Structural studies of three plant lectins showing specificity for complex sugars and their structure function comparison along with recombinant Jacalin.

THESIS SUBMITTED TO

THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

BY

Urvashi

DIVISION OF BIOCHEMICAL SCIENCES

NATIONAL CHEMICAL LABORATORY

PUNE 411 008

INDIA

May 2011

Dedicated to…

My Grandmother, Mummy

& Papa

राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग पुणे - 411 008. भारत
NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411 008. India.

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Structural studies of three plant lectins showing specificity for complex sugars and their structure function comparison along with recombinant Jacalin."** is the result of investigations carried out by **Ms. Urvashi** under my supervision. The material obtained from other sources has been duly acknowledged in the thesis. The results presented in this thesis have not previously formed the basis for the award of any other diploma, degree or fellowship.

Date: 10 - 5 - 2011

Dr. C. G. Suresh

(Research Supervisor)

Scientist

Division of Biochemical Sciences

National Chemical Laboratory

Pune 411 008.

	☎	FAX	WEBSITE
Communication Channels	+91-20-2590 2000 +91-20-2590 (DID)	+91-20-2590 2601(DIR) +91-20-2590 2660 (COA)	www.ncl-india.org

DECLARATION

I hereby declare that the research work reported in the thesis entitled "**Structural studies of three plant lectins showing specificity for complex sugars and their structure function comparison along with recombinant Jacalin**" is entirely original and was carried out by me under the supervision of Dr. C. G. Suresh at the Division of Biochemical Sciences, National Chemical Laboratory (NCL), Pune, India. The thesis is submitted by me to the University of Pune for the degree of Doctor of Philosophy in Biotechnology.

I further declare that the scientific contents of this thesis have not been the basis for award of any degree, diploma, fellowship, associateship or any other similar title of any University or Institution. The material obtained from other sources has been duly acknowledged in the thesis.

Date: 10/05/2011

Urvashi

ID number-10678

Division of Biochemical Sciences

National Chemical Laboratory

Pune -411 008, INDIA.

Acknowledgements

My tenure as a graduate student at National Chemical Laboratory has been a great learning experience in several aspects. The present thesis is a result of five years of work where many people have assisted in numerous ways to make this thesis possible. I now take the opportunity to express my gratitude to all of them.

I owe huge gratitude to Dr. C. G. Suresh, my research supervisor, holding whose fingers I stepped into the world of structural biology. I thank sir for providing me an opportunity to work under his able guidance and for his constant encouragement in warm and considerate ways. He has always been supportive and patient during many ups and downs of my time as research student. He has given me all the freedom to think and plan my experiments independently however; whenever I was stuck he was always available with his advice and critical suggestions which I needed. Above all, it was his faith in my capabilities that has helped me to deliver the best out of me. I am also grateful to Indu madam for her care and treating me like part of family. I often miss the delicious food which we had during numerous dinner parties at their home. I wish to thank Shruti and Nithi for sharing nice time with us.

I wish to express my thanks to Dr. M.V. Krishnasastry, NCCS who provided us the recombinant Jacalin clone for structural studies. He has been generously supportive in providing all the facility needed to purify the lectin. He has also shown keen interest in the progress of my work and helped me during the initial purification of lectin.

I would especially like to thank Dr. Sushama Gaikwad, for teaching me the biochemical and biophysical techniques and the most difficult analysis part. Her friendly guidance and timely suggestions have contributed a great deal for the smooth progress of my work.

I am lucky to be associated with Dr. P. R. Rajamohanan during the NMR studies on MoL. I thank him for sparing time, in spite of his busy schedule, for the analysis of the results and various discussion sessions without which the work on MoL would not have been possible. I will always remain grateful for all the help he has offered.

I also wish to thank Dr. M. I. Khan under whose collaboration the work on chickpea lectin was initiated. He also provided us the sequence of CAL which finally helped us to elucidate the structure of the protein.

I thank Dr. Mahesh Kulkarni, Centre for Material Characterization, NCL, for help in MALDI-TOF and LCMS analysis of my protein samples.

My sincere thanks goes to Dr. Raghavan Vardarajan, MBU, IISc for permission to use the DSC facility. I thank Dr. Rajendra for helping me to record the DSC thermograms.

I also wish to thank Dr. Arwind Sahoo, NCCS for allowing us to use his CD polarimeter and Mr. Vivekanand Yadav for booking the instrument and other help during the experiments.

I acknowledge Dr. Krishnamoorthy at TIFR, Mumbai, for allowing me to access their Time resolved spectrometry facility. I also thank Dr. Madhuri Ugaonkar for her help during the experiments.

I wish to thank Dr. B. Gopal for allowing the access to cryo-crystallographic facility at MBU to ship the Dewar for data collection at ESRF, Grenoble France. I also thank Lalima, and other members of Dr B. Gopal's group for their cooperation during my stay at IISC.

I thank Dr. Hassan Belrhali for his help in data collection at ESRF, France. He took all the care to make me feel comfortable at ESRF.

I thank Dr. Rajesh Singh for his timely help and valuable suggestions in sorting out various software related issues and introducing me to other new crystallographic programs.

Special thanks are due to Gayathri for her help in setting up robotic crystallization experiments at MBU, IISC, Bangalore and also for teaching me the experimental phasing with some of my datasets when she was at Pune.

I thank my seniors at NCCS for their hospitality and generous help when I was working there: Anagh, Aizaj, Soumya, Amita, Neesar, Pani and Renu. It was Aizaj who taught me the recombinant protein expression and other molecular biology techniques. Soumya and Amita have always been friendly to introduce me to their lab facilities.

I thank my seniors from NCL, Dr. Suresh, Priya, Satya, Uma and Poorva for their help and useful discussion at several times. Especially, I thank Uma who has always shown her concern for the progress of my work and inspired me with new ideas even through mails. I had a wonderful company in lab with my colleagues, Nishant, Manas, Tulika, Payal, Priyabrata, Prachi, Ranu, Ruby, and Deepak. I will always remember the help offered by Manas & Tulika at many times in the lab as well as in the hostel. I appreciate the help offered by Ruby and Prachi who have joined our group recently, in conducting some of the experiments for me. I thank Tulika, Prachi and Deepak for pointing out the typographical errors in my thesis. I also thank the short-term visitors of our lab, Varun, Soundarya, Sharath and Mithila.

Harshada, has been my best friend ever and a nice roommate since the initial days. She has always comforted me whenever I was at the lowest. I will always cherish the memories of those days.

I thank my M.Sc friend Priti for listening to all my problems time to time. I am grateful to her for all the support and encouragement.

Gyan, deserves a special mention for the numerous ways he has helped me in the lab as well in the hostel. I will remain grateful to his generosity.

I thank Sumita for being a nice friend. She has been a wonderful company in the hostel where we have shared nice time together. I will miss the amazing dinner offered by her at home. I also thank Santosh for his help several times and the delicious biryani cooked by him for our dinner party at their home.

I also wish to thank my friends from NCL in other groups who have helped me in one or the other way for my research work: Sridevi, Shashi, Avinash, Madhurima, Sonali, Manasi, Vishnu, Ashwini, Varsha, Priti, Fazal, Radhika, Ram, Prashant, Ruby, Rishi, Somesh, Arun, Sheetal, Gayatri, Pallavi, Ramya, Manaswini, Mahesh, Sachin, Nishita and Nitasha.

I thank my friends at hostel Munmun, Neha, Sameera and Dimpal for being a homely company. In the last few months I have come to know Dimpal who has been a close friend at the most difficult time.

I acknowledge Mrs. Gupte, Mrs. Puranik and Mrs. Indira, for their help in administrative work. Mr. Trehan, Mr. Ramakant and Mr. Giri are acknowledged for their technical assistance.

I thank the Director, National Chemical Laboratory and the Head, Division of Biochemical Sciences, for allowing me to pursue my Ph.D work at the Division of Biochemical Sciences, NCL. University Grants Commission (UGC), India, is acknowledged for financial support. I acknowledge Department of Biotechnology, Govt. of India for funding to ESRF Visit.

My younger sister Upasana and brother Kislaya, have been greatest support during each and every walk of my life. I thank them for their love, affection, prayers, support and encouragement.

My parents have been my rocks, I don't know how this all could have been possible without their blessing and everlasting support. I have no intention to thank them formally.

Urvashi

Abbreviations

- ACL-*Arisaema curvatum* lectin
- AAL- *Aleuria aurantia* lectin
- Å- Angstrom
- ° -Degree
- µ-Micro
- α-Alpha
- β-Beta
- γ-Gamma
- δ- Delta
- χ-Chi
- σ-Sigma
- τ-Tau
- µM-Micromolar
- 1D- One dimensional
- 2D-Two dimensional
- 3D -Three dimensional
- BAPNA- benzyl-DL-arginine para-nitroanilide
- BME-β-mercaptoethanol
- C- Celsius
- CCP4- Collaborative crystallography programs 4
- CD- Circular dichroism
- CAL-*Cicer arietinum* lectin
- CPTI-Chickpea trypsin inhibitor
- Con A- Concanavalin A
- COSY- Correlation spectroscopy
- CPB-Citrate phosphate buffer
- DEPT-Distortionless enhancement by polarization transfer
- DSC-Differential scanning calorimetry
- FPLC -Fast protein liquid chromatography
- HSQC- Hetero-nuclear Single Quantum Coherence
- HMBC- Hetero-nuclear Multiple bond Correlation
- kDa-Kilo Dalton
- *E. coli- Escherichia coli*
- gm-gram
- JRLs - Jacalin related lectins

- K- Kelvin
- LLG- Log likelihood gain
- m- Milli
- M- Molar
- MAD- Multiwavelength anomalous dispersion
- MIR -Multiple isomorphous replacement
- MR-Molecular replacement
- mM- Millimolar
- MoL-*Moringa oleifera* lectin
- NCS- Non-crystallographic symmetry
- nm- Nanometer
- njacalin - native Jacalin
- NMR -Nuclear magnetic resonance
- NOE- Nuclear overhauser effect
- NOESY- NOE Spectroscopy
- OD-Optical density
- PBS-Phosphate buffer saline
- PDB-Protein data bank
- PEG-Polyethylene glycol
- PIs-Protease inhibitors
- PPIs- Plant protease inhibitors
- RF - radio frequencies
- rpm- Revolutions per minute
- RSL- *Ralstonia solanacearum* lectin
- SDS-PAGE - Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
- SIR -Single isomorphous replacement
- SIRAS- Single isomorphous replacement and anomalous scattering
- TOCSY-Total correlation spectroscopy
- Tris- tris-hydroxymethyl amino methane
- RMSD- Root mean square deviation
- rjacalin-Recombinant Jacalin

Abstract

Lectins mediate a variety of biological processes such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilization, cancer metastasis, growth and differentiation; all these by the recognition of glycoconjugates and cell surface sugars. The affinity of lectins for oligosaccharides is enhanced by the clustering of simple binding sites. The crystal structures of lectin-carbohydrate complexes turned out as excellent model systems to study protein-carbohydrate interactions. More than 50 % of the three-dimensional structures of lectins and their complexes with sugars listed in the 3-D Lectin Database (http://www.cermav.cnrs.fr/lectins) are from plants, the remaining ones are from animals, bacteria, fungi and viruses. The most thoroughly studied group is that of plant lectins and in that the largest and the best-characterized family is from the seeds of leguminous plants. Most of the plant lectins are specific for simple sugars and their hemagglutination activity is inhibited by monosaccharides or oligosaccharides. In addition, there are lectins whose hemagglutination activity is inhibited only by complex glycoproteins. Very few studies of these lectins showing complex sugar specificity are available.

In the present thesis the structural and functional studies of three plant lectins and a modified single chain recombinant jacalin have been investigated for elucidating their structures as well as interactions contributing towards the sugar recognition and specificity. The three hemagglutinins showing complex sugar specificity studied are: one from the seeds of a legume *Cicer arietinum* (CAL), another extracted from the tubers of *Arisaema curvatum* (ACL) belonging to Araceae family and the third lectin isolated from the seeds of *Moringa oleifera* (MoL) a member of Moringaceae family. Single chain

recombinant form of galactose-specific jacalin from *Artocarpus integrifolia* showed reduced sugar affinity compared to wild type. A protein, identified as a trypsin inhibitor, has been purified along with CAL and it has also been crystallized. Preliminary characterization of these protein crystals is described.

The thesis has been organized into the following chapters:

Chapter 1: General introduction and review of lectins.

Chapter 2: Three-dimensional structure of a seed lectin from *Cicer arietinum* (CAL) showing specificity for complex sugars

Chapter 3: Biochemical and biophysical characterization of *Arisaema curvatum* lectin (ACL).

Chapter 4: Biophysical studies of a hemagglutinin from *Moringa oleifera* using NMR, circular dichroism (CD) and fluorescence spectroscopy.

Chapter 5: Structural study of single chain recombinant jacalin with reduced sugar affinity.

Chapter 6: Purification and preliminary X-ray crystallographic analysis of a trypsin inhibitor protein (CPTI) from *Cicer arietinum*.

Chapter 1: General introduction and review of lectins

This chapter presents a general overview of lectins with reference to their occurrence, classification and biomedical applications. Plant lectin families have been accounted in detail with reference to their three-dimensional structures and sugar binding properties.

Chapter 2: Three-dimensional structure of a seed lectin from *Cicer arietinum* (CAL) showing specificity for complex sugars

Purification and preliminary X-ray characterization of CAL has already been reported from our laboratory. Two crystal forms of CAL were grown, orthorhombic ($P2_12_12$) with unit cell dimensions a=70.9, b=73.3, c=86.9 Å and trigonal (P3) with a=b=80.2, c=69.1 Å and $\beta=120^o$. An iodine derivative of orthorhombic form was also characterized (Katre U V, PH.D thesis, University of Pune, 2007). However, structural studies remained inconclusive due to lack of sufficient information including protein sequence. The previously reported diffraction data (2.2 Å resolution) for orthorhombic and trigonal crystals grown in the presence of iodine salts are used for structural analysis.

Previous attempts to identify models using the N-terminal sequence and to try molecular replacement (MR) method did not succeed in determining the structure. When we were struggling for a second heavy atom derivative the full sequence information became available from our collaborator Dr. M. I. Khan. A fresh search in protein data bank using full length sequence identified a closely related and recently submitted three-dimensional structure of pA2 albumin from *Lathyrus sativus* (pdb id: 3lp9). A MR calculation in *PHASER* of the *CCP4* suit (Collaborative Computational Project, Number 4, 1994) using this model for data of the orthorhombic and trigonal forms gave positive solutions. The structures in both crystal forms were refined using refmac 5.0 of *CCP4* suit. The final R-factors were R_{free} 22.3% and R_{work} 18.2% for orthorhombic form and 22.2 % and 17.0 % for trigonal form using data in the resolution range 20-2.2 Å and 35-2.2 Å, respectively.

The CAL was found to exist as dimer in solution. The asymmetric unit in the crystals also turned out to be dimer. Each subunit is organized into four β-propeller blades related by a pseudo four-fold axis and a channel in the middle of the monomer. This channel provides ligand binding sites for a calcium ion and water molecule. In iodine derivative there is an extra iodine atom in the channel. There is no evidence for involvement of metal-ligand in sugar binding. Identification of the sugar-binding site was attempted using chemical modification studies.

Chapter 3: Biochemical and Biophysical characterization of *Arisaema curvatum* lectin (ACL).

The second lectin showing complex sugar specificity, ACL, was supplied in a purified form by our collaborators from Gurunanakdev University, Amritsar, Punjab. The purified sample was dialyzed against deionized water and concentrated to 20 mg ml^{-1} using centrivap concentrator. Crystallization trials were unsuccessful. However, other biophysical experiments provided information on structure folding and hemagglutination activity of ACL.

Only a single tryptophan was detected in this protein. Chemical modification studies indicated contributions from tryptophan, histidine and carboxylate groups of aspartic and glutamic acid towards sugar binding. Far UV CD spectra of ACL were recorded to estimate the secondary structural elements. The analysis of CD data using CDSSTR program of CDPro suite predicted 4.9 % α-helix, 45 % β-sheet, 18 % turns and 31 % random coil.

The fluorescence maximum of 353 nm for native ACL at pH 7.0 indicated fully exposed polar environment of tryptophan. This was probed using solute quenching

studies at pH 7.2 and 3.0 with quenchers such as acrylamide, potassium iodide and cesium chloride, confirming the highly solvent-exposed state and strongly electropositive environment of the tryptophan residue at neutral pH. It also showed a shift towards hydrophobic environment at acidic pH.

ACL is most stable at neutral pH and retains hemagglutination activity in a wide pH range 1-10 at least for 3 hrs. There is considerable loss of ACL native secondary structure at acidic pH. These acid-induced partially unfolded forms also show enhanced ANS binding. The ANS binding information and the near UV CD data recorded at pH 3 predicted the existence of an acid induced molten globule structure for ACL. The stability of ACL structure was tested using thermal and chemical denaturation studies at pH 7.0 and 3.0.

Chapter 4: Biophysical studies of a hemagglutinin from *Moringa oleifera* using NMR, circular dichroism (CD) and fluorescence spectroscopy.

The purification and characterization of a 14 kDa hemagglutinin from *Moringa oleifera* (MoL) was previously reported from our group (Katre *et al.,* 2008). Further work on MoL was undertaken to determine its three-dimensional structure by X-ray crystallography and to study its structure-function relationship. Crystallization trials have remained unsuccessful presumably due to the high solubility of protein.

MoL was also characterized using NMR, CD and fluorescence spectroscopic techniques. The destabilizing effect of reducing agents on MoL was known. In the present study the protein was incubated with varying concentrations of β-mercaptoethanol (BME) at pH 2.0, 7.0 and 10.0. Both the near and far UV CD profiles showed loss of native structure of MoL at alkaline pH 10.0 in the presence of BME.

Although there was a 40% loss of secondary structure at pH 7.0 in the presence of reducing agent the effect was more pronounced at pH 10.0. A relative increase of fluorescence intensity at pH 7.0 was observed for ANS binding in the presence of BME, with the retention of a blue shift (10 nm). This suggested a change in the environment of hydrophobic residues. Unlike at pH 7.0 a decrease in the fluorescence intensity was observed upon ANS binding at pH 10.0. There was significant difference in the NMR signals of aromatic and aliphatic regions between the native and the BME treated MoL.

The high thermostability of MoL was tested using differential scanning calorimetric measurements at pH 7.0, 2.0 and 10.0. MoL showed high melting temperatures of 86 and 89ºC at pH 2.0 and 10.0, respectively, whereas no transition was detected at pH 7.0.

Chapter 5: Structural study of single chain recombinant Jacalin with reduced sugar affinity.

The gene of jacalin consisting of the α-chain and β-chain of active lectin linked by T-S-S-N loop was cloned to get a single chain recombinant Jacalin. The expression of the clone obtained from our collaborator Dr. M. V. Krishnasastry was standardized. The previously reported purification strategy was also modified to increase the yield.

The purified protein crystallized from a solution of 30 mg ml^{-1} in conditions 10 and 28 of Crystal Screen 1 (Hampton Research, USA) using the hanging-drop vapor-diffusion method. The quality of crystals grown in condition 10 (0.2M ammonium acetate, 0.1M sodium acetate buffer pH 4.6, 30 % w/v PEG 4K) was improved by substituting PEG 8K for PEG 4K. The crystals used for data collection appeared and grown to full size in 6 days. They belonged to the space group C2 having unit cell

dimensions a = 118.9, b = 42.3, c = 73.6 Å, β = 122.3°. The three-dimensional structure was determined using molecular replacement inputting native structure 1JAC as model. The structure was refined for final R-factors 18 and 21 % using data in 56-2.0 Å resolution.

The overall structure of recombinant jacalin matched well with the wild type structures except for the absence of bound sugar. The superposition with wild type structure showed identical sugar binding site. Interestingly, the electron density for the "T-S-S-N" loop was absent in the map of single chain recombinant jacalin implicating high flexibility of this loop region.

Chapter 6: Purification and preliminary X-ray crystallographic analysis of a trypsin inhibitor protein (CPTI) from *Cicer arietinum*.

Plant lectins and protease inhibitors are proposed to have a role in plant defense. Ye and Ng, 2002 (Ye and Ng, 2002) reported one 18 kDa chickpea protein (CLAP) resembling cyclophilin-like proteins possessing antifungal and anti-HIV-1 reverse transcriptase activities. The presence of a Kunitz-type trypsin inhibitor of mass 20 kDa was also reported in chickpea (Srinivasan *et al.,* 2005). During CAL preparation from chickpea seeds we have identified one 18 kDa protein possessing trypsin inhibitor activity. Crystals were grown for this protein.

The trypsin inhibitor protein from *Cicer arietinum* (CPTI) present in the ammonium sulfate precipitated fraction of the seed extract was loaded successively in two ion exchange chromatographic columns of DEAE-cellulose and SP-sephadex resins. Inhibitory action against trypsin from bovine pancreas was tested using the artificial substrate BAPNA (benzyl-DL-arginine para-nitroanilide). The CPTI showed IC50 value

of 2.5 µg. The minimum inhibitory concentration estimated was 0.8 µg/ml. The inhibitory action of CPTI estimated was 114 TIU/mg of protein.

CPTI crystallized from a 25 mg ml^{-1} protein solution in trials using commercially available crystallization kits supplied by Hampton research (USA) and Molecular dimensions Ltd (UK). Three orthorhombic crystal forms were obtained in different conditions. Orthorhombic form A ($P2_12_12$) was obtained from condition 18 of JCSG I screen (0.1 M citrate-phosphate buffer pH 4.2, 40 % ethanol, 5 % w/v PEG 1K) having unit cell dimensions a=37.2, b=41.2, c=104.6 Å. They diffracted up to 2.0 Å resolution. Orthorhombic form B ($P2_12_12_1$) with a=41.1, b=50.1, c=75.1 Å, was obtained from condition 12 of JCSG II screen (0.1 M imidazole pH-8.0, 10 % w/v PEG 8K). They diffracted only up to 2.8 Å. Thin needle crystals of orthorhombic form C ($P2_12_12$) obtained from condition 9 of JCSG screen I (0.2 M ammonium chloride, 20 % PEG 3350) had unit cell dimensions a= 44.4, b= 75.7 and c= 133.7 Å and diffracted not beyond 3.5 Å resolution.

Due to lack of sequence information no model for molecular replacement method could be identified. Preparation of heavy atom derivative for multiple isomorphous replacement (MIR) method was tried. Out of several heavy atom derivatives attempted only a derivative of iodine with low occupancy sites could be characterized using data collected in BM14 Beamline at ESRF, Grenoble.

Chapter 7: Conclusions

This last chapter sums up the conclusions and new findings of this research.

List of publications

- Urvashi Sharma, Sushama M. Gaikwad, C. G. Suresh, Vikram Dhuna, Jatinder Singh, Sukhdev Singh Kamboj. Conformational Transitions in *Arisaema curvatum* Lectin: Characterization of an Acid Induced Active Molten Globule. Journal of Fluorescence. DOI 10.1007/s10895-010-0766-2.
- Urvashi Sharma and C. G. Suresh. Purification, crystallization and X-ray characterization of a trypsin inhibitor protein from the seeds of chickpea (*Cicer arietinum*) (accepted for publication in Acta F).
- Urvashi Sharma, Neesar Ahmed, M.V Krishnasastry and C.G. Suresh. Crystal Structure of Single Chain Recombinant Jacalin Showing a Highly Dynamic Posttranslational Excision Loop and Reduced Sugar Affinity. (to be communicated)
- Urvashi Sharma, Sushama M. Gaikwad, P. R. Rajmohanan and C. G. Suresh. NMR and other spectroscopic studies of *Moringa oleifera* lectin (MoL) reveals a BME induced stable molten globule conformation. (to be communicated)
- Urvashi Sharma, Sushama M. Gaikwad, M.I. Khan and C. G. Suresh. Crystal structure of *Cicer arietinum* lectin (CAL) showing specificity for complex sugars. (in preparation)

Poster Presentations

- Structure of a chickpea lectin (CAL) with four bladed β-propeller fold and specificity for complex sugars. Indo-US Symposium/Workshop on "**Modern trends in macromolecular structures**" at department of chemistry, IIT Bombay from 21-24th February, 2011.

- Crystal structure of recombinant Jacalin at 2.0 Å. **37th National Seminar on Crystallography (February 6-8, 2008), Jadavpur university, Kolkata.**

Seminar/Conferences/ Workshop attended

- From crystal to structures: **CCP4 Seminar cum Workshop**, February 15-19th, 2010. New Delhi, India.
- EMBO World lecture course, "**Recent Developments in Macromolecular Crystallography**. November 9- 14th, 2008, NCL, Pune, India.
- INSA-CAS workshop on structural biology. 21st -23rd December 2007 at IISC, Bangalore, India.
- 11th ADNAT Convention "**Advances in Structural Biology and Structure Prediction**". (23-25th February, 2007), CCMB, Hyderabad, India.

CONTENTS

Chapter 1

General introduction and review of lectins.

1.1. Rationale of the study

Biological recognition involves cells and proteins. Carbohydrates are the most prominently exposed structures on the surface of living cells with flexible chains and many potential binding sites. Proteins that interact with carbohydrates non-covalently occur widely in nature. Interactions between protein and carbohydrates play important role in many biological processes, such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilization, cancer metastasis and growth and differentiation. Lectins belong to the well known class of multivalent carbohydrate binding proteins, which recognize diverse sugar structures reversibly with a high degree of stereo-specificity in a non-catalytic manner (Sharon and Lis, 1989; Lis and Sharon, 1998). Lectins have got ubiquitous occurrence in nature ranging from microorganisms to plants, insects and animals and have been found to evoke a variety of biological responses simply by binding to their cognate carbohydrate ligands. The crystal structures of the lectin-carbohydrate complexes turn out as excellent model systems to study protein-carbohydrate interactions.

More than 50 % of the three-dimensional structures of lectins and their complexes with sugars listed in the 3-D Lectin Data Base (http://www.cermav.cnrs.fr/lectines/) are from plants, the remaining ones are from animals, bacteria, fungi and viruses. The most thoroughly studied groups of lectins are from plants and the largest and the best-characterized family is that from seeds of leguminous plants. Most of the plant lectins are specific for simple sugars and their hemagglutination activity is inhibited by monosaccharides or oligosaccharides. In addition, plant lectins with complex specificity that are inhibited only by complex glycoproteins and not by simple sugars have also been

reported. Very few studies on this type of lectins are reported, especially limited is the information on their three-dimensional structures.

This thesis deals with the structure-function studies of three plant lectins showing specificity for complex sugars. The first hemagglutinin is from the seeds of the legume *Cicer arietinum* (CAL), second one belongs to the Araceae family extracted from the tubers of *Ariesema curvatum* (ACL) and the third one is MoL isolated from the seeds of *Moringa oleifera* a member of Moraceae family. To investigate the active site interactions and other factors which contribute towards generating specificity and recognition for monosacharides a galactose specific lectin Jacalin, from *Artocarpus integrifolia* has been chosen. The recombinant Jacalin was prepared to study the role of posttranslational excision loop in deciding sugar specificity and comparing with the three-dimensional structure of CAL, for understanding the differences which contributes towards complex sugar specificity. Along with the structural studies on plant lectins the purification, characterization and preliminary X-ray crystallographic studies of an 18 kDa trypsin inhibitor (CPTI) protein isolated from the same seeds of *Cicer arietinum* are also included.

