Studies on the Structure and Interactions of a Plant Lectin and a Plant Protease Inhibitor

Thesis Submitted to the University of Pune

For The Degree of

DOCTOR OF PHILOSOPHY

IN

CHEMISTRY (BIOCHEMISTRY)

NARASIMHARAO. KRISHNAMURTHY. (K. N. RAO)

Division of Biochemical Sciences,
National Chemical Laboratory,
Pune 411 008 (India)

k**MAY2002**k

Dedicated to

Annaiah and Amma

CERTIFICATE

Certified that the work incorporated in this thesis entitled, '**Studies on the Structure and Interactions of a Plant Lectin and a Plant Protease Inhibitor**" submitted by Mr. **Narasimharao. K.** was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

(Dr. C. G. Suresh) Date:

Research Guide

DECLARATION

I hereby declare that the thesis entitled "**Studies on the Structure and Interactions of a Plant Lectin and a Plant Protease Inhibitor**" submitted for Ph.D. degree to the University of Pune has not been submitted by me to any other university for a degree or diploma.

Narasimharao. K.

Biochemical sciences division
National Chemical Laboratory
Homi Bhabha Road
Pune 411 008

Date:

CONTENTS

ACKNOWLEDGEMENTS

I take this opportunity to acknowledge with gratitude, the guidance and the support of my supervisor Dr. C. G. Suresh. I am highly indebted to him for his constant encouragement, and his very caring interest in me.

I am very thankful to Dr. Shama Barnabas for her support, valuable suggestions. Discussions with her has been very stimulating and helped to make logical plans for my life.

My sincere thanks to Dr. M. I. Khan for his encouragement. Scientific discussions with him were exciting and helpful in designing experiments. Dr. Sushma Gaikwad deserves a special mention for her support and useful discussions on lectins. I specially thank Drs. Khan, Sushma and Madura for providing pure lectin whenever I needed it for crystallization. I also thank Drs Pundle, Prabhune & Bode for their advice and support.

I wish to express my gratitude to Director Dr. Paul Ratnasamy and deputy director Dr. Sivaram and HODs Drs. K. N. Ganesh, Ranjekar and Aditi Pant for providing the opportunity, support and encouragement for carrying out research in N. C. L. and for their administrative help. I thank CSIR for the financial assistance

I specially thank Profs. Vijayan, Suguna, MRN Murthy and Savitri of IISc, Bangalore, for providing X-ray data collection and sequencing facility at the institute. I thank students at IISc Saikrishna, Biswal, Kiran, J. parakash, S.prakash and Michael for their help and cooperation. I sincerely thank Dr. Hosur and labmates at BARC for providing FPLC facility. I greatly thank Prof. Guy Dodson, Eleanor Dodson, Drs. Richard Lewis & C S Verma of University of York U.K. for collecting X-ray data on BBI crystals. My special thanks to Dr. Peter Flecker of University of Mainz, Germany, for his valuable advice, support and encouragement.

I acknowledge 'the never say no' help and cooperation of my labmates Rajesh, Bharat, Manish, Priya, Satya, KK Singh, Suresh Kumar, Arati, Krishna, Satish, Kosali, and Shravanti.

I am grateful to my close friends- Savita, Bhushan, Katti, Manoj, Neelam, Rajashri, Rajesh, Shashi Pujar, Vrinda, Sridevi, Arun, Thulsi, Loknath, Deepak. They will always hold a special position and respect in my heart.

I thank my colleagues and friends- Drs. Hegde, Sathi, Abhay, A Giri, Ashok, Amutha, Gaurav, Madura, Rangu, MNRao, Sami, Raju, Shamim, Venkat, Veena and Vincent from whom I learnt experimental techniques and turned for advice. I thank my friends Ajay, Anil, Atul, Shivshankar, R.Reddy, S. Niture, Shivappa, Suhas, Yanju and all hostelmates during my stay in the hostel.

I enjoyed the company of Keshavachar, Lata, Parag, Pujar, Sham, Revaiah, Ramesh, Govind, whose company refreshed mind and with them I shared jokes, narration of naughty incidents and interesting novels in my mother tongue. I wish to specially remember here my Gauribidanur college friends, Raghu, Ramnath, Ravikiran, Subbu, Murhty, Giri and Hari whose encouragement and companionship brought me to this stage.

I acknowledge the timely help from Karunakaran, Indira, Usha, Satyali, Mari, and Jagtap and technical support from Mr. Modak, Mr. Kamate, Mr. Karanjkar and Mr. Trehan.

I express my heartfelt gratitude to my parents, brother, sisters and brothers-in-law, uncles, aunts and cousins for their love, untiring support and encouragement.

Finally, I thank all my teachers, friends, relatives and well wishers, whose names may not have been mentioned here, but whose good wishes and support have enabled me to complete this work.

Narasimharao. K.

LIST OF ABBREVIATIONS

AMoRe	:	Automated Molecular Replacement
BBI	:	Bowman-Birk protese inhibitor
BBS	:	bromelain binding site
BPTI	:	bovine pancreatic trypsin inhibitor
BT	:	bovine trypsin
Bt	:	bascillus thuringenisis
BTCI	:	black-eyed pea trypsin/chymotrypsin inhibitor
BSA	:	bovine serum albumin
ß-me	:	ß-mercaptoethanol
CBS	:	chymotrypsin binding site
CCP4	:	Colloborative Computational Project No. 4
Con A	:	Concanavalin A
CPTI	:	cowpea trypsin inhibitor
CRD	:	carbohydrate recognition domain
DEAE	:	Diethyl amino ethyl
EcorI	:	*Erythrina corollodendron* lectin
ES-MS	:	electrospray-mass spectrometry
FPLC	:	fast protein liquid chromatography
Gal	:	galactose
GalNAc	:	N-acetyl-D-galactosamine
GlcNAc	:	N-acetyl-D-glucosamine
GK	:	Greek key
GS IV	:	*Griffonia simplicifolia* IV
h/hr	:	hour
HMW	:	high molecular weight
IEF	:	iso-electric focussing
IP	:	image plate
lit	:	litre
LMW	:	low molecular weight
m/min	:	minute
M	:	molar
MAD	:	multi-wavelength anomalous dispersion
MB	:	megabytes
MBP	:	mannose binding protein
me-a-gal	:	methyl--a-D-galactose
MIR	:	multiple isomorphous replacement
ML	:	mistletoe lectin
Man	:	mannose
MR	:	molecular replacement
Mol. Wt./Mr	:	molecular weight
MWCO	:	molecular weight cut off
MPA	:	*Maclura pomifera* agglutinin
NCS	:	non-crystallographic symmetry
O. D.	:	optical density

PAG	:	polynucleotide adenosine glycosidase
PAGE	:	poly-acrylamide gel electrophoresis
PDB	:	Protein DataBank
PEG	:	poly ethylene glycol
PIs	:	protease inhibitors
PNA	:	peanut agglutinin
PRPs	:	pathogenesis related proteins
PVA	:	penicillin V acylase
RIP	:	ribosome inactivating protein
rpm	:	revolution per minute
SBA	:	soybean agglutinin
s/sec	:	second
SDS	:	sodium dodecyl sulphate
Serpin	:	serine protease inhibitor
T-antigen	:	Thomsen-Fredenreich antigen
TBS	:	trypsin binding site
TCA	:	trichloro acetic acid
Tris	:	tris-hydroxymethyl amino methane
VMO-I	:	vitelline membrane outer layer protein-I
WGA	:	wheat germ agglutinin
w. r. t.	:	with respect to
URL	:	uniform resource locater

Å	:	Angstrom
a	:	alpha
a (h,k,l)	:	phase angle
F(hkl)	:	structure factor
s	:	sigma
S	:	summation
?	:	integration
°C	:	degree centigrade
µg	:	microgram
µl	:	microlitre
µM	:	micromolar
S-S	:	disulfide

ABSTRACT

1. INTRODUCTION

Many biological processes in nature are triggered and nurtured by protein-carbohydrate recognition and protein-protein interactions. Lectins mediate cell-cell and host-pathogen interactions through the specific recognition of carbohydrates present on the cell surface. Protein protease inhibitors regulate proteolytic activity through protein-protein interactions. The structures of lectins complexed with carbohydrates are model systems for studying protein-carbohydrate recognition whereas inhibitors complexed with the cognate proteases are useful for studying protein-protein interactions. Lectins and protease inhibitors are thought to act as defense proteins in plants. The amino acid sequences of several lectins and protease inhibitors are reported. The three-dimensional structures of some of the lectins and protease inhibitors determined through X-ray crystallographic method are also available. The three-dimensional structures of lectins and protease inhibitors have provided information at atomic level on protein-carbohydrate and protein-protein interactions. Since lectins and protease inhibitors have been implicated in several disease conditions they have applications in diagnosis and biomedical research.

In this thesis author has presented the determination of three-dimensional structure of a galactose specific lectin from the seeds of *Artocarpus hirsuta*, a tropical tree, and its interaction with the sugar molecule, methyl-α-D-galactose. Also described in detail the purification, characterization and X-ray crystal structure study of a Bowman-Birk protease inhibitor (BBI) from the seeds of *Vigna unguiculata* (cow pea). The putative interactions between *vigna* inhibitor and bovine trypsin have been probed using a docked complex simulated through the 3D-Dock docking programs.

2. RESEARCH FINDINGS

The purified lectin was available to us through research collaboration. The lectin was crystallized using hanging drop vapour diffusion method and ammonium sulfate as precipitant. The crystals belonged to two orthorhombic and two hexagonal forms. Since hexagonal crystals were of poor diffraction quality, data from them were not used for structure refinement. The orthorhombic forms I and II diffracted X-rays up to 2.8 and 3.0Å resolution, respectively. The diffraction data collected at room temperature was processed using DENZO and SCALEPACK programs. The structures were solved using molecular replacement technique. The search model was Jacalin dimer, a lectin from jackfruit seeds. The program AMoRe was used for molecular replacement calculations. The structures were refined using maximum-likelihood method implemented in REFMAC program. The final models in the two orthorhombic forms were refined to crystallographic R and Rfree values of 14.4% and 21.4% for form I, and 15.2% and 24.3% for form II, respectively.

The structure contained only ß-sheets and no a-helices as secondary structure. The lectin molecule has folded into three four stranded anti-parallel ß-sheets. These ß-sheets are organized around the faces of a prism and the fold is classified as 'ß-prism-I fold'. This organization has imparted a local three-fold symmetry to the molecule, the axis of pseudo-symmetry oriented parallel to the ß-sheets. However, no such internal symmetry is found in amino acid sequence. Two of the three ß-sheets (second and third) form Greek key motifs of topology 1,1,-3. The connectivity of strands forming the Greek key motif of the remaining ß-sheet is different. In this case the two outer strands come from the N-terminus of a and ß-chains, whereas the two inner strands are from C-terminus of a-chain. In all probability, the post-translational modification is responsible for the cleavage of the chain into two, a long a-chain and a short ß-chain. The difference Fourier electron density map (Fo-Fc, at 3σ contour level) has clearly shown the presence of methyl- a-D-galactopyranoside bound at the N-terminus of each of the four lectin molecules in the asymmetric unit. The N-terminal amino group and atoms belonging to residues Gly121, Asp125, Tyr78, Tyr122 and Trp123 of a-chain interact with the sugar molecule. The three-dimensional structure and protein-ligand interactions have been discussed in the context of available information on other lectins belonging to the jacalin family. The structure of *hirsuta* lectin compares closely with jacalin structure and thus resembles other jacalin-related structures.

The Bowman-Birk protease inhibitor from the seeds of *Vigna unguiculata* has been purified using a combination of traditional chromatographic techniques. The crystallization attempts were hampered by the presence of a number of isoinhibitors in the purified protein preparation. Six isoinhibitors in various concentrations could be separated by the application of FPLC technique in the final stages of purification The isoinhibitors were characterized by their molecular weight, isoelectric point and difference in specific activity towards trypsin and chymotrypsin. The N-terminal was sequenced for one of the isoinhibitors.

A major isoinhibitor crystallized from the purified preparation using hanging drop vapour diffusion method. The protein concentration of 50mg/ml, 0.1 M citrate-phosphate buffer of pH 4.0 and 30% saturated ammonium sulfate were used for crystallization. Tiny crystals of size 0.2 mm only could be obtained. X-ray diffraction data collected at 2.5Å resolution under cryo conditions were used for structure solution and refinement. The inhibitor crystal was flash cooled under a liquid nitrogen stream in the presence of 30% glycerol for data collection. The unit cell dimensions were a=32.4, b=61.8, c=32.9 Å, a= ß= 90?, ?=114.5?, in space group $P2_1$. Mathew's number calculated was 1.95 $Å^3$/D, assuming one dimer in the crystal asymmetric unit, which corresponded to 37% solvent content.

The complete amino acid sequence of a Bowman-Birk inhibitor from *Vigna unguiculata* seeds was reported. Initially this sequence was assumed for the crystallized inhibitor. The structure was solved using molecular replacement method using the reported tracey bean BBI structure as search model. Molecular replacement calculations were performed using the program AMoRe. The structure has two molecules of the inhibitor in the asymmetric unit. Thus a molecular averaging of the electron density was possible using the non-crystallographic symmetry. The averaged electron density helped

to improve the map. The initial model was fitted into this improved map using QUANTA. The model thus obtained was subjected to several cycles of refinement using maximum likelihood method implemented in REFMAC program. Towards the final stages of refinement, the non-crystallographic constraint was released and waters were added using difference Fourier electron density. The final values of R factor and R free were 19.1% and 25.8%, respectively.

The BBI monomer structure is similar to that of other known BBI structures. However, the asymmetric unit is a dimer of two closely interacting monomer molecules related by a pseudo 2-fold axis. Each monomer has two domains formed by two anti parallel β-strands and an associated small strand. These domains position the trypsin and chymotrypsin specific insertion loops away from each other. The disulfide bridges formed by 14 cystine residues in the sequence stabilize the individual domains. There are four disulfide bridges in the first domain and three in the second. Most of the side chains could be identified from the electron density map. The amino acid sequence based on electron density map differed in a few places with respect to the reported sequence.

The chymotrypsin insertion loop is poorly defined in the electron density map, compared to that of trypsin loop. The stability of the latter loop, seen by its good electron density, is presumably due to its occurrence at the interface of the monomers in the dimer. The restricted movements of these loops could be due to their close juxtaposition in the dimer. The mutual disposition of the trypsin-specific loops in the dimer suggests a protective role for dimer formation. The mutual interaction and close disposition of the loops in the dimer prevent their exposure to undesirable proteolytic action.

The structural studies of inhibitor were also intended to understand its interaction with cognate proteases. Unfortunately attempts to crystallize the inhibitor complexed with trypsin or chymotrypsin did not succeed. Hence docking of the Vigna BBI inhibitor with bovine trypsin through simulation using 3D-Dock docking program was tried. We have made use of the known interactions from available crystal structures of complexes for the simulation of BBI - trypsin complex.

Appendix 1 provides an account of the preliminary crystallization experiments and crystal characterization of penicillin V acylase. Appendix 2 discusses the structural investigations of some modified carbohydrates.

CHAPTER-1

GENERAL INTRODUCTION

1. 1. INTRODUCTION

Molecular recognition, interaction and binding are central events in biology. Many biological processes in nature are triggered by protein-carbohydrate recognition and protein-protein interactions. Cell-cell recognition is vital for the initiation and control of many celllular functions. Lectins mediate cell-cell and host-pathogen interactions through the specific recognition of carbohydrates present on the cell surface. The regulation of the activity of proteolytic enzymes by protein protease inhibitors is through protein-protein interactions. The structures of lectin-carbohydrate complexes turned out to be excellent model systems for studying protein-carbohydrate interactions. Similarly the complexes of protease inhibitors with their cognate proteases facilitated the study of protein-protein interactions. Lectins and protease inhibitors are thought to act as defense proteins in plants against pests and pathogens. In this thesis the author has presented determination of the three-dimensional structure of a galactose specific lectin from the seeds of *Artocarpus hirsuta* (Fig. 1. 1. C), a tropical tree (Fig. 1. 1. A), and its interaction with the sugar molecule, methyl-α-D-galactose. Also described in detail are the purification, characterization and crystal structure study of a Bowman-Birk type protease inhibitor from the seeds of *Vigna unguiculata* (Fig. 1. 1. D). A modelled structure of the complex of this protease inhibitor with bovine trypsin is also presented.

1. 2. SEED PROTEINS

More than half of the world's human consumption of proteins for energy is supplied by seed proteins. Seed proteins were among the first plant components to be studied systematically. Seeds from many important crops are of immense economic, social and nutritional importance to human kind. Understanding seed proteins in order to improve their composition, properties and concentration will continue to be important objective of future research.

A

B

C

D

Fig. 1. 1. The images of, A) *Artocarpus hirsuta* tree, B) ripen fruit of *hirsuta*, C) *hirsuta* seeds, D) *Vigna unguiculata* seeds.

Osborne was the first biochemist to apply a systematic approach towards the classification, characterization and nomenclature of plant proteins. He classified the simple plant proteins into four major groups depending on their solubility (Osborne, 1924). However, modern classification of plant proteins is

based on two aspects; protein function and molecular/biochemical relationships. The functional proteins present in seeds are further classified into three main classes (Shewry & Casey, 1999). They are: 1. **Storage proteins** which store nitrogen, carbon and sulfur, 2. **Structural and metabolic proteins** that are essential for the growth and structure of the seed, 3. **Defense or protective proteins** involved in plant defense.

Plants generate various inter and intra cellular signals to activate genes for the induction of various substances in response to an external attack (Bowles, 1990). They include antibiotics, alkaloids, terpenes, and defense proteins such as protease inhibitors, lectins, etc, (Ryan, 1990). A wide variety of other proteins are considered as defense proteins (Bowles, 1990; Osborn & Broekaert, 1999; Van Loon, 1985; Vierling, 1991). These include, structural proteins such as hydroxy proline-rich glycoproteins (HRGPs) and glycine rich proteins (GRPs), amylases, thionins, chitinases, phytoalexins and pathogenesis related" proteins (PRPs). Defense proteins play a key role in protecting plants against microbial pathogens, invertebrate pests, environmental stress, etc,. Lot of research has gone into the study of plant defense by protease inhibitors and lectins (Bowles, 1990; Chrispeels & Raikhel, 1991; Inbar & Chet, 1997; Ryan, 1990; Shewry & Lucas, 1997; Valueva & Mosolov, 1999). The expression of genes of protease inhibitors and lectins in important crop plants to protect them against pests and pathogens is gianing much attention in recent times (Hilder & Boulter, 1999; Jouanin *et al*., 1998; Schuler *et al*., 1998; Ussuf *et al*., 2001). The Bowman-Birk protease inhibitor and the *hirsuta* lectin, discussed in this thesis, could be classified as defence proteins (Eckelcamp *et al*., 1993; Gurjar *et al*., 2000; Sakal *et al*., 1989). In fact, the *hirsuta* lectin was recently studied for its insecticidal activity against the larvae of red flour beetle (*Tribolium castaneum*) indicating its role in plant defense (Gurjar *et al*., 2000).

1. 3. LECTINS

Stillmark in 1888 while investigating the toxic effects of castor bean extract (*Ricinus communis*) on blood noticed that the red blood cells (RBCs) are agglutinated, the phenomenon named as hemagglutination. A toxic protein from

castor beans called “ricin” capable of agglutinating the RBCs of human and animals was identified (Stillmark, 1888). Initially the ‘agglutinins’ were found only in plants. Later when they were found in other organisms and agglutinating cells other than erythrocytes the term “lectin” (from Latin verb ‘legere’ means to choose) was proposed by William Boyd (Boyd & Slapeigh, 1954). Present definition of lectins is "proteins possesing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide" (Peumans & Van Damme, 1999). Their detection and quantification are based on the agglutinating property and purification is using specific immobilized sugar affinity columns (Gabius & Gabius, 1993; Goldstein & Poretz, 1986; Lis & Sharon, 1984).

Lectins are classified according to the overall structure of the primary gene translation products (Peumans & Van Damme, 1999). They are merolectins, hololectins, chimerolectins and superlectins. Merolectins are monovalent, have a single carbohydrate binding domain, hence do not precipitate glycoconjugates or agglutinate cells. Hololectins are composed of at least two carbohydrate binding domains, identical or homologous, and bind structurally similar sugar(s). Hololectins agglutinate cells and/or precipitate glycoconjugates owing to their multiple binding sites. Most of the plant lectins belong to this group. Chimerolectins are fusion proteins comprising of one or more carbohydrate binding domain(s) arrayed in tandem with an unrelated domain. The unrelated domain can have well defined biological activity and functions independently of the carbohydrate binding domain. Chimerolectins can behave as merolectins or hololectins depending upon the number of carbohydrate binding domains. Superlectins are also fusion proteins having two carbohydrate binding domains, arranged in tandem, but differ in terms of structure and specificity.

Three-dimensional structures of large number of lectins belonging to variety of folds have been reported. Recent reviews have described the three-dimensional structures and related aspects of lectins in detail (Bouckaert *et al*., 1999; Drickamer, 1999; Rini, 1999; Vijayan & Chandra, 1999). Over 230 three-dimensional structures of lectins and their complexes are listed in 3D Lectin

Data Bank (http://www.cermav.cnrs.fr/databank/lectine). More than half the number is from plants, and the rest from animals, bacteria and viruses. The known structures of lectins can be classified into a finite number of unique folds (around 17). Among them the legume lectin fold and the C-type lectin fold are the most frequently observed ones. Two types of lectin carbohydrate recognition domains (CRD) are identified depending on their sizes (Vijayan & Chandra, 1999). One is the 120-250 residue CRD seen in legume lectins, jacalins, galectins, C-type lectins and influenza virus hemagglutinin. The other is the 40-50 residue CRD seen in bulb lectins and cereal lectins. The available lectin structures, classified according to monomer fold, along with their PDB codes are given below in table. 1. 1. The information is based on 3D lectin database accessible at the above mentioned URL.

1. 3. 1. Plant lectins

1. 3. 1. a. Classification based on sequence and three-dimensional structure

The plant lectins are a heterogeneous group of carbohydrate binding proteins which differ in their molecular structure and carbohydrate binding specificity (Barre *et al.*, 2001; Bouckaert, *et al.*, 1999; Rini, 1995; Van Damme *et al.*, 1998). They are found in seeds and in other tissues such as leaves, bark, stems, rhizomes, bulbs, tubers, etc. (Etzler, 1985 and 1992; Peumans & Van Damme, 1998 and 1999). Because of the broad distribution and ease of isolation, plant lectins are the largest and the best characterized group among lectins. They occur in many species belonging to different taxonomic groups but are not so ubiquitous. Lectins have been identified in less than 500 plant species which belong to a small number of plant families. Rarely, lectins that differ in their sugar specificity have been found in the same plant, especially in the case of legumes and Gramineae (Lis & Sharon, 1986a,b). At present seven distinct lectin families have been identified on the basis of protein sequences (Peumans & Van Damme, 1999). They are the legume lectins, monocot mannose–binding lectins, jacalins, chitin-binding lectins, type2 RIPs, amaranthins, and cucurbitaceae phloem lectins.

Table. 1. 1. Classification of lectin structures based on sequence homology and subunit fold along with their PDB codes.

TYPE	SOURCE	PDB ID
1. PLANT LECTINS		
(i) **Legume lectins** (Legume lectin fold or 'Jelly roll' fold)	*Arachis hypogea* (Peanut agglutinin, PNA)	2PEL 2TEP 1BZW 1CIW 1QF3
	Canavalia ensiformis (Jackbean- Con A)	2CNA 2CTV 3CAN 1NLS 1JBC 1VLN 1CN1 1APN 1ENS 1ENQ 1CES 1CON 1SCR 1SCS 1ENR 2ENR 1DQ0 1DQ1 1DQ2 1DQ3 1DQ4 1DQ5 1DQ6 4CNA 5CAN 1GIC 1CVN 1ONA 1TEI 1VAM 1VAL 1CJP 1QGL 1BXH 1DC 1QDO
	Canavalia brasiliensis (Con B)	1AZD
	Dolichos biflorus lectin (DBL)	1LUL 1BJQ 1LU1 1LU2
	Erythrina corallodendron lectin (EcorL)	1AXY 1LTE 1AXZ 1AX0 1AX1 1AX2
	Griffonia simplicifolia (GS4)	1LEC
	Lathyrus ochrus (Isolectin, LOL I & II)	1LOE 1LOA 1LOB 1LOC 1LOD 1LGB 1LGC
	Lens culinaris (Lentil lectin)	2LAL 1LEN 1LES 1LEM
	Phaseolus lunatus (Limabean agglutinin)	ND
	Pisum sativum (Pea lectin, PSL)	2LTN 1BQP 2BQP 1RIN
	Psophocarpus Tetragonolobus (Winged bean lectin,WBL)	1WBL 1WBF
	Ulex europeans agglutinin	1QNW 1QOO 1QOS 1QOT

	(UEA Isolectin I & II)	
	Dioclea grandiflora lectin (DGL)	1DGL
	Dolichos lablab (Legume lectin- Fril)	1QMO
	Glycine max (Soyabean agglutinin, SBA)	2SBA 1SBD 1SBE 1SBF
	Phaseolus vulgaris (Phytoheamagglutinin, PHA-L)	1FAT
	Vicia faba (Favin- FVA)	ND
	Vicia villosa isolectin B4 (VVLB4)	ND
(ii) **Monocot mannose-binding (Bulb) lectins** (ß Prism-II fold)	*Galanthus nivalis* agglutinin (Snowdrop lectin)	1MSA 1NIV 1JPC
	Hippeastrum hybrid (Amaryllis-HHA)	ND
	*Scilla campa*nulata (Bluebell bulb lectin)	1B2P ND ND ND
	Narcissus pseudonarcissus (Daffodil lectin)	1NPL
	Allium sativum (Garlic bulb lectin)	1BWU
(iii) **Cereal or Chitin-binding lectins** (Havein domain fold)	*Havea brasiliensis* (Havein)	1HEV ND2 ND3
	Triticum vulgaris (Wheat germ agglutinin isolectin I, II & III, WGA1, WGA2, WGA3)	7WGA 1WGC 2CWG 9WGA 2WGC ND ND 1WGT
(iv) **Jacalin lectins** (ß prism-I fold)	*Artocarpus integrifolia* (Jacalin)	1JAC
	Maclura pomifera Agglutinin (MPA)	1JOT

	Helianthus tuberosus lectin (Heltuba)	1C3K 1C3M 1C3N
(v) **Type II RIPs** ('ß- trefoil' fold)	*Ricinus communis* (Ricin B)	2AAI
	Abrus precatorius (Abrin- α)	1ABR
(vi) **Amaranthin lectins** ('ß- trefoil' fold)	*Amaranthus caudatus,* (ACA)	1MMC 1JLY 1JLX
	Viscum album (Mistletoe lectin I)	ND
2. ANIMAL LECTINS		
(i) **C- type lectins** (C-type lectin fold)	*Homo sapiens*	1ESL 1KJA 1KJB 1KJD
	(E-Selectins	1BO8
	Human lung surfactant D Protein	1HUP
	Mannose –binding proteins	1TN3 1HTN
	Tetranectins tunicates)	1TLG 1BYF
	Rattus norvegicus	1AFD 1BUU 1KMB 1YTT
	(Mannose Binding Protein, MBP-AI)	1AFA 1AFB 1BCH 1BCJ 2KMB 3KMB 4KMB 1BV4
	(Mannose Binding Protein, MBP-C)	1MSB 1RTM 2MSB
	Rattus rattus MBP-A, MBP-C	1RDO 1RDI 1RDJ 1RDK 1RDL 1RDM 1RDN
(ii) **Galectins or S-type lectins** (legume lectin fold or 'jelly roll motif')	*Bos taurus* (Galectin-1,Galaptin)	1SLA 1SLB 1SLC 1SLT
	Bos arenarum (Toad ovary galectins)	1GAN 1A78
	Homo sapiens (Galectin- 2, Galectin CN)	1HLC 1A3K
	(Galectin-3, Human Galectin-3 CRD)	1BKZ 2GAL 3GAL 4GAL
	(Galectin- 7, Human	5GAL

	galectin –7 CRD)	
	Conger myriaster	1C1F 1C1L
	(Congerin I)	
	Homo sapiens	1LCL
	(Charcot-Lenden crystal protein)	
	Gallus gallus	1QMJ
	(Chicken galectin)	
(iii) **I- type lectins**	*Mus musculus*	1QFO 1QFP
(I-lectin fold)	(Sialoadhesin)	
(iv) **Pentraxins**	*Homo sapiens*	1SAC 1LGN
(legume lectin fold)	(Serum amyloid P-Component)	
	(C-reactive protein)	1CRV 1GNH
	Mesocricetus auratus	1HAS
	Limulus polyphemus	1LIM
(v) **P-type lectins**	*Bovin*	1M6P 1C39
(P-lectin fold)	(Cation dependent mannose 6 phosphate receptor)	
(vi) **Tachylectins**	*Tachyphleus tridentatus*	TL2
(Tachylectin fold)	(Tachylectin)	
(vii) **Toxins**	*Selenocosmia huwena*	1QK7
	(Spider toxins lectin I)	
3. BACTERIAL LECTINS		1PRT 1BCP 1PTO
(i) **AB_5 toxins or ADP-ribosylating toxins**	*Bordetella pertussis*	
	(Pertussis toxin)	1XTC 1CHP 1FGB 1CHB
	Vibrio cholerae	2CHB 1CT1
	(Cholera toxin)	1LTR 1LTS 1LTB 1LTA
	E coli	1LTT
	(Heat labile enterotoxin)	1BOV 4ULL 1BOS 2BOS
	(Verotoxin 1, Shiga-like toxin-1)	1SE3 1SE4

(ii) **Other toxins**	*Staphylococus aureus* (Staphylococcal enterotoxin B)	
4. VIRAL LECTINS		
(i) **Coat proteins**	*Mouse polyoma virus*	1VPN 1VPS 1SID 1SIE
(ii**) Hemagglutinin**	Foot and mouth disease virus	1FHP
	Influenza virus hemagglutinin strain X-31	1HGF 1HTM 2HMG 3HMG 1HGD 1HE 1HGG 1HGH 1HGI 1HGJ 1QFU 1VIR 4HMG 5HMG
(iii) **Tailspike like proteins**	Phage P22 tailspike proteins	1TSP 1TYV 1CLW 1QA1 1QA2 1QA3 1TYU 1TYW 1TYX

Legume lectins

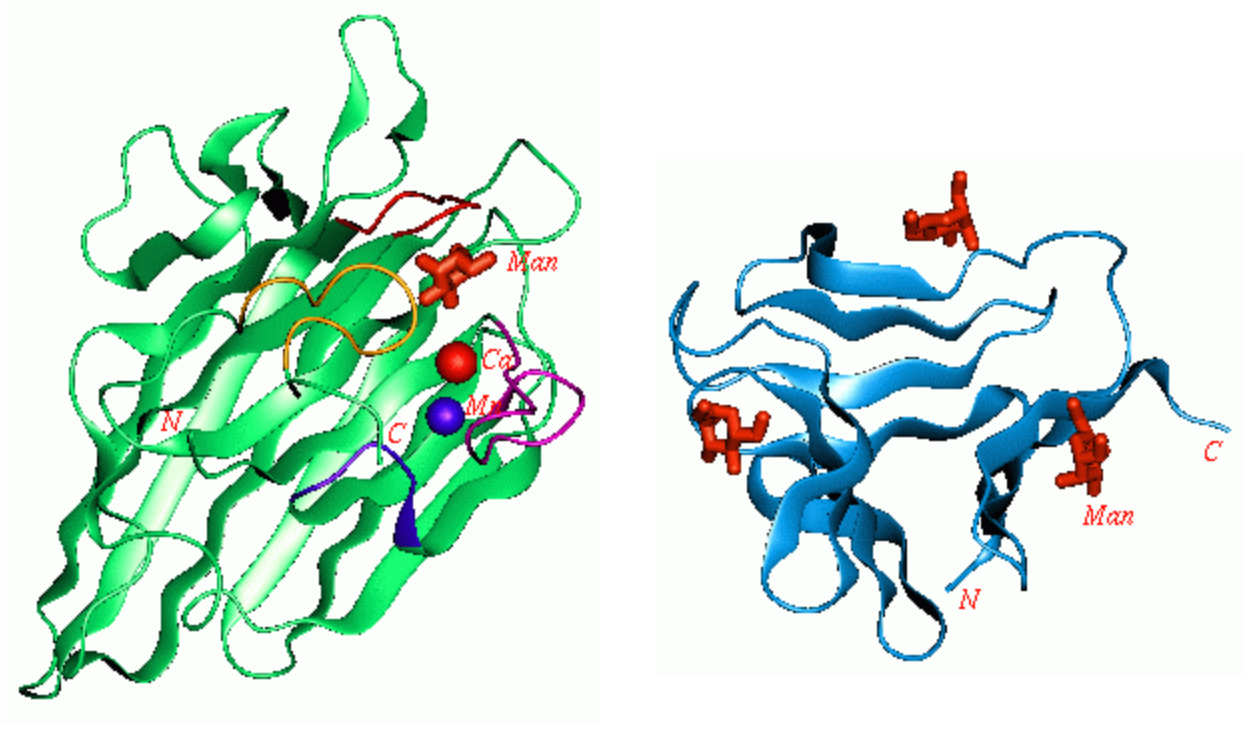
Lectins are present in several hundreds of species of legumes. Many of them have been purified and characterized with respect to their molecular structure, sugar binding specificity and biological role. Around 50 of them have been sequenced, mostly by protein sequencing method (Peumans & Van Damme; 1999; Sharon & Lis, 1990; Van Damme *et al*., 1998).

The legume lectins share extensive sequence homology and three-dimensional structural similarity but differ in carbohydrate specificity (Rouge *et al*., 1991; Sharon & Lis, 1990). The crystal structures of more than 15 legume lectins have been reported (Bouckaert *et al*., 1999; Loris *et al*., 1998) (table. 1. 1). The first legume lectin structure to be solved was of concanavalin A (con A) (Baker *et al*., 1975; Hardman & Ainsworth, 1972). The strutcure belonged to 'jelly- roll motif' commonly found in viral coat proteins. It has a nearly flat six-stranded rear ß-sheet and a curved seven-stranded front ß-sheet. A short five membered ß-sheet on top holds the two larger sheets together. In addition, a number of loops connecting the strands are also present (Fig. 1. 2. A). Two metal ions, a calcium and a transition metal, found in all the legume lectin

structures are essential for the carbohydrate binding. The metal ions are bound to four protein atoms and two water molecules. The legume lectin fold is also found in some animal lectins such as galectins, pentraxins and a few other proteins from animal and microbial sources (Srinivasan *et al.*, 1996).

One interesting aspect about legume lectin structures is their different quaternary association, independent of carbohydrate specificity. It is observed that small sequence dependent variations in tertiary structure lead to large changes in quaternary association (Manoj & Suguna, 2001; Vijayan & Chandra, 1999). The dimeric structure found in the crystals of ConA, Pea lectin, Lentil lectin and Lathyrus lectin is termed as 'canonical dimer'. This dimer is formed by side-by-side arrangement of two monomers in such a way that the two rear ß sheets form a contiguous 12-stranded ß sheet. However, this type of canonical dimer is missing in the crystal structures of Ecorl, GSIV and Peanut lectins. The dimers of these lectins posses back to back arrangement ('hand shake dimer') of the subunits (Srinivas *et al*, 2001).

All legume lectins have similar carbohydrate-binding sites (Loris *et al.*, 1998). The four loops A, B, C, D, shown in red, purple, yellow and blue color in Fig. 1. 2. A, associated with the concave face of the seven-stranded curved ß-sheet at the top front side of the subunit form the binding site (Sharma & Surolia, 1997). The amino acid residues which bind metal ions (Ca^{2+} and Mn^{2+}) are highly conserved and the residues which constitute the sugar binding site are less conserved but similar in properties (Lis & Sharan, 1998). The conserved residues Asp and Gly (or Arg) are from loops A and B, whereas Asn and the hydrophobic residues (Phe, Tyr, Trp or Leu) are from loop C. The size of the backbone of this loop is the primary determinant of carbohydrate specificity in legume lectins. The overall conformational changes observed upon sugar binding are small in legume lectins. The aligned sequences of legume lectins show 4 to 7 gaps in the binding loop D, admitting variation in the loop size (Sharma & Surolia, 1997). This partly explains the broad specificity of legume lectins compared to other lectin families.



A B

C

Fig. 1. 2. Crystal structures of subunits of, A) ConcanavalinA bound to mannose (Man), Ca and Mn are calcium and manganese ions, B) *Galanthus nivalis* agglutinin (GNA) bound to mannose at three sites, C) Wheat germ agglutinin bound to a-N-Acetyl Neuraminyl Lactose (ANL), disulfide bonds are shown in yellow.

Monocot mannose-binding (bulb) lectins

Bulb lectins are present in families such as Amaryllidaceae, Alliaceae, Araceae, Orchidaceae, Liliaceae and Bromeliaceae (Van Damme *et al.*, 1998). They are detected in various vegetative tissues but not in seeds. Monocot mannose-binding lectins, as the name indicates, exhibit exclusive specificity towards a-D-mannose. They share high similarity in sequence and overall structure (Barre *et al.*, 1996; Van Damme *et al.*, 1987).

The structure of monocot mannose-binding lectins, the 'ß prism-II' fold, which is the third structure among ß prism fold lectins, the other two being 'ß prism-I' fold of jacalins and 'ß-trefoil' of type2 RIPs and amaranthins (Wright, 1997). The snowdrop lectin (*Galanthus nivalis* agglutinin, GNA) from amaryllidaceae family complexed with a-D-mannose is the first crystal structure of a bulb lectin sugar complex (Hester *et al.*, 1995). The available crystal structures (Chandra *et al.*, 1999; Chantalat *et al.*, 1996; Hester & Wright, 1996; Saueborn *et al.*, 1999; Wood *et al.*, 1999; Wright & Hester, 1996) are listed in table 1. 1. The polypeptide chain is organized into three subdomains of flat four-stranded antiparallel ß-sheet with a local three-fold symmetry, and arranged to form three faces of a triangular prism (Fig. 1. 2. B). The strands of the 12-stranded ß-barrel run perpendicular to the barrel axis, while the axis itself coincides with the pseudo 3-fold axis. There are about 15 conserved hydrophobic side chains at the center of the barrel. Each of the ß-sheets has a carbohydrate recognition domain (CRD1, CRD2, CRD3) and they share sequence homology (Van Damme & Peumans, 1991). The subunits of lectins from GNA, amryllis, bluebell and daffodil form tetramers with 222 symmetry. The garlic lectin forms a heterodimer of two closely related subunits, and an additional binding site found in the dimer makes it 7 binding sites per dimer. In the case of daffodil lectin there exists a fourth binding site, which makes it 16 binding sites per tetramer. The exclusive specificity for a–D-mannose may be the result of complete burial and inaccessiblity of its 2-OH group, highly immobilized through hydrogen bonds with three side chains, as observed in lectin-mannose complexes.

Chitin-binding lectins

Chitin-binding lectins are a large family of proteins containing at least one hevein domain (Raikhel *et al*., 1993). The hevein domain is a chitin-binding polypeptide of about 40 amino acid residues that has sequence similarity with hevein, the 43 residue chitin-binding protein from the latex of rubber tree. The legume and cucurbitaceae phloem lectins also bind chitin but lack the hevein domain and have no sequence similarity with chitin-binding lectins possessing hevein domain. The chitin-binding lectins belong to the class of merolectins, hololectins and chimerolectins. Hitherto the presence of chitin-binding lectins have been observed in Gramineae, Solanceae, Phytolaccaceae, Urticacea, Papavaracaceae and Viscaceae (Peumans *et al*., 1996; Raikhel *et al*., 1993) indicating their occurrence in both monocots and dicots. They are found in seeds as well as in other vegetative tissues. The lectins of Gramineae are the best characterized chitin-binding seed lectins. These lectins exhibit a marked specificity towards GlcNAc and GalNAc-oligomers but they also bind sialic acid. All Gramineae seed lectins are dimeric proteins composed of identical or similar protomers of around 17 kDa size.

Wheat germ agglutinin (WGA), a representative member of chitin-binding lectins family, exists in three isoforms (WGA1, WGA2, WGA3) that differ in 5-8 amino acid residues out of a total of 171 residues in the polypeptide chain (Smith & Raikhel, 1989). The high resolution crystal structures have been reported for all the three isolectins in native form (Harata *et al*., 1995; Wright, 1987 and 1989) and WGA1 and WGA2 in various complexes with sugar molecules (Wright, 1992; Wright & Jaeger, 1993). The 18 kDa WGA subunit is made up of four homologous domains of 43 residues each folded into a compact globule through four disulfide linkages and five or six ß-turns (Fig. 1. 2. C) The four domains are organized into a helical assembly. Monomers associate in a head-to-tail fashion to form dimers such that the pairs of domains in contact are quasi two-fold related. From the crystal structure of the complex four potential sugar-binding sites per dimer have been identified in the contact regions of opposing monomers (Wright, 1990).

Jacalins

Lectins from Moraceae (jacalin, MPA, *hirsuta*, artocarpin), Convulvulaceae (calsepa, conarva), Asteraceae (heltuba), Gramineae (barley and wheat lectins) and Musaceae (banana lectin) together constitute the jacalin lectin family (Bourne *et al*., 1999). They occur in seeds as well as in vegetative tissues. The *hirusta* lectin (Moraceae), discussed in this thesis, also belongs to this family. Jacalins exhibit low sequence similarity and differ with respect to their carbohydrate binding specificity. The crystal structures of jacalin lectins are listed in table 1. 1. The fold and structure are discussed in detail in the 3rd chapter in the context of *hirsuta* lectin structure.

Type2 RIPs (Ribosome Inactivating Proteins)

These lectins are capable of catalytically inactivating eukaryotic ribosomes (Barbieri *et al*., 1993) using the catalytic toxic domain A. They are present in both monocots and dicots, such as Euphorbiaceae, Fabaceae, Sambucaceae, Viscaceae, Ranunculaceae, Lauracaceae, Passifloraceae, Iridaceae and Liliaceae. They occur both in seeds and in other vegetative tissues. They are chimero lectins composed of a PAG (Polynucleotide adenosine glycosidase) domain called A-chain (Endo *et al*., 1987), linked to a galactose specific domain called B-chain (Lord *et al*., 1994). Type2 RIPs share a high degree of sequence and structural similarity, but differ in catalytic activity, toxicity and carbohydrate binding capacity. The first lectin 'ricin', to be discovered in castor beans, is composed of two subunits of Mr. 32 kDa held by a S-S (disulfide) bridge and has specificity for Gal/GalNAc (Lord *et al*., 1994) (Fig. 1. 3. A). Seeds of *Abrus precatorius* (jequirity bean) contain several isoforms of abrin and *Abrus precatorius* agglutinin. Abrin is a heterodimer of subunits (Mr. 34 kDa and 32 kDa) joined together by single S-S bridge (Hedge *et al*., 1991; Wu *et al*., 2001). Similar type2 RIPs have been isolated from the seeds of camphor tree, bitter gourd and mistletoe (Peumans & Van Damme, 1999).

The two classes of lectins, type2 RIPs and amaranthins share the same 'ß-trefoil fold'. This fold was first observed in the crystal structure of Kunitz

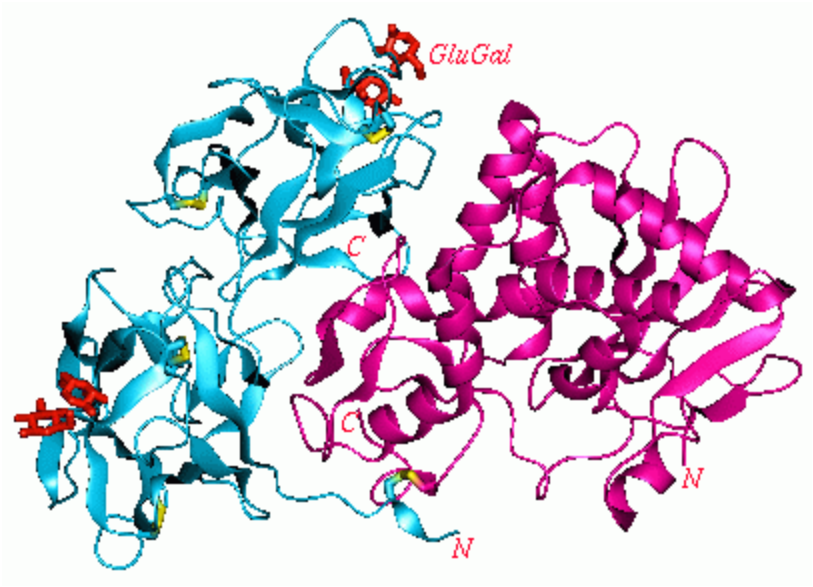
soybean trypsin inhibitor (Murzin *et al*., 1992). The ß-trefoil fold consists of a 12-stranded ß-sheet which forms six hairpins (Fig. 1. 3. A). Three of the hairpins connect 6 strands that form a barrel structure tilted at 56º to the barrel axis. The crystal structures of type2 RIPs (Krauspenhaar *et al*., 1999; Rutenber & Robertus, 1991; Tahirov *et al*., 1995) are listed in table. 1. 1. Although, the members of this family have different phylogenetic origin they share similar basic structural architecture.

Amaranthins

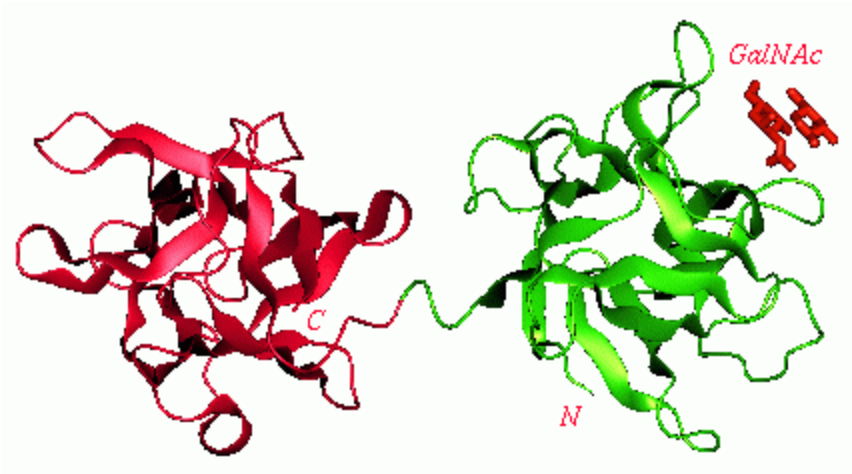
Amaranthins are a small family of lectins, found in the seeds of Amaranthus species. These are homodimeric proteins composed of two identical subunits of Mr around 30 kDa and exhibit specificity towards GalNAc (Rindrele *et al*., 1990). Amaranthin, the lectin from the seeds *Amaranthus caudatus* specific for GalNAc has been studied in detail, including its three-dimensional structure (Transue *et al*., 1997). This protein is a homodimer of subunit size 33 kDa and has a 'ß-trefoil fold', just like type2 RIPs. It has no sequence homology with B chain of the latter one. Its carbohydrate binding site is also different from that of RIPs. Amaranthin subunit folds into two ß-trefoil domains, one of them is involved in carbohydrate binding and the other in oligomerisation (Transue *et al*., 1997) (Fig. 1. 3. B). Amaranthin shows affinity for T- and Tn-anitgen.

Cucurbitaceae phloem lectins

Cucurbitaceae phloem lectins are found only in the phloem exudates of many cucurbitaceae. These lectins are specific for oligomers of GlcNAc (Wang *et al*., 1994). They have high sequence similarity (Peumans & Van Damme, 1999). Seed lectins have been found in some *Trichosanthes, Telfaira* and *Mamordica* species. Seeds of serpent cucumber (*Trichosanthes kirilowii*) contain a glycosylated galactose specific lectin composed of disulfide-linked subunits of Mr. 37 kDa and 25 kDa (Falasca *et al*, 1989). Similar seed lectins from *Trichosanthes anguina* (snake gourd) (Komath & Swamy, 1998) and *Trichosanthes cucumerina* species (Padma *et al*., 1999) have been isolated and



A



B

Fig. 1. 3. Crystal structures of subunits of, A) Ricin (type2 RIP) lectin bound to Glu/Gal, B) *Amaranthus caudatus* agglutinin (ACA) bound to GalNAc. The two domains are shown in different colors.

characterized. Another seed lectin from fluted gourd (*Telfaira occidentalis*) is a hexameric protein with three pairs of disulphide-linked subunits of Mr. 30 kDa (Peumans & Van Damme, 1999). A galactose specific seed lectin from *Mamordica charanita* species has also been characterized (Das *et al.*, 1981). This lectin is glycosylated also. No three-dimensional structre is yet available for

this family. However, recent premilinary X-ray studies of lectins from serpent cucumber (Li *et al*., 2000) and snake gourd (Manoj *et al*., 2001) indicate structural homology with type2 RIPs.

1. 3. 2. Animal lectins

Drickamer in 1988 classified the animal lectins into C-type and S-type lectins. The C-type (calcium dependent) or integral membrane lectins require detergents for their extraction (Barondes, 1981 and 1984). These lectins show broad specificity towards sugars such as mannose, L-fucose, mannose-6-phosphate, GalNAc etc. (Ashwell & Hardford, 1982; Monsigny *et al*., 1983). Membrane lectins are known to be present on the cell surfaces of mouse and human lymphocytes, 3LL Lewis lung carcinoma, mouse leukemia (Monsigny *et al*., 1984; Roche *et al*., 1983), baby hamster kidney cells (Stojanovic & Hughes, 1984) and metastatic B16 melonoma cells (Raz & Lotan, 1981). The galactose specific receptor of mammalian, also the mannose/GalNAc specific receptor of avian species and the receptor for mannose-6-phosphate present in a variety of cells are the best characterized lectins of this class (Lis & Sharon, 1986a,b).

Majority of the S-type lectins are of mammalian origin, around 30 kDa size and form dimers of identical subunits (Allen *et al*., 1987). The S-type (thiol dependent) or S-Lac (soluble lactose binding) lectins are ubiquitous, found in different tissues and cells (Caron *et al*., 1990). Unlike the C-type lectins S-type lectins exhibit considerable sequence homology (Drickamer, 1988). Most of the soluble lectins isolated from vertebrate tissues are specific for ß–galactoside (Lotan, 1992). They require a reducing agent to maintain their carbohydrate binding activity. Like in plants, isolectins are found in animals also. The best studied are a group of dimeric proteins found in organisms such as the electric eel, calf, chicken, rat, man and in snake venom. (Barondes, 1984). Of the soluble vertebrate lectins the ß-galactoside specific lectin from the electric organ of eel *electricus electricus* was the first one to be purified (Levi & Teichberg, 1981). These lectins have Mr. in the range 13.5-16.5 kDa, and are homodimers.

Many invertebrates also have lectins present mainly in the hemolymph and sexual organs, for example, on the membranes of hemocytes albumin glands and eggs (Gilboa-Garber *et al.*, 1985). Only few lectins, mainly specific for sialic acid, such as those from snail (*Helix Pomatia*), *crabs (Limulus polyphemus and carcinoscopus rotunda cauda*), lobster (*Homarus americanus*), clam (*Tridacna maxima*) and slug (*Limax flavus*) have been characterized (Lis & Sharon, 1986a,b). The available three-dimensional structures (Drickamer, 1999; Rini, 1999; Weis & Drickamer, 1996) and fold classification of animal lectins are given in table. 1. 1.

1. 3. 3. Microbial lectins

Surface lectins from several bacterial species have been purified and characterized (Bertels *et al.*, 1991; Gilboa-Garber & Avichozer, 1993; Jann & Jann 1990; Sharon, 1987). Bacterial lectins are often seen as filamentous appendages called fimbriae or pili. Fimbriae are usually 5-7 nm in diameter and 100-200 nm in length. The well characterized lectins are (1) type-I fimbriae from *E.coli*, specific for oligomannose and hybrid oligosaccharides of animal cell surface glycoproteins, and (2) type-P fimbriae specific for Gal-a-1,4-Gal (Lis & Sharon, 1986a). Another set of lectins identified as specific for N-acetylneuraminyl-a-2-3-galactosides found in several strains of *E.coli* and are called S fimbriae (Parkkinen *et al.*, 1983).

Virus is an important system in the study of lectin-carbohydrate interactions. More than 100 strains of influenza virus, mainly the A & B type, were tested for their ability to agglutinate the erythrocytes carrying N-acetyl neuraminic acid, the sialic acid lacking an O-acetyl group. Influenza A & B viruses and paramyxoviruses bind gangliosides containing N-acetyl neuraminic acid, whereas influenza C and bovine corona viruses bind receptors containing N-acetyl-9-O-acetyl neuraminic acid, the 9-O-acetyl group of which is critical for mediating cellular attachment (Schultze *et al.*, 1993).

Different saccharides inhibit amoebic infection suggesting the role of lectin-sugar interactions in amoebic adherence. Among the two lectins isolated and characterized from *E.hystolytica*, one is specific for ß1-4 linked oligomers of

GlcNAc (Kobiler & Mirelman, 1981) and the other for Gal and GalNAc (Petri *et al*., 1989). Cellular slime molds can synthesize lectins in such large quantities like $5x10^6$ lectin molecules per cell (Barondes, 1984). These lectins, irrespective of the species from which they are isolated, have a subunit size of about 25 kDa and can bind ß-galactoside. The extensively studied ones in this group are from *Dictyostelium discodium* that synthesizes two tetrameric ß-galactoside binding lectins, discoidin-I and discoidin-II (Fraizer *et al*., 1975). In table. 1. 1., the available three-dimensional structures of microbial lectins (Lis & Sharon, 1998; Weis, 1997) classified according to fold is given.

1. 3. 4. Biological role and applications

Cell recognition ability of lectins forms the basis of their function, hence their role in biology. Role of lectins differs from organism to organism and also in different organs or tissues of the same organism (Singh *et al*, 1999). Variation in lectin binding is observed during embryonic differentiation, cell maturation, malignant transformation, and under pathological state such as lysosomal storage disease, inflammatory bowel disease, psoriasis and pnuemococcal meningitis (Lis & Sharon, 1986a,b; Weis *et al*., 1998). Since lectins have specific carbohydrate binding sites, they are involved in the recognition of complementary oligosaccharide receptors (Sharon & Lis, 1989). Mitogenic stimulation, agglutination and signal transduction are induced through lectin-receptor interactions. Examples are, the cross linkage of ConA and SBA with glycoconjugates that lead to mitogenic activities (Nicolson, 1976), molecular sorting of glycoproteins in the secretary pathway of cells and signal transduction mechanism of certain glycoprotein hormones (Sairam, 1989). The soluble endogenous lectins, analogous to mitogenic lectins of plant origin are capable of inducing signal transducing events after binding to cell surface glycoproteins (Kornberg *et al*., 1991). Galectins play role in cell adhesion. They are thought to be essentail for the normal development and differentiation of all multicellular animals. The endocytic lectins (a class of C-type lectins) are membrane bound receptors. The mammalian hepatic asialoglycoprotein receptor is known to facilitate the clearance of complex oligosaccharide binding

glycoproteins from circulation (Ashwell & Hardford, 1982). Kupffer cells are responsible for the physiological clearance of old erythrocytes from the circulatory system of humans and other animals (Bladier *et al*., 1980). The recognition between lymphocytes and lymph nodes is dependent on lectin-sugar interactions. Selectins, another family of C-type lectins, control the leukocyte trafficking to sites of inflammation and the homing of lymhocytes to specific lymphoid organs (Lasky, 1995; Springer, 1995). Lectins on human and murine metastatic tumor cells reported to be involved in the formation of emboli (Lotan & Raz, 1988).

Mainly two functions are ascribed to plant lectins (Lis & Sharon, 1998; Sharon & Lis, 1989): a) as mediators of symbiosis of plants and microorganisms, b) in protecting plants against phytopathogens. Attachment of nitrogen-fixing bacteria to legumes is very specific and suggested to be mediated by lectins like SBA (Keen & Stasakwicz, 1988). The toxicity of plant lectins for animals and their growth inhibition on fungi are the basis for considering them involved in defense function (Lis & Sharon, 1998).

The adhesion of virus, bacteria and other microbes, a prerequisite for infection, is mediated by lectins (Lis & Sharon, 1998; Sharon & Lis, 1989). Infection by influenza virus is through the recognition of sialic acid by influenza virus heamagglutinin. The role of lectins in causing diarrhea by *E.coli* K99, urinary and gastrointestinal tracts infection by type I of *E.coli* and *K.pneumoniae* and colonization of lung, liver and kidney by *P. aeruginosa* is recognized. Macrophage membrane bound mannose and GalNAc specific lectins are implicated in the binding and phagocytosis (i. e., lectinophagocytosis) of microorganisms (Ofek & Sharon, 1988). The endogenous lectin from *Dictyostelium discoideum*, synthesized as slime aggregates, functions to promote cell substratum attachment and ordered cell migration during morphogenesis (Spriner *et al*., 1984).

Lectins have use in blood typing and mitogenic stimulation of lymphocytes (Lis & Sharon, 1977 and 1986a). Lecitn derivatives are employed as histochemical and cytochemical reagents for the detection of glycoconjugates in tissues, cells and subcellular organells (Rhodes & Milton,

1998). Lectin–Lymphocyte interaction could be used as a tool for understanding of lymphocyte activation and its control (Kilpatrick, 1991). The lectins with ready availability and diverse specificity are used in the study of simple and complex carbohydrate structures in solution and on cell surface (Lis & Sharon, 1984 and 1986a,b). Lectins are used for the identification and separation of various types of cells. PNA, in particular, has been used for the identification and separation of immature thymocytes from mature thymocytes of mice and man, and for the identification of germinal center cells and human thymocytes (Reisner & Sharon, 1984). SBA is used for the separation of mouse T and B splenocytes (Reisner *et al*., 1976).

Plant lectins are employed in the purification and isolation of glycoproteins using affinity chromatography on immobilized lectins (Carlson, 1994), especially bulb lectins which have strict specificity for a-D-mannose (Gillijam, 1993). Some of the lectins can separate glycoproteins having slight difference in glycosylation or oligosaccharides with minor structural difference. This property is used for identifying the pathological condition as a consequence of glycosylation changes under diseased state (Sumar *et al*., 1993). Some of the bulb lectins are also known to be effective inhibitors of HIV infectivity because of binding to high mannose glycans present on the HIV glycoprotein GP120 (Balzarini *et al*, 1991).

Mitogenic lectins, Con A and PHA, find use in assesing the immunocompetence of patients suffering from diverse diseases (Borrebaeck & Carlsson. 1989). Mistletoe lectin (ML-1) bind to tumor & inflammatory host cells in human lung carcinoma. ML-1 has immuno-stimulatory effect on Natural killer (NK) cells, and causes the secretion of cytokines (Gabius *et al*., 1992; Hajto *et al*., 1990; Joshi *et al*., 1991). Certain lectins are known to bind to tumor cells (Louis & Wyllie 1981; Louis *et al*., 1981; Sela *et al*., 1970). *Pisum sativum* (garlic) lectin specifically binds Ehrlich ascites tumor cells (EAT). This property of lectins is useful in elucidating the changes in cell surface architecture upon malignant transformation. Thus lectins are promising tools in cell biology, as drug carriers for therapy, in diagnosis of diseases such as tumor and in tumor imaging.

