.7 Stability of CaTI2 in organic solvents

To check the effect of polar environment on stability of CaTI2, alcohols with different polarity were used. Organic solvent-stable proteins offer an interesting alternative to enzymatic/ chemical reactions in organic-solvent systems. For effective industrial applications, proteins must be active and stable at high concentrations of organic solvents. Interestingly, CaTI2 showed tolerance towards high concentrations [25% (v/v)] of alcohols methanol, ethanol and propanol.

CD studies: We observed that CaTI2 was quite stable in presence of alcohols at 25% (v/v) concentration as the overall secondary structure of CaTI2 was retained even after 12hr. However, the minima in the spectrum displayed a shift of 8nm from 202 to 210nm (Figure 13B). Further, the CONTINLL analysis suggested a decrease in β -sheets with a simultaneous increase in α -helical content of the protein in presence of alcohols (Table 3). The organic solvent stability of a protein is often imparted by its structural composition. Reports suggest that high β -pleated content in the proteins lead to low solvent stability (28). Moreover, alpha helical structures are induced in proteins/ enzymes treated with alcohols, for instance, methanol treated lysozyme (29). Thus, it is the alteration in structure, rather than loss of CaTI2 structure, that occurs in presence of non-polar groups.

Organic solvent dependent thermal unfolding of CaTI2: The thermal unfolding of the protein in presence of organic solvents also indicated the stable nature of CaTI2. The melting temperature of CaTI2 decreased with increasing concentration of solvents (12% > 25%) as seen in the table 6. Furthermore, the T_m of the protein also decreased with decreasing solvent polarity (water > methanol > ethanol > isopropanol) which correlates well with CD analysis (Table 3). This could be due to the altered hydrophobic interactions of the protein with the polar environment which prove to be favorable for stability.

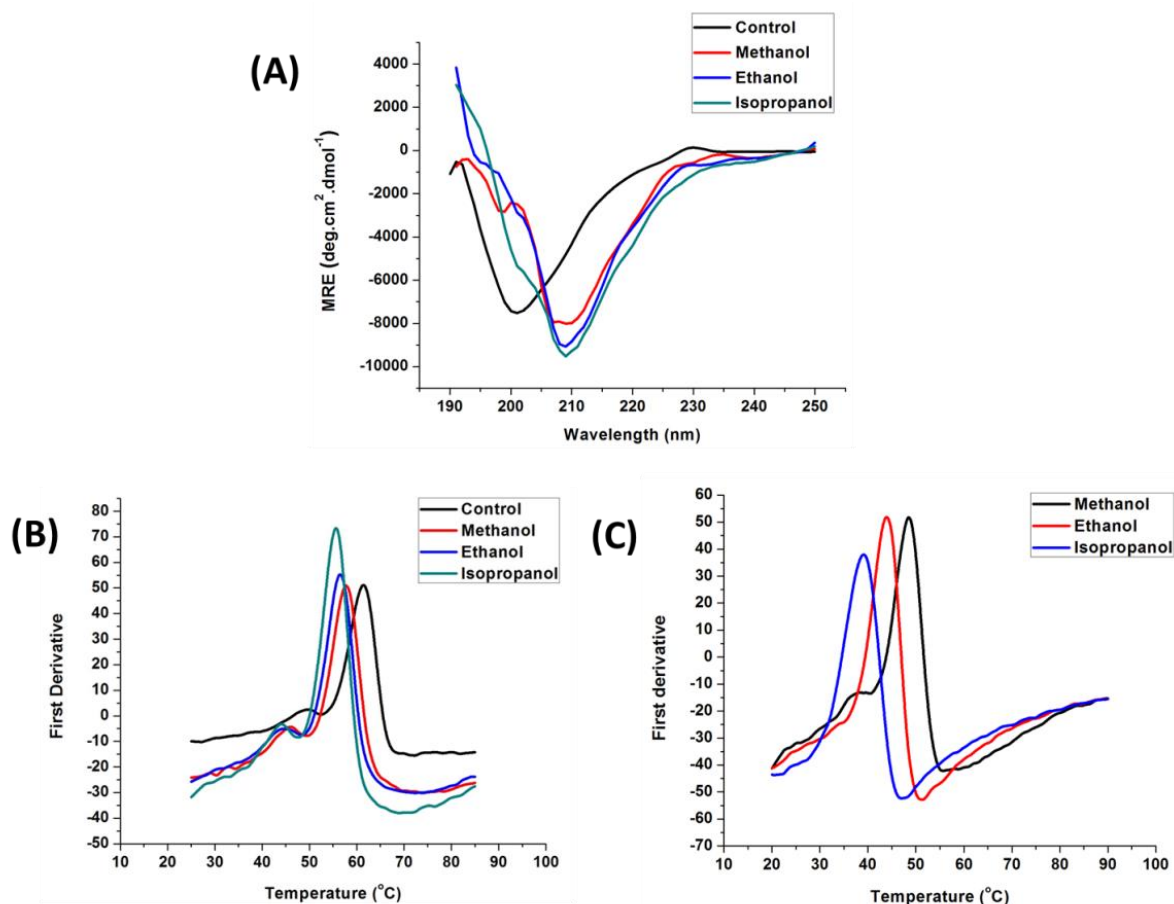


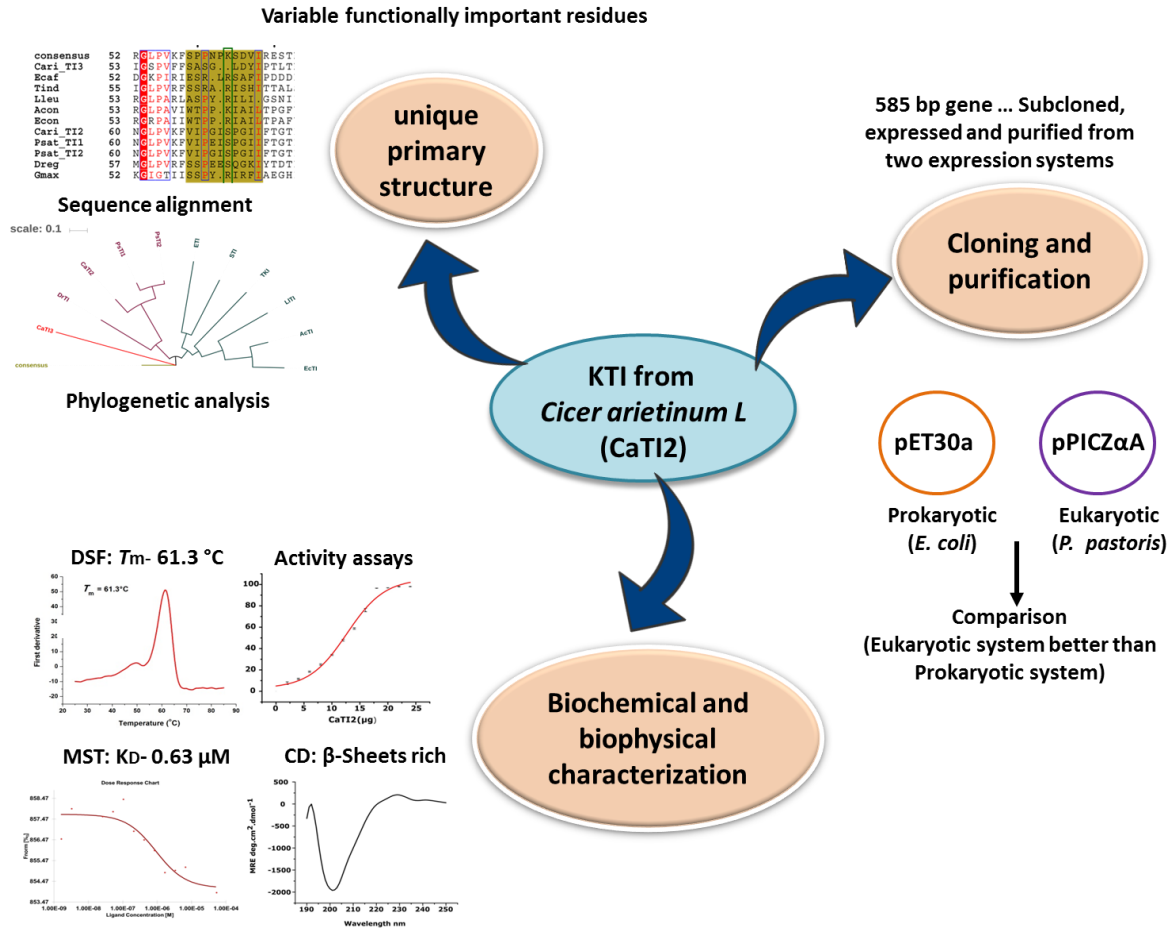
Figure 13: (A) Far-UV CD spectra of CaTI2 at 24hr incubation in alcohols. (B) Thermal unfolding curve, indicating the denaturation midpoints of CaTI2 using DSF at different alcohols at concentrations 12.5% and (C) 25%, 12hr incubation.

Table 6: Thermal unfolding of CaTI2 in presence of alcohols

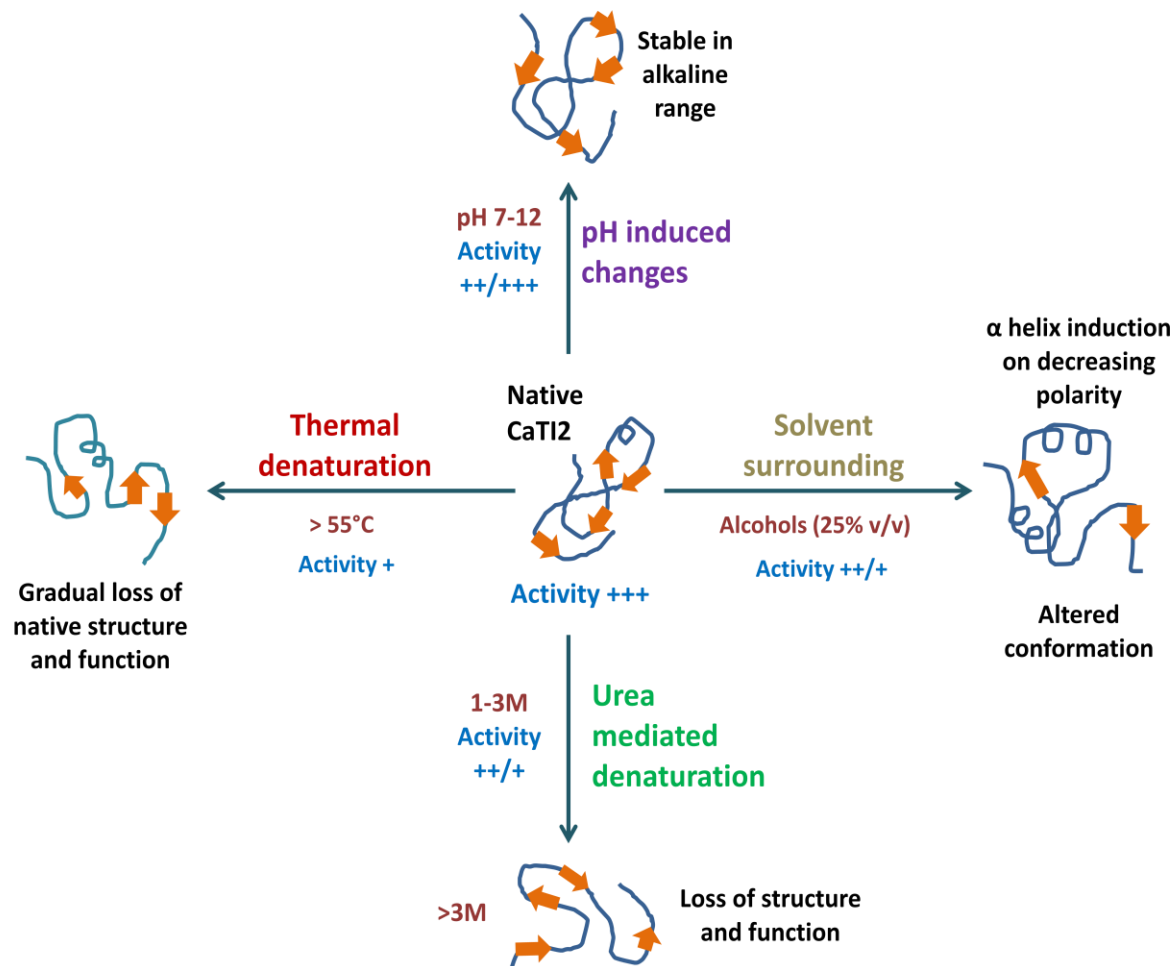
Alcohol	T_m (°C, 12.5% v/v, 12hr)	T_m (°C, 25% v/v, 12hr)
Methanol	57.6 ± 0.3	48.8 ± 0.2
Ethanol	56.4 ± 0.2	44.4 ± 0.3
Isopropanol	55.6 ± 0.4	41.6 ± 0.3

Graphical Summary

Section 1: Schematic depiction of cloning of CaTI2 and its characterization



Section 2: Schematic diagram summarizing the effect of various denaturants (urea, organic solvents, pH and temperature) on structure and function of CaTI2



References

1. Redden RJ, Berger JD. History and origin of chickpea 2007. 1–13 p.
2. Chickpea production in 2016, Crops/Regions/World list/Production Quantity (pick lists)". UN Food and Agriculture Organization, Corporate Statistical Database (FAOSTAT). 2017. Retrieved 23 February 2018.
3. Nene, Y. L., V. K. Shelia, and S. B. Sharma. "A world list of chickpea and pigeonpea diseases." Legume Pathology Progress Report 7 (1989).
4. Ghosh, Raju, et al. "Occurrence and distribution of chickpea diseases in central and southern parts of India." American Journal of Plant Sciences 4.4 (2013): 940-944.
5. rinivasan, Ajay, Ashok P. Giri, and Vidya S. Gupta. "Structural and functional diversities in lepidopteran serine proteases." Cellular & molecular biology letters 11.1 (2006): 132.
6. Hilder, Vaughan A., et al. "A novel mechanism of insect resistance engineered into tobacco." Nature 330.6144 (1987): 160.
7. Srinivasan, Ajay, et al. "A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae." Plant Molecular Biology 57.3 (2005): 359-374.
8. Sharma, Ranu, and C. G. Suresh. "Genome-wide identification and structure-function studies of proteases and protease inhibitors in *Cicer arietinum* (chickpea)." Computers in biology and medicine 56 (2015): 67-81.
9. Hernández-Nistal, Josefina, et al. "Two cell wall Kunitz trypsin inhibitors in chickpea during seed germination and seedling growth." Plant Physiology and Biochemistry 47.3 (2009): 181-187.
10. Verma, Mohit, et al. "CTDB: an integrated chickpea transcriptome database for functional and applied genomics." PloS one 10.8 (2015): e0136880.
11. Nair, Meera, and Sardul Singh Sandhu. "A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts an antimicrobial effect on *Fusarium oxysporum* f. sp. *ciceris*." Agricultural Sciences 4.11 (2013): 585-594.
12. Bendtsen, Jannick Dyrlov, et al. "Improved prediction of signal peptides: SignalP 3.0." Journal of molecular biology 340.4 (2004): 783-795.
13. Robert, ESPript. "Gouet, 2014 <http://esprpt.ibcp.fr/ESPript>."

14. Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets." *Molecular biology and evolution* 33.7 (2016): 1870-1874.
15. Letunic, Ivica, and Peer Bork. "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." *Nucleic acids research* 44.W1 (2016): W242-W245.
16. Gupta, Sonika, et al. "Secretome analysis of chickpea reveals dynamic extracellular remodeling and identifies a Bet v1-like protein, CaRRP1 that participates in stress response." *Scientific reports* 5 (2015): 18427.
17. Nair, Meera, Sardul Singh Sandhu, and Anita Babbar. "Purification of trypsin inhibitor from seeds of *Cicer arietinum* (L.) and its insecticidal potential against *Helicoverpa armigera* (Hübner)." *Theoretical and Experimental Plant Physiology* 25.2 (2013): 137-148.
18. De Meester, Patrice, et al. "Structure of the Kunitz-type soybean trypsin inhibitor (STI): implication for the interactions between members of the STI family and tissue-plasminogen activator." *Acta Crystallographica Section D* 54.4 (1998): 589-597.
19. Strutz, Wyatt. "Exploring Protein Stability by NanoDSF." *Biophysical Journal* 110.3 (2016): 393a.
20. Wienken, Christoph J., et al. "Protein-binding assays in biological liquids using microscale thermophoresis." *Nature communications* 1 (2010): ncomms1093.
21. Greenfield, Norma J. "Using circular dichroism spectra to estimate protein secondary structure." *Nature protocols* 1.6 (2006): 2876.
22. Harsulkar, Abhay M., et al. "Characterization of *Helicoverpa armigera* gut proteinases and their interaction with proteinase inhibitors using gel X-ray film contact print technique." *Electrophoresis* 19.8-9 (1998): 1397-1402.
23. Dietz, Albert A., Herbert M. Rubinstein, and LaVerne Hodges. "Measurement of alpha1-antitrypsin in serum, by immunodiffusion and by enzymatic assay." *Clinical Chemistry* 20.3 (1974): 396-399.
24. Sreerama, Narasimha, and Robert W. Woody. "Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR

- methods with an expanded reference set." *Analytical biochemistry* 287.2 (2000): 252-260.
25. Luthy, James A., et al. "Detailed Mechanism of Interaction of Bovine β -Trypsin with Soybean Trypsin Inhibitor (Kunitz) I. STOPPED FLOW MEASUREMENTS." *Journal of Biological Chemistry* 248.5 (1973): 1760-1771.
26. Jerabek-Willemsen, Moran, et al. "Molecular interaction studies using microscale thermophoresis." *Assay and drug development technologies* 9.4 (2011): 342-353.
27. Salvi, Giovanni, Paolo De Los Rios, and Michele Vendruscolo. "Effective interactions between chaotropic agents and proteins." *Proteins: Structure, Function, and Bioinformatics* 61.3 (2005): 492-499.
28. Ogino, Hiroyasu, et al. "Stabilities and conformational transitions of various proteases in the presence of an organic solvent." *Biotechnology progress* 23.1 (2007): 155-161.
29. Kamatari, Yuji O., et al. "The methanol-induced transition and the expanded helical conformation in hen lysozyme." *Protein science* 7.3 (1998): 681-688.

Chapter 3



Structural features of CaTi_2

In this chapter, we have described the crystallization technique and overall structure of CaTI2. We have attempted to gain insights on the possible mode of inhibition of trypsin by CaTI2 and to get crystal data for CaTI2-BPT complex.

Section 1: Crystal structure determination and analysis of CaTI2

What is meant by the structure of a protein?

When it is asked “what is its structure?” we describe how the various components of any object are placed in real space. We describe its architecture; we describe its shape, its size and certain other similar features. But how do we describe structures that are not visible even with the microscopes? For example, how can one illustrate structures of cellular components?

The first step in answering this question is to make such tiny objects visible. In earlier days microscopic techniques were not advance enough to visualize salts, sugars and biomolecules. However people used their atomic properties to caste their impressions. To build the image and finally the 3D structure, people used Cartesian co-ordinate systems. People had understood that in ground state the atoms would be invariant (1). Thus, in order to build a protein’s inherent structure, scientist chose spatial co-ordinate system and defined each atom’s position along the three imaginary axes. This method became popular because despite the change of origin, the structure would remain same. Each atom was given specific co-ordinate in space $\{x_j, y_j, z_j\}$. The set of such co-ordinates of all atoms in a protein would define the structure of a protein (2). The structure of DNA was determined by Watson and Crick in 1953 on these same principles and the world got to know that DNA exists as a double helix (3). Similarly, the first structure of a protein, myoglobin was determined by X-ray crystallography for which, the noble prize was awarded in 1960 to Max Perutz and John Kendrew in chemistry (4).

In many instances experimental information is not enough to solve the structure of a protein. We often need additional knowledge about the protein, such as its primary amino acid sequence, bond lengths, and bond angles. This information brings about the consistency between collected experimental data and the expected outcome of the atomic model (5). X-ray crystallography, Cryo-electron microscopy and Nuclear magnetic resonance (NMR) spectroscopy are the most popular techniques used for macromolecular structure determination (6). NMR spectroscopy allows us to understand the structure in

solution state under nearly physiological conditions. But the technical difficulty in NMR is the correct and accurate assignment of NMR peaks to underlying atoms (7). Cryo-electron microscopy has got lot of potential but the technique is still in budding condition. As of today, X-ray crystallography is the most widely used technique for macromolecular structure determination. Although its primary requisite is obtaining a protein in crystallized form, it is the only technique that allows precise direct placement of atoms in the protein. With the help of this technique, over 140,000 structures have been deposited in RCSB-PDB database over the past 90 years, and this number is on the rise (8).

The diffraction from a single molecule is too weak to be measurable. Therefore, we use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. A single, well ordered protein crystal contains hundreds of copies of one protein molecule that could give good diffraction pattern, which could lead one to its atomic structure. The whole process is divided into the following steps (Figure 1) (9):

1. **Protein preparation:** To begin with crystallization, we need highly pure, monodisperse and still active protein sample. The sample needs to be highly concentrated (~10mg/ml protein), preferably free of salts. If the sample meets these criteria, one can proceed with crystallization.
2. **Crystallization:** Crystallization, in simple words means removal of excess water molecules from protein in solution form to bring the protein into metastable crystalline form. For this purpose a protein solution is incubated with a buffered precipitant (such as polyethylene glycol or ammonium sulfate) at different temperatures. This could be achieved by using various methods such as vapor diffusion, microdialysis, and micro-batch techniques. If the diffraction is poor, we may have to modify the crystallization condition by changing the pH, or by addition of additives etc. This is usually termed as crystal optimization.
3. **X-ray data collection:** For a selected crystal, the crystal symmetry, resolution range, unit cell parameters and crystal orientation is ascertained during the testing, before proceeding for data collection. The crystal is rotated through a small angle, and the X-ray diffraction pattern data is collected. The diffraction spots should be properly separated. This is repeated until the crystal has moved through at least 30 degrees and sometimes as much as 180 degrees depending on the crystal symmetry. For a crystal with lower symmetry, more data is required.

- Structure solution:** If we already have the coordinates of a similar protein, the structure is solved using Molecular Replacement method. In this method, the model is rotated and translated into the new crystal system until we get a good match to the experimental data. Then, amplitudes and phases are calculated from this solution, which is later combined with the data to produce an electron density map. If a starting model is not available, then the Isomorphous Replacement method is utilized whereby one or more heavy atoms are introduced into the unit cell without perturbing the crystal lattice. Heavy atoms are electron dense and give rise to measurable differences in the intensities of the spots in the diffraction pattern. By measuring these differences for each reflection, an estimate of the phase angle is derived using vector summation methods. A Fourier transform is used to calculate a map.
- Model building and refinement:** In this step, the electron density map is filled and fitted with actual amino acid residues from sequence of our protein. Once residues from the sequence are fitted in the electron density map, refinement of the structure is done for improving the phases which results in better map. It takes several cycles of refinement for final structure solution.

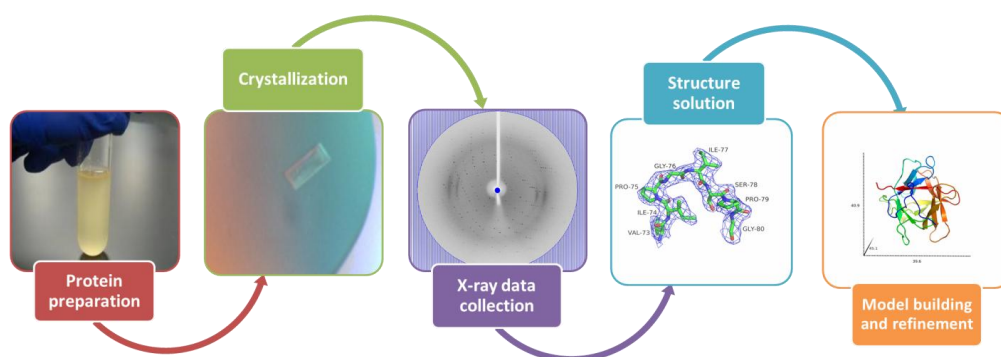


Figure 1: Key steps involved in protein structure determination, starting from protein preparation till final structure solution.

1.1 Materials

All chemical reagents used were purchased from Sigma-Aldrich (India). Readymade commercial crystallization screens were from Hampton research (USA), Qiagen (USA) and Molecular Dimensions (UK). The Micro well SD2 plates, CryoLoops, CryoCanes and other crystallography tools were of Hampton Research (USA) make. The high resolution data were collected at Synchrotron, INDUS-II, BL21 beam, Indore with MAR

CCD detector. To collect diffraction data at low temperature, the crystal was flash cooled in a nitrogen stream produced by X-stream (Rigaku-MS, USA). Various modules of CCP4i (version 7.0.057) and Phenix (version 1.11.1) were used for structure solution. Data integration was attempted using iMOSFLM modules of CCP4i. Merged images were processed using AIMLESS and POINTLESS of CCP4i. Molecular replacement was done by MOLREP of CCP4i suite and Phaser-MR of Phenix suite. Refinement of the generated model was done using REFMAC5 module of CCP4i suite. Electron density and structure was visualized using WinCoot (version 0.8.2). The final model was validated using MolProbity server (<http://molprobity.biochem.duke.edu/>) and RCSB PDB's own validation server (<https://validate-rcsb-2.wwpdb.org>).

1.2 Methodology

1.2.1 Heterologous expression and purification: As per method described in chapter 2, the recombinant CaTI2 inhibitor protein was purified to homogeneity in Tris-Cl buffer pH 7.8, containing 100mM NaCl. The protein was concentrated to the concentration of 8-20mg/ml using Amicon ultra-15ml 10kDa cutoff centrifugal filters (Millipore, USA) at 4°C with 4000rpm. The purity of the protein was checked using SDS-PAGE and the protein was quantified using Bradford reagent.

1.2.2 Protein crystallization and crystal growth: CaTI2 protein was screened against several commercially available crystallization screening kits such as, Crystal screens I & II, PEG/Ion HT, Index, PEGRx HT (Hampton Research) Nextal PACT, CubicPhase I, Protein Complex suite (Quiagen) and JCSG+ (Molecular Dimensions). Screening was performed with Vapor-diffusion sitting-drops containing 300nL of protein and 200nL of screen solution, were set up in 96-well Corning plates containing 50µL screen solution in reservoir well using a Mosquito Crystal Nanolitre protein crystallization robot (TTP Labtech, UK). When crystals were obtained, the crystallization was reproduced using the same condition in a 24 well plate using siliconized coverslips with hanging drop technique. The buffer composition in respective screens was mimicked using manually prepared screen solution. The crystals formed were checked under UV-microscope to differentiate protein crystal from salt crystals.