1.2. Lectins

1.2.1. Historical perspective

At the turn of 19th century Stillmark discovered lectin as a proteinaceous agent in the extract of castor bean (*Ricinus communis*) which was capable of agglutinating animal erythrocytes (Stillmark, 1888) and named it as ricin. In 1891, Hellin discovered abrin, (Hellin, 1891) the toxic protein from jequirity beans. During the 1890's, Ehrlich worked

with ricin and abrin, and discovered immunospecificity and reversibility of the antigen-antibody reaction as some of the fundamental principles of immunology. In 1898, Elfstrand introduced the term 'Blutkörperchenagglutinin' (hemagglutinin) as a common name for all plant proteins that cause clumping of cells (Elfstrand, 1898). Subsequently nontoxic lectins were discovered in the several legumes such as *Phaseolus vulgaris* (bean), *Pisum sativum* (pea), *Lens culinaris* (lentil) and *Vicia sativa* (vetch) (Landsteiner and Raubitschek, 1907). Later on it was observed that certain hemagglutinins could distinguish between different human blood-groups (Renkonen, 1948; Boyd and Reguera, 1949) leading to the discovery of the term 'lectin' (comes from Latin word, 'legere', which means to select or choose) by Boyd and Shapleigh in 1954.

1.2.2. Definition

The first definition of lectin was proposed by Goldstein as "sugar-binding proteins or glycoproteins of non-immune origin which agglutinate cells and/or precipitate glycoconjugates" (Goldstein *et al.*, 1980). This definition presumed polyvalency of lectins. With the discovery of lectins with toxic or hormone like activities, this definition was found to be inadequate. To overcome these shortcomings, Kocourek and Horejsi (1983) proposed a modified version of the definition as "Lectins are proteins of non-immunoglobulin nature capable of specific recognition and reversible binding to carbohydrate moieties of complex carbohydrate without altering the covalent structure of any of the recognized glycosyl ligands". Barondes (1988) later proposed a simpler definition: "lectins are carbohydrate-binding proteins other than enzymes or antibodies". Lectins are presently defined as the "proteins possessing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide" (Peumans & Van

Damme, 1995). This definition includes all previous definitions and comprises a broad range of proteins with different agglutination and/or glycoconjugate precipitation properties.

1.2.3. Classification

In the light of the data obtained mainly from molecular cloning of lectins and lectin related proteins, even this broad definition seems to be inadequate. Therefore, lectins were subdivided into following types based on the overall structure and properties of the mature proteins (Peumans and Van Damme, 1995):

I. Merolectins

These are small proteins, incapable of precipitating glycoconjugates or agglutinating cells because of their monovalent nature. They have a single carbohydrate binding domain. A typical example of a merolectin is Hevein, the small chitin-binding protein from the latex of the rubber tree (*Hevea brasiliensis*) (Van Parijs *et al.*, 1991). Other members of this group include monomeric mannose binding proteins from orchid (Van Damme *et al.* 1994 a and b) and class I chitinases (Collinge, 1993).

II. Hololectins

They are exclusively made up of carbohydrate-binding domains. However, unlike merolectins, they contain two or more carbohydrate-binding domains which are identical or very homologous and bind either the same or structurally similar sugars. By definition they are di- or multivalent hence they are fully capable of agglutinating cells and/or

precipitating glycoconjugates. Majority of plant lectins fall under this category (Van Damme *et al.*, 1998).

III. Chimerolectins

These are fusion proteins composed of a carbohydrate binding domain tandemly arrayed with an unrelated domain with well defined catalytic activity which functions independently of the former. Depending upon the number of binding sites, chimerolectins behave as merolectins or hololectins. For example, the type 2 ribosome inactivating proteins (RIPs) ricin and abrin consist of a toxic A chain (which has the N glycosidase activity characteristic of all RIP's) and a carbohydrate binding B chain (Barbieri *et al.*, 1993) with two carbohydrate binding sites to agglutinate cells function as hololectins whereas, class I chitinases with a single chitin binding domain that do not agglutinate cells behave as merolectin (Peumans and Van Damme, 1995).

IV. Superlectins

Superlectins are also fusion proteins having at least two carbohydrate binding domains, arranged in tandem, but they are able to recognize structurally unrelated sugars. A typical member is tulip lectin TxLCI, which contains an N-terminal mannose-binding domain tandemly arrayed with an unrelated GalNAc-binding domain (Van Damme *et al.*, 1996).

With increasing number of the three-dimensional structures of lectins reported, a more general classification of lectins was proposed by Lis and Sharon, (1998). According to this classification lectins are divided into three classes as described below:

A. Simple lectins

This class of lectins contains subunits of less than 40 kDa molecular weight. The monomers, though few in number, may or may not be identical, comprising domains in addition to the sugar binding site(s). This class includes practically all known plant lectins as well as the galectins.

B. Mosaic (or multidomain) lectins

They are all composite molecules with a wide range of molecular weights, consisting of several kinds of protein modules or domains, only one of which possesses a carbohydrate binding site. Included in this group are diverse proteins from different sources, viral hemagglutinins as well as animal lectins of the C-, P-, and I-type. By nature they are monovalent, but since they are embedded in membranes, they act in a multivalent fashion.

C. Macromolecular Assemblies

These kinds of assembly are commonly found on the bacterial cell surfaces in the form of fimbriae (or pili). These are filamentous, heteropolymeric organelles 3-7 nm in diameter and 100 to 200 nm in length, consisting of helically arranged subunits (pilins) of several different types, assembled in a well-defined order (Ofek and Sharon, 1990; Ofek and Doyle, 1994; Gaastra and Svennerholm, 1996). Only one, among many subunits, exhibits carbohydrate recognition property which gives binding and sugar specificity to the fimbriae.

1.2.4. Detection

Agglutination is the most easily detectable manifestation of the interaction of a lectin with cells. The ability to agglutinate cells distinguishes lectins from other sugar binding macromolecules such as glycosidases, glycosyltransferases, antibodies etc. Hemagglutination is routinely used to detect the presence of lectins in a biological source (Burger, 1974). The erythrocytes used (human or animal) may be plain or treated with papain, trypsin, or neuraminidase (Sharon and Lis, 1972). Other types of cells (Burger, 1974) and polysaccharides (Nakamura *et al.*, 1960) have also been used for the detection of lectins. Techniques like affinity electrophoresis (Horejsi and Kocourek 1974; Ohta *et al.,* 1998) and enzyme multiplied assay (Ghosh *et al*., 1979) too have been employed.

1.2.5. Occurrence

Lectins are ubiquitous in nature ranging from microorganisms to the plants and animals. More than 50% of the three-dimensional structures of lectins and their complexes with sugars listed in the 3-D Lectin Data Base (http://www.cermav.cnrs.fr/lectines/) are from plants, the remaining ones are from animals, bacteria, fungi and viruses. The most thoroughly studied group of lectins are those extracted from plants, perhaps due to their abundant availability and ease of isolation. However till date there is no direct evidence for the biological role played by them in plants. In contrast to that, the function of animal, bacterial and viral lectins is well elucidated. The occurrence and structural diversity of plant lectins are described briefly in the following sections.

1.2.6. Plant lectins

Plant lectins belongs to the heterogeneous group of carbohydrate binding proteins which differ in their molecular structure, biochemical properties and carbohydrate binding specificity (Van Damme *et al.*, 1998a). The richest sources for most lectins are the seeds, the storage organs of plants. However, roots (Urtica, Phytolacca, Sambucus, Trichosanthes, Calystegia), tubers or bulbs (Solanum, Galanthus, Scilla, Allium, Crocus, Tulipa, Iris), bark (Sambucus, Sophora, Robinia, Maackia, Laburnum, Cytisus, Cladrastis, Hevea, Abies) and leaves (Aloe, Lactuca, Vicia unijuga, Viscum album) are also reported to be lectin rich. The amount of lectin present in seed and vegetative tissues vary significantly. Seed lectins account for 1-10 % of total seed protein whereas in case of vegetative tissues such as bark, bulbs, tubers, rhizomes and corns they constitute 1-20 % of the total protein. Plant lectins occur in many species belonging to different taxonomic groups but are not so widespread.

1.2.7. Structural folds of plant lectins

On the basis of biochemical, structural and molecular analysis plant lectins are classified into seven different families of structurally and evolutionarily related proteins (Goldstein and Poretz, 1986, Van Damme *et al.*, 1998a). Four of these families, namely, the legume lectins, the type 2 ribosome inactivating proteins (RIPs), the chitin binding lectins containing hevein domains and the monocot mannose binding lectins are considered to be "large" families. The amaranthins, the Cucurbitaceae phloem lectins and the Jacalin related lectins comprise, at present, only a small number of individual lectins and accordingly are considered "small" families. The folds adopted by plant lectins, their quaternary structures and mode of interactions with a variety of carbohydrate ligands are so diverse that it is difficult to review all of them in a short introductory chapter like the

present one. Moreover, the thesis is exclusively concerned with plant lectins; the discussion here thus is restricted to them. A brief description of each of the plant lectin family is given below with special emphasis to their occurrence and structural properties.

1.2.7.1. Legume lectins

The legume lectins belong to the best characterized family of homologous carbohydrate binding proteins. Legume lectin family comprises more than a 100 well characterized individual lectins which have been isolated from over 70 species belonging to various taxonomic groups. Most legume lectins have been isolated from mature seeds where they account for 1-10 % of the total soluble seed protein. Several legumes contain two or more different seed lectins. Many of them have been purified and characterized with respect to their molecular structure, sugar binding specificity and biological role. The first plant lectin gene to be sequenced was that encoding the soybean seed lectin. (Vodkin *et al.*, 1983). Numerous legume lectins have been sequenced since then by chemical methods of protein sequencing (Peumans & Van Damme, 1999; Sharon & Lis, 1990; Van Damme *et al.,* 1998).

Concanavalin A (ConA) from the legume *Canavalia ensiformis* (Jackbean) was the first plant lectin to be purified and crystallized (Sumner and Howell, 1936). Amongst the lectins ConA happened to be the first one whose primary and three-dimensional structures were resolved (Edelman *et al.*, 1972; Hardman and Ainsworth, 1972). Legume lectins exhibit strong similarity at the level of amino acid sequence and tertiary structures, despite the wide variation in carbohydrate binding specificities and quaternary structures (Sharma and Surolia, 1997).

The legume lectin monomer has a molecular weight of about 25-30 KDa, composed of single-chain polypeptide of roughly 250 amino acid residues. In some cases proteolytic processing of the primary translation product yields a two chain lectin (Van Damme *et al.*, 1998). Members of this family exist as a dimer or tetramer, subunits of which are held together by noncovalent interactions. Each subunit possesses divalent cations at specific metal-binding sites mainly Mn^{2+} and Ca^{2+} which are involved in carbohydrate binding. The amino acids involved in sugar binding are found to be highly conserved in all legume lectins. Legume lectins exhibit fine specificity for di-, tri- and tetrasaccharides, with association constants 1000-fold higher than that for monosaccharides. They bind to the sugar only in the pyranose form in D-configuration, an exception is fucose which binds in L-form. Not all, but most of the legume lectins are glycosylated and carry up to three asparagine-linked oligosaccharides per subunit (Imberty *et al.*, 2000). Glycosylated lectin polypeptide possesses one or two glycan chains some which are of high mannose type for example, oligosaccharide side chain of soybean lectin (Van Damme *et al.*, 1998) while others are of complex type. Both high-mannose and complex type glycan may be found on a single chain polypeptide for example *Phaseolus vulgaris* agglutinin, PHA.

The most widely observed fold in lectins is the legume lectin fold, first observed for ConA and referred as the "jelly-roll motif" fold which is commonly found in viral coat proteins also. The architecture of the legume lectin monomer in this fold consists of three β-sheets: a flat 6-stranded back sheet, a curved 7-stranded front sheet and a short 5-stranded sheet (S-sheet) which holds two large sheets to stay together. The sheets are interconnected by turns and loops to form a flattened dome-shaped structure (Fig.1.1).

The main hydrophobic core is located between the back and the front sheet. A calcium ion and a manganese ion placed in the close proximity (9-13 Å) are essential for the carbohydrate binding. The metal ions are bound to four protein atoms and two water molecules. Demetallization induces large changes in the structure of ConA, leading to loss of carbohydrate binding ability (Bouckaert *et al.*, 1995).

The carbohydrate-binding site consists of four loops A, B, C and D at the top-front of the subunit in the form of shallow depressions on the surface of the protein (Fig.1.1). The amino acid residues which bind metal ions (Ca^{2+} and Mn^{2+}) are highly conserved while the residues at sugar binding sites have similar properties but are less conserved (Lis & Sharon, 1998). The conserved residues Asp and Gly (or Arg) are from loops A and B, respectively, whereas Asn and the hydrophobic residues (Phe, Tyr, Trp or Leu) are present in loop C.

The backbone atoms from loop D contribute three hydrogen bonds thereby serving as the primary determinant of carbohydrate specificity in legume lectins. The sequence analysis of carbohydrate binding loops of legume lectins show 4 to 7 gaps in the binding loop D providing an explanation for the broad specificity of legume lectins compared to other lectin families (Sharma & Surolia, 1997). The sites appear to be preformed because few

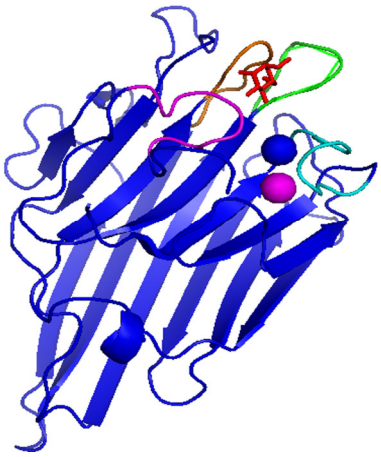

Figure 1.1. A monomer of ConA showing the bound mannose (red stick) and Mn and Ca ions (blue and magenta spheres respectively). The four loops important for sugar binding are shown in green, orange, magenta and cyan.

conformational changes occur upon carbohydrate binding.

As for its carbohydrate specificity, lectins belonging to diverse specificity groups are reported in legume family such as. Glc/Man specific lectins *canavalia ensiformis* lectin (ConA), *Pisum sativum* lectin (PSL), *Lathyrus ochrus* isolectin I and II, Lentil lectin (LenL); Gal/GalNAc specific lectins include *Erythrina corrallodendron* lectin (ECorL), Peanut agglutinin (PNA), Soybean agglutinin (SBA) Winged bean agglutinins (WBA I and II), *Dolichos biflorus* lectin (DBL) etc. Apart from this, other sugar specificity groups of fucose or promiscuous specific lectins I and II from *Ulex europeaus* (Audette *et al.*, 2000; Loris *et al.*, 2000) also occur exclusively in legume lectins. The complex carbohydrate binding lectins such as lectin IV from *Griffonia Simplificolia* (GSIV) (Delbaere *et al.*, 1993) and PHA-L (Hamelryck *et al.*, 1996a) are also reported in this group.

Most of the lectins from the Vicieae tribe (PSA, LCA, LoLI) form bivalent dimers, whereas many other legume lectins are tetrameric structures with four monosaccharide- binding sites. The oligomerization in legume lectin family involves the 6-stranded back β-sheet. Dimerization in ConA involves a side by side arrangement of the two monomers across two-fold symmetric axis, such that the two back β-sheets form a contiguous 12-stranded β-sheet. The two facing monomers associate by their flat bottoms to form a 'canonical dimer' placing the monosaccharide-binding sites at both ends of the dimer (Fig 1.2 A). This canonical mode of dimerization has been observed in AZD, DGL, PSL, LOLI, LOLII, LenL, SBA, UEA-I and UEA-II. The dimeric Lectin IV from *Griffonia simplicifolia* (GSIV) showed first evidence for the non-canonical mode of

dimerization involving face to face arrangement of the back sheets with the β-strands running perpendicularly to each other. It was proposed that this mode of dimerization was adapted to avoid the burial of important glutamic acid residue in the lectin. The non-canonical mode of dimerization was again observed in case of EcorL (Fig.1.2B) (Shaanan *et al.*, 1991).

Tetrameric association of two dimers of legume lectins can be achieved in different ways. The two canonical (side-by-side) dimers become associated through their back walls in such a way that two monosaccharide-binding sites occur at both sides of the tetramer for example conA (Fig 1.2 C). The tetramer in legume lectins possesses 222 (D2) symmetry with perfect tetrahedral arrangement (Uma and Suresh, 2009). An exception is PNA, which has an open quaternary structure involving association of two back–to-back dimers and it does not exhibit the expected 222 (D2) or a fourfold (C4) symmetry (Fig 1.2 D) (Banerjee *et al.*, 1994).

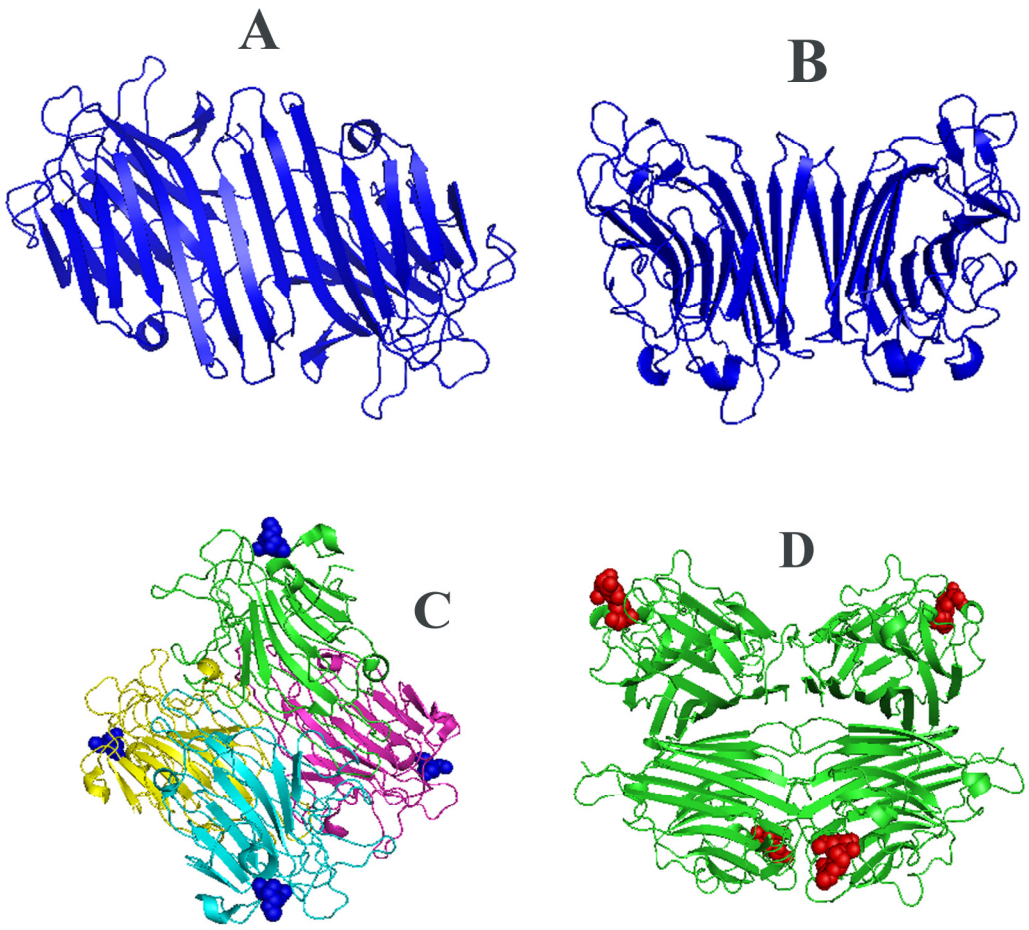

Figure 1.2. Mode of oligomerization in legume lectins: A canonical dimer of ConA (A), EcorL dimer (B), perfect tetrahedral shape of ConA tetramer (PDB ID-5CNA) with bound mannose (blue spheres) (C) and tetrameric assembly of peanut agglutinin (PDB ID-2PEL) with bound lactose (red spheres) (D). The coordinates of the lectins were obtained from PDB and figures were prepared using PyMOL (DeLano, 2002).

The central cleft formed by tetrameric association in legume lectins provides binding sites for non-carbohydrate ligands such as adenine and adenine-related plant hormones. Binding sites for hydrophobic molecules like 1,8-anilinonaphthalenesulfonic

acid (ANS), 2,6-toludinylnaphthalenesulfonic acid (TNS), adenine and phytohormones like cytokinin (Roberts and Goldstein, 1983a, b) have also been reported.

The legume lectin structural family also contains two proteins with no carbohydrate recognition activity; the α-amylase inhibitor (α-AI) and arcelin, a seed defense protein, both from *Phaseolus vulgaris* (Hamelryck *et al.*, 1996b). Expectedly, the metals are absent in them. Most of the residues involved in metal-binding and carbohydrate recognition in other legume lectins are not conserved in α-AI and arcelin (Bompard-Giles *et al.*, 1996).

1.2.7.2 Chitin–binding lectins containing hevein domains

This family of protein consists of one or more functional hevein domain(s), which is a glycine/cysteine rich structural motif of about 43 amino acid residues isolated from the latex of the rubber tree (*Hevea brasiliensis*) (Waljuno *et al.,* 1975). They are considered to be one of the most widespread lectin families in plant kingdom. The chitin binding lectins include the class I chitinases, Heveins, Gramineae, Phytolacca, Solanaceae, Urticacea, Papavaracaceae and Viscaceae families spread over monocots and dicots. There are other chitin binding lectins without hevein domain, for example chitin binding legume lectins and Cucurbitaceae phloem lectins, which have no sequence similarity to the hevein domain and hence are not considered in this family (Van Damme *et al.*, 1998). These lectins are found in seeds and vegetative tissues. The lectins of Gramineae are the best characterized chitin-binding seed lectins.

Most of the pioneering work of chitin-binding lectins was done with wheat germ agglutinin (WGA) which was the first lectin of this family to be isolated, characterized

(Nagata and burger, 1972; leVine *et al.* 1972), three-dimensional structure determined (Wright 1977) and cloned (Raikhel and Wilkins, 1987). The three-dimensional structure of wheat germ agglutinin is a dimer built from the noncovalent association of two monomers, I and II, each of which consists of four structurally similar hevein domains (Wright 1987). Except for a very short α-helix of only five residues, each hevein domain is devoid of regular secondary structures and consists mainly of coil and turn structures. WGA exists in three isoforms (WGA-1, WGA-2 and WGA-3) which differ by 5-8 amino acid residues. The high resolution crystal structures have been reported for all the three isolectins in native form as well as with sugar molecules (Wright, 1992; Wright & Jaeger, 1993). The crystal structure of the complex reveals four potential sugar-binding sites per dimer (Fig.1.3).

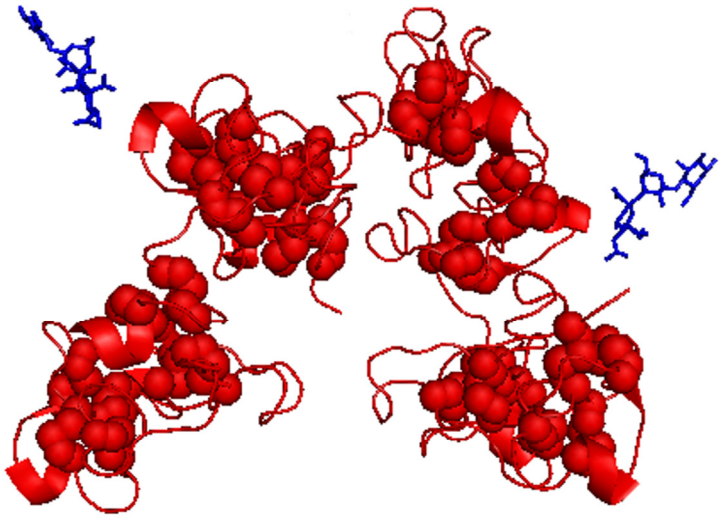

Figure 1. 3. Crystal structure of WGA (PDB ID-1WGC) in complex with N-acetyl neuraminyl lactose (ANL) shown in blue sticks. The disulphide bridges are shown as spheres.

Chitin-binding lectins are heterogeneous family of proteins constituting merolectins, hololectins and different types of chimerolectins. The simplest chitin-binding lectins are the merolectins composed of a single hevein domain of which two types have been reported. The first type belong to hevein and the hevein like protein whereas the second type of single domain chitin-binding lectins are the chitin-binding antimicrobial peptides from *Amaranthus caudatus* (AcAMP) (Broekaert *et al.,* 1992). Hevein contains three strands of β-sheet and two short α-helices.

Hololectins in this family are composed of polypeptides containing two, three, four, or seven tandemly arrayed hevein domains. At present, several hololectins built up of polypeptide chains with two hevein repeats have been described for example the *Urtica dioica* (stinging nettle) agglutinin (UDA) is a monomer of a single polypeptide chain of 89 amino acid residues that consists of two hevein repeats separated by a four amino acid hinge region (Peumans *et al*., 1984; Beintema and Peumans, 1992). The dimeric Wheat Germ Agglutinin (WGA) and other Graminae lectins containing four hevein domains (Raikhel *et al.*, 1993) are also reported.

Chimeric lectins belonging to this family are of two types. In Class I chitinases the single *N*-terminal chitin binding domain is linked to a catalytically active chitinase domain (Collinge *et al.*, 1993; Beintema, 1994). The second type consists of dimeric lectins from Solanaceae family for example, *Lycopersicon esculentum* agglutinin and *Solanum tuberosum* agglutinin. These proteins consists of chimeric polypeptides containing an *N*-terminal chitin binding domain and three hevein repeats linked to a highly *O*-glycosylated serine-hydroxyproline-rich domain (Kieliszewski *et al.*, 1994; Allen *et al.*, 1996). Structurally, none of them have been characterized.

The sugar-binding activity and specificity of the chitin-binding lectins are determined exclusively by their hevein domains. These lectins exhibit a marked specificity towards GlcNAc and GalNAc -oligomers; they also bind sialic acid. The overall specificity of the chitin-binding lectins has been conserved during evolution indicating that the binding to chitin- or (GlcNAc)-containing glycoconjugates is probably essential for their function.