1. 3. 5. Lectin-Carbohydrate interactions and specificity

Although lectins were discovered some 100 years ago, their involvement in carbohydrate recognition has been seriously dealt with only in the past two decades (Hughes, 1992; Lis & Sharon, 1998; Sharon & Lis, 1989 and 1993). Now it is well known that all cells carry carbohydrates on their surface in the form of glycoconjugates (Cook *et al.*, 1986) and lectins mediate cell-cell recognition by binding surface carbohydrates (Lis & Sharon, 1998). These surface carbohydrates on a cell serve as points of attachment for other cells, and pathogens. Sharon and Lis (1989) have proposed that the lectin mediated cell-cell interaction takes place through mechanisms like: 1) Cell surface lectins binding to soluble glycoproteins can create inter cell bridges. 2) Lectins can bind carbohydrates of insoluble components of the extracellular matrix that promote cell adhesion. 3) Soluble lectins binding to carbohydrates on a pair of opposing cells can act as bridge for these two cells.

The atomic basis of lectin-carbohydrate interactions has been elucidated through X-ray crystallographic analysis of a variety of lectin-carbohydrate complexes The forces, which hold the lectin-carbohydrate complexes together, are hydrogen bond networks, metal coordination, van der Walls and hydrophobic interactions (Drickamer, 1995; Elgavish & Shannan, 1997).

Lectins apply various strategies to generate carbohydrate specificity, evident from crystal structures of lectin-sugar complexes. They are: 1) The variation in quaternary structure generates carbohydrate specificity in the case of legume lectins and bulb lectins (Chandra *et al.*, 1999; Vijayan & Chandra, 1999). In both cases, the small alterations in essentially same tertiary structure lead to different quaternary structures that discriminate between different sugar ligands. 2) Post-translational modification generating carbohydrate specificity has been observed in the structures of jacalin (Sankaranarayanan *et al.*, 1996) and related lectins (Vijayan & Chandra, 1999). 3) Water molecules generating sugar specificity is seen in sugar complexes of PNA, Ecorl and other lectins (Elgavish & Shannan, 1998; Ravishankar *et al.*, 1997). 4) Through their extended binding sites (subsite) and subunit multivalency.

Although lectins bind monosaccharides weakly (in mM range), they show enhanced binding capacity for oligosaccharides (in the nM range) through extended binding site or subunit multivalency (Elgavish & Shannan, 1997). The extended binding sites are seen in the crystal structures of some legume lectins, galectins, bulb lectins, Lathyrus orchus lectin in complex with fucosylated oligosaccharides (Bourne *et al.*, 1994), galectin in complex with lactose and N-acetyl lactosamine and snowdrop lectin complexed with mannopentose (Wright & Hester, 1996).

When multivalent sugar ligands are present, the interaction may be limited to only a terminal sugar residue. For example, in the crystal structures of influenza virus haemagglutinin complexed with sialyllactose, only the terminal sialic acid is in contact with lectin. Similarly in mammalian MBP complexed with an oligomannose asparaginyl-oligosaccharide (Weis *et al.*, 1992) only the terminal mannose is in contact with protein. In such cases higher affinity can result from a clustering of identical binding sites through protein oligomerization and the resultant subunit multivalency. In addition to the clustering strategy the orientation of the binding sites is also crucial for higher affinity and may have functional implication also. In the crystal structures of trimeric influenza haemagglutinin, pentameric bacterial toxins and trimeric fragment of MBP, the sugar binding sites are oriented in one direction and are spaced wide apart (Drickamer, 1995).

1. 4. BOWMAN-BIRK PROTEASE INHIBITORS (BBIs)

Protein inhibitors (PIs) of proteases are ubiquitous in nature. Animals, plants and microorganisms contain a number of 'protease inhibitors' which form reversible, stochiometric protein-protein complexes with various proteolytic enzymes (Bode & Huber, 1992; Laskowski & Qasim, 2000). They are present in multiple forms in different tissues of organisms. For the last several years PIs have been investigated for various reasons which include their utility in the study of protein-protein interactions (Helland *et al.*, 1999; Hubbard *et al.*, 1991; Krystek *et al.*, 1993). Their gross physiological function is to prevent unwanted proteolysis and thereby to control the protein turnover and metabolism. The

most abundant source of PIs is plants and their presence in plants was known since 1938 (Richardson, 1977; Ryan, 1973). The PIs are widely distributed throughout the plant kingdom, especially in Leguminosae, Gramineae and Solanaceae families, and are usually present in the storage tissues such as seeds and tubers (Richardson, 1991; Shewry, 1999). They are usually localized inside the vacuolar protein bodies of the cell. A large number of PIs have been isolated and characterized (Valueva & Mosolov, 1999). Mostly they are low molecular weight proteins (4-20 kDa), soluble in water and their polypeptide chains are non-glycosylated.

Protease inhibitors have been classified into four families based on their specificity for the four mechanistic classes of proteolytic enzymes, serine, cysteine, aspartic and metallo proteases (Ryan, 1990). A particular inhibitory reactive site usually inhibits proteases belonging to one of the four mechanistic classes. The inhibition is strictly competitive. Majority of PIs are specific for serine proteases, the most widespread and most extensively studied enzymes (Ryan, 1981; Ryan, 1990). Plants contain different types of serine protease inhibitors (serpins), that inhibit trypsin and trypsin-like enzymes of animal, fungal and bacterial origins. Many serpins are multi headed type as a result of internal gene duplication and fusion (Laskowski & Kato, 1980). 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 structure includes more than 15 families (Bode & Huber, 1992; Laskowski & Kato, 1980; Laskowski & Qasim, 2000). Among them the Bowman-Birk family and the Kunitz family PIs are the most extensively studied ones.

1. 4. 1. Occurrence and properties of BBIs

The Bowman-Birk Protease inhibitor (BBI) was isolated for the first time from soybeans by Bowman (1946) and its biochemical properties were studied by Birk *et al*., (1963). Subsequently many BBIs were isolated and characterized from legumes (only from Fabaceae family), Gramineaes and many other plants (Ikenaka & Norioka, 1986). They are present in seeds, tubers and roots of

monocots and dicots (Birk, 1985). The special features of dicot members of this family include (Sreerama & Gowda, 1997): 1) their small molecular weights, range between 6-9 kDa; 2) contain seven disulfide bridges that stabilize their active configurations; 3) have two tandem homologous domains, each possessing an insertion loop, capable of inhibiting two protease molecules simultaneously and independently, thus are called “double headed” inhibitors.

BBIs from dicotyledonous seeds are of 8 kDa size and double headed. In contrast, the monocots have 8 kDa single headed and 16 kDa double headed inhibitors (Prakash *et al.*, 1996). Dicot BBIs would have evolved from single headed ancestral BBI via internal gene duplication, mutation and fusion (Odani & Ikenaka, 1978). It is suggested that during evolution, one of the reactive sites of 8 kDa double headed BBI became non-functional, that resulted in a 8 kDa single headed BBI in monocots (Song *et al.*, 1999). Subsequently the 16 kDa double headed BBI must have evolved from this 8 kDa one by gene duplication and fusion. It may be significant that the two reactive site loops and disulfide bridges are well conserved among monocot and dicot BBIs which also show 35% sequence homology. Generally the double headed dicotelydenous BBIs inhibit trypsin at the first site and chymotrypsin, elastase, subtilisin, etc. at the second site (Laskowski & Kato, 1980), whereas monocots inhibit only trypsin (Song *et al.*, 1999). The members of the dicot BBI family show homology of 50-80% in their primary structure with the disulfide bridges fully conserved (Ferrason *et al.*, 1995; Ikenaka & Norioka, 1983); aligned sequences are shown in Fig. 1. 4.

Most of the BBI type PIs exist in various isoforms called ‘isoinhibitors’ that differ slightly in sequence and/or length. The presence of mutiple isoinhibitors is attributed to the encoding of BBIs by multigenes and the post translational modification of a few amino acids at the N- or C-terminal of the inhibitors (Quillien *et al.*, 1997). BBIs usually self associate to form dimers, trimers and more complex oligomers (Gennis & Canter, 1976; Laskowski & Kato, 1980). Several BBIs have been widely studied for their anticarcinogenic activity (Kennedy, 1993) and for their immune stimulating properties (Hams-Ringdahl *et al.*, 1979). They show high resistance towards acidic pH in the

digestive systems of humans and animals and are known to be stable at cooking temperatures (Birk, 1987).

```
                                                                  P3  P1'    P5'
S1  ---------------QVIKSGD------------HH-EATDEPSESSEACCDRCECTKSIPP--QC
S2  -------------------SGD------------HHQDFTDEPSE— EACCDQCECTKSIPP--QC
S3  -------------------SGH----------- HZB—STBZASZSSKPCCRZCACTKSIPP--ZC
S4  ------------------LSFAANVVNARFDSTSFITQVLSNGDDVKSACCDTCLCTKSNPP—-TC
S5  ----------------------------------------GDDVKSACCDTCLCTKSNPP—-TC
S6  MELMNKKVMMKLALMVFLLSFAANVVNARFDSTSFITQVLSNGDDVKSACCDTCLCTKSDPP--TC
S7  -------------------SGHR----------HESXBSTBXASXSSKPCCBHCACTKSIPP--QC
S8  ----------------------------------------GDDVKSACCDTCLCTRSQPP--TC
S9  ----------------------------------------GDDVKSACCDTCLCTKSEPP--TC
S10 --MVVLKVCLVLLFLVGGTTSANLRLSKLGLLMKSDHQHSNDD—ESSKPCCDQCACTKSNPP--QC
S11 ------------MCIL-------------SFLKSDQSSSYDDDEYSKPCCDLCMCTRSMPP--QC
S12 -------MVVLKVCFLVLFLVGVTNAHMELDLFKSDHSSS— DDESSKPCCDLCMCTASMPP--QC
S13 ----------------------------------HEHSSDESSESSKPCCDLCTCTKSIPP--QC
S14 ----------------------------------DHHHSTDEPSESSKPCCDECACTKSIPP--QC
S15 -------------------SG-------------HHDETTDEPSESSKPCCDQCSCTKSMPP--KC
S16 -------------------SG-------------HHEDTTDEPSESSKPCCDQC-CTKSMPP--KC
S17 --------------------------------SVHHQDSSDEPSESSHPCCDLCLCTKSIPP--QC
S18 ----------------------------------DHHQSTDEPSESSKPCCDQCTCTKSIPP--QC
S19 -------------------SGH----------- HEH—STDZPSZSSKPCCBHCACTKSIPP--QC
S20 ------------------------------------SHDEPSESSEPCCDSCDCTKSKPP--QC
S21 --------------------------------BHHZSSBBZPSZSSPPCCBICVCTASIPP--QC
S22 --------------------------------BHHZSSBBZPSZSSPPCCBICVCTASIPP--QC
S23 ------------------------------------------SSGPCCDRCRCTKSEPP--QC
S24 --------------------------------DDDHSDDEPRESESSKPCCSSC-CTRSRPP--QC
S25 ----------------------------------DHSDD---ESESSKPCCDECKCTKSEPP--QC
S26 ------------------------------------------TSACCDKCFCTKSNPP--IC
S27 ----------------------------------------TKSTTTACCDFCPCTRSIPP--QC
S28 ------------------------------------------TTACCNFCPCTRSIPP--QC
S29 ------------------------------------EASSSSDDNVCCNGCLCDRRAPPYFEC
S30 ----------------------------------------AASD---CCSACICDRRAPPYFEC
                                                    **  * *    **   *
```

```
                        P3  P1'   P6'
S1   RCSDVRLNSCHSACKSCACTFSIPAQCFCGDIND-SC-YKPCKSSSHDDDDWDK 88
S2   RCSDVRLNSCHSACKSCACTFSIPAQCFCGDIND-FC-YKPCKSDSHDD----- 78
S3   RCSZVRLNSCHSACKSCACTFSIPAZCFCGBIBB-FC-YKPCKSSHSBBBBWN- 83
S4   RCVDVR-ETCHSACDSCICAYSNPPKCQCFDTHK-FC-YKACHNSEVEEVIKN- 96
S5   RCVDVG-ETCHSACLSCICAYSNPPKCQCFDTQK-FC-YKQCHNSELEEVIKN- 72
S6   RCVDVG-ETCHSACDSCICALSYPPQCQCFDTHK-FC-YKACHNSEVEEVIKN- 114
S7   RCSBLRLNSCHSECKGCICTFSIPAQCICTDTNN-FC-YEPCKSSHGPBBNN-- 85
S8   RCVDVG-ERCHSACNHCVCNYSNPPQCQCFDTHK-FC-YKACHSSEKEEVIKN- 72
S9   RCVDVG-ERCHSACNSCVCRYSNPPKCQCFDTHK-FC-YKSCHN---------- 63
S10  RCSDMRLNSCHSACKSCICALSYPAQCFCVDITD-FC-YEPCKPSEDDKEN--- 110
S11  SCEDIRLNSCHSDCKSCMCTRSQPGQCRCLDTND-FC-YKPCKSRDD------- 83
S12  HCADIRLNSCHSACDRCACTRSMPGQCRCLDTTD-FC-YKPCKSSDEDDD---- 103
S13  HCNDMRLNSCHSACKSCICALSEPAQCFCVDTTD-FC-YKSCHNN-AEKD---- 76
S14  RCTDVRLNSCHSACSSCVCTFSIPAQCVCVDMKD-FC-YAPCKSS-HDD----- 76
S15  RCSDIRLNSCHSACKSCACTYSIPAKCFCTDIND-FC-YEPCKSSRDDD--WDN 82
S16  RCSDIRLDSCHSACKSCACTYSIPAKCFCTDIND-FC-YEPCKSSRDDD--WDN 81
S17  HCANIRLNSCHSACKSCICTRSMPGKCRCLDTDD-FC-HKPCKSRDKD------ 78
S18  RCTDVRLNSCHSACSSCVCTFSIPAQCVCVDMKD-FC-YAPCKSS-HDD----- 76
S19  RCTDLRLDSCHSACKSCICTLSIPAQCVCBBIBD-FC-YEPCKSSHSDDDNNN- 83
S20  HCANIRLNSCHSACKSCICTRSMPGKCRCLDTDD-FC-YKPCESMDKD------ 72
S21  QCADIRLDSCHSACKSCMCTRSMPGQCRCLDTHD-YC-YKSCKSBSGZBB---- 79
S22  VCTBIRLBSCHSACKSCMCTRSMPGKCRCLBTTB-YC-YKSCKSBSGZBB---- 79
S23  VCTBIRLBSCHSACKSCMCTRSMPGKCRCLBTTB-FC-HEPCKSSGDDED---- 67
S24  QCQDVRLNSCHSACEACVCSHSMPGLCSCLDITH-FC-YKPCKSSGDDDZZ--- 80
S25  QCTDVRLNSCHSACKSCMCTFSDPGMCSCLDVTD-FC-YKPCKSGGGDEDD--- 76
S26  QCVDTRLESCHSACKLCLCALSFPAKCRCVDTTD-FC-YDKCDSDS-------- 61
S27  QCRDVG-ETCHSACKFCICALSYPAQCHCLDQNT-FC-YPSCR---------- 62
S28  QCTDVR-EKCHSACKSCLCTLSIPPQCHCYDITD-FC-YPKCN---------- 58
S29  RCTDIG-ETCHSACKTCLCTKSIPPQCHCADITNGRCPVTECRS---------- 70
S30  VCVDTF-EHCPASCNSCVCTRSNPPQCRCTDKTQGRCPLTPCA---------- 63
      *       *   *  * *    *  * *       *    *
```

Fig.1. 4. Multiple sequence alignment using CLUSTAL W (Higgins & Gibson, 1995) of several representative 'double headed' dicot PIs of Bowman-Birk family. The * indicates conserved residues and the two binding site loop residues are indicated through Pn to Pn' (Schetcher & Berger, 1967). Only those sequences with less than 90% sequence homology are included in the

list. The sequences are, S1 & S2-Cowpea (Isoinhibitors), S3-Cowpea (isoinhibitor), S4-BBI Pea seeds, PSTI-IVa (precursor), S5-Pea seeds (Chain A, crystal structure), S6-Garden pea, TI12-36, S7-Garden bean, PVI-3, S8-*Vicia angustifolia* (common vetch), S9-fava bean, S10- Soybean, D-II (Precursor), S11-Soybean, PI-IV, S12- Soybean CII (precursor), S13-*Macrotyloma axillare* seed DE-4, S14-*Macrotyloma axillare* seed DE-3, S15-adzuki beans PI-I, S16-adzuki beans, PI-II, S17-adzuki bean-IA, S18-horse gram seeds, S19-Lima beans, S20-Mung bean beans, S21-Kidney bean-II, S22-Garden bean-II, S23-*Dioclea glabra*, DGTI-I, S24-apple leaf seed-DE4, S25-*Canavalia lineata*, CLTI-II, S26- *Erythrina variegata* seeds, EBI, S27- snail medic seeds, MSTI, S28-alfalfa leaves, S29- Peanut A-II, S30-Peanut B-II.

Kunitz type PIs (Kunitz, 1947), the other class of extensively studied serpins, are usually single headed types with a molecular mass of 20 kDa. They consist of one or two polypeptide chains with four disulfide bridges. The three-dimensional structure of Bovine pancreatic trypsin inhibitor (BPTI), in complex with trypsin, was the first structure of a Kunitz family PI to be solved by X-ray crystallography (Huber *et al.*, 1970 and 1974). It contained six two-stranded ß-sheets connected through hairpin loops classified as ' ß-trefoil fold'.

1. 4. 2. Three-dimensional structure

Many BBIs have been isolated, characterized and amino acid sequences were determined. The three-dimensional structures of few dicot BBIs and only one from moncot have been reported. Table. 1. 2., lists the known structures of BBIs. Except for one structure, of soybean BBI determined in solution by NMR spectroscopy, all others have been determined by X-ray crystallography.

The fold of the polypeptide chain of BBIs is described as “bow tie motif” (Chen *et al.*, 1992). This has a central core of two pseudo-symmetry related domains each consists of three antiparallel ß-strands, which associate into two short ß-sheets connected by four loops, and the N- and C-terminal segments fold into an extended conformation (Fig. 1. 5. A). The three-dimensional structure is described in detail in the chapter 5 in relation to *Vigna* BBI structure.

BBIs are found as monomers (Figs. 1. 5. A and C), dimers (Figs. 5. 8. A-B and 5. 9. A-B) and tetramers in their crystal structures. The dimer forms the asymmetric unit in the crystal structure of pea BBI (de la Sierra *et al.*, 1999). The monomers of peanut A-II BBI form tetramer with an approximate 222 symmetry in the crystal structure (Suzuki *et al.*, 1993).

There are two crystal structures, one of a 16 kDa double headed monocot BBI from barley seeds (Song *et al.*, 1999) and another of a small 14 residue peptide inhibitor from sunflower seeds (Luckett *et al.*, 1999), worth discussing in the context of the BBI structures. Former is the highest resolution structure (1.9 Å) available to date of a BBI. This BBI is a single polypeptide chain consists of 125 amino acids with two trypsin inhibitory loops. The structure is made up of 11 ß-strands and the loops connecting them (Fig. 1. 5. C). It has no a-helices. The structure is closer to tracey bean BBI structure of dicot family. The polypeptide chain folds into two compact domains (N and C), the fold of each domain resembles that of 8 kDa dicot BBIs, and consists of two antiparallel ß-sheets of different size. Each domain has five disulfide bridges including the one at the reactive site loop. These five disulfide bridges are a subset of the seven bridges of the classical dicot BBIs. The two missing ones are located near the chymotrypsin binding loop corresponding to dicots BBIs. The loss of this disulfide bond is suggested to be responsible for this loop losing activity in each subdomain of the monocot BBIs (Song *et al.*, 1999).

A BBI type PI from sunflower seeds is the smallest naturally occurring plant protease inhibitor reported to date (Luckett *et al.*, 1999). This novel cyclic peptide shows both sequence and conformational similarity with the trypsin reactive site loop of BBIs. This one is considered a potent PI of trypsin (Ki=100pM) and its high potency is attributed to structural rigidity arising out of the cyclic nature of the chain fold and due to a disulfide bond present. The peptide consists of two antiparallel ß-strands connected by an extended reactive site loop at one end and a hairpin loop at the opposite end (Fig. 1. 5. D). The peptide is unique in being monofunctional, cyclic, shorter and thus differ from the classical BBIs.

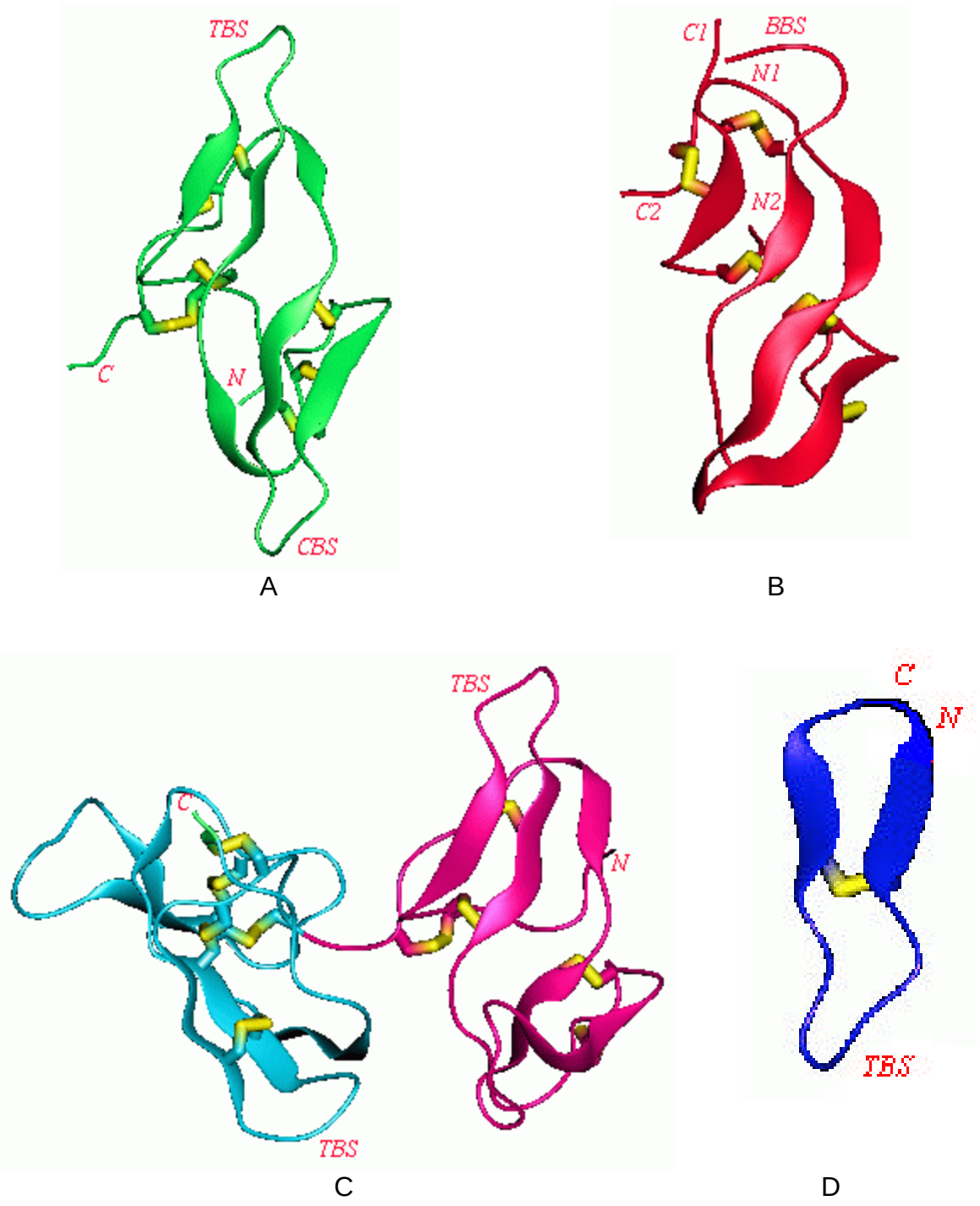


Fig. 1. 5. Monomer structures of, A) dicot BBI from tracey bean, B) Bromelain inhibitor IV from pineapple stem, C) monocot BBI from barley showing two domains in different colors, D) small peptide inhibitor from sunflower seeds. Disulifde bridges are shown in yellow, TBS/CBS-trypsin/chymotrypsin binding site, BBS- Bromelain binding site.

Table. 1. 2. The list of available three-dimensional structures of Bowman-Birk protease inhibitors. (ND: not deposited).

Protein source /PDB code	**Structure**	**Structure determina-tion method**	**Resolut -ion (Å)**	**Reference**
Adzuki beans/ 1TAB	Monomer in complex with bovine tyrpsin	MIR	2.3	Tsunogae *et al*., 1986
Tracey soybean/ 1PI-2	Native dimer	MIR	2.5	Chen *et al*., 1992
Soybean seeds / 1BBI, 2 BBI	Native monomer in solution	NMR	_	Werner & Wemmer, 1992a,b
Peanut A-II/ ND	Native tetramer	MIR	2.3	Suzuki *et al*., 1993
Mung bean / 1SMF	Monomer in ternary complex with porcine trypsin	MR	2.5	Lin *et al*., 1993
Soybean seeds/ 1K9B	Native monomer	MR	2.8	Voss *et al*., 1996
Pea seeds/ 1PBI	Native dimer	MR	2.7	De la Sierra *et al*., 1999
Soybean seeds/ 1D6R	Ternary complex with bovine trypsin	MR	2.3	Koepke *et al*., 2000
Mung bean/ 1DF9	Ternary complex with Dengue virus NS3 protease	MAD	2.1	Krishnamurthy *et al*., 2000
Barley seeds/ 1C2A	Native monomer	MAD	1.9	Song *et al*., 1999
Sunflower seeds 14-residue peptide/ 1SFI	Complex with bovine ß-trypsin	MR	1.6	Luckett *et al*., 1999

1. 4. 3. The binding loops and inhibition

The two protease binding loops of BBI consist of a conserved disulfide bridge and nine to eleven amino acid residues that are largely conserved (Fig. 1. 4). The conformations of the two reactive site loops are highly conserved across all known BBIs. In most BBIs the first reactive site inhibits trypsin and the second chymotrypsin. The P1 (Schetcher & Berger, 1967) residue in BBIs has to be arginine or lysine for trypsin binding, leucine or phenylalanine or tyrosine for chymotrypsin and alanine for elastase binding (Laskowski & Kato, 1980).

Serpins, in general, block the access of the substrates to the catalytic site of their cognate proteases through steric hindrance (Bode & Huber, 1994 and 2000). The reactive sites of BBI, like any other PI, assume a characteristic 'canonical' conformation to bind to the active site of the enzyme in the form of a productively bound substrate. At the same time BBI can bind like an inhibitor rather than a substrate due to its conformational rigidity and structural complementarity with the active site of the enzyme. Unlike in the case of a good substrate the energy barrier for the hydrolysis of the inhibitor is large and unfavorable, hence the rate of hydrolysis is very slow (Read & James, 1986). The interaction between PI and enzyme can be described by the following minimal scheme proposed by Bode and Huber (1992).

$$E + I \rightleftharpoons EI \rightleftharpoons E + I^{*}.$$

The stable enzyme-inhibitor complex (EI) is rapidly formed but dissociates very slowly into free enzyme (E) and virgin (I) or modified (I*) inhibitor. The modified inhibitor (I*) is specifically cleaved at the scissile P1-P1' peptide bond. In most cases the protease-inhibitor interaction are characterized by large K_{cat}/ K_m, however, both K_{cat} and K_m are usually very small at neutral pH and this leads to very tight binding (low K_m) and slow hydrolysis (low K_{cat}).

Four crystal structures of the BBI-trypsin complexes are available (table. 1. 2.), but none with chymotrypsin. Information on BBI-chymotrypsin interactions are obtained through modeling based on reported crystal structures of trypsin-inhibitor complexes (de Freitas *et al.*, 1997; de la Seirra *et al.*, 1999). The details of the interactions are discussed in the fifth chapter in the context of modeled *Vigna* BBI-bovine trypsin complex.

1. 4. 4. Comparison of BBIs with other serpins

Despite lack of similarity in sequence and overall structure, the members of the family of serpins share a common conformation at the reactive site loop defined by residues from P3 position till P3' (Hubbard *et al.*, 1991; Laskowski & Qasim, 2000). The geometry of the region surrounding the reactive site is also virtually identical. The reactive site loop protrudes on the surface of the protein, thus placing it easily accessible to proteases. But the inhibitor structure diverges out from the binding loop towards both N- and C-terminus. The similarity between trypsin binding loop of BBIs and that of BPTI belonging to Kunitz family and of PSTI (Porcine soybean trypsin inhibitor) belonging to Kazal family has been shown (Chen *et al.*, 1992). Similarity exists for the chymotrypsin binding loop of BBI with the binding loops of potato chymotyrpsin inhibitor-1, turkey ovomucoid inhibitor third domain and barley chymotrypsin inhibitor-II (Werner & Wemmer, 1992b).

When considering the differences between BBIs and other serpins, BBIs are double headed and bind two protease molecules simultaneously, whereas the other serpins are usually single headed hence bind only one protease molecule at a time. BBIs have seven disulfide bridges and a hydrophilic core and a hydrophobic surface in contrast to other serpins which have a hydrophobic core and less than seven disulfide bridges (Werner & Wemmer, 1992a; Voss *et al.*, 1996).

Interestingly, a cysteine PI known as bromelain inhibitor VI from pineapple stem, was found to share a similar fold and dislufide connectivity pattern with the BBIs but not with the cystatin superfamily PIs (Hatano *et al.*, 1996). This PI is made up of a 11 residue light chain and a 41 residue heavy

chain and five disulfide bridges. This structure has two distinct domains each containing a three-stranded ß-sheet (Fig. 1. 5. B). In terms of its position the bromelain binding site corresponds to the chymotrypsin binding site of BBIs. In the other site the PI lacks a type 'VIb' turn and on the whole is a weak PI of both trypsin and chymotrypsin. The strutcural similarities indicate that these proteins might have evolved from a common ancestor and diverged in function during the course of molecular evolution (Hatano *et al*., 1996).

1. 4. 5. Role of BBIs in plants and their applications

Protease inhibitors in general are known to play an important role in nature by regulating the proteolytic activity of their target proteases (Bode & Huber, 1992). The prevention of many unwanted proteolysis includes, preventing premature activation of trypsinogen in the vertebrate pancreas, avoiding excessive proteolysis in lung tissues, delimiting blood clotting in mammalian plasma, controlling hormone production, etc., (Laskowski & Kato, 1980).

Three possible functions have been identified for PIs in plants (Bode & Huber, 2000; Ryan, 1990; Valueva & Mosolov, 1999; Wilson, 1981). They are: (1) defense against the digestive enzymes of microbial, avian, or mammalian predators; (2) storage of sulfur containing amino acids in seeds; (3) inhibition of endogenous proteases and regulation of proteolytic activity.

BBIs were found to be wound inducible (EckelKamp *et al*., 1993) and inhibit digestive enzymes of various insects (Sakal *et al*., 1989). Also they are more inhibitory to exogenous than endogenous enzymes. In a recent report, the antifungal, anti-HIV and mitogenic activity of a BBI from broad beans (*Vicia faba*) has been demonstrated for the first time (Ye *et al*,. 2001). Thus, implicating their role in plant defense and suggesting their potential to act as biological insecticides. PIs protect the plant from the invading pests by inhibiting the proteolytic process of pathogens and thereby limiting the availability of amino acids for growth and multiplication of the pathogen (Ryan, 1990). These proteins are also known to be involved in the protection of the plant seeds against environmental hazards during the initial phases of germination and in

the establishment of seedling. The defense role of BBI was demonstrated through the transgenic expression of BBI gene from cowpea in tobacco plant (Ghoshal *et al*., 2001; Hilder *et al*., 1987; Ussuf *et al*., 2001).

BBIs are known to be potent cancer chemopreventive agents, implicated by virtue of their chymotrypsin inhibitory activity (Flecker, 1993; Kennedy, 1998). However, the mechanism of action is not yet clear, but it is likely that they act by recognizing a protease. The inhibitors are effective at extremely low levels of dosage (nM) and act in an irreversible manner without toxicity (Yavelow *et al*, 1985). The striking difference between the BBIs and other chemical anticarcinogens is the ability of the former to prevent different types of cancers. Human populations consuming a large amount of BBI in their diet have been shown to exhibit lower rates of colon, breast, prostate and skin cancers (Birk, 1993). Several studies have shown that the soybean derived BBI can prevent carcinogenesis *in vivo* and malignant transformation *in vitro* (Kennedy, 1994: Kennedy *et al*., 1996). The study by Kennedy *et al*., (1996) demonstrated that soybean BBI can suppress carcinogenesis in mice known to be genetically predisposed to the development of intestinal tumors. Also they suggest the possibility of preventing the cancer in humans who are genetically predisposed to colon cancer.

Among the ultimate goals of enzymology the elucidation of high resolution structures of transition state of enzyme-substrate complexes is one. The lifetime of the enzyme-inhibitor complexes is long, hence can serve as excellent models for studying transition state intermediates (Bode & Huber, 2000; Laskowski & Qasim, 2000). The study of BBI-serine protease interaction provides a way to characterize the functional details of proteases. Information obtained from such complexes could be used in the discussions of hydrolytic mechanism of serine proteases (Read & James, 1986) and in the study of kinetic versus thermodynamic stability of the enzyme-inhibitor complexes (Flecker, 1995; Jensen *et al*., 1996). BBIs have been used in protein folding studies to understand the effect of various factors on the protein folding pathway (Flecker, 1995; Philip *et al.*, 1998).

BBIs, owing to their specific properties including high stability, defense role, specificity and mode of action, hold considerable interest in medical and agricultural field (Birk, 1985) Clinical trials in the light of the anticarcinogenic activity of BBIs critically dependent on the production of highly pure protein in large quantities. Using modern molecular biology techniques the cDNA of BBI was cloned, expressed in *E.coli* and the recombinant BBI obtained (Flecker, 1987). It is now feasible to prepare many variants of BBI for various studies to understand structure-function relationships, to tune its anticarcinogenic action, to produce more potent PIs and for rational drug design. The PI genes of several legume seeds are thought to be target candidates for improving plant resistance against arthropod parasites in an economical and environmental friendly manner (de Freitas *et al*., 1997).

1. 4. 6. Protein-Protein interactions

Many biological phenomena in nature depend on specific interactions among proteins. Vast majority of the proteins interact with one another at some time or other during their existence and their ability to form specific complexes is fundamental to biological existence itself (Bogan & Thorn, 1998). The life cycles of cells are controlled by interactions of proteins present in metabolic and signaling pathways, in the molecular machines that synthesize and use ATP, in replication and translation of genes, or in the formation of cytoskeletol infrastructure. The role of protein-protein interactions in the regulation of enzyme catalysis, in the control of gene expression, in some neurological disorders such as Cruetzfold-Jacob and Alzheimer's disease and in drug-receptor interactions are well established (Feeney, 1999).

The X-ray crystal structure studies have greatly contributed towards the understanding of relationships between sequence, structure, dynamics, energetics and function of protein-protein interactions (Helland *et al*., 1999; Hubbard *et al*., 1991; Krystek *et al*., 1993). The three-dimensional structures have contributed to extensive characterization of the protein-protein interface (Conte *et al*., 1999; Jones & Thornton, 1997). The kinetic (Janin. 1997; Northrup & Erickson, 1992; Schreiber & Fersht, 1996) and thermodynamic

(Janin, 1995; Janin & Chothia, 1990) studies based on theoretical and experimental work have also helped in understanding the affinity and specificity of protein-protein recognition. Affinity determines the magnitude of protein-protein interactions and specificity decides the molecular selection among related chemical species at the cellular level. The general reversible binding features, van der Walls forces, electrostatic interactions and hydrophobic forces contribute to protein-protein interactions, just as they do in many biological systems. But the specific binding features on protein surfaces are not yet fully understood (De Lano *et al.*, 2000).

The simple form of protein-protein interactions is found in the association of monomeric subunits in oligomeric proteins. The highest number of monomeric associations (60, 180 and 240) occur in viral coat proteins (Johnson, 1996). The other examples are the protease-inhibitor and antigen-antibody complexes, proteins of electron and energy transfer reactions, proteins of signal transduction pathways and proteins involved in control of gene expression. Protease-inhibitor complexes provide one of the best models for studying protein-protein interactions and represent the structurally well studied system (Hubbard *et al.*, 1991). Some of the fundamental properties of the interface that characterize the protease-inhibitor complexes include, the shape and size of the interface, affinity and forces that stabilize the complexes, nature and propensity of amino acid residues, conformational changes upon complexation, segmentation and secondary structure of the polypeptide chain at the interface (Jones & Thornton, 1996). Important features that characterize the interface of a protein-protein complex are shown in table. 1. 3. The data is based on 24 protease-inhibitor complexes and 19 antigen-antibody complexes (Conte *et al.*, 1999). More details of protein-protein interactions is discussed in chapter 5.

Table 1. 3. Statistics of parameters that characterize the protein-protein interface.

Parameter	**Protease–Inhibitor complex**	**Antigen–antibody complex**
Interface Area (B, Å^2)	1530	1680
Average number of hydrogen bonds at the interface	9	9
Average number of water molecules at the interface	18	20
Average number of amino acids	50	50
Average number of protein atoms	200	190
Chemical character (% interface area)		
Non polar	61	51
Polar	30	34
Charged	9	15

1. 5. GENESIS OF THE THESIS

The role of lectins and protease inhibitors as defense proteins in plants has been demonstrated in several studies. Lectins mediate cell-cell interactions through the recognition of specific carbohydrates present on the cell surface. Protease inhibitors regulate the activity of proteolytic enzymes by tightly binding to them in a substrate like manner. The role of protein crystallography has been to determine the unique structural features which confer on them the

recognition function to bind their ligands specifically. The three-dimensional structures of lectins and PIs have attracted the attention of structural biologists for understanding the structural basis of protein-carbohydrate recognition in the case of lectins and protein-protein interactions in the case of PIs. The structure can also help in elucidating the mechanism of action and protein evolution. Since lectins and protease inhibitors have been implicated in several diseases they find application in pharmaceutical industry, the three dimensional structures are models for rational drug design.

Here we have studied in detail the three-dimensional structure of a lectin from the seeds of the plant *Artocarpus hirsuta* and carried out the purification, characterization and X-ray structural studies of a Bowman-Birk protease inhibitor from *Vigna unguiculata* seeds. The lectin belonged to jacalin family of which only one structure was available when we started this work. Similarly no dimeric structure of a BBI was known when starting the work on *Vigna* BBI.

Since lectin crystallized in different crystal forms we could analyse the difference in aggregation at subunit level. Also we obtained sugar bound molecule in every crystal form of lectin structure. This has elucidated the lectin-galactose interactions in particular and tentative reasons for galactose specificity. In the case of BBI structure, the dimer in the asymmetric unit helped us to understand the stabilizing interactions for its dimer formation. Since we could not obtain crystals of the BBI with cognate protease we carried out modeling of such a complex and compared it with other experimentally observed complex structures.

CHAPTER-2

MATERIALS AND METHODS

2. 1. INTRODUCTION

In this chapter details of the materials, methods and strategies used by the author for purification and characterization of BBI, crystallization, X-ray data collection and data processing, structure solution and refinement of both lectin and BBI are described. The software used for the analysis of structures and for structure comparison with the proteins belonging to their related families has been detailed. The purification using FPLC (Fast Protein Liquid Chromatography) was attempted for separating the isoinhibitors of BBI. Hanging drop vapor diffusion technique was used for growing single crystals of the BBI, lectin and also penicillin V acylase (PVA) described in Appendix 1. The X-ray diffraction data were collected on MAR image plate detector mounted on Rigaku rotating anode X-ray generator and processed using DENZO and SCALEPACK programs (Otwinowski, 1993; Otwinowski & Minor, 1997). The structures of both lectin and BBI were solved by Molecular Replacement (MR) method implemented in AMoRe program (Navaza, 1994; Navaza & Saludijan, 1997; Navaza, 2001). The program REFMAC (Murshudov *et al.*, 1997) was used for refinement and QUANTA (1997) for display and model fitting of the structures. Programs from CCP4 suite (Collaborative Computational Project, Number 4, 1994) were used for other calculations.

2. 2. MATERIALS

Methyl-a-D-galactose, Tris(hydroxymethyl)aminomethane, DEAE-Sephadex, Sephadex G-50 and G-100, Trizma, NaCl, Polyethylene glycol (PEG), Ammonium sulfate, Bovine ß-trypsin, Porcine a-chymotrypsin, Molecular weight marker kits for both gel filtration and SDS-PAGE were purchased from Sigma chemical company, St. Louis, USA and multiwell trays for crystallization from Becton Dickson and Company. All other chemicals used were obtained locally and mostly of analytical grade. The matured *hirsuta* seeds (Fig. 1. 1. C) were collected towards the end of summer or onset of monsoon from Kerala in South India. The tree and ripen fruit of *hirsuta* fruit are shown in Figs 1. 1. A and

B. The dry *Vigna* (cowpea) seeds (Fig. 1. 1. D) were procured from the local market. The collected seeds were stored at 4 ?C till use.

Silicon graphics Octane workstation was used for graphics display, for running the crystallographic and other programs and for related calculations. The slab gel electrophoresis unit and the tube gel isoelectric focussing (IEF) unit were from Biotech, India. The density gradient single glass column IEF unit was developed in our laboratory (Chinnathambi *et al*., 1995). The FPLC machine (Gradient programmer GP-250 plus) and MonoQ column were of Pharmacia make, Uppsala, Sweden. Most of the software used for crystallographic calculations were from 'CCP4 suite'. The program QUANTA is a product of Molecular Simulations Incorp. (MSI), (presently Accelrys Inc., San Diego, USA). The 'NACCESS' program for surface accessibility calculations was obtained from Dr. Hubbard of Dept of Bimolecular sciences, Manchester, U.K. The '3D-Dock' docking programs were provided by Prof. Graham Smith of Bimolecular Modeling laboratory, Imperial Cancer center (U. K.) Other programs used were freeware down loaded from internet.

2. 3. METHODS USED FOR CRYSTALLOGRAPHY

2. 3. 1. Crystallization

The first step considered crucial in protein structure determination is the growth of diffraction quality single crystals. In the absence of any single concrete theory behind the mechanism of crystallization, we have treated the protein crystallization as a trial and error procedure invoking experience and crystallization reports as guiding principles. It is accepted that the presence of impurities, ionic strength, pH, temperature, precipitating agent and several unspecified factors play role in crystallization process. In crystallization experiments carried out by us, a precipitant such as, PEG, salt, or an organic solvent is diffused into a pure protein solution maintained at particular pH and temperature such that diffraction quality crystals grow from the solution (Drenth & Haas. 1998). Among various crystallization techniques known, hanging drop vapor diffusion method is widely adopted and has produced more crystallized proteins than all other methods combined (Chayen, 1998). This method is

simple, consumes less protein and it is easy to monitor the progress of crystallization. In a typical experimental set up using multiwell trays, 2 µl of protein solution was placed on a siliconized cover slip, mixed it with 2 µl of the precipitant solution and allowed to slowly equilibrate against 1 ml reservoir solution of the precipitant.

The concentrated solutions of pure lectin, BBI and PVA were prepared in saline solution or buffers before setting up crystallization. The concentrations of proteins were estimated as described by Lowry *et al*., (1951). Initial crystallization trials of lectin did not yield crystals. Subsequently it was found that the presence of methyl-α-D-galactose (me-α-gal) is essential for growing crystals. The standardized protocol for crystallization had concentrated lectin mixed with 0.5 M me-α-gal solution in 10:1 proportion prior to setting up crystallization. Different precipitants such as ammonium sulfate, PEGs of different molecular weights, MPD (2-methylpentan-2,4-diol) were tried for crystallization.

2. 3. 2. Data collection

After growing protein crystals of suitable size and quality the immediate step in an X-ray diffraction experiment is to measure the intensities of Bragg reflections. Data collection is best performed as a highly interactive process. The materials required are a crystal, an X-ray source with shutter, an area detector, and a goniometer to orient and rotate the crystal. Protein crystals diffract X-rays much less than do the small molecule crystals; hence the diffraction intensity is weak. Thus protein data collection requires high intense X-ray source and high sensitivity area detector. However, in the last decade the combination of powerful tools such as, X-ray Synchrotron sources, Image plate and area detectors and software's used for collecting and processing the data, have transformed protein crystallography into a powerful structural tool in the era of structural genomics (Beauchamp & Isaacs, 1999; Hendrickson, 2000; Blundell *et al*., 2002).

The X-ray storage-phosphor image plate (IP) is considered to be the most suitable detector for acquiring protein data with a home source, as well at

synchrotron radiation facility. IP is a very sensitive detector, has wider dynamic range, has high spatial resolution and high count rate capacity, which are the fundamental requirements of an X-ray area detector (Amemiya, 1997). The IP has enabled protein crystallographer to obtain very accurate data sets with reduced X-ray dosage and exposure time. The plate can be erased by exposure to intense white radiation and can be used repeatedly. The IP has a radius of 300 mm mounted on a solid base and an INDY computer can control its movements through a controller box. Each collected image has a size of 2 MB, which can be stored in the computer. The crystal alignment was done through the CCD camera and TV monitor assembly. The processes of exposure, data collection, readout and storage of data are carried out automatically via a VME-based microprocessor system.

All the diffraction data of crystals except that of BBI have been collected at IISc, Bangalore, India. The data used for the structure of BBI were collected by Richard Lewis at the University of York, U.K. High Intensity X-ray radiation usually damages protein crystals during longer exposure time of data collection. Now it is routine to collect the macromolecular data at cryogenic temperatures. This technique of flash cooling protein crystals at liquid nitrogen temperatures and collecting data offers several benefits. Some of them are: reduces radiation damage of the crystal on exposure to X-rays, invariably improves the limit of resolution, decreases thermal parameters, allows storage and reuse of crystals, helps to overcome the scaling problem by enabling the completion of entire data collection using only one crystal (Garman & Schneider, 1997). The BBI crystal was frozen under the cryostream of liquid nitrogen (120 K) and cryo protected using 30% glycerol. Similar strategy to collect cryo-temperature data on lectin crystals in presence of 30% glycerol that diffracted up to 2.5 Å resolution was hampered due to ice formation over the crystal after 30 frames of data collection. Different cryoprotectants such as glycerol, ammonium sulfate, and low molecular weight PEGs, with varied concentration were unable to protect the crystal from the ice formation. Hence we could collect only the room temperature data up to 2.8 Å resolution using crystals of different size.

During data acquisition the crystals were oscillated about an axis perpendicular to the X-ray beam, with a chosen, relatively small angle of oscillation, usually closer to 1° per frame. Crystal to detector distance chosen, based on the longest unit cell dimension, mosaic spread, etc., so that the intensity spots are well resolved, is approximately equal to the longest crystal cell dimension. The exposure time depends on the quality of crystal and oscillation range, larger the oscillation range, longer the exposure time required. However, the situation where the intensities crossing the limit of the dynamic range of image plate has been avoided. One way to minimize such errors is to apply more than one oscillation of phi (f) per image for exposure times greater than 600 sec. In our data collection, exposure times less than 600 sec per frame only was used. The full range of rotation angle in order to acquire complete data set was chosen according to the symmetry of the crystal.

2. 3. 3. Data Processing

Indexing, processing, scaling and merging of the raw X-ray data were carried out using the programs DENZO and SCALEPACK, respectively. Denzo provides numerical analysis of each oscillation image, whereas Scalepack provides overall statistics for the whole data set. Denzo accepts peaks for autoindexing only from a single oscillation image and makes a complete search of all possible indices of all reflections. The program calculates the distortion index for all 14 Bravis lattices, user is given the choice to select the lattice and space group. The interactive mode of indexing is also an option in Denzo. Our processing procedure involved interactive mode. The peaks were selected by defining a non-overlapping box. The program Scalepack performs the scaling and merging of data from all images and does a global refinement of crystal parameters. The Scalepack method reduces the bias existing in other programs toward reflections with intensity below the average.

2. 3. 4. Mathew's number

Once the space group and unit cell dimensions of the crystal are known, it is possible to estimate the number of molecules in the crystallographic

asymmetric unit and the solvent content of the protein crystals with the knowledge of the molecular weight of protein. The following equations (2.1 and 2.2) are used (Matthews, 1968).

$$Vm = \text{Unit cell volume} / (\text{Mol.Wt.X } n\text{X } z) \quad (2.1)$$

$$Vsolv = 1 - (1.23/ Vm) \quad (2.2)$$

Where Vm is the Mathew's number, n is the number of molecules per asymmetric unit and z is the Avogadro's number; (Vsolv) is the solvent content of protein crystals.

The Matthew's number and the solvent content were calculated for all the three crystal forms of lectin. The solvent content of the BBI crystal was calculated by assuming a dimer molecule in the asymmetric unit.

2. 3. 5. Structure solution

Multi-wavelength Anomalous Dispersion (MAD), Multiple Isomorphous Replacement (MIR) and Molecular Replacement (MR) are the three methods widely used in protein crystallography to solve structures. MAD technique is currently most popular owing to the recent technical advances made in the field of synchrotron radiation, image plate and CCD detectors for data collection and molecular biology techniques that provide selenium derivative for any protein (Ealick, 2000). MIR technique requires more than one heavy atom derivative.

Molecular replacement (MR) technique is the simplest of all and can be used when a homologous protein with structural similarity is available in the database. The pioneering studies of Rossman & Blow (1962) laid the foundation of Molecular Replacement method (Rossman, 2001). Owing to the rapid expansion of Protein Data Bank (PDB; Berman *et al.*, 2000) with the increase in the number of models available, MR method is now routinely used in protein structure determinations.

The chosen model for MR can be related to the unknown structure by a rotation matrix [R(O)] and a translation vector [T(?)]. The molecular replacement

method outputs six parameters, three rotation angles (Evans, 2001) and three translation components which can be used for correctly placing the model in the unit cell of the unknown (Brunger, 1997). If X and X' represents the coordinates of the unknown structure and the model, respectively, then their relationship is given by

$$X = [R(O)]\, X' + [T(?)] \qquad (2.3)$$

The rotation matrix is effectively calculated in the Patterson space (Patterson, 1934 and 1935) using the diffraction data collected of unknown structure and calculated data of the model. If Px(r) represents the Patterson function of unknown structure and Pm(r) that of model, then for the correct superposition of the two structures, a function termed rotation function [R(O)] defined below will be maximum (Drenth, 1994).

$$R(O) = \int_u Px(r)[R]Pm(r)\, dV \qquad (2.4)$$

Where u represents the volume of integration in which the structures are compared, and [R] is 3X3 rotation matrix described using three Eulerian angles a, ß and ?, and r is the integration variable.

A translation function T(?), defined below is used for calculating the translation vector and for placing the correctly oriented model in the correct position of the unit cell of unknown structure.

$$T(?) = \int_u Px(u+?)Pm([R]u-?)\, du \qquad (2.5)$$

This function T(?) will have the maximum value (Rossman & Blow, 1964; Tollin, 1966; Crowther & Blow, 1967) when it corresponds to the correct placement of the model - oriented correctly - in the unit cell.

A set of routines implemented in an **Automated Molecular Replacement** (AMoRe) program was used for all the calculations. The input to the program is the coordinates of the model and unit cell and diffraction data of the target protein.

The important subroutines of AMoRe are outlined below.

1. SORTING- sorts, packs and assesses the quality of experimental reflection data.

2. TABLING- Structure factor calculation recurs very often during molecular replacement procedure. Tabling algorithm efficiently calculates the continuous Fourier coefficients from the model placed in the artificial cell.

3. ROTING- is used to calculate the cross rotation function using Crowther's algorithm (Crowther, 1972). The program expands the Patterson function in spherical harmonics as in Crowther's formulation, but the radial variables are treated separately, leading to more accurate results. ROTING also exploits several properties of Bessel functions and samples the radial functions depending on the resolution of the data and the interval of integration so as to enhance the resolution of the rotation peaks. The standard output contains all the peaks greater than 50% of the highest peak.

4. TRAING- calculates the translation function. Input is a whole list of peaks from the rotation function output. Four translation functions, two overlap (based on the overlap of observed and calculated Patterson intensity functions) and two correlation coefficient based, are computed under TRAING. Experience has shown that correlation coefficient based calculations overrides the other type as they can be evaluated as a Fourier summation.

5. NCS- If the asymmetric unit has more than one identical molecule then a series of N-body translation functions are computed. After each cycle, a potential solution of the previous step provides fixed contribution to the N-body translation function. This N-body translation option makes AMoRe more powerful than other molecular replacement programs.

6. FITTING- is the last step in which the potential solutions are obtained for the molecule by refining the orientational and translational parameters. The molecule is treated as a rigid body as proposed by Huber & Schneider (1985),

and the algorithm used in AMoRe was that of Castellano *et al*., (1992). The function is with respect to positional parameters, overall scale and temperature factors.

Homologous protein structures under their respective sequence based classification were available for both lectin and BBI. Hence MR method was employed for the determination of their three-dimensional structures. The reported tracey bean BBI structure (PDB entry 1PI2) (Chen *et al*., 1992) was used as search model for the structure solution of the BBI. Jacalin dimer (chains EF & GH), PDB entry 1JAC (Sankaranarayanan *et al*., 1996), was used for lectin structure solution.

2. 3. 6. Structure refinement

Refinement of a protein structure involves minimizing a function involving observed (Fo) and calculated structure factors (Fc) and along with applying restraints such as minimizing the difference between the refined parameters and ideal parameters of the stereo chemistry (Drenth, 1994). Refinement techniques generally fall into two categories, depending upon whether the calculations are performed either in the real space or in the reciprocal space. Refinement methods based on reciprocal space is preferred over real space because the former ones are computationally less expensive. To prevent the model from going into local minimum, interactive graphics was used for checking the fit of the model to the electron density.

To monitor the progress of refinement the crystallographic parameter called R factor is used. This is defined as,

$$R = \frac{\sum_{hkl} ||Fo(hkl)| - |Fc(hkl)||}{\sum_{hkl} |Fo(hkl)|} \qquad (2.6)$$

where h, k, l, are the Miller indices of Bragg reflections, and the summation is over all the reflections.

2. 3. 6. a. Rigid body refinement

Rigid body refinement is the first step in a protein structure refinement procedure that fixes the gross features of a molecule for further refinement (Head-Gordon & Brooks, 1991). The adjustment of the model consists of refining the three positional parameters and one temperature factor of all the atoms in the structure except hydrogen atoms. This procedure minimizes the R factor value by refining three rotational and three translational degrees of freedom of the subunits or groups. It is possible to regard the entire molecule as a rigid entity and refine its position and orientation in the unit cell. Each molecule or subunit is treated as a continuous mass distribution located at the center of mass position defined by,

$$R_J = 1/M_J \, ?_J \, m_l r_l \qquad (2.\ 7)$$

where $M_J = ?_J \, m_l$, m_l are atomic masses and J labels the rigid bodies.

Positional refinement was done using constraints and restraints. The ratio of the number of parameters to be determined to the number of observations is very high in a protein structure determination. This ratio could be improved in two ways, one is by reducing the number of parameters by the use of constraints, and the other by increasing the number of observations by the addition of restraints. The stereochemical information such as bond lengths, bond angles, conformational angles, planarity etc., obtained from small molecule structures, are used for applying restraints and constraints. Restraints are considered when a specific parameter has restricted freedom, limited to a range of values, whereas, the parameter is constraint, when it can assume only a specific value.

2. 3. 6. b. Refinement by maximum-likelihood method

The initial models of lectin and BBI obtained from MR calculations were refined using the program **REFMAC** (implemented in CCP4) which makes use of maximum-likelihood equations. Read (1990) and Bricogne (1991) have suggested a maximum-likelihood target that should be based on various

probability distributions. One of the expected advantages of maximum likelihood refinement is a decrease in refinement bias, as the calculated structure-factor amplitudes will not be forced to match the observed amplitudes (Read, 1997). Use of appropriate likelihood targets through the incorporation of the effect of measurement of error and the use of cross-validation data to estimate the s (sigma) values are the key ingredients in the likelihood refinement. Verification tests have shown that for refinement, maximum likelihood method is more than twice as effective compared to least-squares method, in improving the model (Pannu & Read, 1996).

REFMAC program can carry out rigid body restrained or unrestrained refinement using X-ray data (Murshudov *et al.*, 1997). The program minimizes the coordinate parameters to satisfy a maximum-likelihood or least squares residual. Before running the REFMAC, **PROTIN** program was run to invoke geometric restraints. PROTIN analyses the protein geometry and produces an output file containing restraint information. REFMAC also produces an output file with extension MTZ (named after three of its progenitors, McLaughlin, Terry and Zelinka) containing weighted coefficients for sA weighted mFo-DFc and 2mFo-DFc maps. About 5% of the reflections were kept aside during refinement to calculate Rfree for cross validation (Brunger, 1992).

NCS averaging is effectively done for an asymmetric unit composed of N similar objects related by non-crystallographic symmetry (NCS). Rossman and Blow (1963) proposed this method, by which the current phases of reflections can be improved by averaging over the electron densities of NCS related objects. NCS averaging requires an accurate estimate of NCS operators and exact information on the position and shape of the objects whose density has to be averaged (Vellieux & Read, 1997). Since both lectin and BBI structures had more than one molecule in their asymmetric units, their initial models were refined taking advantage of the presence of non-crystallographic symmetry. The NCS restraint was removed and subunits refined independently in the final cycles of refinement.

2. 3. 7. Graphics, visualization and electron density maps

After phasing the reflections obtained from X-ray diffraction, using one of the methods discussed above, an electron density map is calculated using Fourier transform. The formula for the Fourier summation to calculate an electron density map is:

$$?(xyz)= (1/V) \sum_{h}\sum_{k}\sum_{l} | F(hkl) | \cos 2p [(hx+ky+lz - a(hkl)], \qquad (2.8)$$

where x, y, z are the fractional coordinates of each point in the unit cell, F(hkl) is the structure factor, V is the unit cell volume and a(hkl) is the phase angle. The h, k, l, are Miller indices.

The difference maps such as 2Fo-Fc and Fo-Fc for electron density are used to identify errors in the model structure and to refine the positional and displacement parameters. Generally the deviations of the model from the reference molecule could be detected in the Fo-Fc difference map. The maps were contoured at 1s for 2Fo-Fc and 3.0s for Fo-Fc. Here s refers to r. m. s. deviation in the mean density in electrons/$Å^3$. Poorly defined regions of the map were examined with the maps contoured at lower levels. Difference electron density maps are important for locating bound ligands in protein structures (Glusker *et al*., 1994). This is the case for lectin structures discussed in the thesis; the location of the bound sugar molecule was clearly visible in the difference density map.

The program **QUANTA** was used for displaying and examining the electron density maps, for displaying atoms, for interactive fitting and optimizing the geometry and also for solvating the structures. The electron density maps along with the molecule were displayed on the screen of "Octane" silicon graphics workstation. The displayed map and the molecule could be rotated and viewed from any direction. Each residue starting from N-terminal to the C-terminal was examined for their optimum fitting in the electron density maps and correct geometry. The deviating ones were corrected using the **X-AUTOFIT** module. A difference Fourier (Fo-Fc) map was calculated to identify the

deviations of the protein from its search model. Both the aligned sequences of the respective families and the observed difference densities were used to ascertain the identity of dissimilar residues. After every cycle of visual fitting of the model to the calculated electron density, it was subjected to several cycles of refinement using REFMAC. After the refinement the changed residues were carefully checked in the new map. The fitted residue was retained if no difference in density (Fo-Fc) was observed and Rfactor and Rfree improved. This procedure was repeated a number of times till all the observed difference densities in the vicinity of protein atoms were accounted and both Rfactor and Rfree values converged.