1.2.3 Data collection and processing: The crystals obtained from the crystal screen were exposed to high intensity X-ray beam to collect diffraction data. When electrons

generated from a heated filament accelerated by high voltage collide with a metal target such as copper anode, X-rays are produced. X-rays are electromagnetic radiation of wavelength about 0.1–10nm (1 -100Å) (10). The selected crystals were soaked in suitable cryoprotectant (based on respective screen composition). A cryoprotectant is the solution that prevents crystal from damage due to extreme low temperatures. When X-rays collide with protein crystals, free radicals are produced that could bring about changes in protein structure, such as breakage of disulphide bonds. Therefore, X-ray diffraction data collection is done at temperatures as low as 100K (11, 12).

Although several crystals of CaTI2 were obtained the structure could be successfully solved with the crystal formed in the screen condition containing 0.1M sodium acetate pH5, 1M ammonium sulfate. The crystals were formed during 30 days of incubation at 20°C in the screen cocktail. A cryoprotectant was prepared using above mother liquor with 20% PEG-400. The crystal was flash cooled under liquid nitrogen. The crystal was diffracted at INDUS-2 synchrotron BEAMLINE PX-BL2, RRCAT, Indore, India (13). Data were collected using CCD RAYONIX MX-225 detector, at 100K temperature. The diffraction data were collected in the form of images (Figure 2). The images were then integrated using iMOSFLM, a graphical interface of program MOSFLM for diffraction data integration (14). iMOSFLM produces a set of indices (*hkl*s) with their associated intensities together with an accurate estimate of the crystal unit-cell parameters from a set of diffraction images. The program works in 3 steps. In first step it indexes the images and determines the unit cell parameters. In second step it refines the unit cell parameters and the mosaicity determined from step one. In the final step, the images are integrated, by predicting the positions of the Bragg reflections on each image and obtaining an estimate of the intensity of each reflection and its uncertainty.

Post data integration, the data was ‘reduced’ using the module ‘POINTLESS’. Here the point-group symmetry and likely space group was determined. In next step the symmetry related measurements were put on common scale and averaged. This was done using ‘AIMLESS’ program, which provides different statistical parameters to measure the data quality and helping in further processing. These parameters include R_{merge} , $I/\sigma I$ etc. It is believed that a good quality, processed data can have R_{merge} of less than 10%. It should be noted that R_{merge} alone is not sufficient to decide the high resolution cut-off since it depends on the redundancy of data (15).

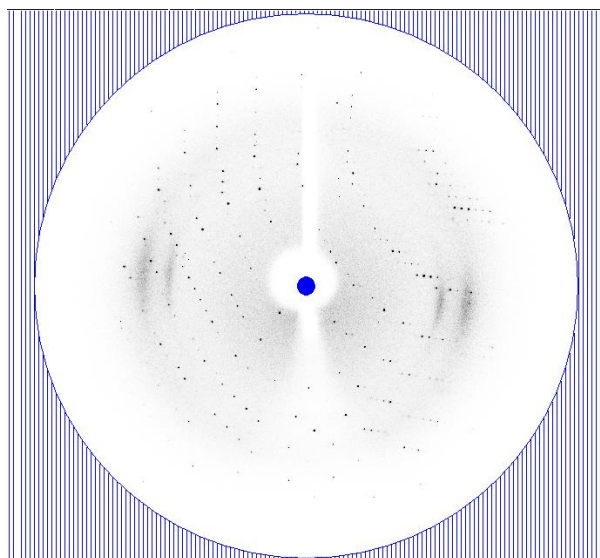


Figure 2: Diffraction pattern of CaTI2 crystal diffracted at PXBL21 beamline, INDUS-II synchrotron, RRCAT, Indore, India.

1.2.4 Matthew's coefficient: A protein crystal does not contain protein molecule alone. It also contains considerable amount of solvent in it. In 1968, Matthews analyzed about 116 crystal forms and observed that the solvent content in different crystals ranged from 27% to 78%. The most observed solvent percentage was about 43. He established a relation between molecular weight and solvent content of a crystal, called Matthews coefficient. The Matthews coefficient (V_M) was defined as the crystal volume per unit of protein molecular weight. It was shown that V_M displays a direct relationship to the fractional volume of solvent in the crystal (16). Using Matthew's coefficient module of CCP4i, it was calculated that the crystal of CaTI2 contained two copies of monomeric inhibitor.

1.2.5 Structure determination: During data collection, the intensities of diffracted waves scattered from a series of possibly imaginary planes passing in all directions through the crystal are collected. The amplitudes of the scattered waves are calculated from the intensities. During the experiment the detector collects only those intensities which hit it. The incoming wave has amplitude and the phase. The information about the phase is lost during the data collection which is of immense importance. This limitation is termed as the 'Phase problem'. This problem could be overcome by several methods such as Multiple Isomorphous Replacement (MIR) - where heavy atoms are inserted into structure, molecular replacement (MR) and Single-wavelength Anomalous Dispersion (SAD). These days, a vast majority of the structures (> 70%), derive the phases using

atomic coordinates of homologous protein structures, by the popular Molecular replacement method (17).

Since the homologous structures from Kunitz legume family were already reported, the approach of molecular replacement was preferred for initial model building. The structure of a Kunitz type inhibitor, DrTI from *Delonix regia* (PDB Id: 1R8N) was found to be the closest among available KTI structures and thus was used as a template for molecular replacement (18). The CaTI2 protein shares about 47% sequence homology with DrTI (Figure 3).

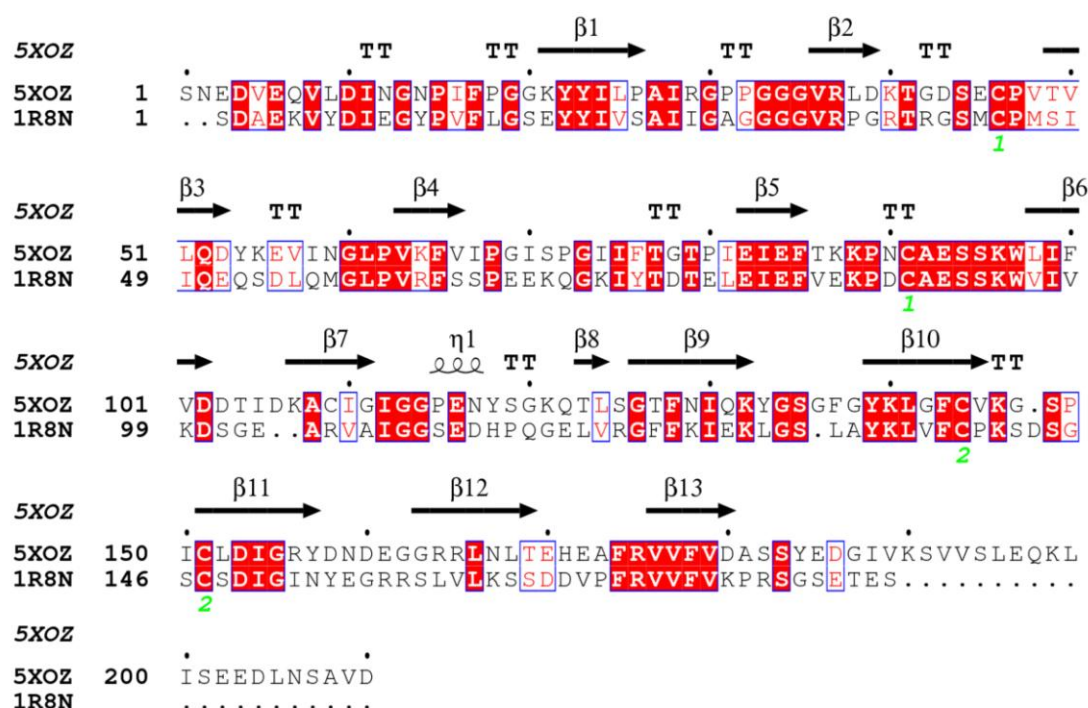


Figure 3: Alignment of CaTI2 (PDB ID: 5XOZ) sequence with DrTI inhibitor (PDB ID: 1R8N). CaTI2 shares maximum sequence identity (~ 47%) with DrTI compared to other KTIs.

Initial model building was attempted using PHASER-MR of Phenix and MOLREP of CCP4i (19). The best model was then refined using REFMAC5 program of CCP4i and Coot (20, 21). Attempts were made to fit almost the residues of CaTI2 in the electron density map during the successive cycles of structure refinement. After every restrained refinement cycle the model was inspected and improved by fitting outlier residues in electron density by manual rebuild using Coot. The model generated at the end of each

cycle of refinement was checked for unusual geometry at each residue, the peptide-flip values, and high values of temperature factors.

1.2.6 Structure validation: Attempts were made to minimize the errors in estimated phases after the repetitive cycles of REFMAC5 and manual rebuilding of the structure. R-factor is a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data. R-factor, also termed as R_{work} is a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data. It is the mathematical representation of the difference between the experimental observations and the calculated values.

$$R = \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$$

Where, F is the structure factor and the sum extends over all reflections measured and their calculated counterparts respectively.



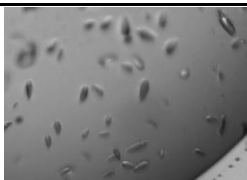
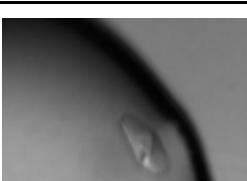

R_{free} is another parameter used to assess the structure quality since R-factor sometimes could be biased (22). Consider an example of a dataset in which there is certain electron density unoccupied. The R-factor would be high initially. The data could be over-fit by introducing water molecules in the electron density map instead of amino acid residues and still the R-factor could come down but it won't mean that the solution to the dataset is correct. The R-value is a measure of how well the simulated diffraction pattern matches the experimentally-observed diffraction pattern. Theoretically R-factor should be 0 but a value of 0.2 is considered good fitting. The free R_{free} value is not only a better statistic but also helps in keeping track of the direction in which structure refinement progresses. A random sample of 5% of total experimental observations (say test set), were set aside before refinement. About 95% in total observations were used for refinement while the test set was used to calculate R_{free} . It is calculated based on how precisely the model predicts the 5% dataset based on 95%. Thus the R_{free} has to be higher than that of the R-factor (not more than 5%) for a well refined data. Ideally if the R-factor is 20%, the R_{free} should not be more than 25%.

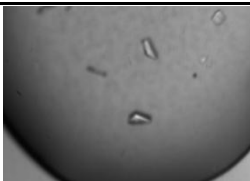


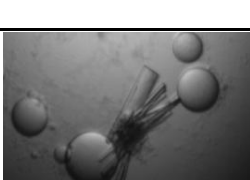
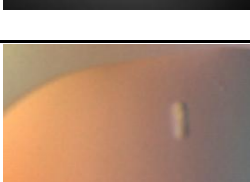

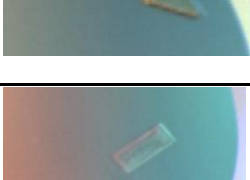

Keeping these factors in mind, the final model was assessed for quality check using MolProbity server and Ramachandran plot analysis, before depositing the final structure in RCSB-PDB (23).

1.3 Results and discussion

CaTI2 crystals were obtained in different crystallization conditions and protein concentrations. Some of these crystals did not diffract, while some died during data collection (Table 1). The best crystal was obtained by the sitting drop method in the F8 condition of Protein Complex screen. The condition composition was 0.1M sodium acetate pH 5, 1M ammonium sulfate. PEG400 (20% w/v) was used as the suitable cryo-protectant for this crystal. The crystal was soaked in mother liquor with cryo-protectant and flash cooled under liquid nitrogen.

Table 1: List of different screen conditions in which CaTI2 crystals were obtained

No	Screen	Condition	Crystal Image
1	Nextal Pact (D8)	0.2M Ammonium chloride 0.1M Tris pH 8 20% (w/v) PEG 6000	
2	Nextal Pact (F6)	0.2M Sodium formate 0.1M Bis Tris propane pH 6.5 20% (w/v) PEG 3350	
3	Nextal Pact (F7)	0.2M Sodium acetate 0.1M Bis Tris propane pH 6.5 20% (w/v) PEG 3350	
4	PEG Rx (C2)	0.1M Imidazole pH 7.0, 20% (v/v) Jeffamine® ED-2001 pH 7.0	
5	CubicPhase (E9)	2.2M Ammonium phosphate pH 7	

6	CubicPhase (G9)	2.2M Ammonium phosphate pH 8.2	
7	Protein Complex (B9)	0.1M Sodium cacodylate pH 6, 15% (w/v) PEG4000	
8	Protein complex (B10)	0.1M MES pH 6, 15% w/v PEG4000	
9	Protein complex (C6)	0.15M ammonium sulphate, 0.1M HEPES pH 7. 20% PEG 4000	
10	Protein complex (E11)	0.1M sodium citrate pH5, 20% PEG 8000	
11	Protein complex (F8)	0.1M sodium acetate pH 5, 1M ammonium sulphate	
12	Protein complex (G5)	1M KCl, 0.1M HEPES pH7, 1M ammonium sulphate	
14	PEG Rx (G1)	0.10% w/v n-Octyl-b-D-glucoside, 0.1M Sodium citrate tribasic dihydrate pH 5.5, 22% (w/v) Polyethylene glycol 3,350	

As mentioned in section 1.2.3, the diffraction data was collected and the structure solution was deposited in RCSB-PDB with PDB ID 5XOZ. The details of data collection and refinement statistics have been tabulated in Table 2.