1.2.7.3. Type II RIP and related lectins

Lectins belonging to this family are capable of catalytically inactivating eukaryotic ribosomes (Barbieri *et al*., 1993). The basic mechanism involves enzymatic removal of specific adenine residue from a highly conserved loop in the large subunit of the ribosomal RNA causing a conformational change in the affected loop. This prevents binding of elongation factor EF2 to the ribosome leading to the arrest of protein synthesis and cell death (Van Damme *et al.,* 1998a)

RIPs are subdivided into two groups: type 1 and type 2. Type 1 RIPs consist of a single polypeptide of about 30 kDa with polynucleotide adenosine glycosidase activity (PAG); whereas type 2 RIPs contain enzymatically active A chain (Endo *et al*., 1987) and a carbohydrate binding B chain (Lord *et al*., 1994) with lectin activity. The similar protomers are linked by a disulfide bridge between the two chains. Mostly, RIPs are single-chain type-1 proteins, but a few members of this family possess a galactose-specific lectin domain that binds to cell surfaces thus forming the chimeric type 2 RIPs. Type 2 RIPs are present in both monocots (for example Iridaceae and Liliaceae) and dicots (such as Euphorbiaceae, Fabaceae, Sambucaceae, Viscaceae, Ranunculaceae,

Lauracaceae, and Passifloraceae). They occur in seeds as well as in different vegetative tissues. Similar type 2 RIPs have been isolated from the seeds of camphor tree, bitter gourd and mistletoe (Peumans & Van Damme,1999). In the last decade prof. Swamy's group has reported several type 2 RIPs in seed lectins from cucurbitaceae family such as *Trichosanthes anguina* (Komath *et al.*, 1996), *T. cucumerina* (Padma *et al.*, 1999) and *T. dioica* (Sultan *et al.*, 2004).

Type 2 RIPs are potent toxins, the best known of which is "ricin" isolated from the seeds of *Ricinus communis*. Ricin was the first type 2 RIP whose primary and three-dimensional structures were resolved (Montfort *et al.*, 1987). This was also the first type 2 RIP to be cloned (Lamb *et al.*, 1985). At present, several type 2 RIP and their corresponding genes have been isolated and characterized.

The three-dimensional structure of ricin was determined at 2.5 Å resolutions by X-ray crystallography (Rutenber *et al.*, 1991; Rutenber and Robertus, 1991). Ricin shows specificity for Gal/GalNAc (Lord *et al.*, 1994) and consists of two subunits A and B of molecular weight ~32 kDa held by a disulfide bridge (Fig. 1.4). The catalytic activity of ricin resides in A chain which possess regular secondary structure consisting of eight α-helices and six β-sheets.

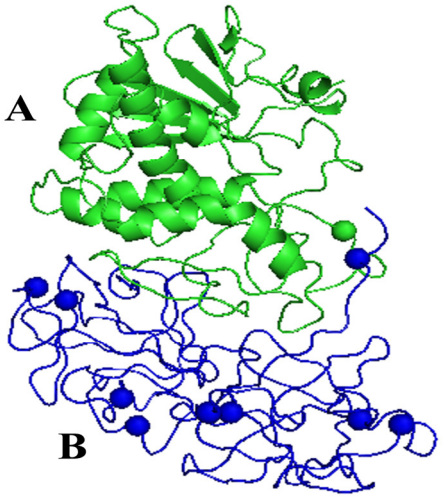

Figure 1. 4. Structure of Ricin (PDB ID-2AAI). The carbohydrate binding B chain is shown in blue. The disulphide bridges are indicated in blue spheres.

The B chain has no well defined regular secondary structures and consists mainly of coil structures linked by turns and loops which is involved in the carbohydrate binding. Four disulphide bridges between cysteine residues 20-39, 63-80, 151-164 and 190-207 stabilize the folding of B chain. The two domains in the B chain show a tertiary folding characteristic of the β-trefoil family (Murzin *et al.*, 1992) which was first observed in the crystal structure of Kunitz type trypsin inhibitor from soybean. The B chain contains two N-glycosylation sites, Asn95–Gly96–Thr97 and Asn135-Asn136–Thr137.

Examples of monomeric type 2 RIPs are the *Ricinus communis* and *Abrus precatorius* agglutinins. The only known tetrameric type 2 RIP is the Siaα2-6Gal/GalNAc-binding lectins from Sambucus sp. (Van Damme *et al.,* 1998). Three-dimensional structure of type 2 RIP from *Abrus precatorius* seeds (abrin) is similar to that of ricin. Abrin is a heterodimer of 34 kDa and 32 kDa subunits joined together by single disulphide bridge (Tahirov *et al.*, 1995; Hedge *et al.*, 1991; Wu *et al.*, 2001).

1.2.7.4. Monocot mannose-binding lectins

These lectins belong to the superfamily of strictly mannose-specific lectins, found exclusively in monocotyledonous plants. Their strict specificity for D-mannose sets them apart from the Glc/Man/Gal specific family of dicotyledonous legume lectins and the C-type mannose binding animal lectins. Monocot mannose-binding lectins are structurally and evolutionary unrelated to the mannose/ maltose-specific Convolvulaceae lectins that are classified as jacalin-related lectins. They are detected in various vegetative tissues such as leaves, flowers, ovaries, bulbs, tubers, rhizomes, roots (Van Damme *et al.*, 1995) but not in seeds. The first lectin of this class was identified in snowdrop (*Galanthus*

nivalis) bulbs belonging to Amaryllidaceae family (Van Damme *et al.,* 1987). Later on they were isolated and characterized in different monocot families from Alliaceae, Araceae, Bromeliaceae, Orchidaceae, Liliaceae and Iridaceae.

Besides their unique specificity for mannose they have marked resemblance with respect to their amino acid composition, sequence, molecular structure and serological properties. However, their tertiary and quaternary structures exhibit significant structural diversity due to number, structure and assembly of protomers. The majority of monocot mannose-binding lectins consist of two or four identical one-domain protomers of 12 kDa that are held together by noncovalent interactions. Others, however, are composed of one, two or four protomers consisting of two similar or dissimilar domains.

Galanthus nivalis agglutinin (GNA) was the first lectin of this family to be crystallized (Wright *et al.*, 1990) and analyzed by X-ray diffraction (Hester *et al.*, 1995). The three-dimensional structure of this lectin shows a β-prism II fold, a characteristic of this family. GNA is a homotetramer composed of four identical noncovalently bound monomers of 109 residues (12 kDa). The monomer consists of three tandemly arrayed subdomains (I, II, and III) each of which consists of a four-stranded β-sheet. The three subdomains have a local three-fold symmetry, and form three faces of a triangular prism. Each of the subdomain has a carbohydrate recognition domain (CRD). They are connected by loops and form a 12-stranded β-barrel which exhibits three mannose-binding sites located in the clefts formed by the three bundles of β-sheet. Two such monomers form tight dimers through hydrogen-bond contacts stabilized by *C*-terminal strand exchange which in turn associate into tetramers through hydrophobic interactions.

Although these lectins are specific towards mannose, they also recognize mannose containing glycoproteins due to which they are able to show inhibitory activity on the *in vitro* replication of retroviruses (Balzarini *et al.*, 1991 and 1992). A comparison of the stretches of dimeric garlic lectin and tetrameric snowdrop lectin showed how oligomerization could be used as a strategy for generating carbohydrate-specificity. The binding site of the carbohydrate component of gp120 straddles the two dimers in tetrameric snowdrop lectin (Fig.1.5 A). This extended binding site cannot obviously exist in dimeric garlic lectin (Fig. 1.5. B), resulting into the loss of binding to gp120. Thus, the tetrameric snowdrop lectin can act as anti-retroviral agent whereas the dimeric garlic lectin cannot (Vijayan and Chandra, 1999).

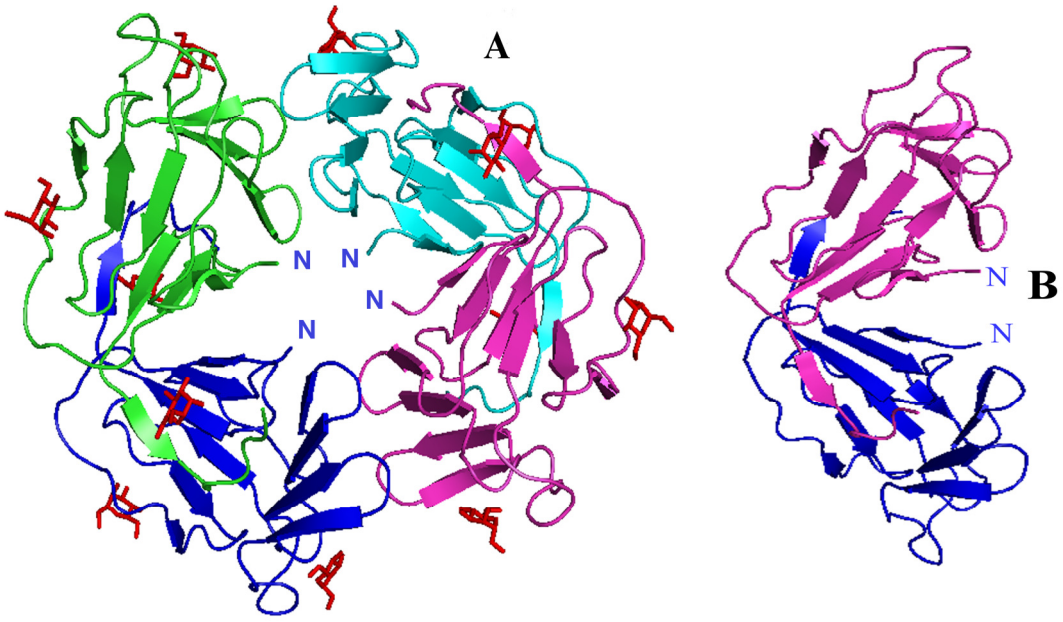

Figure 1.5. Representative members of monocot mannose binding lectin family. **(A)** Crystal structure of tetrameric snowdrop lectin (GNA) with bound mannose shown in red sticks (PDB ID-1MSA) and **(B)** dimeric garlic lectin (PDB ID-1KJ1).

Apart from mannose-specific lectins, certain lectins with complex-sugar specificity have also been reported to occur in some monocots, for example *N*-acetyl-D-lactosamine (LacNAc) specific lectins from *Arisaema flavum* (Singh *et al.*, 2004), *Alocasia cucullata* (Kaur *et al.*, 2005a), *Arisaema tortuosum* (Dhuna *et al.*, 2005) and *Arundo donax* (Kaur *et al.*, 2005b) etc. The crystal structure of one such complex-sugar specific lectin from *Scilla campanulata* bulbs (SCAfet) has been reported (PDB code 1DLP; Wright *et al.*, 2000). Unlike most single domain monocot mannose-binding lectins, for example, GNA and SCAman, SCAfet contains two domains with approximately 55% sequence identity, joined by a linker peptide. Each domain is made up of a 12-stranded β-prism II fold, with three putative carbohydrate-binding sites, one on each subdomain (Fig. 1.6). The lack of interaction of SCAfet with simple sugars could be due to the replacement of key amino acid residues (Asn, Asp, Gln, Tyr) within the monosaccharide-binding pocket by hydrophobic residues.

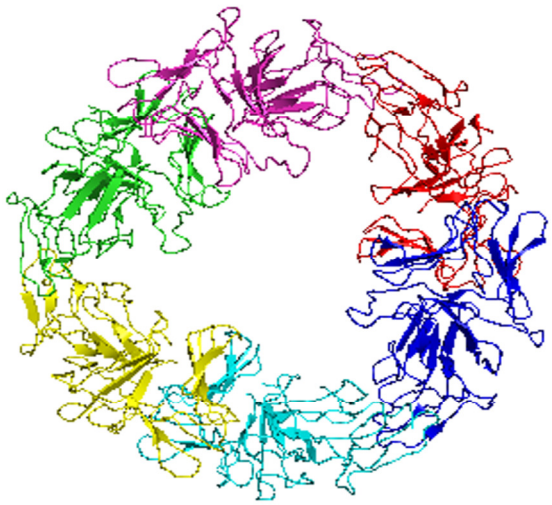

Figure 1.6. Crystal structure of hexameric SCAfet lectin from *Scilla campanulata* (PDB-ID1DLP) showing binding to fetuin though grouped into the family of monocot mannose binding lectin.

1.2.7.5. Jacalin related lectins (JRLs)

Jacalin related lectins derive their name after the lectin jacalin, from *Artocarpus integrifolia*, the first member of the family that has been identified from the seeds of jackfruit. Jacalin also happens to be the first member of this family to be studied by X-ray crystallography which revealed a novel lectin fold termed as the β-prism I fold (Sankaranarayanan *et al.*, 1996).

Jacalin related lectins have been reported from Moraceae (jacalin, artocarpin, MPA, hirsuta), Convulvulaceae (calsepa, conarva), Asteraceae (*helianthus tuberosus*), Gramineae (barley and wheat lectins) and Musaceae (banana lectin) family. Moraceae lectins are usually considered typical seed proteins. Jacalin is the most abundant seed protein comprising more than 50% of the total soluble protein (Kabir, 1998). Within the Convolvulaceae family, lectins have been isolated from rhizomes of *Calystegia sepium* (hedge bindweed) and *Convolvulus arvensis* (bindweed). In addition, a jacalin like lectin has also been reported from sweet potato (*Ipomea batatas*).

Jacalin is a tetrameric protein, each subunit of which is made up of a heavy α-chain of 133 amino acids and a light β-chain of 20 amino acids. The crystal structure of jacalin indicated that each of its subunits exhibited a type-1 β-prism fold comprised of three Greek keys (4-stranded β-sheet) giving rise to a 3-fold symmetric β-prism which are arranged like the three faces of a prism (Sankaranarayanan *et al.*, 1996). The crystal structures of other lectins in this family, artocarpin from *Artocarpus integrifolia* (Pratap *et al.*, 2002), heltuba from *Helianthus tuberosus* (Bourne *et al.*, 1999), MPA from *Maclura pomifera* (Lee *et al.*, 1998) and calsepa from *Calystedia sepium* (Bourne *et al.*,

2004), confirm this fold to be characteristic of the family (Fig. 1.7), although significant differences in quaternary associations are observed. These crystal structures also indicate one carbohydrate binding site per subunit and the residues forming the binding site emerge from different loops at one end of the prism. Sugar binding properties and the details of the structural fold are described in detail in chapter 5 of the thesis along with structure of recombinant Jacalin.

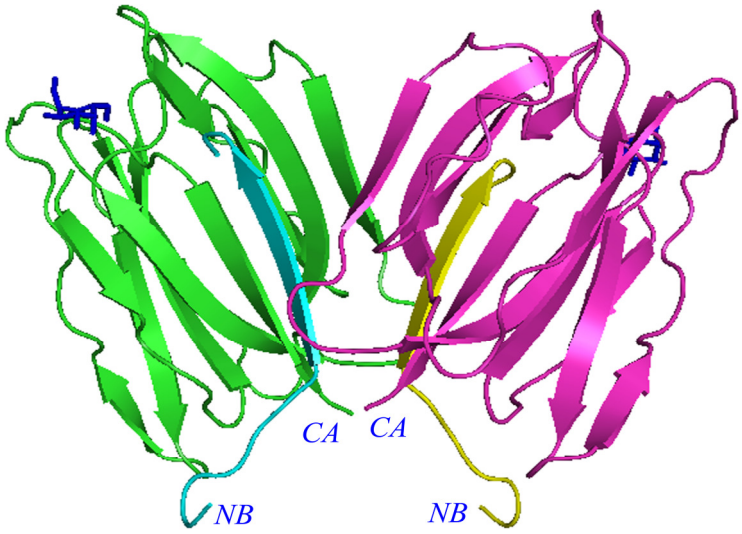

Figure 1.7. A dimer of Jacalin (PDB-ID1JAC) with methyl α-D-galactose bound near the N-terminus of glycine 1 in each monomer shown in blue sticks.

1.2.7.6. Amaranthins

The name Amaranthin lectin family was deduced from the first identified lectin of this class from *Amaranthus caudatus* seeds. Subsequently various lectins were identified in several species of Amaranthaceae family e.g., *A. caudatus*, *A. spinosus*, *A. leucocarpus* and *A. cruentus*. The term 'Amaranthins' is now collectively used for the closely related

GalNAc-specific seed lectins from various Amaranthus species. The amaranthins are small lectin family which does not show similarity with any other known protein. Amaranthin lectins are homodimeric in nature comprised of two identical non-glycosylated subunits of about 33 kDa (Rinderle *et al.*, 1989). Amaranthins were inhibited best by N-acetylgalactosamine and fetuin (Rinderle *et al.*, 1990). Lectin from *A. caudatus* was also found to react specifically with the T-antigenic disaccharide and its α-linked glycosides.

Amaranthus caudatus lectin was the first from this family whose primary and three-dimensional structures were determined by X-ray crystallographic analysis (Transue *et al.*, 1997). The 299 amino acid long polypeptide of the lectin consists of two homologous domains (called N- and C-domains) linked by a short helix, each having a β- trefoil structure similar to that of the two domains of the ricin B chain. However, unlike the ricin B chain (which has no extended secondary structure), the domains of amaranthin consist of six strands of antiparallel β-sheet capped by three β-hairpins into a β-barrel. The two domains are linked by a 3_{10} helix. Two monomers associate in head to tail fashion to form a dimer where the N-terminal domain of one monomer faces the C-terminal domain of the other monomer. The dimer is held together by extensive noncovalent contacts between the monomers. This dimeric organization provides two surface-exposed carbohydrate binding sites as shallow depressions formed at the interface between the N- and C-terminal domains of the two facing monomers.

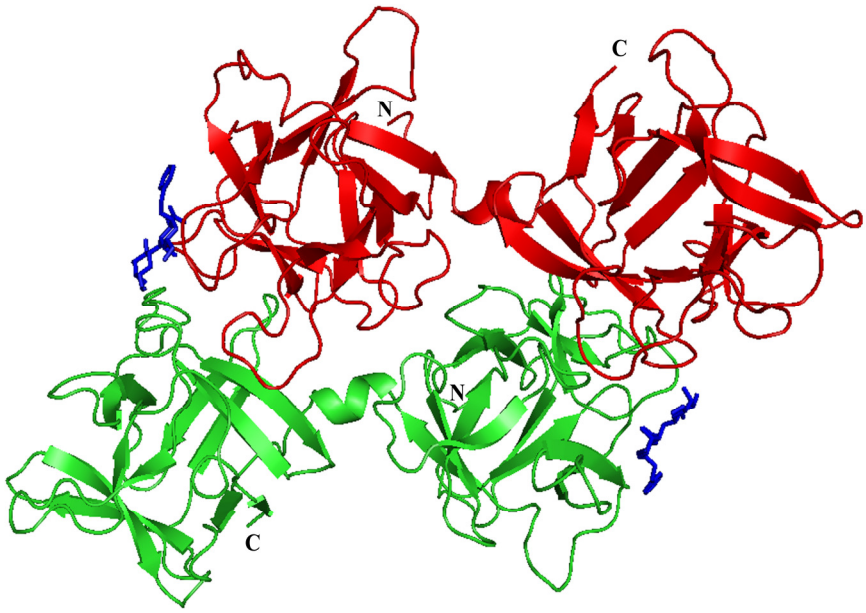

Figure 1.8. The β-trefoil fold of *Amaranthus caudatus* lectin. (PDB ID-1JLX). The N and C-terminal domains are held by short helix. The two monomers are shown in red and green color. The benzyl T-antigen disaccharide bound to the lectin is shown in stick representation (blue).

1.2.7.7. Cucurbitaceae phloem lectins

The Cucurbitaceae phloem lectins are a small family of chitin-binding agglutinins found in the phloem exudate of Cucurbitaceae species. They are not related to other Cucurbitaceae lectins (e.g., the type 2 RIP from *Trichosanthes kirilowii*) and do not contain hevein domains. The species identified to possess these lectins belong to the genera Citrullus, Coccinia, Cucumis, Cucurbita, Luffa and Sechium.

All these phloem lectins are dimeric with two non-glycosylated subunits of 25 kDa each (Wang *et al.*, 1994). Chitin oligosaccharides have been found to be strong inhibitors of this class of lectin in terms of their sugar binding activity while N-

acetylglucosamine is a weak inhibitor of lectin activity. The Cucurbitaceae phloem lectins exhibit specificity towards oligomers of GlcNAc. Studies on *Luffa acutangula* lectin have demonstrated that the inhibitory potency of the GlcNAc-oligomers strongly increases with chain length up to five residues. The *Luffa acutangula* and the *Cucurbita maxima* lectins also recognize the internal di-N-acetylchitobiosyl sequences of N-linked glycan chains from fetuin, ovalbumin, and soybean lectin (Van Damme *et al.*, 1998). Being the major protein of the phloem exudate they may have an anti-parasitic function. There is no report available on the three-dimensional structure of these lectins.

1.3. Structural basis of lectin carbohydrate interactions

Detailed crystallographic studies on lectin-carbohydrate complexes and thermodynamic measurements have provided a wealth of information on the structural basis of lectin–sugar interactions. The affinity of lectins for monosaccharides is usually weak (with Ka in the mM range) yet they are highly selective (Lis and Sharon, 1998; Rini, 1995; Weis & Drickamer, 1996). In general, lectins that bind to Gal do not recognize Glc/Man and *vice versa*. Within a subgroup, they exhibit different affinities for different derivatives. Lectins that bind Gal usually bind GalNAc, generally with a stronger affinity, but there are exceptions for example PNA. GS4 and PHAL are two legume lectins that do not bind to monosaccharides at all.

The atomic features of lectin-carbohydrate interactions have been studied in considerable detail to identify common recognition principles (Lis & Sharon, 1998; Loris *et al.*, 1998; Bouckaert *et al.*, 1999). Mechanisms for sugar recognition have evolved independently but bear some common basic features, namely, the hydrophobic (van der Waals) interactions and hydrogen bonds with shared portions of sugars, such as the ring

oxygen which stabilize the sugar-lectin interaction. Selectivity for different sugars is achieved through a combination of hydrogen bonding to the sugar hydroxyl groups with van der Waals interaction, often including packing of a hydrophobic sugar face against aromatic amino acid side chains as well as metal coordination bonds to key distinguishing hydroxyl groups. Higher selectivity of binding is achieved by extended binding sites through additional direct and water mediated contacts between oligosaccharides and the proteins surface. (Drickamer, 1995; Elgavish and Shannan, 1998). Dramatically increased affinity for oligosaccharides result from clustering of simple binding sites in oligomers of the lectin polypeptides (Weis and Drickamer, 1996). Along with clustering strategy the orientation of the binding sites is also essential for higher affinity and functional role.

Lectins apply various strategies to generate carbohydrate specificity, as evident from crystal structures of lectin-sugar complexes. They are: 1) The variation in quaternary structure generates carbohydrate specificity in the case of legume and bulb lectins (Chandra *et al.*, 1999; Vijayan and 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) Proteolytic processing to generate free amino terminus involved in carbohydrate specificity has been observed in the case of jacalin (Sankaranarayanan *et al.*, 1996) and related lectins (Vijayan and Chandra, 1999). 3) Water molecules generating sugar specificity is seen in sugar complexes of PNA, Ecorl and other lectins (Elgavish and Shannan, 1998; Ravishankar *et al.*, 1997). 4) Lectins show enhanced binding capacity for oligosaccharides (in the nM range) through extended binding site or subunit multivalency (Elgavish and Shannan, 1997). The

extended binding sites are observed 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 in complex with mannopentose (Wright and Hester, 1996).

1.4. Biological role of plant lectins

Lectins are present abundantly in many plant sources. For example, as much as 10% of the total content of many mature plant seeds are made up of lectins. Due to their abundance they are considered to be genuine storage proteins. Despite this abundance, their precise biological roles in the plants to which they belong, are not well understood. The available evidences suggest two main roles for them.

Mediators of symbiosis

Legume lectins are considered to be mediators of symbiosis in plants and microorganisms (Diaz *et al.*, 1989; Brewin and Kardailsky, 1997). Lectins localized at the root hairs are the entry sites for rhizobia. The lectins then aggregate the rhizobia in the root nodules and make them immobile. Type specificity of hostparasite interactions between leguminous plants and particular strains of rhizobia infecting them is determined by lectins (Hirsch, 1999; Gage, 2004). The expression of the pea lectin gene in white clover roots enabled them to be nodulated by a rhizobium strain specific for the pea plant (van Eijsden *et al.*, 1995). It was suggested that rhizobial attachment to plant root occurs by direct interaction between bacterial surface carbohydrates and lectins present in the roots.

Protection of plants from predatory animals and phytopathogens

A major argument used in favor of defensive role for plant lectins is the observation that some plant lectins bind glycoconjugates not found in plants but is present in other organisms. Legume lectins show preferential specificity towards typical animal glycans implicating their role in the plant's defense against insects and/or predating animals (Chrispeels and Raikhel, 1991; Peumans and Van Damme, 1995). Sialic acid binding lectins from elderberry bark (Sambucus species) bind sialic acid which is absent in plants but forms a major constituent of animal glycoproteins. Chitin binding plant lectins recognize a carbohydrate that is a typical constituent of the exoskeleton of invertebrates and not found in plants. Moreover, most plant lectins show a marked stability under unfavorable conditions and in this respect resembles other defense related proteins like protease and α-amylase inhibitors. The preferential association of lectins with those parts of the plant that are most susceptible to attack by foreign organisms is also an argument in favor of their protective role (Peumans and Van Damme, 1995). The anti-insect activity of many plants has been attributed to the presence of lectins in them. For example, PHA, peanut agglutinin (PNA), WGA, Maclura pomifera agglutinin (MPA) and lectins from potato, thorn apple and osage orange show anti-insect activity against cowpea weevil.

Class I chitinases possess a biological activity that is unrelated to their carbohydrate-binding activity due to the presence of a non-catalytic domain. They are implicated for the plant's defense against fungi (Collinge *et al.*, 1993). The physiological role of the nonenzymatic chitin-binding lectins is still unclear. Some chitin binding lectins may also be involved in the plant's defense against bacteria, for example the seed

lectin from *Datura stramonium* (Broekaert and Peumans, 1986) and insects (Murdock *et al.*, 1990; Huesing *et al.*, 1991).

RIPs exhibit antiviral activity *in vitro* against plant viruses and hence can play role in plant's defense against these viruses (Barbieri *et al.*, 1993; Kumar *et al.*, 1993). They can also be linked to antibodies or ligands to form immunotoxins or conjugates specifically toxic to a given type of cell.

1.5. Application of Lectins:

Sugars are the information bearing molecules with a vast potential for structural diversity. Lectins are specialized in deciphering this information by offering complementary surfaces. This decoding is the basis for a large repertoire of functions of lectins in biological systems. Lectins have the capability to serve as recognition molecules within cells, between cells or between organisms by virtue of their binding specificity and thus find extensive applications in biochemical and medical research. Few of the common applications of lectins in various disciplines as categorized by Rüdiger and Gabius (2001) are listed below:

A. Biochemistry

- Detection of defined carbohydrate epitopes of glycoconjugates in blots or on thin-layer chromatography plates.
- Purification of lectin-reactive glycoconjugates by affinity chromatography.
- Glycan characterization by serial lectin affinity chromatography.

- Glycome analysis (glycomics).
- Quantification of lectin-reactive glycoconjugates in enzyme-linked lectin-binding assays (ELLA).
- Quantification of activities of glycosyltransferases/glycosidases by lectin-based detection of products of enzymatic reaction

B. Cell biology

- Characterization of cell surface presentation of glycoconjugates and their preceding intracellular assembly and routing in normal and genetically engineered cells.
- Analysis of mechanisms involved in correct glycosylation by lectin-resistant cell variants. Fractionation of cell populations.
- Modulation of proliferation and activation status of cells.
- Model substratum for study of cell aggregation and adhesion.