The solvent molecules were placed wherever the Fo-Fc density was observed above 3s level and when the water molecules made reasonable hydrogen bonds. The module **X-SOLVATE** in QUANTA was used for adding water molecules to each of the structures through interactive mode. The addition of water molecules was started only when the refinement of the model reached an R factor below 25%. The water search was limited to distance of 6.0 Å from protein, and the distance between any two waters not less than 2.5 Å. Initially water molecules were added at a higher sigma level, later when the structures were refined, this progressively was reduced to lower levels, up to 3s.

2. 3. 8. Analysis and comparison of structures

The program **PROCHECK** was used for checking the stereochemistry and quality of the model. This program is written by Laskowski *et al.*, (1993) and forms part of the CCP4 suite of programs. Once the model has improved, after every cycle of model fitting and run of REFMAC it was subjected to stereochemical quality check. The program compares and assesses the quality of the model with the available structures of similar or better resolution than the reference structure. The output contains a comprehensive residue by residue listing of the parameters and their graphical representation. The program highlights the regions of the structure where conformations are unusual. These

are due either to interesting properties of the structure or possible errors in interpretation, and require further investigation during rebuilding step.

Hydrogen bonds were calculated using the **CONTACT** program of CCP4 suite. Hydrogen bonded contacts were considered for distances less than 3.5 ? between donor and acceptor and angle N-H….O (type) greater than 120^{o} . Similarly for O….O contacts, in addition to the distance criteria, the angle C-O….O greater than 90 was considered.

The **NACCESS** program (Hubbard & Thornton, 1993) was used to calculate the atomic accessible surface defined by rolling a probe of given size around a van der Walls surface. This program is an implementation of the method of Lee and Richards (1971).

The program **ALIGN** (Cohen, 1997) was used for comparison of the lectin and the BBI structures with the homologous protein structures.

2. 4. PURIFICATION AND CHARACTERIZATION OF BBI

2. 4. 1. Proteases and protease inhibitors assay

Proteases and protease inhibitors were assayed by Kunitz's method (Kunitz, 1947) using casein as substrate and by measuring the absorbance of the trichloroacetic acid (TCA) filtrate of reaction solution at 280 nm. 2 ml reaction mixture contained protease or a mixture of protease and PI, 0.1 M potassium phosphate buffer, pH 7.5, and 10 mg of casein diluted in appropriate proportion. This was incubated for 15 min at $37^{?}$C, and 3 ml of 5% TCA was added to terminate the reaction. The mixture was centrifuged for 10 min at 10,000 rpm and the absorbance read at 280 nm. The readings were corrected for the residual absorbances of casein, enzyme and PI.

One unit of protease is defined as the amount of protease required for the hydrolysis of 1 µmol/min of casein. One unit of protease inhibitor is defined as one protease unit inhibited. Specific activity of the PI is defined as the activity per mg of the inhibitor. Protein concentration was estimated by Lowry method using Bovine Serum Albumin (BSA) as standard (Lowry *et al.*, 1951).

2. 4. 2. Purification

The purification procedure used was based on the one reported by Rele *et al*., and Vartak *et al.,* (1980), with appropriate modifications including an additional FPLC step to obtain a homogeneous preparation.

Step1. Extraction of BBI from seeds

One kg of cowpea seeds was cleaned and used for one batch. Grossly only the seeds whose weight exceeded 200 mg were selected. The seeds were washed in tap water, rinsed and soaked in 2 lit of distilled water and allowed to swell for an hour at room temperature. The remaining steps were carried out in cold (4$^{?}$C) unless otherwise mentioned. The seeds were blended in mixer-grinder with a solution of buffer-salt-HCl comprising 1 lit 50 mM acetate buffer, pH 5.7, 50 g KCl and 1 lit 0.6 M HCl. The volume of the protein solution from this step was 3 lit and had a pH of 3.2. This solution was kept in cold for one hour and filtered through muslin cloth. The pH of the filtrate was raised to 6.7 by the addition of 180 ml potassium bicarbonate (2.0 M) and again kept in cold for one more hour. The remaining plant material was removed by centrifuging at 8,000 rpm for 15 min in batches.

Step2. Ammonium sulfate fractionation

The fractionation using ammonium sulfate precipitation has the advantage of intermediate removal of unwanted proteins, simultaneously the protein of interest will be concentrated. The fine powder of ammonium sulfate (560 g) was slowly added to 2.7 lit supernatant obtained from the previous step, so that 40% saturation is reached in about 20 min with continuous stirring. This solution was kept in cold (4$^{?}$ C) for about 4 hr for precipitation. The precipitated proteins were centrifuged for 10 min. at 8,000 rpm in batches. Again ammonium sulfate (900 g) was added to the supernatant (3 lit) in 30 min to bring the saturation to 80%. The solution was allowed to stand overnight in cold. This was centrifuged at 8,000 rpm for 15 min. The residue obtained was dissolved in about 30 ml of 10 mM phosphate buffer of pH 7.5 and dialyzed against the same buffer giving three changes in intervals of 3 hr.

Step3. Acid precipitation

After re-dissolving and dialyzing the ammonium sulfate precipitated protein, bringing down the pH of the solution resulted in precipitation without loss of PI activity. The precipitation could be due to starch and other molecules from plant tissues. The inhibitor solution at pH 7.5 was reduced to 6.3 by the addition of 0.1M acetic acid and the precipitate formed was removed by centrifugation. The pH of the supernatant was restored to 7.5 using 0.1 M NaOH. This solution was then concentrated from 320 to 150 ml using amicon YM-3 membrane of molecular weight cut off (MWCO), 3000. The amicon method was adopted for the concentration of large volumes, whereas concentration of volumes less than 2 ml was by lyophilization.

Step4. Ion-exchange chromatography using DEAE-Sephadex

Proteins, due to surface charge, bind to ion-exchangers. These reversibly adsorbed proteins can be eluted out either using a pH or a salt gradient (Jacoby, 1974; Scopes & Cantor, 1982). DEAE (diethyl amino ethyl) is a commonly used anion exchanger in protein purification system. It is prepared by attaching DEAE groups to matrices of Cellulose, Fractogel, Sephadex, Sepharose, etc,. DEAE-Sephadex was found effective for BBI purification. A glass column of size 4X24 cm was chosen. Slurry of DEAE-Sephadex was prepared in pre-equilibrating buffer and poured into the column. The material was allowed to settle and column was equilibrated with 10 mM phosphate buffer at pH 7.5. The concentrated protein from the previous step was loaded on to this column and left for an hour. The column was then washed (3 times the column volume) using the same buffer at a flow rate of 50 ml/hr. After that a buffer-salt (phosphate-NaCl) gradient of five times the column volume was run. 4 ml fractions were collected and their optical density (O. D.) at 280 nm and PI activity measured. The active fractions were pooled and concentrated to 3 ml.

Step5. Gel filtration on Sephadex G-50

The gel filtration, also called gel permeation or ‘molecular sieving’, is used for protein separation (Porath & Flodin, 1959; Scopes & Cantor, 1982).

The gel used for both protein separation and for molecular weight determination (Andrews, 1965) is the cross-linked dextran based Sephadex gels. The Sephadex G-50 filled glass column of size 2X85 cm was prepared as described in the Ion-exchange chromatography section. The column was pre-equilibrated with 50 mM phosphate buffer at pH 7.5 containing 0.1 M NaCl. The inhibitor (1.5 ml) was loaded on to the column and eluted at a flow rate of 15 ml/hr. Initial 240 ml corresponding to void volume (Vo) was discarded. Subsequent 3 ml fractions were collected and checked for protein absorbance (?=280 nm) and PI activity. The active fractions were pooled and stored at -20$^{?}$C till further use.

Step6. FPLC using MonoQ column

The MonoQ HR 5/5 (1ml) anion exchange column, with the charged group on the gel -CH2-N+$(CH_3)_3$, based on a beaded hydrophilic resin with one of the narrowest particle size distributions available was used. The column is specially designed for fast, high resolution chromatography of proteins and other biomolecules.

The BBI sample from gel filtration step was extensively dialyzed against 10 mM Tris-HCl, pH 8.5. The BBI was injected to the pre-equilibrated MonoQ column and eluted at 1.0 ml/min flow rate by applying salt (NaCl) gradient. The gradient program was standardized using trial runs (1hr run) to achieve maximum separation of isoinhibitor peaks with minimum dilution. The isoinhibitor peaks were collected separately. It was required to reload some of the adjacent peaks 2-3 times to get complete homogeneity. The column required enzymatic cleansing after every 10 cycles of usage.

2. 4. 3. Purity of the protein

Purity of the BBI preparation at different stages was checked using **gel electrophoresis, isoelectric focussing and X-ray film-contact print** techniques.

Native PAGE (Polyacrylamide gel electrophoresis) for the BBI was carried out at pH 8.9 in 12% gel and SDS-PAGE (sodium dodecyl sulfate-PAGE) in a 15% gel. The gels were run at 200 V following Lammelli's (1970)

procedure. The gels were stained using coomassie brilliant blue R-250 and de-stained using water/ methanol/ acetic acid at 5:4:1 (v/v) composition.

Isoelectric focussing (IEF) was carried out in tube gels of polyacrylamide (Vesterberg *et al.*, 1977) and in density gradient single column IEF unit (Chinnatambi *et al.*, 1995). The BBI was focussed with a broad pH gradient ranging from 3-7 in the first run and narrowed the gradient between pH 4-6 in the subsequent run on IEF unit as well on tube gels containing 7% acrylamide. The exposure was at 200 V and 4$^{?}$C for 5 hr on IEF unit and overnight on tube gels. The focussed fractions from IEF unit were collected in a multi well assay plate (2 drops/well) and diluted to 1 ml with deionized water. The pH and PI activity of the fractions were checked. The tube gels were stained using coomassie blue, R-250. The stained gels were destained using water/methanol/acetic acid (5:4:1) to observe protein bands. The gel images were stored in computer.

The X-ray film-contact print technique was used for checking the presence of isoinhibitors. After running native PAGE, the acrylamide gels were processed for PI activity by staining using X-ray film-contact print method (Pichare & Kachole, 1994). After electrophoresis the gel was equilibrated in 0.1 M Tris-HCl buffer at pH 7.8, followed by incubation in 0.1 mg/ml trypsin or chymotrypsin solution for 15 min on a shaking water bath. Later the gel was washed briefly with the same buffer and placed on a piece of undeveloped X-ray film. After 5 min the gel was removed and hydrolysis of gelatin coating on the film was monitored visually by washing the film under tap water. Inhibitor bands appeared as unhydrolyzed gelatin against the background of hydrolyzed gelatin. The other side (back) of the film was cleared with protease and developed. The bands on the negative film became translucent against an opaque background, and after contact printing showed up as black bands on the white background.

2. 4. 4. Determination of molecular weight

The molecular weight of BBI was determined using **gel filtration, SDS-PAGE and mass spectrum** techniques.

The relative molecular weight of the BBI was estimated using Sephadex G-100 column (1.2X100 cm) according to the method of Andrews (1965). The column was equilibrated with 50 mM phosphate buffer, pH 7.5, containing 0.2 M NaCl. The void volume (Vo) was estimated by running blue dextran (Mr. 200 kDa). The elution volumes (Ve) were estimated by running standard molecular weight markers in succession followed by the BBI.

In SDS-PAGE electrophoresis, a linear plot over a wide range of molecular weights can be obtained if mobility (R_f) is plotted against logarithm of molecular weight (log Mr.) (Weber & Osborn, 1969). The apparent molecular mass of the BBI subunit was determined using the method of Lamelli (1970), The acrylamide gel (15%, w/v) was run at 200 V for 4 hr. Protein bands were visualized by staining using coomassie brilliant blue, R-250.

In mass spectrometry the molecular weight of individual molecules were determined by transforming them into ions *in vacuo* and then measuring the response of their trajectories in electric or magnetic fields or both (Fenn *et al*., 1989). Electrospray ionization has recently emerged as a powerful technique for producing intact ions *in vacuo* from large complex biological molecules (Mann *et al*., 2001). The electrospray mass spectrometric (ES-MS) method has been used to determine accurately the mass of the isoinhibitors. The isoinhibitors were extensively dialyzed against deionized water and 5 µl containing 10 µg sample were injected into the ES chamber through a stainless needle. The needle is maintained at a few Kilo-Volts (KV) relative to the walls of the chamber. The resulting field at the needle tip charges the surface of the emerging liquid sample and disperses it into a fine spray of charged droplets. The spectrum was scanned to cover a wide range of m/z (mass to charge ratio) in order to detect all the ions.

2. 4. 5. Amino acid (N-terminal) sequence determination

One of the major isoinhibitors (labeled as P-IV) in our preparation was found to be comparatively free from other isoinhibitors -based on FPLC gradient profile and Isoelectric focussing on acrylamide gel- was used for complete amino acid sequencing. The single crystals were also grown from this isoinhibitor. 5 mg of the isoinhibitor was extensively dialyzed against deionized water and concentrated. The sample (approx. 500 pmol/run) was passed through a C18 RP HPLC column using an acetonitrile/ water solvent system with 0.1% trifluoroacetic acid (TFA) as the ion-pairing agent. The peak obtained was concentrated using a Savant speedvac system and the concentrated protein peak was loaded onto an automated gas phase protein sequencer (Shimadzu, PSQ-1) and N-terminal sequence determined.

2. 4. 6. Preparation of BBI-protease complex

The BBI and bovine trypsin were allowed to react at 37°C for 30 min at 1:1 molar ratio. The reacted mixture was then run on a Sephadex G-100 column (same column used for molecular weight determination of the BBI) to separate the complex from mixture. The column was equilibrated and eluted with 50 mM phosphate buffer at pH 7.5 containing 0.2 M NaCl. Fractions of 2 ml were collected and the peaks identified by measuring O. D. at 280 nm. The peaks were checked for both protease and PI activity. The molecular weight of BBI-protease complex was determined using the same gel filtration column. Similar procedure was followed in the preparation of BBI-chymotrypsin complex. The complexes were screened for various crystallization conditions.

CHAPTER-3

THE STUDY OF THE CRYSTAL STRUCTURE AND CARBOHYDRATE BINDING OF A LECTIN FROM *ARTOCARPUS HIRSUTA* SEEDS.

3. 1. INTRODUCTION

The genus *Artocarpus* constitutes a group of plants whose fruits and timber are useful, from which the first jacalin family lectin isolated (under the species *Artocarpus integrifolia*; jack fruit) (Kumar *et al.*, 1982) and characterized (Kabir & Daar, 1994) and has been named 'jacalin'. Subsequently, many lectins have been isolated from the seeds belonging to various *Artocarpus* species (Young *et al.*, 1989). Jacalin, the representative member of the jacalin family lectins, was the first to be sequenced, crystallized and three-dimensional structure determined in complex with methyl-α-galactose (Sankaranarayanan *et al.*, 1996). Subsequently a few more jacalin related lectins have been crystallized and three-dimensional crystal structures for four of them are available. These include, *Maclura pomifera* Agglutinin (MPA) complexed with T-antigen disaccharide (Lee *et al.*, 1998), *Helianthus tuberoses* lectin (heltuba) in native and in complex with Manα1-3Man and Man-α1-2Man (Bourne *et al.*, 1999) and artocarpin from jack fruit in complex with methyl-α-D-Man (Pratap *et al.*, 2002). Jacalin and MPA belong to galactose specific group, whereas heltuba and artocarpin to mannose specific group. Jacalin and MPA are two chain lectins, containing a 133 residue heavy α-chain and a 20 residue small β-chain, whereas heltuba and artocarpin are single chain lectins. Jacalin and MPA show exclusive specificity for tumor associated T-antigen disaccharide, whereas heltuba and artocarpin do not. However, all the four lectins share similar monomer fold and carbohydrate binding site topology. In both lectins, the functional unit is a homo-tetramer. According to a recent report, the galactose specific and mannose specific jacalin lectins were found to be located in different subcellular compartments (Peumans *et al.*, 2000). Galactose specific lectins are cytoplasmic proteins, whereas mannose specific ones are located in storage vacuoles.

Crystallization and preliminary X-ray studies are reported for mannose specific jacalin lectins, KM+ lectin (Oliviera *et al.*, 1997), calsepa (Wright *et al.*, 1997) and Morniga M lectin (Rabjins *et al.*, 2001) from *Morus nigra* (mulberry). In an endeavor of knowledge based structure prediction, the three-dimensional

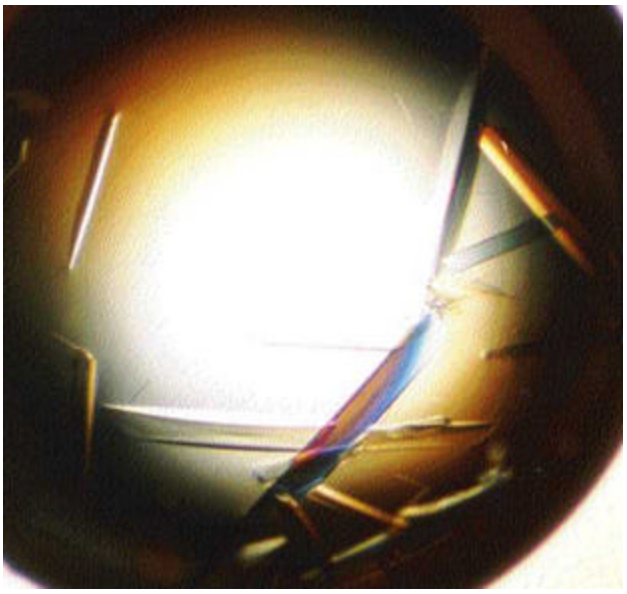
structure of KM+ lectin was predicted using the coordinates of jacalin. However, this model failed to predict unambiguously the structural determinants involved in sugar binding (Rosa *et al.*, 1999).

The lectin from *Artocarpus hirsuta* seeds belongs to the family Moraceae. The lectin was isolated, purified and characterized in terms of its sugar specificity and sugar binding site (Gurjar *et al.*, 1998). Also the lectin was analysed for its detailed binding affinity towards galactose and its derivatives using fluorescence spectroscopy (Gaikwad *et al.*, 1998). The *hirsuta* lectin was also studied for its insecticidal activity against the larvae of red flour beetle (*Tribolium castaneum*) indicating its role in plant defense (Gurjar *et al.*, 2000). The report of this lectin subjected to both chemical and thermal denaturation studies has appeared recently (Gaikwad *et al.*, 2002). This basic lectin belonged to the galactose specific group of jacalin family lectins. The purified *hirsuta* lectin in homogeneous form has been obtained through collaboration.

3. 2. RESULTS AND DISCUSSION

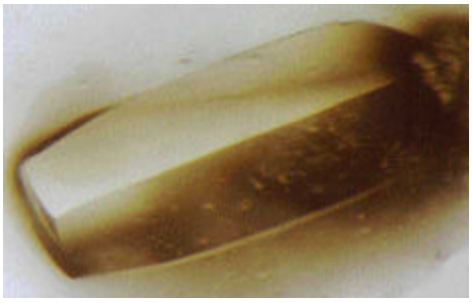
3. 2. 1. Characterization of lectin crystals

The pure lectin was concentrated to 60 mg/ml by lyophilization and mixed with 0.5 M methyl-α-D-galactose (me-α-gal) in 10:1 proportion prior to setting up crystallization. Crystals were grown from solutions of 0.2 M phosphate buffer in the pH range 6.0-7.0 at room temperature using 35-40% saturated ammonium sulfate. Two orthorhombic crystal forms with little difference in morphology used to grow under these conditions (Figs. 3. 1. A). Hexagonal crystals were also obtained in identical conditions (Figs. 3. 1. B). Crystals could be grown in the absence of sugar molecule as well under the same conditions but in these experiments 10% glycerol was added to the well solution. Large size crystals (Fig. 3. 1. C) used to grow in a week's time. They belonged to another hexagonal form but were of poor quality (Rao *et al.*, 1999a).



A

B



C

Fig. 3. 1., A) orthorhombic crystals of *hirsuta*, typical size 0.6 X 0.1 X 0.1 mm, B) hexagonal form-I, crystals of typical size 0.5 X 0.25 X 0.25 mm, C) big size hexagonal form-II, crystals grown in the absence of sugar but with 10% glycerol.

Orthorhombic form-I

Crystals selected for data collection had maximum dimension of 0.5 mm. Crystal to detector distance chosen was 170 mm and the frames were collected at $1^{?}$ oscillations. Indexing of the oscillation frames gave orthorhombic cell of unit cell dimensions a=92.9, b=99.8, c=166.2 Å. 100 frames used in data processing using DENZO provided reflection data set complete by 99.15% up to 2.8 Å resolution. Rmerge was estimated 0.143. From the systematic absences of axial reflections the space group was determined as $P2_12_12_1$.

Orthorhombic from -II

The crystals were of similar size as orthorhombic form-I and can be distinguished only by unit cell parameters. Data (130 frames) were collected by keeping the crystal to detector distance at 170 mm and oscillations $1^{?}$ per frame. Reflections were indexed and cell dimensions estimated as a=89.9, b=121.9, c=131.6 Å. Space group inferred from systematic absences was $P2_12_12_1$. The processed data was 98.9% complete up to 3.0 Å with a Rmerge of 0.096%.

Hexagonal form-I

These crystals exhibited hexagonal bipyramidal morphology with largest dimension of 1 mm. Data were collected keeping crystal to detector distance of 300 mm and giving $0.5^{?}$ oscillation. More than 100 frames were collected. Indexing gave a unit cell of dimensions a=b=84.1 and c=271.7 Å. Data processed from 100 frames were complete to 94.3% at 3.4 Å resolution and the Rmerge for equivalent reflections was 0.059. The processing was done in space group P622.

Hexagonal form-II

Big crystals of hexagonal cylindrical morphology were obtained in this form. The crystals had typical size of 1.5 mm along the cylindrical axis. Initial data collection trials showed that one of the cell dimensions was very large. In order for the reflection spots to be resolved along the longest axis a large value

for the crystal to detector distance was chosen. 22 frames were collected at crystal to detector of 380 mm and 0.3$^{?}$ oscillation for each frame. Indexing of these frames gave a cell with dimensions a=b=111.46 and c=648.75 Å. This hexagonal unit cell parameters were in confirmity with the clear hexagonal morphology of crystals. Due to the large c-axis further data collection could not be carried out with the available set up.

3. 2. 2. Solvent content of crystals

The Mathews number (Vm) (Matthews, 1968), was calculated assuming two tetramers in the asymmetric unit, in all the three crystal forms. The molecular weight of the lectin (tetramer) determined by electrophoresis and gel filtration methods was 55 kDa (Gurjar *et al*., 1998). The estimated Vm turned out to be 3.25, 3.00 and 2.31 Å^3/D which corresponded to 60%, 59% and 47% solvent content in the crystals of two orthorhombic and one of the hexagonal forms, respectively. However, the models obtained from molecular replacement calculations gave only one tetrameric molecule in the asymmetric unit of the unit cell of both the orthorhombic forms. As a result, the solvent content of these two crystal forms turned out to be as high as 80% and 79%, respectively. In the case of hexagonal form the model placed in the unit cell and displayed in QUANTA confirmed that the tetramer resulted from the crystallographic 2-fold symmetry. No short contacts were detected. However, since the asymmetric unit has only one dimer, the calculated solvent content turned out to be 73%. The *hirsuta* crystals showed unusually high solvent content, outside the range compiled by Mathews (1968), and rarely observed in protein crystals. There are very few reports of proteins with such high solvent content (Lindqvist & Barnden, 1980; Mattevi *et al*., 1992).

3. 2. 3. Structure solution by MR method

For all the three crystal forms the search model used was the dimer (chains EF & GH) of jacalin. The data in the resolution range, 20.0-2.8 Å for orthorhombic form-I and 20.0-3.0 Å for orthorhombic form-II, 15.0-3.4 Å for hexagonal, were used for MR calculations. The solutions with highest

correlation coefficients (Cc) in the rotation function calculation turned out to be the correct ones in all three crystal forms (table. 3. 1. a). Both the orthorhombic forms had two unique solutions in the rotation function calculation (O1S1, O1S2 and O2S1, O2S2). In the hexagonal form only one solution with highest Cc for the rotation function was obtained (HS1). The translation function calculated (table. 3. 1. b), using the solutions of rotation function, placed the molecule in correct positions in their respective unit cells. The application of rigid body fit in AMoRe program improved the solutions in all the cases (table. 3. 1. c).

The processed data for the hexagonal cell did not have the necessary axial reflections for determining the correct space group. The correct space group including the chirality of the screw axis were confirmed through the translation function calculations (table. 3. 2). The best Cc and Rfactor were obtained only for calculations in 6_122 space group.

Table. 3. 1. a. Rotation function solution for the three crystal forms. Cc is the correlation coefficient, a, ß, ? are the Eulerian angles.

Crystal form	**Solution**	**a(?)**	**ß(?)**	**?(?)**	**Cc(%)**
Orthorhombic-I	O1S1	49.83	76.92	255.43	20.1
	O1S2	125.17	78.39	285.35	17.4
Orthorhombic-II	O2S1	168.60	56.15	266.78	17.5
	O2S2	23.44	56.51	266.66	13.5
Hexagonal-I	HS1	34.22	84.86	322.51	12.3

Table. 3. 1. b. Translation function (Tx, Ty, Tz) solution for the three crystal forms. Cc and R (Rfactor) shown against O1S2 and O2S2 are for the two dimers together in each case.

Solution	Tx	Ty	Tz	Cc(%)	R(%)
O1S1	0.3993	0.2179	0.4156	49.2	46.6
O1S2	0.0237	0.2826	0.7882	79.6	29.9
O2S1	0.1204	0.4811	0.3471	43.2	50.1
O2S2	0.8418	0.0701	0.1475	69.0	37.7
HS1	0.2122	0.7778	0.0420	62.5	38.1

Table. 3. 1. c. Results of the rigid body fit carried out using program in AMoRe package.

Solution	a(?)	ß(?)	?(?)	Tx	Ty	Tz	Cc(%)	R(%)
O1S1	49.35	77.03	255.74	0.3993	0.2182	0.4156	-	-
O1S2	125.78	78.24	285.29	0.0238	0.2822	0.7880	81.5	29.0
O2S1	168.50	55.97	266.57	0.1204	0.4809	0.3470	-	-
O2S2	23.52	55.89	266.45	0.8415	0.0698	0.4180	71.2	37.3
HS1	33.06	84.64	322.22	0.2124	0.7780	0.0419	68.6	35.4

Table. 3. 2. Results obtained from translation function calculation in different hexagonal space groups in order to determine the correct one. The first one in bold with the highest Cc and lowest R is identified as the correct space group.

Space group	Cc(%)	R(%)
P $6_1$22	**62.5**	**38.1**
P $6_5$22	28.8	50.9
P $6_4$22	18.0	54.7
P $6_3$22	36.3	48.8
P $6_2$22	25.8	52.5
P 622	19.1	55.0

3. 2. 4. Refinement of the structure

The refinement was carried out only in the case of orthorhombic forms. The initial atomic coordinates for the molecule were obtained by operating on the jacalin model the rotation matrices and translational vectors. When carrying out the refinement the reflection data set was divided into two parts; a working set and a test set for cross validation tests. 1854 reflections were set aside for Rfree calculation in case of form-I and 1276 for form-II. The form-I was first taken up for refinement since it diffracted to a better resolution than form-II. The initial phases obtained from molecular replacement using the search model were improved by subsequent rigid body refinement retaining jacalin sequence. This was followed by several cycles of positional refinement using the data in the resolution shell of 9.0-2.8 ?, resulting in the improvement of both Rfactor and Rfree. Individual B-factors were refined individually in the final cycles. Electron density maps were calculated at this stage. Side chain conformations were progressively changed, fitted into the observed density and their positions refined. The non-crystallographic symmetry (NCS) present in the structures was used to improve the map quality. The asymmetric unit has four identical subunits. The 8 chains belonging to the four subunits were labeled differently.

The output of REFMAC refinement contained, 1) the transformation matrices and the relative rotation angles f , ?, ?, for each NCS related pair of chains. 2) the number of positions and the number of atoms which differ by more than 2s (sigma) from the average position. Similar procedure was followed in the refinement of orthorhombic form-II using the data between 9.0-3.0 Å resolution.

The module X-SOLVATE of QUANTA was used interactively to add solvent molecules. The acceptance of solvent molecules was decided carefully as mentioned in chapter 2. After a few cycles of refinement those water molecules which tend to reach the limiting value 100, chosen for B-factor, were deleted from the coordinate file. In the regions where the map was not well defined or the residue was not fitting in the density, fresh maps were calculated by omitting the residues and then the correct residues were chosen based on the shape of the calculated difference density.

As the refinement progressed, difference Fourier maps (Fo-Fc) clearly revealed the position of sugar bound to N-terminal of a-chain of all the four subunits. In addition, a residual electron density was observed in the difference map, at the lectin face opposite to the side where the sugar binding site was located. Some main chain polar atoms and side chain hydrophobic groups surround this density (Fig. 3. 2.) However, this electron density was observed in form I and not in the corresponding position of form II. Lectins, especially legume lectins, in addition to their carbohydrate binding site posses a hydrophobic site that binds nonpolar compounds such as adenine and indole acetic acid (Gegg *et al.*, 1992; Roberts & Goldstein, 1983). However, this is not yet observed in the known structures of jacalin family lectins. The binding of hydrophobic compounds by *hirsuta* lectin is under investigation.

After several cycles of refinement of the lectin structures in the two crystal forms, the Cc and Rfactor values improved, and remained with insignificant changes in the final cycles (table. 3. 3). In the final refined model, form I had 4548 protein atoms, 52 sugar atoms and 763 solvent atoms, whereas form II had 4560 protein atoms, 52 sugar atoms and 534 solvent atoms.

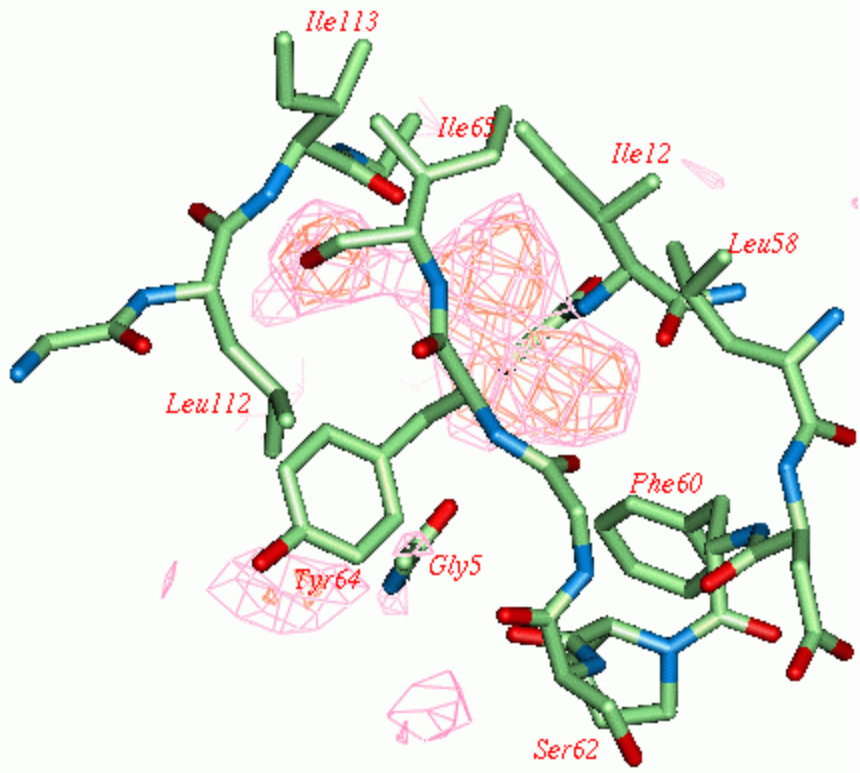


Fig. 3. 2. The diagram showing the difference electron density (Fo-Fc, 3s level) observed at the face opposite to the sugar binding site of *hirsuta*. The hydrophobic residues surrounding the density are shown.

Although there is close sequence and structural similarity between *hirsuta* lectin and jacalin, during the final stages of refinement the difference in residues - in terms of side chain and conformation - were apparent. Such residues were eventually replaced and fitted according to the observed electron density and based on information from multiple alignment of jacalin family lectins (Fig. 3.4). The changed residues are V31T, K45T, T49S, M66T, V79T, L124M, all belong to a-chain; S14P belong to ß-chain. Residues Pro61 and Pro91 of a-chain and Pro14 of ß-chain have cis-conformation. Assuming similar length for *hirsuta* and jacalin lectin genes, electron density corresponding to three N-terminal and three C-terminal residues of the ß-chain could not be found. The density for the side chain of Val19 of ß-chain was not defined in three subunits of form I and in one subunit of form II and hence was treated as

alanine in only these subunits. All the 133 residues of a-chain could be traced in the electron density map. No electron density was observed for any covalently linked carbohydrate.

During final cycles of refinement the stereochemistry and the geometry of the models were checked using PROCHECK. The Ramachandran (f , ?) plot (Ramachandran *et al*, 1968) showed that the residues were placed in the most favored region or allowed region of the map (Figs. 3. 3. A-B). The statistics of the refinement is given in the table. 3. 3.

Table. 3. 3. Refinement statistics of orthorhombic forms.

Parameter	**Form-I**	**Form-II**
Resolution range of the data used for the refinement in Å	9.0-2.8	9.0-3.0
Total number of reflections used in the refinement	35,073	23,747
Number of reflections set aside for Rfree calculation	1854	1276
Correlation coefficients, Cc & Cc free (%)	95.78 & 90.62	95.46 & 87.48
Rfactors, R & Rfree (%)	14.43 & 21.47	15.30 & 23.30
Residues in the Ramachandran plot		
Residues in the; a) most favored region (%)	87.1	84.7
b) allowed region (%)	12.9	15.3

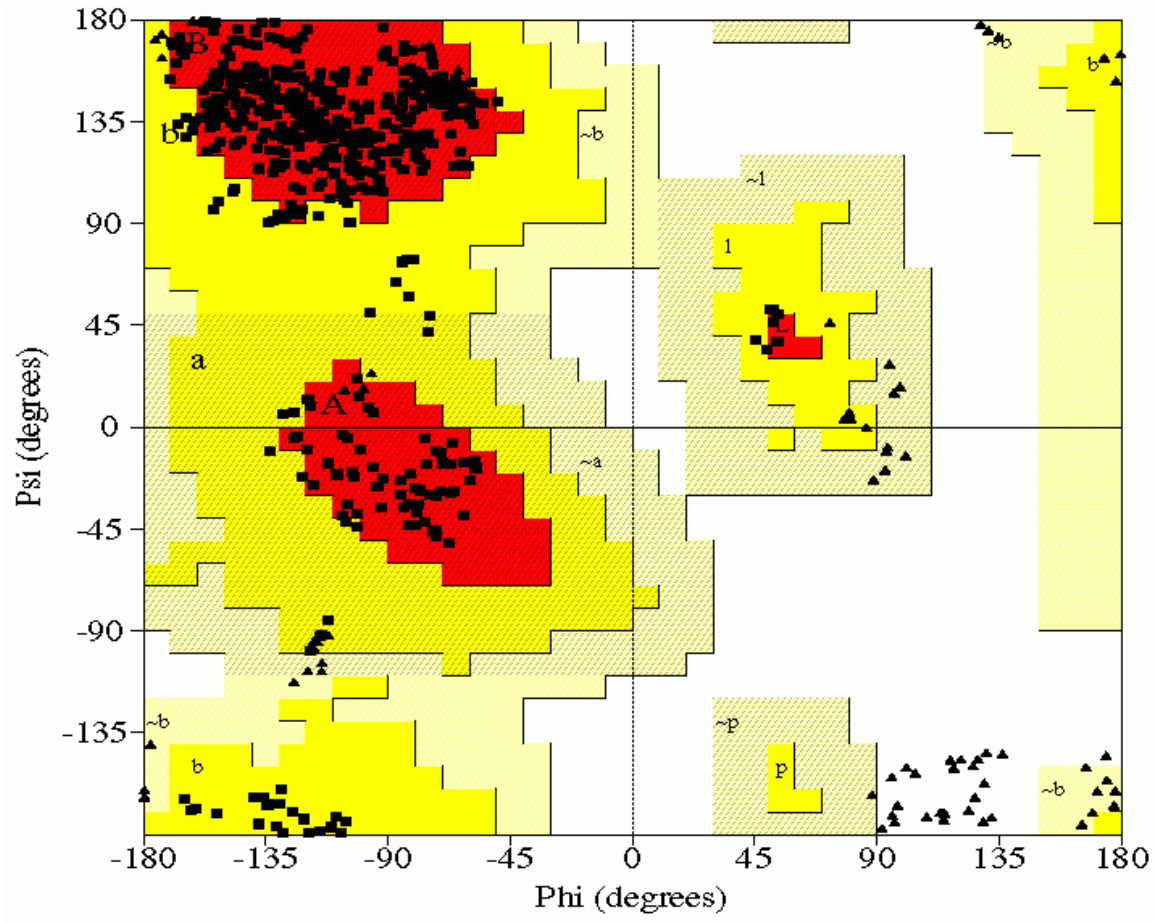


Fig: 3. 3. A. Ramachandran plot corresponding to orthorhombic form I generated using PROCHECK. Triangles and squares represent glycine and non-glycine residues respectively.

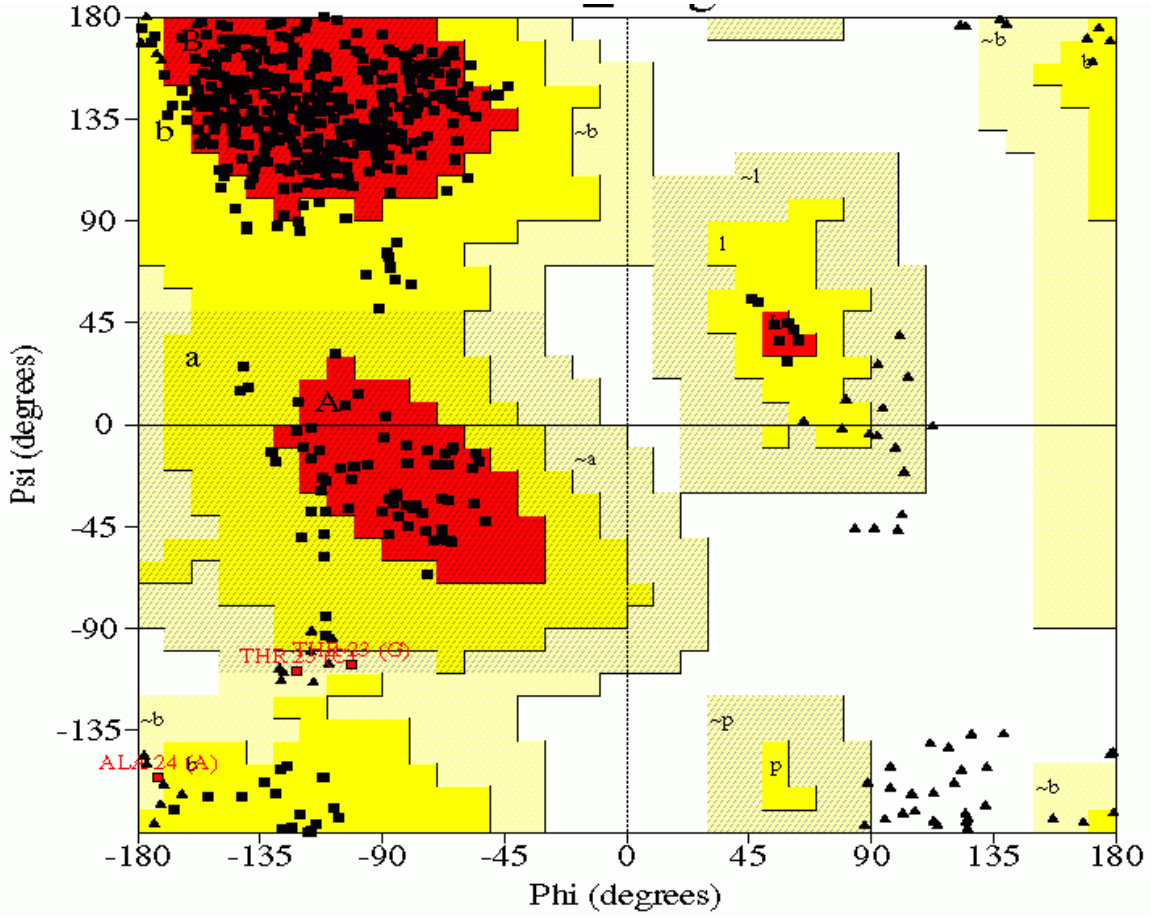


Fig. 3. 3. B. Ramachandran plot corresponding to orthorhombic form II. The conventions are similar to Fig. 3. 3. A.

3. 2. 5. Primary Structure

The linear amino acid sequence of *hirsuta* lectin derived from the electron density map was aligned against the sequences of related lectins jacalin, MPA, heltuba, artocarpin and KM+ (Fig. 3. 4), whose three-dimensional structures have been reported. The glycine content of these proteins is slightly more than usual, and it is distributed throughout the sequence rather than forming a collagen like or any other unusual structure. The sequence similarity among these lectins varied between 12 to 91%, with maximum similarity exists between *hirsuta* lectin and jacalin and minimum between artocarpin and heltuba (table. 3. 4). The lectins of mannose specific group heltuba, artocarpin and KM+ revealed poor sequence similarity (12-68%) among themselves, whereas lectins of galactose specific group *hirsuta*, jacalin and MPA exhibited very high sequence similarity (75-91%). In total, 28 residues out of the total, are invariant among these lectins (may be a requirement for the integrity of ß-prism fold) and contain mainly glycine and hydrophobic residues that play a key role in the packing of the three Greek key motifs. Further, mannose binding lectins are devoid of any aromatic residues in their sugar binding region (Fig. 3. 14. B), whereas galactose specific ones have aromatic residues, Tyr78, Tyr122 and Trp123 at their sugar binding site (Figs. 3. 13. A-B and 3. 14. A). Interestingly, the sequence alignment of these 6 proteins shows that none of the above 4 aromatic residues are found in the corresponding positions of the sequences of mannose specific lectins and hence near the carbohydrate binding site of mannose specific lectin, heltuba (Fig. 3. 14. B). However, it is apparent from the sequence analysis that the jacalin family lectins have evolved from a common ancestor. The residue numbering referred above corresponds to the crystal structure of *hirsuta* and to the sequence of *hirsuta* in the aligned sequences below.

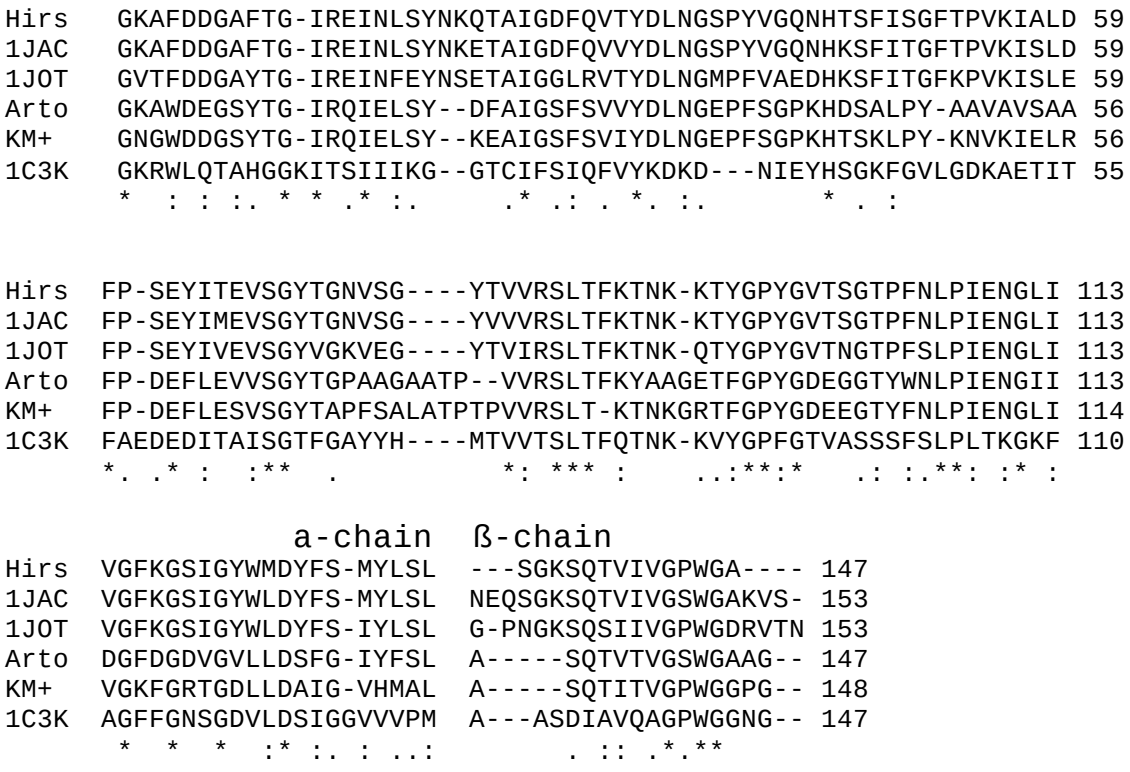


Fig. 3. 4. Multiple sequence alignment of *hirsuta* with jacalin-related lectins using CLUSTAL W. The sequences are named as Hirs- *hirsuta* lectin , 1JAC-jacalin, 1Jot-MPA, Arto-artocarpin, KM+- KM+ lectin, 1C3K-heltuba lectin. The symbols *, : and . point to the conserved residues, substitution by a similar type of amino acid and substitution by a non-similar type of amino acid residues, respectively. In the case of galactose specific lectins (*hirsuta*, jacalin and MPA) there is a break in the polypeptide chain due to post-translational modification, hence they are two chain (a and ß-chain) lectins. Members of mannose specific group (heltuba, artocarpin and KM+) are single chain lectins.

Table. 3. 4. Pair wise sequence alignment score (in %) of *hirsuta* with other jacalin-related lectins.

Lectins	1JAC	1JOT	Arto	KM+	1C3K
Hirs	93	77	24	53	52
1JAC		79	22	56	51
1JOT			22	48	48
Arto				12	22
KM+					69

3. 2. 6. Secondary Structure

The secondary structure of the protein is inferred from the Ramachandran angles and the intra-molecular hydrogen bonding of peptide backbone (Ramachandran *et al*., 1968). The secondary structure of *hirsuta* lectin contains three four stranded anti-parallel ß-sheets (Figs. 3. 5 and 3. 7. A). There are many main chain-main chain hydrogen bonds in each of the eight subunits considered, mostly involved in ß-sheet formation. Listed in table. 3. 5 are the number of the various types of observed hydrogen bonds in the two crystal structures. The main chain hydrogen bonds between the strands of the three sets of ß-sheets are shown in a schematic representation in Fig. 3. 5. There were very few inter ß-sheet hydrogen bonds.

70 % of the protein residues are involved in the ß-sheet formation. The residues in the strands involved in making these sheets are: the residues Lys2-Phe9, Leu112-Gly121, Trp123-Ser132, of a-chain and Val10-Val17 of ß-chain form the first ß-sheet; the residues Thr10-Asn20, Thr23-Asn35, Gly36-Thr45, Phe51-Leu58 of a-chain form second sheet; the residues Tyr64-Asn74, Gly77-Asn89, Lys90-Thr99, Thr102-Ile108 of a-chain form third ß-sheet (Fig. 3. 5). The second & third ß-sheets form Greek key (GK) motifs with topology 1,1,-3 (Richardson, 1981). The first sheet also forms a GK motif but the strands

connectivity is different from the other two and hence the topology of this Greek key differs. The two outer strands come from both the N-terminus of a and ß-chains whereas the two inner strands are from the C-terminus of the a-chain. The first Greek key contains 39 residues, Gly1-Phe9, and Leu112-Leu133 of a-chain and Val10-Ala17 of ß-chain. Greek key 2 contains 52 residues, Thr10-Pro61 of a-chain. Greek key 3 contains 46 residues, Glu63-Ile108 again of a-chain (Fig. 3.5). The three GK motifs and their superposition are shown in the Figs. 3 .6. A-D.

Table. 3. 5. The number of intra-subunit hydrogen bonds classified according to their types: main chain-main chain (Mc-Mc), main chain- side chain (Mc-Sc), side chain-side chain (Sc-Sc) and protein-water (Prot-Wat), observed for each subunit in the two orthorhombic forms of *hirsuta*.

Interaction **Subunits**	**Mc-Mc**	**Mc-Sc**	**Sc-Sc**	**Prot-Wat**
Form I				
A1	92	38	17	130
B1	87	36	19	125
C1	89	40	18	138
D1	96	41	15	132
Form II				
A2	92	36	17	108
B2	91	28	17	100
C2	92	38	15	102
D2	92	32	16	105

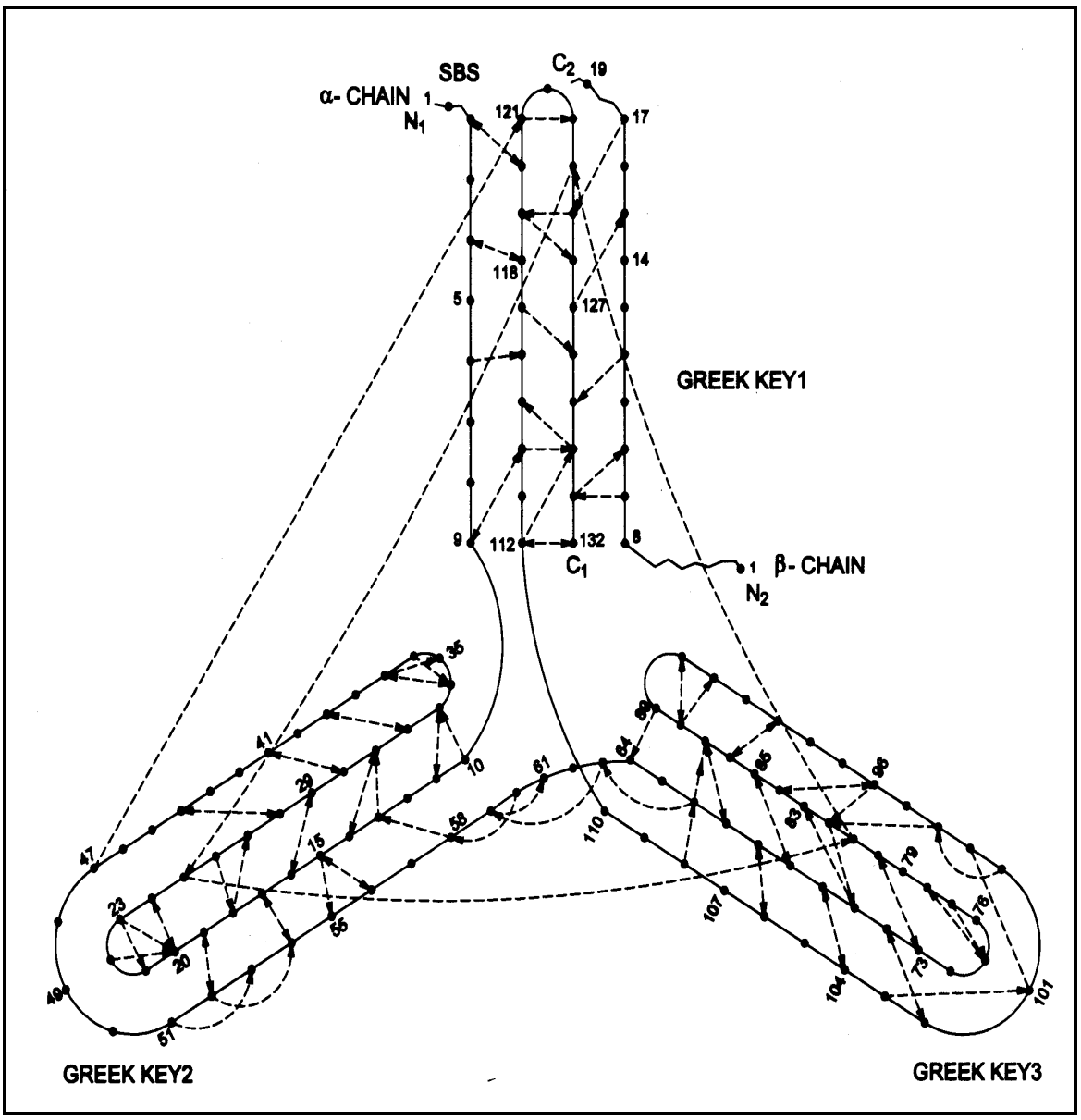
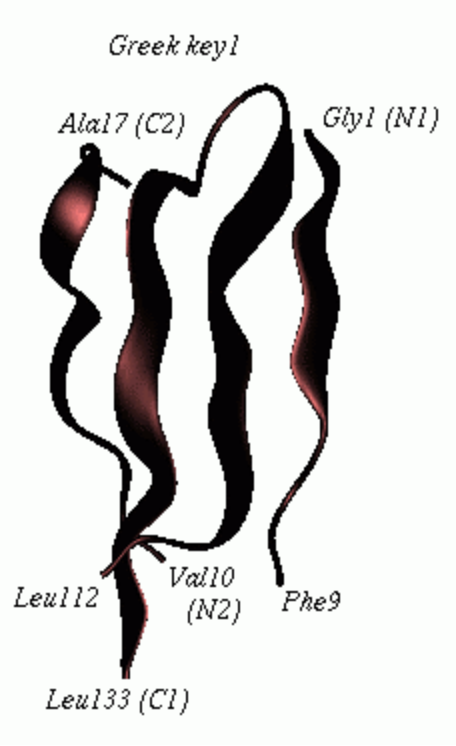
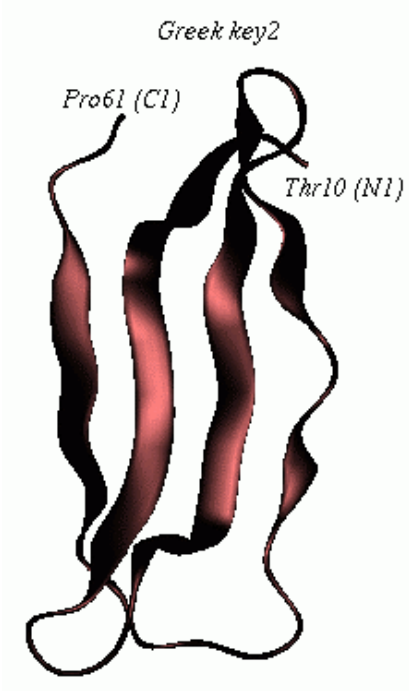


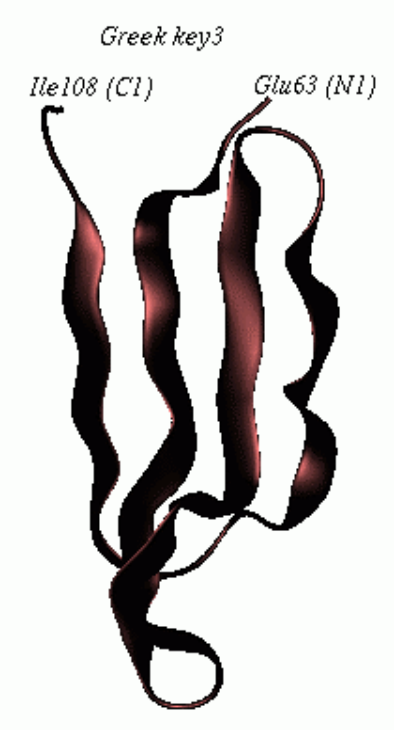
Fig. 3. 5. Schematic diagram showing the hydrogen bonds between main chain atoms of three ß-sheets and the organization of the three Greek key motifs of subunit of *hirsuta*. N1, N2 & C1, C2 are the N & C termini of a & ß-chain respectively, SBS indicates sugar binding site. The numbering in the diagram corresponds to the residue numbering in the crystal structure.



A



B



C

D

Fig. 3. 6. Schematic representation of the three Greek key motifs, A) Greek key1, B) Greek key2, C) Greek key3, D) Superposition of the three Greek keys. The terminal residues are labeled.

3. 2. 7. Tertiary Structure

The three ß-sheets of *hirsuta* lectin are arranged like the faces of a triangular prism (Figs. 3. 5 and 3. 7. A-B). The overall fold of *hirsuta* lectin is very similar to that of jacalin and other jacalin-related lectins and the fold is classified as 'ß prism-I fold' (Chothia & Murzin, 1993). This fold was observed for the first time among lectins in jacalin (Sankaranarayanan *et al*., 1996) and is named as 'jacalin fold' under the classification of lectin folds.

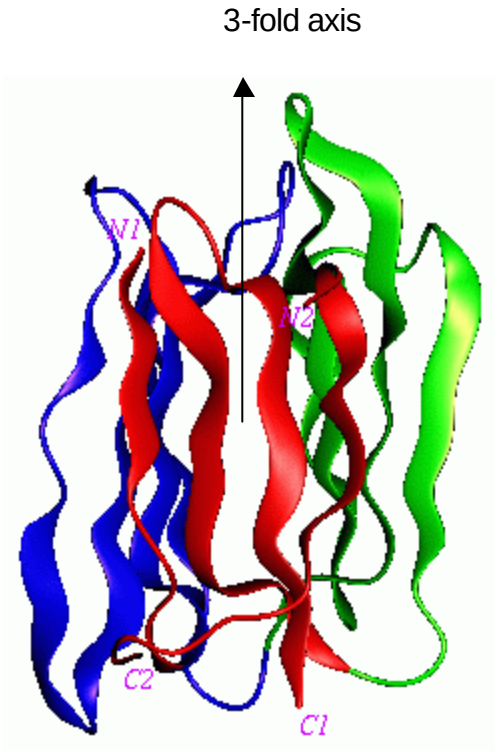
The structures of subunits in both the crystal forms superpose well among themselves. The r. m. s. deviation in Ca positions of the subunits on superposition is given in the table. 3. 6. From the table it can be found that the subunits of form I superpose among themselves better than those of form II. Inter-crystal form subunits also superpose well. The subunit A1 of form I was selected for further analysis. The results of the analysis hold true for all other subunits of both the crystal forms. The core of the ß-prism is occupied partly by hydrophobic residues and partly by polar residues that form hydrogen bonds, and interactions among these residues stabilize the three Greek key motifs with in the subunit. Accessibility of each residue and its relative accessibility, defined as percentage of its accessibility compared to the accessibility of that residue type in an extended Ala-X-Ala tripeptide (Hubbard *et al*., 1991), were computed for the subunit using NACCESS program. 33 residues were found to have relative accessibility values less than 5% and constitute the hydrophobic core of the molecule. They are, Ile15, Leu17, Ser18, Ala24, Ile25, Gly26, Phe28, Val30, Tyr32, Asp33, Ile56, Glu63, Ile65, Val68, Gly70, Val80, Val81, Ser83, Leu84, Phe86, Thr88, Ile113, Val114, Gly115, Phe116, Gly118, Ser119, Ile120, Gly121, Met124, Phe127, Ser128, Met129. Expectedly majority of the residues are hydrophobic in character.