Table 2: Diffraction data collection and structure refinement parameters

X-ray source (wavelength)	INDUS-II (0.97947 Å)
Space group	P 1 2 ₁ 1
Resolution range	36.77–2.8 Å
Unit cell parameters	$a = 38.73 \text{ Å}$, $b = 88.22 \text{ Å}$, $c = 59.34 \text{ Å}$; $\alpha = \gamma = 90^\circ$, $\beta = 108.32^\circ$
Molecules per asymmetric unit	2 monomers
Matthews coefficient	1.97 Å ³ /Da
Solvent content (%)	38.69%
Total No. of reflections	35,322 (5,183)
No. of unique reflections	9,333 (904)
Multiplicity	3.8 (3.8)
Completeness (%)	98.98 (99.12)
Average I/s(I)	12.0 (8.3)
R-merge (%)	0.057 (0.101)
Structure refinement	
R-factor	0.1941
R-free	0.2124
Number of atoms	
Protein	21,696
Water	0
Average B-factors	
Chain A	30.362
Chain B	30.867
Ramachandran plot (%)	
Most favorable + additional allowed region	94.62 + 5.38
Disallowed region	0

1.3.1 Overall structure of CaTI2: In the asymmetric unit of CaTI2, two molecules were observed forming a crystallographic dimer. The two molecules were arranged along the imaginary axis of rotation as shown in figure 4.

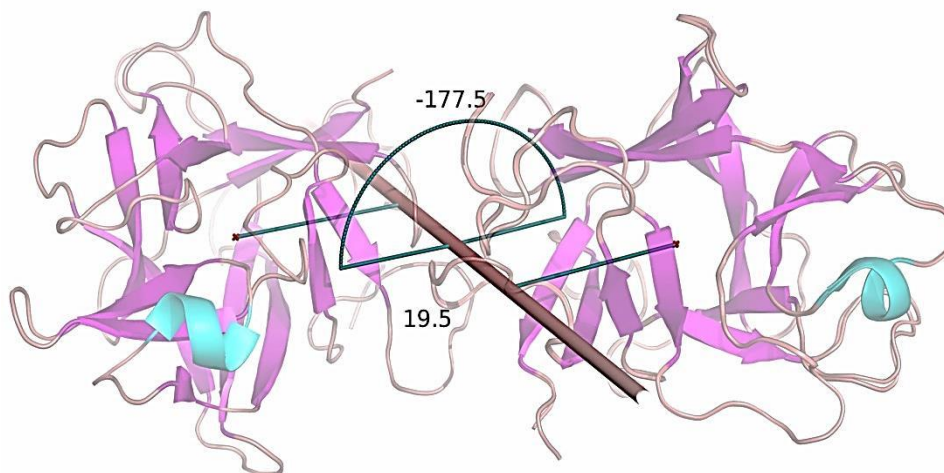


Figure 4: An axis of rotation with two units of CaTI2 in asymmetric unit forming a crystallographic dimer. The structure co-ordinates have been deposited in RCSB-PDB under ID: 5XOZ

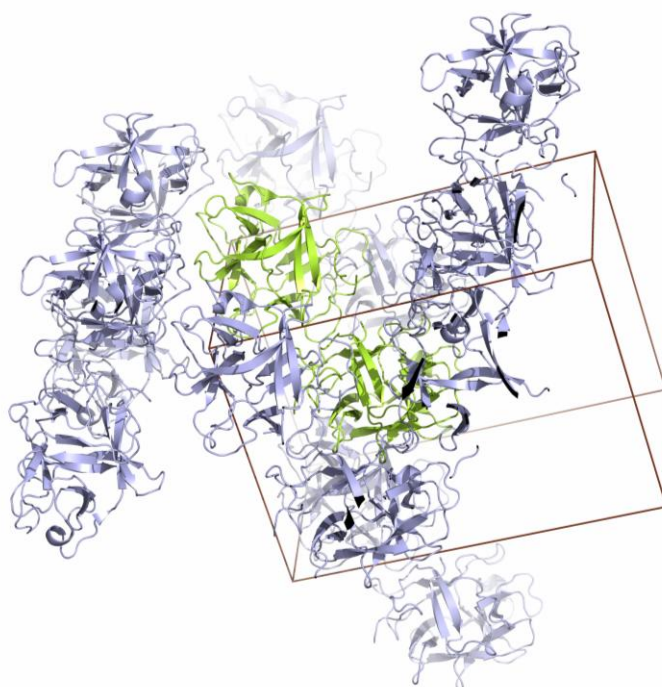


Figure 5: Packing of CaTI2 molecules inside a unit cell. A unit cell contains two copies of CaTI2

The approximate dimensions of the protein are 45.11 x 40.87 x 39.6Å. Since the crystals of CaTI2 contain two molecules in the asymmetric unit, it is possible to evaluate the influence of crystal packing on its structure (Figure 5). As observed in all KTIs, CaTI2 also displays β -trefoil fold (Figure 6). β -trefoil is also seen in various proteins such as cytokines, agglutinins, ricin B-like lectins, fibroblast growth factors, interleukins, tetanus, and botulinum neurotoxins (24, 25). The fold has the arrangement of twelve antiparallel β strands (labeled β 1 to β 12) arranged into three structurally similar units related by pseudo-threefold symmetry.

The characteristic feature of this fold is that most of the β regions are well conserved. The main variations are observed in the loop regions. A β barrel creates a significant hydrophobic core. The root mean square deviation (RMSD) between the coordinates of the two molecules is 0.141Å, calculated for 145 pairs of superimposed C α atoms. The coordinates for CaTI2 structure have been deposited in the PDB under the accession codes 5XOZ.

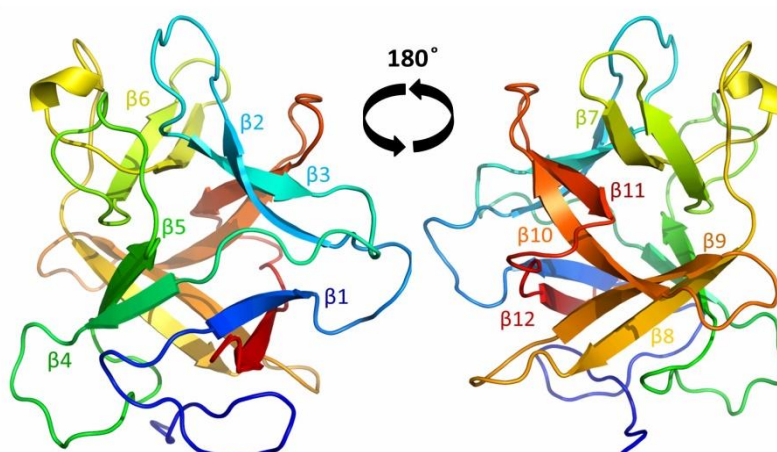


Figure 6: Cartoon representation of secondary structural elements in CaTI2 color-ramped from N- to C-terminus from the front and rotated 180°.

Various interactions that hold the quaternary conformation in place such as hydrogen bonds, ionic interaction networks were calculated using iRDP server have been listed in table 3 and 4. Like other KTIs, CaTI2 also contains 5 Cys residues which are highly conserved. Out of the five Cys residues, four are involved in disulfide bridge formation as listed in table 5.

Table 3: List of intra-chain hydrogen bonds (Cutoff 3Å). Bond type abbreviations: S- Side chain, M- Main chain

Sr. No.	Donor residue	Acceptor residue	Bond type	Distance (Å)
1	Lys47	Asp46	MS	2.87
2	Lys62	Asp60	MS	2.78
3	Thr111	Asp109	MS	2.78
4	Asp167	Asp165	MS	3
5	Ser189	Asp187	MS	2.68
6	Arg163	PRO39	SM	2.84
7	Arg171	PRO39	SM	2.8
8	Lys95	ALA99	SM	2.7
9	Arg182	PHE144	SM	2.49
10	Arg163	GLN59	SS	2.85
11	Arg171	GLN59	SS	2.75
12	Arg182	SER142	SS	2.8
13	Tyr61	ASP165	SS	2.34
14	Asn166	GLU168	SS	2.94
15	Tyr61	ASP165	SS	2.34
16	Cys53	Ser51	MS	2.99
1	Gly80	Ser78	MS	2.82
18	His178	Thr176	MS	2.88
19	Gln59	Gly41	SM	2.91
20	Ser51	Cys53	SM	2.68
21	Thr86	Phe83	SM	2.94
22	Sre102	Phe92	SM	2.67
23	Tyr125	Ile119	SM	2.72
24	Tyr140	Ala180	SM	2.81
25	Ser155	Val152	SM	2.76
26	Asn174	Leu173	SM	2.99

Table 4: List of Ion-pair interactions observed in native structure of CaTI2 (Cut off 4Å)

Sr. No.	Residue 1	Residue 2	Distance (Å)
1	Asp46	Arg44	3.1
2	Asp60	Arg44	3.23
3	Asp 109	Arg172	2.86
4	Asp 109	Arg172	3.09
5	Asp160	Lys147	3.16
6	Asp160	Lys147	2.78
7	Glu168	Arg172	3.43
8	Glu168	Arg172	2.83
9	Glu89	Lys103	3.17
10	Glu89	Lys103	3.53
11	Glu91	Lys71	3.05

Table 5: Out of five Cys residues in CaTI2, four are engaged in disulfide bond formation as per list

Sr. No.	Residue 1	Residue 2	Distance (Å)
1	Cys53	Cys98	2.03
2	Cys151	Cys158	2.05

1.3.2 Comparison with homologous structures: The CaTI2 structure was compared with other KTI structures using the ConSurf server, and significant conservation was observed within the β -strand regions, while loop regions are highly variable (figure 7) (26).

This variation could be due to insertion or deletion of the residues in loops. Like other KTIs, CaTI2 also lacks any significant α helices. The STI has reactive loop between 4th and 5th β strands. The FFT (Fast Fourier Transform) program was used to calculate Fouriers, difference Fouriers, double difference Fouriers to ensure the overall placement of inhibitory loop (Figure 8). The loop consists of about 7 amino acid residues viz. Ser60-Pro-Tyr-Arg-Ile-Arg-Phe66. Kink producing Pro residues are seen near the start and the end of the loop. The P1 Arg63 is flanked by Tyr62 at P1' and Ile64 at P2 position (27).

The CaTI2 has corresponding loop between residues 77 and 88. The Arg63 in STI is replaced with Ile77 in CaTI2.

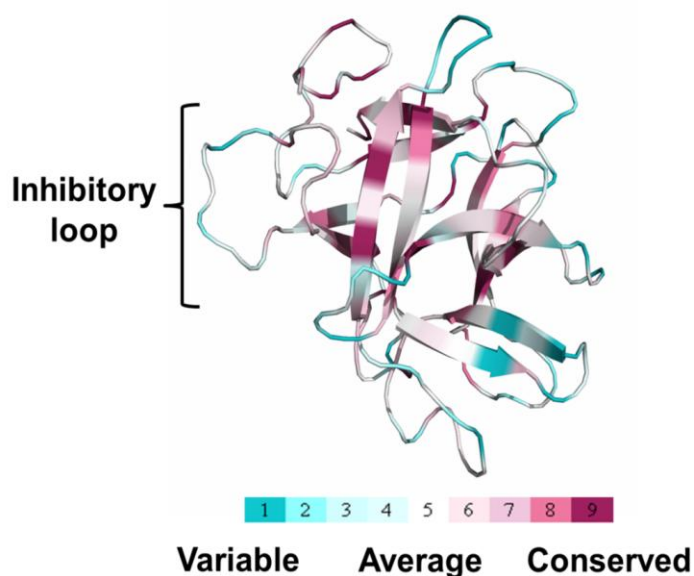


Figure 7: ConSurf analysis result of amino acid wise conservation representation. Blue color represents highly variable amino acid while magenta represents highly conserved

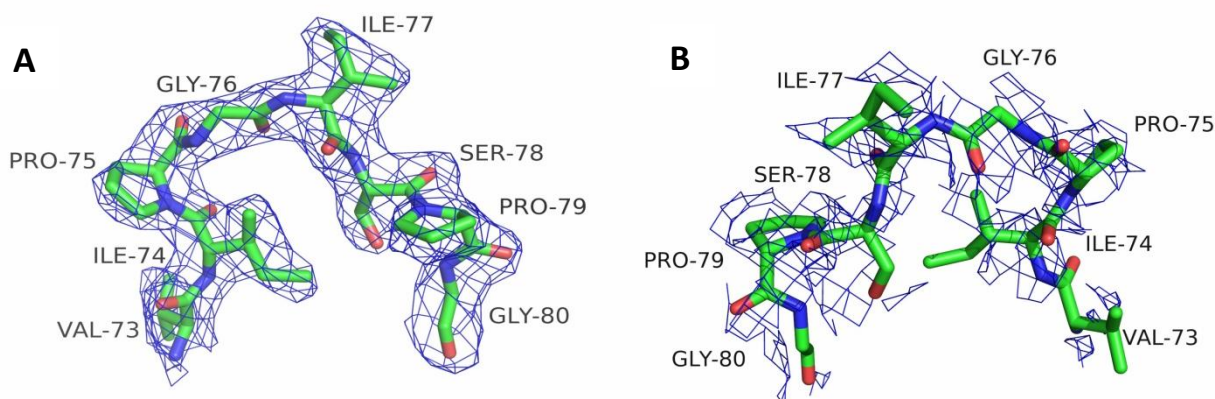


Figure 8: FFT map of inhibitory loop. **A)** 2Fo-Fc **B)** Fo-Fc

The catalytic triad of His57, Asp102, and Ser195, in the active pocket of trypsin is responsible for cleavage in the substrate protein at the carbonyl oxygen of Arg/Lys residues (28). The proteinaceous inhibitor binds to active site pocket of serine protease, mimicking substrate binding, i.e. the Arg/Lys residue from inhibitor interacts with Ser195 of trypsin just like substrate proteins, which slows down the hydrolysis of the inhibitor

peptide about 10^7 times in comparison to substrate protein binding (29). Hydrogen bonds formed on either sides of scissile Arg63 residue stabilize the cleaved products of inhibitor together in active pocket of trypsin resulting in enzyme inhibition (30). In typical KTIs, the conserved Asn occupies the 13th position (STI numbering) which helps in re-ligation of cleaved inhibitor (31). But Asn13 in CaTI2 is replaced by Pro25. This highlights the fact that the absence of Arg residue in the inhibitory loop of CaTI2 leads it to its binding to trypsin in a non-substrate like manner; thus, conserved Asn13 might have evolved to Pro25.