C. Medicine

- Detection of disease-related alterations of glycan synthesis.
- Blood group typing and definition of secretor status.
- Quantification of aberrations of cell surface glycan presentation, for example in malignancy.

- Cell marker for diagnostic purposes including infectious agents (viruses, bacteria, fungi, parasites).

Plant lectins have been utilized extensively in biomedical research. These applications primarily employ precipitations and aggregations. A brief account for few of them is presented below.

I. ***Blood typing:*** The oldest application of plant lectins is in blood typing. Fucose specific lectins from *Lotus tetragonolobus* and *Ulex europaeus* are employed for identifying blood type O cells. *Dolichos biflorus* agglutinin is used to distinguish between A1 and A2 subgroups whereas *Vicia graminea* lectin is able to select out blood type N from a mixture of M and N type cells. Peanut lectin and *Vicia villosa* aggregate T and Tn cell types, respectively. An N-acetylgalactosamine specific lectin from lima bean agglutinates A red cells. Basic winged bean agglutinin (WBAI) is specific to A and B blood group substances while acidic winged bean agglutinin is specific to O blood group substances (Slifkin and Doyle, 1990; Sharon and Lis, 2004; Sharon, 2005).

II. ***Analysis and purification of glycoconjugates:*** The ability to bind specific carbohydrates makes lectins an obvious tool for differentiating glycoconjugates on the basis of sugar present in them. Glycoproteins containing mannose, galactose or N-acetylneuraminic acid are purified in laboratories using concanavalin A (ConA), PNA and wheat germ agglutinin, respectively (Liener *et al.*, 1986). Jacalin, in addition to its ability to selectively bind IgA1 subclass (Hagiwara *et al.*, 1988), binds a number of human plasma glycoproteins. They are IgD, C1-inhibitor, C4-

binding gp120, hemopexin, plasminogen, α1-antitrypsin, α2-macroglobulin, 8S α3-glycoprotein and α2-HSG (Kondoh *et al.*, 1986; Aucouturier *et al.*, 1987; Hiemstra *et al.*, 1987, Hortin and Trimpe, 1990; To *et al.*, 1995).

III. ***Histochemical and cytochemical probes:*** Cell surface carbohydrates undergo well marked changes during the process of development. Lectins have been used as markers to characterize these changes (Gabius, 1991). The ability to agglutinate specific cell types is used for cell separation. Peanut agglutinin, soybean agglutinin (SBA) and jacalin are particularly utilized in this regard (Sharon and Lis, 1989a). The ability of PNA to distinguish between mature and immature thymocytes has been used in bone marrow transplantations (Reisner, 1987). The exclusive specificity of PNA for the Thomson-Friedenreich antigen (T-antigen) is exploited widely for monitoring its differential expression for both the prognosis and diagnosis of malignancies (O'Keefe and Ashman, 1982; Zabel *et al.*, 1983; Zebda *et al.*, 1994). Jacalin has been used as a histochemical reagent to study tissue-binding properties in benign and malignant lesions of the breast and thyroid (Remani *et al.*, 1989; Vijayakumar *et al.*, 1992). Jacalin has also been utilized to distinguish malignant cells from benign reactive cells in serous effusions (Sujathan et al., 1996). Jacalin binds efficiently to different kind of oral carcinomas (Pillai *et al.*, 1996). Lectins derivatised with gold particles, fluorescent dyes or enzymes are employed as histochemical and cytochemical reagents for detection of glycoconjugates in tissue sections, on cells and subcellular organelles, and in investigations of intracellular pathways of protein glycosylation (Lis and Sharon, 1998; Christiane *et al.*, 2004).

IV. ***Mitogenic stimulation of lymphocytes:*** Certain lectins act as potent mitogens, by activating lymphocytes and inducing them to divide. For example, phytohemagglutinin (PHA-L) and ConA, stimulate T lymphocytes, while pokeweed mitogen (PWM) stimulates both T and B cells (Di Sabato *et al.*, 1987; Ashraf and Khan, 2003). PNA could agglutinate lymphocytes from rat, mouse and human only after their treatment with neuraminidase. Mitogenic simulation by lectins provides easy and simple means to assess the immunocompetence of patients suffering from a diversity of diseases including AIDS, and to monitor the effects of various immunosuppressive and immune therapeutic manipulations. Jacalin is the only lectin known to be selectively mitogenic for human $CD4^+$ T cells and this property has been applied in AIDS research (Pineau *et al.*, 1989, 1990; Corbeau *et al.*, 1995; Favero *et al.*, 1993; Lafont *et al.*, 1994, 1996). Since $CD4^+$ T lymphocytes act as a receptor of HIV-1, jacalin also has been used to investigate the proliferation of PBMC in HIV-1 infected patients (Pineau *et al.*, 1989; Tamma *et al.*, 1996). *Canavalia brasiliensis* (ConBr), *Pisum arvense* (PAA) and *Artocarpus integrifolia* (KM+) lectins are used as immunostimulatory molecules in vaccination against *Leishmania amazonensis* infection. They induce IFγ, enhance the expression of MHC II, CD80, and CD86 and thereby reduce the level of parasite (Teixeira *et al.,* 2006).

V. ***Mapping of neuronal pathways****:* Lectins also find application in neuroanatomy research. Lectins are being used as tracers for mapping neuronal connections; WGA was the first to be used in this context (Gerfen and Sawchenko, 1985). Lectin-HRP conjugate is readily taken up by neurons and transported along the axon thereby

help in tracing the central neuronal pathways. The same strategy when employed with a cytotoxic lectin (e.g. ricin) function as suicide transporters offering a new means to tackle neurobiological disorders. PNA and WGA are used to label the interphotoreceptor matrix in cryosections of retinal tissue. PNA readily labeled the cone-associated matrix however its binding is weak to the rod-associated matrix; WGA labeled both the rod- and cone-associated matrices efficiently (Kristina *et al.,* 1991).

VI. ***Lectin mediated drug targeting and delivery:*** Based on the fact that oligosaccharides encode biological information, the biorecognition between lectinised drug delivery systems and glycosylated structures in the intestine could be exploited for improved peroral therapy (Gabor *et al.*, 2004). Basic research revealed that lectins such as WGA, ConA, PNA and jacalin can mediate mucoadhesion, cytoadhesion and cytoinvasion of drugs (Yi *et al.*, 2001). Entering the vesicular pathway by receptor mediated endocytosis, part of the conjugated drug is accumulated within the lysosomes. In addition, part of the drug is supposed to be transported across the epithelium. As exemplified by lectin-grafted prodrug and carrier systems, this strategy is expected to improve absorption and probably bioavailability of poorly absorbable drugs, peptides and proteins as well as therapeutic DNA (Gabor *et al*., 2004). Conjugation of lectins with suitable drugs or vice versa, enhances the drug delivery to the epithelial cells. Lectins from mistletoe (ML I-III), stinging nettle (UDA), tomato (TL, LEA), and WGA have been used to target human intestinal epithelial Caco-2 cells. ConA, WGA and TL are used to deliver the drug formulation to the intestinal cells. UEA-I, DBA, WGA and GS-I-

B4 are targeted to M cells in the nasal cavity. With its higher affinity towards cerebral endothelium cells, WGA has been known to cross the blood-brain barrier, without disturbing the brain function (Christiane *et al.*, 2004; Gao *et al.,* 2006).

1.6. Plant lectins with complex sugar specificity

The crystal structures of several plant lectins, from different sources and possessing varying specificities have been determined (Loris *et al.*, 1998). Most of the lectins are specific for simple sugars and their hemagglutination activity is inhibited by monosaccharides or oligosaccharides. However, with advances in lectin studies, it was observed that many lectins do not show specificity towards simple sugars and are inhibited by glycoproteins like fetuin, asialofetuin, thyroglobulin, fibrinogen, ovalbumin etc. and their corresponding glycopeptides. In the last decade several lectins showing specificity for complex sugars have been reported from different plant families (table 1.1).

In few of the cases especially with legumes and Gramineae family, lectins that differ in their sugar specificity have been found in the same plant (Lis & Sharon, 1986). Few of the lectins of different sugar specificity present in the same plant are listed in table1.2.

Table 1.1: Few of the plant lectins showing specificity for complex sugars.

Lectin	Plant family	Source	References
Phaseolus vulgaris, (PHA-E and PHA-L)	Leguminoseae	Seed	Kamemura *et al.*, 1993 Kaneda *et al.*, 2002
Cicer arietinum	Leguminoseae	Seed	Kolberg *et al.*, 1983
Gonatanthus pumilus(GPL)	Araceae	Tuber	Dhuna *et al.*, 2007
Arisaema curvatum (ACL)	Araceae	Tuber	Shanghary *et al.*, 1995
Arisaema flavum	Araceae	Tuber	Singh *et al.*, 2004
Arisaema helleborifolium	Araceae	Tuber	Kaur *et al.*, 2006a
Sauromatum venosum	Araceae	Tuber	Singh Bains *et al.*, 2005
Arisaema tortuosum Schott	Araceae	Tuber	Dhuna *et al.*, 2005
Arisaema jacquemontii Blume	Araceae	Tuber	Kaur *et al.*, 2006b
Acacia constricta	Araceae	Seed	Guzmán-Partida *et al.*, 2004
Scilla campanulata (SCAfet)	Liliaceae	Bulb	Wright *et al.*, 1999
Amaranthus viridis	Amaranthaceae	seed	Kaur *et al.*, 2006c
Ficus cunia	Moraceae	Seed	Ray *et al.*, 1993
Ficus bengalensis	Moraceae	Seed	Singha *et al.*, 2007
Saraca indica	Fabaceae	Seed	Ray and Chatterjee, 1995
Salvia sclarea	Lamiaceae	Seed	Piller *et al.*, 1986
Salvia bogotensis	Lamiaceae	Seed	Vega *et al.*, 2006
Glechoma hederacea	Lamiaceae	Seed	Singh *et al.*, 2006
Moringa oleifera	Moringaceae	Seed	Katre *et al.*, 2008a

Table 1.2: Examples of lectins with different sugar specificity present in the same plant source.

Plant Source	Lectin	Sugar-specificity	Reference
Ulex eurpoeus	I II	L-fucose GlcNAc	Goldstein and Hayes, 1978
Artocarpus integrifolia	Jacalin Artocarpin	Galactose and its derivatives Mannose and its derivatives	Sankarnarayanan *et al.*, 1996 Pratap *et al.*, 2002
Vicia Cracca	I II	GalNAc Man, Glc	Baumann *et al.*, 1982
Cicer arietinum	CAA-II CAL	Gal, GalNAc Fetuin	Qureshi *et al.,*2006 Kolberg *et al.*, 1983
Scilla campanulata	SCAman SCAfet	Mannose Fetuin	Wright *et al.*,1999
Morus nigra	MornigaM MornigaG	Mannose Galactose	Van Damme *et al.,*2002

Amongst plant lectins legume lectins which are specific for simple sugars are well characterized and crystal structures of most of them are available. Purification, characterization and biological role of several lectins showing specificity for complex sugars (table 1.1) are reported. However, structurally they are not well characterized and the only structure available is for a lectin from *Scilla campanulata* (SCAfet) (Fig. 1.6), showing hemagglutination inhibition with fetuin (PDB code 1dlp; Wright *et al.*, 2000).

Studies on lectins showing specificity for complex sugars are fewer and there is a need to characterize them in detail to understand the molecular basis of their biochemical properties and differences contributing towards the sugar recognition.

Few of the Araceae lectins listed in table 1.1 also possess anti-insect (Kaur *et al.*, 2006a, b) as well as anti-fungal (Kaur *et al.*, 2006c) properties, suggesting their possible use in protecting crop plants from the attack of insects and fungal pathogens. Some of the recently discovered lectins from Araceae family have shown strong mitogenic and/or anti-proliferative effect against human cancer cell lines (Singh *et al.*, 2004; Dhuna *et al.*, 2007), hence can be utilized in cancer research and therapy.

1.7. Scope for the Work

Structural studies on few lectins showing specificity for complex sugars were started in our laboratory to get a detailed insight into differential mode of sugar recognition by these lectins. Studies on structure-activity relationship and recognition of complex sugars has been previously reported for some of them such as *Cicer arietinum* lectin (CAL) and *Moringa oleifera* lectin (MoL) (Katre *et al.*, 2005; Katre *et al.,* 2008a and b). Dharkar *et al.,* (2009) had reported biophysical characterization of two Araceae lectins involved in recognition of complex sugars.

The present thesis describes the structure-function studies of three hemagglutinins from different plant families; all showing specificity for complex sugars. The first hemaglutinin is from the seed of legume *Cicer arietinum* (CAL), second belongs to the Araceae family extracted from the tubers of *Arisaema curvatum* (ACL) and the third one is MoL isolated from the seeds of *Moringa oleifera* a member of Moringaceae family.

Purification and preliminary X-ray characterization of CAL has already been reported from our laboratory (Katre *et al.*, 2005). Two crystal forms of CAL were obtained: orthorhombic ($P2_12_12$) with unit cell dimensions a=70.9, b=73.3, c=86.9 Å and trigonal (P3) with a=b=80.2, c=69.1 Å and $\beta=120^{o}$. An iodine derivative of orthorhombic form was also characterized (Katre U V, Ph.D thesis 2007). However, structural studies remained inconclusive due to various difficulties encountered in getting a second heavy atom derivative as well as the lack of sufficient information including protein sequence. CAL is basically a plant albumin which also shows hemagglutination activity. Various cultivars of chickpea were screened for sufficient amount of lectin for structural studies. CAL purified from a particular cultivar Pusa-256 showed maximum hemagglutination activity and gave diffraction quality crystals. Despite several attempts a second heavy atom derivative could not be obtained. Meanwhile the structure of a PA-2 albumin was reported recently in PDB. Thus the Structure of CAL was determined using molecular replacement method and the newly reported structure. The analysis of the CAL structure is described in chapter 2 of the thesis.

During the last decade, lectins with interesting properties have been isolated and characterized from various monocot families including, Araceae. Most of the lectins from these families belong to a single monocot mannose-binding lectin superfamily as revealed by their molecular structure, sequence homologies, and exclusive specificity for mannose. However, a few other studies revealed the occurrence of some monocot lectins having specificity for complex glycoproteins (table1.1) and not for mannose. *Arisaema curvatum* lectin (ACL) is a 13 kDa protein isolated from the tubers of *Arisaema curvatum* showing hemagglutination against rabbit, rat and sheep RBC's only (Shanghary *et al.*,

1995). The hemagglutination activity of the lectin is inhibited by asialofetuin alone, while simple sugars/derivatives including chitin, porcine mucin and fetuin did not react. The lectin showed mitogenic potential for human blood lymphocytes. There is no report available for the structure-function studies of this lectin so far. Hence, biochemical and biophysical characterizations as well as attempt to crystallize the protein were undertaken in order to determine the three-dimensional structure of the protein.

The third hemagglutinin studied is MoL, obtained from the seeds of drumstick. Purification and characterization of the lectin was reported earlier from our lab (Katre *et al.*, 2008a). MoL is a tetrmeric protein of subunit molecular weight 7 kDa and highly basic in nature due to high arginine content. The protein is very stable and active in a wide range and at extremes of pH. Presence of three disulphide bonds confers unusual high thermostability to MoL. Drastic change in the structure of MoL was observed upon treatment with reducing agents. The details of the conformational flexibility and high thermal stability shown by the lectin were not explored previously. It was interesting to study the conformational changes occurring to MoL in the presence of reducing agent. To investigate the dynamics of flexible regions of MoL under native and unfolded conditions, solutions studies using NMR and other spectroscopic tools were carried out.

Jacalin is a tetrameric two chain glycoprotein of 66 kDa, isolated from the seeds of jackfruit. Each monomer of Jacalin consists of a heavy chain (α) of 133 amino acid residues and a light chain (β) of 20 amino acid residues. The reported crystal structure has suggested that the newly generated amino terminus in the α-chain of jacalin due to post-translational cleavage at glycine residue may be contributing to the higher affinity for galactose and its α- linked derivatives (Sankarnarayanan *et al.*, 1996). Interestingly,

jacalin has several other family members with different sugar specificity. Jacalin can distinguish even minor changes in the orientation of hydroxyl groups of hexoses. Taking into consideration the multi chain requirement of jacalin and its proteolytic processing to generate carbohydrate specificity/affinity it was highly relevant to clone and express jacalin in single chain unglycosylated form to investigate the multimeric or multichain form and glycosylation requirement of jacalin in generating its sugar specificity/affinity. The gene of jacalin was cloned as a single chain (where the α chain and β chain are linked by T-S-S-N loop) in pT_7Nc vector between Nco I and EcoR I, unlike in native protein, to make a single chain protein in recombinant jacalin (rjacalin) (Sahasrabudhe *et al*., 2004). The rjacalin showed 100-fold less affinity for galactose in the single chain form. The rjacalin failed to bind Mannose, though the similar lectin Artocarpin which is also unglycosylated and single chain shows binding to mannose. To provide the structural basis for the reduction in sugar recognition/affinity shown by the rjacalin, the clone of rjacalin was obtained from our collaborator Dr. M.V. Krishnasastry and rjacalin was expressed, purified and used for structural studies.

Apart from lectins protease inhibitors are also implicated for their action in plant defense against predators. Protein proteinase inhibitors are one of the most studied classes of plant defense proteins. During our studies on CAL, an 18 kDa protein co-purified along with CAL from the chickpea seeds, which was identified later as a trypsin inhibitor (CPTI) belonging to the soybean type kunitz family. The CPTI showed no homology to other kunitz type inhibitors reported from legumes. Investigating the details of functional properties of this protein will have significant bearing because the complete chickpea genome is not yet sequenced and the mature seed contains many hypothetical

proteins whose structural and functional studies have not been initiated, apart from the usual storage proteins such as albumins and globulins. Structural studies were initiated on CPTI for detailed characterization. The crystallization and preliminary X-ray characterization studies of CPTI are reported in the chapter 6 of the thesis with a short introduction on plant protein protease inhibitors.

Chapter 2

Three-dimensional structure of a seed lectin from *Cicer arietinum* (CAL) showing specificity for complex sugars

2.1. Summary

Cicer arietinum lectin (CAL) shows specificity towards complex sugars and has crystallized in two crystal forms, orthorhombic ($P2_12_12$) with unit cell dimensions a=70.9, b=73.3, c=86.9 Å and trigonal (P3) with a=b=80.2, c=69.1 Å and β=120°. The three-dimensional structure of CAL is reported in this chapter. The structure is determined using molecular replacement module *PHASER* of the *CCP4* suit (Collaborative Computational Project, Number 4, 1994). The coordinates of pa 2 albumin (PDB-ID 3LP9) from *Lathyrus sativus* were used for phasing protein reflections. *REFMAC 5* of *CCP4* suit was used for the refinement and the CAL model in orthorhombic crystal form could be refined to an R_{free} of 22.3 % and R_{work} of 17.2 % in the resolution range of 19.9 to 2.2 Å. The refined coordinates of the orthorhombic form of CAL were used to obtain the structure in the trigonal form. The refined structures in orthorhombic and trigonal forms were compared. The differences in the packing arrangement of molecules in both the structures have also been analyzed.

Structurally the CAL exists as a stable dimer. The single monomer is organized as four β-propeller blades exhibiting a pseudo four-fold axis of symmetry passing through the centre of the monomer which forms a channel. The channel of each monomer provides ligand binding sites which are occupied tightly by the calcium ion followed by a water molecule and iodine atom. However none of these ligands are found near the sugar binding site. To get an idea about the putative sugar binding site and to identify the residues involved in sugar binding the chemical modification studies were carried out. However, the sugar specificity of the lectin for any monomer sugar could not be detected

in hemagglutination assays. Also based other studies we consider the lectin to have complex sugar specificity.

2.2. Introduction

Chickpea (*Cicer arietinum* L.) is the world's third most important pulse crop (Food and Agriculture Organization of the United Nations, 1993) and an important source of vegetable proteins in many countries, extensively grown in India, Mexico, and the Mediterranean region. It is a rich source of carbohydrates and proteins, accounting for 50 and 25% (w/w), respectively (Vioque *et al.*, 1999). Apart from this, they are also a good source of soluble and insoluble fiber, vitamins and minerals. Soluble fibers are known to control the blood cholesterol level and hence reduce the risk of heart diseases. Insoluble fiber helps preventing digestive disorders. Seeds contain low amounts of fats most of which are polyunsaturated, their intake helps in preventing cardiovascular diseases. India produces 75 % of the total world's supply of chickpea. There are mainly two types of chickpea varieties i.e. Desi and Kabuli. The Desi ones (Fig.2.1A) are relatively smaller in size having a thicker seed coat, they appear dark brown and are consumed in a variety of ways. The fresh seeds are eaten as raw snack while the seed flour (*besan*) is utilized in many Indian recipes. Kabuli chickpeas (Fig.2.1B) have a whitish-cream color, relatively bigger in size and have a thinner seed coat. They are generally used in soups /salads. Kabuli type is cultivated in temperate regions while the Desi type is cultivated in the semi-arid tropics (Malhotra *et al.*, 1987). The Desi and Kabuli type contributes to around 80 and 20% of the total production of chickpea. The chickpea used for the purification of CAL in the present research belongs to the Desi variety.

Figure 2.1. Mature seeds of chickpea (*Cicer arietinum*). Desi variety **(A)** and Kabuli variety **(B).**

Chickpea contains two major protein fractions: globulins and albumins. Globulins, the storage proteins, are mainly constituted by legumin and vicilin and represent 60-80 % of the extractable proteins. The albumin fraction, less abundant than globulins, represents 15-25 % of the total cotyledonary proteins. The albumins form a heterogeneous class of unrelated proteins which play essential role in the seeds because they include most of the metabolic and enzymatic and/or regulatory proteins. In addition, albumins are of nutritional importance because of their high content of lysine and sulphur containing amino acids.

Plant albumins are known to act as lipoxygenases (Clemente *et al.*, 2000), glycosidases (Agrawal and Bahl, 1968), proteinases (Schlereth *et al.*, 2001), proteinase inhibitors (Kochhar *et al.*, 2000), and lectins (Harley and Beevers, 1986). PA2 albumin is one of the most abundant albumins in chickpea (Vioque *et al.*, 1998) and in pea seeds (Croy *et al.*, 1984). PA2 albumin from chickpea was described as a lectin that agglutinates enzymatically treated erythrocytes (Kolberg *et al.*, 1993; Vioque *et al.*, 1998; Katre *et al.*, 2005). In legumes, most lectins are constituted from four identical peptides,

or a combination of two light and two heavy chains (Loris *et al.*, 1998). In contrast to other legume lectins, PA2 is not a glycoprotein and does not require Mn^{2+} or Ca^{2+} for agglutination. The agglutination is unstable at room temperature and requires previous treatment of erythrocytes with papain or pronase (Kolberg *et al.*, 1993: Vioque *et al.*, 1998; Katre *et. al.*, 2005). Thus, considering its molecular weight, amino acid sequence, quaternary structure, absence of glycosylating sugars, and other physicochemical requirements for agglutination, the pa2 albumin from chickpea although can be considered as a lectin (CAL) is different from commonly found lectins in legumes. Chickpea pa2 albumin has been known to bind hemin in a saturable manner, a property that may be important when considering its possible physiological roles (Pedroche *et al.*, 2005).

There is a previous report of the isolation of a lectin from chickpea (Kolberg *et al.*, 1983) but no further studies have been reported for this lectin. X-ray crystallographic studies were initiated in our lab on *C. arietinum* lectin (CAL) possessing complex-sugar specificity isolated from mature chickpea seeds. Crystallization and preparation of one heavy atom derivative of this protein was previously reported (Katre U V, 2007). However, these studies remained inconclusive due to various difficulties encountered in getting pure protein preparation and reproducing good quality crystals which can withstand heavy atom soaking.

To understand the basis of complex specificity of this lectin towards glycoproteins structural details are prerequisite which might reveal a new lectin fold itself. Fresh purification and X-ray crystallographic studies were carried out on CAL to elucidate its structure.

2.3. Materials

Mature seeds of *Cicer arietinum* (chickpea), cultivar PUSA-256, were obtained from the Pusa agriculture university, Pusa, Bihar, India. Sodium chloride, tris base, ammonium sulphate, potassium dihydrogen phosphate, dipotassium hydrogen phosphate, sodium citrate and sodium acetate were purchased from SRL, India. DEAE-cellulose, SP-sephadex, sodium cacodylate, polyethylene glycols 200, 300, 400, 600, 1K, 2K, 4K, 5K, 6K, and 8000 (PEG 8K), glycerol, ethylene glycol, p-Hydroxy mercury benzoate (PCMB), mercuric acetate and MPD were purchased from Sigma, USA. Sodium molybdate, $Pb(NO_3)_2$ and KI were purchased from SRL, India. All chemicals used were of analytical grade.

Crystallization buffers for appropriate pH were prepared in the lab. Glass capillaries (1.5 mm) and nylon loops used to mount the crystals were purchased from Hampton Research, USA. The 24-well tissue culture trays used for hanging-drop vapor-diffusion method of crystallization were from Axygen and Corning. The circular coverslips (Φ = 19 mm) were of Blue Star make. The Cu-Kα radiation was generated using a rotating anode X-ray generator from Rigaku-MSC, USA which was equipped with a confocal mirror focusing system. X-ray diffraction data were collected on an R-AXIS IV^{++} image plate. To collect diffraction data at low temperature, the crystal was flash cooled in a nitrogen stream produced by X-stream (Rigaku-MSC, USA).

Different crystallographic softwares such as *DENZO* and *MOSFLM* to integrate the diffraction images and *SCALEPACK* or *SCALA* to scale the data were used during data processing. A Silicon Graphics workstation (Octane) with Irix 6.5 as the operating

system and HP workstations with Linux as the operating system were used to run these programs.

2.4. Methods

2.4.1. Purification of *Cicer arietinum* lectin (CAL)

CAL was purified by using two successive column chromatographic steps on DEAE-cellulose and SP-sephadex columns according to the procedure described by Katre *et al.*, (2005). The pure protein solution was dialyzed against deionized water and concentrated to ~ 20 mg ml^{-1} prior to setting up crystallization experiments.

2.4.2. Hemagglutination assay

Rabbit RBCs were washed 5 times with PBS (phosphate buffer saline, 10 mM potassium phosphate buffer pH 7.2 containing 150 mM NaCl). A 3% (v/v) suspension of the erythrocytes in the same buffer was prepared. The hemagglutination assay was performed according to the method of Gurjar *et al.,* (1998).

2.4.3. Chemical modification studies

Modification of the single tryptophan present in CAL was carried out by treating the protein (50 μg) with freshly prepared N-bromosuccinimide (1-5 mM) according to the method of Spande and Witkop, (1967). The residual activity was determined by hemagglutination assay for modfified protein separately.