The interaction between the a and ß-chains are through eight inter-strand hydrogen bonds between residues 9 to 17 of the ß-chain and residues 125 to 131 in the C-terminal segment of the a-chain (Fig. 3. 5). Together, these strands make up half of the first Greek key subdomain. The N-terminal of a-chain is created by the post-translational cleavage that removes four linking

residues between a and ß-chains. This processing creates a key ligand-binding residue, the amino group (NH_3^+) of the N-terminal glycine of a-chain.

Table. 3. 6. The r. m. s. deviation in Ca positions upon pair wise superposition of all possible combinations of subunits of the two orthorhombic forms. A1-D1 represents subunits of form I, and A2-D2 subunits of form II.

Subunit	B1	C1	D1	A2	B2	C2	D2
A1	0.151	0.170	0.160	0.273	0.290	0.261	0.290
B1		0.192	0.180	0.294	0.318	0.282	0.316
C1			0.152	0.271	0.291	0.277	0.284
D1				0.275	0.275	0.266	0.274
A2					0.260	0.245	0.238
B2						0.245	0.283
C2							0.220



A

B

Fig. 3. 7. The side and top views of *hirsuta* subunit, A) parallel to the 3-fold symmetry axis direction B) down the 3-fold axis.

3. 2. 8. The 3-fold symmetry of the subunit structure

The *hirsuta* monomer contains three identical subdomains related by 3-fold symmetry and form faces of a prism. The symmetry axis passes through the center of the prism. The three Greek key motifs superposed are shown in Fig. 3. 6. D. The subunit viewed from the side and from the top, indicating the approximate three-fold axis is shown in Figs. 3. 7. A-B.

Just like in the case of jacalin, the internal 3-fold symmetry is not reflected in the amino acid sequence of the Greek key motifs. The manual sequence alignment of the residues belonging to the three Greek keys is shown in Fig. 3. 8. It is clear from the sequence alignment that no significant sequence identity exists among the three motifs. The only conserved residue (Ile, shown in bold in the aligned sequences below) among these motifs is at positions a113 (GK1), a12 (GK2) and a65 (GK3). The 3-fold pseudo symmetry of the subunit structure seen in *hirsuta* and other jacalin lectins is also found in monocot mannose-binding (bulb) lectins, type2 RIPs and amaranthins (Wright, 1997). However, the 3-fold axis runs perpendicular to the three ß-sheets in bulb lectins, whereas it is at 56$^{?}$ to the barrel axis in type2 RIPs and amaranthins.

```
        a 113                                  a133, ß10          ß17,a1              a9
GK1:   L I V G FKG S  I G Y WMDY F SMYLS L, V I V  G   PWGA,GKAF DDGAF
GK2: TG I R E I N L S Y N K E T A I G DF QVTY  D LN GS PY VG QNHT S F I SGFTPVKIS LDFP
GK3: EY I T E VS G Y T G N V S GYT VV R SLT  K FT  NKKT YG PY GVT SGT PFNLPI
```

Fig. 3. 8 . The manual sequence alignment of three Greek key motifs, GK1, GK2 and GK3. Residues belonging to the discontinuous regions in Greek key1 are labeled.

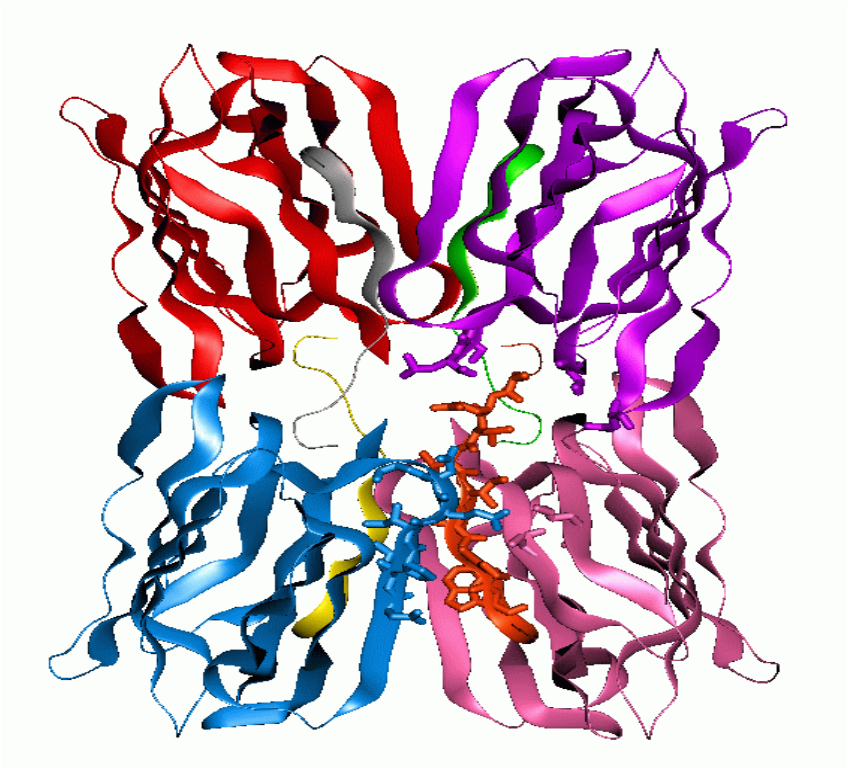
3. 2. 9. The Quaternary Structure

Gel filtration experiments indicated that *hirsuta*, like other lectins of jacalin family, exists as a tetramer in solution (Gurjar *et al.*, 1998). The asymmetric unit of *hirsuta* crystals also contained a functional tetramer (Fig 3. 9. A). This is in contrast to jacalin structure, where the asymmetric unit is formed from two dimers belonging to two different functional tetramers (Fig 3. 9. B). Structurally the four subunits of *hirsuta* are nearly identical (table. 3. 6). The four subunits associate into a flat assembly with overall dimensions of 20 x 60 x 80 Å (Fig. 3. 10). Subunits A and C are related by a non-crystallographic 2-fold axis (P) to subunits B and D. Similarly, subunits A and B are related to C and D by another 2-fold axis (Q). The third 2-fold axis R, is generated by P and Q, at the point of their intersection, the three axes being mutually perpendicular to each other (Fig 3. 10).

Buried surface area

The difference between the sum of the accessible surface areas of the individual subunits and the accessible surface area of the oligomer gives the surface area buried on oligomerization. The surface area buried upon association of any two subunits were computed in order to understand the nature of interactions among the subunits. The computation was performed using the method of Lee and Richards (1971) by employing a probe radius of 1.4 Å. Table. 3. 7, lists the total surface area buried and the non-polar surface area buried with the percentage of non-polar surface area estimated for a pair of subunits. Each subunit lost the same amount of total surface area, around 1760 $Å^2$, out of which around 1050 $Å^2$ is contributed by non-polar residues. When the four subunits associated to form tetramer, they buried a total surface area of 7065 $Å^2$, of which 60% is hydrophobic. The surface area buried at the interface of A-B and C-D subunits is more than that buried at A-C and B-D (table 3. 7), implying that A-B and C-D dimers formed first, before the formation of tetramer. Thus *hirsuta* lectin can be considered a tetramer of two tightly bound dimers. The short stretch of ß-chain which forms an integral part of the tertiary structure also plays a crucial role in the subunit association and stabilization (tables 3. 8 and 3. 9).

A



B

Fig. 3. 9. Diagrammatic representation of asymmetric unit in crystal, A) functional tetramer of *hirsuta*, B) non-functional tetrmer of jacalin. The residues at the inter-subunit region are in liquorice.

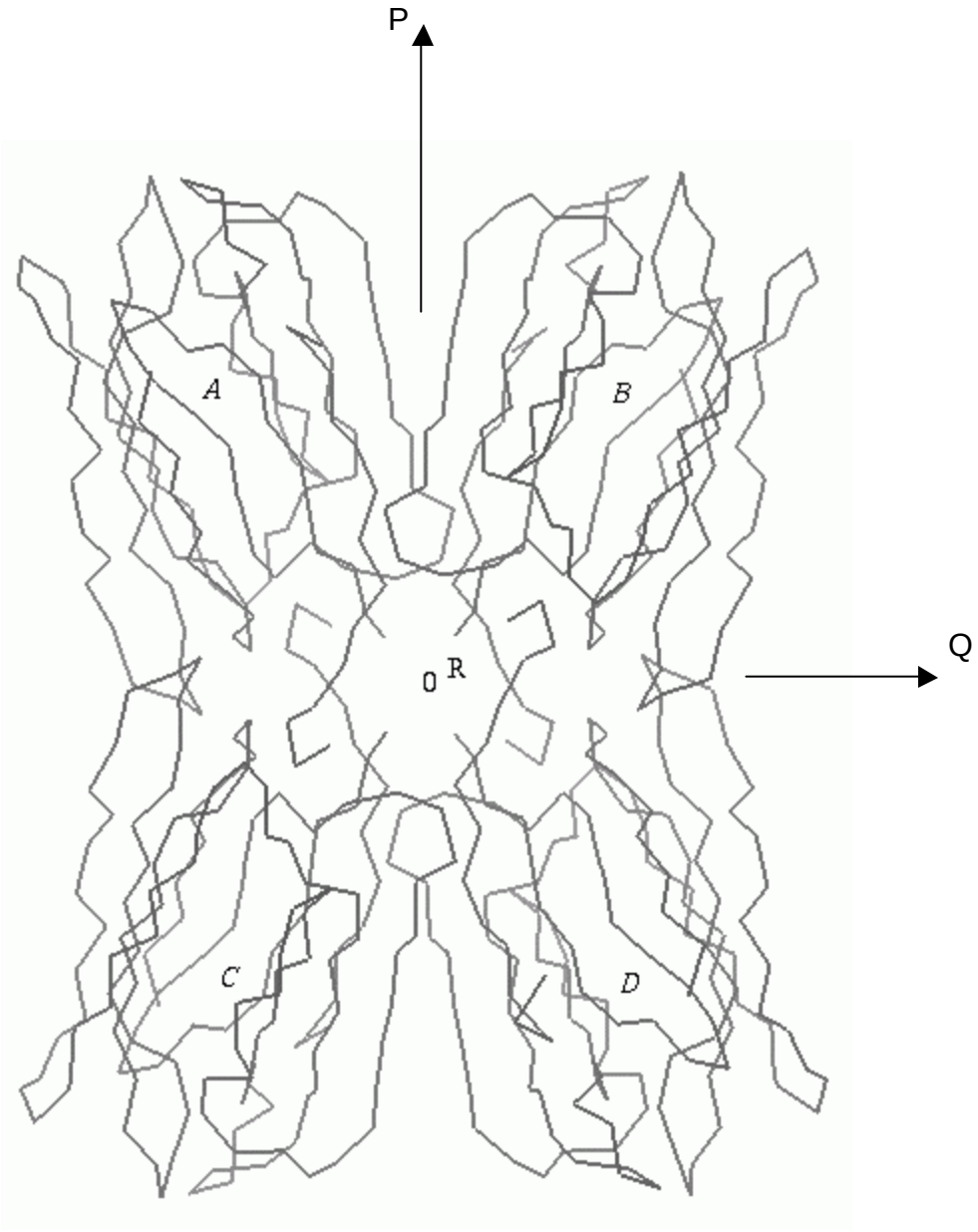


Fig. 3. 10. The Ca trace of *hirsuta* tetramer showing the 222 symmetry. The directions of three NCS 2-fold axes, P, Q, R are also shown. The four subunits are indicated by letters *A, B, C, D*.

Table. 3. 7. Results of buried surface area calculations in the two structures.

Subunits	**Total surface area buried ($Å^2$)**	**Non-Polar surface area buried ($Å^2$)**	**Percentage of non-polar surface area buried**
Orthorhombic form I			
A1-B1	1845	1105	60
A1-C1	1590	960	60
B1-D1	1675	1035	62
C1-D1	1840	1095	58
Tetramer (total)	7065	4250	60
Orthorhombic form II			
A2-B2	1850	1150	62
A2-C2	1600	1010	63
B2-C2	1665	1000	60
C2-D2	1900	1145	60
Tetramer (total)	7140	4380	61

The inter-subunit hydrogen bonds observed between subunits A1 and B1 and A1 and C1 are listed in tables. 3. 8 and 3. 9., classified according to the type. Majority of the hydrogen bonds found between subunits A1 and B1 are between the N-terminal strand of a-chain and the C-terminal strand of ß-chain of Greek key1. Similarly, the residues involved in subunit interactions between A1 and C1 are listed in table. 3. 9. This assembly of subunits is similar to that observed in jacalin structure.

Table. 3. 8. Hydrogen bonds between the subunits A1 and B1 as observed in the orthorhombic form- I.

Nature of the bond	**Subunit A1** Chain, Residue number, amino acid & Atom	**Subunit B1** Chain, Residue number, amino acid & Atom	**Distance** (Å)
Main chain - Main chain	a 110 Asn N	ß 11 Ile O	3.16
	a 110 Asn O	ß 11 Ile N	2.90
	ß 11 Ile N	a 110 Asn O	3.00
	ß 11 Ile O	a 109 Glu N	2.73
	ß 11 Ile O	a 110 Asn N	3.10
	a 107 Pro O	ß 13 Gly N	2.80
	a 61 Pro O	ß 5 Gly N	3.25
Main chain - Side chain	a 105 Asn O	a 8 Ala NE1	3.04
	a 110 Asn OD1	ß 9 Thr O	3.18
	ß 9 Thr O	a110 Asn ND2	3.08
	ß 15 Trp NE1	a 105 Asn O	3.32
Side chain - Side chain	a 105 Asn OD1	ß 15 Trp NE1	3.24
	a 109 Glu OE1	a 117 Lys NZ	3.22
	a 109 Glu OE2	a 128 Ser OG	2.81
	a 110 Asn OD1	ß 8 Gln NE2	3.14
	ß 15 Trp NE1	a105 Asn OD1	3.12
	a 128 Ser OG	a109 Glu OE2	3.16

Table. 3. 9. Hydrogen bonds between the subunits A1 and C1 of orthorhombic form-I.

Nature of the hydrogen bond	**Subunit A1** Chain, Residue number, amino acid & Atom	**Subunit C1** Chain, Residue number, amino acid & Atom	**Distance** (Å)
Main chain – Main chain	a 10 Thr O a 109 Glu N a 133 Leu O ß 7 Ser N	ß 7 Ser N ß 11 Ile O ß 8 Gln N ß 10 Thr O	2.90 2.80 3.00 3.00
Main chain – Side chain	a 8 Ala N a 8 Ala N a 35 Asn OD1 a 35 Asn ND2 a 105 Asn O	a35 Asn OD1 a 35 Asn ND2 a 8 Ala N a 8 Ala O ß15 Trp NE1	3.27 3.23 2.86 3.00 3.04
Side chain – Side chain	a 133 Leu OXT ß 8 Gln OE1	ß 8 Gln NE2 a133 Leu OXT	3.43 3.32

3. 2. 10. Solvent structure

As already mentioned the solvent content of both the crystal forms were found to be very high. Thus the refined structures of *hirsuta* contained large number of solvent molecules. With the search criteria mentioned earlier 763 water molecules were obtained for form I and 534 for form II. Those solvent molecules, found within 2.5-3.5 Å distance from protein atoms, were considered to form the primary solvent shell. These water molecules directly interact with the lectin molecule and involved in making hydrogen bonds with protein atoms. The solvent molecules present in the range of 3.5-6.0 Å distance away from the protein atoms were considered to form the secondary hydration shell. More than 2/3rd of the solvent molecules were found in the primary solvent shell. The orthorhombic form-I contained 552 water molecules in the primary solvent shell and established a total of 525 hydrogen bonds with the protein atoms. The numbers in the orthorhombic form-II were 462 waters and 415 hydrogen bonds, respectively. In total, there were 220 water-water hydrogen bonds in the crystal form-I and 150 in form-II. The solvent molecules present in the secondary solvent shell were generally in contact among themselves and with solvent atoms of the primary shell.

A comparison of the hydration of the structures in two orthorhombic forms was carried out. The structurally significant (invariant) water molecules that are common to both the structures were identified. A water molecule was considered invariant, if it interacted with at least one protein atom common in any two given subunits, and when on superposition of the two subunits along with their hydration shell, the oxygen atoms of the two waters were closer than 1.8 Å (Sadasivan *et al.*, 1998). The details of invariant water molecules interacting with their common protein atoms in form I and form II and the hydrogen bonding distance are listed in the table. 3. 10. The invariant water molecules of subunit A1 of form I are shown in Fig. 3. 11.

Table. 3. 10. The invariant water molecules in the subunits of the two orthorhombic forms. The waters, protein atoms and the hydrogen bonding distance shown here are for subunits A1 and A2 of the two crystal forms. C1 and C2 are the subunits C of orthorhombic forms I and II. The residues of the ß-chain are shown with '(ß)'.

Common protein atom Residue ID atom	**Invariant water molecules of subunit A1**	**Distance (Å)**	**Invariant water molecules of subunit A2**	**Distance (Å)**
Asn105 N	W9	3.04 (A1)	W518	3.15 (A2)
Asn105 O		2.61 (A1)		2.94 (A2)
Asn105 N		3.11 (C1)		2.92 (C2)
Asn105 O		2.70 (C1)		2.67 (C2)
Asp6 OD1	W14	2.68	W35	2.80
Glu42 O		2.90		2.84
Val30 N		2.60		2.73
His44 NE2		3.30		3.00
Thr79 O	W25	2.97	W16	2.67
Asp125 OD1		2.76		2.94
Pro103 O	W52	2.77	W14	2.94
Asn195 OD1		2.61		3.04
Asn105 OD1	W90	2.66	W535	3.30
Thr102 O	W151	2.60	W493	2.75
Thr 72 O		3.10		2.66

Asn 6 OD1	W5	2.98	W632	2.60
Gln 42 N		3.00	W165	2.60
Leu131 O	W19	2.80	W116	2.50
Ala 10 N (ß)		2.66		3.00
Gly111 O	W20	2.75	W647	2.70
Ile 65 O		3.13		2.90
His 44 ND1	W26	2.73	W35	3.80
Ser 7 O (ß)	W27	3.22	W756	3.77
Phe 4 O	W40	3.02	W466	2.78
Gly118 O		3.23		3.00
Thr 9 N (ß)	W50	3.04	W521	3.00
Thr 9 O (ß)		3.40		3.00
Ile108 O	W96	2.80	W335	3.00
Thr 85 OG1	W155	2.80	W572	2.80
Ser 83 OG		3.00		3.95
Asn74 N	W246	3.25	W566	3.90
Ala 3 O	W251	3.08	W777	3.97
Lys 2 NZ		3.10		4.00
Gln 42 NE2		3.37		4.30
Gly13 O (ß)	W874	2.63	W203	2.84
Asn 20 ND2	W58	3.02	W79	2.66
Lys 18 N	W777	2.81	W225	3.20
Trp123 NE1	W72	2.90	W520	2.74
Ile 120 O	W8	2.72	-	-
Gly 70 N	W82	3.23	-	-
Lys 2 O	W152	2.87	-	-
Gly16 N (ß)	W244	2.90	-	-

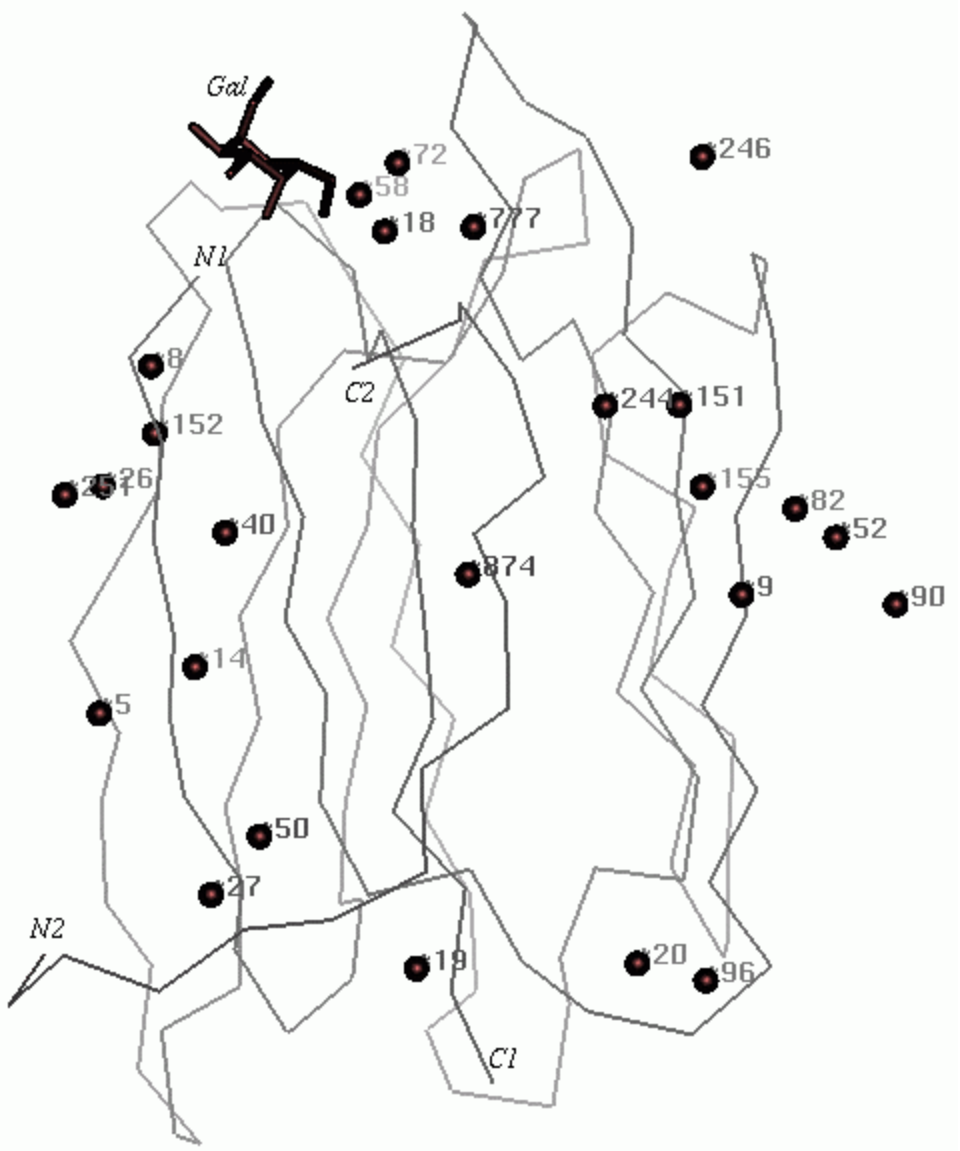


Fig. 3. 11. The diagram showing the invariant water molecules (van der Walls spheres) as seen in the sugar bound (at the top left, Gal) subunit A1 of orthorhombic form I. The water molecules are labeled according to residue ID number.

In total 21 solvent molecules were found common to both the structures in the two crystal forms. Four solvent molecules (the last four listed in table. 3. 10), though invariant in the subunits of orthorhombic form I, were not found in the sites of subunits in orthorhombic form II. The invariant water W9 of form I interacts with atoms N and O of Asn 105 of a-chain from both the subunits A1 and C1. Similarly, W518 of form II, interacts with N and O of Asn 105 of a-chain from both the subunits A2 and C2. The invariant water W25 of form I and W16 of form II interact with the sugar binding residue Asp 125 OD1 (table. 3. 10). These two water molecules interact also with W901 and W125, respectively. W 901 and W125 in turn interacts with two other waters W661 and W453 which make hydrogen bonds with O3 atom of the sugar in the orthorhombic forms I and II, respectively, (table. 3. 11 and Figs. 3.13. A-B). Similar network of water mediated interactions were observed in the crystal structure of another galactose specific jacalin lectin, MPA-T-antigen complex (Lee *et al*., 1998), thus indicate their importance in sugar binding.

3. 2. 11. Crystal packing

The lectin crystallized in the two orthorhombic forms had the same space group $P2_1\ 2_12_1$, but different unit cell dimensions. It is interesting to note here that one of the jacalin forms also crystallized in $P2_12_12_1$ space group with cell dimensions, a=92.5, b=98.7 and c=164.5 ?, very similar to that of orthorhombic form I of *hirsuta* (Dhanaraj *et al*., 1988). Unfortunately, the three-dimensional structure of this form is not available for further comparison between *hirsuta* and jacalin. The asymmetric unit of both the *hirsuta* forms contained a tetrameric molecule. Since the unit cell dimensions of the two orthorhombic forms were different molecular arrangement in their crystals differs. The molecular packing diagrams viewed along the three unit cell axes of the orthorhombic forms are shown in Figs. 3. 12. A-C The interacting amino acid residues at the intermolecular contact region that stabilize the crystal packing in form I are A76 Ser and A22 Glu with 71Tyr and 100Ser of a subunit from a symmetry related neighbor tetramer. Similarly in the form II C59 Asp interacts with a 78Tyr and C13 Arg with O2 atom of the sugar from the symmetry related neighbor subunit.

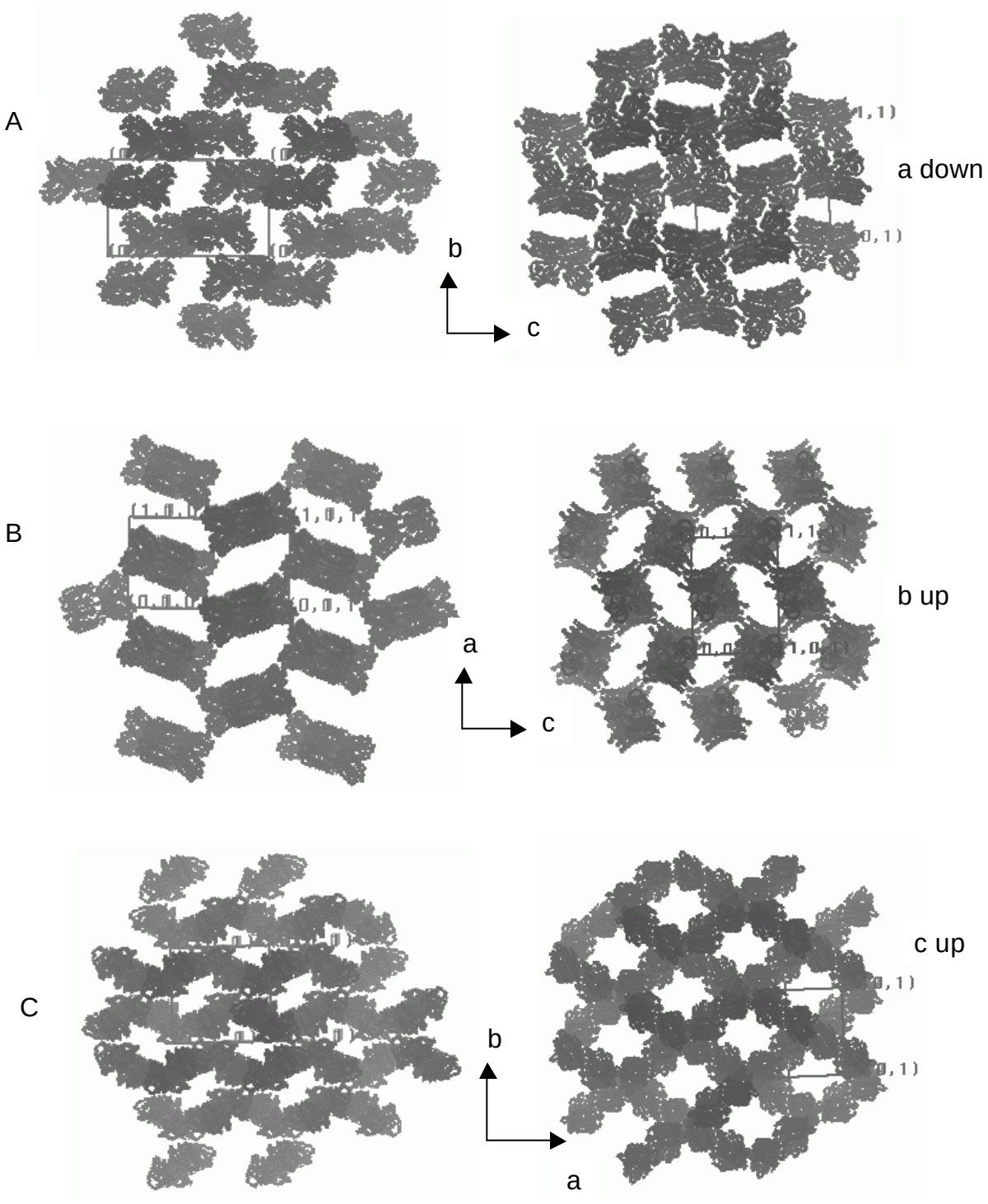


Fig. 3. 12. The crystal packing diagrams showing the differences in packing between the orthorhombic form I (left three) and orthorhombic form II (right three). The packing is viewed along a, b, c, directions in (A), (B), (C), the directions of two perpendicular axes are also indicated.

3. 2. 12. Interactions with the sugar methyl-a-D-galactopyranoside

One of the main goals of structural studies of the *hirsuta* lectin was to elucidate the atomic basis of its carbohydrate recognition. The difference Fourier map (Fo-Fc) clearly showed the presence of methyl-a-D-galactose (me-a-gal) bound to each of the four subunits. As implicated by the biochemical and spectroscopic studies (Gaikwad *et al.*, 1998), tryptophan and tyrosine residues were in contact with the sugar molecule, through their main chain amino and carbonyl groups. Interactions predominantly occur through the net work of main chain hydrogen bonds and van der Walls interactions with the cluster of aromatic amino acid residues at the binding site. Figs. 3. 13. A-B shows the binding of me-a-gal by *hirsuta* lectin as seen in orthorhombic form I and II, respectively, and table 3. 11, lists the corresponding interactions.

The carbohydrate binding site of each subunit is generated by the loops, which connect the inner strands of Gree keys 1 and 3. Residues that are involved in binding are Tyr 78 and Val 80 from GK3 and Gly1, Gly121, Tyr122, Trp123, and Asp125 from GK1. All these residues which belong to α-chain are within 4.0 Å distance from the sugar molecule (Fig. 3. 13. A-B). These residues are conserved in the homologous jacalin and MPA structures as well and involved in the sugar binding (Fig. 3. 14. A). Although the galactose sugar is polar, the carbon atoms at five epimeric positions in the ring and one at the adjacent exocyclic position form a continuous hydrophobic region. This region is found packed against the aromatic side chains of the lectin. Shielding of this hydrophobic surface from solvent contact adds to the stability of the complex. The aromatic side chains, Tyr78, Tyr122, Trp123 were found within van der Walls distance from C1, C3, C4 and C5 carbons of sugar. It is now well established that the galactose binding always involves stacking interaction with an aromatic residue against the B face of the sugar (Pratap *et al.*, 2002). This stacking interaction is between an extended patch of partially positively charged aliphatic protons on the B-face of the ring and the p-electron cloud of the aromatic residue. The side chain of Tyr78 stacks against the B-face of the galactose ring. Most of the hydrogen bonds with the sugar molecule come from the loop of Greek Key 1. Six of the nine hydrogen bonds involve main-chain

atoms. The terminal amino group of the α-chain (Gly1) hydrogen bonds to O3 and O4, the NH group of Tyr122 to O5 and O6, and both N and O of Trp123 to O6 and OH of Tyr78 with O1 atom of sugar molecule. The side chain of Asp125 makes 3 hydrogen bonds with the sugar molecule, wherein OD1 bonds to both O4 and O6, and OD2 to O4. The side chain carboxyl group of Asp125 is held in position through interaction of OD1 with water W25 (table. 3. 11.) and OD2 with O$^{?}$ of Ser119. In addition to the above interactions, an Arg13 from a neighboring subunit contacts the O2 atom of the sugar molecule in form I. This interaction is not observed in other galactose specific jacalins. This is the only interaction the O2 atom of sugar has with lectin. This additional interaction and the resultant subunit association, if stable in solution, could block the binding of *hirsuta* to T-antigen and other polysaccharides. It may be noted that *hirsuta* lectin was found highly specific for monomer sugars (Gaikwad *et al*., 1998). In contrast to that observed in orthorhombic form I, the O2 atom of the sugar is not involved in interaction with any protein atoms in orthorhombic form II and is exposed to solvent, similar to what observed in the jacalin structure. Even in mannose binding lectins the situation is same when O2 is in axial configuration. This could be the reason that the binding affinity of *hirsuta* for GalNAc is not very different from that of a-Gal (Gaikwad *et al*., 1998). The O2 atom of bound sugar molecule in *hirsuta* is observed as interacting (blocking) in orthorhombic form I and solvent exposed in orthorhombic form II. In both cases we have tried to provide tentative reasons why *hirsuta* bind ligands such as T-antigen and GalNAc.

As already mentioned the O4 axial hydroxyl group of galactose forms three hydrogen bonds, two with the side chain of Asp125 and one with the N-terminal amino group of Gly1 of a-chain. When this O4 is changed to the equitorial configuration, as in mannose, only one of the hydrogen bond with Asp125 will still be retained but that with Gly1 will be lost. It has been postulated that the galactose specific group of jacalin lectins, acquired galactose specificity through the creation of free N-terminal for glycine (a-chain) by post-translational modification (Shankaranarayanan *et al*., 1996). Thus the specificity is considered chiefly determined by post-translational processing. In addition,

modeling studies have shown that the carbohydrate binding site of jacalin can accommodate, but may not bind mannose (Pratap *et al*., 2002).

The higher affinity of the *hirsuta* lectin towards me-a-gal compared to a-galactose (Gaikwad *et al*., 1998) could be explained in terms of the stacking of methyl group between the phenyl rings of Tyr78 and Tyr122 and the resultant favorable interactions (Figs. 3. 13. A-B).

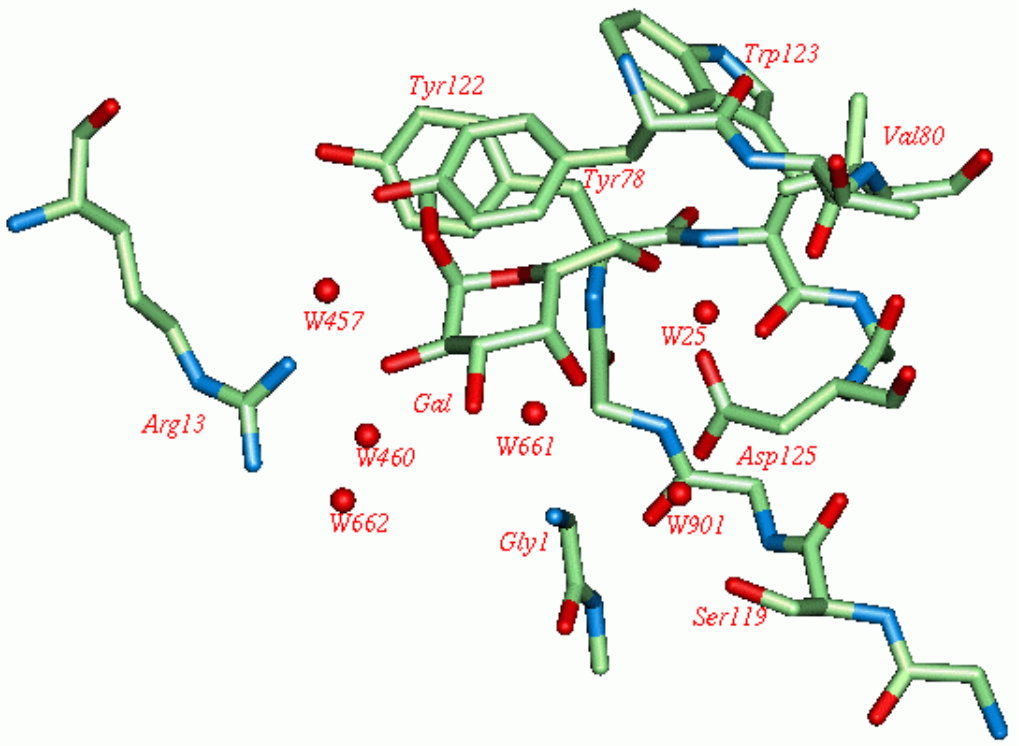
Table. 3. 11. The polar interactions involving sugar-protein and sugar-water atoms at the carbohydrate binding sites of subunits A1 and A2 of the two forms.

Sugar atom	**Protein/ Water atom** Residue ID atom	**Distance (Å) in subunit A1 of form I**	**Distance (Å) in subunit A2 of form II and the interacting water**
O1	Tyr78 OH	3.33	3.25
	W71 O	4.30	3.92 (W259)
	W457 O	3.12	-
	W619 O	4.30	-
O2	Arg13 NH1	2.74	-
	Arg13 NH2	4.23	-
	W457 O	2.80	3.63 (W72)
	W460 O	3.31	3.35 (W259)
	W517 O	3.44	2.83 (W767)
	W662 O	3.17	-
O3	Gly1 N	3.10	2.71
	W241 O	3.84	2.54 (W453)
	W460 O	2.64	3.15 (W767)
	W661 O	2.64	4.31 (W684)
	W662 O	3.84	3.47 (W902)

O4	Gly1 N	3.14	3.41
	Gly121 N	4.10	4.40
	Tyr122 N	4.34	4.37
	Asp125 OD1	2.84	2.63
	Asp125 OD2	3.18	3.75
	W25 O	4.34	4.16 (W16)
	W661 O	3.80	4.30 (W453)
O5	Tyr122 N	3.04	3.08
O6	Gly121 N	4.18	4.20
	Tyr122 N	3.01	3.26
	Trp123 N	2.92	3.40
	Trp123 O	3.22	3.22
	Asp125 OD1	2.99	3.06
	Asp125 OD2	4.48	4.45
			4.32 (W16)

Comparison of *hirsuta* lectin with jacalin considering sugar recognition per se, confirms the similarity in sugar recognition between these proteins. The differences between the structures of jacalin and *hirsuta* are: (1) *hirsuta* contain a functional tetramer in its asymmetric unit, whereas jacalin tetramer is made up of two dimers drawn from the two symmetry related functional tetramers (Figs. 3. 9. A-B). (2) However, unlike jacalin and MPA, *hirsuta* does not bind T-antigen, we have seen that in form I an Arg residue from a neighboring subunit is involved in sugar binding (Fig. 3. 13. A). This could effectively block a disaccharide from binding. Since jacalin does not have this interaction (Fig. 3. 14. A), the difference between T-antigen specificity of jacalin and *hirsuta* could perhaps be explained. (3) The observed difference density in region opposite to sugar binding site of *hirsuta* imply a bound small molecule (Fig. 3. 2), whereas nothing is reported to bind at this position in the jacalin structure.

A



B

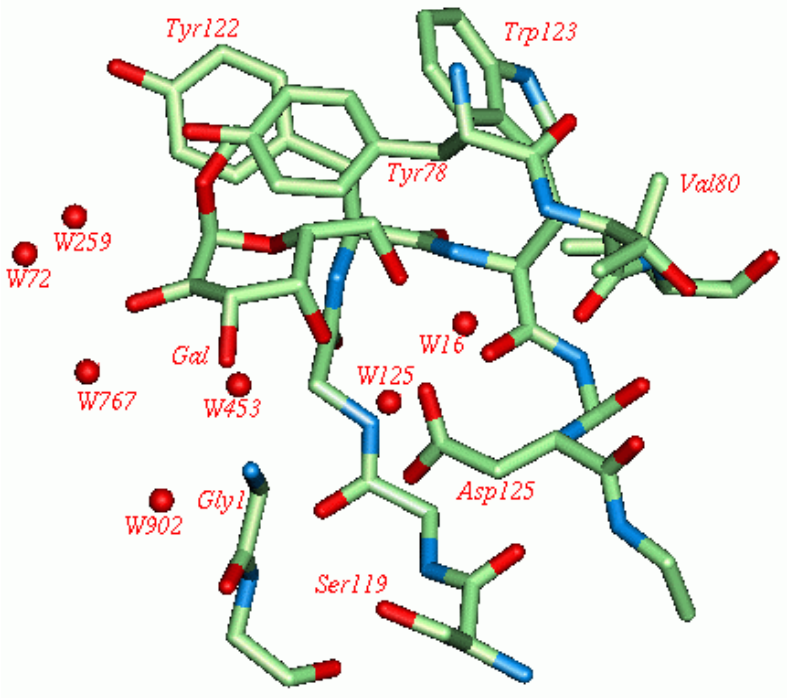
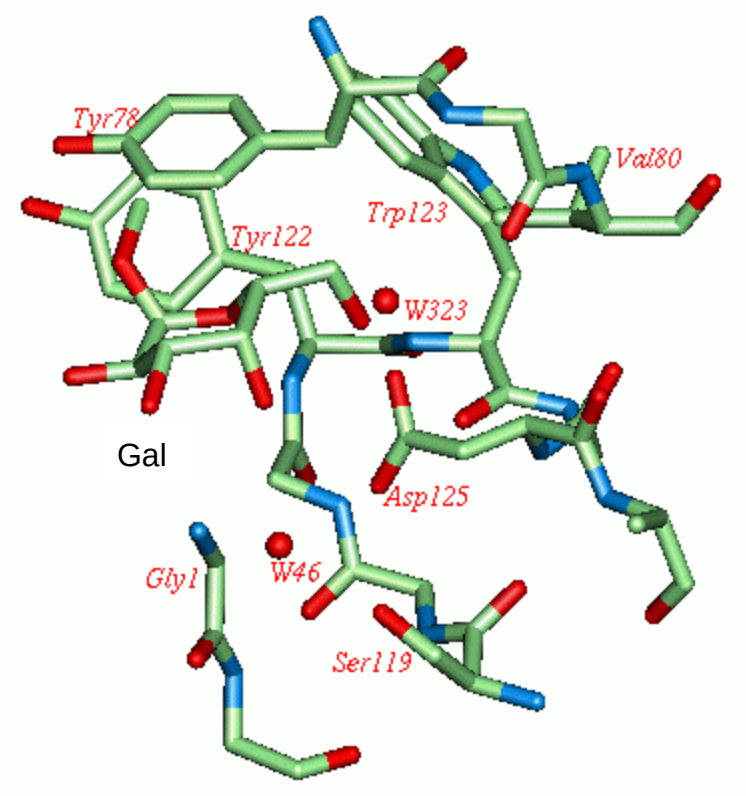


Fig. 3. 13. The carbohydrate binding sites of *hirsuta*, A) orthorhombic form I, B) orthorhombic form II, bound to me-a-gal (Gal), showing the amino acids and solvent molecules around the recognition site. The Arg13 interaction from the neighbor subunit is seen only in orthorhombic form I.

A



B

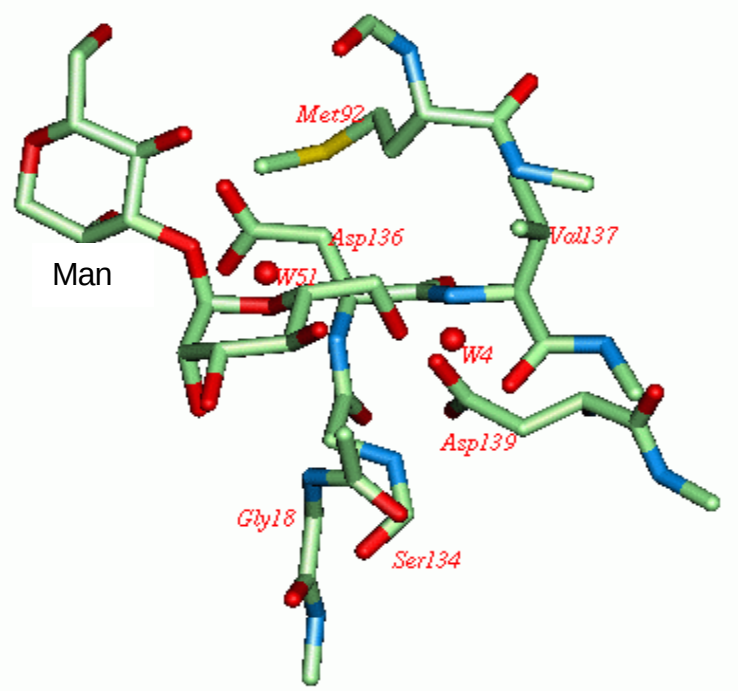


Fig. 3. 14. The carbohydrate binding sites of, A) jacalin bound to me-a-gal (Gal), B) heltuba bound to Man a-1-3Man (Man).

3. 2. 13. Structural comparison with Jacalin family lectins and other homologous proteins of DALI data base

Comparison of *hirsuta* structure with the DALI database of protein structures (Holm & Sander, 1995) revealed 11 homologous protein structures (table. 3. 12). Three of them (jacalin, MPA, heltuba) belong to jacalin family lectins. The *hirsuta* lectin structure superposed well with the jacalin structure (Fig. 3. 15. A). The two monomers superpose well all along both a and ß-chains. When MPA and *hirsuta* structures were superposed (Fig. 3. 15. A) the Ca positions had r. m. s. deviation of only 0.30 Å, indicating that the two proteins possessed similar folds. Heltuba, another plant lectin, represents the first three-dimensional structure determined for a member of the mannose specific group of jacalin lectins (Bourne *et al*., 1999). The heltuba octamer assembly represents the highest number of subunits association observed among jacalins and it contains eight exposed carbohydrate binding sites. Superposition of *hirsuta* and heltuba monomers showed r. m. s. deviation of 1.5 Å for all the Ca atoms. The r. m. s. deviation was reduced to 1.1 Å between the two monomers after excluding the N-terminal segment (1-10 residues of a-chain, 4-9 residues of ß-chain) and two C-terminal residues of ß-chain. Large deviations of up to 4 Å was observed in the loop regions between the structures. The superposed Ca trace of the two structures is shown in Fig. 3. 15. B.

Two insecticidal toxins from *Bacillus thuringiensis* (Bt) and a vitelline membrane outer layer protein I (VMO-I) from hen egg also belong to ß prism family and overlaps over the entire region of a-chain of *hirsuta*. In fact the ß prism-I fold was first observed in the crystal structure of domain II of d-endotoxin (CryIIIA) from Bt at 2.5 Å resolution (Li *et al*., 1991). Another insecticidal protein from Bt, cry toxin (Cry2Aa), whose crystal structure has recently been reported (Morse *et al*., 2001) also contained the similar fold as Cry IIIA. These toxin structures comprise of three domains, the domain I is a seven helix bundle, domain II has three anti-parallel ß-sheets and domain III at

the C-terminus is a sandwich of two anti-parallel ß-sheets. Domain II has two four stranded Greek key motifs and a three stranded ß-sheet along with a two turn a-helix, packed around a hydrophobic core. There is no identity among the residues of the two Greek key motifs except for five residues. The internal strands of the two greek key motifs of toxins are lengthier than those of *hirsuta*, hence the overlap in these regions is very poor. The superposed structures of *hirsuta* and d-endotoxin is shown in Fig. 3. 15. C. The residues in those regions around which the two structures deviate are numbered in Figure. 3. 15. C.

The crystal structure of VMO-I at 2.2 Å resolution exhibited the ß prism-I fold (Shimuzu *et al.*, 1994). This fold is made up of three anti-parallel ß-sheets that form three Greek key motifs. The ß-sheets are made up of six long ß-strands and eight short ß-strands. The three ß-sheets organize to form a triangular prism structure. Interestingly, this is the only structure in which the 3-fold pseudo internal symmetry of the molecule is reflected also in its sequence. Out of 49 residues 11 are identical in all the three Greek key motifs. When compared to d-endotoxin the three-fold internal symmetry is more regular in VMO-I (Shimuzu *et al.*, 1994). The superposition of *hirsuta* structure on to VMO-I structure is shown in Fig. 3. 15. D. The difference between the two structures is more in the loop regions. The overlap of rest of the five proteins (the last five listed in table. 3. 12) is confined to a small domain region of around 40 residues.

Table. 3. 12. The r. m. s. deviation in Ca positions of *hirsuta* with the homologous protein structures of DALI databse. The region of the overlap is shown in the last column.

Homologous Protein (source)/ PDB Code	**The number of Ca positions superposed/ the total number of residues in the protein**	**r. m. s. deviat-ion in Å.**	**The residues in the superposed region,** *Hirsuta* lectin = homologous protein.
Jacalin (jack fruit)/ 1JAC	147/147	0.25	All residues
MPA (Osage orange)/ 1JOT	147/147	0.30	All residues
Heltuba (*Helianthus tuberoses*)/ C3K	147/149	1.50	All residues
d-endotoxin (*Bacillus thuringenisis*)/ 1ClY	130/577	2.30	Phe4-Leu133 = Asn290-His456
Insecticidal crystal protein (*Bacillus thuringenisis*, Hd-1)/ 1I5P	133/625	2.80	Gly1-Leu133 = Ala287-Ala470
VMO-I (Hen egg,	133/163	2.75	Gly1-Leu133 = Gly18-Lys163

Gallus Gallus)/ 1VMO			
Cytochrome f (*Brassica rapa, Turnip*)/ 1HCZ	49/250	2.90	Gly1-Tyr19=Asn168-Glu186; Ile25-Asn55=Lys187-Ala197; Gly36-Gly41=Glu200-Asp205; Asn43-Phe47=Ile206-Gly210; Leu112-Gly115=Gly218-Ile221; Phe116-Ser119=Pro226-Ser229
CD4 (domains 3 & 4, rat)/ 1CID	40/177	3.20	Thr66-Val75=Ser25-Ala34; Ser76-Val98=Ala37-Thr59; Gly101-Pro107=Gly87-Leu93
Gelation actin binding protein 120 (Slime mold)/ 1QFH	39/212	3.40	Thr66-Val75=Phe782-Glu791; Gly77-Tyr93=Val812-Ile828; Gly94-Val98=Ser831-Gln835; Gly101-Pro107=Ser849-Thr855
Charcot-Leyden Crystal protein (human eosinophil)/ 1LCL	38/141	3.60	Thr66-Val75=Gly17-Pro26; Ser76-Val80=Asp85-Phe89; Arg82-Tyr93=Glu90-Gln101; Pro95-Val98=Val102-Asn105; Gly101-Pro107=Thr133-Tyr139
Myosin fragment mutant (*Dictyostellium discoidin*, slime mold) / 1LVK	42/743	4.40	Thr66-Val77=Thr79-Asn87; Gly97-Asn110=Ser443-Ser456; Phe127-Leu133=Asn649-Cys655;

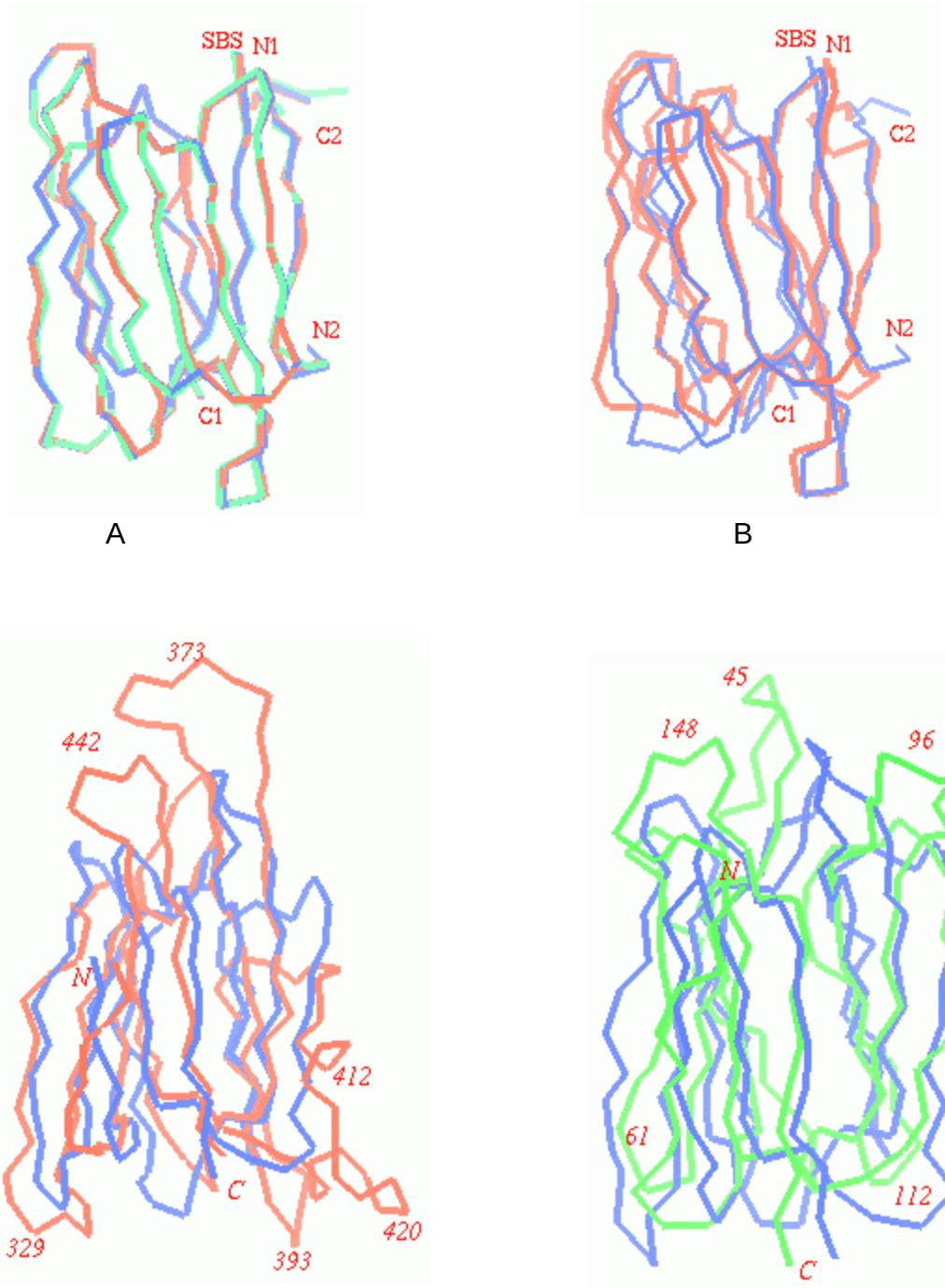


Fig. 3. 15. The superposition using Ca positions of *hirsuta* subunit (blue) on to those of, A) jacalin (red) and MPA (green), B) heltuba (red), C) CRD of d-endotoxin (domain II; red), D) VMO-I (green). The residues around which the two structures (of d-endotoxin and VMO-I) show maximum deviation are labeled.

3. 3. CONCLUSIONS

Crystal structure determination of galactose specific lectin from *Artocarpus hirsuta* has shown that it belonged to jacalin family and shared close structural and sequence similarity with jacalin and MPA. The sequence determined from electron density shows a close evolutionary relationship with jacalin. *Hirsuta* crystals have comparatively high solvent content. Unlike jacalin the asymmetric unit of *hirsuta* lectin is a functional tetramer. Rationale for galactose specificity lies in the creation of a free N-terminal at a-chain through post-translational modification. *Hirsuta* lectin has a hydrophobic pocket on the side away from sugar binding side. The mechanism of carbohydrate binding is very similar to that in jacalin.

In conclusion, the present structural study of *hirsuta* lectin bound to me-a-gal has again demonstrated the mode of carbohydrate recognition by lectins particularly by the Moraceae family lectins. The three-dimensional structure of *hirsuta* lectin provide structure for another member of jacalin family lectins and lectins in general. The structure could account for the specificity and affinity of the lectin for its different ligands. The structure satisfactorily explains the observations based on the spectroscopic methods, carried out on the binding of lectin to the various galactose sugar derivatives by lectin in solution.

CHAPTER-4

ISOLATION, PURIFICATION AND CHARACTERIZATION OF BOWMAN-BIRK TYPE ISOINHIBITORS FROM *VIGNA UNGUICULATA* SEEDS.

4. 1. INTRODUCTION

Cowpea, also called as 'black-eyed pea' (*Vigna unguiculata* (L). Walp) is an important grain legume crop and provides as one of the main source of dietary protein. The plant grows in dry land in India, sub-Saharan Africa, Brazil and elsewhere. The plant is known to be highly susceptible to some post-flowering insect pests such as pod borers (*Maruca vitrata*) and pod sucking bugs (*Calvigralla spp*) which cause substantial damage to grain yield. Genetically engineering cowpeas using biotechnological approaches could help overcome these grain lose (Pellegrineschi, 1997).

The presence of various serine protease inhibitors -including trypsin/chymotrypsin inhibitors, and other types of protease inhibitors- in the *Vigna unguiculata* seeds has been reported (Gatehouse *et al.*, 1980; Gennis & Canter, 1976; Rele *et al.*, 1980; Vartak *et al.*, 1980). These reports deal with the purification, characterization and other aspects of these protease inhibitors, including the Bowman-Birk type inhibitors. In a series of reports by Gennis & Canter (1976), the purified double headed BBIs from black-eyed peas have been characterized in terms of amino acid composition, optical properties, subunit interactions, energetics of protease-inhibitor interactions and solution structure studies of their complexes. The complete amino acid sequence of one of the black-eyed pea trypsin/chymotrypsin inhibitor (BTCI) has also been reported by Brazilian researchers (Morhy & Ventura, 1987). Recently, studies pertaining to the stability aspects of BTCI has been reported (Da Silva *et al.*, 2001). The protein and cDNA sequences of Bowman-Birk type PIs from cowpea has been reported by Hilder *et al.*, (1989). In a different study, Hilder and group have shown the field resistance of cowpea BBIs to its major insect pest, brudchid beetle; *Callosobruchus maculantus,* (Gatehouse *et al.*, 1979) and the anti-metabolic activity of PIs towards a wide range of herbivorous insect species (Gatehouse & Hilder, 1988). In recent times the cowpea trypsin/chymotrypsin (CpTI) gene expressed from various transgenic plants found to enhance resistance to herbivorous insect pests (Ghoshal *et al.*, 2001; Hilder *et al.*, 1987; Ussuf *et al.*, 2001).

In this chapter author has described the purification and characterization of Bowman-Birk type isoinhibitors from the cowpea seeds. The purification of BBI to homogeneity was faced with problems since the seeds contained multiple isoinhibitors, obtaining a homogenous preparation for any one isoinhibitor was not easy. Finally few of the isoinhibitors were purified to considerable level of homogeneity using FPLC technique. Our aim was to obtain each of the isoinhibitors in homogeneous form, then to sequence and crystallize them, in order to carry out three-dimensional structural studies using X-ray diffraction. This chapter also reports the N-terminal amino acid sequence of one of the purified isoinhibitors.

4. 2. RESULTS

4. 2. 1. BBI purification

The details of various techniques used for purification of BBI is described in chapter2. The cowpea seeds were crushed to homogeneity using buffer-salt solution to dissolve the proteins. In the next step (0-40% ammonium sulfate fractionation), the precipitate obtained was discarded since it did not contain protease inhibitor (PI) activity. However, the 40-80% fractionation step precipitated most of the inhibitor activity (table. 4. 1). The precipitated PI was redissolved in a minimum volume (30 ml) of 10 mM phosphate buffer and dialyzed against the same buffer to remove ammonium sulfate. The volume of the PI solution after dialysis increased to 320 ml. The inhibitor's stability in acidic medium was exploited for removing contaminating proteins through acid precipitation without altering PI activity. Significant reduction in the total protein concentration while retaining the activity of the PI (table. 4. 1) and improvement in OD 280 to 260 ratio indicated that acid precipitation is a useful step in the purification of BBI.