Arg63 is surrounded by a highly electropositive environment. Similar pattern is observed in β regions of CaTI2 but it is less prominent as compared to STI (Figure 9). Another notable observation is that the corresponding reactive residue, Ile77 seems to be electro-neutral. The electrostatic surface charge distribution pattern of STI revealed that the core β region appears to be highly electronegative (Figure 10).

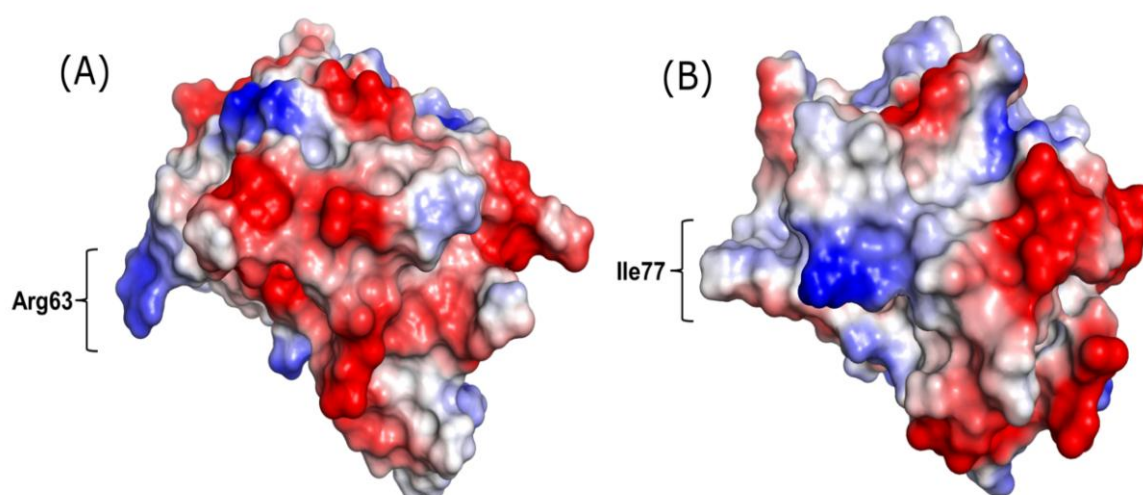


Figure 9: Distribution of surface charge. Blue color indicates electropositive region while Red color indicates electronegative region. **(A):** In STI, the core region appears highly electronegative with P1 Arg as highly electropositive. **(B):** The CaTI2 does not follow any particular trend in surface charge distribution. The predicted P1 Ile77 appears neutral.

RMSD was calculated for CaTI2 with four established structures (Table 6) to estimate the structural similarity among the known KTI structures.

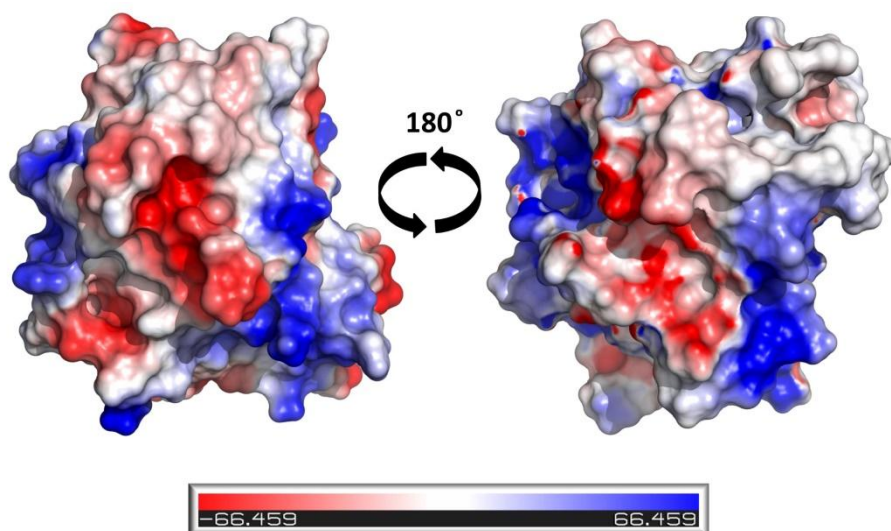


Figure 10: Surface charge distribution representation in CaTI2

Table 6: C_{α} RMSD values of CaTI2 with structures of STI, ETI, EcTI and TKI for complete native structure and for the respective inhibitory loops regions

Protein (PDB Id)	C_{α} RMSD (Å)	
	Complete structure	Inhibitory loop region
STI (1AVU)	1.68	2.16
ETI (1TIE)	2.88	2.25
EcTI (4J2K)	3.99	1.4
TKI (4AN6)	1.53	1.61

It can be observed that the C_{α} RMSD is relatively less when compared with STI and TKI than that with ETI and EcTI (32, 33, 34, 35). Also, RMSD in the region of inhibitory loops is very high for all four KTIs which directly implies that the variations in sequence at inhibitory loop regions of CaTI2 is reflecting in its overall geometry and dynamics of binding conformation (Figure 11).

Since the soybean inhibitor, STI was reported to be allergenic in nature (36), CaTI2 was checked for presence of any structural domains that could be allergenic in nature. Using AlgPred server it was confirmed that the CaTI2 structure did not contain allergenic domains. The approach of IgE epitope search and HMMER-MAST search was used for prediction (37).

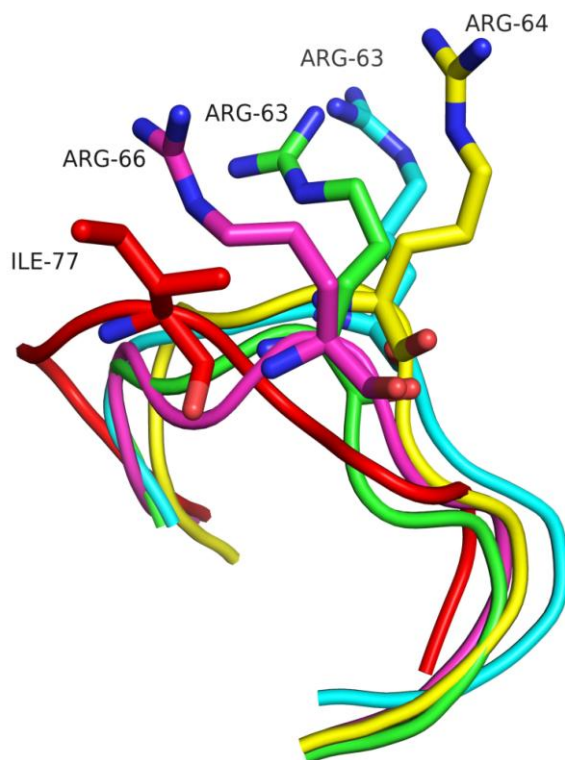


Figure 11: Superposition of inhibitory loops from 4 canonical KTIs and CaTI2. Red: CaTI2, Pink: TKI (4AN6), Green: STI (1AVU), Blue: ETI (1TIE), Yellow: EcTI (4J2K).

Section 2: Assessment of interactions between CaTI2 and Trypsin

Enzyme-inhibitor complexes are considered as one of the model systems for understanding protein–protein interactions (38). There are very few such complex structures available. When there is lack of complex crystal structures, one can study protein-protein interactions by docking available crystal structures and predict the possible interactions. Such studies help in predicting the interface regions between the two proteins and thus provide useful information about the association of the two molecules.

Attempts have been made to deduce the mode of action of various inhibitors and establish the mechanism of the trypsin inhibition. Based on biochemical and biophysical studies it was known that usually a particular element of the inhibitor protein structure binds to trypsin active pocket and bring about its inhibition (39). Inhibitors from different families may have different residues involved in the inhibition process, but usually the underlying mechanism of inhibition of serine proteases is the same. The inhibitory loop from the inhibitor binds to the trypsin reactive pocket just like the manner in which substrate

protein binds to trypsin during cleavage. The Bowman Birk and Pin II family inhibitors contain two inhibitory loops. In the first inhibitory loop of Bowman Birk inhibitor (e.g. PsTI from *Pisum sativum*), Lys residue interacts with Ser195 from trypsin, while in case of Pin II inhibitors (e.g. *Solanum lycopersicum* inhibitor II), Arg residue at P1 position in loop 1 interacts with a nucleophilic Serine and a network of hydrogen bonds between trypsin active site pocket and the inhibitory loop, conferring overall inhibition (40, 41).

According to the mechanism proposed by Radisky E. S. et al, the inhibitor binds to trypsin just like the cognate substrate (Figure 12) (29). It is hypothesized that trypsin hydrolyzes the peptide backbone of inhibitor. The nucleophilic Ser195 of trypsin attacks the P1 residue in the inhibitor which resembles the substrate. A Michaelis complex is formed which is converted to an acyl-enzyme complex. At this stage enzyme hydrolyzes the inhibitor at P1 residue; but the rate of hydrolysis is extremely slow such that the reaction appears to be static and the inhibitor does not leave the trypsin active pocket. The resulting acyl-enzyme complex appears to be stable and prevents trypsin from accessing other substrate proteins to cleave.

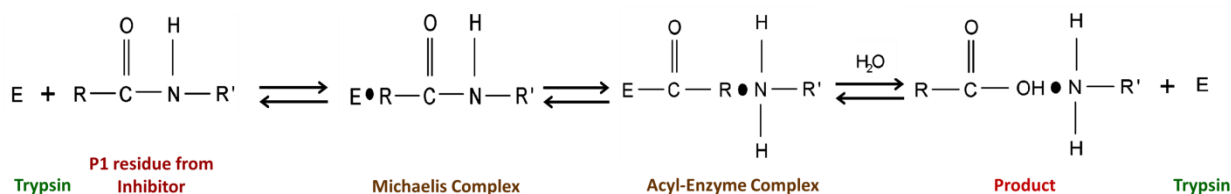


Figure 12: Reaction mechanism of trypsin inhibition by canonical KTIs.

In the current study we have attempted to analyze the CaTI2-trypsin complex by molecular docking and enzyme-inhibitor complex crystallization.

2.1 Methodology

2.1.1 Molecular docking of CaTI2 with BPT: The structure coordinates for bovine pancreatic trypsin were downloaded from RCSB PDB database (PDB ID 4I8L) (42). Using protein preparation utility of Maestro 10.4 (Schrodinger ver15.3), hydrogen atoms were added to both CaTI2 and trypsin structures to carry out restrained minimization (43). The root mean square deviation (RMSD) of the atomic displacement for terminating the minimization was set as 0.3Å. The rigid body docking of processed structures were performed using program ZDock 3.0.2 (44). Among several model complexes generated

by ZDock, the top 10 highest ranking poses were selected to study the enzyme inhibitor interactions and RMSD was calculated (Table 7). The resultant models were compared with complex structures available and the best pose was chosen based on inter-chain interactions observed. The protein-protein interaction parameters were calculated using iRDP server and CoCoMaps tools (45, 46).

Table 7: RMSD values of top models of docking

Docked model	RMSD (Å)
Complex.0	0 (Stationary model)
Complex.1	0.02
Complex.2	0.573
Complex.3	0.68
Complex.4	0.782
Complex.5	0.787
Complex.6	0.812
Complex.7	0.719
Complex.8	0.692
Complex.9	0.767

2.1.2 Crystallization of CaTI2 and BPT complex: Equimolar concentrations of CaTI2 and BPT were mixed together and allowed to complex at room temperature for about 30 minutes. The complex mixture was purified using gel filtration chromatography using a PrintEnrich TMSEC 650 column. As mentioned in section 1.2.2, Using different commercial screens, the crystallization was set up in Vapor-diffusion sitting-drops containing 300nL of protein and 300nL of screen solution, in 96-well Corning plates containing 50µL screen solution in reservoir well using a Mosquito Crystal Nanolitre protein crystallization robot (TTP Labtech, UK).