Modification of histidine residues with diethyl pyrocarbonate (DEPC) was carried out according to the method of Ovadi *et al.*, (1967). The reagent was prepared in absolute ethanol just prior to use. The lectin (50 μg) in 50 mM phosphate buffer, pH 7.2

was treated with 2-10 mM DEPC separately for 30 mins and the residual hemagglutination activity was determined in each case.

Modification of carboxylate groups with Woodward's reagent K (WRK) was carried out as described elsewhere (Sinha and Brewer, 1985). CAL (200 μg) was incubated in 50 mM citrate-phosphate buffer pH 6.0, with different concentrations (5-20 mM) of WRK. Aliquots were removed after 15 mins in each case and the residual activity was determined.

Arginine residues of CAL were modified according to the method of Takahashi, (1968). The lectin (50 μg) incubated in 50 mM phosphate buffer pH 8.0 was treated with phenylglyoxal (prepared in methanol) in the range of 0.5-5 mM, for 30 min at 25 °C and the residual hemagglutination activity was determined for each aliquot. The methanol concentration in the reaction mixture did not exceed 2 % (v/v) and had no effect on the activity and stability of the lectin during the incubation period.

Modification of tyrosine residues of CAL were with N-acetyl imidazole. The reaction was carried out as described by Riordan *et al.*, (1965). To 50 μg of CAL incubated in phosphate buffer pH 7.2, *N*-acetyl imidazole was added in the concentration of 2-10 mM separately and hemagglutination assay was performed for the modified protein.

Modification of Serine with Phenylmethylsulfonyl fluoride (PMSF) was by the method of Gold and Fahrney, (1964). The lectin (200 μg) in 50 mM Tris-HCl buffer, pH 8.0 was incubated with 5-10 mM PMSF, at RT, for 60 min. Aliquots were removed after 15 mins and residual activity was determined by hemagglutination assay.

Modification of lysine with Trinitrobenzenesulphonic acid (TNBS) was carried out as previously described (Habeeb, 1966). The reagent was dissolved in 0.1 M Tris buffer, pH 8.0 to make a stock solution of 0.1 M. 50 μg of protein in 0.1 M tris buffer, pH 8.0 was allowed to react for 1 hr with TNBS up to final concentration of 2, 4, 5 and 10 mM and the residual activity was determined in each case.

The native lectin without any reagent was used as a control for hemagglutination activity in each case of chemical modification.

2.4.4. X-ray crystallographic studies of CAL

The three-dimensional structure of biological macromolecules can be studied by various experimental methods such as X-ray crystallography, Nuclear magnetic resonance (NMR) and cryo-electron microscopy. Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for three-dimensional structure determination of proteins and nucleic acids at atomic resolution, since the NMR data can be recorded in solution. Unlike X-ray crystallography, NMR studies do not require crystals, at the same time the dynamics of the structure can be analyzed. The limitation is that NMR can be used to determine structures of proteins having only molecular weights up to ~ 30kDa. The Cryo-electron microscopy has been utilized in interpreting and visualizing unstained biological complexes of very large size (200 kDa or larger) such as ribosomes or viruses.

X-ray crystallography is a powerful tool to obtain atomic resolution details of the three-dimensional structure of biological macromolecules. Crystallographic techniques find wide applications in the understanding of macromolecular assembly, enzyme

mechanism, mode of action and activation of enzymes, substrate-specificity, ligand-binding properties, domain movement, etc.

The structure determination of proteins using X-ray crystallography consists of the following steps: 1. protein purification, 2. crystallization, 3.data collection and processing, 4. calculation of phases and electron density calculation 5. map interpretation and model building, 6. model refinement and 7. structure validation and analysis. A graphical summary of the methods utilized in X-ray crystallography to obtain the final model of a protein is shown in Fig. 2.2. The methods employed for structure determination of CAL starting from crystal to final model will be briefly discussed, emphasizing crystallization, cryo-crystallography, data collection, molecular replacement method and refinement.

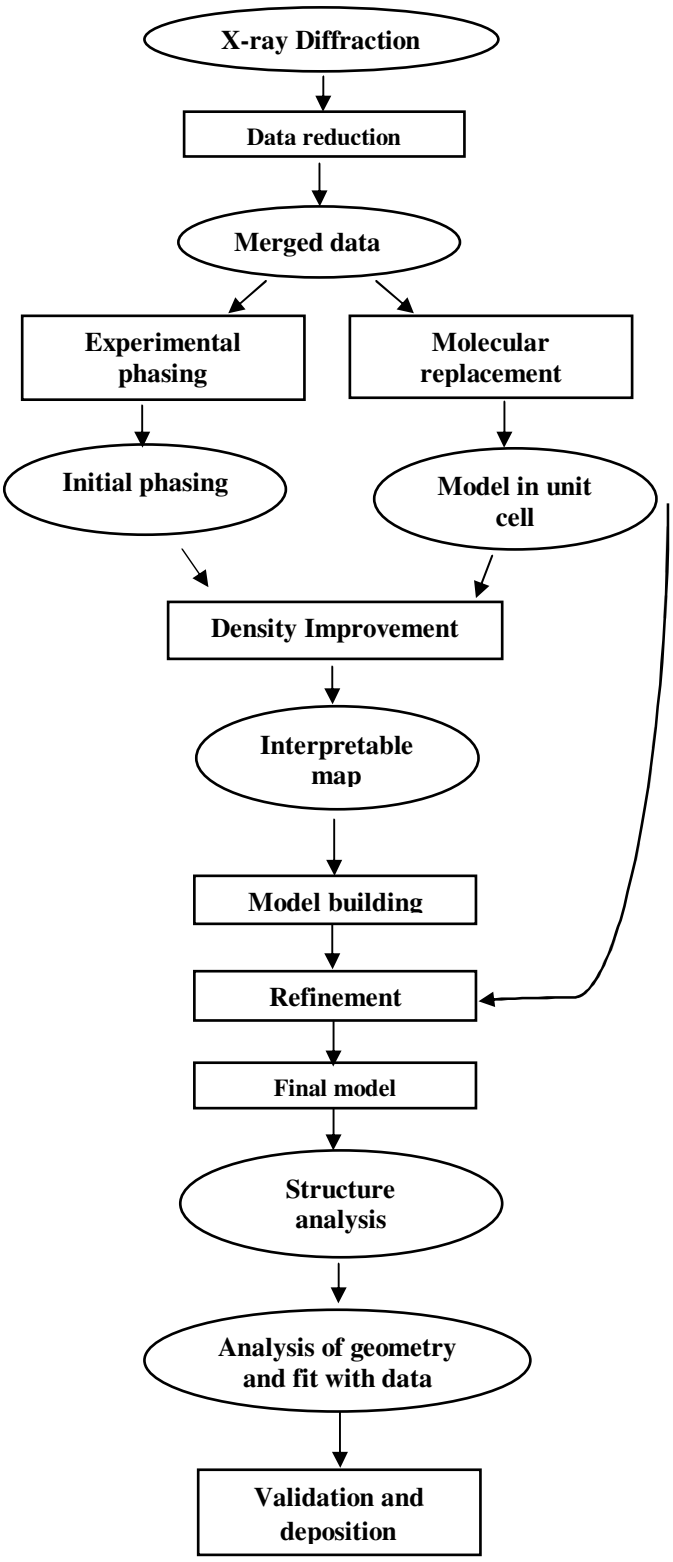

Figure 2.2 Schematic representation of the methods used in protein X-ray crystallography (Adapted from CCP4 tutorial, http://www.ccp4.ac.uk/dist/ccp4i/help/roadmaps/index.html).

2.4.4.1. Crystallization

Crystallization is the first crucial and most often the rate-limiting step in structure determination process. The success of a protein crystal structure determination depends heavily on the quality of crystal grown. In the absence of any well established theory behind the mechanism of crystallization, protein crystallization is generally considered to be a trial and error procedure and more as an "art" rather than a "science". Several factors are known to play important role in crystallization process and thus affect the quality of crystals obtained, such as temperature, pH, precipitating agents, ionic strength of the solution, presence of impurities and nucleating agents, additives and also several other unknown factors. Crystallization is known to lower the free energy of proteins by ~3-6 kcal/mole relative to the solution state (Drenth and Haas, 1992).

Over the years, a number of methods have been developed to grow quality protein crystals. Of these, vapour-diffusion, free interface diffusion, batch and dialysis methods are most commonly used (McPherson, 1982; Giege and McPherson, 2001). The protein to be crystallized often dictates the method of choice. Among various crystallization techniques known, hanging-drop vapor-diffusion method is widely adopted and has produced more protein crystals than all other methods combined (Chayen, 1998). This method is simple, consumes less protein and it is easy to monitor the progress of crystallization.

The vapour-diffusion method is based on the diffusion of vaporisable species between solutions of two different concentrations in achieving supersaturation of the solute protein. Two most popular methods used to achieve this state are the hanging drop and the sitting drop methods. In a typical hanging drop experiment set up in multiwall

trays, the protein solution (2 µl) is mixed with a precipitant solution (2 µl) containing buffer and additives on a siliconised glass cover slip and suspended over the well of a tissue culture plate containing the reservoir solution and kept undisturbed for equilibration. The volume of the reservoir solution, which contains the precipitant and additives at a higher concentration, is much larger than the drop volume and thus, while equilibration supersaturation is achieved favoring the crystal growth. The sitting drop method involves a very similar procedure except that the protein drop is kept seated over the reservoir buffer instead of being suspended from the coverslip. Consequently, a larger volume of protein solution is required for growing the crystals. The benefit of using the vapour diffusion method in particular, the hanging drop method, is the low amount of protein required and large number of conditions that can be screened.

In the present research the hanging- drop vapour-diffusion method have been used for crystallization of CAL and other proteins. CAL crystals were reported previously (Katre *et al.*, 2005) from condition no. 28 of crystal screen 1(Hampton research, USA) containing 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5 and 30% (w/v) polyethylene glycol (PEG) 8000). In the present preparation of protein sample, CAL purified from cultivar PUSA-256, these conditions were explored further to reproduce good quality crystals by varying PEG8k concentrations as well as with additives like L-arginine (5-25 mM), L-glycine (5-25 mM), L-histidine (5-25 mM), NaCl (25-200 mM) and ammonium sulphate (25-200 mM). To overcome the difficulties due to growth of extremely small crystals, seeding was also tried. Several cryoprotectants were screened for CAL crystals to collect data at low temperature, namely, PEG's 200, 300, 400, 600, 1K, 2k, 4k, 5k, 6k, and 8K in the range of 20-40 %. In addition to PEG's Glycerol (10-25

%), Ethylene glycol (10-30 %), various combination of Paraffin and Silicon oil (1:1, 1:2, etc), MPD (20-35 %) were also tried. These cryoprotectants were incorporated to the mother liquor. The crystal was briefly soaked (1-2 seconds) in the cryoprotectant solution and flash freezed immediately in the liquid nitrogen jet from X-stream.

2.4.4.2. Preparation of heavy atom derivatives of CAL crystals

To prepare heavy atom derivatives of CAL co-crystallization as well as quick soaking approach were tried. To co-crystallize CAL very low concentration of the freshly prepared heavy metal salt in the range of 1-5 mM was mixed with protein solution prior to setting up crystallization experiments. Alternatively, pre-formed crystals of CAL were soaked in the mother liquor containing heavy metal salts upto 10 mM to check the maximum concentration crystal withstands. The time of soak was varied from few seconds to several hours. Salts of various heavy metals, namely $HgCl_2$, mercuric acetate, $Pb(NO_3)_2$, KI, p-Hydroxy mercury benzoate (PCMB) and sodium molybdate were used in the crystallization and soaking experiments.

2.4.4.3. Data collection

After obtaining crystals of suitable size and quality the second step towards crystal structure determination involves obtaining diffraction pattern of the crystal by exposure to monochromatic X-ray beam.

Diffraction intensities of CAL crystals were recorded at the Macromolecular X-ray diffraction facility available at Division of biochemistry, National chemical laboratory (NCL), Pune, India. X-ray diffraction data were collected using Cu-Kα X-rays from a Rigaku rotating anode X-ray generator operated at 50 kV and 100 mA and a R-

Axis IV^{++} image plate detector. The crystal alignment was done by viewing the image captured using a CCD camera on a TV monitor. The different processes such as exposure, data collection, readout and storage of data were carried out automatically via the Crystal clear program supplied by Rigaku-MSC. Crystal-to-detector distance was chosen based on the longest unit cell dimension and mosaic spread etc, so that the diffraction spots could be well resolved. For CAL crystals the detector distance was kept at 150 mm to achieve reasonable resolution of the data. During data collection, the crystals were oscillated through an angle of 0.5^0 about an axis perpendicular to the direction of the X-ray beam. The exposure time for data collection was chosen depending on the intensity of spots. For room temperature data collection exposure time given was 2-3 mins, whereas for low temperature it was 4-8 mins.

2.4.4.4. Data processing

The analysis and reduction of single crystal diffraction data collected using an imaging plate detector consists of six major steps:

1. Visualization and preliminary analysis of the original, unprocessed, diffraction pattern.
2. Indexing of the diffraction pattern and determining of the crystal orientation
3. Refinement of the crystal and detector parameters
4. Integration of the diffraction intensities
5. Refining the relative scale factors between equivalent measurements

6. Precise refinement of crystal parameters using the whole dataset

7. Merging and statistical analysis of the measurements related by space group symmetry.

The first four steps of data processing were carried out using *DENZO* and *XDisplayF* while the steps 5-7 were executed using *SCALEPACK*. All these three modules are part of HKL suit of programs (Otwinowsky and Minor, 1997) and were run on a Silicon Graphics Octane workstation. *DENZO* reads the diffraction data and outputs a file containing the intensity of spots, the estimated errors and the corrected intensities (Otwinowsky and Minor, 1997). The scaling and the merging of the reflections and post-refinement are done using *SCALEPACK*, and it outputs the final intensities and the standard deviation and also the data quality statistics. *MOSFLM* can also be used for data processing. *MOSFLM* (Powell, 1999) performs the functions similar to that of *DENZO*, while *SCALA* of the *CCP4* suite is used to scale the reflections (CCP4, 1994).

2.4.4.5. Data quality statistics

The quality of the data collected and the resolution limit are decided based on the data quality statistics. The quality parameters output by *SCALEPACK* includes χ^2, R_{merge} and $\langle I / \sigma I \rangle$. These parameters are defined by the equations given below (Otwinowsky and Minor, 1997).

$$R_{merge} = \Sigma_{hkl} \Sigma_j | I_{hkl.j} - \langle I_{hkl} \rangle | / \Sigma_{hkl} \Sigma_i I_{hkl.j} \quad (2.1)$$

$$\chi^2 = \Sigma [(Ij - \langle Ij \rangle)^2 / (\sigma I^2 N / (N-1))] \quad (2.2)$$

Where $I_{hkl,i}$ is the intensity of ith observation of (*h k l*) reflection, and $< I_{hkl} >$ is the mean intensity of all measured symmetry equivalents, σ(I) is the error in the measurement of I and N is the number of reflections. Σ_j is over the different measurements of I_{hkl} on its symmetry equivalents.

One of the least defined criteria in data collection is the resolution limit of diffraction (Dauter, 1997). It is important to restrict the resolution to the point below which more than half the reflection intensities are higher than 2σ. This assumes that the estimate of the errors in measurement of intensities is correct. For a good data set, Rmerge will be less than 10%. The Rmerge (also called Rsym when only symmetry equivalents of a single dataset are used) is the most widely accepted statistical parameter to indicate data quality in macromolecular crystallography. A good refinement will have χ^2 values closer to 1.0 in the different resolution bins. Very large values of χ^2 may be an indication of serious problems with data processing. The high resolution cut-off is decided such that the highest resolution bin has a Rmerge value less than 50% with at least 50% completion and <I/σI> value more than 2.0. Rmerge alone is not sufficient to decide the high resolution cut-off because of its dependence on the redundancy of data (Weiss, 2001; Weiss and Hilgenfeld, 1997).

2.4.4.6. 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 (2.3 and 2.4) are used for this calculation (Matthews, 1968).

$$V_m = (V\ X\ z) / (MW\ X\ n) \quad (2.3)$$

$$V_{solv} = 1- (1.23/ V_m) \quad (2.4)$$

Where V_m is the Matthews number, V is the unit cell volume, MW is the molecular weight of the protein in Daltons as part of the asymmetric unit, n is the number of asymmetric units per unit cell and z is the Avogadro's number; V_{solv} is the solvent content of the protein crystals.

In the case of CAL, the number of molecules in the crystallographic asymmetric unit was estimated using Matthews Probability Calculator. The solvent content was calculated assuming a dimer in the asymmetric unit.

2.4.4.7. Structure solution

In the diffraction experiment, we measure the intensities of waves scattered from Miller (hkl) planes of the crystal lattice which is related to the structure factor amplitudes. However, no phase information can be obtained by this experiment. Different methods are used to estimate the phases, essential for calculating the electron density using inverse Fourier transformation of the structure factors as given below.

$$\rho(xyz) = 1/V \Sigma\ m|F_{hkl}| \exp(i\alpha_{hkl})\exp[-2\pi i(hx + ky + lz)] \quad (2.5)$$

where, V is the volume of the unit cell, $|F_{hkl}|$ and α_{hkl} are the amplitude and the phase of the reflection hkl, and m is the figure of merit representing the probable phase error.

For phasing reflections in macromolecular crystallography three methods are used namely, Multiple Isomorphous Replacement (MIR), Multiple Anomalous Dispersion (MAD) and Molecular Replacement (MR). MAD technique is currently the most popular one owing to the availability of variable wavelength synchrotron radiation sources and sensitive detectors like Image plate and CCD to collect anomalous data on selenium derivative that can be prepared using molecular biology techniques for any protein (Beauchamp & Isaacs, 1999). Another advantage is that MAD needs only one heavy atom derivative compared to MIR technique that requires more than one. When the structure of a similar or homologous molecule is known, MR is the simplest and most successful technique to determine the target structure. The crystal structures reported in this thesis were solved using Molecular Replacement (MR) technique.

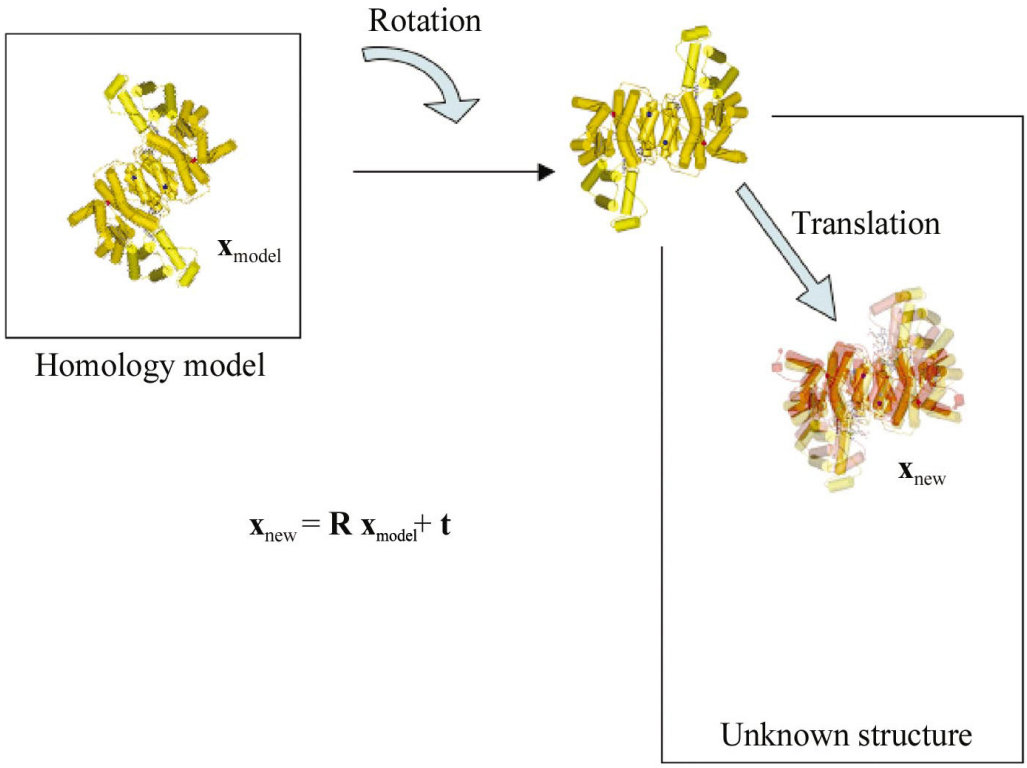

Figure 2.3. Molecular replacement method (Reproduced from Taylor, 2003)

The model has to be placed in the correct orientation and position in the target cell (Figure 2.3). This is carried out by a three-dimensional rotational search for molecular orientation, followed by a three-dimensional translation search for the molecular position. Patterson methods (Rossmann and Blow, 1962) are used to calculate the rotation function, to obtain the orientation of the model in the new unit cell, and then the translation function calculated using model in correct orientation (Crowther and Blow, 1967) helps to place the model in the new unit cell. As a rule of thumb, a sequence identity >25% is normally required and an r.m.s. deviation of <2.0 Å between the C-α atoms of the model and the final new structure, although there are exceptions to this. The

commonly used programs for molecular replacement are *AMoRe* (Navaza, 1994), *PHASER* (McCoy *et al*., 2005; McCoy, 2007; Storoni *et al.*, 2004) and *MOLREP* (CCP4, 1994).

AMoRe uses the fast rotation and translation functions for molecular replacement. It is reasonably fast and ranks the solution based on the Rfactor R and correlation coefficient CC for a translation t. Usually, Rfactor less than 50% and correlation coefficient greater than 30% are indicators of a correct solution, but the final test is always the quality of the electron density map.

MOLREP (CCP4, 1994) uses Crowther's fast rotation function, but differs from AMoRe in the translation function used. The modifications to the translation function results in an increase in the contrast. The incorporation of the packing function removes peaks of incorrect solutions with bad packing. It identifies correct solutions based on the contrast function.

PHASER (McCoy, 2007; Storoni *et al.*, 2004) uses maximum likelihood based methods for performing the rotation and translation searches. It has been found to be useful where many other molecular replacement programs fail to give a solution. It has been especially successful in problems involving high symmetry space groups with large number of molecules in the unit cell and those involving low homology models. The success of a *PHASER* run is determined based on the parameters, log likelihood gain (LLG) and Z-score. The LLG is the difference between the likelihood of the model and the likelihood calculated from a Wilson distribution. It is a measure of how much better the data can be predicted with the model than with a random distribution of the same

atoms. Z-score stands for the number of standard deviations over the mean value of the peaks in the rotation or translation functions. A Z-score value greater than 8 usually indicates that a solution has been obtained.

In the present investigation the program *PHASER* of *CCP4* suit was used for the structure calculation of CAL.

2.4.4.8. Structure refinement

The final step of protein crystallography is structure refinement. Crystallographic refinement aims at optimising the agreement of an atomic model with both observed diffraction data and chemical restraints. This is achieved by minimizing the discrepancy between the calculated (Fcalc) and observed structure factors (Fobs). To remove the many small local conformational errors remaining after rebuilding, the model is subjected to global restrained reciprocal space refinement. The parameters of the model atoms (coordinates x, y, z, and in later stages also individual B-factors except in very low resolution structures) together with overall parameters such as scale factor and overall B-factors, bulks solvent corrections, and anisotropy corrections are refined against all experimental data. One of the problems in macromolecular refinement is the poor observation to parameter ratio. Hence, restraints are included during refinement. The geometric restraints are bond angles, bond lengths, planarity, chirality, etc. (Engh and Huber, 1991). The refinement is carried out by the global minimization of a refinement target which is the residual between the observed experimental structure factor amplitudes and the model structure factor amplitudes. Most macromolecular refinement

programs have implemented maximum likelihood target functions, which are less susceptible to model errors and incompleteness than least squares target functions.

Refinement of structures reported in this thesis was carried out using restrained maximum likelihood refinement implemented in the program *REFMAC 5* (Murshudov *et al.*, 1996; Murshudov, 1997). Each cycle of the program can be considered to carry out grossly two steps by estimating the overall parameters of likelihood using the free set of reflections and using these parameters to build the likelihood function and refine the atomic parameters. To refine the atomic parameters only a working set of reflections were used. Maximum likelihood method of refinement performs the calculation of the first derivative and makes an approximation of the second derivative of the likelihood function with respect to refinement parameters and then estimating the shifts to be added to the parameters. The values of standard deviation are given for the geometric restraints used during the refinement. These values are also the estimated standard deviations that determine the relative weights of the corresponding restraints. The restraints used during refinement by *REFMAC 5* include distances, bond angles, torsion angles, peptide planarity, chiral volumes, Van der Waals radii and B values.

The progress in refinement is monitored by the Rwork which gives a measure of the difference between the observed structure factors |Fobs| and the structure factors calculated from the model |Fcalc|.

$$Rwork = \Sigma \mid |Fobs| - |Fcalc| \mid / \Sigma |Fobs| \qquad (2.6)$$

The free R value or Rfree was introduced for cross-validation, and it is a reliable indicator of model quality (Brunger, 1992; Brunger, 1993). In cross validation, the data is randomly divided into working set and test set. The test set contains about 5-10% of the reflections that are not included in the refinement process. The rest of the data forms the working set and refinement is carried out using only these reflections. Rfree is calculated according to equation (2.6) using only the reflections from the test set. If the model is correct, the Rfactor and Rfree should be close, as the changes to the model should affect both of them similarly. Thus, monitoring of Rfree serves as a check against over-fitting of the data. A difference between R and Rfree of about 6% is accepted.

Symmetry between the molecules in the asymmetric unit is termed as non-crystallographic symmetry (NCS). The presence of NCS is helpful during the structure determination process. NCS information can also be used as restraints during initial refinement. Since the lectins under present investigations 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 were refined independently in the final cycles of refinement.

2.4.4.9. Visualization and interpretation of electron density maps

After every refinement cycle, the model was inspected and manual rebuilding was carried out with the help of $2F_o$-F_c and F_o-F_c electron density maps. A 2Fo–Fc map gives more weightage to the observed structure factors, and hence reduces some of the model bias. It is normally viewed at a contour level of around 1.0σ. Fo–Fc map is examined at both positive and negative contours. Presence of a positive peak is an indication of

unaccounted density, while a negative peak shows that the model contains features not conforms to the actual structure. The program *COOT* ((Emsley and Cowtan, 2004) was used for displaying and examining the electron density maps, for displaying atoms, interactive fitting and optimizing the geometry and also for adding water molecules. The addition of water molecules was done at the final stages of refinement of the model. Initially water molecules were added at a higher sigma level where (Fo-Fc density > 3 σ) and by analyzing reasonable hydrogen bonds between them. Later, when the structures were refined solvent molecules were added up to3σ level, carefully checking the interactions.