After the acid precipitation step pH of the PI solution was readjusted from 6.3 to 7.5 with NaOH and concentrated to 150 ml. This solution with protein concentration of 150 mg/ml was loaded on to DEAE-Sephadex pre-equilibrated (10 mM phosphate buffer, pH 7.5) column in two batches. The initial (wash) fractions contained subtilisin inhibitory activity as reported in the literature (Rele

et al., 1980) was discarded. The BBI was eluted from the column by applying buffer-salt combined gradient containing 10-50 mM phosphate buffer at pH 7.5 and 0.0-1.0 M NaCl. The elution profile is shown in Fig. 4. 1. The PI activity eluted in the gradient range of 20-30 mM phosphate buffer along with 0.3-0.5 M NaCl. Fractions having high activity were pooled and concentrated to 3 ml using amicon unit. Most of the impurities including plant hormones (coloring matter) were washed out during DEAE purification step. Since major part of the other proteins were removed in this step a significant improvement in the specific activity of the BBI was achieved (table. 4. 1).

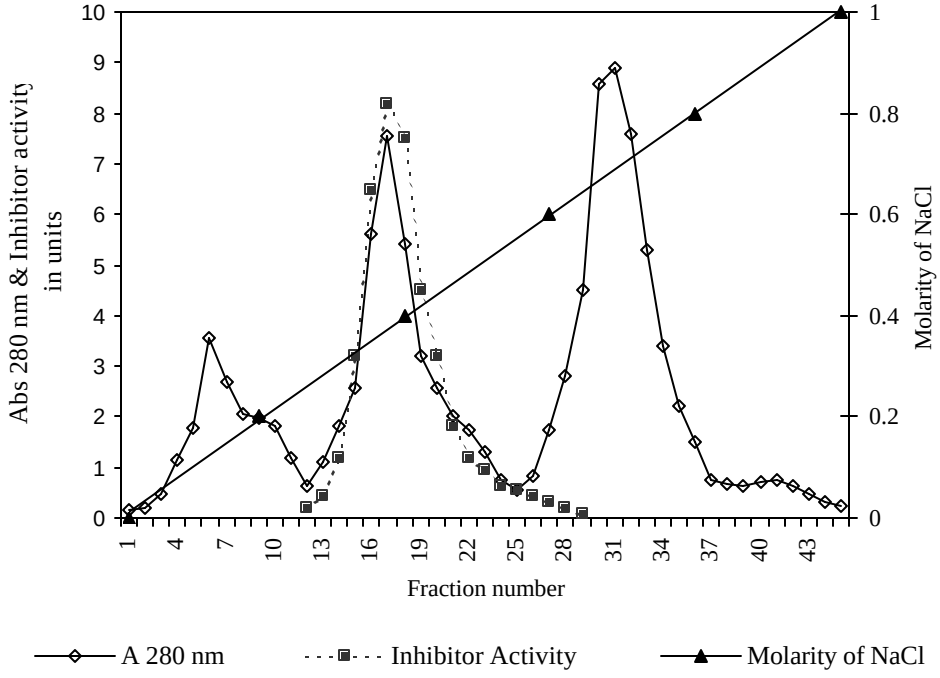
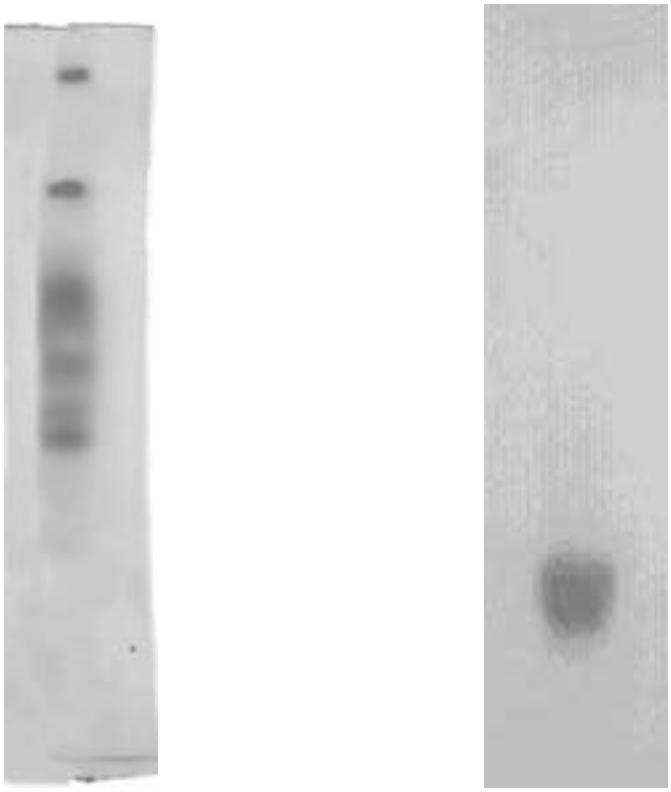


Fig. 4. 1. Elution profile of the BBI during DEAE-Sephedex ion-exchange chromatography.

The purity of the preparation from the DEAE-Sephadex step when checked using SDS-PAGE showed multiple bands (Fig. 4. 2. A). Thus the PI was further purified using gel filtration technique on Sephadex G-50 column. The concentrated PI solution of 3 ml was loaded on to the column and eluted at a flow rate of 15 ml /hr. In this step, the PI peak eluted well separated from a major high molecular weight protein and other proteins (Fig. 4. 3.) Fractions having inhibition activity were pooled.



A B

Fig. 4. 2. The SDS-PAGE run of BBI sample showed, A) multiple bands (12% gel) after DEAE-Sephadex purification, B) single but diffused band (15% gel) after gel filtration purification.

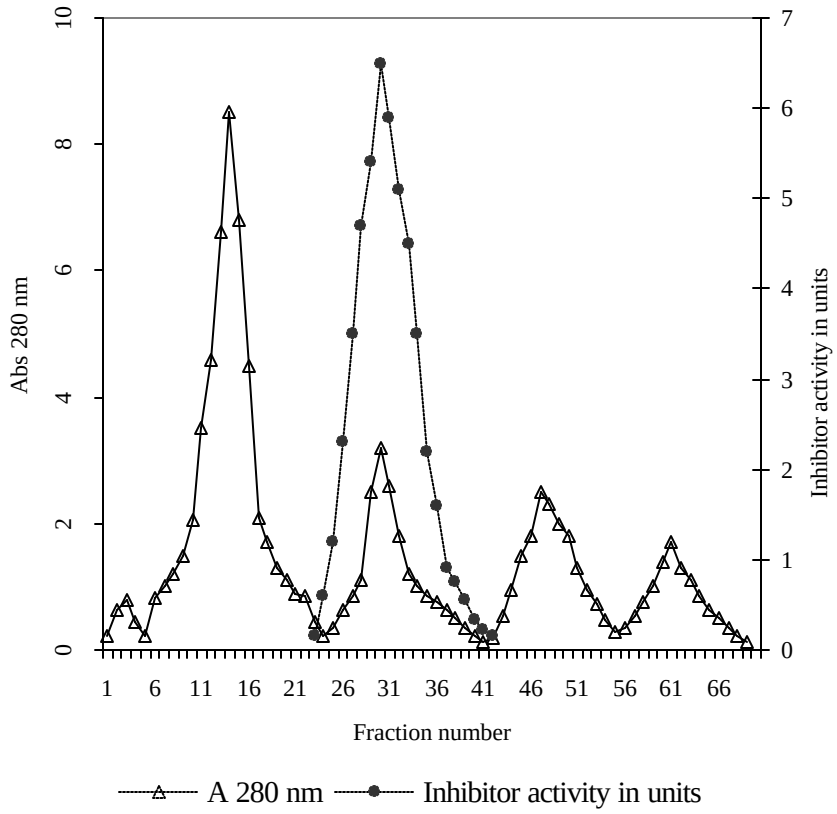


Fig. 4. 3. Elution profile of the BBI in Sephadex G-50 gel filtration chromatography.

4. 2. 2. Homogeneity of the BBI preparation

The homogeneity of the BBI after gel filtration purification was tested using different methods as described in Chapter 2. The SDS-PAGE was carried out using 15% gel in a slab gel electrophoresis unit. From the single band obtained in the electrophoresis, it was confirmed that the preparation was reasonably pure (Fig. 4. 2. B); but the diffused band indicated the likely presence of multiple isoinhibitors with nearly same molecular weights. It was later confirmed, like in many other plant PIs, this preparation also had several isoinhibitors in it. Both the gel isoelectric focussing and X-ray film-contact print experiments showed the presence of at least 8 isoinhibitors (Fig. 4. 4. A-B). The

visualization of well separated bands of different intensities indicate that the isoinhibitors carry different charges and their concentration varied. In the gel isoelectric focussing run, all the PIs got focussed between the pH range 4-6. Further attempts were made to separate the isoinhibitors for obtaining diffraction quality single crystals. The high purity was also required for amino acid sequencing, to compare it with another published sequence of a isoinhibitor from the same seeds (Morhy & Ventura, 1987).

4. 2. 3. FPLC using MonoQ

Further purification of the BBI was carried out by FPLC technique using anion exchanger MonoQ column. The samples injected to the column usually contained 25 mg of protein in 500 µl volume. The gradient program used for separation was standardized as described earlier in chapter 2. The bound proteins were eluted using the buffer-salt gradient containing 10 mM Tris-HCl at pH 7.5 and 1 M NaCl. The isoinhibitors were eluted in the range of 10-25% B (where 'B' represents the concentration of NaCl). Six major peaks, which showed PI activity, have been pooled separately. These BBIs were labeled as P-I through P-VII in the order of the sequence they were eluted in the gradient. The isoinhibitors were then dialyzed and re-injected and elution was carried out as previously. It was required to repeatedly reload at least three times to achieve homogeneity for an isoinhibitor. Fig. 4. 5 shows the typical elution profile of the PI during the first loading. The elution profile corresponding to successive loading in the case of one isoinhibitor (P-IV) is shown in Figs. 4. 6. A-C.

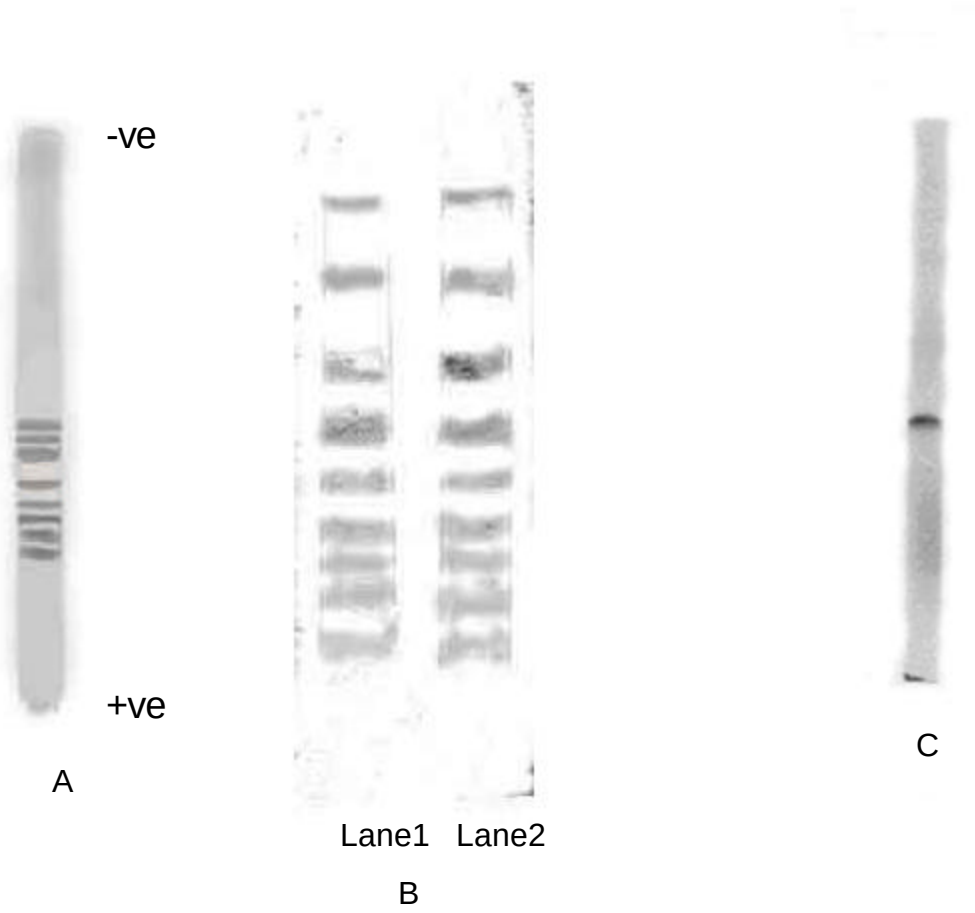


Fig. 4. 4. A) Band pattern of multiple isoinhibitors run on 7.5% polyacrylamide gel during gel isoelectric focussing using pH-gradient 3.0- 7.0.

B) Band pattern of isoinhibitors exposed on X-ray film-contact print technique, Lane1- stained with trypsin , Lane2- stained with chymotrypsin.

C) The single band obtained for isoinhibitor of P-IV on gel IEF (pH-gradient 4.0-6.0) after FPLC purification of the sample.

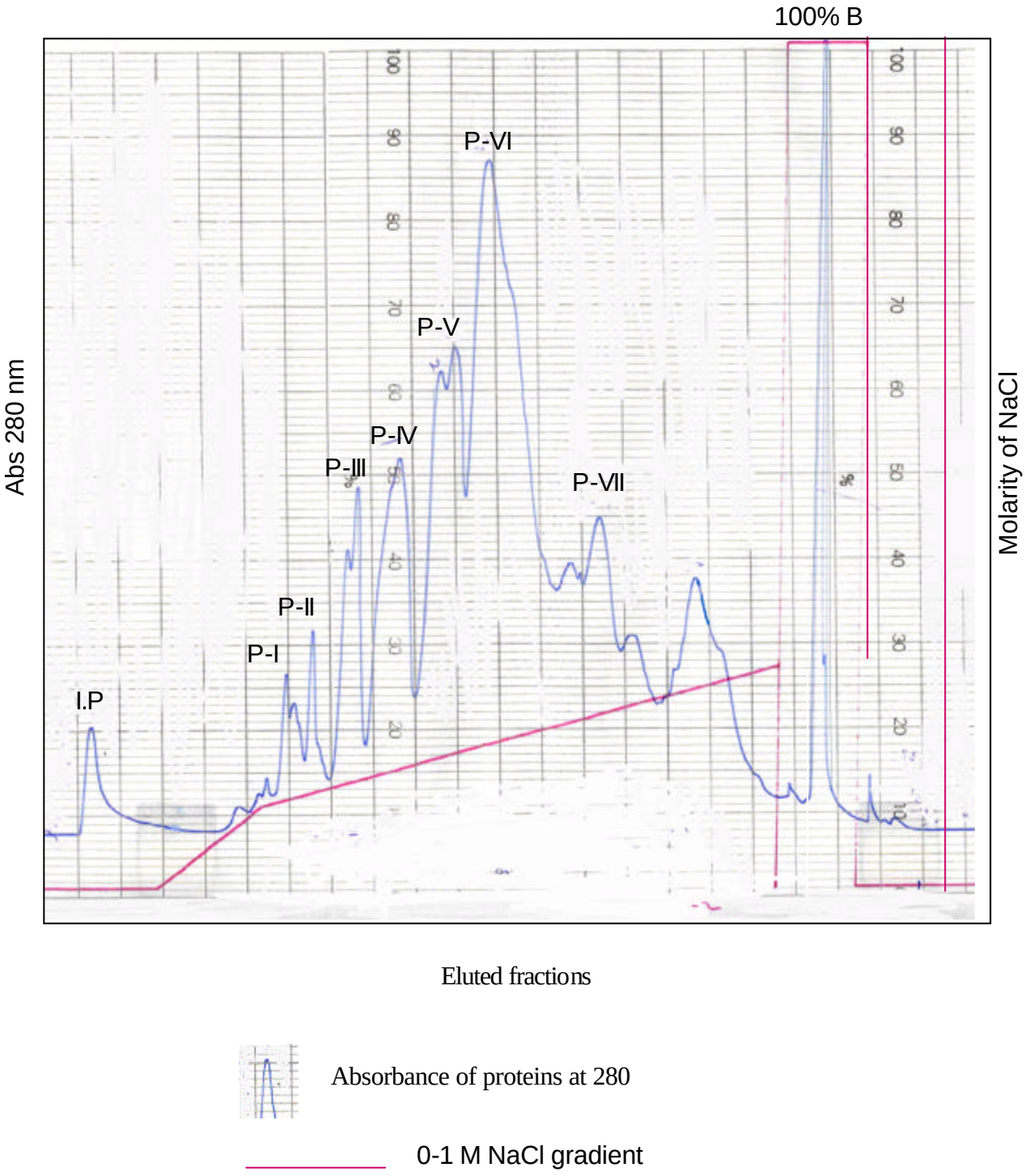


Fig. 4. 5. Elution profile of the BBI during first injection on FPLC using MonoQ column. In the diagram I.P indicates the injection peak, the isoinhibitors eluted in sequence were labeled P-I through P-VII, 100% B is 1M NaCl.

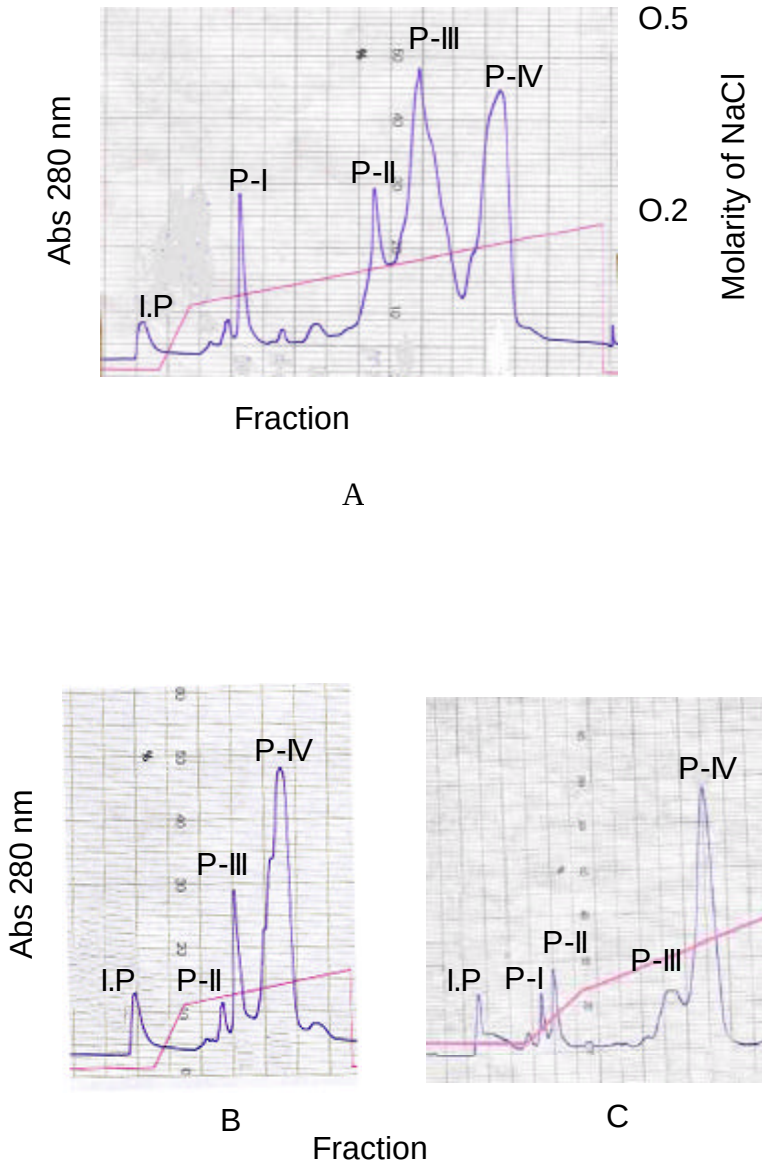
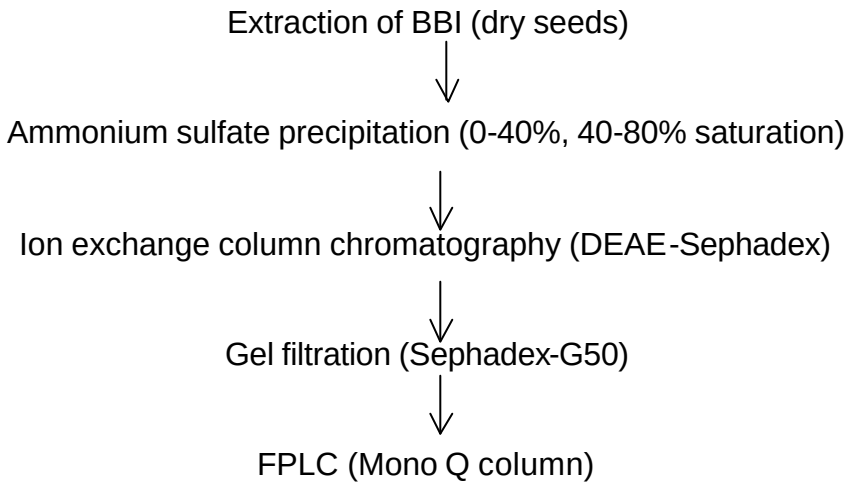


Fig. 4. 6. Purification of the isoinhibitor, P-IV by reloading 3 times, (A), (B), (C), in succession. It can be noticed that the P-IV peak is progressively getting separated when going from A-C.

Table. 4. 1. Details of the purification of BBI using 1 kg of seeds.

Purification step	**Total Volume in ml at the end of the step**	**Total activity in units**		**Protein concen-tration in mg**	**Specific activity in units / mg**	
		Trypsin	Chymo-trypsin		Trypsin	Chymo-trypsin
0-40% Ammonium sulfate	3,000.0	11,500	3,550	43,400	0.265	0.082
40-80% Ammonium sulfate followed by dialysis	320.0	10,400	3,150	37,950	0.274	0.083
Acid precipitation	323.0	9,150	2,760	22,500	0.407	0.123
DEAE-Sephadex	180.0	7,750	1,800	920	8.424	1.957
Sephadex-G50	70.0	6,300	1200	310	20.323	3.871
FPLC						
Isoinhibitor Label						
1 PI	3.5	790.0	52.0	1.3	607.0	40.0
2 PII	4.0	620.0	85.0	1.5	413.0	57.0
3 PIII	6.0	760.0	150.0	2.2	345.0	68.0
4 PIV	4.5	1080.0	130.0	2.5	432.0	52.0
5 PV	5.5	745.0	185.0	2.4	310.0	77.0
6 PVI	4.0	880.0	120.0	2.7	326.0	44.0

4. 2. 4. The complete purification flow chart of BBI



4. 2. 5. Characterization of Isoinhibitors

The six purified isoinhibitors showed differences between each other in their specific activity towards trypsin and chymotrypsin (table. 4. 1). The specific activity w. r. t. trypsin improved from 0.26 units/mg in the 0-40% ammonium sulfate precipitation step and varied between 310-607 units/mg after FPLC step for the six isoinhibitors. The corresponding values for chymotrypsin inhibition improved from 0.08 to in the range of 40-77 units/mg for various isoinhibitors. Thus the PIs have undergone about 1190-2330 fold purification w. r. t. trypsin and 500-960 fold purification w. r. t. chymotrypsin. After applying the FPLC purification step, the specific activity of PIs markedly improved from the previous gel filtration step value of 20 units/mg to 310-607 units/mg w. r. t. trypsin and improved from 3.8 to the range 40-77 units/mg in the case of chymotrypsin. The overall yield of the purified PIs (adding together all) was around 12 mg from 1kg of seeds. Because of the presence of minor isoinihibitors and due to multiple purification steps there was extra loss during final purification steps.

The molecular weights of the isoinhibitors were determined using three different techniques described in Chapter 2. The molecular weight of isoinhibitor P-IV determined by gel filtration technique using standard molecular weight markers kit was 14.8 kDa (Fig 4. 7). The molecular weight of the same

isoinhibitor determined by SDS-PAGE method was 15.2 kDa (Figs. 4. 8. A-B). However, this estimate was not in agreement with that generally observed for Bowman-Birk type or Kunitz type PIs (Tan-Wilson & Wilson, 1986). But the isotopically averaged masses measured by electro-spray mass spectrometry (ES-MS) showed that they were closer to 8 kDa (Fig. 4. 9). This molecular weight is the typical of classical double headed dicot Bowman-Birk type protease inhibitors (Birk, 1985). The individual molecular weights of isoinhibitors from mass spectrometry is given in table. 4. 2. However, the molecular weight of the BBI P-IV estimated by gel filtration method is double this value. Thus gel filtration experiment suggests the existence of BBI as dimer in solution. The molecular weight determined by SDS-PAGE method correspond to undissociated diimer band. Initially this deviation was thought to have resulted from abnormal migration, as previously observed in other cases (Wu & Whitaker, 1990; Gatehouse *et al.*, 1980). It may be noted that the dimer band remained even after boiling the sample in the presence of ß-mercaptoethanol (ß-me) and sodium dodecyl sulfate (SDS) (Fig. 4. 8. A), supporting the possibility of a very stable dimer.

Table. 4. 2. The % of NaCl (B) at which the isoinhibitors eluted in MonoQ column, the isoelectric pH values estimated using IEF unit and the molecular weights determined by mass spectrum method for the six isoinhibitors.

Isoinhibitor	**%NaCl (B)**	**Isoelectric pH values**	**Mol. Wt. in Dalton**
PI	12.5	4.45	7761
PII	16.0	4.68	8050
PIII	17.5	4.95	7890
PIV	19.0	5.16	8180
PV	20.2	5.38	8005
PVI	21.5	5.71	8418

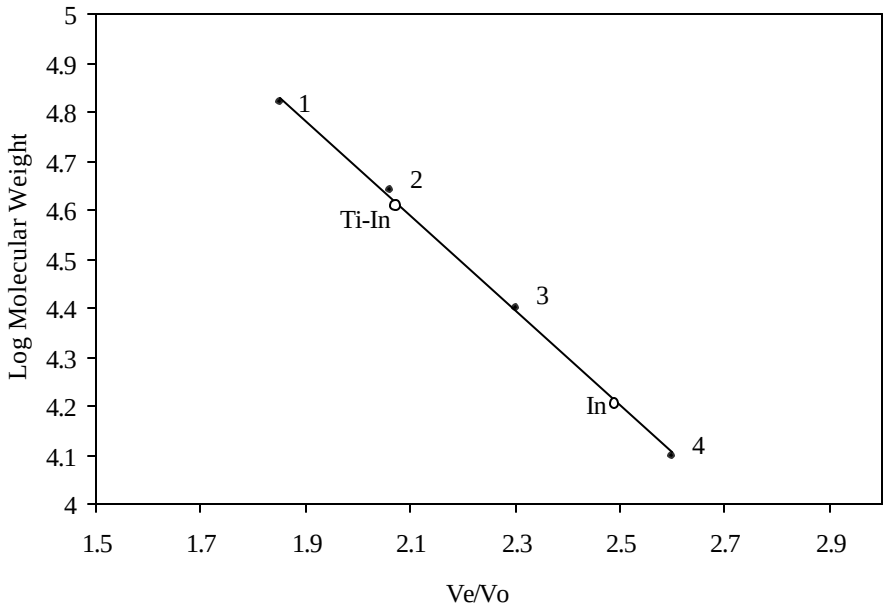
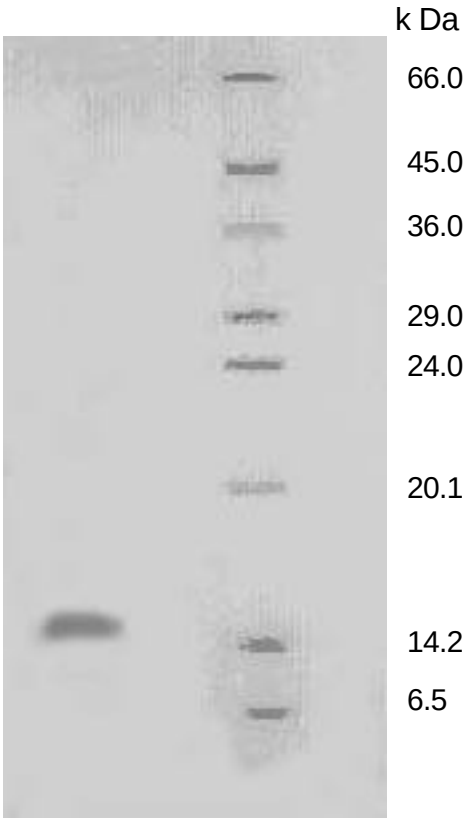


Fig. 4. 7. Molecular weight determination of the purified isoinhibitor P-IV [In] and its trypsin complex [Ti-In] by gel filtration.

The molecular weight markers used for calibration are (1) Bovine serum albumin, 66 kDa, (2) Ovalbumin, 43 kDa (3) Chymotrypsinogen, 25 kDa (4) Ribonuclease, 12.4 kDa.

The molecular weight estimated for BBI is 14.8 kDa and that for BBI-trypsin complex is 38.5 kDa.



A

Fig. 4. 8. A Molecular weight determination of the purified isoinhibitor P-IV by SDS-PAGE.

A) The BBI (Lane-1) along with the molecular weight markers (Lane-2) (50 µg) was electrophoresed on a 15.0% (w/v) polyacrylamide gel in the presence of ß-me and SDS.

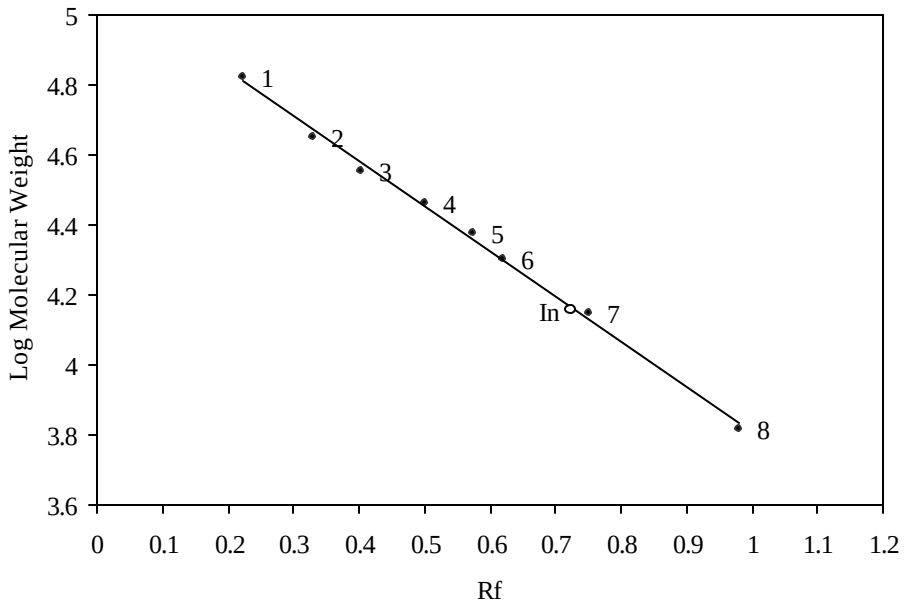


Fig. 4. 8. B. Molecular weight determination of the purified isoinhibitor P-IV [In] by SDS-PAGE (continued from previous page).

B) Relative mobilities (R_f) of the reference proteins were plotted against the log of molecular weights. Reference proteins used were (1) Bovine serum albumin, 66 kDa (2) Egg albumin, 45 kDa (3) Glyceraldehyde 3-phosphate dehydrogenase 36, kDa (4) Carbonic anhydrase, 29 kDa (5) Trypsinogen 24 kDa (6) Trypsin inhibitor 20.1 kDa (7) a-lactalbumin 14.2 kDa (8) aprotinin 6.5 kDa.

The molecular weight estimated for BBI is 15.2 kDa.

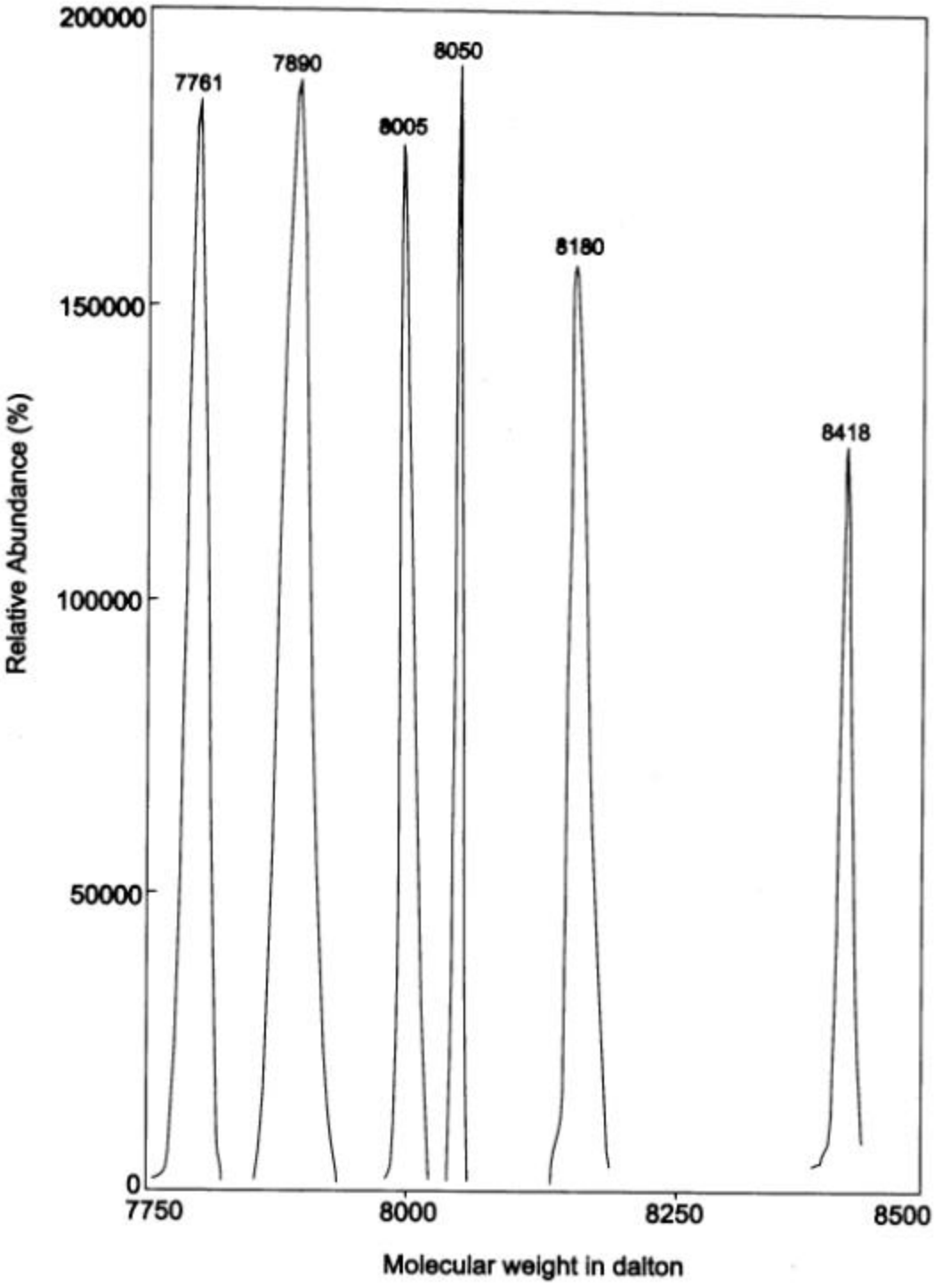


Fig. 4. 9. The spectra showing the accurate molecular weight of different isoinhibitors determined by ES-MS method. The m/z values were converted here into molecular weights in Dalton. The spectrum is recorded separately for each isoinhibitor and pasted together.

4. 2. 6. N-terminal Sequence

The knowledge of the amino acid sequence of BBI will help in sequence comparison with other protease inhibitors. The complete sequencing of one major isoinhibitor (P-IV), for which the three-dimensional structure reported in this thesis is in progress. The N-terminal sequence is presently available. The sequence is, E-P-S-E-S-S-E-P-X-X-X-S-, X-is unknown residue. This sequence is similar to the N-terminal sequence derived from our electron density map (Fig. 5. 4) and also comparable with N-terminal sequences of many BBI type PIs (Fig. 1. 4).

4. 2. 7. BBI-trypsin complex preparation

Purified *Vigna* BBI (isoinhibitor P-IV) and bovine trypsin were mixed at 1:1 molar ratio, the reaction was allowed to take place for 15 min at 37$^{?}$ C, the complex was isolated using the gel filtration column. The same Sephadex G-100 column calibrated to estimate the molecular weight of BBI was used for isolation and identification of the complex. Fig. 4. 10, shows the elution profile of the complex. The peak with highest O. D. was assumed to be that of complex, since this peak could be assigned neither to BBI nor to trypsin. Correspondingly reduction or disappearance of the protease or BBI or both peaks is also an indication of complex formation. Thus the formation of *Vigna* BBI-trypsin complex was reasonably confirmed. The purity and stability of the prepared complex was checked by a repeat run using the same column. The peak eluted corresponded to the elution volume of the assumed complex (Fig. 4. 10) and the estimated molecular weight, 38.5 kDa. A plot of log molecular weight v/s Ve/Vo (Fig. 4. 7) showed that the size of the purified complex is close to the sum of the molecular weights of BBI (dimer) and trypsin (38.5 kDa), indicating that the trypsin and BBI dimer have combined in 1:1 molar ratio. Hanging drop vapor diffusion method was used for screening crystallization conditions of the complex. Several trials did not yield any crystals. Crystallization of the complex between *Vigna* BBI-chymotrypsin complex has also been tried in a similar manner, but met with no success.

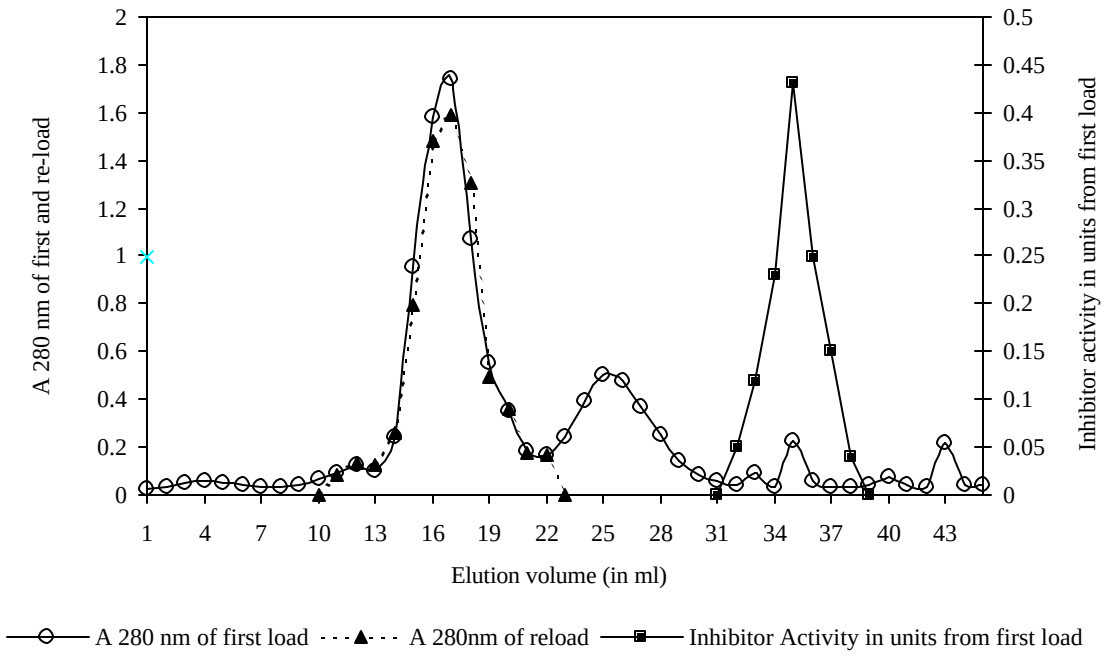


Fig. 4. 10. Elution profile of the BBI-trypsin complex in Sephadex G-100 gel filtration column.

4. 3. DISCUSSION AND CONCLUSIONS

Nature has used diverse approaches to achieve proteinase inhibition. Protease inhibitors are one of the most important tools of nature for regulating the proteolytic activity of their target proteases. They are synthesized in biological systems and they play a critical role in controlling a number of diverse physiological functions. Owing to their interesting and unique properties as described in the first chapter research on these small proteins is intriguing.

In the current study, few Bowman-Birk type protease inhibitors from the seeds of *Vigna unguiculata* have been isolated and characterized. Results shown in tables 4. 1. and 4. 2 on BBI lead to the following conclusions. All the six isoinhibitors showed inhibition activity towards both trypsin and chymotrypsin and hence all of them are 'double headed' PIs. Because of the

presence of multiple PIs the yield of each of the six isoinhibitors was poor. These isoinhibitors are closely related among themselves and to the general class of low molecular weight double headed Bowman-Birk PIs from legumes.

The isoinhibitors showed single band during SDS-PAGE (Fig. 4. 2. B), and a single PI activity peak during gel filtration (Fig. 4. 3). Indication from SDS-PAGE and gel filtration experiments is that the isoforms have similar molecular weights and exist as dimers in the solution phase. It is commonly observed that BBIs in general and those from black-eyed pea in particular, show tendency to form oligomers in solution (Gennis & Cantor, 1976; Kakade *et al*., 1970). Similarly the resistance of the dimers of cowpea trypsin/chymotrypsin PI to denaturation has been documented (Gatehouse *et al*., 1980). In its crystal structure also, described in the following chapter, the asymmetric unit in the unit cell consists of two tightly associated BBI molecules that form a compact dimer. The exposed hydrophobic patches on the surface, a typical structural feature of BBIs, is thought to be responsible for the commonly observed phenomenon of self association in BBIs (Voss *et al*., 1996).

One of the reasons for undertaking the separation and characterization of the isoinhibitors is that they could be characterized based on their physico-chemical properties. Indeed the six isoforms could be distinguished by their variation in specific activity or affinity towards trypsin and chymotrypsin, in terms of their isoelectric points and molecular weights determined by mass spectrum method (table 4. 1 and 4. 2). These closely related isoproteins differ by charge differences since multiple bands were seen on isoelectric focussing gels (Fig 4. 4. A) and thus were separated on anion exchanger MonoQ column (Figs 4. 5 and 4. 6. A-C). It is possible that these isoforms differ by a few addition/deletion of amino acids at their termini and also with respect to their amino acid sequence. The difference in the observed mass between the isoforms, as shown by ES-MS (Fig. 4. 9), suggests post-translational modifications. The presence of multiple isoinhibitors in plants was attributed either to expression of multi gene family and/or to post-translational modification (Quillien *et al*., 1997). Establishing the complete amino acid sequence of individual isoinhibitors could only clearly reveal relationship among these isoforms. Physiological role and

the reason for the presence of sevaral isoinhibitors are not very clear. However, it is now widely believed that they are defense proteins. It has been postulated that many PIs are products of multi-gene families and it is common to find several isoinhibitor species exhibiting different specifities towards proteases (Laskowski *et al*., 1988). In a co-evolving system, plants and insects evolve with new forms of PIs and proteases to counter each other's defense mechanisms (Hasulakar *et al*., 1999). Therefore It is possible that the plants evolved numerous isoinhibitors against diverse digestive enzymes of various parasites in nature (Bown *et al*., 1997; Felton, 1996; Ryan, 1990). Plants also seem to have simplified the process by producing multi-domain PIs (Jongsma & Bolter, 1997).

The N-terminal sequence of the P-IV has been determined presently, the complete sequencing is in progress. This sequence compares with the N-terminal sequence of classical BBIs. Considerable information about the structure-function relationship of the BBIs can be obtained by determining the complete sequence. The protease-BBI complex was prepared with both trypsin and chymotrypsin enzymes. These complexes were found to be very stable, as upon reloading on gel filtration column they were eluted as a single peak at a position corresponding to its preparative step. The characterization of the complexes through determining their molecular weights showed that one molecule of protease was bound to a dimer of BBI, indicating the high stability of the BBI dimer. The dimer BBI combining with a protease molecule has been observed in the case of black-eyed pea BBIs (Gennis & Cantor, 1976). Unfortunately, both the trypsin and chymotrypsin complexes did not yield crystals.

Vigna BBIs can be subjected to further biochemical and structural studies. They could be exploited for anticarcinogenic, antiviral and antifungal properties. They are models for designing peptide drugs. Due to comparatively smaller size, BBIs are ideal candidates for elucidating the principles of protein folding, protein-protein interactions and protein evolution.

CHAPTER-5

CRYSTALLIZATION AND STRUCTURAL CHARACTERIZATION OF A DOUBLE HEADED ISOINHIBITOR FROM *VIGNA UNGUICULATA*

5. 1. INTRODUCTION

Presently, sequences of more than 100 Bowman-Birk type protease inhibitors from plant sources are available at the PLANT-PIs database accessible at http://bighost.area.ba.cnr.it/ PLANT-PIs (De Leo *et al*., 2002). The three-dimensional structures of BBIs from only 7 plant seeds (5 structures in native and 4 in complex form) have been reported to date (table. 1. 2). Only crystallization and no structures are reported for some BBIs long back. These include the BBI from horse gram seeds (Prakash *et al*., 1994) and a ternary complex of soybean BBI with trypsin and chymotrypsin (Gaier *et al*., 1981). Homology based model structures have been reported for the complex of black-eyed pea BBI with chymotrypsin (de Freitas *et al*., 1997) and a ternary complex of pea BBI with trypsin and chymotrypsin (de la Sierra *et al*., 1999). In this chapter author has described the three-dimensional structure of a BBI from cowpea. Also a structure of complex this inhibitor with bovine trypsin simulated through modeling is presented.

5. 2. RESULTS AND DISCUSSION

5. 2. 1. Crystallization and Crystal characterization

Crystallization trials on the sample prior to FPLC purification were unsuccessful. The BBI could be crystallized only after separating it from the isoinhibitors and purifying to a high level of homogeneity. Except for one, attempts to crystallize few other FPLC purified isoinhibitors did not yield any crystals and they precipitated in the hanging drop. The isoinhibitor P-IV, the fourth isoinhibitor peak from the start of the FPLC gradient (Figs. 4. 6. A-C), gave several tiny crystals of maximum size 0.2 mm (Fig. 5. 1). Attempts to grow big size crystals were not successful. P-IV was extensively dialyzed against 10 mM citrate-phosphate buffer and concentrated at 50 mg/ml. Crystals were grown against well solutions containing 0.1 M citrate-phosphate buffer at pH 4.0 and 30% saturated ammonium sulfate as precipitant (Rao *et al*., 1999b).

The X-ray intensity data on BBI crystals were collected up to 2.5 Å resolution under cryogenic conditions. The data could be processed using DENZO and SCALEPACK programs. The unit cell determined was monoclinic

with cell dimensions a=32.4, b=61.8, c=32.9 Å, ß=114.5$^{?}$ and space group $P2_1$. Based on the molecular weight of the BBI obtained by mass spectrum method (8,180 Dalton), and assuming that the asymmetric unit is a dimer, the Mathew's number calculated was 1.95 $Å^3$/D which correspond to 37% solvent content in the crystal (Rao *et al.*, 1999b). Table. 5. 1. gives the data collection statistics.

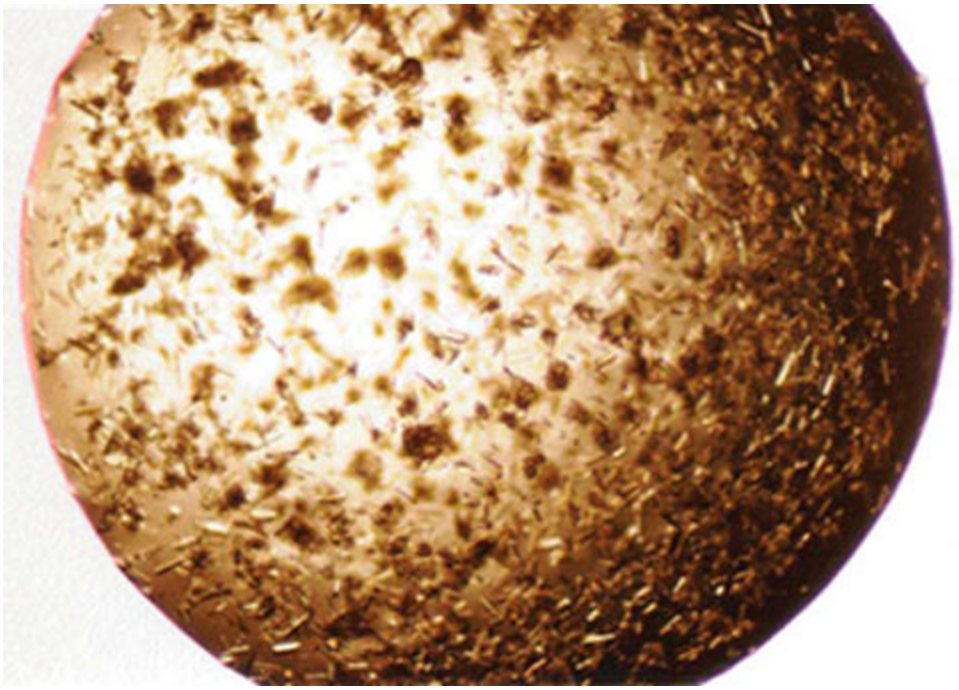


Fig. 5. 1. Monoclinic crystals of V*igna* BBI. Maximum dimension of crystals was 0.2 mm.

5. 2. 2. Structure solution

The amino acid sequence of a BBI from *Vigna* seeds was available (Morhy & Ventura, 1987). This reported sequence had 83 amino acids, corresponded to 9278 Da size, which is slightly different from the estimated molecular weight (8180 Da) by ES-MS of the V*igna* isoinhibitor P-IV. It was assumed that the sequence of the presently crystallized isoinhibitor P-IV was close to the above reported sequence.

Table. 5. 1. Data collection statistics (Values in the parenthesis refer to data in the highest resolution shell, 2.59-2.5 Å).

Parameter	**Value**
Resolution range of the data.	20.0-2.5 Å
R_{sym}#	0.083
I/SsI	13.9(5.8)
No. of measurements	12514
No. of unique reflections	4072
Completeness (%)	99.0(94.7)

R_{sym}= ? $_{hkl}$? $_{i}$|I_i(hkl)-<I(hkl)>|/ ? $_{hkl}$? $_{i}$ I_i(hkl), where I_i(hkl) is the intensity of i^{th} measurement of hkl reflection and <I(hkl)> is its mean.

When looking for a model for molecular replacement method we compared the reported sequence with those for which three-dimensional structures were available. Through alignment search we found out that 38 residues out of 57 were similar (70% homology) between *Vigna* (reported sequence) and tracey bean BBI. The high homology seen between the two sequences indicated that the two structures could be very similar. Thus the three-dimensional structure of tracey bean BBI - PDB entry 1PI2 - (Chen *et al*., 1992) was used as the search model for molecular replacement.

For molecular replacement calculations, the amino acid Ala was replaced in those positions where the residues in the reported *Vigna* BBI sequence differed from the tracey bean sequence. The calculations were performed using AMoRe program. The resolution of diffraction data used in the calculation of rotation function, translation function and rigid body fit are shown below.

Rotation function solution (S) - 10.0-3.0 Å
Translation function solution (T) - 10.0-2.5 Å
Rigid body fit (R) - 17.0- 2.5 Å

Solution of rotation and translation function calculations is provided in table. 5. 2. Two solutions emerged for rotation (S1, S2) and translation function calculations (S1-T, S2-T). The combined solution corresponding to dimer molecule was refined in rigid body fit (S1-R, S2-R).

Table. 5. 2. Results of molecular replacement calculations. Where a, ß, ? are Eulerian angles, Tx, Ty, Tz translations along X, Y, Z, axes. Cc is the correlation coefficient, Rfactor is the crystallographic residual, and they are estimated for two solutions together in the case of S2-R.

Solution	a(?)	ß(?)	?(?)	Tx	Ty	Tz	Cc(%)	R(%)
S1	220.93	97.54	57.34	-	-	-	15.7	-
S2	160.94	100.85	32.96	-	-	-	15.3	-
S1-T	-	-	-	0.4074	0.0000	0.2589	34.4	55.0
S2-T	-	-	-	0.9669	0.3996	0.5185	46.0	49.9
S1-R	219.21	100.35	56.41	0.4042	-0.0020	0.2606	-	-
S2-R	161.10	101.35	33.47	0.9668	0.3995	0.5214	53.5	48.4

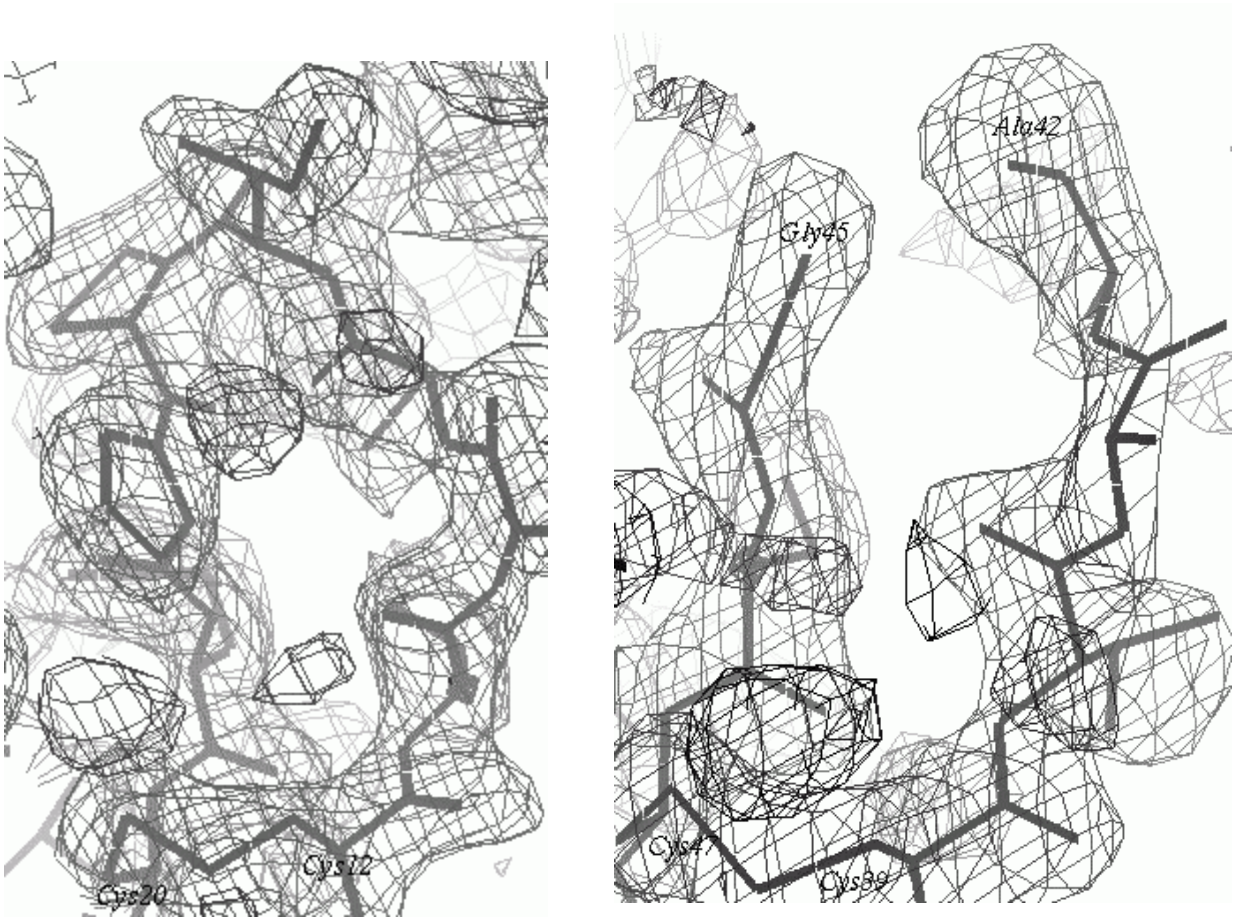
5. 2. 3. Refinement

The rotation matrix and translation vectors obtained from molecular replacement calculations were used to place the model in the correct orientation and position in the unit cell of the V*igna* BBI. Out of the total 3763 reflections used in refinement 177 were kept for Rfree calculations. Initially rigid body refinement was carried out in REFMAC using data in the resolution range 15-3.0 Å. This was followed by several cycles of positional refinement using the data in the same resolution. Electron density maps were visualized at this stage. Since the molecule was a dimer in the asymmetric unit, molecular

averaging was tried making use of the presence of non-crystallographic symmetry (NCS). The averaged electron density map looked much better and the initial model could be improved. Side chains were progressively fitted as the map quality improved and the positions where the side chain density was ambiguous alanines were retained. After checking through the whole protein chain using QUANTA and getting an improved fit of the residues for the observed electron density, the structure was further refined using the data between 8.0-2.5 Å along with individual B factors. Solvent molecules were added progressively to the structure using X-SOLVATE module of QUANTA. Those water molecules whose B-factors reached the limit of 100 in a few test cycles of refinement were rejected. Like this, a total of only 58 water molecules could be added, further attempts at addition of water molecules drastically increased Rfree despite a reduction in Rfactor. At the end of the refinement most of the residues were identified unambiguously except for those belonging to chymotrypsin loop region and termini. Omit maps were calculated for poorly defined regions of electron density and the residues were rebuilt.

The missing electron density for some N-terminal and C-terminal residues in both the subunits may be due to disorder. Also density for the side chains of residues14 and 41-44 of subunit A and residues 31, 41 and 43 of subunit B was not clear and hence these residues were treated as alanines. In fact, electron density was poorly defined in the region between the residues 42 and 45 in the chymotrypsin loop of subunit A (Fig. 5. 2. B). Since chymotrypsin inhibition was not affected, this could be due to disorder rather than cleavage.

In the final stages, after every cycle of refinement the stereochemistry and the geometry of the models were checked using PROCHECK program. The Ramachandran f , ?, plot (Ramachandran *et al*., 1968) showed that 84.4 % of the residues belong to most favored region and 12.5 % to allowed region of the map and the rest in generously allowed region (Fig. 5. 3). The refinement statistics is shown in table. 5. 3.



Cys20 Cys12 Cys47 Cys39

A B

Fig. 5. 2. The diagrams showing the electron density, A) near the trypsin binding loop region and B) near the chymotrypsin binding loop region, showing the break in the density between the residues Ala42-Gly45 in the A subunit of *Vigna*.