2.2 Results and Discussion

Since the amino acid composition in the CaTI2 inhibitory loop is different from that of STI, it was necessary to study the inter-chain interactions displayed by the P1 residue of the inhibitor when it is bound to the active site pocket of trypsin. As of now, structures for 3 KTIs in complex with the protease are available (PDB Ids.: 1AVW, 4AN7 and 4J2Y)

(32, 34, 35). Among them the first two i.e. STI and TKI are complexed with porcine pancreatic trypsin, while EcTI is complexed with BPT. The P1 Arg in all three complexes was found to be involved in extensive network of hydrogen bonds in the active site pocket of the enzyme, primarily with the nucleophilic Ser195. From sequence alignment only it is difficult to predict the P1 residue. Also, the absence of Arg or Lys in the inhibitory loop region made it necessary to obtain deeper information about the cross talk between trypsin and CaTI2. We carried out molecular docking with the two molecules as per the methodology discussed above. It was observed that in all poses, the predicted inhibitory loop fitted into the reactive center cavity of BPT as seen in figure 13.

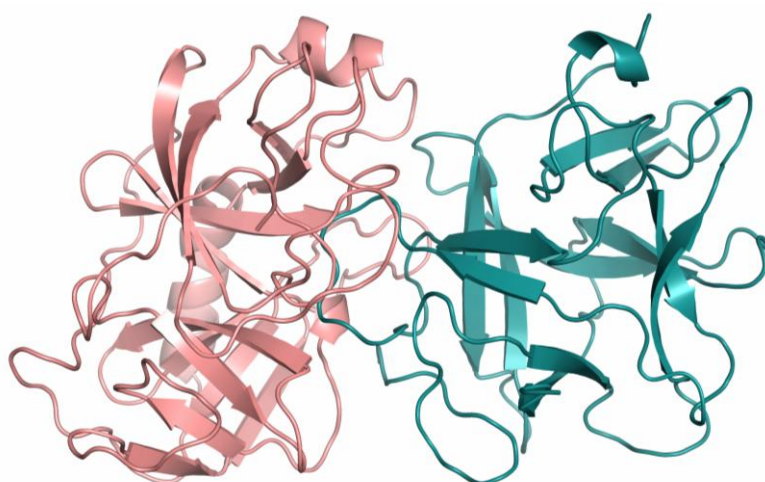


Figure 13: A complex of CaTI2 (Blue) and trypsin (Pink). The inhibitory loop of the CaTI2 protrudes inside the active pocket of the trypsin.

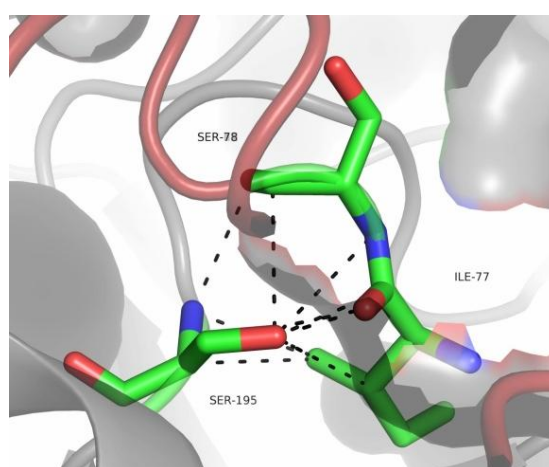


Figure 14: Interactions between Ser195 from trypsin and Ile77, Ser78 from CaTI2 within 3.5Å

Our docking experiments showed that Ile77 from CaTI2 formed hydrogen bond with Ser195 of BPT. In addition, Ser78 of CaTI2 also formed hydrogen bonds with Ser195 (Figure 14). This suggests the absence of Arg residue in P1 position does not mimic the substrate for trypsin and thus the re-ligation step is omitted. Overall inter-chain interactions between trypsin and CaTI2 make the complex stable.

The distance between the residues of the CaTI2 and the trypsin were plotted along with the respective different non-covalent interactions (Figure 15). We found that the predicted inhibitory loop region was found to be in close contact with the trypsin active pocket region. When non-covalent interactions were considered, the residues ~73-79 from the inhibitory loop region were showing interaction with the region surrounding the Ser195 of the trypsin.

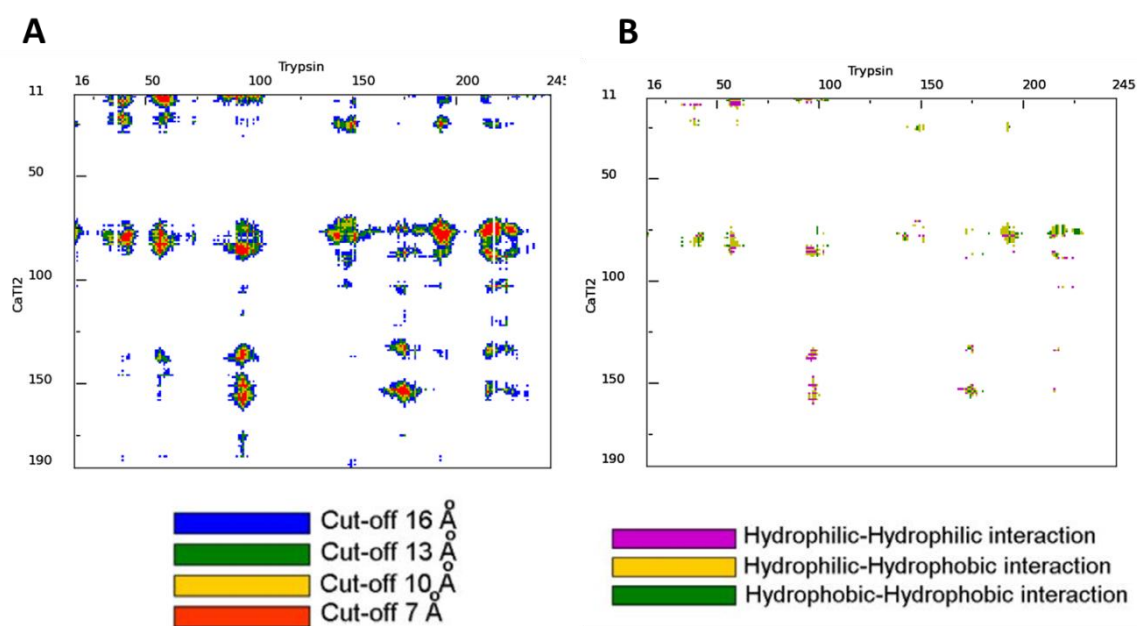


Figure 15: (A) Distance between the inter-chain residues. (B) Non covalent interactions between inter-chain residues.

The shape complementarity between the members of the protein complexes has been used to evaluate the strength of the complex structures (47). The interface area was calculated for both STI-BPT and CaTI2-BPT complexes using the method illustrated by C. Chothia (48). The interface area for the STI-BPT complex was found to be about 870.2Å and the same for CaTI2-BPT complex was calculated to be 1222.35Å. This indicated that the

CaTI2 might be forming stronger complex with trypsin as compared to STI since the interface area for CaTI2-BPT complex is much larger. The actual structure of the complex would throw more light on this observation. Therefore, we tried to solve the structure of CaTI2-BPT complex crystal as per method described in section 1.2.3. As mentioned in section 2.1.2, the CaTI2-BPT complex was purified and visualized on silver stained SDS-PAGE gel. The complex protein from FPLC fraction was loaded on SDS-PAGE with and without β -mercaptoethanol, heat treatment. When treated with heat and β -mercaptoethanol, the complex sample showed two distinct bands, one near 25kDa corresponding to CaTI2 and one near 21kDa corresponding to BPT as seen in lane 2 of figure 16(C). When the sample was loaded without any treatment, it showed three bands as seen in lane 3. The band near 40kDa corresponds to complex while two bands of CaTI2 and BPT were observed near 25kDa and 21kDa respectively.

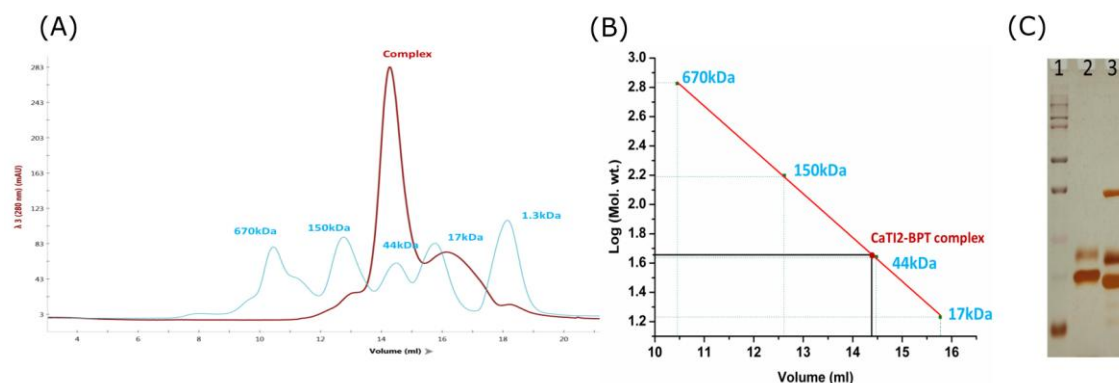


Figure 16: (A) Gel filtration profile of CaTI2-BPT complex in brown color. Sky blue colored profile indicates the peaks of proteins present in molecular weight marker. (B) Plot of logarithm of molecular weight against volume eluted. The molecular weight of complex was found to be about 45kDa. (C) FPLC fraction of purified complex on gel. Lane 1: Bio-Rad dual color marker. Lane 2: Complex fraction treated with β -mercaptoethanol and heat. Lane 3: Complex fraction without heat and β -mercaptoethanol treatment.

Four crystals of the CaTI2-BPT complex were obtained in screen conditions mentioned in table 8. The attempts were made to collect diffraction data for all four crystals. But three crystals died during data collection due to radiation damage. We could collect data for the crystal obtained from the screen condition, 1.8M Ammonium sulfate, 0.1M BIS-TRIS pH 6.5, 2% v/v PEG monomethyl ether 550 at RRCAT synchrotron facility. The structure solution work is still on going.

Table 8: List of screen conditions in which CaTI2-BPT complex crystals were obtained

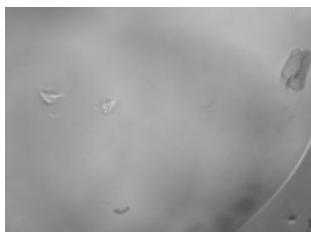
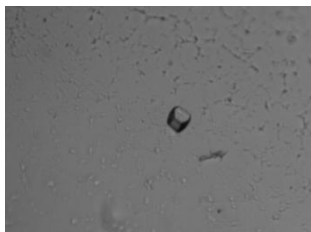
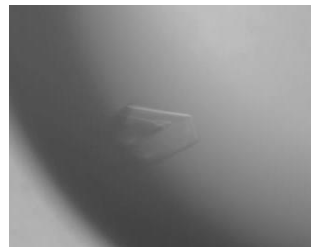
Sr. No	Screen	Condition	Crystal images	Diffraction quality
1	PEG-Rx (D3)	0.1M BIS-TRIS propane pH 9.0, 30% (w/v) EG 6,000 (2 crystals)		Died during data collection
2	Protein Complex (C7)	0.1M Sodium Citrate pH 5.6, 20% (w/v) PEG 4,000, 20% (v/v) Isopropanol		Died during data collection
3	PEG-Rx (E11)	1.8M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 2% (v/v) PEG monomethyl ether 550		2.9Å

Table 9: Preliminary diffraction data statistics of CaTI2-BPT complex crystal

Protein	CaTI2-BPT
Space Group	I4
Resolution (Å)	48.53-2.9
X-Ray Source	Indus II
Temperature	100K
Wavelength	0.97947

Unit Cell Parameters (Å)	$a = b = 165.66, c = 291.16 \alpha = \beta = \gamma = 90^\circ$
Matthews Coeff.	2.66 Å ³ /Da
Solvent Content	47.94
Observed Reflections	200836(6378)
Unique Reflections	43292 (1392)
R _{merge} (All I+ & I-)	0.196
Mean I/σ(I)	11.9 (9.8)
Completeness	99% (98.7%)
Multiplicity	4.6 (4.6)

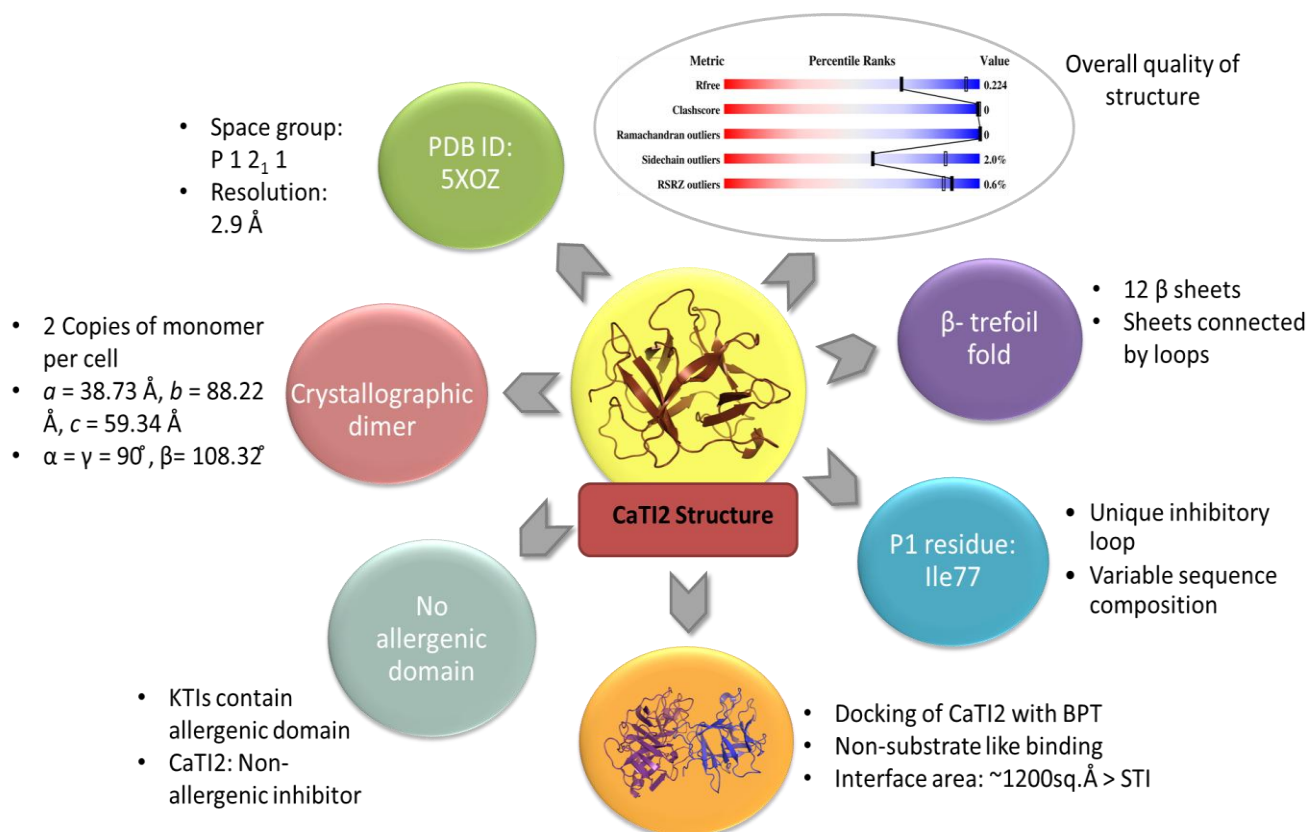
The overall amino acid composition, inhibitory loop organization and the findings of the docking studies strongly suggest that the CaTI2 must be binding to the trypsin reactive pocket in a non-substrate like manner, plausibly through a novel mode of inhibition of serine protease.