2.4.4.10. Structure validation and analysis

Quality of the refined model is expressed in terms of the geometry and stereochemistry and the average B-factors of the model. Ramachandran plot (Ramachandran and Sasisekharan, 1968) serves as a powerful validation tool. A good model will have more than 90% of the residues in the allowed regions of the Ramachandran plot, and none will be in the disallowed region. 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. The program output consists of a comprehensive residue- by- residue listing of the parameters and its graphical representation. The program highlights the regions of the structure which 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 Å between donor

and acceptor. The interfaces between subunits of an oligomeric protein molecule were characterized using the *MSDPISA* server (Krissinel and Henrick, 2007), which provides details of all interfaces including the list of interacting pairs of residues, the hydrogen bonding distances between them, the salt bridges, and the area of interaction. All the figures in the thesis have been generated using *PyMOL* (DeLano, 2002).

2.5. Results and discussion

2.5.1. Protein purification, crystallization and data collection

One of the major difficulties faced during structure determination of CAL was the availability of pure protein for crystallization experiments. To start with three different cultivars of chickpea were screened for the lectin, one obtained from the local market, another BDN9-3 from Jalna Agricultural University and a third one, Pusa-256 from Pusa Agricultural University out of which Pusa-256 gave maximum yield and diffraction quality crystals. CAL crystals were consistently obtained with PUSA-256 cultivar. CAL was purified to homogeneity showing a single band corresponding to the subunit molecular mass of 25 kDa on 12 % SDS-PAGE (Fig.2.4). These fractions also showed high hemagglutination activity and were pooled together, dialyzed extensively against DIW and concentrated using centrivap concentrator for setting up crystallization experiments.

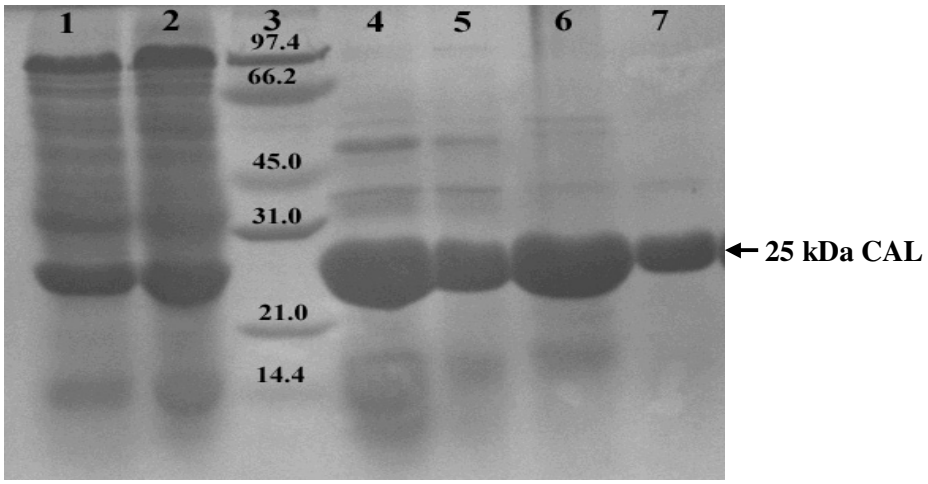

Figure 2.4. **Purification profile of CAL.** Lane 1-Crude extract, Lane 2-ammonium sulfate fraction, Lane-3: Low range molecular weight marker (Biorad) (numbers on the band corresponds to the MW of the protein), Lane 4-6:DEAE unadsorbed fraction, Lane7- purified CAL (25 kDa) eluted from SP-sephadex.

Purified CAL of concentration 20 mg/ml was used in crystallization trials using crystallization solution consisting of 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5 and 30% w/v polyethylene glycol 8000 based on the previous report (Katre *et al.*, 2005). The quality of crystals was further improved by varying polyethylene glycol 8000 concentration from 12-22 %. The crystals appeared with 14, 16, and 18 % PEG 8K and grew to full size in 2-3 days. However, with the lectin preparation from cultivar PUSA-256 crystals belonging to orthorhombic forms were never obtained and only trigonal crystals (with unit cell dimensions a=b=80.2, c=69.1 Å and $\beta=120^{o}$) could be grown. When these trigonal crystals were put in the X-ray beam for data collection the best quality crystal diffracted only up to 3 Å. Crystals obtained were of small size so

microbatch method of crystallization was tried (crystallization under oil with 1:1 mixture of silicon and paraffin oil) using protein and precipitant solution 3 µl each. Large size crystals of CAL appeared after overnight incubation but they were very fragile and thin and tend to break while mounting in capillary for data collection (Fig. 2.5C).

Different additives were tried to increase the size and quality of the crystal as it was suspected that the protein is very prone to aggregation. L-arginine an amino acid has been reported to prevent aggregation of proteins (Shiraki *et al.*, 2002; Lange and Rudolph, 2009) so it was included in the crystallization buffer at a concentration of 20 mM. As a result comparatively bigger and fewer crystals in each drop of CAL were obtained which diffracted upto 2.2 Å (Fig.2.5 A & B). Further optimization of L-arginine concentration was tried out to improve the crystal stablility for better diffraction which can withstand even heavy metal soaking as one more heavy atom derivative was required for phase determination of CAL in absence of homologous structure, apart from the iodine derivative reported previously (Katre UV, 2008).

Since the crystals obtained in the conditions of 0.1 M sodium cacodylate, 0.2 M sodium acetate and 20 % PEG 8K were always found to be small in size and inconsistently appearing, seeding was tried to grow bigger crystals. Although, few bigger crystals could be grown belonging to the trigonal form problems were faced in reproducing the same. When seeding was tried in crystallization condition 0.1 M Na-cacodylate, 0.2 M Na-acetate and 18 % PEG 8K two morphologically different crystal forms, trigonal and hexagonal shapes, appeared (Fig.2.5) but they were found to belong to trigonal form. Diffraction data for P3 space group could not be collected at room temperature because crystals used to die fast. Further data collection was tried at low

temperature. Various cryo-protectants were tried out such as PEG's 200, 300, 400, 600, 1K, 2k, 4k, 5k, 6k, and 8K in the range of 20-40 %. In addition to PEG's Glycerol (10-25 %), Ethylene glycol (10-30 %), various combination of Paraffin and Silicon oil (1:1, 1:2, etc), MPD (20-35 %) were also tried out. Amongst them, MPD in the range of 28-30 % turned out to be a better cryo-protectant of choice which offered appropriate cryo-protection with minimum mosaicity for CAL crystals.

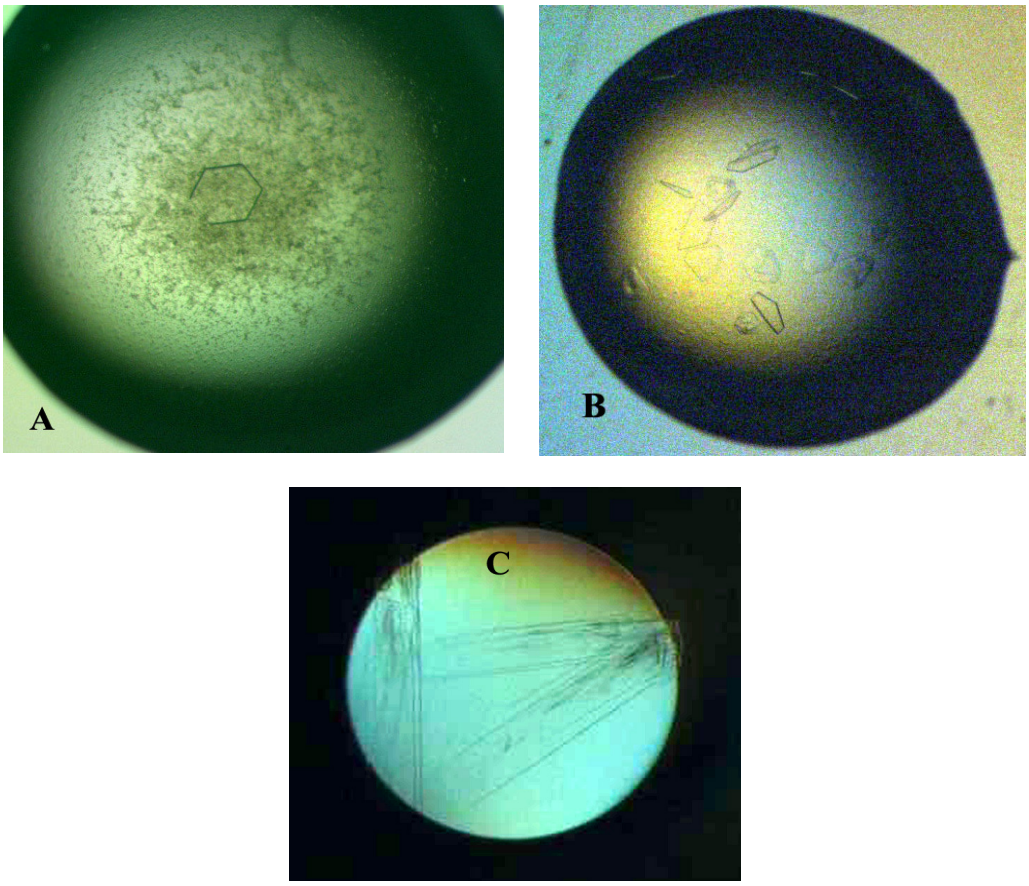

Figure 2.5. Bigger Crystals of CAL obtained by seeding A (hexagonal shape) and B (trigonal shape). These crystals were grown in 0.1 M Na-cacodylate pH 6.5, 0.2 M Na-acetate and 18 % PEG 8K in presence of L-arginine (20mM) and (C) Thin plate shaped crystals obtained with microbatch crystallization method.

For the preparation of heavy atom derivative quick soaking approach gave better results instead of co-crystallization method as with later approach even at very low concentration of heavy atom (3mM or beyond) heavy precipitate was observed immediately after setting up crystallization experiment. With quick soaking approach complete data could be collected in presence of some of the heavy metals namely, Hg, Mo and Pb with different concentrations and for longer soaking times (upto few hrs) as in the case of iodine derivative of CAL poor occupancy was observed.

Highly redundant dataset of CAL crystals soaked in presence of 50-0.5 M KI was collected at home source since iodine was known to bind to this lectin. The data collection statistics of native crystals and the crystals grown/soaked in the presence of KI and other heavy metal atoms are summarized in table 2.1.

Table 2.1. Data collection statistics for the native and derivative crystals of CAL belonging to space group P3 at low temperature.

Heavy atom salt used	Native	Iodine (KI) (50 mM)	Pb-nitrate (10 mM)	Na-Molybdate (10 mM)	PCMB (25mM)
Time of soak (min)	-	20	5	60	90
Temp ^{o}C	-150	- 150	- 150	- 150	- 150
Crystal-to-detector Distance mm	200	200	200	200	200
No. of frames	1-184	1-446	1-357	1-366	1-390
Unit cell dimension a = b Å	80.1	79.7	80.4	80.2	80.1
Unit cell c Å	69.1	69.0	69.2	69.1	69.0
Unit cell angle $\alpha=\beta^{o}$	90	90	90	90	90
Unit cell angle γ^{o}	120	120	120	120	120
Unit cell volume $Å^3$	384622	380134	387827	386024	384162
Resolution Å (last shell)	80-2.55 (2.59-2.55)	50-2.50 (2.54-2.50)	50-2.5 (2.54-2.50)	50-2.5 (2.54-2.50)	80-2.70 (2.75-2.70)
Reflections, total	28810	236178	198129	223468	170030
Reflections, unique	15509	16994	17373	19452	13610
Multiplicity	2.9	13.9	9.72	10.5	12.9
Completeness %	95.5 (95.6)	100 (100)	100 (100)	100 (100)	100 (100)
R_{merge}	4.6 (12.8)	9.1 (27.8)	6.7 (18.7)	7.2 (16.2)	9.3 (19.1)
Matthews Coefficient $Å^3$/Da	2.56	2.53	2.59	2.57	2.56
Solvent content %	52	51	52	52	52

2.5.2. Structure determination of CAL

Despite these attempts a second heavy atom derivative of CAL could not be obtained. None of the above mentioned dataset showed the presence of bound heavy metal atom in the Patterson map. Phasing attempts tried using SAD approach with the reported iodine derivative in orthorhombic form and available phasing programs such as *PHENIX* (Adams *et al.*, 2010) and *CRANK* (CCP4, 1994) was also not successful. While we were struggling for a second heavy atom derivative the full sequence information of CAL became available from our collaborator Dr. M. I. Khan. A *BLAST* search in protein data bank using full length sequence of CAL identified a closely related and recently submitted three-dimensional structure of pA2 albumin from *Lathyrus sativus* (PDB-ID:3LP9).

The three-dimensional structure of CAL was determined using molecular replacement calculation using *PHASER* module of the *CCP4* suit (Collaborative Computational Project, Number 4, 1994) which gave positive solutions with Z-score of 22 for dataset in orthorhombic crystal form (diffraction data used from Katre UV, 2007). Coordinates of pa2 albumin, LS24 (PDB-ID: 3LP9) from *Lathyrus sativus* were used for modelling. Refined coordinates of orthorhombic crystal form of CAL was used for finding the structure of trigonal crystal form. The data collection and processing statistics of these two crystal forms are summarized in table 2.2.

Table 2.2. Data collection statistics for the iodine derivative crystals of CAL belonging to space group P3 and P21212 used for structure determination.

Heavy atom salt used	Iodine	Iodine (Reproduced from Katre UV, Ph.D thesis, 2007)
Space group	P3	P21212
Temp ^{o}C	22	22
Crystal-to-detector Distance mm	150	150
No. of frames	1-194	1-300
Unit cell dimension a Å	81.8	71.2
Unit cell dimension b Å	81.8	73.3
Unit cell dimension c Å	69.5	87.1
Unit cell angle $\alpha = \beta$ ^{O}	90	90
Unit cell angle γ ^{O}	120	90
Unit cell volume Å^3	404425	455621
Resolution range (last shell) Å	50-2.2 (2.24-2.20)	20.00-2.20 (2.28-2.20)
Reflections, total	48875	598857
Reflections, unique	26032	23739
Completeness %	98.1(97.7)	99.9 (99.6)
Rmerge	7.4 (32.7)	10 (29)
Matthews Coefficient Å^3/Da	2.69	2.28
Solvent content %	54	46

2.5.3. Refinement of CAL structure

Structure of CAL was refined in both the orthorhombic and trigonal crystal forms using *REFMAC 5* of *CCP4* suit. To carry out the refinement the reflection data set was divided into two parts; a working set and a test set for cross validation tests. 5 % of the reflections were set aside for Rfree calculation in case of orthorhombic (1144) and trigonal (1236) crystal forms. The models were subjected to several cycles of iterative model building and fitting using the sequence of CAL in *COOT* (Emsley and Cowtan, 2004). Each model building cycle was followed by several cycles of restrained refinement in *REFMAC 5* (CCP4, 1994). The non-crystallographic symmetry (NCS) present in the structures was used as a restraint in the initial cycles of refinement to improve the map quality. In the final stages of refinement the subunits were refined individually leading to significant drop in R_{free} value. As the refinement progressed, difference Fourier maps (Fo-Fc) showed the unaccounted blobs of electron density > 10σ which could be allocated to the naturally bound calcium, chloride and sulfate molecules to the lectins in orthorhombic crystal form. Apart from the bound metal atoms, in case of the iodine derivatives of CAL in orthorhombic as well as trigonal crystal form an iodine atom could also be placed in the unmodelled blob corresponding to the 20σ contour level and further refinement was carried out. The solvent molecules were added according to the Fo-Fc maps first at 5 σ level followed by a few more molecules at 3 σ. A total of 123 and 139 solvent molecules could be placed in the structure of CAL in orthorhombic and trigonal forms respectively.

In both the orthorhombic and trigonal crystal forms CAL extra density in Fo-Fc map were observed near Cys 163, revealing slow oxidation of the residue.

For the final refined model the R-factors in the orthorhombic form were: R_{free} 22.7 % and R_{work} 18.2 % for data in the resolution range 20-2.2 Å. In case of trigonal form the final R-factors for the refined model were 22.2 % and 17.0 % using data in resolution range of 35-2.2 Å. The final refined model for orthorhombic and trigonal crystal forms had 3753 protein atoms. The refinement statistics are summarized in table 2.3. After final cycles of refinement the stereochemistry and the geometry of the models were checked using *PROCHECK*. The Ramachandran plot showed most of the residues in the favored or allowed regions of the plot in case of orthorhombic crystal form (table 2.3). However, in case of trigonal crystal form 0.3 % of residues fall in the disallowed region of the Ramachandran plot. The torsional angles in the residue range 202-208 could not be regularized despite several refinement cycles owing to the poor electron density in these regions for the orthorhombic and trigonal crystal forms. These residues are part of the sulphate binding loop, which might me the reason for the improper geometry.

Table 2.3. Data collection and refinement statistics for CAL in orthorhombic and trigonal crystal forms.

Parameters	Orthorhombic form	Trigonal form
Space group	P21212	P3
Unit cell parameters (Å)		
a	71.2	81.8
b	73.3	81.8
c	87.1	69.5
α, β, γ ^{0,}	90	90, 90, 120
Z	2	2
Resolution range (Å)	20.00-2.20 (2.28-2.20)	50-2.20 (2.24-2.20)
Mathews coefficient	2.28	2.69
Solvent content (%)	46.04	54.35
No. of unique reflections	23739	26032
Crystallographic R_{factor}	0.18	0.17
No of reflections used for refinement	22432	24713
R_{free}	0.22	0.22
No. of reflections for R_{free}	1144	1236
No. of protein atoms	3589	3588
No. of water molecules and inorganic ligands	123, 17	139, 18
Ramachandran Plot Residues in the allowed regions (%)	85.6	83.8
Residues in additionally allowed regions (%)	10.8	12.6
Residues in generously allowed regions (%)	3.6	3.3
Residues in disallowed regions	0.0	0.3

2.5.4. Structural features of CAL

CAL exists as a stable homo dimer (Mol. Wt.-50kDa) in solution under native conditions. Biological assembly of CAL was also found to be a dimer. The polypeptide chain in each monomer of CAL is organized in four β sheets I-IV, which are arranged almost radially around a central pseudo four- fold axis giving the appearance of a 4-bladed propeller (Fig. 2.6).The 4-bladed propeller fold has been characterized for the first time for hemopexin and therefore also designated as hemopexin like fold (Paoli *et al*., 1999). Sheets I-IV are arranged in consecutive order around the central axis. The pseudo four- fold axis of symmetry passing through the centre of the monomer forms a funnel shaped channel of approximate dimension 20 X 8 Å (Fig. 2.6). This channel provides ligand binding sites for a calcium ion, chloride ion and a water molecule exactly in a queued fashion one followed by the other. An iodine atom is bound near the bottom of the channel in case of the iodine derivatives of orthorhombic and trigonal crystal forms. A velcro closure brings together the C- and N-terminus as part of the blade-I which is supposed to play an important role in stabilization of the circular structure (Fülöp and Jones, 1999), also observed to be a conserved feature in several proteins containing the β-propeller motif.

Each blade in the monomer comprises of four stranded twisted β-sheet. The innermost β-strands, towards the centre of the channel are most regular whereas the outer ones are slightly irregular. The adjacent blades are connected by a linker region consisting of three short α-helices. A region consisting of α-helices extends from the blade IV toward the N terminus of the subunit, thereby maintaining cementing interactions between blade I and blade IV similar to those at the other three interfaces.

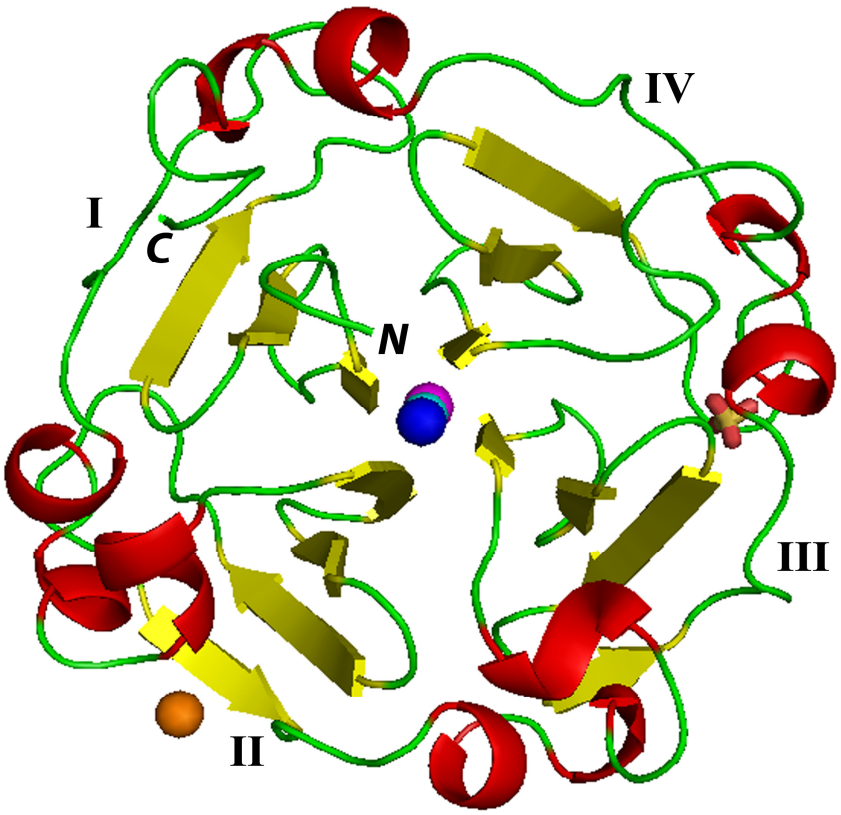

Figure 2.6. The four bladed β-propeller fold in CAL monomer (orthorhombic form) viewed across the four fold axis. The four blades are numbered as I, II, III and IV. The bound calcium, chloride and iodide ions are shown in blue, cyan and magenta spheres in a queue in the channel. Orange sphere represents the chloride ion, near blade II which is at the dimerization interface. Bound sulphate molecule is shown in stick between blade III and IV. The cartoon was prepared using PYMOL (Delano, 2002).

The β-strands of the adjacent blades together form a β-sandwich enclosing a hydrophobic cavity predominated by phenyl alanines (10, 68, 76, 78, 124, 135, 177 and 188) and tyrosines (5, 31 and 133). These hydrophobic residues enclose the central

channel towards the outermost side while the innermost part of the channel is maintained hydrophilic by the presence of various charged residues namely, Asn 7, Asp 65, Asp 69, Asp 121, Asp174, Ser 67, Arg 11 and Arg 125.

2.5.5. Analysis of the metal ion binding site of CAL

The identification and correct placement of metal atoms has been done based on the electron density, B-factors and chemical environment surrounding them in both the orthorhombic and trigonal crystal forms. The four ions/water molecule are arranged along the pseudo four-fold axis shown as $Ca^{2+,}$ Cl^{-}, H2O and iodide, respectively in the orthorhombic form (Fig.2.7 A). In case of the trigonal form electron density for Cl^{-} ion was missing and an extra water molecule near the iodine atom could be located (Fig.2.7 B). The occupancy of iodine atom was also poor in trigonal form compared to the orthorhombic form. Interestingly, the CAL crystals grown/soaked with iodine salt diffracted better. This suggests that iodine imparts stabilization to the CAL structure. This is possible because iodine atom is liganded to four main chain peptide nitrogens provided by Arg 11, Asp 69, Arg 125 and Ala 178. In addition to these metal atoms a clear electron density above 12σ was detected in the Fo-Fc map assigned to a sulphate molecule in the loop regions connecting the β-blade III and IV. This sulphate might have bound during the ammonium sulfate fractionation step of the purification.

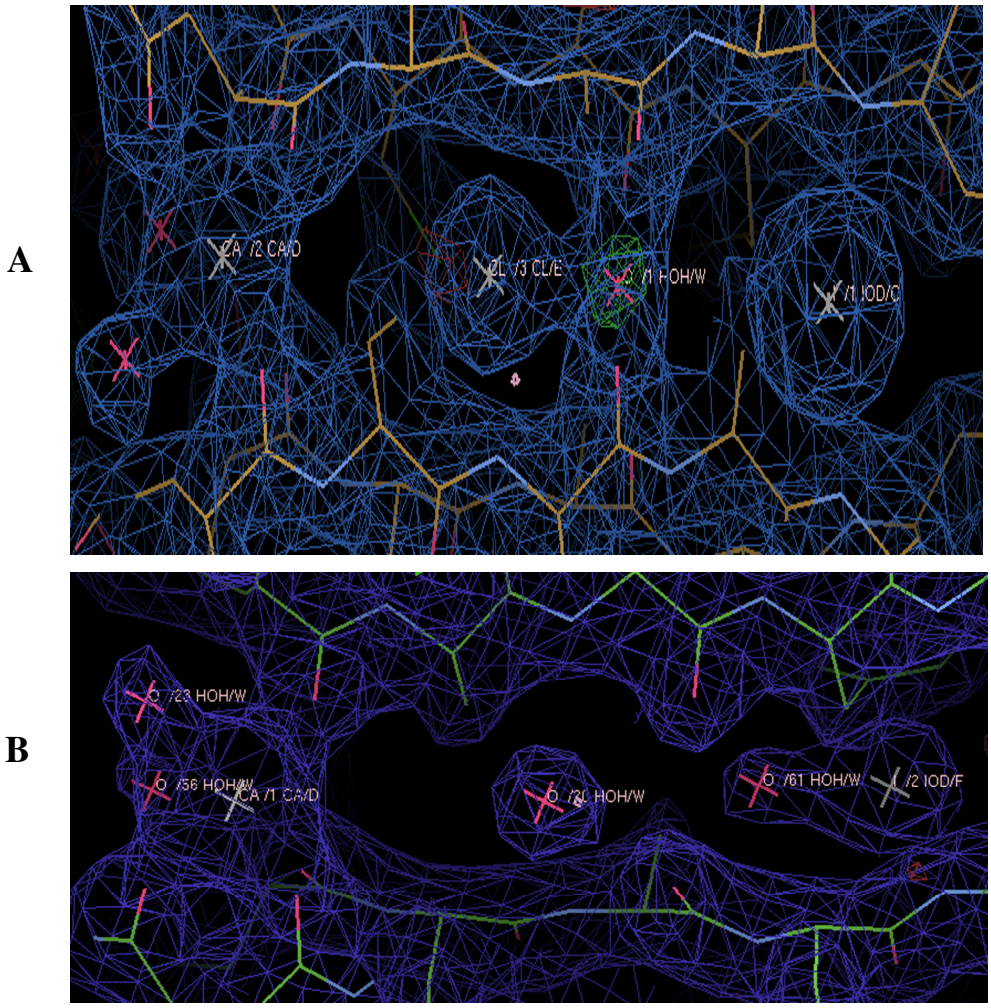

Figure 2.7. Metal atoms bound to CAL in the channel shown fitted in 2Fo-Fc map. (A) orthorhombic form and (B) trigonal form.