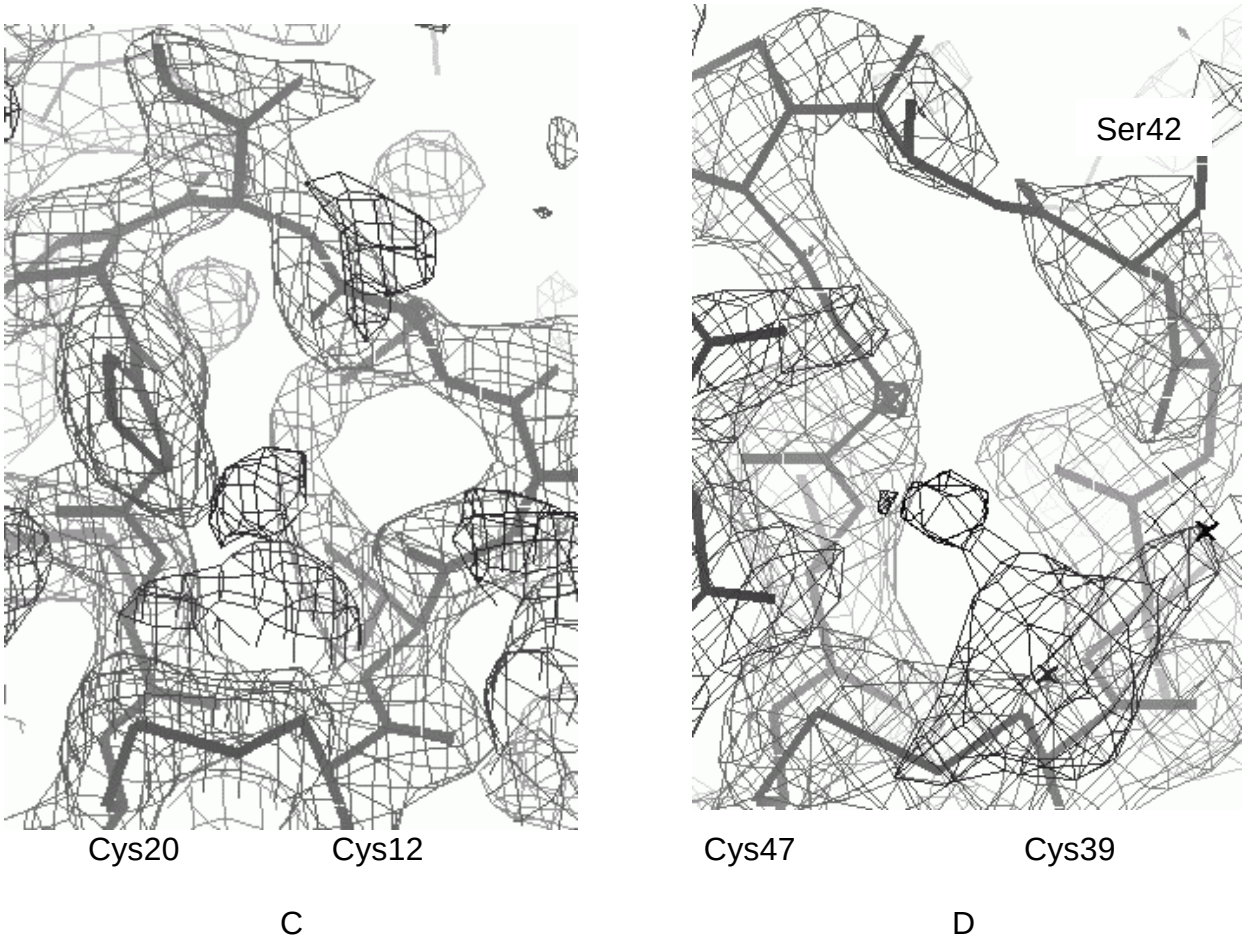


Fig. 5. 2. The diagrams showing the electron density, C) near the trypsin binding loop region and D) near the chymotrypsin binding loop region in the B subunit of *Vigna*. The density around Ser42-Ala43 is poorly defined.

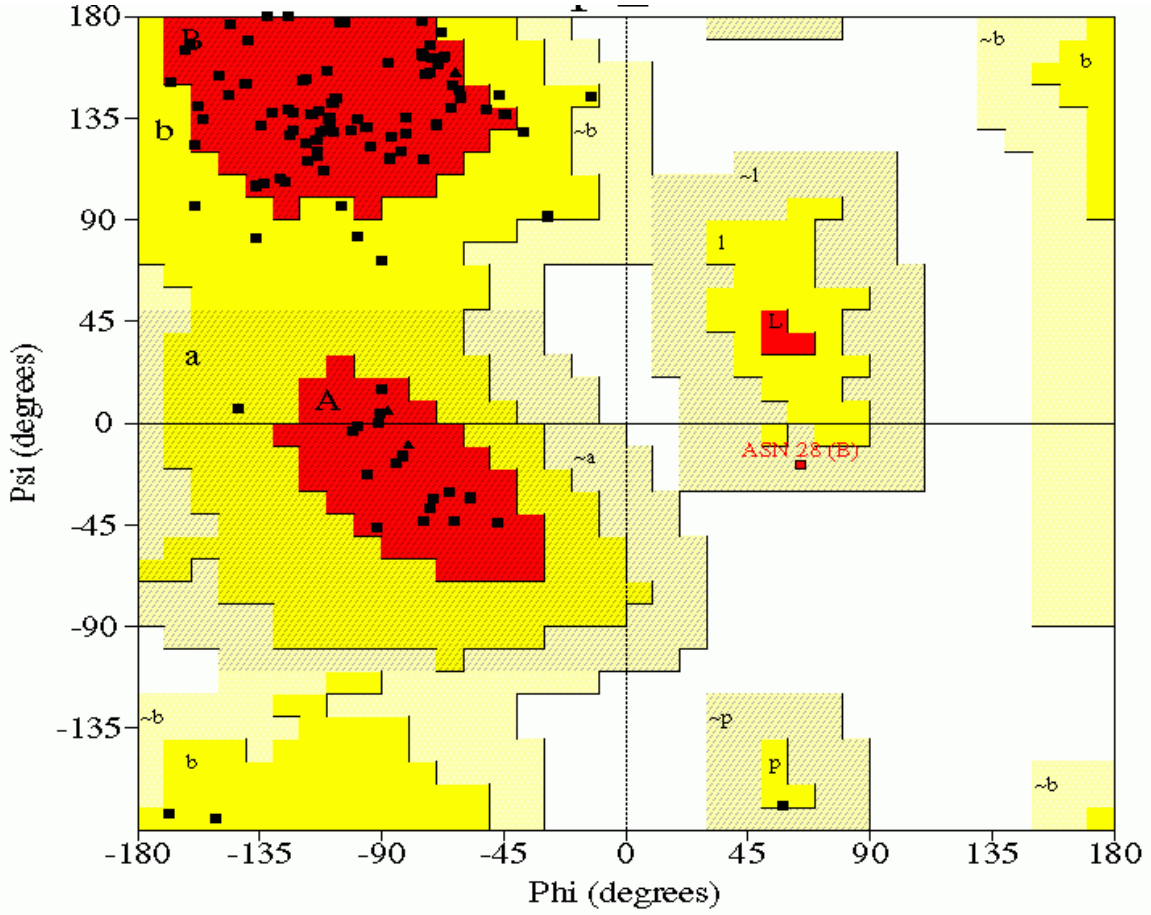


Fig. 5. 3. Ramachandran plot for BBI structure generated by PROCHECK. Triangles and squares represent glycine and non-glycine residues, respectively.

Table. 5. 3. Statistics of refinement.

Parameter	**Value**
R factor & R free (%)	19.1 & 25.8
Cc & Ccfree (%)	87.6 & 92.8
Number of reflections	3763
Number of reflections for Rfree	177
Average B-factor for all atoms	46.4
Number of protein atoms	798
Number of solvent molecules	58
RMS deviation from ideality	
a. bond lengths(Å)	0.01
b. bond angles (º)	3.1
Ramachandran plot	
Residues in: a) most favored region (%)	84.4%
b) additional allowed regions (%)	12.5%

5. 2. 4. The primary structure

The sequences of *Vigna* BBI (reported), tracey soybean BBI (search model) and the sequence derived from our electron density map (e. d.) were aligned as shown in Fig. 5. 4. The N-terminal sequence of *Vigna* isoinhibitor P-IV (Nt-IV), the sequencing of which is in progress, is also shown. The alignment shows that the sequences are highly homologous, especially the cysteine positions are conserved throughout the sequence. The three sequences share 65-80 % homology.

The trypsin binding site is in the first domain and the residue involved in binding is lysine in *Vigna* and arginine in tracey bean. The second domain carries the chymotrypsin binding site and the residue involved is phenylalanine in *Vigna* (reported sequence), whereas it is arginine in tracey bean BBI. The

conservation of residues in the trypsin and chymotrypsin binding loops is significantly higher.

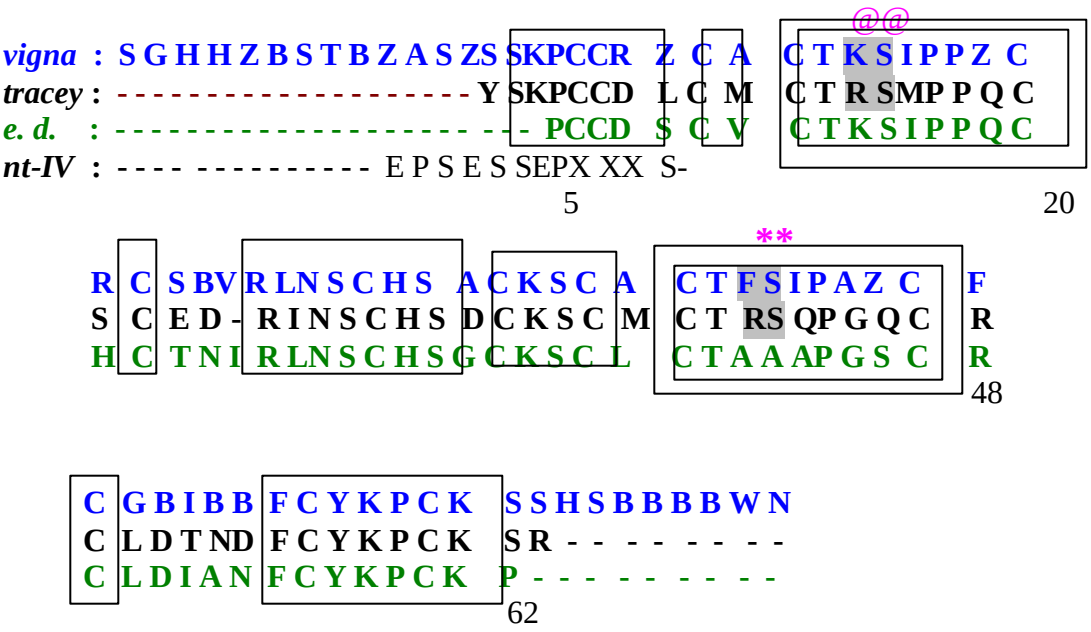


Fig. 5. 4. The sequence alignment of V*igna* BBI derived from electron density (also its N-terminal by protein sequencing) against the reported sequence and the sequence of tracey bean BBI. @@ indicates trypsin binding site, ** indicates chymotrypsin binding site, B for -Asp/Asn, Z for –Glu/Gln, X for –unknown residue, conserved residues are indicated by single boxes, the trypsin and chymptrypsin binding loop residues are indicated by double boxes. Alanines are placed near the chymotrypsin binding site where the electron density corresponding to the 3 (41-43) residues is poorly defined. The numbering of residues is by considering the first (N-terminal) cystine as 6th residue and the last (C-terminal) cystine as 60th residue(Fig. 5. 5).

5. 2. 5. The subunit structure

The secondary structure consists of ß-sheets and devoid of any a-helices. The monomer structure of BBI can be divided into two domains, an N and a C-domain. Each domain has a two stranded anti-parallel ß-sheet and an

associated short strand (Figs. 5. 5 and 5. 6. A-B). The monomer fold of this BBI, is the same "bow-tie motif", described in the case of other BBI structures (de la Sierra *et al*., 1999; Koepke *et al*., 2000). The ß-sheets consists of residues Ser9-Thr13, Gln19-Thr23 and Ile52-Asn54 in the N-domain, and Ser36-Ala41, Ser46-Leu50 and Ile25-Leu27 in the C-domain. The main chain hydrogen bonds between the ß-sheets, disulfide connectivity and the overall organization of the molecule is depicted in Fig. 5. 5. The two ß-sheets place the trypsin and chymotrypsin binding loops towards two opposite ends (Figs. 5. 5 and 5. 6. A-B). The subunit structure is stabilized by a number of hydrogen bonds. Each subunit contains more than 60 intra subunit hydrogen bonds (table. 5. 4). Most of them are of the same type in both the subunits and predominantly of main chain-main chain interactions. In addition, hydrogen bonds involving protein and solvent atoms are also present. These numerous hydrogen bonds (compared to the protein's smaller size) and the seven disulfide bridges provide high stability to the molecule. Table. 5. 4. lists the number of different types of hydrogen bonds in the structure.

Table. 5. 4. The number of hydrogen bonds classified according to their types: main chain-main chain (Mc-Mc), main chain- side chain (Mc-Sc), side chain-side chain (Sc-Sc) and protein-solvent (Prot-Sol) atoms, observed in the two subunits.

Interaction **Subunit**	**Mc-Mc**	**Mc-Sc**	**Sc-Sc**	**Prot-Sol**
A	40	29	7	94
B	41	23	7	93

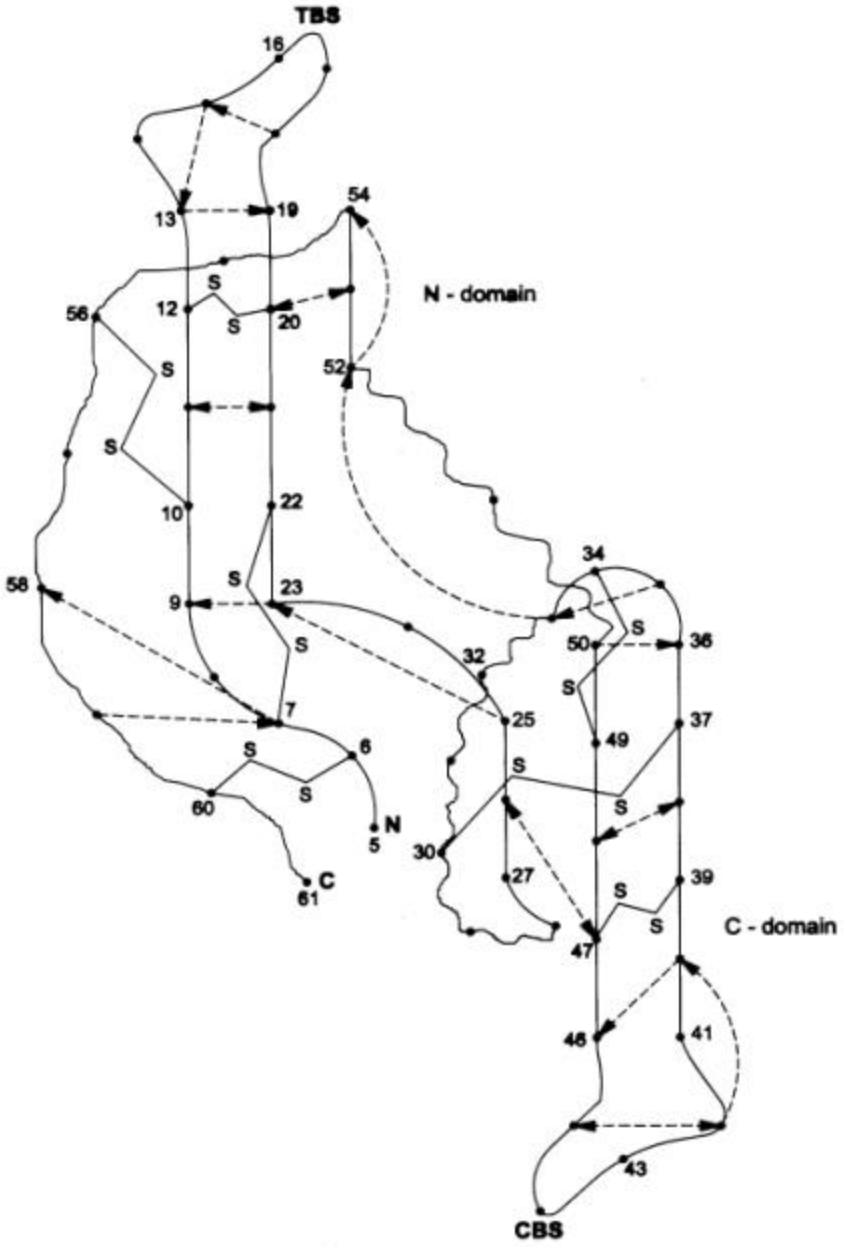


Fig. 5. 5. Diagrammatic representation of the structure of *Vigna* BBI. The dotted arrows indicate the direction (donating) of hydrogen bonds, disulfide bridges are shown using three lines (Z-type), ß-sheets are shown in straight lines, loops are shown by irregular lines. The upper left part of the diagram is N-domain and the lower right is C-domain. TBS/ CBS are trypsin/ chymotrypsin binding sites.

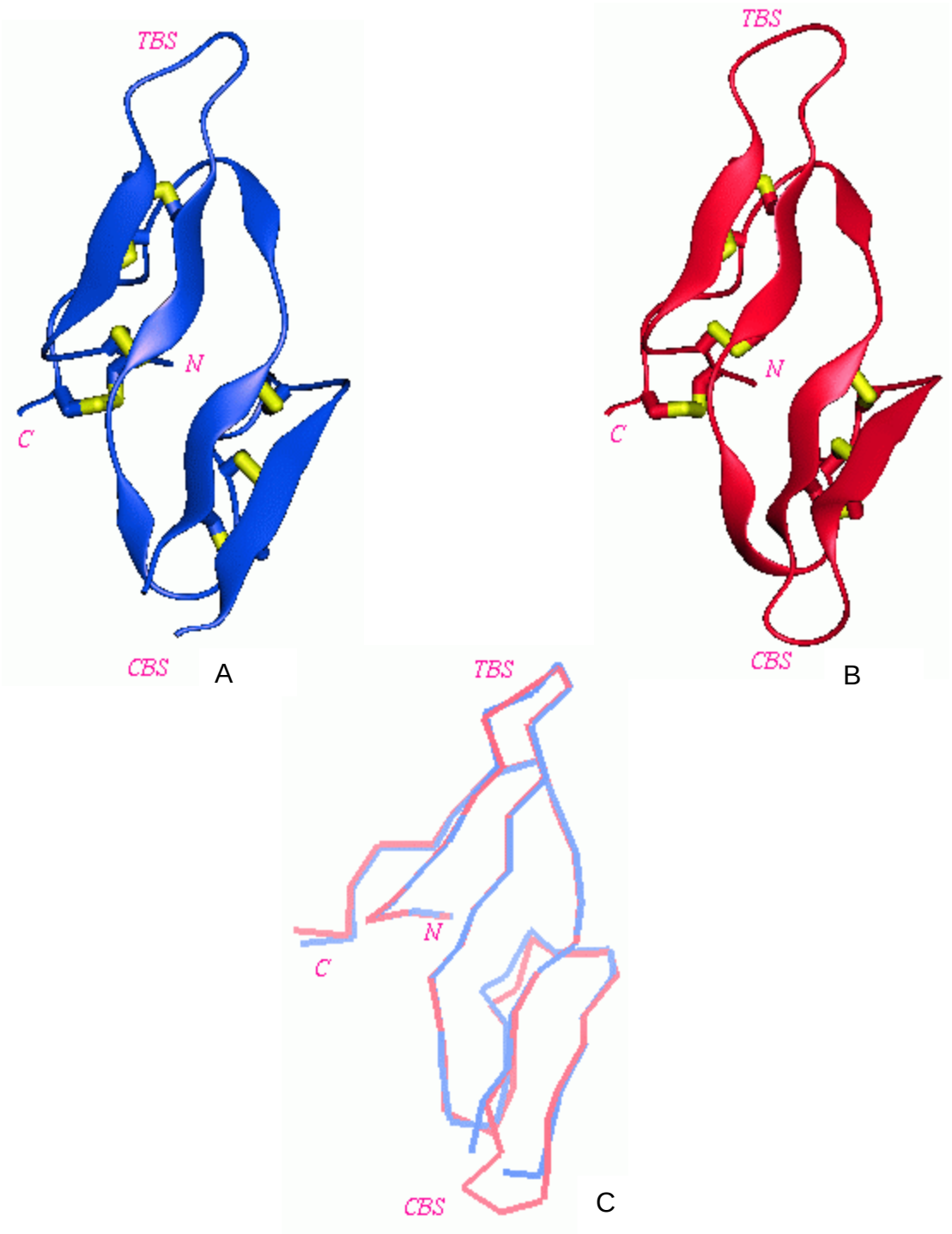


Fig. 5. 6. The subunit structures of V*igna* BBI, A) subunit A with chymotrypsin loop broken, B) subunit B, the disulfide bridges are shown in liquorice and yellow, C) superposition of Ca trace of the subunit B (red) on to that of subunit A (blue).

As already mentioned, only a couple of residues before the former and after the latter could be seen in the electron density map. The trypsin-binding domain (N domain) is formed from two peptide chain segments (residues Pro5-Thr23 and Asp51-Lys61) separated along the sequence, whereas the chymotrypsin-binding domain (C domain) is formed by a continuous segment of residues Asn24-Leu50. The interactions of the domains are through hydrogen bonds between main chain atoms and hydrophobic contacts between side chains. The individual domains are further stabilized by disulfide bridges. There are four such bridges in the first domain and three in the second. The disulfide bridges are Cys6-Cys60, Cys7-Cys22, Cys10-Cys56, Cys12-Cys20 of the N domain, and Cys30-Cys37, Cys34-Cys49, Cys39-Cys47 of the C domain (Fig. 5. 5). The seven disulfide bridges provide high rigidity to the structure and may explain the observation that proteolytic cleavage of the two reactive sites causes a conformational change of only loops while the rest of the structure remain unchanged (Thornton, 1981).

Accessibility of each residue and its relative accessibility were computed for the two inhibitor subunits using NACCESS program. 7 residues were found to have relative accessibility values less than 5% and constitute the core of the molecule. They are, Cys7, Cys20, Cys22, Cys34, Gly45, Cys47, Cys49. Majority of these residues are cystines and no hydrophobic ones in the core. In contrast the side chains Pro5, Pro17, Ile16, Phe55, Ile52 and Pro59 have been found exposed to solvent, constitute the hydrophobic patches on the surface of the protein. Another notable feature is the buried charged residues Gln19, Asn24, Asp51 and Asn54. Similarly the charged side chains of Asp8, Asn24, Arg48, Asp51 form an electrically charged cluster of amino acids at the inter-domain region. The interactions of the side chain atoms of three invariant aspartates and a arginine in both monomers, with main chain atoms, are listed below in table 5. 5 and shown in Fig. 5. 7. A. These interactions presumably stabilize the association and orientation of the two domains.

Table. 5. 5. List of the residues involved in interaction at the inter-domain region of V*igna* BBI.

Side chain	*Main chain atom*
Asp8	*N of Cys60 and Lys61*
Asn24	*N of Asp8 and Ser9*
Arg48	*O of Cys22*
Asp51	*N of Lys35 and Ser36*

The exposed hydrophobic patches on the surface of the protein is the unusual structural feature of the BBI. However, this feature is commonly observed among BBIs (Voss *et al*., 1996). The lyphophilic side chains on the surface of the protein are thought to be responsible for the commonly observed phenomenon of self-association in BBIs (Kakade *et al*., 1970). The structural patterns, which comprise of hydrophobic cores and charged residues on the surface, observed here are in contrast with that of typical soluble proteins. The structural stability of the BBI monomer does not seem to have arisen from the hydrophobicity as normally the case with most of the proteins. The structural stability appears to be contributed by combination of numerous hydrogen bonded contacts, electrostatic attraction at the inter-domain region and by the rigid framework of disulfide bridges and a good proline content

5. 2. 6. Dimer structure

The asymmetric unit in the crystal structure contains two BBI molecules - the subunits A and B - that closely interact to form a compact dimer (Figs. 5. 8. A and 5. 9. A). A pseudo 2-fold axis whose direction is different from the unit cell axes directions relates the two subunits (Fig. 5. 8. A). The approximate dimensions of the dimer are 40 X 53 X 30 Å. The two subunits superpose in Cα positions with r. m. s. deviation of 0.520 Å (Fig. 5. 6. C). Main contribution to the deviation comes from residues in the chymotrypsin insertion loop.

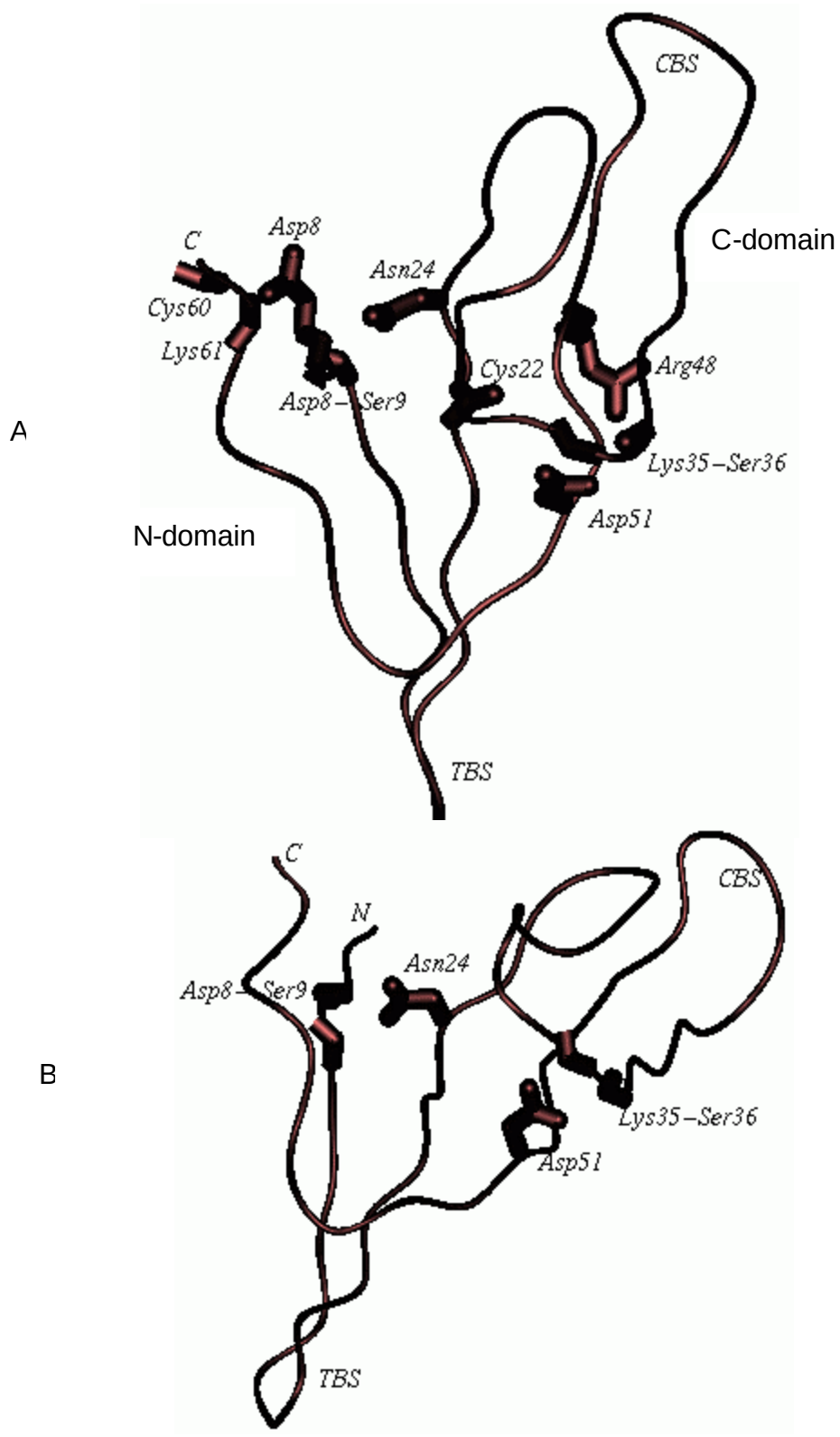


Fig. 5. 7. A) The interacting residues (listed in table. 5. 7) at the inter-domain region of Vigna monomer, B) residues, Asn 24 and Asp 51, interacting with main chain atoms of Asp8-Ser9 and Lys35-Ser36 at the extremities of the two loop regions.

This difference is also reflected in the B-factors of atoms in the chymotrypsin specific loops of the subunits. As mentioned earlier, there is a break in the electron density of chymotrypsin specific loop of subunit A .

The surface area buried due to the association of two subunits, calculated using the 'NACCESS' program, is 1095 $Å^2$, out of which 645 $Å^2$ is contributed by non-polar and rest by polar groups. In other words, 60% of the interface is hydrophobic and the remaining hydrophilic. Again, an interface area of 915 $Å^2$ (84%) is occupied by main chain atoms; the remaining by side chain atoms. The residues which lost surface accessibility on dimer formation are Pro5, Pro17, Gln19, Gly33, Lys35, Cys39, Asp51, Ile52, Ala53, Asn54, Phe55, Cys56, Tyr57. Thus, both the hydrophilic and hydrophobic residues are involved in dimer formation. The hydrophobic residues found exposed to solvent are now buried.

The core of the dimer interface involve extensive net work of hydrogen bonds involving the residues Gln19, Gly33, Lys35, Asn54, Tyr57 and Cys56 (Fig. 5. 9. A). The interface region has some 11 hydrogen bonds. Only two of them involve main chain atoms, four are between main chain and side chain atoms and five are between side chain and side chain atoms. The details of the inter subunit hydrogen bonds, classified according to their type, is given in table. 5. 6. No solvent molecule was found near the interface of the dimer.

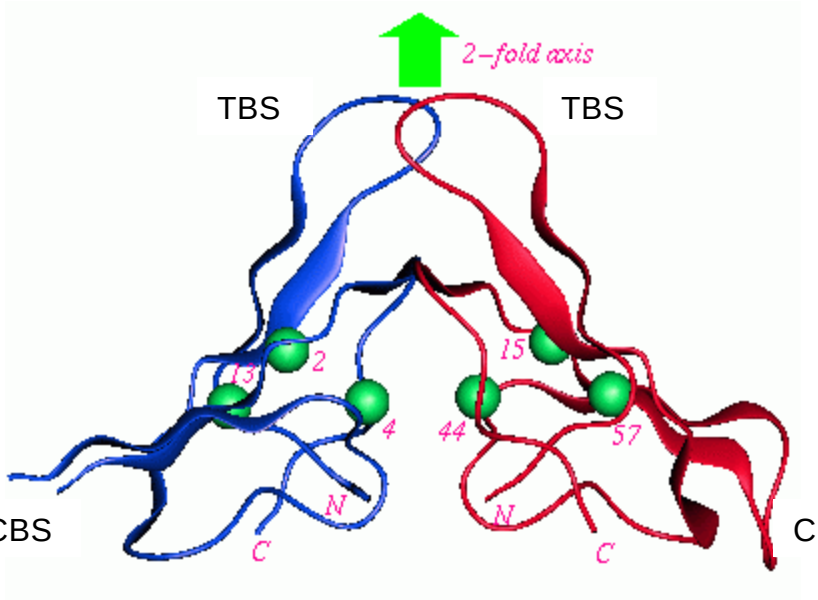
The strands 52-54 of both monomers come very close to each other forming an anti-parallel β-sheet (Fig. 5. 9. A), with amino and carbonyl of Ile52 and Asn54 forming two hydrogen bonds. The two monomers are symmetrically positioned about this β-sheet. Mainly the trypsin-specific loops of the two molecules are positioned on one side of the β-sheet plane and the rest of the molecules on the other side. Interactions on the trypsin loop side of the sheet plane are between the side chains of Gln19 and Asn54. On the opposite side of the β-sheet, side chains of Lys35 and Tyr57 interacts with the carbonyl groups of Cys56 and Gly33, respectively.

The dimeric form of a BBI was first observed in the crystal structure of pea BBI. The structure of pea BBI dimer did not match with that of *Vigna* BBI, however, their monomers superposed well, indicating the difference in their

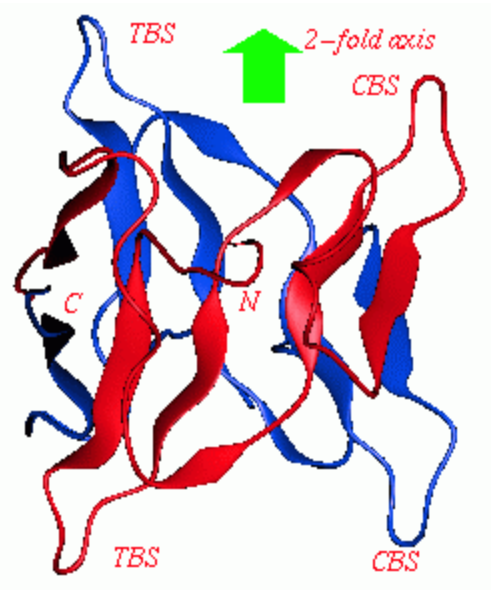
quaternary structures. In pea BBI the two monomers orient almost perpendicular to each other (Fig. 5. 8. B), whereas in *Vigna* they associate face to face (Fig. 5. 8. A). The residues involved in interactions at the interface of *Vigna* BBI and pea BBI are shown in Figs. 5. 9. A and B, respectively. Unlike the *Vigna* dimer where the trypsin binding sites come near to each other, the four binding sites of pea BBI can imagined to be placed at the four corners of a square (Fig. 5. 8. B). However, it is still unlikely that this dimer is a functional unit of the molecule. The model of the ternary complex of the inhibitor dimer/trypsin/chymotrypsin indicates that simultaneous binding to maximum only two proteases is possible (de la Sierra *et al*., 1999).

Table. 5. 6. Hydrogen bonds between the subunits A and B of *Vigna* classified according to the type of interaction.

Nature of the hydrogen bond	**Subunit A** Res ID atom	**Subunit B** Res ID atom	**Distance (Å)**
Main chain – Main chain	Ile 52 O Asn 54 N	Asn 54 N Ile 52 O	2.93 2.96
Main chain – Side chain	Gly 33 O Lys 35 NZ Cys 56 O Tyr 57 OH	Tyr 57 OH Cys 56 O Lys 35 NZ Gly 33 O	2.56 2.99 2.58 2.76
Side chain - Side chain	Gln 19 OE1 Gln 19 NE2 Asn 54 OD1 Asn 54 ND2 Asn 54 ND2	Asn 54 ND2 Asn 54 OD1 Gln 19 NE2 Gln 19 OE1 Gln 19 NE2	2.81 2.48 2.52 3.10 3.30



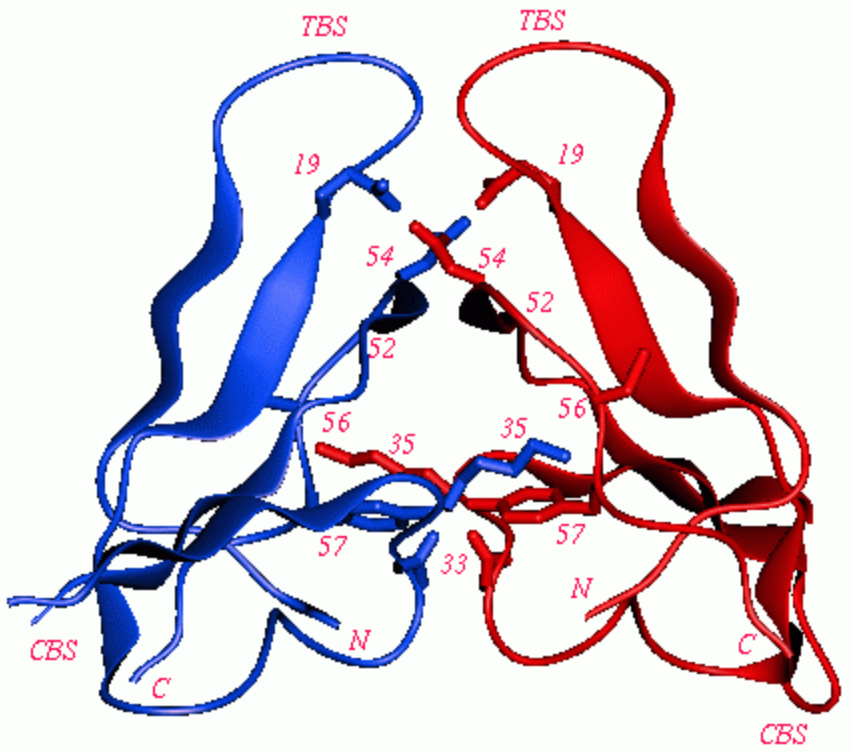
A



B

Fig. 5. 8. The dimer structures of A) V*igna* BBI showing the 2-fold NCS symmetry and the four binding sites, also shown are the six invariant water molecules (listed in table 5.7) in van der Walls spheres, B) pea BBI showing the arrangement of 4 binding sites at the corners of a square.

A



B

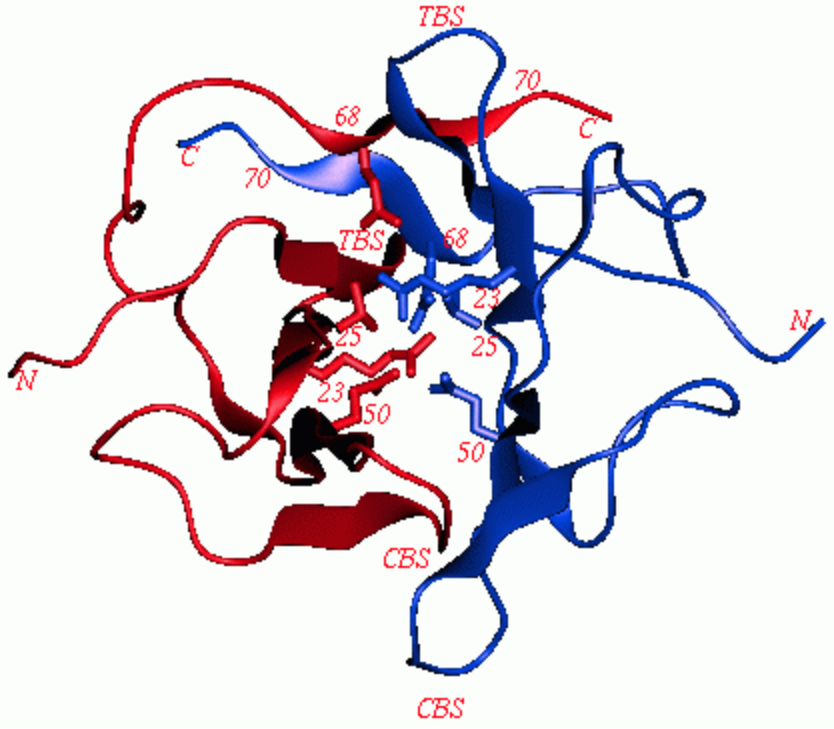


Fig. 5. 9. Another view of dimer structures, A) V*igna* BBI, B) pea BBI, showing the residues interacting at the inter-subunit region. The residues numbered here for V*igna* are listed in table 5. 6, and that for pea BBI are from de la Seirra *et al*., (1999).

The *Vigna* dimer crystallized in the space group $P2_1$. The molecular packing in the crystal viewed along the unique axis b and two perpendicular axes a and c are shown in Figs. 5. 10. A-C below.

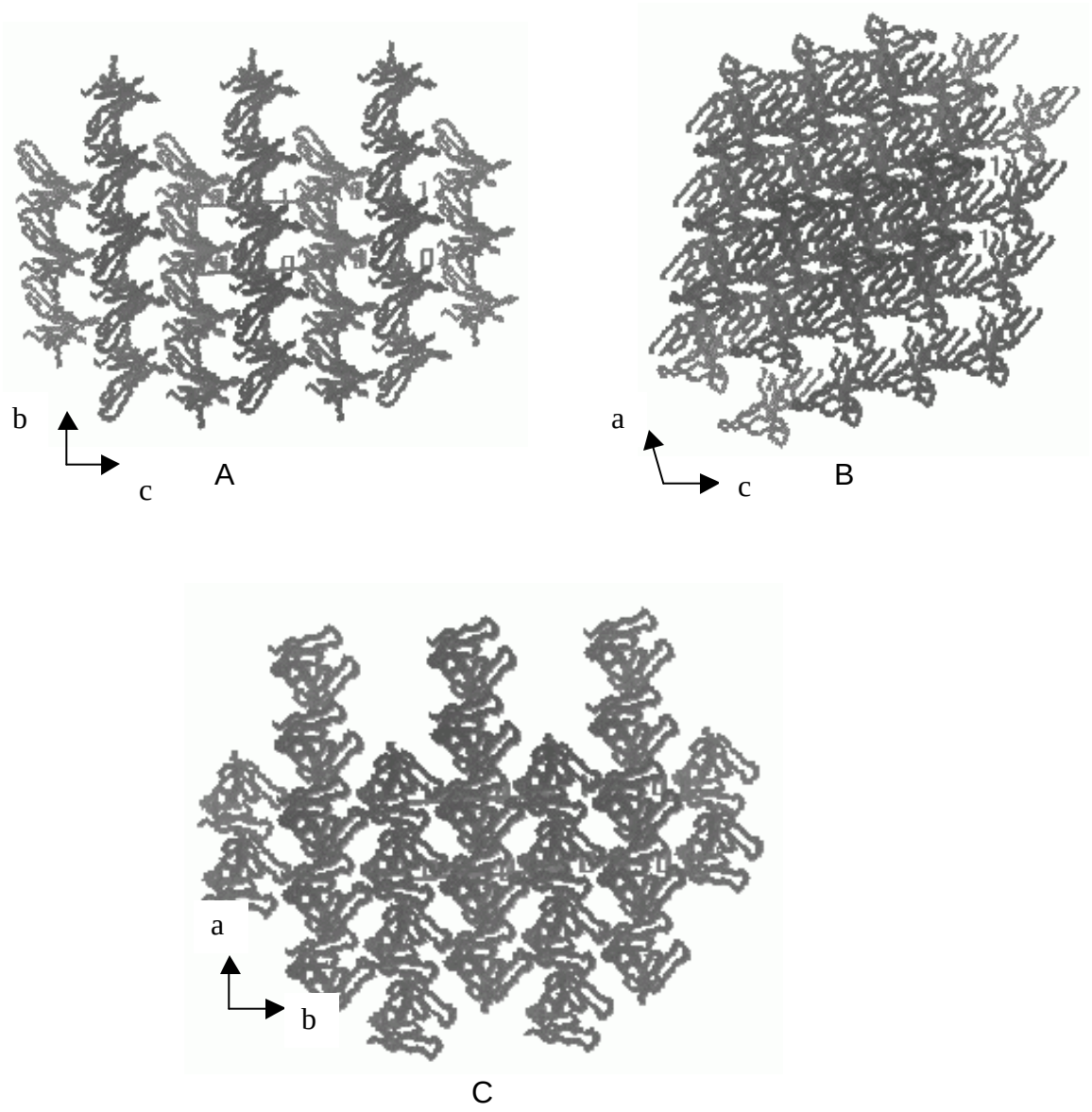


Fig. 5. 10. The crystal packing diagrams of *Vigna* BBI in monoclinic space group P21 viewed along the three unit cell axes directions, a (A), b (B) and c (C), the directions of axes are shown in arrows.

5. 2. 7. Structure of the reactive site loops

The two protease binding sites of BBI are located in the turn region of the two pairs of anti-parallel ß strands that make the two domains of the molecule (Figs. 5. 5 and 5. 6. A-B). In each subunit the two reactive site loops are at two symmetrically related ends along the longest dimension of the subunit, 36 Å apart. This arrangement of the reactive sites facilitates simultaneous inhibition of two protease molecules. The two loops project out from the BBI so as to be easily accessible by the active site of the proteolytic enymes. The chymotrypsin insertion loop is poorly defined in the electron density map of both the monomers (Figs. 5. 2. B and D). The trypsin loop is clearly located in the map of both the subunits (Figs. 5. 2. A and C). The latter loop in each monomer has been stabilized presumably due to the mutual close disposition of them in the dimer (Fig. 5. 8. A).

The reactive site loops are constrained by disulfide bridges, Cys12-Cys20 in the N-domain and Cys39-Cys47 in the C-domain. The disulfide bridge is close to the Pn (Schetcher & Berger., 1967) side of the reactive site loop. The disulfide bridges could be playing the role of limiting the conformational freedom of the loops. In addition, the hydrogen bonds and van der Walls contacts between residues on the N-terminal and C-terminal sides of the loops also stabilize the rigid structure of the reactive site loops. The extremities of the two loops are tethered by two symmetry related strong polar interactions, the side chains of Asn24 and Asp51 are hydrogen bonded to the main chain nitrogen atoms of Asp8-Ser9 and Lys35-Ser36 (Fig. 5. 7. B). It may be interesting to note that Asn24 and Asp51 are invariant among BBIs (de la Sierra *et al*., 1999) and their positions related by 2-fold pseudo-symmetry, and thus may be crucial for stabilizing the structure. A conserved cis-proline residue at the P3' in each binding region creates a turn classified as type 'VI b' (Richardson, 1981). The residues at P3' and P4' positions and the internal hydrogen bond net work between Thr at P2 and Ser at P1' positions help the loop to acquire this conformation. Due to the interactions described above, the main chain of the P1 residue protrudes towards the outer direction, and also results in restrain of the conformation around the binding site. The P1-P1' reactive site of each domain

is part of a bulge in the type VI b turn of the hairpin. This feature is observed throughout the serine protease inhibitor families, despite differing topologies and sequence variation within the reactive sites themselves (Hubbard *et al.*, 1991; Laskowski & Qasim, 2000). The pointed shape of the 'VI b' turns reflect the lock and key motif thought to be the mode of binding of serpins to their cognate enzymes (Bode & Huber, 1992; Read & James, 1986).

One notable difference between the two reactive loops is at P3'-P4' in the sequence. In domain I these positions are occupied by two prolines, whereas in domain II it is Pro-Gly in this position as found in most BBIs (Fig. 1. 4). One cause for the observed differences in specificity of loops may be attributed to the conformational differences caused by the above difference in residues. The nature of the residue in the crucial position P1 is satisfied by the specificity of the loop for proteases inhibited (Lu *et al.*, 1997). Thus the P1 residue is lysine for trypsin binding, while the residue could not be identified in chymotrypsin inhibitory loop because of poor electron density. It is known that in addition to the P1 residue, the residues at P2 to P2' are also important for binding (Gariani *et al.*, 1999; Pavone *et al.*, 1994; Terada *et al*, 1980). These residues are: Thr at P2, Ile at P2' and Ser at P1' (Lin *et al.*, 1993; McBride *et al.*, 1998). The P2 position in natural BBIs is most conserved and occupied by Thr. Usually a large aliphatic side chain, frequently Ile, is found at P2'. The residues of the trypsin binding loop seem to be more conserved than those of chymotrypsin loop (Fig. 1. 4., for aligned sequences). The loops are solvent exposed hence their surface accessibility and B-factors have high values. The small size of the reactive loop of BBI (9 residues) is in contrast with the large reactive site loop of the Kunitz type PI, which contains as many as 50 residues (Wlodawer *et al.*, 1987) suggesting that, the differences in molecular weight and cystine content reflect the sizes of the reactive site loops in these PIs.

5. 2. 8. Solvent structure

The final refined structure had a total of 58 ordered solvent molecules. Most of the solvent molecules were found on the surface of the protein. 20 water molecules were found in contact with protein atoms. They are W2, W3,

W4, W13, W14, W40, W45, W46 and W51, interacting with atoms of subunit A, and W1, W6, W8, W13, W15, W16, W23, W43, W44, W57 and W58 were interacting with the subunit B. Four of them, W3, W14, W27, W58, formed hydrogen bonds among themselves also. Rest had interactions among themselves only and not directly with protein. Also no waters are involved in bridging the two subunits.

An analysis of water molecules common in the two subunits was carried out. The procedure for selection of invariant waters is same as described in chapter 3 for lectin. The invariant water molecules W2, W4 and W13 of subunit A were substituted by W15, W44 and W57 at the corresponding positions in subunit B (Fig. 5. 8. A). The water molecules W2 and W15 interact with N of Cys 22 and O of Cys 49 in the two subunits; whereas W4 and W44 interact with O and N of the same two residues (table. 5. 7). W4 and W44 also interact with N and O of Asn 24. Also the water molecules W2 and W15 interact with O of Asp51 while W13 and W57 interact with O of Gly33. The B-factors of these water molecules are lower than the average value implying that they are tightly bound to protein atoms and may have a role in stabilizing the structure. Similar water mediated interactions involving homologous residues are observed in other BBI structures (de la Sierra *et al*., 1999). Thus it may be reasonable to conclude that these invariant water molecules are integral part of the structure, essential for preserving the tertiary structure of the BBI. The invariant water molecules and their interactions are listed in table 5. 7.

5. 2. 9. Comparison with other PI structures

The reported structures of BBIs from legumes are listed in table. 1. 2. When the atomic coordinates of the *Vigna* BBI were subjected to DALI database search three different types of homologous PI structures were found. They are the classical dicot BBIs, the monocot BBI from barley and the solution structure of bromelain inhibitor VI. The alignment of the amino acid sequences of several dicot BBIs is shown in Fig. 1. 4. The sequence identity among these BBIs is more than 50%, which imply that these BBIs may share a common tertiary fold. Also these proteins show intra-molecular sequence identity of 55%

Table. 5. 7. The Invariant water molecules, their B-factors and interacting common protein atoms in the two subunits of *Vigna* dimer.

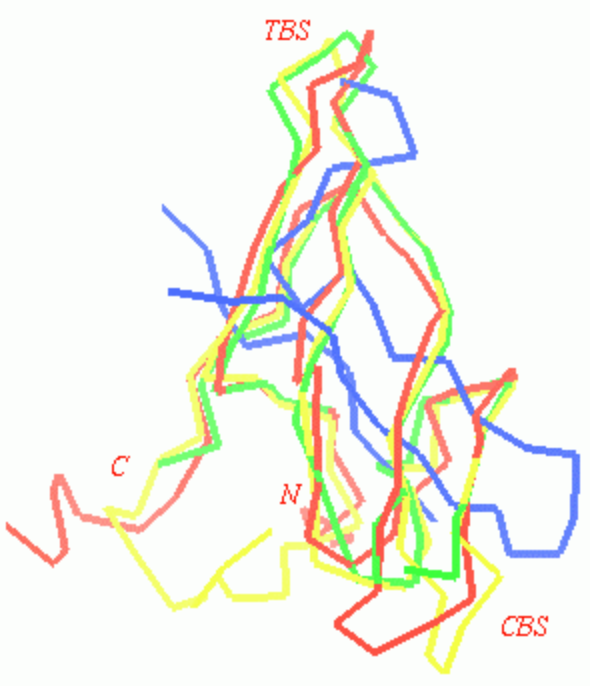
Atoms of subunit A and B Residue ID atom	**Interacting water in subunit A and Distance in Å**	**B-factor in Å^2**	**Interacting water in subunit B and Distance in Å**	**B-factor in Å^2**
Cys 22 N	W2 2.83	26.5	W15 2.73	30.0
Cys 22 O	W4 3.08	-	W44 2.55	-
Asn 24 N	W4 2.78	34.0	W44 3.43	32.0
Asn 24 O	W4 3.21	-	W44 2.90	-
Gly 33 O	W13 2.98	42.3	W57 3.10	46.7
Cys 49 N	W4 2.80	-	W44 2.97	-
Cys 49 O	W2 2.80	-	W15 2.90	-
Asp 51 O	W2 2.58	-	W15 2.60	-

between the N and C domains. The *Vigna* BBI structure was compared with all the known Bowman-Birk family PI structures. The program ALIGN was used for the super-positioning of structures. Table 5. 8., lists the r. m. s. deviation in Ca positions when the *Vigna* structure was superposed onto other PI structures. In the case of barley monocot BBI the N-terminal domain (residues 1 to 66) and C-terminal domain (residues 67 to 121) were separated and used for alignment with the subunits of *Vigna*.

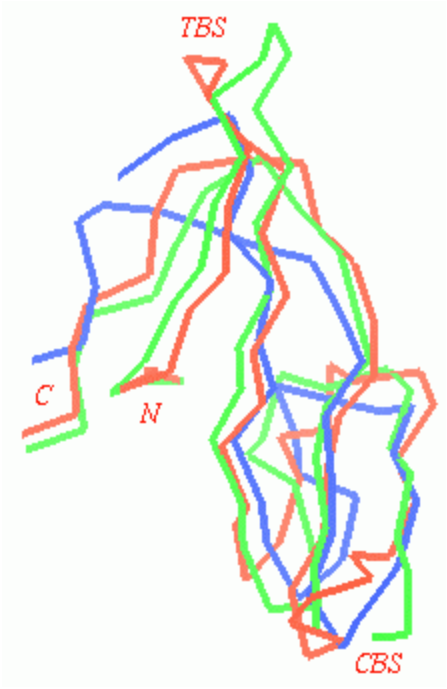
The overall fold of BBIs ("bow-tie motif") is conserved with only minor differences. The *Vigna* structure compares well with the structures of BBIs from tracey soybean, soybean, mung bean and pea seeds (Fig. 5. 11. A). The difference of *Vigna* BBI with the other structures is mainly in the terminal and the loop regions, specifically with the chymotrypsin binding loop, which is poorly defined in *Vigna*. It deviates from the structure of BBI from adzuki bean. The monomer of *Vigna* BBI overlap best with the pea monomer and least with the adzuki bean. The coordinates of tetrameric peanut BBI is not yet available for comparison. The superposition of *Vigna* over the monocot barley BBI and the bromelain PI is shown in Fig. 5. 11. B. The structures deviate much in the loop at the chymotrypsin binding site regions.

Table. 5. 8. The result of the superposition of subunits of *Vigna* BBI with the homologous PI structures of DALI database. In the case of pea BBI both its subunits (A and B) were compared with each subunit of *Vigna*. ND and CD are N and C domains, respectively. The PDB codes of structures used for alignment are also given.

PI /PDB Code	**Subunit A**	**Subunit B**
Tracey-bean/ 1PI-2	0.81	0.98
Soybean/ 2DR6	1.72	1.60
Mungbean/ 1DF9	1.13	1.17
Pea seeds/1PBI	0.68 (A) 0.68 (B)	0.80 (A) 0.81 (B)
Barley/ 1C2A	1.85 (ND) 2.40 (CD)	1.78 (ND) 2.33 (CD)
Bromelain PI/ 1BI6	2.54	2.30
Adzuki bean/ 1TAB	4.16	4.64



A



B

Fig. 5. 11. Superposed Ca trace structure of *Vigna* BBI (green) over those of BBIs from, A) soybean (red), adzukibean (blue) and mungbean (yellow), B) barley (red) and the bromelain PI (blue).

5. 2. 10. Docking of *Vigna* BBI on to Bovine Trypsin (BBI-BT) using 3D-DOCK

Unfortunately, our attempts to grow single crystals of the complex of BBI with neither trypsin nor chymotrypsin succeeded. Hence we decided to carry out modeling studies by simulating the BBI-trypsin complex using 3D-DOCK docking programs. The model was constructed assuming that the binding site loop of the BBI bind to the catalytic residues of the enzyme in a manner similar to a productively bound substrate, as observed in crystal structures (Lin *et al.*,1993; Koepke *et al.*, 2000; Krishnamurthy *et al.*, 2000). The atomic coordinates of *Vigna* BBI and bovine trypsin (from PDB code 1D6R) were used to simulate the complex structure. To facilitate easy comparison with observed structure, the trypsin coordinates were chosen from PDB file 1D6R, the bovine trypsin - soybean BBI crystal structure (Koepke *et al.*, 2000).

3D-DOCK is a suite of programs (Gabb *et al.*, 1997) designed for computational prediction of protein-protein complexes. Starting from the individual structures of the combining proteins, which are known to form a biological complex, it is possible to predict a limited set of possible models for the complex using the 3D-Dock programs. The suite consists of the modules FTDOCK, RPSCORE, FILTER and MULTIDOCK, which can be run sequentially. FTDOCK descretises any two given proteins onto orthogonal grids and performs a global scan in translational and rotational space for possible association of the two molecules. The search is limited by surface complimentarity and electrostatic filter and each complex will be ranked based on these two parameters. RPSCORE uses an empirical pair potential matrix to score each docked complex (Moont *et al.*, 1999). FILTER reduces the number of possible complexes with the knowledge of contacting residues from either side of the interface. Energy minimization and removal of any steric clashes between residues at the interface is done using MULTIDOCK (Jackson *et al.*, 1993).

Input files were prepared from normal PDB files of *Vigna* BBI and bovine trypsin using a PERL program. In the first step ftdock was run, which used to take a few days on silicon graphics Octane workstation, and output the file

'ftdock_global.dat' containing 10000 records. Each record is identified by an 'ID' number and arranged in the descending order of surface complimentarily score (Sc score). When a group of complexes have the same Sc score the favorable one is selected based on electrostatic score ratio. The output file contained three translational components and three rotational angles of each pair. In the next step RPSCORE was run; its output in the file ftdock_rpscored.dat also contained same 10000 records but with a new score. The complexes are now ordered according to their residue level pair potential score (RPSCORE). Each complex is now identified by a new ID number and the ID of ftdock run is also carried over as 'PreID' in order to trace back the ranking of a complex as the successive programs are run. The knowledge of the residues of BBI involved in binding trypsin from the known structures of BBI-trypsin complexes, was used for selecting a few from the 10000 records. The output of this step was ftdock_filtered.dat. After applying the above constraints the number of selected complexes drastically reduced to 23. The constraints applied is the contact between the three catalytic residues His57, Asp102, Ser195 of the enzyme with the residues of the BBI. Finally, the program 'build' was used to obtain the modeled complexes in the PDB format. The output file of complexes are named Complex_xg.pdb, where x stands for the record ID number from the input file. One of these complexes, found closer to the crystal structure in orientation of binding and similar interactions, was selected for structural analysis. This complex was ranked at 4395 after FTDOCK run, 14 after RPSCORE and 5 after FILTER run.

5. 2. 11. Homology modeling

The predicted complexes by 3D-DOCK docking program did not contain all the interactions generally observed in the crystal structures of BBI-trypsin complexes. Also the orientation of BBI after binding to the trypsin did not match with that observed in the crystal structure (Fig. 5. 12. A). Moreover these complexes were found energetically less favorable by about 30 kcal/mol when compared to the crystal structure of soybean BBI-bovine trypsin complex, from which the trypsin coordinates were chosen.

Thus homology modeling was carried out based on the crystal structure 1D6R. To start with, the *Vigna* BBI structure was superposed on to soybean BBI using the ALIGN program. The trypsin coordinates were taken from 1D6R. The rigid body energy minimization was carried out using MULTIDOCK on the above set. This energy minimized complex structure (Fig. 5. 12. B) had energy, -62 kcal/mol, about -30 kcal/mol favorable energy than the 3d-dock docked complex but less favorable by 20 kcal/mol when compared to the crystal structure (table. 5. 9). The trypsin binding loops of *Vigna* BBI and soybean differed at only one residue, Ile 16 in *Vigna* is replaced by Asn in soybean, which had hydrogen bonded contacts with trypsin. Among other BBIs we found that mungbean BBI had similar trypsin binding loop sequence. The structure of mungbean BBI in complex with NS3 protease of dengue virus, PDB code 1DF9, was available. Similar to *Vigna* BBI, a complex was simulated between mungbean BBI and trypsin from 1D6R. This complex after energy minimization had an interaction energy -77 kcal/mol, very close to that of the crystal structure 1D6R (-83 kcal/mol). Also the mungbean BBI-bovine trypsin complex was simulated using 3D-Dock docking program. The complex was energy minimized and had energy of -67 kcal/mol. The energetics of these interactions are listed in table. 5. 9 below.