Note: The work presented in this chapter has been published in the following research article-

Ameya D. Bendre, Dhanasekaran Shanmugam, C. G. Suresh, and Sureshkumar Ramasamy. "Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from *Cicer arietinum L.*", Journal of Biomolecular Structure & Dynamics (In press). DOI: 10.1080/07391102.2018.1494633

Schematic summary

Following is the schematic representation which summarizes the key findings of this chapter.



References

1. Keating, P. N. "Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure." *Physical Review* 145.2 (1966): 637.
2. Ladd, Marcus Frederick Charles. *Symmetry in molecules and crystals*. Halsted Press, 1989.
3. Watson, James D., and Francis HC Crick. "The structure of DNA." *Cold Spring Harbor symposia on quantitative biology*. Vol. 18. Cold Spring Harbor Laboratory Press, 1953.
4. Kendrew, John C. "Myoglobin and the structure of proteins." *Science* 139.3561 (1963): 1259-1266.
5. Blundell, Tom L., and Louise N. Johnson. "Protein crystallography." (1976).
6. McPherson, Alexander. "Crystallization of biological macromolecules. Vol. 586. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1999.
7. Bax, Ad. "Two-dimensional NMR and protein structure." *Annual Review of Biochemistry* 58.1 (1989): 223-256.
8. Kouranov, Andrei, et al. "The RCSB PDB information portal for structural genomics." *Nucleic Acids Research* 34.suppl_1 (2006): D302-D305.
9. Rupp, Bernhard. "Biomolecular crystallography: principles, practice, and application to structural biology".
10. Garland Science, 2009. Zachariasen, W. H. "A general theory of X-ray diffraction in crystals." *Acta Crystallographica* 23.4 (1967): 558-564.
11. Ravelli, Raimond BG, and Elspeth F. Garman. "Radiation damage in macromolecular cryocrystallography." *Current Opinion in Structural Biology* 16.5 (2006): 624-629.
12. Ravelli, Raimond BG, et al. "Specific radiation damage can be used to solve macromolecular crystal structures." *Structure* 11.2 (2003): 217-224.
13. Kumar, Ashwani, et al. "Protein crystallography beamline (PX-BL21) at Indus-2 synchrotron." *Journal of Synchrotron Radiation* 23.2 (2016): 629-634.
14. Powell, Harold R., Owen Johnson, and Andrew GW Leslie. "Autoindexing diffraction images with iMosflm." *Acta Crystallographica Section D* 69.7 (2013): 1195-1203.
15. Weiss, M. S., and R. Hilgenfeld. "On the use of the merging R factor as a quality indicator for X-ray data." *Journal of Applied Crystallography* 30.2 (1997): 203-205.

16. Matthews, Brian W. "Solvent content of protein crystals." *Journal of Molecular Biology* 33.2 (1968): 491-497.
17. Rossmann, MICHAEL G. "The molecular replacement method." *Acta Crystallographica Section A* 46.2 (1990): 73-82.
18. Krauchenco, Sandra, et al. "Crystal structure of the Kunitz (STI)-type inhibitor from *Delonix regia* seeds." *Biochemical and Biophysical Research Communications* 312.4 (2003): 1303-1308.
19. Vagin, Alexei, and Alexei Teplyakov. "MOLREP: an automated program for molecular replacement." *Journal of Applied Crystallography* 30.6 (1997): 1022-1025.
20. Murshudov, Garib N., et al. "REFMAC5 for the refinement of macromolecular crystal structures." *Acta Crystallographica Section D: Biological Crystallography* 67.4 (2011): 355-367.
21. Lohkamp, Bernhard, Paul Emsley, and Kevin Cowtan. "Coot news." CCP4 Newsletter 42 (2005): 3-5.
22. Free, R. "value: a novel statistical quantity for assessing the accuracy of crystal structures Brunger AT." *Nature* 355.6359 (1992): 472-5.
23. Chen, Vincent B., et al. "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallographica Section D: Biological Crystallography* 66.1 (2010): 12-21.
24. Hatakeyama, Tomomitsu, et al. "C-type lectin-like carbohydrate recognition of the hemolytic lectin CEL-III containing ricin-type β -trefoil folds." *Journal of Biological Chemistry* 282.52 (2007): 37826-37835.
25. Murzin, Alexey G., Arthur M. Lesk, and Cyrus Chothia. " β -trefoil fold: patterns of structure and sequence in the kunitz inhibitors interleukins-1 β and 1 α and fibroblast growth factors." *Journal of Molecular Biology* 223.2 (1992): 531-543.
26. Ashkenazy, Haim, et al. "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids." *Nucleic Acids Research* 38.suppl_2 (2010): W529-W533.
27. Song, Hyun Kyu, and Se Won Suh. "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." *Journal of Molecular Biology* 275.2 (1998): 347-363.

28. Leung, Donmienne, Giovanni Abbenante, and David P. Fairlie. "Protease inhibitors: current status and future prospects." *Journal of Medicinal Chemistry* 43.3 (2000): 305-341.
29. Radisky, Evette S., and Daniel E. Koshland. "A clogged gutter mechanism for protease inhibitors." *Proceedings of the National Academy of Sciences* 99.16 (2002): 10316-10321.
30. Dattagupta, Jiban K., et al. "Refined crystal structure (2.3Å) of a double-headed winged bean α -chymotrypsin inhibitor and location of its second reactive site." *Proteins: Structure, Function, and Bioinformatics* 35.3 (1999): 321-331.
31. Dasgupta, Jhimli, et al. "Spacer Asn determines the fate of Kunitz (STI) inhibitors, as revealed by structural and biochemical studies on WCI mutants." *Biochemistry* 45.22 (2006): 6783-6792.
32. Song, Hyun Kyu, and Se Won Suh. "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." *Journal of Molecular Biology* 275.2 (1998): 347-363.
33. Onesti, Silvia, Peter Brick, and David M. Blow. "Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina coffra* seeds." *Journal of Molecular Biology* 217.1 (1991): 153-176.
34. Zhou, Dongwen, et al. "Crystal structures of a plant trypsin inhibitor from *Enterolobium contortisiliquum* (EcTI) and of its complex with bovine trypsin." *PloS one* 8.4 (2013): e62252.
35. Patil, Dipak N., et al. "Structural basis for dual inhibitory role of tamarind K unitz inhibitor (TKI) against factor Xa and trypsin." *The FEBS journal* 279.24 (2012): 4547-4564.
36. Moroz, Leonard A., and William H. Yang. "Kunitz soybean trypsin inhibitor: a specific allergen in food anaphylaxis." *New England Journal of Medicine* 302.20 (1980): 1126-1128.
37. Saha, Sudipto, and G. P. S. Raghava. "AlgPred: prediction of allergenic proteins and mapping of IgE epitopes." *Nucleic acids research* 34.suppl_2 (2006): W202-W209.
38. Jones, Susan, and Janet M. Thornton. "Principles of protein-protein interactions." *Proceedings of the National Academy of Sciences* 93.1 (1996): 13-20.

39. Luthy, James A., et al. "Detailed Mechanism of Interaction of Bovine β -Trypsin with Soybean Trypsin Inhibitor (Kunitz) I. STOPPED FLOW MEASUREMENTS." *Journal of Biological Chemistry* 248.5 (1973): 1760-1771.
40. de la Sierra, Inés Li, et al. "Dimeric crystal structure of a Bowman-Birk protease inhibitor from pea seeds." *Journal of Molecular Biology* 285.3 (1999): 1195-1207.
41. Barrette-Ng, Isabelle H., et al. "Structural basis of inhibition revealed by a 1: 2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg." *Journal of Biological Chemistry* 278.26 (2003): 24062-24071.
42. Liebschner, Dorothee, et al. "On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin." *Acta Crystallographica Section D: Biological Crystallography* 69.8 (2013): 1447-1462.
43. Schrodinger, Schrödinger Release. "4: Maestro, version 10.4." Schrödinger, LLC, New York, NY (2015).
44. Pierce, Brian G., Yuichiro Hourai, and Zhiping Weng. "Accelerating protein docking in ZDOCK using an advanced 3D convolution library." *PloS one* 6.9 (2011): e24657.
45. Panigrahi, Priyabrata, et al. "Engineering proteins for thermostability with iRDP web server." *PloS one* 10.10 (2015): e0139486.
46. Vangone, Anna, et al. "COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes." *Bioinformatics* 27.20 (2011): 2915-2916.
47. Lawrence, Michael C., and Peter M. Colman. "Shape complementarity at protein/protein interfaces." *Journal of Molecular Biology* (1993): 946-950.
48. Poulos, Thomas L., et al. "The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450." *Journal of Biological Chemistry* 260.30 (1985): 16122-16130.

Chapter 4



Summary and conclusion

Section 1: Summary

Kunitz inhibitors (KTIs) from both plants and animal origin hold great agronomical and pharmaceutical importance which has been highlighted in the introduction of the thesis. The inhibitor which has been characterized in this study in detail is CaTI2, which belongs to the Kunitz legume superfamily of protease inhibitors, members of which are structurally related but display diversity in their primary structure. However, key conservations in the inhibitory loop region are noticeable. KTIs are known to inhibit a broad spectrum of serine proteases. Main objective of our investigations was to understand the structure-function relation of a KTI from *C. arietinum L* i.e. chickpea, that has got unique primary structure. The study involved biochemical, biophysical, structural and computational analysis to understand static as well as in solution characteristics of CaTI2 structure.

We analyzed the sequence alignment with known KTIs and found that most of the regions that were conserved in canonical KTIs were conserved in CaTI2 too, but functionally important residues were highly variable in case of CaTI2 and CaTI3. The conclusions made from sequence analysis reflected in phylogenetic analysis of KTIs. It showed that all canonical KTIs with conserved inhibitory loops cluster together in the same clade, while unique KTIs were segregated from the canonical KTIs.

CaTI2 is an important KTI which regulates the trypsin activity in chickpea crop when exposed to pests and pathogens (1). It is also involved in seed germination and embryo axis development (2). The gene coding for CaTI2 from *C. arietinum L* (a Desi Chickpea) was fished out using gene-specific primers and was then cloned, expressed and purified from two expression systems, *E. coli* [Strain: BL21*(DE3)] and *P. pastoris* (Strain: GS115) expression host cells (3). CaTI2 inhibitor was earlier reported to exhibit activity towards insect gut proteases specifically from *H. armigera* (4). It was observed that the inhibitor was generously expressed in eukaryotic *P. pastoris* system as compared to prokaryotic *E. coli* and in our work; we checked its activity against Bovine pancreatic trypsin (BPT).

The biochemical and biophysical characterization of any macromolecule accelerates its development for potential application. For example, serpins have been evaluated for various commercial and pharmaceutical applications (5). The efficacy of CaTI2, against insect gut

proteases has already been demonstrated (4). It was observed that while insect gut proteases developed resistance to other KTIs, in case of CaTI2 resistance has not been reported (6). This property is very crucial for any potential application. Thus, we estimated the dynamic properties of CaTI2 using DSF, CD spectroscopy and biochemical activity measurements at different conditions such as various temperatures, pH, urea concentrations and in presence of solvents with varying polarity. The knowledge about such properties help for development of formulations It was observed that the inhibitor was maximally active up to 45 °C, with a T_m of about 61.3 °C. The inhibitor was found to be active in pH range 7-12. In increasing concentrations of urea, with time CaTI2 lost its activity and native structural conformation. It retained both structure and activity up to 2M urea concentration. As the polarity of the solvent decreases, a decrease in T_m was observed indicating alterations in native secondary structure elements which was also reflected in the CD profile. The dissociation constant was calculated using the highly sensitive microscale thermophoresis technique. The nanomolar K_d value observed indicates that the inhibitor has a high affinity for bovine pancreatic trypsin, resulting in a very strong CaTI2-BPT complex.