Such kind of arrangement of calcium binding site in the interior of the tunnel has also been observed in other proteins containing four bladed β-propeller motif such as, hemopexin (Paoli *et al.*, 1999), human gelatinase A (MMP2) (Gholke *et al.*, 1996), porcine synovial collagenase (Li *et al.*, 1995) and pa2 albumin from *lathyrus sativus* and cowpea (Gaur *et al.*, 2010 and 2011). Most of the known propeller proteins use their central tunnel or the entrance to that tunnel to coordinate a ligand or to carry out some

catalytic function (Baker *et al.,* 1997). However unlike legume lectins, there is no evidence for involvement of metal-ligand in sugar binding in CAL, since hemagglutination activity of the lectin remains unaffected by treatment with chelating agents, such as EDTA

The calcium binding site was compared with those reported in other proteins. Several groups have studied the calcium binding sites in proteins (Pidcock and Geoffrey, 2001; Deng *et al.*, 2006). According to a survey carried out using Cambridge structural database for calcium containing coordination complexes a total 162 Ca^{2+} binding sites were identified out of which 38 % of the cases have Ca^{2+} coordination number of 6, in 30 % of the cases coordination is 7, and in 22 % it is 9 (Pidcock and Geoffrey, 2001). Three general types of calcium binding sites have been identified based on the structures in PDB. They are labeled as: type I where the Ca^{2+} coordinated with atoms from a continuous short sequence of amino acids (PNA, SBA, β-trypsin, alpha-lactalbumin etc), type II is where one ligand comes from a part of the amino acid sequence far removed from the main binding sequence (Lipase, hydrolase, cellulose, Galactose binding protein etc.) and type III where all the ligands are supplied by amino acids remote from one another in the sequence. Ca^{2+} binding sites in case of CAL seems to be of type III where ligands to the Ca^{2+} ion were suppled by Asn 7, 174 and Asp 65, 121. The geometry of Ca^{2+} binding sites in CAL was of pentagonal bypiramidal type, with Ca^{2+} coordination number 7.0 in both the crystal forms. Ligands to the calcium ions and their corresponding bond lengths are listed in table 2.4 for orthorhombic and trigonal crystal forms.

Table 2.4. Atoms in the CAL coordinating with the calcium ion in orthorhombic and trigonal crystal forms

CAL residues ligating to Ca^{2+}	Coordination bond length (Å)
Orthorhombic form	
OD1-Asn 7	2.52
O-Asn 7	2.41
O- Asp 65	2.39
O-Asp 121	2.35
O- Asp 174	2.33
O-H_2O, 83	2.48
O-H_2O, 87	2.43
Trigonal form	
OD1-Asn 7	2.39
O-Asn 7	2.24
O- Asp 65	2.41
O-Asp 121	2.41
O- Asp 174	2.44
O-H_2O, 21	2.38
O-H_2O, 50	2.20

2.5.6. Subunit association and interface analysis

The crystallographic asymmetric unit of CAL is a homodimer. The oligomerization has been analyzed in the solution state as well. Although CAL migrates

as a single band corresponding to subunit molecular weight of ~ 25 kDa on SDS-PAGE (Fig. 2.4), gel filtration chromatography with sephacryl S-200 clearly indicated two peaks corresponding to the monomer and dimer (data not shown). The major peak was of dimer. Oligomeric state of CAL was also monitored at acidic pH 4.0 where again the major population was of dimer. The analysis of the interface was carried out by submitting the structure to *MSDPISA* server. Based on the orthorhombic and trigonal crystals forms *PISA* suggested the existence of a highly stable dimer in solution associated with it a high free energy value of interface 105.0 (orthorhombic) and 91.9 Kcal/mol (trigonal).

The interface of the dimer in orthorhombic and trigonal crystal forms were analyzed separately. Program *CONTACT* was used to calculate the interatomic distance at the interface of protein in both the crystal forms. A comprehensive analysis of interface was carried out using *PISA* and *PROFACE*. The H-bonds and salt bridges stabilizing the quaternary structure of CAL in orthorhombic and trigonal crystal forms are summarized in table 2.5. Interestingly, the CAL in crystal shows a tetrameric association like those found in Con A. Thus, there might be a possibility for the existence of a weak tetramer with poor interface interaction and stability. The oligomerization observed in the case of CAL makes it clearly distinct from the tetrameric legume lectins possessing specificity for mono/oligosaccaharides.

Table.2.5. Interactions at the dimerization interface of CAL in orthorhombic and trigonal form.

Residue in Chain A	Residue in Chain B	Interatomic distance (Å) Orthorhombic form	Interatomic distance (Å) Trigonal form
56:LYS / NZ	100:GLY / O	2.69	3.15
56:LYS / CE	100:GLY / O	3.31	3.12
56:LYS / C	101:PRO / CD	3.48	-
56:LYS / O	101:PRO / CD	3.32	3.35
61:THR / CB	101:PRO / O	3.27	3.43
61:THR / CG2	101:PRO / O	3.48	3.49
61:THR / OG1	101:PRO / O	2.58	2.82
62:TYR / OH	103:LYS / N	2.79	2.67
62:TYR / OH	106:ASP / CG	3.44	3.34
62:TYR / CE1	106:ASP / OD2	3.28	-
62:TYR / CZ	106:ASP / OD2	3.28	-
62:TYR / OH	106:ASP / OD2	2.49	-
62:TYR / CE1	106:ASP / OD1	-	2.99
62:TYR / CZ	106:ASP / OD1	-	3.09
62:TYR / OH	106:ASP / OD1	-	2.39
62:TYR / OH	102:LYS / CA	-	3.36
62:TYR / OH	102:LYS / C	-	3.47
79:TYR / OH	103:LYS / CE	3.48	-
79:TYR / OH	103:LYS / NZ	3.30	-
97:ILE / O	56:LYS / NZ	-	3.36
100:GLY / O	56:LYS / NZ	2.62	-
101:PRO / CD	56:LYS / O	3.40	-
101:PRO / CD	57:THR / C	3.49	-
101:PRO / O	61:THR / CB	3.33	3.25
101:PRO / O	61:THR / OG1	2.67	2.64
101:PRO / O	61:THR / CG2	3.50	3.38
103:LYS / CB	62:TYR / OH	3.41	3.47
103:LYS / N	62:TYR / OH	2.79	2.78
103:LYS / CE	79:TYR / OH	3.32	3.35
103:LYS / NZ	79:TYR / OH	3.18	2.83
106:ASP / OD2	62:TYR / CE1	3.18	3.27
106:ASP/OD2	62:TYR/CZ	3.26	3.34
106:ASP / CG	62:TYR / OH	3.40	-
106:ASP/OD2	62:TYR/OH	2.50	2.57
80:GLU / OE2	103:LYS / NZ	2.72	2.55
103:LYS / NZ	80:GLU / OE2	2.77	2.78
80:GLU / CD	103:LYS / NZ	-	3.46
80:GLU / OE2	103:LYS / CE	-	3.37

The interface between two protein subunits provides a method to understand the principles of molecular recognition in complexes between independent polypeptide chains, or oligomeric assembly of subunits. The program *PROFACE* (Saha *et al.*, 2006) has been used to dissect the protein-protein interface which gives various physicochemical parameters to characterize the interface. It has been shown that many large interfaces are not contiguous, but built of spatially demarcated surface patches. Such segregation into patches is also indicative of the location and distribution of water molecules held in the interface. Additionally, it divides the interface residues into core and rim regions based on their difference of solvent accessibilities and chemical properties which are quite distinct in each region. Such kind of analysis helps to understand the degree of conservation of interface residues in a family of homologous proteins and this represents an important signature of protein interaction sites. The physicochemical parameters derived from the analysis can distinguish the true oligomeric state (dimer, in particular) from the lattice contacts observed in protein crystals. The various parameters analyzed at the dimeric interface of CAL using *PROFACE* are summarized in table 2.6. The analysis revealed that the interaction in orthorhombic crystal form with total interacting surface area of 1330.68 **Å^2** is slightly less than for the trigonal form (1339.61 **Å^2**) (Table 2.6 A)**.** Probably the crystal packing might be compensating this at the interface in trigonal form. For both the crystal forms residues in the surface area corresponding to the non-solvent accessible rim regions was found to be six times more than the core regions approximately (Table 2.6 B).

Table. 2.6. (A) Analysis of the physiochemical parameters of dimerization interface in orthorhombic and trigonal crystal forms of CAL

Parameters	Orthorhombic form			Trigonal form		
	Chain A	Chain B	Total	Chain A	Chain B	Total
Interface Area (Å^2)	669.21	661.47	1330.68	680.41	659.20	1339.61
Interface Area / Surface Area	0.07	0.07	0.07	0.07	0.07	0.07
Number of atoms	64	66	130	70	67	137
Number of residues	17	18	35	18	19	37
Fraction of non-polar atoms	0.70	0.70	0.70	0.73	0.69	0.71
Non-polar interface area (Å^2)	404.55	405.30	809.85	431.01	398.22	829.23
Fraction of fully buried atoms	0.30	0.29	0.29	0.37	0.33	0.35
Residue Propensity Score	-0.22	-0.25	-0.47	-0.25	-0.30	-0.55
Local Density	34.38	35.30		38.43	36.87	

Table. 2.6. (B) Core and rim residues assignment of the dimerization interface in orthorhombic and trigonal crystal forms of CAL

CAL	Chain	CORE			RIM			Total		
		Atoms	Residues	Area	Atoms	Residues	Area	Atoms	Residues	Area ($Å^2$)
Orthorhombic form	A	19	12	588.13	45	5	81.08	64	17	669.21
	B	19	10	445.58	47	8	215.89	66	18	661.47
Trigonal form	A	26	12	597.98	44	6	82.43	70	18	680.41
	B	22	11	528.77	45	8	130.43	67	19	659.2

2.5.7. Identification of the putative sugar binding site

Since CAL showed hemagglutination inhibition only with complex glycoproteins such as fetuin and desialylated fetuin and fibrinogen, the saccharide binding specificity of the lectin could not be identified which imposed a major hurdle in complexing the lectin with mono/di-saccharides to get a sugar bound structure for characterization of the binding site. Other than site directed mutagenesis, chemical modification is an alternative method which provides reasonably reliable information about the residues involved in the saccharide binding. Chemical modification studies on CAL were carried out to get an idea about the residues involved in sugar recognition. The amino acid modification was carried out according to the standard protocol as mentioned in the methods section and

effect of modification at the sugar binding site was estimated by hemagglutination assay. The results obtained are listed in table 2.7.

Table. 2.7. Summary of the results obtained upon chemical modification studies of CAL. The residues present within 5.0 Å distances of two histidines of CAL in the four blades are listed. The residues shown in bold and italics are those which are affecting the hemagglutination activity of CAL upon modification while those in bold are found near the putative binding site but has no effect on activity upon treatment with modifier.

Chemical modifier	**Buffer used**	**% Residual activity**	**Blade I**	**Blade II**	**Blade III**	**Blade IV**
CAL	PBS pH 7.2	100	NA	NA	NA	NA
NBS	CPB pH5.0	100	-	-	-	**W 217**
TNBS	Tris pH 8	50	-	-	-	***K 181***
CITRACONIC ACID	PBS pH 7.2	0	-	-	-	-
NAI	Tris pH7.2	100	Y 19, **31**	-	-	-
WRK	CPB pH 6	50	***N 15,16*** E 17,	-	-	-
DEPC	Tris pH7.2	0	-	H 91	-	***H 180***
PG	PB pH7.2	100	**R 11,** 14	-	-	-
PMSF	Tris pH8	100	**S 13**	-	-	-

Two probable sugar binding sites could be possible as the protein has two histidines, modification of which results into complete loss of hemagglutinating activity. However, out of the four blades the lysine and carboxylate groups affecting the activity to significant extent and are present near the histidine 181 is found in the loop regions of blade IV and I respectively. The probable sugar binding site might be in the short loop of blade IV containing His 180, Lys181 and carboxylate group of Asn 15 and 16 which are the part of the loop from blade I (Fig. 2.8). In blades II and III all these residues were not present in the vicinity. Thus it seems that possibly there is only a single sugar binding site per monomer of CAL and the binding site is very shallow and towards the solvent exposed region on the surface of the protein. However, it may be premature to conclude about the details of sugar binding. More insight into the binding can be achieved with a sugar bound structure of CAL.

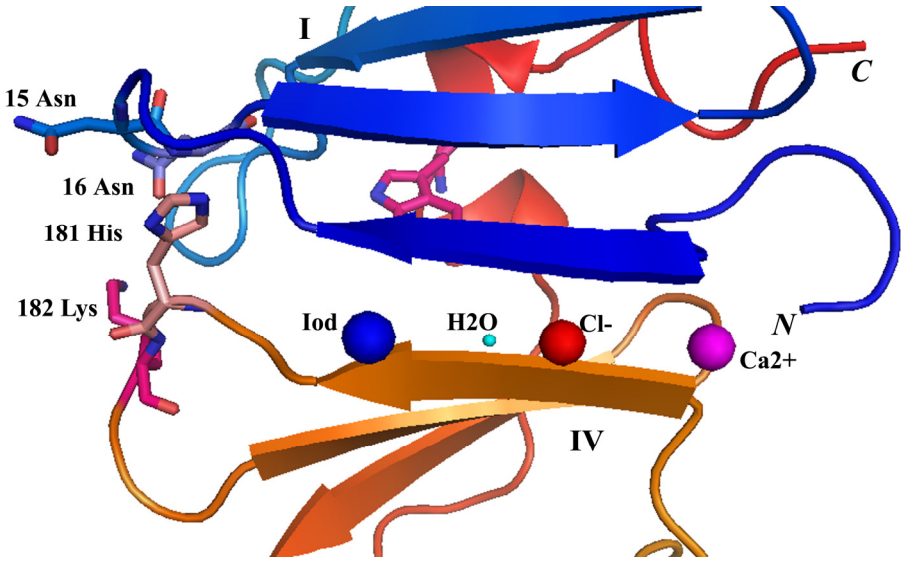

Figure 2.8. Putative residues involved in sugar recognition by CAL inferred using chemical modification studies. The His 180 and Lys 181 are from blade IV while the caaroxylate groups are contributed by blade I. All these residues are labeled accordingly. Also shown are the bound metal ions in the central channel.

2.5.8. Structural diversity, stability and determinants of β-propeller fold in proteins

Amongst all β-folds, the group of the β-propeller proteins is especially interesting; these structures are modular in nature, have extreme diversity in sequence and function, and are found in organisms with very different phylogenetic origins such as bacteria, viruses, fungi and eukaryotes including humans. The modular nature of this fold is based on a simple building block: a four-stranded antiparallel β-sheet which is repeated 4–8 times and radially arranged around a central axis. Because of the twist in the strands of the β-sheet, these modular units appear like the blade of a propeller, which is the reason for the name given to the fold (Paoli, 2001). Despite of striking differences observed in their sequences and length of loops, the general conformation of the propeller's backbone is conserved throughout the known β-propeller structures.

In the last decade a number of entries in the PDB with this folding architecture have been reported, providing unique structures of varied complexity and with fascinating properties. Several structure shave been reported with different modular construction of the propeller fold with diverse function such as 4-bladed fold has been reported for hemopexin (N& C domains, PDB code- 1QHU), collagenase C-domain (PDB code- 1FBL), gelatinase C-domain (PDB code-1RTG), pa2 albumin from *lathyrus sativus* (PDB code-3LPN) and cowpea (CP4) (paoli *et al.*, 1999; Li *et al.*, 1995; Gholke *et al.*, 1996, Gaur *et al.*, 2010 and 2011). The five bladed propellers have been identified in only one protein which is Tachylectin, a lectin from Horse shoe crab (PDB code-1TL2) (Beisel *et al.,* 1999). Six bladed propellers have been reported for bacterial phytase (PDB code-1CVM) and sialidase (PDB code- 2SIL), viral neuraminidase (PDB code-1NN2, ToLB etc. (PDB code- 1CRZ, 1C5K). The seven bladed propellers have been

observed for *methylamine dehydrogenase* (PDB code-2BBK), galctose oxidase (PDB code-1GOG), prolyl oligopeptidase (PDB code-1QFS), nitrous oxide reductase etc (PDB code- 1QNI). The eight bladed propellers have been identified in only two proteins till date they are: Methanol dehydrogenase (PDB code-3AHH) and Nitrite reductase (PDB code-1AOF) (Paoli, 2001).

The general features of the β-propeller fold have been reviewed earlier (Fülöp and Jones, 1999). The twisted β-sheets are radially arranged around a central tunnel and they pack face-to-face. The predominantly hydrophobic interaction between the sheets provides most of the required structural stability, although it is not the only stabilising force. Majority of the known propeller proteins have evolved ways to 'close the circle' between their first and last blades. In 5–8-bladed propellers, all but one of the blades of the propeller domain are built up regularly, with the polypeptide chain progressing outwards from the central axis, forming four-stranded antiparallel β- sheets (Fülöp and Jones, 1999). These blades are joined in succession around the central pseudo-rotation axis. In the last blade, ring closure is achieved by forming the four antiparallel β- strands from both termini of the propeller domain. This arrangement, in which one of the β-sheets tightens the circular belt, is described as molecular 'Velcro' required for the stabilization of the structure.

The structure-based alignments of modular β-sheets in most of the propeller structures have revealed a conserved aspartate residue, which often makes capping interactions to the exposed peptides of a structurally conserved turn region. Apart from the conserved aspartates, such kind of electrostatic interactions can also be mediated by side chains of threonine, asparagine, serine and even tyrosine.

In some cases, the electrostatic interactions do not involve a turn but part of a strand instead as reported for the 16 modular sequences from the 4-bladed haemopexin-like domains where a highly conserved aspartate is engaged in interactions with a conserved β-turn. This feature is unique and has been found important for stability and folding of 4-bladed propellers (Paoli, 2001).

In case of CAL the alignment of the four blades were carried out using CLUSTAL W server. The alignment is shown in fig.2.10. Conserved Asp's were detected in the alignment at two places which were the part of four different blades in loops (Asn 7, Asp's 65, 121 and 174, shown in red in fig. 2.9) and conserved β-turns (Asp 48, 106, 158 and 215, colored blue in fig. 2.9) respectively which has been reported important for the stability of such folds.

```
Blade-III  -VFENGIDAAFRSTKGKEVYLFKGDKYARIDYL----TNRLVQNIKYISDGFPCLRGTIFEA
Blade-IV   -----GMDSAFASHKTNEAYLFKGEYYARINFTPGS-TNAIMGGVKKTLDYWPSLRGIIPLE
Blade-I    -TKTGYINAAFRSSRNNEAYLFINDKYVLLDYAPGTSNDPVLYGPSFVRDGYKSLAKT----
Blade-II   IFGTYGIDCSFDTEYN-EAFIFYENFCGRIDYAPHSDKDKIISGPKKIADMFPFFKGT----
                ::.:* :    *.::*  :    :::     .: :: . .   * :  :
```

Figure 2.9. Multiple sequence alignment of the amino acids present in four blades of CAL (CLUSTAL W (1.81). Conserved residues are shown in red and blue color.

2.5.9. Comparison of the structure of CAL with β-propeller fold containing lectins

There was no apparent similarity in the sequence repeat of the four blades of CAL, despite this they were identical to each other except for the few variations in the regions of loops, helices and turns. The rmsd of the four blades superposed on each other is summarized in table 2.8.

Table 2.8. Superposition of four blades of CAL on each other with rmsd.

	Blade I (RMSD Å)	**Blade II (RMSD Å)**	**Blade III (RMSD Å)**	**Blade IV (RMSD Å)**
Blade I (56AA)	-	1.17	1.33	1.41
Blade II (57AA)		-	1.16	1.22
Blade III (57AA)			-	1.22
Blade IV (55 AA)				-

Root mean square deviations are comparatively more between blade I and IV suggesting asymmetry in the structure in these regions. These blades might be involved in sugar recognition as supported by the chemical modification studies.

Lectin structures based on a β-propeller architecture are rare. β-propeller fold are comparatively new to lectins and have been identified for three different lectins from bacteria (six bladed *Ralstonia solanacearum* lectin, RSL, Kostla´nova *et al.*, 2005), fungi (six-bladed *Aleuria aurantia* lectin , AAL Wimmerova *et al.*, 2003) and eukaryotes (five bladed Tachylectin-2, Beisel *et al.*, 1999). There is no evidence of any other lectin belonging to this fold so far. The structure of Tachylectin-2 consists of five blades with highly similar repeats and five GlcNAc-binding sites located between the blades. The overall structures and carbohydrate binding sites of AAL and tachylectin-2 are distinct. In case of the fugal and bacterial lectins AAL and RSL the ligand-binding sites are very similar. The interactions at the binding site are characterized by numerous hydrogen bonds to the side chains of polar amino acids as well as by strong hydrophobic

interactions between aromatic residues (Imberty *et al.*, 2005). The similarity at the level of sequence, structure and binding site between AAL and RSL could be due to gene exchange allowed by close contact of the two organisms in the soil.

2.6. Conclusions

The β-propellers have extreme diversity in function. Till date only three lectins have been reported in PDB having β-propeller architecture, from bacteria (six bladed *Ralstonia solanacearum* lectin, RSL, Kostla´nova *et al.*, 2005), fungi (six-bladed *Aleuria aurantia* lectin, AAL) and eukaryotes (five bladed Tachylectin-2). Structure of CAL will be the first amongst the plant lectins comprising of the four bladed β-propeller fold. The structure resembles more with the pa2 albumins from plant than the known lectins containing β-propeller folds. Unlike legume lectins, comprising of the typical "jelly-roll motif" structure of CAL has four bladed β-propellers. The metal binding site and oligomerization mode observed in CAL was also different from that observed in legume lectins. Thus the differences in the structure, oligomerization pattern and metal ion requirement of CAL might be contributing towards complex sugar specificity of CAL, compared to simple sugar specific lectins known from legumes. Plants being the highly evolved ones might have utilized the β-propeller fold for dual function: in sugar recognition and as albumins. The four to eight-fold pseudo-symmetry offered by β-propeller architecture seems to be ideal for multivalent sugar binding sites in lectins thus in the future more lectins having β-propeller folds are expected.

Chapter 3

Biochemical and biophysical characterization of *Arisaema curvatum* lectin (ACL).

3. 1. Summary

Structure-activity relationship of a lectin from *Arisaema curvatum* (ACL) showing specificity for complex sugars was investigated using steady state as well as time resolved fluorescence and CD spectroscopy under various denaturing conditions. Chemical modification studies reveal a tryptophan at the binding site of the lectin involved in the recognition of glycoproteins. Phase diagram of the fluorescence data analysis of ACL during chemical denaturation with GdnHCl showed presence of an intermediate which exhibited altered tryptophan microenvironment, pronounced secondary structure and unchanged hemagglutination activity at a concentration of 0.25 M GdnHCl. An acid induced molten globule like structure possessing activity and higher thermostability was also detected. Transition of the lectin to the molten globule state was found to be reversible. The lectin retained hemagglutinating activity even after incubation at 95 °C. Both chemical and thermal unfolding of the lectin were found to consist of multistate processes. Fluorescence quenching of ACL was strong with acrylamide and KI. The single tryptophan residue present in ACL was found to be surrounded by a high density of positively charged amino acid residues as shown by a tenfold higher K_{sv} for KI compared to that for CsCl. The average lifetime of tryptophan fluorescence increased from 1.24 ns in the native state to 1.72 ns in the denatured state.

3. 2. Introduction

Lectins are reported to be abundant in Araceae family constituting 70-80 % of the tuber storage proteins (Van Damme *et al.*, 1995). Araceace lectins are classified under the superfamily of monocot mannose binding lectins according to their sequence homology and sugar specificity (Wright *et al.,* 1996). Other members of Man-binding lectins from Amaryllidaceae, Alliaceae (Van Damme *et al*., 1998), and Orchidaceae (Peumans *et al.*, 1991) family are well studied. A few other studies has revealed the occurrence of some monocot lectins having specificity for complex glycoproteins and not for D-mannose. With respect to their specificity, the Araceae lectins resembled the previously isolated Man-binding lectins from these families except that they interact more strongly with glycoproteins (Van Damme *et al.*, 1995). Though grouped into the mannose binding lectin family, non-mannose binding monocot lectins which are showing specificity for complex sugars are also purified and characterized from the tubers of the different aracaeous plants. Few of the lectins belonging to this category are: *Sauromatum guttatum Schott* (SGA), *Arisaema consanguineum* Schott (ACA), *Arisaema curvatum Kunth* (ACmA), *Arisaema tortuosum schott* (ATL), *Gonatanthus pumilus* (GPA) (Shangary *et al.,* 1995) and *Alocasia indica* (AIL) (Singh *et al.*, 1993).

Arisaema curvatum is a Himalayan shrub belonging to the family Araceae. It is found in Eastern Asia and in Himalayas ranging from Shimla to Sikkim, Bhutan and Nepal. Roots and tubers are the known edible parts of the plant and consumed especially in Nepal (Manandhar *et al.*, 2002). In addition to being edible the tubers have also got insecticidal properties. The roots have been used as a vermifuge in cattle (Chopra *et al.*, 1986). The juice of the tubers is applied to the wounds of cattle in order to kill the

parasites (Manandhar *et al.*, 2002). The dried powder of tubers is applied to cure snake bites. The seeds are mixed with salt and used to treat colic in sheep.

The purification of a 13 kDa monocot lectin from the tubers of *Arisaema curavtum* (ACL) was reported earlier (Shangary *et al.*, 1995). ACL is glycosylated with 1.58 % sugar content. The lectin hemagglutination of the lectin is inhibited by asialofetuin and not by simple monosaccharides indicating the distinct mode of sugar recognition as compared to that of other monocot mannose binding lectins. The lectin also possesses mitogenic potential for human blood lymphocytes. In the present study, biochemical and biophysical characterization of *Arisaema curvatum* lectin (ACL) has been reported in relation to its structure and activity.

3. 3. Structural studies of ACL

To study the structure-function relationship of a protein the basic understanding of the active site environment are essential. X-ray crystallography gives the detailed knowledge of the spatial arrangement of essential amino acid residues important for protein function. However, crystallographic analysis has its limitations such as difficulties encountered in obtaining good quality protein crystals, understanding the dynamics of protein function etc. In addition to site directed mutagenesis, chemical modification is an alternative method which provides reasonably reliable information about the residues involved in the saccharide binding. Chemical modification studies on ACL were carried out to get an idea about the residues involved in sugar recognition.This chapter describes the structure-function relationship of the ACL studied under different denaturing conditions by using spectroscopic methods as described below.

CD Spectroscopy

The circular dichroism (CD) spectrum of a biological macromolecule is sensitive to its secondary structure. Proteins are optically active due to their structural asymmetry and amino acid configuration. The differential absorption of the left and right circularly polarized light by proteins results in a characteristic CD pattern. CD spectroscopy is widely employed to study the various aspects of protein function which are listed below:

1. Studying the folding pattern of a protein in terms of its secondary and tertiary structures

2. Comparing the structures of related proteins obtained from different sources (e.g. species or expression systems)

3. To study the conformational stability of a protein under stress induced by varying the temperature, pH or by introducing chemical denaturants

4. Analyzing the conformational change of protein during protein-protein or protein-ligand interactions.

Fluorescence based studies

The three aromatic amino acid residues namely tryptophan, tyrosine and phenylalanine present in a protein act as intrinsic fluorescent probes. Upon absorption of light energy (photons) the electrons in the fluorophores are excited to a highly unstable state. While returning to the ground state the excess energy of the electron is dissipated by means of longer wavelength fluorescence emission. Tryptophan exhibits stronger fluorescence than the other two aromatic residues. Thus, the dynamics of a protein in

solution can be studied using the tryptophan residues present in it which will be highly sensitive to their microenvironment.