Table. 5. 9. The minimized energies in kcal/mol of various BBI complexes, in the crystal structures and in the modeled complexes simulated by homology modeling and 3D-DOCK program.

Complex	**Crystal structure**	**Homology modeling**	**3D-DOCK**
Soybean BBI-BT	-83.5	ND	ND
Vigna BBI-BT	NA	-62.0	-57.0
Mungbean BBI-BT	NA	-80.0	-67.0
Mungbean BBI-NS3 protease	-75.0	ND	ND
Vigna BBI-NS3 protease	NA	-43.0	ND

NA: Not available; ND: Not done.

5. 2. 12. Structure of the *Vigna* BBI-BT complex

Discussion below of BBI-protease interactions relates substantially to the *Vigna* BBI-BT complex modeled by homology modeling. The two proteins associate using a broad contact surface area. There is a change of 1270 Å^2 in the solvent accessible area due to complex formation, which is considered to be the minimum size required to acquire stability of the structure and specificity (Conte *et al*., 1999). Individually, the enzyme lost 680 Å^2 and BBI 590 Å^2 of solvent accessible area. About 1040 Å^2 (82%) of interface area is contributed by main chain atoms, and rest by side chain atoms. Naturally, the largest change in the accessible surface area occurred at the trypsin insertion loop (Cys12-Cys20) of the BBI and at the S1 catalytic pocket residues of the enzyme (Asp189-Gly219).

Around 15 residues of BBI and 25 residues of trypsin are involved in making the intermolecular contacts at the interface. The chemical character of the interface is almost similar to that of an average protein interior containing 63% non-polar and 36% polar atoms. The recognition site has 14 hydrogen bonds involving 3 main chain-main chain, 2 main chain-side chain and 7 side chain -side chain atoms. The interface parameters estimated above for the model are comparable with those observed in many enzyme-inhibitor complexes (table. 1. 3).

The nine residue trypsin binding loop (Cys12-Cys20) was found to be in contact with the trypsin S1 primary binding site residues. The loop residues Cys12, Thr13, Lys14, Ser15, Ile16 and Gln19 were in contact with the trypsin residues Phe41, His57, Asp189, Ser190, Gln192, Gly193, Ser195, Ser214, Trp215, Gly216, Gly219 with in 4.0 Å distance (Fig. 5. 13. B-C). The binding is mainly centered about the residue P1 (Lys14) on the BBI side and the residues of the S1 specificity pocket (Asp189-Ser195 and Ser214-Gly216) on trypsin side. The side chain of Lys14 is hydrogen bonded to side chains of Asp 189 and Ser190 of trypsin. The main chain carbonyl group of Lys 14 is hydrogen bonded to N and side chain O$^?$ atoms of catalytic Ser195. The carbonyl carbon of P1 is in close proximity with nucleophilic O$^?$ of Ser195 (table. 5. 10). As per the mechanism known, the substrate cleavage is initiated by the nucleophilic

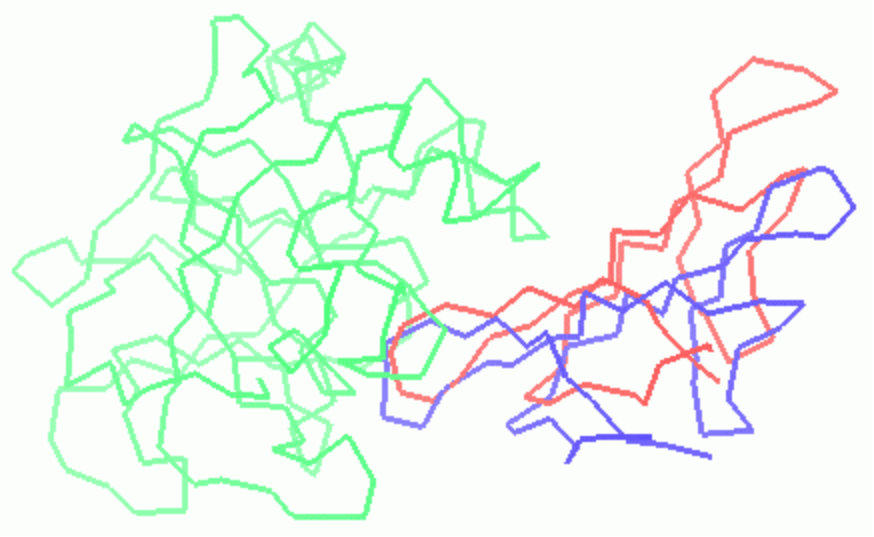
attack of $O^{?}$ (Ser195) on the carbonyl carbon of P1 residue. The residues at P2, P1' and P2' positions also interacts with trypsin atoms (table. 5. 10). No major conformational change upon complex formation is observed. The r. m. s. deviation of the backbone atoms of BBI before and after complexation is only 0.0007 Å. The interaction of BBI with the cognate protease, modeled here, compares with the reported crystal structures of BBI-potease complexes (Figs. 5. 12. A-B and Fig. 5. 13. A-C).

Our model gave a reasonable insight into structural aspects of the BBI-enzyme complex, although we excluded crystallographic waters, which might have left out some minor details of the interactions. The comparison of contact distances in the model and crystal structure (table 5. 10) suggest that the interaction of *Vigna* BBI with trypsin is very similar to that observed in crystal structure. Contacts made by BBI with trypsin imply that the enzyme has been inhibited by the standard mechanism that has been postulated for the serpins (Laskowski & Kato, 1980). Because of structural rigidity the BBI conformation remains unchanged with the required adjustments being made on the trypsin. Interaction of the main chain of BBI with the protease active site is similar to that seen in serine protease inhibitor complexes (Helland *et al*, 1999; Laskowksi & Qasim, 2000). The observed intra molecular interactions of the inhibitor's binding site with the inhibitor's core and inter-molecular interactions with the enzyme's binding site mutually stabilize each other and are also strong enough to prevent dissociation. These mutually stabilizing interactions apparently prevent reactive site deformation presumed to be important for peptide bond cleavage, and thus slow down catalytic processing. The two interacting partners have structures complimentary to each other and cavities rarely occur at the interface as it is closely packed like amino acids in crystals. The type of interface we have observed here usually form stable protein-protein complexes (Conte *et al*., 1999). Structural analysis of these types of complexes is essential for identifying the general features of a protein surface that are important in protein-protein interactions. The results presented in this study is an assessment of the important interactions that helps in the formation of protein-protein complexes.

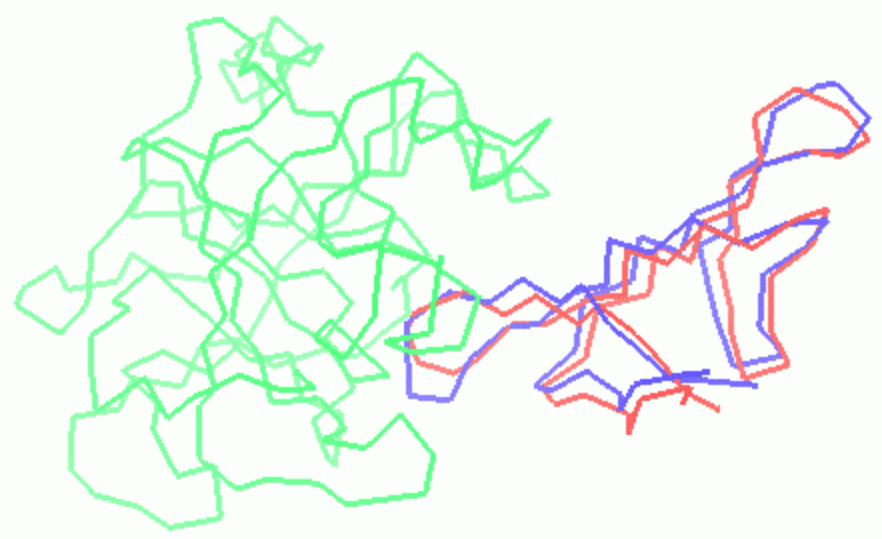
Table. 5. 10. The polar contacts in various BBI-BT complexes. The crystal structures of soybean BBI-BT (SBBI-BT) and the complexes of *Vigna* BBI-BT and mungbean BBI-BT by homology modeling, (H-VBBI-BT & H-MBBI-BT) and 3d-dock docked complex of *Vigna* BBI-BT (3D-VBBI-BT) are included. Distances (Å) beyond 5 are indicated by a '–' sign. The inhibitors residue numbering corresponds to the *Vigna* BBI. With reference to residue numbering 'n' in *Vigna*, the residue numbering in the crystal structures of soybean is n+2 and in mungbean is n+5.

BBI atom Res ID atom	**Trypsin atom** Res ID atom	**SBBI-BT**	**HVBBI-BT**	**3D-VBBI-BT**	**HMBBI-BT**
Mc-Mc					
Cys12 N	Gly216 O	3.10	3.96	3.93	2.95
Cys12 O	Gly216 N	3.06	3.68	4.42	3.03
Thr13 N	Gly216 O	4.67	-	-	-
	Ser214 O	-	2.82	4.18	4.65
Lys14 N	Trp215 N	4.68	4.42	3.84	4.21
	Ser214 O	2.93	2.82	3.01	-
	Gly216 N	-	-	-	-
Lys14 O	Gly193 N	2.92	3.58	-	3.07
	Ser195 N	2.89	-	3.82	2.67
	Gln192 N	-	-	4.50	4.58
Ser15 O	Gly193 N	4.31	4.80	-	4.50
Ile 16 N	Phe 41 O	3.42	3.67	-	3.28
Mc-Sc					
Cys10 O	Ser217 OG	2.72	-	-	2.78
Thr13 O	Gln192 OE1	3.30	-	-	3.64
Lys14 N	Ser195 OG	2.93	-	4.23	2.86
Lys14 O	Ser195 OG	2.84	2.76	3.93	2.66
Lys14 NZ	Ser190 O	3.14	2.97	2.85	-

	Gly219 O	4.33	3.94	3.71	-
	Lys224 O	-	-	5.00	-
Ser15 N	Ser195 OG	3.34	4.20	3.50	2.58
Ser15 O	Gln192 OE1	-	-	2.73	-
	Gln192 NE2	-	-	3.96	-
Ser15 OG	His 57 O	4.86	4.70	2.86	4.30
His21 ND1	Asn97 O	-	-	2.00	4.60
His21 NE2	Asn97 O	2.87	-	-	4.45
Arg23 NH1	Asn97 O	3.13	-	-	-
Arg23 NE	Asn97 O	3.40	-	2.93	-
Cys49 O	Asn97 ND2	-	-	3.77	-
Leu50 N	Asn97 ND2	-	-	3.13	-
Leu50 O	Asn97 ND2	-	-	-	4.17
Sc-Sc					
Asp8 OD1	Gln175 NE2	3.32	2.75	3.09	4.60
Asp8 OD2	Gln175 NE2	3.26	3.11	2.90	2.63
Thr13 OG1	His 57 ND1	4.71	3.53	-	4.65
	His 57 NE2	3.93	4.22	-	3.88
Lys14 NZ	Asp189 OD1	3.64	2.82	2.00	-
	Asp189 OD2	3.58	3.07	3.53	4.56
	Ser190 OG	3.55	-	3.93	-
Ser15 OG	His 57 NE2	4.17	4.53	3.75	3.26
	His 57 ND1	-	2.66	-	4.81
Gln19 NE2	His 57 NE2	-	3.45	4.65	-
His21 ND1	Asn 97 OD1	-	-	3.03	-
	Asn97 ND2	4.41	-	3.86	-
Arg23 NH2	Asn97 OD1	3.40	-	3.65	-
Arg48 NE	Asn97 OD1	-	-	3.65	-



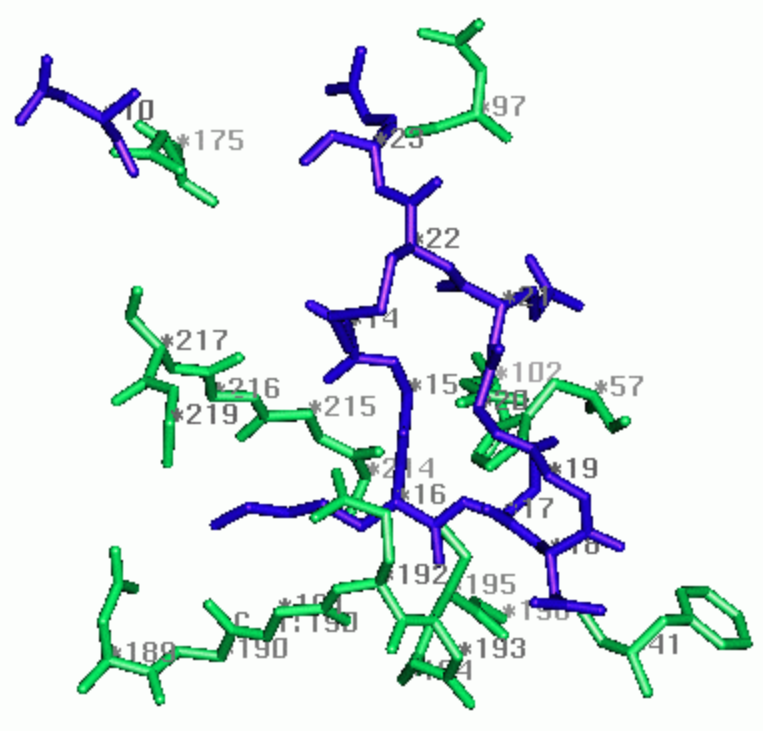
A



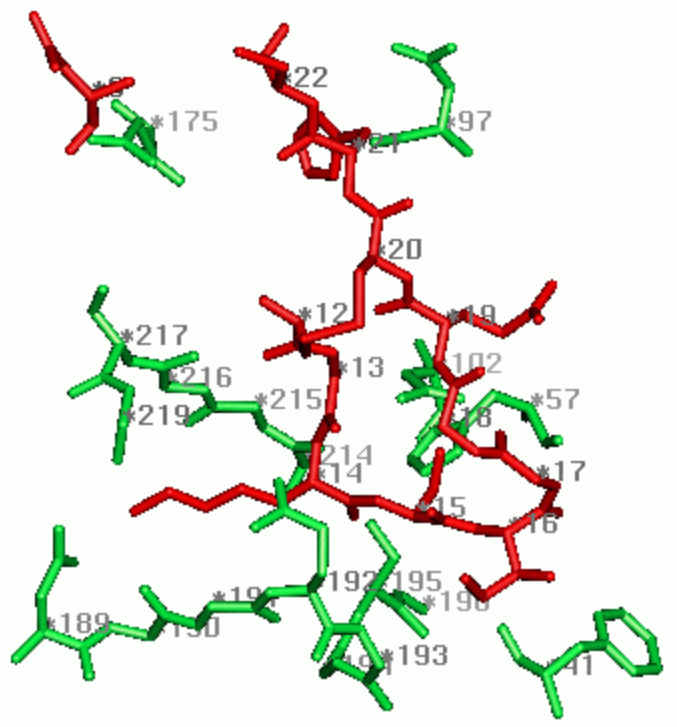
B

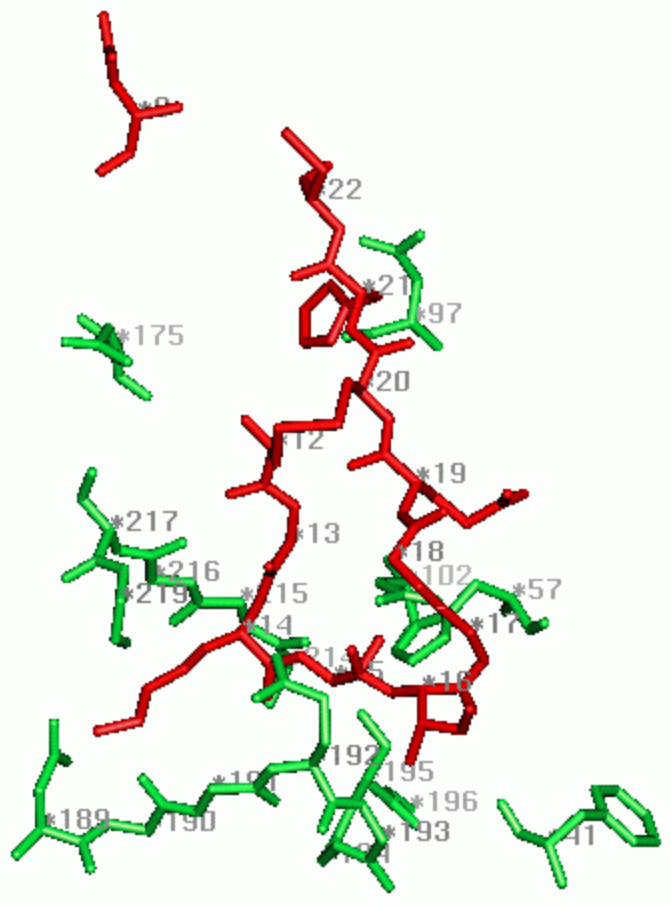
Fig. 5. 12. Comparison of the modeled *Vigna* BBI-BT complex with the crystal structure 1DR6, using Ca positions, A) soybean BBI (blue), and *Vigna* BBI by 3D-Dock (red) bound to bovine trypsin (green); B) soybean BBI (blue) and *Vigna* BBI by homology modeling (red) bound to bovine trypsin.

A


*175
*97
*22
*217
*216
*219
*215
*102
*57
*19
*16
*192
*195
*193
*190
*41

B


*175
*22
*97
*20
*217
*216
*219
*215
*13
*102
*57
*214
*17
*15
*16
*192
*195
*193
*190
*189
*41



C

Fig. 5. 13. The binding sites of A) soybean BBI bound to bovine trypsin in the crystal structure 1DR6; B) 3D-Dock based modeled complex of *Vigna* BBI-bovine trypsin; C) homology based modeled complex of *Vigna* BBI-bovine trypsin. The enzyme residues are shown in green and the BBI residues in red and blue. The residues are numbered by their Cα positions.

5. 3. NOVELTY AND CONCLUSIONS

Although, large variations in the overall structure of serine protease inhibitors exist, loop regions display remarkable similarity in conformation. In order to determine the general principles of inhibition by protein inhibitors, the structural study of a variety of complexes will be prudent. Indeed, an increasing number of complex structures analyzed using X-ray crystallography are reported. However, the details of the mechanism and fine specificity of complex formation still remain obscure. In this context it is interesting to note that out of some 70 structures of protease inhibitor-enzyme complexes determined to date only a few belong to BBI (Helland *et al*., 1999).

Many BBIs from plants have been isolated, characterized and sequences determined. Despite these extensive studies, the three-dimensional structures determined for these proteins are very few. The structures of these small double headed PIs is intriguing owing to their unique properties and applications. Double headed PIs are thought to have evolved from a single headed common ancestral protein and through subsequent mutation processes. The intra molecular sequence homology of *Vigna* BBI render credence to this hypothesis. Like the pea BBI structure, *Vigna* BBI is also a tight dimer in the asymmetric unit, but their quaternary structures are different. In *Vigna* the dimers associate face to face along the long direction, whereas in pea BBI they associate perpendicular to each other. Out of the four binding sites of *Vigna* dimer, two are located close to the interface. Hence such dimers can not be expected to represent true double headed functional PI. The three-dimensional structure of the ternary complex of by-functional soybean BBI monomer shows that it can bind two trypsin molecules at the same time (Koepke *et al*., 2000). Similarly, a modeled ternary complex of pea BBI with trypsin and chymotrypsin indicates that the simultaneous binding of two proteases is only possible with a monomeric molecule (de la Seirra *et al*., 1999). It is not very clear whether nature has selected this type of dimer formation to protect one set of protease inhibition sites from any unwanted protease attack.

Understanding the molecular recognition on the basis of three-dimensional structure is significant from both academic point of view and in applied research. Protein-protein recognition is a central event in the functional biology and needs deep study using variety of examples. On the other hand, there are relatively few structures of true protein-protein complexes available and hence the principles governing these interactions are yet not fully understood. Here we have made a modest attempt to understand protein-protein interactions by modeling a complex structure of *Vigna* BBI with bovine trypsin. The complex was found to contain typical interactions -hydrogen bonds, electrostatic and hydrophobic forces- involved in recognition process. The interface area is occupied by a large non-polar component which provides hydrophobic free energy in favor of association as it becomes buried in complexes, and a polar component providing hydrogen bonds.

Further, the structural work on *Vigna* BBIs can be continued by growing better crystals and improving diffraction and getting better electron density for chymotrypsin binding loop region. The BBI monomer and dimer combining with more than one protease molecule simultaneously could be exploited to understand protein-protein interactions either by growing crystals of their complexes or through modeling studies.

BIBLIOGRAPHY

Allen. H. J., Cywinski. M., Palmberg. R. & DiCioccio. R. D., (1987). Comparative analysis of galactoside binding lectins isolated from mammalian species. *Arch. Biochem. Biophys.*, **256**; 523-533.

Amemiya. Y., (1997). X-ray storage-phosphor imaging plate detectors: High sensitivity X-ray area detector. *Methods Enzymol*., **276**; 233-243.

Andrews. P., (1965). The gel filtration behavior of proteins is related to their molecular weights over a wide range. *Biochem. J*., **96**; 595-600.

Ashwell.G. & Hardford. J., (1982). Carbohydrate-specific receptors of the liver. *Ann. Rev. Biochem*., **51**; 531-534.

Baker. J. W., Reeke. G. N., Wang. J. L., Cunningham. B. A. & Edelman. G. M., (1975). The covalent and three-dimensional structure of concanavalinA. *J. Biol. Chem*., **250**; 1513-1524.

Balzarini. J., Scholas. D., Neyts. J., Van Damme. E. J. M., Peumans. W. & Declerq. E., (1991). a(1,3) and a(1,6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections *in vitro*. *Antimicrob. agents & chemother*., **31**; 410-416.

Barbieri. S., Batelli.G. & Stirpe. F., (1993). Ribosome inactivating proteins from plants. *Biochim. Biophy. Acta*., **1154**; 237-282.

Barondes. S. H., (1981). Lectins: Their multiple endogenous cellular functions. *Ann. Rev. Biochem*., **50**; 207-231.

Barondes. S. H., (1984). Soluble lectins: A new class of extra cellular proteins. *Science* **223**; 1259-1264.

Barre. A., Bourne. Y,. Van Damme. E. J. M., Peumans. W. J., & Rouge. P., (2001). Mannose binding plant lectins: Differential structural scaffolds for a common recognition process. *Biochimie*., **83**; 645-651.

Barre. A., Van Damme. E. J. M., Peumans. W. J. & Rouge. P., (1996). Structure-function relationship of monocot mannose-binding lectins. *Plant Physiol*., **112**; 1531-1540.

Berman. H. M., Westbrook. J., Feng. Z., Gilliland. G., Bhat. T. N., Weiseeig. H., Shindyalov. I. N. & Bourne. P. E., (2000). The Protein Data Bank. *Nucleic Acids Res*., **28**; 235-242.

Bertels. A., De Greve. H. & Lintermans. P., (1991). Function and genetics of fimbrial and non-fimbrial lectins from E.coli. In: *Lectin Reviews*., Kilpatric. D. C., Van Driessche. E. & Bog-Hansan. T. C. eds, **Vol.1**. Sigma Chem. Co., St Louis MO, USA, pp.53-67.

Beuchamps. J. C. & Isaacs. N. W., (1999). Methods for X-ray diffraction anlysis of macromolecular structures. *Curr. Opin. chem. Biol*., **3**; 525-529.

Birk. Y., (1985). Comprehensive review on the Bowman-Birk inhibitor. *Int. J. Pept. Prot. Res*., **25**; 113-131.

Birk.Y., (1987). Proteinase inhibitors. In: *Hydrolytic Enzymes*., Neuroberger & Brocklehurst. eds, elsevier science publications. B.V., Amsterdam; pp.257-300.

Birk.Y., (1993). Protease inhibitors of plant origin and role of protease inhibitors in human nutrition. In: *Protease inhibitors as cancer chemopreventive agents*., Troll. W. & Keendy. A. R. eds., Plenum press, New York., pp.97-106.

Birk. Y., Gertler. A. & Khalef. S., (1963). A pure trypsin inhibitor from soybeans. *Biochem. J*., **87**; 281-284.

Bladier. D., Gattengo. L., Fabia. F., Perret. G. & Cornillot. P., (1980). Individual variations of the seven carbohydrate components of human erythrocyte membrane during aging *in vivo*. *Carbohyd. Res*., **83**; 371-376.

Blundell. T. L., Jhoti. H. & Abell. C., (2002). High-throughput crystallography for lead discovery in drug design. *Nature Rev*., **1**; 45-54.

Bode. W. & Huber. R., (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem*., **204**; 433-451.

Bode. W. & Huber. R., (1994). Proteinase-protein inhibitor interactions. *Fibrinolysis*., **8**; 161-171.

Bode. W. & Huber. R., (2000). Structural analysis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta*., **1477**; 241-252.

Bogan. A. A. & Thorn. K. S., (1998). An analysis of hot spots in protein interfaces. *J. Mol.Biol*., **280**; 1-9.

Borrebaeck. C. K. A. & Carlsson. R., (1989). *Adv. Lectin. Res*., **2**; 10.

Bouckaert. J., Hamelryck. T., Wyns. L., & Loris. R., (1999). Novel structures of plant lectins and their complexes with carbohydrates. *Curr.Opin.Stru.Biol*.,**9**; 572-577.

Bourne.Y., Mazurier. J., Legrand. D., Rouge. P., Montreuil. J., Spik. G. & Cambillau. R, (1994). Structure of a legume lectin complexed with the human lactotransferrin N2 fragment, and with an isolated biantennary glycopeptide: Role of the fucose moiety. *Structure*., **2**; 209-219.

Bourne. Y., Zamboni. V., Barre. A., Peumans. W. J., Van Damme. E. J. M. & Rouge P., (1999). *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure*., **7**; 1473-1482.

Bowles, D. J., (1990), Defense related proteins in higher plants. *Ann. Rev. Biochem*., **59**, 873-907.

Bowman. D. E., (1946). Differentiation of soybean anti-tryptic factors. *Proc. Soc. Expt. Biol. Med*., **63**; 547-550.

Bown. D. P., Wilkinson. H. S. & Gatehouse. J. A., (1997). Differentially regulated inhibitor-sensitive and insensitive protease genes form the phytophagous insect pest, *Helicoverpa armigera,* are members of complex multigene families. *Insect Biochem. Mol. Biol*., **27**; 625-638.

Boyd. W. C. & Slapeigh. E., (1954). Specific precipitating activity of plant agglutinins (lectins). *Science*., **119**; 419.

Bricogne. G., (1991). A multisolution method of phase determination by combined maximization of entropy and likelihood. IV. The *ab initio* solution of crystal structures from their X-ray powder data. *Acta Crystallogr*., **A47**; 830-841

Brunger. A. T., (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature*., **355**; 472-475.

Brunger. A. T., (1997). Patterson correlation searches and refinement. *Methods Enzymol*., **276**; 558-580.

Caron. M., Bladier. D. & Joubert. R., (1990). Soluble galactoside-binding vertebrate lectins: a protein family with common properties. *Int.J.Biochem*., **22**; 1379-1385.

Carlson. S. R., (1994). In: *Glycobilogy, a practical approach.,* Fukuda. M. & Kobata. A. eds., IRL Press; Oxford, pp. 1.

Castellano. E., Oliva. G. & Navaza. J., (1992). Fast rigid body refinement for molecular replacement techniques. *J. Appl .Crystllogr*., **25**; 281-284.

Chandra. N. R., Ramachandraiah. G., Bachawat., Dam. T. K., Surolia. A. & Vijayan M., (1999). Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. *J. Mol. Biol*., **285**; 1157-1168.

Chantalat. L., Wood. S. D., Rizkallah. P. & Reynolds. C. D., (1996). X-ray structure solution of Amaryllis lectin by molecular replacement with only 4% of the total diffracting matter. *Acta Crystllogr*., **D52**; 1146-1152.

Chayen. N. E., (1998). Comparative studies of protein crystallization by vapour-diffusion and Microbatch techniques. *Acta Crystallogr*., **D54**; 8-15.

Chen. P., Rose. J., Love. R., Wei. C. H. & Wang. B. C., (1992). Reactive sites of an anticarcinogenic Bowman-Birk inhibitor are similar to other tyrpsin inhibitors. *J. Biol. Chem*., **267**; 1990-1994.

Chinnathambi. S., Lachke. A. & Radhakrishnan. S., (1995). An extracellular α-L-arabinofuranosidase from *Sclerotium rolfsii*. *J. Chromatogr. A.,* **705**; 400-402.

Chothia. C. & Murzin., (1993). New folds for all ß-proteins. *Structure*., **1**; 217-222.

Chrispeels. M. J., & Raikhel. N. V., (1991). Lectins, lectin genes, and their role in plant defense. *Plant Cell*., **3**; 1-9.

Cohen. G. E., (1997). ALIGN: A program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystgr*., **30**; 1160-1161.

Collaborative Computational Project Number. 4., (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystllogr*., **D50**; 760-763.

Conte. L. L., Chothia. C. & Janin. J., (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol*., **285**. 2177-2198.

Cook. G. M. W., (1986). Cell surface carbohydrates: molecules in search of a function?. *J. Cell. Sci*., **4**; 45-70.

Crowther. R. A. (1972). The fast rotation function. In: *The Molecular Replacement Method*., Rossman. M. G. eds., Gordon & Breach publications, New York, pp.174-178.

Crowther. R. A. & Blow., (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta. Crystallogr*., **23**; 544-548.

Das. M. K., Khan, M. I. & Surolia, A., (1981). Fluorimetric studies of the binding of *Momordica charantia* lectin with ligands. *Biochem. J*., **195**; 341-343.

Da Silva. L. P., Leite. J. R. S. A., Bloch. C. & De Freitas. S. M., (2001). Stability of a black-eyed pea trypsin/chymotrypsin inhibitor. *Protein Pept. Lett*., **8**; 33-38.

de Freitas. S. M., de Mello. L. V., da Silva. M. C. M., Vriend. G., Neshich. G. & Ventura. M. M., (1997). Analysis of the black-eyed pea trypsin and chymotrypsin inhibitor-a-chymotrypsin complex. *FEBS. Letters*., **409**; 121-127.

DeLano. W. L., Ultsch. M. H., de Vos. A. M. & Wells. J. A., (2000). Convergent solutions to binding at a protein-protein interface. *Science*., **287**; 1279-1283.

de la Sierra. I. L., Quillien. L., Flecker. P., Gueguen. J. & Brunie. S., (1999). Dimeric crystal structure of a Bowman-Birk protease inhibitor from pea seeds. *J. Mol. Biol*., **285**; 1195-1207.

De Leo. F., Volpicella. M., Licciulli. F., Liuni. S., Gallerani. R. & Ceci. L. R., (2002). PLANT PIs: a database for plant protease inhibitors and their genes. *Nucleic Acids Res*., **30**; 347-348.

Dhanaraj. V., Patanjali. S. R., Suro;ia. M. & Vijayan. M., (1988). Preparation and preliminary X-ray studies of two crystal forms of anti-T lectin from jackfruit (*Artocarpus integrifolia*). *J. Mol. Biol*., **203**; 1135-1136.

Drenth. J., (1994). Principles of protein crystallography. Springer-Verlog; New York, pp. XIII + 305.

Drenth. J. & Haas. C., (1998). Nucleation in protein crystallization. *Acta Crystallogr*., **D54**; 867-872.

Drickamer. K., (1988). Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem*., **263**; 9557-9560.

Drickamer. K., (1995). Multiplicity of lectin-carbohydrate interactions. *Nat. Struct. Biol*., **2**; 437-439.

Drickamer. K., (1999). C-typ lectin-like domains. *Curr. Opin. Struct. Biol*., **9**; 585-590.

Ealick. S. E., (2000). Advances in mutiple wavelength anomalous diffraction crystallography. *Curr. Opin. chem. Biol*., **4**; 495-499.

Eckelkamp. C., Ehmann. B. & Schopfer. P., (1993). Wound-induced systemic accumulation of a transcript coding for a Bowman-Birk trypsin inhibitor-related protein in maize (*Zea mays* L.) seedlings. *FEBS. Letters*., **323**; 73-76.

Elgavish. S. & Sahnnan. B., (1997). Lectin-Carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci*. **22**, 462-467.

Elgavish. S. & Sahnnan. B., (1998). Structure of *Erythrina corallodendron* lectin and of its complexes with mono and disaccharides. *J. Mol. Biol*., **277**; 917-932.

Endo. Y., Mitsui. K., Motizuki. M. & Tsurrugi. K., (1987). The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and characteristics of the modification in 28S ribosomal RNA caused by the toxins. *J. Biol. chem*., **262**; 5908-5912.

Etzler. M. E., (1985). Plant lectins: molecular and biological aspects. *Ann. Rev. Plant physiol*., **36**; 209-234.

Etzler. M. E., (1992). Plant lectins: molecular biology, synthesis and function. In *Glycoconjugates: composition, structure and function*., Allen. H. J. & Kisailus. E. C. eds, Marcel Dekker, New York, pp. 521-539.

Evans. P. R., (2001). Rotations and rotation matrices. *Acta.Crystallogr*.,**D57**;1355-59.

Falasca. A. I., Abbondanza. A., Barbieri. L., Bolognesi. A., Rossi. C. A. & Stirpe. F., (1989). Purification and partial characterization of a lectin from the seeds of *Trichosanthes kirilowii Maximowicz*. *FEBS Letters*., **246**; 159-62.

Feeney. J., (1999). NMR studies of protein-ligand and protein-protein interactions involving proteins of therapeutic interest. In: *NMR in supramolecular chemistry*., *NATO. ASI. Ser*., Ser.C, **526**, Kluwer publications., pp. 281-300.

Felton. G. W., (1996). Nutritive quality of plant protein: Sources of variation and insect herbivore response. *Arch. Insect Biochem. Physiol*., **32**; 107-130.

Fenn., M. M., Meng. C. K., Wong. S. F., & Whitehouse. C. M., (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*., **246**; 64-71.

Ferrason. E., Quillien. L. & Guguen. J., (1995). Amino acid sequence of a Bowman-Birk proteinase inhibitor from pea seeds. *J. Prot. Chem*., **14**; 467-475.

Fersht. A. R., (1985). Enzyme Structure and Mechanism, Freeman. W. H. & Co, New York, pp. 305-307.

Flecker. P., (1987). Chemical synthesis, molecular cloning and expression of gene coding for a Bowman-Birk type protease inhibitor. *Eur.J.Biochem*.,**166**;151-156.

Flecker. P., (1993). Analysis of structure-activity relationships of the Bowman-Birk inhibitor of serine proteinases. Toward a rational design of new cancer chemopreventive agents. In: *Protease inhibitors as cancer chemopreventive agents*., Troll. W. & Kennedy. A. R., eds., Plenum press, N. Y., pp.161-176.

Flecker. P., (1995). Template-directed protien folding into a metastable state of increased activity. *Eur. J. Biochem*., **232**; 528-535.

Fraizer. W. A., Rosen. S. D., Reitherman. R. W. & Barondes. S. H., (1975). Purification and compression of two developmentally regulated lectins from *Dictyostelium discoidum J. Biol. Chem*., **250**; 7714-7721.

Gabb. H. A., Jackson. R. M. & Sternberg. M. J. E., (1997). Modeling protein docking using shape complementarily, electrostatistics and biochemical information. *J. Mol. Biol*., **272**., 106-120.

Gabius. H. J., & Gabius. S. eds., (1993). In: *Lectins and Glycobiology*., Springer-Verlag Berlin Heidelberg., pp. 3-500.

Gabius. H. J., Walzel. H., Joshi. S. S., Kruip. J., Kojima. S., Goke. V., Kratziu. H., & Gabius. S., (1992). The immuno-modulatory galactoside specific lectin from misteltoe: partial sequence analysis, cell and tissue binding, and impact on intracellular bio-signaling of monocytic leukemia cells. *Anticancer. Res*., **12**; 669-676.

Gaier. J. R., Tulinsky. A., & Leiner. I. E., (1981). Formation, crystallization and preliminary X-ray crystallographic data of the ternary complex of a-chymotrypsin, ß-trypsin, and the Bowman-Birk inhibitor. *J. Biol. Chem*., **256**; 11417-11419.

Gaikwad. S. M., Gurjar. M. M. & Khan. M. I. (1998). Fluorimetric studies on saccharide binding to the basic lectin from *Artocarpus hirsuta*. *Biochem. Mol. Bio. Int*., **46**; 1-9.

Gaikwad. S. M., Gurjar. M. M. & Khan. M. I. (2002). *Artocarpus hirsuta* lectin: Differential modes of chemical and thermal denaturation. *Eur. J. Biochem*., **269**; 1-5.

Gariani. T., McBride. J. D., & Laetherbarrow. R. J., (1999). The role P2' position of Bowman-Birk proteinase inhibitor in the inhibition of trypsin: studies on P2' variation in cyclic peptides encompassing the reactive site loop. *Biochim. Biophy. Acta*., **1431**; 232-247.

Garman. E. F. & Schnieder. T. R., (1997). Macromolecular Crystallography. *J. Appl. Crysllogr*., **30**; 211-237.

Gatehouse. A. M. R. & Hilder. V. A., (1988). Introduction of genes conferring insect resistance. In: *Proceedings of the Brighton crop protection conference*., **3**; 1245-1254.

Gatehouse. A. M. R., Gatehouse. J. A., Dobie. P., Kilminster. A. M. & Boulter. D., (1979). Biochemical basis of insect resistance in *Vigna unguiculata*. *J. Sci. Food Agric*., **30**; 948-958.

Gatehouse. A. M. R., Gatehouse. J. A. & Boulter. D., (1980). Isolation and characterization of trypsin inhibitors from cowpea (*Vigna unguiculata*). *Phytochemistry*., **19**; 751-756.

Gegg. C. V., Roberts. D. D., Sege. I. H. & Etzler. M. E., (1992). Characterization of adenine binding sites of two *Dolichos biflorus* lectins. *Biochemistry*.,**31**;6938-42.

Gennis. L. S. & Cantor. C. R., (1976). Double headed protease inhibitors from black-eyed peas. *J. Biol. Chem*., **251**; 734-775.

Ghoshal. D., Sen. S. K. &. Goyal. A., (2001). Introduction and expression of cowpea trypsin inhibitor (CpTI) gene in transgenic tobacco. *J. Plant Biochem. Biotechnol*., **10**; 19-24.

Gilboa-Garber. N. & Avichezer. D., (1993). In: *Lectins and Glycobiology*., Gabius. H. J., & Gabius. S. eds., Springer-Verlag Berlin Heidelberg., pp. 380-395.

Gilboa-Garber. N. Susswein. A. J., Mizrahi. L., & Avichezer. D., (1985). Purification and characterization of the gonad lectin of Aplysia depilans. *FEBS letters*., **181**: 267-270.

Gilljam. G., (1993). Envelope glycoproteins of HIV-1, HIV-2 and SIV purified with *Galanthus nivalis* agglutinin induce strong immune responses. *Aids Res. & Human. Retrovir*., **9**; 431-438.

Glusker. J. P., Lewis. M. & Rossi. M., (1994). Crystal structure analysis for Chemists and Biologists. VCH publishers Inc., New York, pp. XVIIII +854.

Goldstein. I. J., & Poretz. R. D., (1986). Isolation, physicochemical characterization, and carbohydrate binding specificity of lectins. In: *The lectins: Properties, Functions and Applications in Biology and Medicine*, Leiner. I. E., Sharon. N. & Goldstein. I. J., eds., Academic Press, Orlando, pp. 33-247.

Gurjar. M. M., Khan. M. I. & Gaikwad. S. M., (1998). Methyl-a-galactoside binding lectin from *Artocarpus hirsuta*: Characterization of the sugar specificity and binding site. *Biochem. Biophys. Acta*., **1381**; 256-264.

Gurjar. M. M., Gaikwad. S. M., Solakhe. S.G., Mukherjee.S. & Khan. M. I. (2000). Growth inhibition and total loss of reproductive potential in *Trilobium cattaneum* by *Artocarpus hisuta* lectin. *Invr. Repro & Dev*. **38**; 95-98.

Hajto. T., Hostanska. K., Frei. K., Rordorf. C. & Gabius. H. J., (1990). Increased secretion of tumor necrosis factor-a, interleukin-1 and interleukin-6 by human mononuclear cells exposed to ß-galactoside specific lectin from clinically applied misteltoe extract. *Cancer. Res*., **50**; 3322-3326.

Hardman. K. D. & Ainsworth. C. F., (1972). Structure of concanavalinA at 2.4 ? resolution. *Biochemistry*., **11**; 4910-4919.

Harms-Ringdahl. M., Forsberg. J., Fedorcscak. I. & Ehenberg. L., (1979). Trypsin inhibitory activity of a polypeptide isolated from red kidney beans that also enhances lymphocyte simulation. *Biochem. Biophys. Res. Commun*.,**86**;492-99.

Harsulkar. A. M., Giri. A. P., Patankar. A. G., Gupta. V. S., Sainani. M. N., Ranjekar. P. K. & Deshpande. V. V., (1999). Successive use of non-host plant proteinase inhibitors required for effective inhibition of *Helicoverpa armigera* gut proteinases and larval growth. *Plant Physiol*., **121**; 497-506.

Hatano. K., Kojima. M., Tanokura. M. & Takahashi. K., (1996). Solution structure of bromaelain inhibitor VI from pineapple stem: Structural similarity with Bowman-Birk trypsin/chymotrypsin inihibitor from soybean. *Biochemistry*. **35**; 5379-5384.

Head-Gordon. T. & Brooks. C. L., (1991). Virtual rigid body dynamics. *Biopolymers*., **235**; 1021-1031.

Hedge. R., Maiti. T. K. & Podder. S. K., (1991). Purification and characterization of three toxins and two agglutinins from *Abrus precatorius* seed by using lactamyl-Sepharose affinity chromatography. *Anal. Biochem*., **194**; 101-109.

Helland. R., Otlewski. J., Sundheim. O., Dadlez. M. & Smalas. A. O., (1999). The crystal structures of the complexes between bovine ß-trypsin and ten P1 variants of BPTI. *J. Mol. Biol*., **287**; 923-942.

Hendrickson. W. A., (2000). Synchrotron crystallography. *Trends Biochem. Sci*., **25**; 637-643.

Hester. G., & Wright. C. S., (1996). The mannose specific bulb lectin from *Galnthus nivalis* (Snowdrop) binds Mono- and Dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 ? and 3.0 ? resolution. *J. Mol. Biol*., **262**; 516-531.

Hester. G., Kaku. H., Goldstein. I. J. & Wright. C. S., (1995). Structure of mannose-specific snowdrop (*Galnthus nivalis*) lectin is representative of a new plant lectin family. Nat*. Struct. Biol*., **2**; 472-479.

Higgins. J. D. & Gibson. T. J., (1995). CLUSTALW: Improving the sensitivity of progressive multiple sequece alignment through sequence weighting position specific gap penalties and weight matrix choice. *Nucleic Acids Res*.**22**;4673-80.

Hilder. V. A. & Boulter. D., (1999). Genetic engineering of crop plants for insect resistance – a critical review. *Crop Protection*., **18**; 177-191.

Hilder. V. A., Barker. A. F., Samour. R. A., Gatehouse. A. M. R. Gatehouse. J. A. & Boulter. D., (1989). Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp). *Plant Mol.Biol*.,**13**;701-10.

Hilder. V. A., Gatehouse. A. M. R., Sheerman. S. E., Barker. R. F., & Boulter. D., (1987). A novel mechanism of insect resistance engineered into tobacco. *Nature*., **330**; 160-163.

Holm. L. & Sander. C., (1995). DALI: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**; 478-480.

Hubbard. S. J., & Thoronton. J. M., (1993). NACCESS, Computer program, Department of Biochemistry and Molecular Biology, University college of London.

Hubbard. S. J., Campbell. S. F. & Thoronton. J. M., (1991). Molecular recognition: Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol*. **220**; 507-530.

Huber. R. & Schneider. M., (1985). A group refinement procedure in protein crystallography using Fourier transforms. *J. Appl. Crystllogr*., **18**; 165-169.

Huber. R., Kukla. D., Bode. W., Schwager. P., Bartels. K., Deisehofer. J., & Steigemann. W., (1974). Structure of the complex formed by bovine trypsin and of bovine pancreatic trypsin inhibitor, II. Crystallographic refinement at 1.9 ? resolution. *J. Mol. Biol*., **89**; 73-101.

Huber. R., Kukla. D., Ruhlmann. A., Epp. O. & Formanek. H., (1970). The basic trypsin inhibitor of bovine pancreas. Structure analysis and conformation of the polypeptide chain. *Naturweisenschaften*., **57**; 389-392.

Hughes. R. C., (1992). Lectins as cell adhesion molecules. *Curr. Opin. Struct. Biol*., **2**, 687-692.

Ikenaka. T. & Norioka. S., (1983). Amino acid sequences of trypsin/chymotrypsin inhibitors from peanut: A discussion on the molecular evolution of legume Bowman-Birk type inhibitors. *J. Biochem*., **94**, 589-599.

Ikenaka. T. & Norioka. S., (1986). Bowman-Birk family serine proteinase inhibitors. In: *Proteinase inhibitors*., Barrett. A. J. & Salvenson. G. eds., Amsterdam: Elsevier., pp. 361-374.

Inbar. J. & Chet. I., (1997). Lectins and Biocontrol. *Critic. Rev. Biotech*. **17**; 1-20.

Jackson. R. M., Gabb. H. A. & Sternberg. M. J. E., (1993). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol*., **276**., 265-285.

Jacoby. W. B., (1974). Affinity techniques: Enzyme purification, part B. *Methods Enzymol*., **34**.

Janin. J., (1995). Principles of protein-protein recognition from structure to thermodynamics. *Biochimie*., **77**; 497-505.

Janin. J., (1997). The kinetics of protein-protein recognition. *Protiens: Struct. Funct. Genet*., **28**; 153-161.

Janin. J & Chothia. C., (1990). The structure of protein-protein recognition sites. *J. Biol. Chem*., **265**; 16027-16030.

Jann. K. & Jann. B., (1990). In: *Current topics in Microbiology and Immunology*. **Vol.151**. Springer-Verlag, Berlin-Heidelberg.

Jensen. B., Unger. K. K., Uebe. J., Gey. M., Kim.Y-M. & Flecker. P., (1996). Proteolytic cleavage of soybean Bowman-Birk inhibitor monitored by means of high-performance capillary electrophoresis. Implications for the mechanism of proteinase inhibitors. *J. Biochem. Biophy. Meth*., **33**; 171-185.

Johnson. J. E., (1996). Functional implications protein-protein interactions in icosahedral viruses. *Proc Nat. Acad. Sci.* (USA)., **93**; 27-33.

Jones. S. & Thoronton. J. M., (1996). Principles of protein-protein interactions. *Proc. Nat. Acad. Sci*. (USA). **93**; 13-20.

Jones. S. & Thoronton. J. M., (1997). Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol*., **272**; 121-132.

Jongsma. M. A. & Bolter. C. J., (1997). The adaptation of insects to plant protease inhibitors. *J. Insect Physiol*., **43**; 885-895.

Joshi. S. S., Komanduri. K. C., Gabius. S. & Gabius . H. J., (1991). Immunotherapeuitc effects of purified misteltoe lectin, ML-I, on murine large cell lymphoma. In: *lectins and cancer*. Gabius. H. J. & Gabius. S. eds., Springer-Verlag, Heidelberg, pp. 207-217.

Jouanin. L., Bonade-B. M., Girard. C., Morrot. G. & Giband. M., (1998). Transgenic plants for insect resistance. *Plant Sci.*, **131**; 1-11.

Kabir. S. & Daar. A. S., (1994). The composition and properties of jacalin, a lectin of diverse applications obtained from jackfruit (*Artocrpus heterophyllus*) seeds. *Immunol. Invest*., **23**; 167-188.

Kakade. M. L., Simons. N. E. & Leiner. I. E., (1970). The molecular weight of the Bowman-Birk soybean proteinase inhibitor. *Biochim. Biophy. Acta*.,**200**;168-69.

Keen. N. T. & Stasakwicz. B., (1988). Host range determinants in plant pathogens and symbiosis. *Ann. Rev. Microbiol*., **42**; 421-440.

Kennedy. A. R., (1993). Anticarcinogenic activity of protease inhibitors. In: *Protease inhibitors as cancer chemopreventive agents*., Troll. W. & Kennedy. A. R., eds., Plenum press, New York., pp. 9-91.

Kennedy. A. R., (1994). Prevention of carcinogenesis by protease inhibitors. *Cancer. Res*., **54**; 1999s-2005s.

Kennedy. A. R., (1998). Chemopreventive agents: Protease inhibitors. *Pharmacol. Theor*., **78**; 167-209.

Kennedy. A. R., Barclay. Y. B., Kinzler. K. W. & Newberne . P. M., (1996). Suppression of carcinogenesis in the intestines of Min Mice by the soybean derived Bowman-Birk inhibitor. *Can. Res*., **56**; 679-682.

Kilpatrick. D. C., (1991). Lectin interaction with human leukocytes: mitogenicity, cell separation, clinical applications, In: *Lectin reviews* **Vol.1**, Sigma Chemical Co., St Louis, pp. 69-80.

Kobiler. D. & Mirelman. D., (1981). Adhesion of *E. histolytica* trophozoites to monolayers of human cells. *J. Infe. Dise*., **144**; 539-546.

Koepke. J., Ermler. U., Warkentin. E., Wenzl. G. & Flecker. P., (2000). Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 ? resolution. Structural basis of janus-faced serine protease inhibitor specificity. *J. Mol. Biol*., **298**; 477-491.

Komath. S. S. & Swamy. M. J., (1998). Further characterization of snake gourd (*Trichosanthes anguina*) seed lectin. *Curr. Sci.*, **75**; 608-611.

Kornberg. L. J., Earp. H. S., Turner. C. E., Prockop. C. & Juliano. R. L., (1991). Signal transduction by integrins: increased protein tyrosine phosphorylation caused by clustering of ß1 integrins. *Proc. Nat. Acad. Sci.*(USA)., **88**; 8392-96.

Krauspenhaar. R., Eschenburg. S., Perbandt. M., Kornilov. V., Konareva. N., Mikailova. I., Stoeva. S., Wacker. R., Maier. T., Singh. T., Mikhailov. A., Voelter. W. & Betzel. C. (1999). Crystal structure of mistltoe lectin1 from *viscum album*. *Biochem. Biophys. Res. Commun*., **257**; 418-424.

Krishna Murthy. H. M., Judge. K., DeLucas. L. & Padmanabhan. R., (2000). Crystal Structure of dengue virus NS3 protease in complex with a Bowman-Birk inhibitor: Implications for flaviviral polyprotein processing and drug design. *J. Mol. Biol.*, **301**; 759-767.

Krystek. S., Stouch. T. & Novotny. J., (1993). Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J. Mol. Biol*., **234**; 661-679.

Kumar. S. G., Appukuttan. P. S. & Basu. D., (1982). a-D-galactose specific lectin from jack fruit (*Artocarpus integrifolia*) seed. *J. Biosci.*, **4**; 257-261.

Kunitz. M., (1947). Crystalline soybean trypsin inhibitor. *J. Gen. Physiol*., **30**;291-310.

Lammelli. U. K., (1970). Cleavage of structural proteins during the assembly of the Head Bacteriophage T4. *Nature*., **227**; 680-685.

Laskowski. M, & Kato. I., (1980). Protein inhibitors of proteases. *Ann. Rev. Biochem*., **49**; 593-626.

Laskowski. M. Jr. & Qasim. M. A., (2000). What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme-substrate complexes. *Biochim. Biophys. Acta*., **1477**; 324-337.

Laskowski. M. Jr., Kato. I., Kohr. W. J., Park. S. J., Tashiro. M. & Wahtley. H. E., (1988). Positive Darwinian selection in evolution of protein inhibitors of serine proteinases. *Cold Spring Harbor. Symp. Quant. Biol*., **52**; 545-553.

Laskowski. R. A., MacArthur. M. W., Moss. D. S. & Thoronton. J. M., (1993). PROCHECK: A program to check the stereo-chemical quality of protein structures. *J. Appl. Crystllogr*., **26**; 283-291.

Lasky. L. A., (1995). Selectin-Carbohydrate interactions and the initiation of the inflammatory response. *Ann. Rev. Biochem*., **64**; 113-139.

Lee. B. & Richards. F. M., (1971). The interpretation of protein structures: estimation of static accessiblity. *J. Mol. Biol*., **55**; 379-400.

Lee. X., Thompson. A., Zhang. Z., Ton-that. H., Biesterfeldt. J., Ogata. C., Xu. L., Johnston. R. A. Z. & Young. N. M., (1998). Structure of the complex of *Maclura pomifera* Agglutinin and the T-antigen Disaccharide, Galß1, 3GalNAc. *J. Biol. Chem*., **273**; 6312-6318.

Levi, G., & Teichberg, V. I., (1981). Isolation and physicochemical characterization of electrolectin, a ß-D-galactoside binding lectin from the electric organ of *Electrophorus electricus*. *J. Biol. Chem*., **256**; 5735-5740.

Li. J., Carrol. J., & Ellar. D. J., (1991). Crystal structure of insecticidal d-endotoxin from *Bacillus thuringinesis* at 2.5 Å resolution. *Nature*., **353**; 815-821.

Li. M., Wang-P. Y., Chai-J. J., Wang-Y. K. & Bi. C., (2000). Molecular-replacement studies of *Trichosanthes kirilowii* lectin 1: a structure belonging to the family of type2 ribosome-inactivating proteins. *Acta Crystgr*., **D56**; 1073-1075.

Lin. G. D., Bode. W., Huber. R., Chi. C. & Engh. R. A, (1993). The 0.25 nm X-ray structure of the Bowman-Birk type inhibitor from mung bean in ternary complex with porcine trypsin. *Eur. J. Biochem*., **212**; 549-555.

Lindqvist. Y & Branden. C. I., (1980). Structure of glycolate oxidase form spinach at a resolution of 5.5 Å. *J. Mol. Biol*., **143**; 201-211.

Lis. H. & Sharon. N., (1977). Lectins: their chemistry and application to immunology. In: *Antigens*., Michael. S. eds., **Vol. 4**;: New York., pp. 429-529.

Lis. H. & Sharon. N., (1984). In: *Biology of Carbohydrates*, Ginsburg.P., Robbins.W. eds., **Vol. 2**; New York, Wiley; pp.1-85.

Lis. H. & Sharon. N., (1986a). Lectins as molecules and as tools. *Annu. Rev. Biochem*., **55**; 35-67.

Lis. H. & Sharon. N., (1986b). In: *The Lectins: properties, functions and applications in biology and medicine*. Liener. I. E., Sharon. N. & Goldstein. I. J. ed., 1986. New York: Academic press.

Lis. H. & Sharon. N., (1998). Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev*., **98**; 637-674.

Lord. J. M., Roberts. L. M. & Robertus. J. D., (1994). Ricin: Structure, mode of action, and some current applications. *The FASEB. J*., **8**; 201-208.

Loris. R., Hamelryck. T., Bouckaert. J. & Wyns. L., (1998). Legume lectin structure. *Biochem. Biophys. Acta.*, **1383**; 9-36.

Lotan. R., (1992). ß-galctoside binding vertebrate lectins: synthesis, molecular biology, function. In: *Glycoconjugates: Composition, Structure and Function*, Allen. H. J. & Kisailus. E. C. eds., Marcel Dekker., Inc., New York, pp. 635-637.

Lotan. R. & Raz. A., (1988). Endogenous lectins as mediators of tumor cell adhesion. *J. Cell. Biochem*., **37**; 107-117.

Louis. C. J. & Wyllie. R. G., (1981); Fluorescein- conA conjugates distinguish between normal and malignant human cells; a preliminary report. *Expertietia*., **37**; 508-509.

Louis. C. J., Wyllie. R. G., Chou. S. T. & Sztynda. T., (1981). Lectin-binding affinities of human epidermal tumors and related conditions. *Am. J. Clin. Pathol*., **75**; 642-647.

Lowry. O. H., Rosenburgh. N. J., Farr. A. L. & Randall. R. J., (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem*., **193**; 265-276.

Lu. W., Apostol. I., Qasim. M. A., Warne. N., Wynn. R., Zhang. W. L., Anderson. S., Chiang. Y. W., Ogin. E., Rothberg. I., Ryan. K. & Laskowski. M. Jr., (1997). Binding of amino acid side chains to S1 cavities of serine proteinases. *J. Mol. Biol*., **266**; 441-461.

Luckett. S., Garcia. R. S., Barker. J. J., Konarev. A. V., Shewry. P. R., Clarke. A. R. & Brady. R. L., (1999). High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol*., **290**; 525-533.

Mann. M., Hendrickson. R. C. & Pandey. A., (2001). Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem*., **70**; 437-473.

Manoj. N. & Suguna. K., (2001). Signature of quaternary structure in the sequences of legume lectins. *Protein Engg*., **14**; 735-745.

Manoj. N., Jayprakash. A. A., Pratap. J. V., Komath. S. S., Swamy. M. J. & Vijayan. M., (2001). Crystllization and preliminary X-ray studies of snake gourd lectin: homology with type II ribosome-inactivating proteins. *Acta Crystallogr*., **D57**; 912-914.

Mattevi. A., Obmolova. G., Schulze. E., Kalk. K. H., Westphal. A. H., Kok. A. D. & Hol. W. G. J., (1992). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science*., **255**; 1544-1550.

Matthews. B. W., (1968). Solvent content of protein crystals. *J. Mol. Biol*., **33**;491-97.

McBride. J. D., Brauer. A. B. E., Nievo. M., & Leatherbarrow. R. J., (1998). The role of threonine in the P2 position of Bowman-Birk proteinase inhibitors: Studies on P2 variation in cyclic peptides encompassing the reactive-site loop. *J. Mol. Biol*., **282**; 447-457.

Monsigny. M., Keida. C. & Rcohe. A. C., (1983). Antibody bearing liposomes targeting *in vivo*. *Biol. Cell*., **47**; 95-110.

Monsigny.M., Rcohe.A.C. & Midoux.P., (1984). Uptake of neo-glycoprotein via membrane lectins of L1210 cells evidenced by quantitative flow cyto-fluriometry of drug targeting. *Biol. Cell*., **51**; 187-196.

Moont. G., Gass. H.A. & Sternberg. M. J. E. (1999). Use of pair potentials across protein interfaced in screening predicted docked complexes. *Proteins*., **35**; 364-373.

Morhy. L. & Ventura. M. M., (1987). The complete amino acid sequence of the *Vigna unguiculata* (L.) walp. seed trypsin and chymotrypsin inhibitor. *An. Acad. Brasil. Ci*., **59**; 71-81.

Morse. R.J., Yamamoto.T. & Stroud. R.M. (2001) Structure of Cry2Aa suggests an unexpected receptor binding epitope. *Structure,* **9**; 409-417.

Murshudov. G. N., Vagin. A. A. & Dodson. E. J., (1997). Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr*. **D53**; 240-255.