The three-dimensional structure of CaTI2 was determined at 2.9Å resolution. The structural analysis and comparison with other KTIs showed key variation in active site binding loop. Canonical KTIs contain Arg/Lys residue at P1 position, which blocks the nucleophilic attack of Ser195 from trypsin, a residue that brings about the digestion of substrate proteins. Thus, it is said that the classical KTIs should bind to a serine protease just like the substrate protein. But in case of CaTI2, the lack of substrate mimicking residues makes it bind to trypsin in a non-substrate like manner. Due to this variation, the geometry of inhibitory loop differs from other KTIs, making CaTI2 a unique inhibitor. In addition, the conserved Asn13 residue which has a role the in inhibitory activity of KTIs, is absent in CaTI2. Thus, we hypothesize that the CaTI2 not only binds to serine protease in non-substrate like mode, but also the underlying mechanism of inhibition might differ from that previously established for KTIs. Docking studies were helpful in predicting the residues from CaTI2 that could interact with the BPT. It was also noted that the interfacial area between CaTI2 and BPT was much higher than that between STI and BPT. This could suggest that the CaTI2 might be forming a stronger complex with BPT as compared to STI. For confirmation of the findings of docking experiments, elucidation of crystal structure of CaTI2-BPT complex is underway.

To summarize, as an outcome of this thesis work, the protease inhibitor CaTI2 isolated from a leading cash crop *C. arietinum L*, was identified to be a highly active KTI. Cloning, expression and characterization of CaTI2 using various biochemical and biophysical techniques, along with computational studies have established the role of structural features in its high specific activity, unique mechanism and promiscuity. Structural and docking studies have shown that the variations in amino acid sequence of the inhibitor are reflected in unique properties and structural organization of this inhibitor molecule. Apart from offering structural insights of an aberrant KTI from chickpea, our findings will also help in protein engineering to produce modified inhibitors with desirable properties.

Section 2: Conclusion

Following are the major highlights of the thesis:

- Sequence analysis: CaTI2 contains Kuntiz domain with variations in key conserved residues.
- Phylogenetic tree: CaTI2 is a divergent form of KTI protease inhibitor.
- Eukaryotic system (*P. pastoris*) proved to be more efficient in the expression of the inhibitor as compared to prokaryotic system (*E. coli*).
- Dissociation constant: The K_d value for trypsin binding by CaTI2 was calculated to be around $0.631\mu\text{M}$
- Native CD spectrum: Minima near 202nm indicates that CaTI2 is a β -sheet rich protein
- NanoDSF studies revealed that the melting temperature (T_m) is 61.3°C and thermal denaturation is irreversible.
- Only one Trp, present in core of the β -trefoil, is the source of intrinsic fluorescence.
- Trypsin inhibitory activity maximal up to 45°C and in basic pH range (7.0-12.0).
- Moderately stable in the presence of chaotropic agent (up to 2M urea).
- CaTI2 exhibits functional and structural stability in organic solvents (up to 25%) where, stability increases with increase in solvent polarity.
- CaTI2 diffracted crystal: Space group - $P 1 2_1 1$, a primitive space group as observed in most of the KTI crystals.


- Native structure was solved at 2.9Å resolution at RRCAT synchrotron, Indore, India.
- Crystallographic dimer with Matthews's coefficient 1.97 Å³/Da.
- The native structure was deposited in RCSB PDB with PDB ID 5XOZ.
- Inhibitory loop present between 4th and 5th β sheets.
- Molecular docking provides insights on enzyme-substrate interactions


In conclusion, current studies presented in this work help us in understanding new dimension of one of the model systems for PPI, an enzyme-inhibitor interaction system. The work has resulted in better understanding of the structural integrity and functional stability at molecular level for a KTI, which can serve as a structural toolbox to improve and engineer these KTIs as better and more efficient defence agents against plant pests in future. *In vitro* and *in silico* studies described in this thesis contribute to our understanding of the stability, activation and functionality of CaTI2 under different conditions which would prove directional in near future for development of formulations of such inhibitors for desired applications. Questions regarding structural details and molecular interaction of pathogenic proteases still remain to be resolved. Future efforts in crystallizing this protein with pathogenic proteases would give a complete picture of its structure-function relationship. Also, structure of such complexes would help establish the mode of inhibition, especially in case of inhibitors like CaTI2.


References


1. Jiménez, Teresa, et al. "The accumulation of a Kunitz trypsin inhibitor from chickpea (TPI-2) located in cell walls is increased in wounded leaves and elongating epicotyls." *Physiologia plantarum* 132.3 (2008): 306-317.
2. Hernández-Nistal, Josefina, et al. "Two cell wall Kunitz trypsin inhibitors in chickpea during seed germination and seedling growth." *Plant Physiology and Biochemistry* 47.3 (2009): 181-187.
3. Bendre, Ameya Dipak, Sureshkumar Ramasamy, and C. G. Suresh. "Chickpea Kunitz Inhibitor: A mechanistic basis for Trypsin Inhibition." *Acta Crystallographica Section A: Foundations and Advances* 70 (2017): C271.
4. Srinivasan, Ajay, et al. "A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae." *Plant Molecular Biology* 57.3 (2005): 359-374.
5. Ilies, Marc A., Claudiu T. Supuran, and Andrea Scozzafava. "Therapeutic applications of serine protease inhibitors." *Expert Opinion on Therapeutic Patents* 12.8 (2002): 1181-1214.
6. Srinivasan, Ajay, et al. "Podborer (*Helicoverpa armigera* Hübn.) does not show specific adaptations in gut proteinases to dietary *Cicer arietinum* Kunitz proteinase inhibitor." *Journal of insect physiology* 51.11 (2005): 1268-1276.


Symposia and workshops


-  International symposium on **‘Proteomics Beyond IDs... and 4th Annual Meeting of Proteomics Society (India)’** held at CSIR- NCL, Pune from 22nd- 24th Nov 2012


-  One day symposium on **‘The Wonderland of Molecular Structure through the Looking - Glass of X-ray Crystallography’** organized at CSIR-NCL, Pune, 23rd September 2013


-  International conference and workshop **‘Indo-Mexico workshop on Biotechnology: Beyond Borders’** at CSIR- National Chemical Laboratory, Pune, held on 7th -9th October 2013


-  **‘42nd National Seminar on Crystallography and International Workshop on Application of The X-ray Diffraction for Drug Discovery’** at Jawaharlal Nehru University, New Delhi, from 21st – 23rd November 2013

-  **‘44th National Seminar on Crystallography’** at Indian Institute of Science Education and Research, Pune, from 10th – 13th July 2016

-  **‘Workshop on Transcriptome Data Analysis’** organized at Venture Center, CSIR-NCL, Pune, from 16th – 18th August, 2016

-  One day **‘Elsevier Author Workshop’** held at CSIR-NCL, Pune, on 9th Dec 2016

-  **‘Workshop on Basic ‘R’ for Life Sciences’** organized at Venture Center, CSIR-NCL, Pune, from 29th June – 1st July, 2017

-  **‘24th Congress and General Assembly of the International Union of Crystallography’** at International Convention Centre, Hyderabad, from 21st – 28th August 2017

Poster presentations

- ✚ Science day poster: **“Protease Inhibitors from Chickpea: Structural, Molecular and Functional Aspects”** (25th and 26th February, 2015)
- ✚ 44th National Seminar on Crystallography: **“Kunitz Trypsin Inhibitor II from Chickpea: Unraveling the β - trefoil”** (10th – 13th July, 2016)
- ✚ 24th Congress and General Assembly of the International Union of Crystallography: **“Chickpea Trypsin Inhibitor- A Mechanistic Basis for Trypsin Inhibition”** (21st – 28th August, 2017)

List of publications

From the thesis work:

- ✚ **Ameya D. Bendre**, Sureshkumar Ramasamy, and C. G. Suresh. “Chickpea Kunitz Inhibitor: A mechanistic basis for Trypsin Inhibition”, *Acta Crystallographica Section A: Foundations and advances* (2017). 70: C271. DOI:10.1107/S2053273317093020
- ✚ **Ameya D. Bendre**, Sureshkumar Ramasamy, and C. G. Suresh. “Analysis of Kunitz inhibitors from plants for comprehensive structural and functional insights”, *International Journal of Biological Macromolecules* (2018). 113: 933-943. DOI: 10.1016/j.ijbiomac.2018.02.148
- ✚ **Ameya D. Bendre**, Dhanasekaran Shanmugam, C. G. Suresh, and Sureshkumar Ramasamy. “Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from *Cicer arietinum L.*”, *Journal of Biomolecular Structure & Dynamics* (In press). DOI: 10.1080/07391102.2018.1494633
- ✚ **Bendre Ameya D.**, Shukla E., Sureshkumar Ramasamy. “Understanding the dynamics of a Kunitz inhibitor from *Cicer arietinum L.*” (Communicated)

From collaborative work:

- ✚ Shukla E., Thorat Leena J, Bhavnani V., **Bendre Ameya D.**, Pal Jayanta K., Nath Bimalendu B., Gaikwad Sushama M. “Molecular cloning and *in silico* studies of physiologically significant trehalase from *Drosophila melanogaster*”, *International Journal of Biological Macromolecules* (2016). 92:282-292. DOI:10.1016 /j.ijbiomac.2016.06.097
- ✚ Pal Jayanta K., **Bendre Ameya D.**, Suresh C. G., Chatterjee S., Berwal S., Bhatia V. “Activation of HRI is mediated by Hsp90 during stress through modulation of

the HRI-Hsp90 complex”, *International Journal of Biological Macromolecules* (2018) (*In press*). DOI: 10.1016/j.ijbiomac.2018.06.204

- ✚ Shukla E., Thorat Leena J., **Bendre Ameya D.**, Jadhav S., Pal Jayanta K, Nath Bimalendu B., Gaikwad Sushama M. “Cloning and characterization of trehalase: a conserved glycosidase from oriental midge, *Chironomus ramosus*” (**Revised version communicated**)



Ameya Dipak Bendre

Correspondence: Lab 1875, Biochemical Sciences Division
CSIR-National Chemical Laboratory
Dr. Homi Bhabha Rd. Pashan
Pune 411008.

+91-9922745129/+91-020-25902212
ameyabenz@gmail.com

- Objective**
- To reach a position which would allow me to primarily apply and expand my research avenues in practical applications
- Skills**
- Biological technical skills: Protein expression and purification techniques from *E.coli* and *Pichia pastoris* systems and purification from natural source such as plant seeds, Nucleic acids Extraction, Molecular cloning, Protein Estimation Assays, UV-visible, Fluorescence and circular dichroism spectroscopy, LC-MS, Protein Crystallization, Protein structure determination using X-ray diffraction
 - Bioinformatics: Protein sequence analysis using alignment, phylogeny, Homology modeling, Protein- Ligand docking and MD-simulation using Schrodinger
 - Operating System: Windows, Linux/UNIX (Basic)
 - Programming: Basics of R
- Education**
- **Doctor of Philosophy in Biological Sciences (CSIR- National Chemical Laboratory, Pune)**
August, 2012 – till date
 - **Master of Biochemistry (61.2%, Class: 1st)**
Chemistry Department, University of Pune, Pune. June 2011
 - **Bachelor of Chemistry with Vocational Biotechnology (71.78%, Class: Distinction)**
Fergusson College (Under University of Pune), Pune. June 2009
 - **Higher School Secondary Certificate, Pune. (72.67%, Class: Distinction)**
N.M.V. Jr. College Pune, May 2005
 - **Secondary School Certificate, Pune (82.13%, Class: Distinction)**
Damle Prashala, Pune, May, 2003
- Professional Experience**
- **Quality control, Charoen Pokphand Foods Pvt. Ltd.**
June-August 2011

Carry out quality check of raw material and outputs at different stages of poultry feed production.

Research Experience

- **Research Fellow, CSIR-National Chemical Laboratory, Pune
July 2012 –Till date**
Project: Structural and functional characterization of kunitz inhibitor from *Cicer arietinum L.*
Role: Identification of Kunit genes from draft chickpea genome. Molecular cloning of identified genes. Biochemical activity assays of inhibitors. Identifying best inhibitor and structure determination and mutagenesis to identify key residues
- **Project Trainee, Department of Chemistry, University of Pune, Pune
June 2010 – May 2011**
Project: Green synthesis of metal and metal salt nanoparticles'.
Role: Synthesis of nanoparticles using small peptides like curcacyclin, oxytocin. Characterization of synthesized nanoparticles using Powder XRD, electron microscopy, EDXS etc.
- **Project Trainee, National Center for Cell Science , Pune
January, 2008 – December, 2008**
Project: Gene characterization of major malarial vector in India, *Anopheles stephensi*
Role: Cloning and sequencing of Hox genes from *Anopheles stephensi*

Awards and Fellowships

- Junior Research Fellowship (CSIR-JRF) and Lectureship in the Joint CSIR-UGC National Eligibility Test conducted by the Council of Scientific and Industrial Research (CSIR).
- GATE 2011
- Scholarship from Jamshetaji Tata Trust under PICC program for undergraduate students for pursuing individual research

Workshops

- Basic R for life sciences (June 2017) organized by Bio Sakshat Inc.
- Transcriptome data analysis (August 2016) organized by Bio Sakshat Inc.