Murzin A. G., Lesk. M. A. & Chothia. C., (1992). ß-trefoil fold: Patterns of structure and sequence in the Kunitz inhibitors interleukins-1ß and 1a and fibroblast growth factors. *J. Mol. Biol*. **223**; 531-543.

Navaza. J., (1994). AMoRe: An automated package for molecular replacement. *Acta Crystallogr*., **A50**; 157-163.

Navaza. J., (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr*., **D57**; 1367-1372.

Navaza. J. & Saludjian. P., (1997). AMoRe: An automated molecular replacement program package. *Methods Enzymol*., **276**; 581-594.

Nicolson. G. L., (1976). Trans-membrane control of the receptors on normal and tumor cells. II Surface changes associated with transformation and malignancy. *Biochim. Biophys. Acta*., **458**; 1-72.

Northrup., S. H. & Erickson. H. P., (1992). Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci*. (USA)., **89**; 3338-3342.

Odani. S. & Ikenaka. T., (1978). Studies on soybean trypsin inhibitors: Linear sequences of two soybean double headed inhibitors, D-II and E-I. *J. Biochem*., **83**; 737-745.

Ofek. I. & Sharon. N., (1988). Lectinophagocytosis: a molecular mechanism of recognition between cell surface sugars and lectins in the phagocytosis of bacteria. *Infect. Immun*., **56**; 539-547.

Oliviera. P. S. L., Garratt. R. C., Mascarenhas. Y. P., Beltramini. L. M., Roque-Barreira. Maria. C., Harvey. I. & Oliva. G., (1997). Crystallization and preliminary crystallographic data of a neutrophil migration-inducing lectin (KM+) extracted from the seeds of *Artocarpus integrifolia*. *Proteins: Struct. Funct. Genet*., **27**; 157-159.

Osborn. R. W. & Broekaert. W. F., (1999). Antifungal proteins. In: *Seed proteins*., Shewry. P. R & Casey. R., eds., Kluwer academic publishers, Dordercht, Holland, pp. 727-751.

Osborne, T. B., (1924). In: *The vegetable proteins*., Longmans, Green & Co. eds., **Vol. 2**. London, pp. 154.

Otwinowski. Z., (1993). DENZO. An oscillation data processing program for macromolecular crystallography. Yale University, New Haven, CT, USA.

Otwinowski. Z & Minor. W., (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*., **276**; 307-326.

Padma. P., Komath, S. S., Nadimpalli. S. K. & Swamy. M. J., (1999). Purification in high yield and characterization of a new galactose-specific lectin from the seeds of *Trichosanthes cucumerina*. *Phytochemistry*., **50**; 363-371.

Pannu. N. S. & Read. R. J., (1996). Improved structure refinement through maximum likelihood. *Acta Crystallogr*., **A52**; 659.

Parkkinen. J., Finne. J., Achtman. M., Vaisanen.V. & Korhonen.T., (1983). *E. coli* strains binding neuraminyl a-2-3 galactosides. *Biochem. Biophys. Res. Commun*., **111**; 456-461.

Patterson. A. L., (1934). A Fourier series method for the determination of the componenets of interatomic distances in crystals. *Phys. Rev*., **46**; 372-376.

Patterson. A. L., (1935). A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift fur Kristallographie*., **A90**; 517-542.

Pavone. V., Isernia. C., Saviano. M., Falcigno. L., Lombardi. A., Paolillo. L., Pedone. C., Buogen. S., Naess. H. M., Revheim. H. & Eriksen. J. A., (1994). Conformational studies on peptides as enzyme inhibitors: chymotrypsin inhibitors using Bowman-Birk type as models. *J. Che. Soc. Perkin. Trans*., **2**; 1047-1053.

Pellegrineschi. A., (1997). *In vitro* plant regeneration via roganogenesis of cowpea [*Vigna unguiculata* (L.) Walp.]. *Plant cell Rep*., **17**; 89-95.

Petri. W. A. Jr., Chapman. M. D., Snodgrass. T., Mann. B. J., Broman. J. & Ravdin. J. I., (1989). Subunit structure of the galactose and N-acetyl -D-galactosamine-inhibitable adherence lectin of *E. histolytica*. *J. Biol. Chem*., **264**; 3007.

Peumans. W. J. & Van Damme. E. J. M., (1998). Plant lectins; Specific tools for the identification, isolation and characterization of O-linked glycans. *Critic. Rev. Biochem & Mol Biol*., **33**; 209-258.

Peumans. W. J. & Van Damme. E. J. M., (1999). Seed lectins., In; *Seed proteins*., Shewry. P. R. & Casey. R., eds., Kluwer Academic publishers, The Netherlands, 657-683.

Peumans. W. J., Verhaert. P., Pfuller. U. & Van damme. E. J. M., (1996). Isolation and characterization of a small chitin binding lectin from misteltoe (*Viscum album*). *FEBS. letters*., **396**; 261-265.

Peumans. W. J., Hause. B. & Van damme. E. J. M., (2000). The galactose-binding and mannose-binding jacalin-related lectins are located in different sub-cellular compartments. *FEBS. letters*., **477**; 186-192.

Philipp. S., MiKim. Y., Durr. I., Wenzl. G., Vogt. M. & Flecker. P., (1998). Mutational analysis of disulfide bonds in the trypsin reactive subdomain of a Bowman-Birk inhibitor of trypsin and chymotrypsin: Cooperative versus autonomous refolding of subdominas. *Eur. J. Biochem*., **251**; 854-862.

Pichare. M. M. & Kachole. M. S., (1994). Detection of electrophoretically separated protease inhibitors using X-ray film. *J. Biochem. Biophys. Methods*., **28**; 215-24.

Porath. J. & Foldin. P., (1959). Gel filtration; A method for desalting and group separation. *Nature*., **83**; 1657-1659.

Prakash. B., Murthy. M. R. N., Sreerama. Y. N., Sarma. P. R. & Rao. D. R., (1994). Crystallization and X-ray diffraction studies on a trypsin/chymotrypsin double-headed inhibitor from horse gram. *J. Mol. Biol*., **235**; 364-366.

Prakash. B., Selvaraj. S., Murthy. M. R .N., Sreerama. Y. N., Rao. D. R., & Gowda. L. R., (1996). Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. *J. Mol. Evol*., **42**; 560-569.

Pratap. J. V., Jeyaprakash A. A., Rani. P. G., Sekar. K., Surolia. A., Vijayan. M., (2002). Crystal structures of Artocarpin, a *Moraceae* Lectin with mannose specificity, and its complex with methyl--D-mannose: Implications to the generation of carbohydrate specificity. *J. Mol. Biol*., **317**; 237-247.

QUANTA (1997). X-ray structure analysis user's reference. San Diego: Molecular simulations.

Quillien. L., Ferrason. E., Molle. D. & Gueguen. J., (1997). Trypsin inhibitor polymorphism: Multigene family expression and post translational modification. *J. Pro. Chem*., **16**; 195-203.

Raikhel. N. V., Lee. H. I. & Borekaert. W. F., (1993). Structure and function of chitin-binding proteins. *Ann. Rev. Plant Physiol. & Plant Mol. Biol*. **44**; 591-615.

Rabijns. C. A., Novoa de Armas. V. H., Van Damme. E. J. M., Peumans. W. J. & De Ranter. C. J., (2001). The crystals of a mannose-specific jacalin-related lectin from *Morus nigra* are merohedrally twinned. *Acta Crystallogr*. **D57**, 609-611.

Ramachandran. G. N. & Sasishekaran. V., (1968). Conformation of polypeptides and proteins. *Adv. Protein. Chem*., **23**; 283-438.

Rao. K. N. Gurjar. M. M., Gaikwad. S. M., Khan. M. I. & Suresh. C. G., (1999a). Crystallization and preliminary X-ray studies of the basic lectin from the seeds of *Artocarpus hirsuta*. *Acta.Crystallogr*. **D55**, 1204-1205.

Rao. K. N. Hegde. S. S., Lewis. R. J., & Suresh. C. G., (1999b). Crystallization and preliminary X-ray diffraction studies of a Bowman-Birk inhibitor from the *Vigna unguiculata* seeds. *Acta.Crystallogr*. **D55**, 1920-1922.

Ravishankar. R., Ravindran. M., Suguna. K., Surolia. A. & Vijayan. M., (1997). Crystal structure of peanut lectin-T-antigen complex: carbohydrate specificity generated by water bridges. *Curr. Sci*., **72**; 855-861.

Raz.A. & Lotan. R., (1981). Lectin-like activities associated with human and murine neoplastic cells. *Cancer. Res*., **41**; 3642-3647.

Read. R. J., (1990). Structure-factor probabilities for related structures. *Acta Crystallogr*., **A46**; 900-912.

Read. R. J., (1997). Model phases: Probabilities and bias. *Methods Enzymol*., **277**; 110-127.

Read. R. J. & James. M. N. G., (1986). Introduction to protein inhibitors: X-ray crystallography. In: *Proteinase inhibitors*., Barret. A. J., & Salvensen. G., eds., Elsevier, Amsterdam, pp. 301-335.

Reisner. Y. & Sharon. N., (1984). Fractionation of subpopulations of mouse and human lymphocytes by peanut agglutinin or soybean agglutinin. *Methods Enzymol*., **108**; 168-179.

Reisner. Y., Ravid. A. & Sharon. N., (1976). Use of soybean agglutinin for the separation of mouse B and T lymphocytes. *Biochem. Biophys. Res. Commun*., **72**; 1585-1591.

Rele. M. V., Vartak. H. G., & Jagannathan. V., (1980). Protease inhibitors from *Vigna unguiculata* subsp. cylindrica. *Arch. Biochem. Biophy*., **204**; 117-140.

Rhodes. J. M. & Milton. J. D., (1998). In : *Lectin methods and protocols*., Humana press Inc., Totowa, N. J. pp. 616.

Richardson. J. S., (1981). The anatomy and taxonomy of protein structure. *Adv. Protein. Chem*., **34**; 167-339.

Richardson. M., (1977). The proteinase inhibitors of plants and microorganisms. *Phytochemistry*., **16**; 159-169.

Richardson. M., (1991). Seed storage proteins: the enzyme inhibitors. *Methods Plant Biochem*., **5**; 259-305.

Rinderle. S. J., Goldstein. I. J. & Remsen. E. E., (1990). Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds. *Biochemistry*., **29**; 10555-10561.

Rini. J. M., (1995). Lectin structure. *Ann. Rev. Biophys. Biomol. Struct.*, **24**; 551-557.

Rini. J. M., (1999). New animal lectin structures. *Curr. Opin. Struct. Biol*., **9**; 578-584.

Roberts. D. D. & Goldstein. I. J., (1983). Binding of hydrophobic ligands to plant lectins: titration with arylamino-napthalenesulfonates. *Arch. Biochem. Biophys*., **224**; 479-484.

Roche. A. C., Barzilay. M., Midoux. P., Junqua. S., Sharon. N. & Monsigny. M., (1983). Sugar-specific endocytosis of glycoproteins by Lewis lung carcinoma cells. *J. Cell. Biochem*., **22**; 131-140.

Rosa. J. C., De Oliverira. P. S. L., Garratt, R., Beltramini, L., Resing. K., Rpque-Barreira. M. C. & Greene. L. J., (1999). KM+, a mannose-binding lectin from *Artocarpus integrifolia*: amino acid sequence, predicted tertiary structure, carbohydrate recognition, and analysis of the ß-prism fold. *Prot. Sci*., **8**; 13-24.

Rossmann. M. G., (2001). Molecular replacement–historical background. *Acta Crystallogr*., **D57**; 1360-1366.

Rossmann. M. G. & Blow. D. M., (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr*., **15**; 24-31.

Rossmann. M. G. & Blow. D. M., (1963). Determination of phases by the conditions of non-crystallographic symmetry. *Acta Crystallogr*., **10**; 24-31.

Rossmann. M. G. & Blow. D. M., (1964). The relative positions of independent molecules within the same asymmetric unit. *Acta Crystallogr*., **17**; 338-342.

Rouge. P., Cambillu. C. & Bourne. Y., (1991). The three-dimensional structure of legume lectins. In: *Lectin Reviews*., Kilpatrick. D. C., Van Driessche. E. & Bog-Hansen. T. C. eds., **Vol.1**, Sigma Chemical co., St.Louis, USA, pp.143-159.

Rutenber. E. & Robertus. J. D., (1991). Structure of Ricin B-chain at 2.5 Å resolution. *Proteins: Struc. Funct. Genet*., **10**; 260-269.

Ryan. C. A., (1973). Proteolytic enzymes and their inhibitors in plants. *Annu. Rev. Plant Physiol*., **24**; 173-196.

Ryan. C. A., (1981). Protease inhibitors, a comprehensive treatise. In: *The Biochemistry of Plants*. Stumpf. P. K., & Conn. E. E., eds., **Vol.6**, Academic Press. N.Y., pp.351-371.

Ryan. C. A., (1990). Protease inhibitors in plants: genes for improving defense against insects and pathogens. *Annu. Rev. Phytopathol*., **28**; 425-449.

Sadasivan. C., Nagendra. H. G. & Vijayan. M., (1998). Plasticity, hydration and accessibility in Ribonuclease A. The structure of a new crystal form and its low humidity content. *Acta. Crystallogr*., **D54**; 1343-1352.

Sairam. M. R., (1989). Role of carbohydrates in glycoprotein hormone signal transduction. *FASEB. J*., **3**; 1915-1926.

Sakal. E., Applebaum. S. W. & Birk. Y., (1989). Purification and characterization of trypsins from the digestive tract of *Locusta migratoria*. *Int. J. Pept. Prot. Res*., **34**; 498-505.

Sankaranarayanan. R., Sekar. K., Banerjee. R., Sharma. V., Surolia. A. & Vijayan. M., (1996). A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a ß-prism fold. *Nat. Struct. Biol*., **3**; 596-603.

Sauerborn, M. K., Wright, L. M., Reynolds, C.D., Grossmann, J. G., Rizkallah, P. J., (1999). Insight into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: the crystal structure at 2 ? resolution in complex with a-1-3 mannobiose, *J. Mol. Biol*., **290**, 185-199.

Schetchter. I. & Berger. M., (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun*., **27**; 157-162.

Schmidt. R. R., (1987). In: *Stereochemistry of Organic and Bioorganic Transformations*, Barman.W. & Sharpless. K. B., eds. VCH., Verlagsgesellschaft, Weinhiem, Federal republic of Germany., pp.169-189.

Schreiber. G. & Fersht. A. R., (1996). Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol*., **3**; 427-431.

Schuler. T. H., Poppy. G. M., Kerry. B. R. & Denholm. I., (1998). Insect-resistant transgenic plants. *Trends Biotechnol*., **16**; 168-175.

Schultze. B., Zimmer. G. & Herrlere. G., (1993). Viral lectins for the detection of 9-O-Actylated Sialic acid on glycoproteins and glycolioids. In: *Lectins and Glycobiology*., Gabius. H. J. & Gabius. S. eds., Springer-Verlag Berlin, Heidelberg., pp. 158-174.

Scopes. R. K. & Cantor. C. R., (1982). Protein purification: principles and practice. Springer-Verlag; N. Y., pp. 1-182.

Sela. B. A., Lis. H., Sharon. N. & Sachs. L., (1970). Different locations of carbohydrate containing sites in the surface membrane of normal and transformed mammalian cells. *J. Memb. Biol*., **3**; 267-279.

Sharma. V. & Surolia. A., (1997). Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary activity. *J. Mol. Biol*., **267**; 433-445.

Sharon. N., (1987). Bacterial lectins, cell-cell recognition and infectious diseases. *FEBS. letters*., **217**; 145-157.

Sharon. N. & Lis. H., (1989). Lectins as cell recognition molecules. *Science*., **246**; 227-234.

Sharon. N. & Lis. H. (1990). Legume lectins –a large family of homologous proteins. The *FASEB. J*., **4**; 3198-3208.

Sharon. N. & Lis. H., (1993). Carbohydrates in cell recognition. *Scientific*. *American*, **Jan**; 74-80.

Shewry. P. R., (1999). Enzyme inhibitors of seeds: types and properties. In: *Seed proteins*., Shewry. P. R & Casey. R., eds., Kluwer academic publishers, Dordercht, Holland, pp. 587-616.

Shewry. P. R., & Casey. R., (1999). Seed proteins. In: *Seed proteins*., Shewry. P. R & Casey. R., eds., Kluwer academic publishers, Dordercht, Holland, pp. 1-10.

Shewry. P. R. & Lucas. J. A., (1997). Plant proteins that confer resistance to pest and pathogens. *Adva. Bot. Res*., **26**; 135-192.

Shimuzu. T., Vassylyev. D.g., Kido. S., Doi. Y. & Morikawa. K., (1994). Crystal structure of the vitelline membrane outer layer protein I (VMO-I): a folding motif with homologous Greek key structures related by an internal 3-fold symmetry. *EMBO. J*., **13**; 1003-1010.

Singh. R. S., Tiwary. A. K. & Kennedy. J. F., (1999). Lectins: Sources, Activities, and Applications. *Critic. Rev. Biotech*., **19**; 145-178.

Smith. J. J. & Raikhel. N. V., (1989). Nucleotide sequences of cDNA clone encoding wheat germ agglutinin isolectins A and D. *Plant Mol. Biol*., **13**; 601-603.

Song. H. K., Kim. Y. S., Yang. J. K., Moon. J., Lee. J. Y. & Suh. S. W., (1999). Crystal structure of a 16kDa double headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *J. Mol. Biol*., **293**; 1133-1144.

Springer. T. A., (1995). Traffic signals on endothelium for lymphocyte recirculation and leukocyte emigration. *Ann. Rev. Physiol*., **57**; 827-872.

Spriner. W. R., Cooper. D. N. W. & Barondes. S. H., (1984). Discoidin I is implicated in cell-substratum attachment and ordered cell migration of *Dictyyostelium discoideum* and resembles fibronectin. *Cell*., **39**; 557-564.

Sreerama. Y. N. & Gowda. L. R., (1997). Anitgenic determinants and reactive sites of a trypsin/chymotrypsin double-headed inhibitor form horse gram (*Dolichos biflorus*). *Biochim. Biophy. Acta*., **1343**; 235-242.

Srinivas. V. R., Reddy, G. B., Ahmad. N., Swaminathan. C. P., Mitra. N. & Surolia. A., (2001). Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta*., **1527**; 102-111

Srinivasan. N., Rufino. S. D., Pepys. M. B., Wood. S. P. & Blundell. T. L., (1996). A superfamily of proteins with the lectin fold. *Chemtracts-Biochem & Mol Biol*., **6**; 149-164.

Stillmark. H., (1888). Uber Ricin ein giftiges Fermentaus den Samen von Ricinus communis. L. und einige anderen Euphorbiaceen. Inagural Disseration Dorpat.

Stojanovie. D. & Hughes. R. C., (1984). An endogenous carbohydrate binding agglutinin of BHK cells. Purification, specificity and interaction with normal and ricin resistant cell lines. *Biol. Cell*., **51**; 197-206.

Sumar. N., Bodman. K. B. & Rudo. P. M., (1993). Lectins as indicators of disease associated glycoforms. In: *Lectins and Glycobiology*., Gabius. H. J. & Gabius. S. eds., Springer-Verlag Berlin, Heidelberg. pp 158-174.

Suzuki. A., Yamane. T., Ashida. T., Norioka. S., Hara. S. & Ikenaka. T., (1993). Crystallographic refinement of Bowman-Birk type protease inhibitor AII from peanut (*Arachis hypogeae*) at 2.3 Å resolution. *J. Mol. Biol*., **234**; 722-734.

Tahirov. T. H., Lu. T., Liaw. Y., Chen. Y. & Lin. J., (1995). Crystal structure of Abrin-a at 2.14 Å . *J. Mol. Biol*., **250**; 354-367.

Tan-Wilson. A. L. & Wilson. K. A., (1986). In: *Nutritional significance and toxicological significance of enzyme inhibitors in food*., **99**; Freidman. M. ed., Plenum press, N. Y. pp. 391-411.

Terada. S. Sato. K., Kato. S. & Izumiya. N., (1980). Studies on the synthesis of proteinase inhibitors. II. Synthesis of cyclic nanopeptide fragments and analogs related to the reactive sites of soybean Bowman-Birk inhibitor. *Int. J. Pept. Protein. Res*., **15**; 441-454.

Thornton. J. M., (1981). Disulfide bridges in globular proteins. *J.Mol.Biol*.,**151**;261-87.

Tollin. P., (1966). On the determination of molecular location. *Acta Crystallogr*., **21**; 613-614.

Transue. T. R., Smith. A .K., Mo. H., Goldstein. I. J. & Saper. M. A., (1997). Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* lectin. *Nat. Struct. Biol*., **10**; 779-783.

Tsunogae. Y., Tanaka. I., Yamane. T., Kikkawa. J., Ashida. T., Ishikawa. C., Watanabe. K., Nakamura. S. & Takahashi. K., (1986). Structure of the trypsin-binding domain of Bowman-Birk type protease inhibitor and its interaction with trypsin. *J. Biochem* (Tokyo)., **100**; 1637-1646.

Ussuf. K. K., Laxmi. N. H. & Mitra. R., (2001). Proteinase inhibitors: Plant-derived genes of insecticidal protein for developing insect-resistant transgenic plants. *Curr. Sci*., **80**; 847-853.

Valueva. T. A. & Mosolov., (1999). Protein inhibitors of proteinases in seeds: 1. Classification, distribution, structure, and properties, 2. Physiological functions., *Russian J. Plant physiol*., **46**; 362-387.

Van Damme. E. J. M. & Peumans. W. J., (1991). Biosynthesis, primary structure and molecular cloning of snowdrop lectin. *Eur. J. Biochem*., **202**; 23-30.

Van Damme. E. J. M., Allen. A. K. & Peumans. W. J., (1987). Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS. Letters*. **215**; 140-144.

Van Damme. E. J. M., Peumans. W. J., Barre. A. & Rouge. P., (1998). Plant lectins: a composite of several distinct families of structurally and evolutionarily related proteins with diverse biological roles. *Crit. Rev. Plant. Sci.*, **17**; 575-692.

Van Loon L. C., (1985). Pathogenesis related proteins. *Plant Mol. Biol*., **4**; 111-116.

Vartak., H. G., Rele. M. V. & Jagannathan. V., (1980). Proteinase inhibitors from *Vigna unguiculata* susp. cylindica. *Arch. Bichem. Biophy*., **204**; 134-140.

Vellieux. F. M. D. & Read. R. J., (1997). Non-crystallographic symmetry averaging in phase refinement and extension. *Methods Enzymol*., **277**; 18-52.

Vesterberg. O., Hansen. L. & Sjosten. A., (1977). Staining of proteins after isoelectric focusing in gels by new procedures. *Biochem. Biophy. Acta*., **491**; 160-166.

Vierling. E., (1991). The roles of heat sock proteins in plants. *Ann. Rev. Plant Physiol & Plant Mol. Biol*., **42**; 579-620.

Vijayan. M. & Chandra. N., (1999). Lectins. *Curr. Opin. Stru. Biol*., **9**; 707-714.

Voss. R. H., Ermler. U., Essen. L., Wenzl. G., Kim. Y. & Flecker. P., (1996). Crystal structure of the bi-functional soybean Bowman-Birk inhibitor at 0.28-nm resolution, structural peculiarities in a folded conformation. *Eur. J. Biochem*., **242**; 122-131.

Wang. M. B., Boulter. D. & Gatehouse. J., (1994). Characterization and sequencing of cDNA clone encoding for phloem protein PP2 of *Cucurbita pepo*. *Plant Mol. Biol*. **24**; 159-170.

Weber. K. & Osborn. M., (1969). The reliability of molecular weight determination by dodecyl sulphate-polyacrylamide gel electrophoresis. *J. Biol. Chem*., **244**; 4406-4412.

Weis. W. I., (1997). Cell surface carbohydrate recognition by animal and viral lectins. *Curr. Opin. Struct. Biol*., **7**; 624-630.

Weis. W. I. & Drickamer. K., (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem*., **65**; 441-473.

Weis. W. I., Drickamer. K. & Hendrickson. W. A., (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*., **360**; 127-134.

Weis. W. I., Taylor. M. E., & Drickamer. K., (1998). The C-type lectin super family in the immune system. *Immunol. Rev*., **163**; 19-34.

Werner. M. H. & Wemmer. D. E., (1992a). Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution. *Biochemistry*., **31**; 999-1010.

Werner. M. H. & Wemmer. D. E., (1992b). Identification of a protein binding surface by differential amide hydrogen-exchange measurements. *J. Mol .Biol*., **225**; 873-889.

Wilson. K. A., (1981). The structure, function and evolution of legume proteinase inhibitors. In: *Antinutrients and natural toxicants in foods*., Ory. R. L. eds., Food Nutr. Press., Westport., Conn., pp. 187-202.

Wlodawer. A., Deisenhofer. J. & Huber. R., (1987). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol*., **193**; 145-156.

Wood. S. D., Wright. L. M., Reynolds. C. D., Rizkallah. P. J., Allen. A. K., Peumans W. J. & Van Damme. E. J. M., (1999) Structure of the native (unligated) mannose-specific bulb lectin from *Scilla Campanulata* (bluebell) at 1.7Å resolution. *Acta Crystallogr*. **D55**; 1264-1272.

Wright. C. S., (1987). Refinement of the crystal structure of the wheat germ agglutinin isolectin2 at 1.8 Å resolution. *J. Mol. Biol*., **194**; 501-529.

Wright C. S., (1989). Comparison of the refined crystal structures of two wheat germ isolectins. *J. Mol. Biol*., **209**; 475-487.

Wright. C. S. (1990) 2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose-Wheat germ agglutinin isolectin complexes. *J. Mol. Biol*., **215**; 635-651.

Wright. C. S., (1992). Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. *J. Biol. Chem*., **267**; 14345-14352.

Wright. C. S., (1997). New folds of plant lectins. *Curr. Opin. Struct. Biol*., **7**; 631-636.

Wright. C. S. & Hester. G., (1996). The 2.0 ? structure of a cross-linked complex between snowdrop lectin and a branched mannopentose: evidence from two unique binding modes. *Structure*., **4**; 1339-1352.

Wright. C. S. & Jaeger. J., (1993). Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorinA. *J. Mol. Biol*., **232**; 620-638.

Wright. L. M., Wood. S. D., Reynolds. C. D., Rizkallah. P. J., Van Damme. E. J. M. & Peumans. W. J., (1997). Crystallization and preliminary X-ray analysis of a novel plant lectin from *Calystegia sepium*. *Acta Crystallogr*., **D53**; 220-221.

Wu. J. A., Herp. A., Chow. L. P. & Lin. J. Y., (2001). Carbohydrate specificty of a toxic lectin, abrinA, from the seeds of *Abrus precatorius* (jequirity bean). *Life Sci*., **69**; 2027-2-38.

Yavelow. J., Collins. M., Birk. Y., Troll. W. & Kennedy. A. R., (1985). Nanomolar concentrations of Bowman-Birk soybean protease inhibitor supress X-ray induced transformation *in vitro*. *Proc. Natl. Acad. Sci.* (USA)., **82**; 5395-5399.

Ye. X. Y., Ng. T. B. & Rao. P. F., (2001). A Bowman-Birk trypsin-chymotrypsin inhibitor from broad beans. *Biochem. Biophys. Res. Comun*., **289**; 91-96.

Young. M. N., Johnston. R. A. Z., Szabo. A. G. & Watson. D. C., (1989). Homology of the D-galactose specific lectins from *Artocarpus intgrifolia* and *Maclura pomifera* and the role of an unusual small polypeptide subunit. *Arch. Bichem. Biophy*., **270**; 596-603.

APPENDIX 1

Crystallization and crystal characterization of the enzyme Pencillin V acylase.

The initial crystallization experiments carried out on the enzyme Penicillin V acylase (PVA) from *Bacillus sphaericus* is described in this part of the appendix. Penicillin G acylase (PGA) and Penicillin V acylase are two industrially used enzymes, for the cleavage of penicillin G and penicillin V, respectively, to yield 6-Aminopenicillanic acid (6-APA), the precursor for the preparation of semi-synthetic penicillins. These two enzymes account for all the enzymatic production of 6-APA by pharmaceutical industry (Valle *et al.*, 1991). The PGA and PVA differ widely in their molecular properties. PVA from *B. sphaericus* is a tetramer of Mr. 148 kDa (Glsson &Uhlen, 1986) while PGA from *E. coli* is a heterodimer of Mr. 90 kDa (Dugglely *et al.*, 1995). The two enzymes show no sequence similarity (Glsson & Uhlen, 1986; Schumacher, 1986). The PGA structure was already known (Dugglely *et al.*, 1995) when the crystallization experiments on PVA were initiated.

After initial crystallization trials on the purified sample of PVA using hanging drop method, the best conditions chosen for crystallization based on the preliminary experiments consisted of phosphate buffer in the pH range 6.4 to 6.8 and 28 – 30% saturated ammonium sulfate as precipitant. The protein was known to contain free cysteins and hence 5 mM dithiothreitol (DTT) was maintained in the crystallization solution. The concentration of protein solution used was 15 mg/ml. The diffraction data was collected on a MAR image plate mounted on a Rigaku rotating anode at IISc. Bangalore. The MARXDS was used for processing the images.

Two types of crystals could be characterized during data collection. The first one gave a hexagonal unit cell of dimensions a= b= 208.6, c=96.2 Å, a= ?= 90, ß=120. The second one, on indexing the images, gave a monoclinic unit cell of dimensions a=135.0, b=132.5, c=54.0 Å, a= ?= 90, ß= 98.2. The details of the data collection are given in table A1-1. The hexagonal crystals (Fig. A1-1.) were of poor quality compared to monoclinic. However, hexagonal crystals were more stable and used to grow consistently in all experiments. Subsequently the

quality of hexagonal crystals were improved and they were used in structure solution and refinement of the structure.

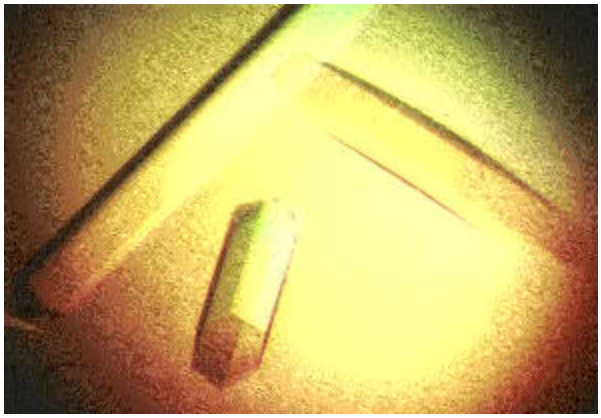


Fig. A1-1. The hexagonal crystals of PVA.

Table A1-1. The data collection statistics on PVA.

Description of parameter	**Hexagonal**	**Monoclinic**
Crystal to detector distance (mm)	250	200
Oscillation width (degrees)	1	1
Exposure time (seconds)	1000	1000
Number of images	54	141
Space group used for processing	P6	C2
Resolution range(Å)	20-5	20-3
Completeness (%)	87	96
Rmerge (%)	24.0	12.4

The crystal structure of this enzyme was solved using Multiple Isomorphous Replacement method using two (Mercury and Platinum) heavy atom derivatives. The structure was refined using synchrotron data at 2.5 Å resolution (Suresh *et al.*, 1999). Structure determination established that there existed similarities between PGA and PVA at structural level and both belonged to recently discovered (Brannigan *et al.*, 1995) Ntn (N-terminal nucleophile) hydrolase family.

References:

Brannigan. J. A., Dodson. G., Duggley. H. J., Moody. P. C. E., Smith. J. L., Tomchick. D. R. & Murzin. A. G., (1995). A protein catalytic framework with an N-terminal nucleophile capable of self activation. *Nature*., **378**; 416-419.

Dugglely. H.J., Tolley S.P., Hill C.P, Dodson, E.J. Dodson .G. & Moody, C.E. (1995). Penicillin acylase has a single-amino-acid catalytic centre. *Nature*, **373**,264-268.

Glsson. A. & Uhlen. M., (1986). Sequencing and heterologous expression of the gene encoding pencillin v amidase from Bacillus sphericus. *Gene*., **45**; 175-181.

Schumacher. G., Sizmann. D., Hang. H., Buckel. P. & Bock. A., (1986). Pencillin acylase from *E.coli*; unique gene-protein relation. *Nucl. Acids Res*., **14**; 5713-5727.

Suresh. C.G., Pundle. A. V., Rao. K. N., SivaRaman. H., Brannigan. J.A., McVey. C.E., Verma. C.S., Dauter. Z., Dodson. E.J., & Dodson. G.G. (1999) Penicillin V acylase crystal structure reveals new Ntn-hydrolase family member. *Nature Struct. Biol.*, **6**; 414-416.

Valle. F., Balbas. P., Merino. E. & Bolivar. F., (1991). The role of penicillin amidases in nature and industry. *Trends Biochem. Sci.*, **16**; 36-40.

APPENDIX 2

Structural studies on the C-2 substitution in a new set of synthetic dideoxyaminosugars; the steric bulk at C-2 influences the puckering of the pyranose ring.

Here we report on the structural investigations of three phenylsulfonyl-modified monosaccharides with amino substitution at C-2. Two of them are a pair of ß- and a-anomers of C-2 isobutyl amino substituted pyranose derivatives (2aa & 2ßa) and the third is a ß-anomer of C-2 tertbutylamino substituted one (2ßc). The pyranose ring in the a, ß pair assumed favourable chair conformation whereas the pyranose ring in the third ß-anomer had a strained boat conformation. Moreover, the synthetic pathway did not achieve the corresponding C-2 addition in a-anomer. It is possible that the addition of bulk groups at C-2 position is influenced by the disposition of the methoxy group at C1, however, further studies will be required to confirm this.

The crystals of 2aa and 2ßc contained one molecule each in the asymmetric unit of their unit cell, whereas 2ßa contained two. The difference between the two molecules in the asymmetric unit of 2ßa is mainly due to the differences in mutual orientations of the two phenyl rings in them, angles 53.2(3) and 73.9(3) degrees for Molecule A and B, respectively (Fig. A2. 2. A and A2. 2. B). When pyranose rings of the three crystallographically independent molecules of 2aa and 2ßa were compared, the differences were found insignificant in the average of either C-C bond length or angles around ring carbon atoms. In case of 2ßc the average values of pyranose ring carbon angles and O5 angle had comparatively higher values. The pyranose ring conformation of 2aa (Fig. A2. 1) is similar to that reported for the structure of 2ab [1]. Similarly the sugar puckering of 2ßc (Figure. A2. 3) closely resembles that of the reported one for 2ßb [1]. However, the pyranose of 2ßa in its structure is in chair conformation. In all these structures the addition of amino group at C-2 is at the equatorial position. Thus in ß-anomers the vinyl sulfone, amino and methoxy groups at C-3,2,1 positions are all equatorial. Even then in cases where bulkier groups are present at C-2, such as in 2ßa and 2ßc, the pyranose ring takes boat conformation to facilitate positioning the methoxy group of C-1 away from the amino group of C-2. The observation of increasing values of

conformational angles O1-C1-C2-N1 in the order 2ßb→ 2ßa →2ßc in their respective ß-anomeric structures (Table 1,[1]) confirms this argument. However, in the case of a-anomers, the mechanism by which to increase this particular conformational angle is limited due to the axial positioning of methoxy group at C-1. Thus a bulkier group, greater than a critical size, could not be accommodated at C-2 position in a-anomer, in the absence of a mechanism to release the resultant strain due to steric hindrance. This could be an explanation for the non-formation of 2ac when the synthesis was started with 1a whereas product formation of 2ßc took place in synthesis with 1ß.

Crystallographic data

Compound (2aa) - $C_{24}H_{31}NO_6S$, M_w = 461.56, monoclinic, $P2_1$, a = 11.297 (2), b = 8.534 (2), c = 13.405 (4) Å, ß = 114.71 (2) °, V = 1174.0 (5) $Å^3$, D_c= 1.306 Mg m^3, Z = 2, µMo = 0.177 mm $^{-1}$, Crystal size 0.56 X 0.34 X 0.20 mm, ?$_{max}$ 24.96 °, Out of 2207 independent reflections, 2123 were with I>2sigma(I), R_{merge} 0.0911. The final residuals were wR(F^2) 0.1020 for all data and R[F^2 > 2s(F^2)] 0.0373, Flack parameter[2]=0.09(10). Data were collected on an Enraf{Nonius CAD-4 diffractometer with ? -2? scans, using Mo Ka radiation (? = 0.71069 Å).

Compound (2ßa) - $C_{24}H_{31}NO_6S$, M_w=461.56, Triclinic, P1, a = 9.476 (3), b=10.010 (3), c=13.935 (3)Å, a=72.47 (2), ß=85.66 (2), ?=70.98 (2)°, V =1191.2 (6) $Å^3$, Z = 2, Dc=1.287 Mg m^{-3}, µ Mo=0.175 mm $^{-1}$, crystal size 0.80 X 0.70 X 0.45 mm, θ_{max}=24.99°, Out of 4114 independent reflections 4223 reflections with I>2sigma. The final residuals were wR(F^2)=0.1540 for all data and R[F^2>2s(F^2)] = 0.0563, Flack parameter[2]=-0,01(10). Data were collected on an Enraf Nonius CAD-4 diffractometer with ?-2? scans, using Mo Ka radiation (?=0.71069 Å).

Compound (2ßc) -$C_{24}H_{31}NO_6S$, M_w= 461.56, Triclinic, P1, a = 8.123 (5), b = 12.485 (6), c=5.9460 (11)Å, a=92.95(3), ß = 94.96 (4), ?=84.23(5)°, V = 597.2(5) $Å^3$, Z = 1, Dc=1.283 Mg m^{-3}, µ Mo=0.174 mm $^{-1}$, crystal size 0.80 X 0.20 X 0.12 mm, ? $_{max}$=32.43°. Out of 4313 independent reflections 2814

reflections with I >2sigma(I). The final residuals were wR(F^2)=0.1372 for all data and R[F^2 > 2s(F^2)]=0.0494, Flack parameter[2]=0.08(9). Data were collected on AFC-7S diffractometer with ? -2? scans, using Mo Ka radiation (? = 0.71069 Å).

H-atoms were located using geometrical considerations and difference Fourier map. They were treated as riding on the heavier atoms to which they were attached. For compounds 2aa and 2ßa, data collection and cell refinement: CAD-4- PC Software [3], data reduction: NRCVAX DATRD2 [4]; for compound 2ßc data collection, cell refinement and data reduction: MSC/AFC Diffractometer Control Software [5]. For all the compounds, programs to solve structure: SHELXS86 [6]; program(s) used to refine structure: SHELXL93 [7]; molecular graphics: PLATON [8]

TABLE 1

Selected conformational angles in the three structures. The six endocyclic conformation angles of pyranosyl ring are in bold.

Conformation angle	(2aa)	(2ßa) Molecule A	(2ßa) Molecule B	(2ßc)
C5-O5-C1-C2	**-57.8 (3)**	**-65.6 (5)**	**-61.9 (5)**	**32.9 (4)**
O5-C1-C2-C3	**-42.0 (3)**	**-47.9 (5)**	**-49.6 (4)**	**16.1 (3)**
C5-O5-C1-O1	65.8 (3)	174.8 (4)	179.6 (4)	-90.4 (3)
O5-C1-O1-C7	73.1 (3)	-78.5 (7)	-79.9 (7)	-62.0 (4)
O1-C1-C2-N1	43.0 (3)	-61.8 (6)	-65.2 (5)	-166.4 (3)
O5-C1-C2-N1	167.4 (2)	179.5 (4)	176.7 (4)	68.3 (3)
C1-C2-C3-C4	**44.0 (3)**	**56.1 (5)**	**52.7 (5)**	**-56.2 (3)**
N1-C2-C3-S1	74.6 (3)	64.7 (4)	67.7 (4)	126.0 (3)
C15-S1-C3-C4	73.9 (2)	48.1 (3)	60.6 (3)	61.4 (3)
C2-C3-C4-O4	172.6 (2)	169.1 (3)	172.7 (3)	156.0 (2)
C2-C3-C4-C5	**52.2 (3)**	**50.0 (5)**	**54.8 (4)**	**39.5 (3)**
C1-O5-C5-C6	-171.9 (2)	-175.0 (5)	-174.1 (4)	147.2 (3)

C1-O5-C5-C4	**69.1 (3)**	**66.1 (5)**	**66.5 (5)**	**27.5 (4)**
O6-C6-C5-O5	-173.6 (2)	-178.8 (4)	-175.1 (4)	-178.0 (3)
O6-C6-C5-C4	-55.1 (3)	-58.6 (6)	-55.9 (5)	
O4-C4-C5-O5	**173.7 (2)**	**-179.9 (4)**	**176.7 (3)**	- 56.6 (4)
C3-C4-C5-O5	-65.6 (3)	-59.0 (5)	-63.1 (4)	**176.6 (3)**
O4-C4-C5-C6	54.6 (3)	60.2 (5)	57.2 (4)	-66.0 (3)
C3-C4-C5-C6	175.2 (2)	-178.9 (4)	177.4 (4)	58.1 (3)
C1-C2-N1-C21	62.5 (3)	86.4 (6)	102.0 (6)	175.6 (3)
C3-C2-N1-C21	-172.7 (2)	-151.6 (5)	-135.9 (6)	101.2 (5)
C2-N1-C21-22	177.9 (3)	166.8 (6)	-170.9 (7)	-134.7 (4)

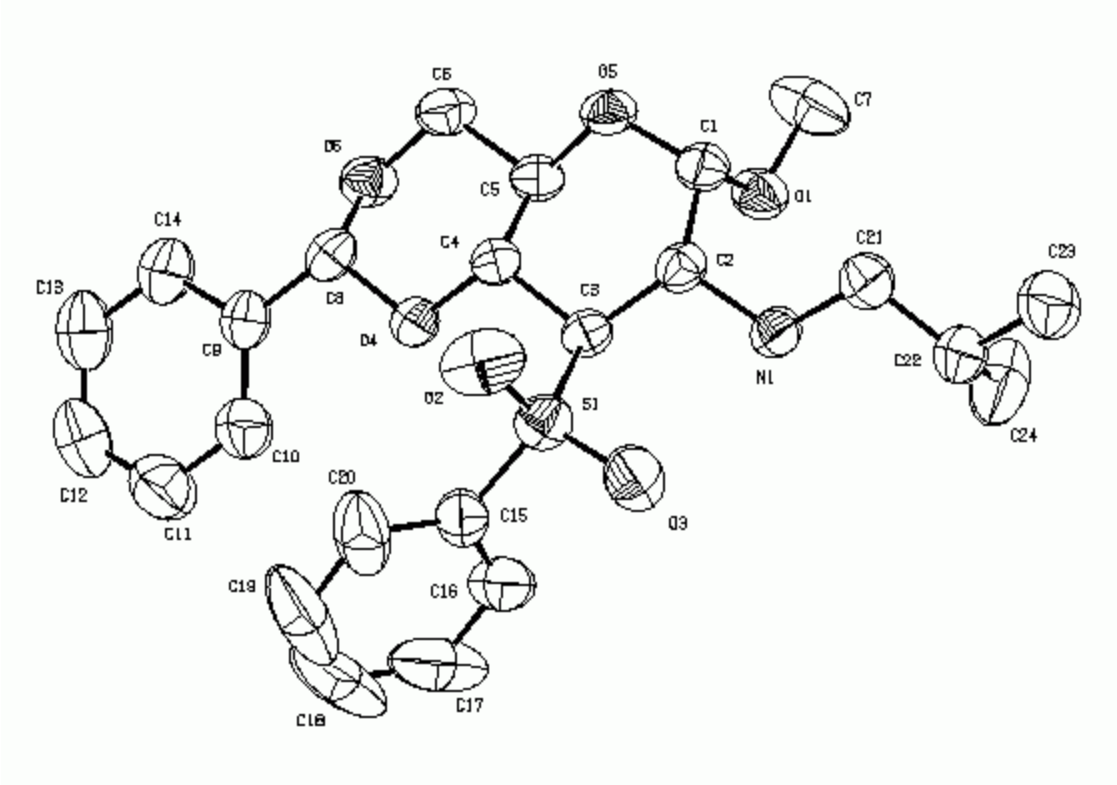


Fig. A2. 1. Perspective view of molecule 2aa

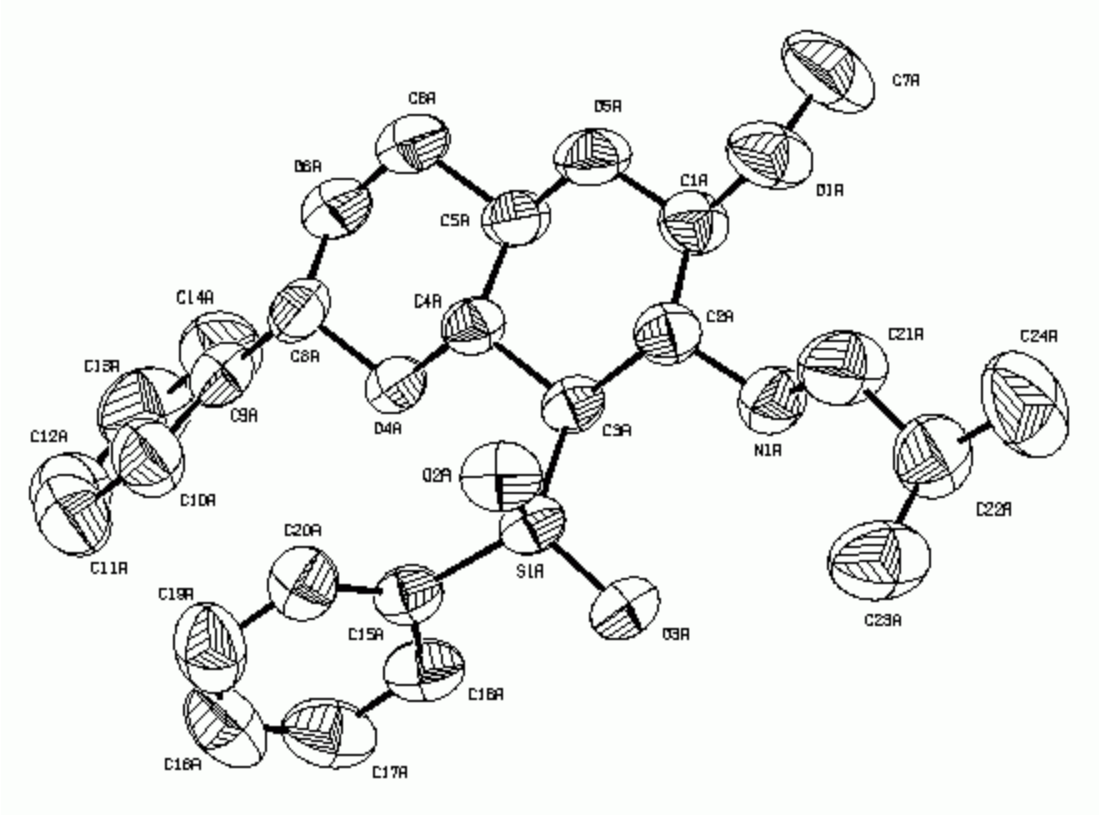


Fig. A2. 2. A. Perspective view of molecule A of 2ßa

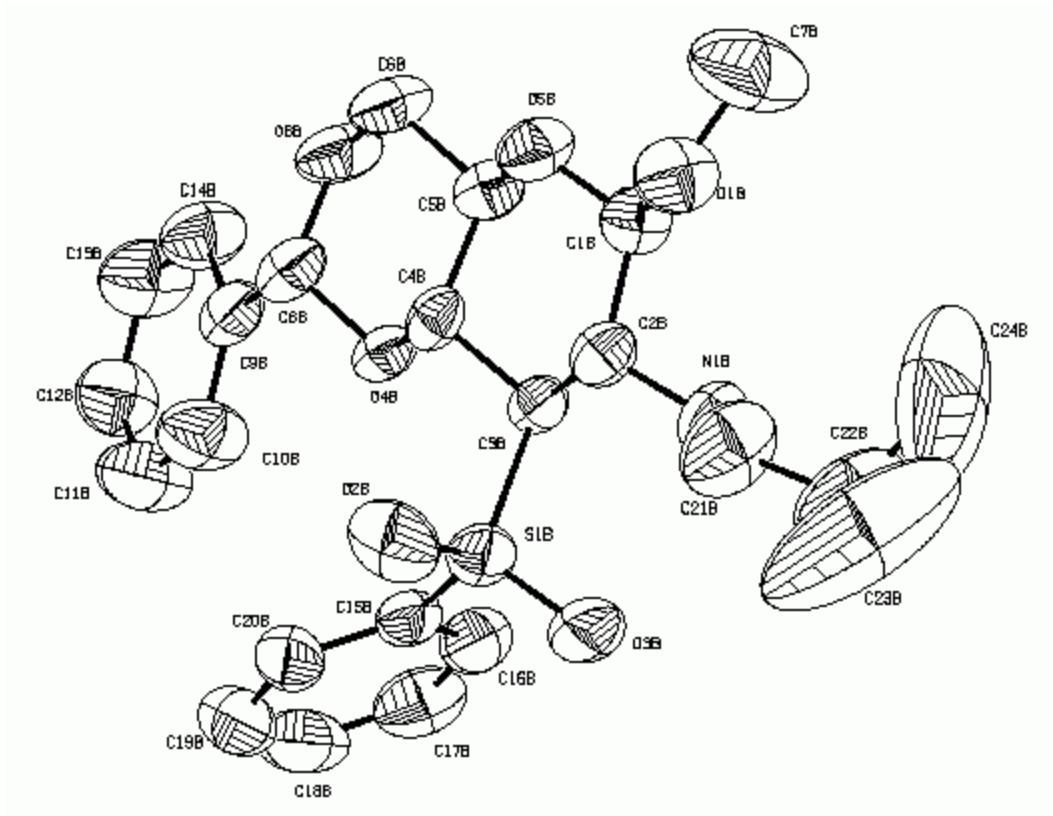


Fig. A2. 2. B. Perspective view of molecule B of 2ßa

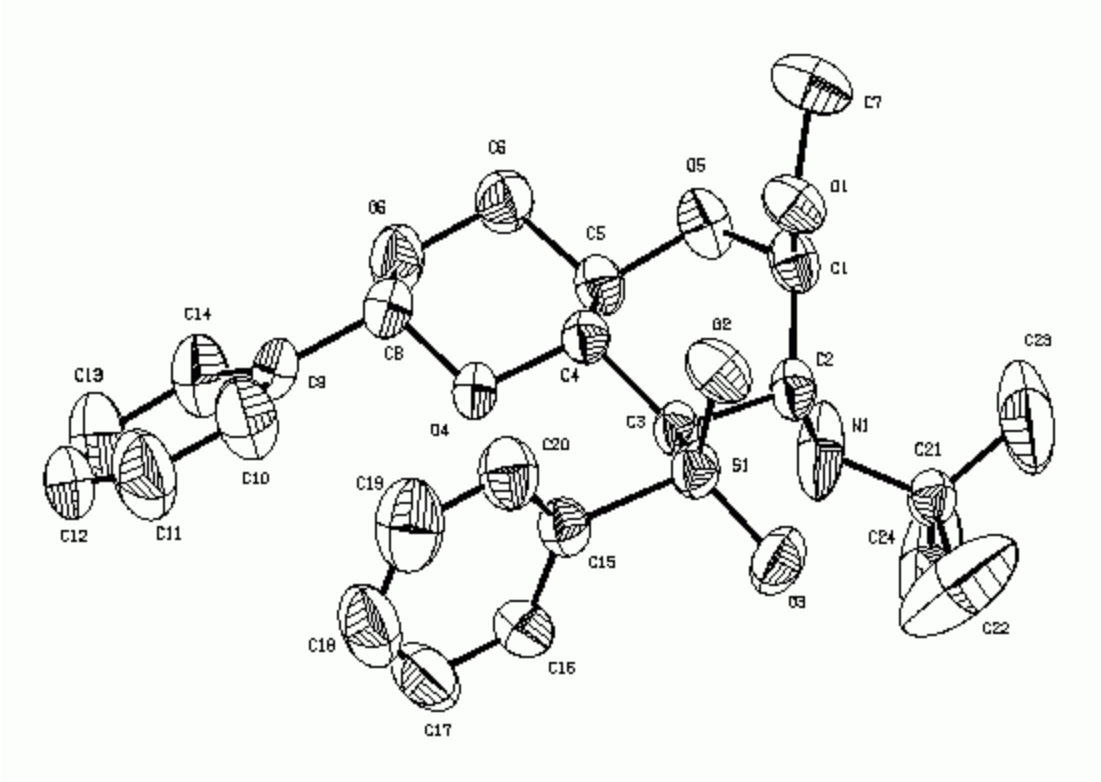


Fig. A2. 3. Perspective view of molecule of 2ßc

References

1. Suresh, C. G; Ravindran, B.; Rao, K. N.; Pathak, T. *Acta Cryst*., 2000, **C56**, 1030-1032

2. Flack, H. D. *Acta Cryst*., 1983, **A39**, 876-881

3. Enraf-Nonius. CAD-4-PC Software. Version 1.2. Enraf-Nonius, Delft, The Netherlands, 1993

4. Gabe, E. J.;LePage, Y.; Charland, J. P.; Lee, F. L.; White, P. S. *J. Appl. Cryst*., 1989, 22, 384-387

5. Molecular Structure Corporation. MSC/AFC Diffractometer Control Software. MSC, 3200 Research Forest Drive, The Woodlands, TX 77381, USA. 1991

6. Sheldrick, G. M. *Acta Cryst*., 1990, A46, 467-473

7. Sheldrick, G. M. SHELXL93., University of Gottingen, Germany, 1993

8. Spek, A. L. PLATON, A Multipurpose Crystallographic Tool, Utrecht University, Utrecht, The Netherlands, 2000

Curriculum Vitae

Curriculum Vitae

Narasimharao. K. (K. N. Rao.)

Date of Birth: January 06, 1972

Academic Qualifications:

Ph.D. (Thesis submitted)*	(2002) University of Pune, Pune.
M.Sc. (Chemistry)	(1994) Bangalore University, Bangalore.
B.Sc. (Physics, Chemistry, Mathematics)	(1992) Bangalore University, Bangalore.

* Title of Thesis: "**Studies on the structure and interactions of a plant lectin and a plant protease inhibitor**."

Address for correspondence
Manikuppe, Naranahalli (Post)
C. S. Pura (Via)
Tumkur (District)-572213.
India.
Phone: 0816-449538
Email: knraos@yahoo.com, krisleela2000@yahoo.co.in

Awards and experience:

Recipient of Junior & Senior research fellowships from Council of Scientific & Industrial Research (CSIR, Govt. of India) in Chemical Sciences.

Research experience in the fields of : a) **Protein X-ray Crystallographic methods**- Protein crystallization, X-ray diffraction data collection, Structure solution, Structure refinement and analysis.
b) **Protein Biochemistry**- Protein purification and characterization techniques.

Specifically- Worked on a lectin and a protease inhibitor from plant seeds using the above mentioned techniques for Ph. D. thesis.

Worked on a Department of Biotechnology, Govt of India, sponsored project on the Structural studies of two lectins from *Artocarpus hirsuta* for 2 years.

Publications/Presentations:

The three-dimensional structures of the lectin and the protease inhibitor are to be published. Some of the results described in the thesis are published in the following publications.

1. **Rao.K.N.** Gurjar.M.M., Gaikwad.S.M., Khan.M.I. and Suresh.C.G., (1999). Crystallization and preliminary X-ray studies of the basic lectin from the seeds of *Artocarpus hirsuta*. *Acta.Cryst.* **D55**, 1204-1205.
2. **Rao.K.N.** Hegde.S.S., Lewis.R.J., and Suresh.C.G., (1999). Crystallization and preliminary X-ray diffraction studies of a Bowman-Birk inhibitor from the *Vigna unguiculata* seeds. *Acta.Cryst.* **D55**, 1920-1922.

3. Suresh.C.G., Pundle.A.V., Sivaraman.H., N **Rao.K.N**., Brannigan.J.A., McVey.C.E., Verma.C.S., Dauter.Z., Dodson.E.J., Dodson.G.G., (1999). Pencillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat.Struct.Biol*. **6**., No.5., 414-416.
4. Suresh.C.G., Ravindran.B., **Rao.K.N** and Pathak.T., (2000). Comparison of the two anomers of methyl 2-(N-benzylamino)-2,3-di-deoxy-4,6-O-phenylmethylene-3-C-phenylsylfonyl-D-glucopyranoside. *Acta.Cryst.* **C56**, 1030-1032.
5. Suresh.C.G., Ravindran.B., Pathak.T., **Rao.K.N**., Prasad.J.S.S., and Lokanath.K., Structural studies on the C-2 substitution in a new set of synthetic dideoxyaminosugars; the steric bulk at C-2 influences the puckering of the pyranose ring and the choice of substitution group leads to anomeric selectivity of the reaction (Communicated).
6. **Rao.K.N**., Lewis.R., Suresh.C.G. and (1998). Crystallization and preliminary X-ray studies on the Bowman-Birk inhibitor from *Vigna Unguiculata* seeds. Presented at the XIX National seminar on crystallography., Univ of Madras, Chennai, Abstract is available in the Book of Abstracts Sec. E *(Biocrystallography*), No.E1.17.
7. **Rao.K.N**., Suresh.C.G. and Lewis.R., (1999). The crystallization and crystal structure of a Bowman-Birk inhibitor from *Vigna Unguiculata*. Presented at the XVIIIth IUCr Congress & General assembly, Glasgow, Scotland., abstract is available in Abstracts book Sec.04, *Crystallography of Biological Macromolecules*, pp.317, No.P09.04.003.
8. **Rao.K.N.** Gurjar.M.M., Gaikwad.S.M., Khan.M.I. and Suresh.C.G., (1999). Crystal structure of a lectin from the seeds of *Artocarpus hirsuta*. J.Biosci. Vol.24. Supppl.1.pp.157, (Presented at the XIII International biophysics congress, New Delhi, India., abstract is available in Abstracts book, Sec. D3: Molecular assemblies, Recognition and Metabolic regulation: *Cell surface interactions*. p.157., No. P. 337.
9. **Rao.K.N.** Gaikwad.S.M., Gurjar.M.M., Khan.M.I. and Suresh.C.G., (2001). Structure of a galactose specific lectin from *Artocarpus hirsuta* in two orthorhombic crystal forms. Presented at the International symposium on Crystallography and Bioinformatics in Structural Biology, IISc, Bangalore, India., Abstract is available in Abstracts book, Sec. S1: *Protein structures*, S1-42.

References:

1. Dr. C. G. Suresh (Ph. D. Supervisor)
Senior Scientist,
Division of Biochemical Sciences,
National Chemical Laboratory,
Pune 411 008, India.
Tel: +91-20-5893034
Fax: +91-20-5884032
E-mail: suresh@ems.ncl.res.in

2. Dr. Shama Brnabas
Senior Scientist,
Division of Biochemical Sciences,
National Chemical Laboratory,
Pune 411 008, India.
Tel: +91-20-5893034
Fax: +91-20-5884032
Email: john@ems.ncl.res.in

3. Dr. M. I. Khan
Senior Scientist,
Division of Biochemical Sciences,
National Chemical Laboratory,
Pune 411 008, India.
Tel: +91-20-5893034, Fax: +91-20-5884032,
E-mail: mikhan@ems.ncl.res.in