Quenchers are small molecules or charged ions used to perturb the fluorophores in proteins thereby producing changes in the fluorescence intensity, quantum yield and lifetime of the fluorescence emission. Fluorescence quenching studies provide a useful means of probing the accessibility of tryptophan residues and thus give information about their structural environments. Commonly used quenchers are acrylamide, succinimide (neutral), iodide (cationic) and cesium (anionic). Being small and uncharged, acrylamide can penetrate the interior of the protein and can quench the fluorescence of buried tryptophans. The charged quenchers like iodide and cesium ions quench the fluorescence of tryptophans present on surface of the protein and influenced by the charge surrounding the tryptophan (Albani, 2004).

The fine analysis of tryptophan fluorescence data of ACL was carried out by constructing the parameter A and phase diagram. The parameter A is a measure of the characteristic shape and position of the fluorescence spectra (Turoverov *et al*., 1976) which is used for monitoring protein conformational changes (Su *et al.*, 2007 and He *et al.,* 2005). The phase diagram is obtained by plotting the fluorescence intensity at 320 vs 365 nm and is widely employed for the detection of folding intermediates (Bushmarina *et al.,* 2001). In the phase diagram a straight line represents an all-or-none process, while the non-linearity in the variation between I_{320} and I_{365} reflects the structural transitions involving folding intermediate(s). The crossover of the two lines corresponds to the appearance of intermediate at that pH/GdnHCl concentration.

A majority of proteins exhibit smooth, non-structured spectra of tryptophan fluorescence, which often contain more than one component. The position of the maximum in the fluorescence spectrum of tryptophan residues in proteins varies from 307 to 353 nm. According to the model of discrete states, there are five most probable spectral classes of tryptophan residues namely A, S, *I*, *II* and *III* (Burstein *et al.,* 1973, Reshetnyak *et al.*, 2001). The spectral property exhibited by each of these spectral forms (Ladokhin, 2006) is briefly described here.

Spectral form A corresponds to the emission of the unperturbed indole chromophore in the extremely nonpolar environment inside the protein molecule and thus showing the emission maximum around 307 nm. Such a spectrum is reported for Azurin, a small globular metal-binding protein containing single tryptophan residue.

Spectral form S corresponds to the emission of the indole chromophore located in the relatively nonpolar environment inside the protein globule and forming a 1 : 1 complex with some polar protein group. The S spectrum has a maximum at 316–317 nm and shoulders at 305–307 nm and 320–330 nm. It is of interest to know that a pure S spectrum has never been reported, but is always accompanied in proteins by a contribution from the class I spectrum which corresponds to a 2 : 1 complex. For example, types S and I contribute almost equally to emission of L-asparaginase.

Spectral form I corresponds to the emission of the indole chromophore in the somewhat polar but perhaps rigid environment inside the protein globule, forming a 2 : 1 complex with two neighboring polar protein groups. The fluorescence spectrum of class I type has a maximum at about 330 nm and width of about 50 nm. The class I spectrum is observed for actin, chymotrypsin and tetrameric melittin under high-salt conditions.

Spectral form II corresponds to the emission of the indole chromophore at the protein surface. It is assumed to be in contact with the bound water and other polar groups. The fluorescence spectrum of class II is also structureless, but red-shifted as compared to class I. The position of the maximum is at about 340 nm and the width is about 55 nm. Many proteins contain tryptophan residues of this class, e.g. human serum albumin and myosin.

Spectral form III corresponds to the emission of the indole chromophore at the protein surface in contact with free water molecules. The spectrum of class III shows maximum at about 350 nm and a width of about 60 nm, coinciding with the spectrum of free tryptophan. The tryptophan residues of spectral class III seldom occur in the native proteins, but are typical for unfolded states.

3.4. Materials and methods

3.4.1. Materials

Purified ACL was supplied by our collaborators from Gurunankdev university, Amritsar, Punjab. Potassium dihydrogen monophosphate, dipotassium hydrogen phosphate, sodium chloride, citric acid, sulfuric acid were procured from Merck, India. Guanidium hydrochloride, ANS, phenylglyoxal, diethyl pyrocarbonate, N-bromosuccinimide, Woodward's reagent K, 5, 5'-dithiobis (2-nitrobenzoic acid) and N-acetyl imidazole were purchased from Sigma, USA.

3.4.2. Hemagglutination assay

Rabbit RBCs were washed 5 times with PBS (phosphate buffer saline, 10mM potassium phosphate buffer pH 7.2 containing 150mM NaCl). A 3% (v/v) suspension of

the erythrocytes in the same buffer was prepared. The hemagglutination assay was performed according to the method of Gurjar *et al.,* (1998). To test the effect of pH on the activity of the lectin, ACL (50 µg) was incubated for 5 hrs in appropriate buffers of pH 1-10 and hemagglutination assay was carried out at standard conditions. To examine the thermal stability, ACL was incubated at various temperatures from 25-95 ^{o}C for 10 mins. The samples were then cooled down to room temperature immediately in ice water and the residual activity was determined.

3.4.3. Chemical modification studies

The tryptophan content of the protein was estimated by titrating 1 ml (300 µg) of ACL at pH 5.0 against freshly prepared N-bromosuccinimide (3 mM) by the method of Spande and Witkop, (1967). Aliquotes of 3-5 µl were added to the protein solution till the decrease in OD_{280} remained constant. By measuring the reduction in OD at 280 nm, the number of modified tryptophan residues was determined, assuming the molar absorption coefficient as 5,500 M–1 cm^{-1}. To determine the total number of tryptophans modified in the unfolded state, the protein sample was incubated in the presence of 6M GdnHCl for 16 hrs and the titration with NBS was carried out as described above.

Modification of histidine residues with diethyl pyrocarbonate (DEPC) was carried out according to the method of Ovadi *et al.*, (1967). The reagent was prepared in absolute ethanol just prior to use. The lectin (60 µg) in 50 mM phosphate buffer, pH 7.2 was treated with 1-10 mM DEPC for 30 min and the residual hemagglutination activity was determined for each aliquot.

Modification of the carboxylate groups with Woodward's reagent K (WRK) was carried out as described elsewhere (Sinha and Brewer, 1985). ACL (200 µg) was incubated in 50 mM citrate-phosphate buffer pH 6.0, with different concentrations (5-20 mM) of WRK. Aliquots were removed after every 15 mins in each case and the residual activity was determined.

Arginine residues of ACL were modified according to the method of Takahashi, (1968). The lectin (60 µg) incubated in 50 mM phosphate buffer pH 8.0 was treated with phenylglyoxal (prepared in methanol) in the range of 0.5-5 mM, for 30 min at 25 °C and the residual hemagglutination activity was determined for each aliquot. The methanol concentration in the reaction mixture did not exceed 2% (v/v) and had no effect on the activity and stability of the lectin during the incubation period.

Tyrosine residues of ACL were modified with N-acetyl imidazole. The reaction was carried out as described by Riordan *et al.*, (1965). To 60 µg of ACL incubated in phosphate buffer pH 7.2, 2-10 mM *N*-acetyl imidazole separately and hemagglutination assay was performed with the modified protein.

Modification of Serine with Phenylmethylsulfonyl fluoride (PMSF) was attempted by the method of Gold and Fahrney, (1964). The lectin (200 µg) in 50 mM Tris-HCl buffer, pH 8.0 was incubated with 5 and 10 mM PMSF, at RT, for 60 min. Aliquots were removed at 15 min intervals, and residual activity was determined by hemagglutination assay for each aliquot.

Modification of lysine With Trinitrobenzenesulphonic acid (TNBS): The reaction was carried out as described earlier (Habeeb, 1966). The reagent was dissolved in 0.1 M

Tris buffer, pH 8.0 to make a stock solution of 0.1 M. 50 µg of protein in 0.1 M tris buffer, pH 8.0 was allowed to react for 1 hr with TNBS up to final concentration of 2, 4, 5 and 10 mM and the residual activity was determined.

The native lectin without any reagent was used as a control for hemagglutination activity in each case of chemical modification.

Effect of chelating agents on activity of ACL

ACL (60 µg) in 50 mM phosphate buffer pH 7.2 was incubated with EDTA in the concentration range of 2-10mM for 10 mins and hemagglutination assay was performed. The residual activity was determined in each condition.

3.4.4. Circular dichroism spectroscopy

CD spectra of ACL samples were recorded at room temperature using a Jasco J-815-150S (Jasco, Tokyo, Japan) spectropolarimeter connected to a Peltier Type CD/FL Cell circulating water bath. For far UV analysis the spectra was recorded in the range of wavelengths 200–250 nm at a scan speed of 100 nm min^{-1} with a response time of 1 s and slit width 1 nm. The sensitivity was 20 mdeg. A rectangular quartz cell of 1 mm path length was used. The lectin concentration used was 0.1 mg ml^{-1} for all the samples. Each spectrum was the average of 5 scans. Secondary structure elements were calculated by using CDpro software (Sreerama, 1999). All the observed values were converted to mean residue ellipticity (MRE) in deg cm^2 $dmol^{-1}$ (eq. 3.1) defined as

$$MRE = M\,\theta_\lambda / 10dcr \qquad (3.1)$$

Where M is the molecular weight of the protein, θ_λ is CD in millidegree, d is the path length in cm, c is concentration of the protein in mg/ml and r is the average number of amino acid residues in the protein.

To study the effect of pH on the secondary structure of ACL, the following buffers 20 mM were used: Glycine-HCl (pH 1.0–3.0), citrate-phosphate (4.0–5.0), phosphate (pH 6.0–7.0), Tris-HCl (pH 8.0-9.0), and glycine/NaOH (pH 10.0–12.0). Protein samples were incubated for 12 hrs in the buffer of appropriate pH before recording the spectra. Corresponding buffer baseline was corrected.

To monitor the effect of chemical denaturation on the ACL, the protein was incubated in GdnHCl in the concentration range of 0-6 M at pH 7.0 and 0-4M at pH 3.0 for 10 hrs. Buffer scans recorded under the same conditions were subtracted from the lectin spectra for further analysis.

To measure the tertiary structure of the ACL at pH 1.0, 3.0, 7.0 and 10.0 the near-UV CD spectra were recorded in the range of 250-300nm using a cuvette of path length 1 cm. The lectin concentration used was 1 mg/ml.

The effect of temperature on the secondary structure of ACL was studied at pH 7.0 and 3.0 by increasing the temperature of the protein samples at the rate of 2 ^{o}C/min from 25-95 ^{o}C. The interval of 10 ^{o}C was used and the ellipticity was recorded between 200-250 nm. The Tm of the protein sample was determined by plotting the ellipticity at 222 nm against temperature. All data were corrected for the respective buffer base lines.

3.4.5. Steady State Fluorescence spectroscopy

Intrinsic fluorescence of ACL was measured using a Perkin Elmer Life Sciences LS50 fluorescence spectrophotometer. The samples were kept in a quartz cuvette, at room temperature. The background emission due to the signal produced by either buffer solution, or buffer containing the appropriate quantity of denaturants was subtracted. The lectin solution (~0.050 mg ml^{-1}) was excited at 295 nm and the emission spectra were recorded between 300-400 nm. Each spectrum was an average of 5 accumulations. Both the excitation and emission spectra were obtained setting the slit-width at 7 nm, and speed 100 nm min^{-1}. To see the effect of pH on fluorescence maximum, ACL was incubated at different pH from 1 to 12 in the respective buffers for 10 hrs. Before analysis the respective buffer spectra were subtracted from the protein spectra.

3.4.6. Steady-state fluorescence quenching of ACL

Fluorescence quenching measurements for native ACL (at pH7.0) and protein incubated at pH 3.0 were carried out with different quenchers like acrylamide (5 M), iodide (5 M) and cesium ions (5 M) by titrating against small amounts of quencher stock solutions. The protein sample (0.05 mg/ml) was prepared at pH 7.0 and pH 3.0 prior to starting the quenching experiments. Sodium thiosulfate (0.2 M) was added to the iodide stock solution to prevent the formation of tri-iodide (I3). Relative fluorescence intensities were measured at the wavelength corresponding to the emission maximum (353 nm) of the protein and volume correction for fluorescence intensities was applied before analyzing the quenching data (Lakowicz and Weber, 1973).

3.4.7. Phase Diagram and Parameter A

The ratio of fluorescence intensity at 320 to intensity at 365 nm of the steady state fluorescence of ACL was plotted as a function of GdnHCl (M) /pH to obtain the parameter A. The phase diagram analysis was also carried out for ACL to detect the folding intermediates. To construct the phase diagram the fluorescence intensity at 320 nm versus 365 nm at different GdnHCl concentrations as well as at varying pH was analyzed as described previously (Su *et al*., 2007 and Bushmarina *et al.,* 2001). The fluorescence intensities were normalized with respect to the corresponding intensity of the spectra for native protein.

3.4.8. Fluorescence Lifetime measurement

Lifetime fluorescence measurements were carried out at TIFR, Mumbai, by employing CW-passively mode-locked frequency-doubled Nd: YAG laser (Vanguard, Spectra Physics, USA)- driven rhodamine 6G dye laser which generates pulses of width ~1 ps. Fluorescence decay curves were obtained by using a time-correlated single-photon counting set up, coupled to a microchannel plate photomultiplier (model 2809u; Hamamatsu Corp.). The instrument response function (IRF) was obtained at 295nm using a dilute Ludox solution. The samples were excited at 295nm and the fluorescence emission was recorded at 353 nm. The slit width of emission monochromator was 7nm. The resultant decay curve was analyzed by a multi-exponential iterative fitting program provided with the instrument. Concentration of ACL samples was 0.3 mg/ml for the experiments.

3.4.9. Binding of Hydrophobic dye ANS to ACL

The binding of the hydrophobic dye 8-Anilino-1-naphthalene sulfonic acid (ANS) was studied by recording the emission spectra in the range 430–550 nm with excitation at 375 nm using steady-state spectrofluorimeter. The protein was incubated at different pH in the range 1-12 in the respective buffers. 5 µl of 25 mM ANS was mixed with 2 ml of protein (0.04 mg/ml). Buffer spectrum with ANS in each of the condition was subtracted from the spectrum of the protein.

3.4.10. Decomposition of fluorescence spectra

The decomposition of tryptophan fluorescence spectra of ACL was carried out using PFAST program (http://pfast.phys.uri.edu/pfast/) based on the SIMS and PHREQ algorithm developed by Burstein *et al.,* (2001).

3.4.11. Crystallization trials on purified ACL

The purified ACL was dialyzed against deionised water and concentrated to 20 mg ml^{-1} using centrivap concentrator. Crystallization trials were conducted using hanging-drop vapor-diffusion method. The commercially available crystallization kits used to screen for the suitable crystallization condition(s) of ACL are listed below:

1. Crystal Screen I and II, PEG Ion screen I and II, Salt Rx Screen I and II from Hampton Research (USA).

2. JCSG Plus screen I and II from Molecular Dimensions Limited (UK).

3.5. Results and discussion

ACL in highly purified from was obtained from our collaborators. The hemagglutination activity of the protein was assayed. Before proceeding for crystallization and biophysical studies purity of the protein was estimated on 15 % SDS-PAGE where ACL migrated as a single band corresponding to 14 kDa (Fig.3.1). Although 99 % pure, attempts to crystallize ACL were not successful and hence further structural studies were carried out using spectroscopic methods described in section 3.3 to investigate the structure-function aspect of the protein.

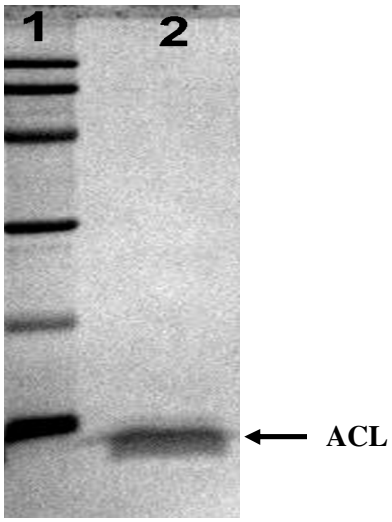

Fig 3.1. Purification profile of ACL (15 % SDS-PAGE). Lane 1: Molecular weight markers (Biorad laboratories). From bottom to top: 14.4, 21.5, 31.0, 45.0, 66.2 and 97.4 kDa respectively and Lane 2: 14 kDa ACL.

The far UV CD spectrum of ACL showed sharp minima around 225 nm and maxima at 210 nm (Fig.3.2). The calculation of secondary structure elements by CDSSTR program from CDPro showed α-helix: 5 %, β-sheet: 45 % turns: 18 % and

random coil: 31 %. Thus ACL is a beta protein containing several turns and unordered elements. The negative minima at 225 nm and maxima around 210 nm is well characterized for type III β-turns in the case of tuftsin a synthetic tetrapeptide (Siddiqui *et al.,* 1996). This could be due to the high content of proline-glycine repeat in ACL as reported by Brahmachari *et al.,* (1982) in case of tripeptides with pro-gly sequences. The type III β turns seems to be a conserved structural feature of araceae lectins also characterized in case of *Arisaema tortuosum* lectin (ATL) and *Sauromatum guttatum* Schott agglutinin (SGA) by Dharkar *et al.,* (2009) from our laboratory.

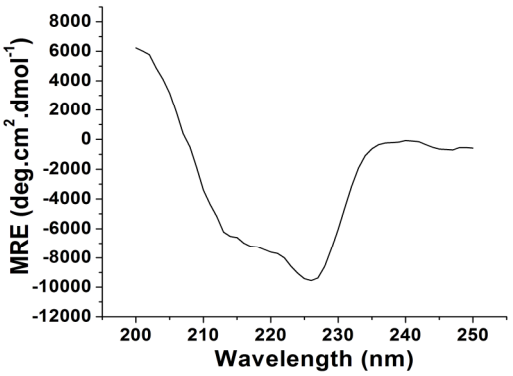

Figure 3.2. Far UV-CD spectrum of ACL at pH 7. The protein concentration used was 0.1mg ml^{-1}.

3.5.1. Chemical modification studies

A single tryptophan residue per monomer of ACL could be detected in native and denatured conditions upon titration with NBS. Modification of the tryptophan residue of ACL under native conditions resulted in the complete loss of hemagglutination activity of the protein (Table.3.1) suggesting the probable involvement of the tryptophan in sugar recognition. 50 % loss in activity was observed upon treatment of the lectin with DEPC and WRK indicating the presence of histidine and aspartate/glutamate near the sugar

binding site of the protein. This observation was in agreement with the 50 % loss of activity upon treatment with KI which usually interacts with the positively charged groups on the surface of the protein. Reducing agent did not affect the activity of ACL presumably due to absence of cysteine residues in the protein, also a common feature of araceae lectins. Involvement of naturally bound metal ion if/any in sugar recognition may be ruled out as the lectin retains hemagglutinating activity upon treatment with chelating agent EDTA at various concentrations tested.

Table 3.1. Effect of chemical modifiers on the hemagglutinating activity of ACL. The lectin was incubated with various chemical modifiers each separately at room temperature for 5-10 mins and the residual activity was determined by hemagglutination assay.

Reagents	Concentration of the modifier (mM)	Buffer	Residual Activity (%)
Native ACL	-	PBS, pH 7.2	100
TNBS	5	Tris-HCl, pH8.0	Do
PMSF	5	Tris-HCl, pH8.0	Do
EDTA	5	PB pH 7.2	Do
N-acetyl Imidazole	10	PB pH 7.2	Do
BME	2	PB pH 7.2	Do
DTNB	1	PB pH 7.8	Do
Phenyl glyoxal	3	PB pH 7.8	Do
DEPC	5	PB pH 7.2	50
WRK	10	CPB pH 6.0	50
NBS	200	CPB pH 5.0	0
GdnHCl	100	PB pH 7.2	100
GdnHCl	250	PB pH 7.2	Do
GdnHCl	500	PB pH 7.2	25
KI	5-10	PB pH 7.2	50

3.5.2. Effect of pH on the structure and activity of ACL

The ACL samples were checked for hemagglutinating activity in appropriate buffers in the pH range 1–10, and the residual activities observed were plotted as a function of the sample pH (Fig. 3.3A). The optimum pH observed for the maximum activity of the lectin was 6-7 owing to the structural stability of the protein at smaller interval between pH 6 and 7. A sharp decrease in agglutination activity was observed both towards acidic and alkaline pH. Minimum activity (18 %) was observed at pH 1-2, while at pH 3.0 ACL retained 40 % activity. At pH 10 the residual activity observed was 37 %. Altogether, ACL remains active in a wide range of pH for 5 hrs. Very few plant lectins are known to be stable in such a wide range of pH, some of them are: lectin from Pinto beans (pH range: 3-12) (Wong *et al.*, 2006); *Alocasia cucullata* lectin (pH range: 2-12) (Kaur *et al.*, 2005) and *Arisaema tortuosum* lectin (pH range: 2-10) (Dhuna *et al.*, 2005). Another araceae lectin isolated from the tuber of *Typhonium divaricatum* (L.) Decne (TDL) was found to be active from pH 5.6-8.6 retaining 50% and 60% agglutination activity at pH 2 and 12 respectively (Luo *et al.*, 2007).

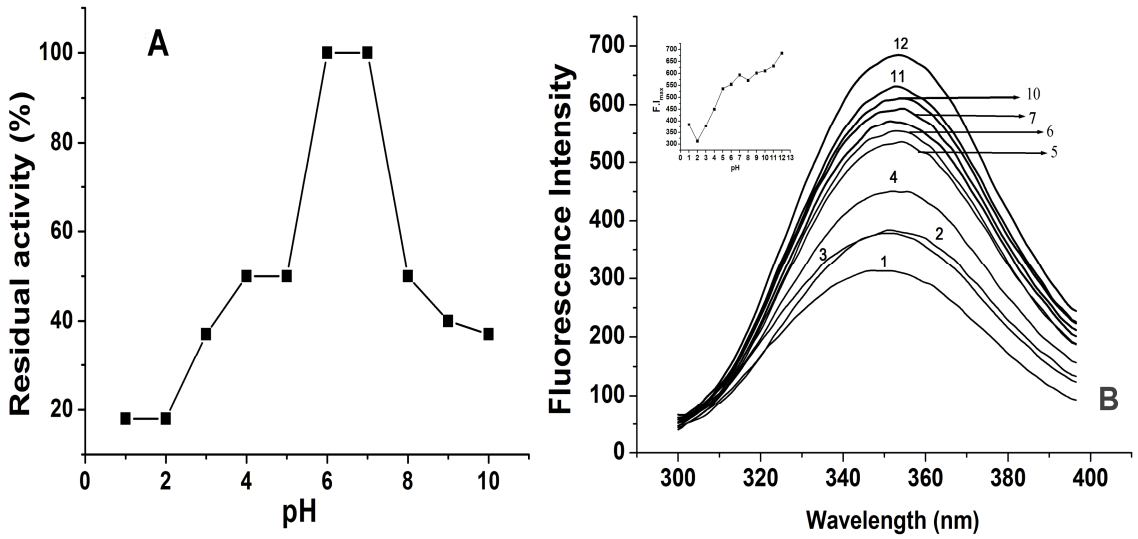

Figure 3.3. Effect of pH on stability of ACL. Hemagglutination activity of ACL as a function of pH (A), Fluorescence intensity of ACL at pH 1-12 (B).The inset shows change in fluorescence intensity vs pH. The numbers on the spectra corresponds to the pH of the protein solution.

The effect of pH on the structure of ACL was further investigated by spectroscopic studies. Native ACL at pH 7.0 showed the maximum at wavelength of 353 nm in the intrinsic fluorescence studies indicating fully solvent exposed tryptophan (Fig.3.3B). The tryptophan in ACL could be classified into spectral class A (313 nm) or III (347 nm) by decomposition analysis of fluorescence spectra with 25.9 and 74.1 % contributions from individual components as obtained by SIMS-2 program. At the extremes of acidic pH 1-3, a three-fold decrease in the fluorescence intensity of the protein was observed associated with a blue shift in the fluorescence maximum upto 3 nm. This shows the increased hydrophobicity of the tryptophan microenvironment of ACL at acidic pH. To confirm this observation hydrophobic dye binding studies were carried out at acidic pH. ANS binding to the ACL is shown in Fig.3.4. The native protein

at pH 7.0 and at pH 4-6 failed to bind ANS. However, ANS binding was maximum at pH 1 where 5-fold increase in the fluorescence intensity with a blue shift in the λ_{max} from 520 to 489 nm was observed indicating the fully exposed hydrophobic patches on the protein. Since the fluorescence intensity was very high and beyond the limit of calculation, the protein concentration was reduced to half for the spectra recorded at pH 1.

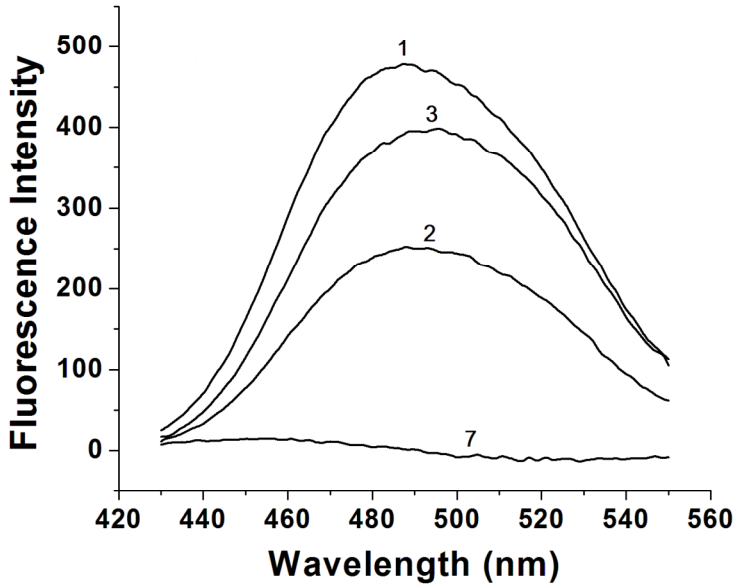

Figure 3.4. Effect of pH on ANS fluorescence emission spectra of ACL. The numbers on the spectra corresponds to the pH of the protein.

The parameter A (Fig. 3.5A) and phase diagram (Fig. 3.5B) for pH-fluorescence profile were plotted for detection of intermediates during pH induced unfolding. Parameter A showed two regions where structural transitions occured in protein, first around pH 4.0-2.0 and second between pH 2.0-1.0. The phase diagram detected the existence of two stable intermediates at pH 3.0 and pH 5.0. The intermediate at pH 3.0 was taken up for further characterization which also possessed exposed hydrophobic

patches. The intermediate at pH 5.0 might be the result of zero net charge on the protein (pI).

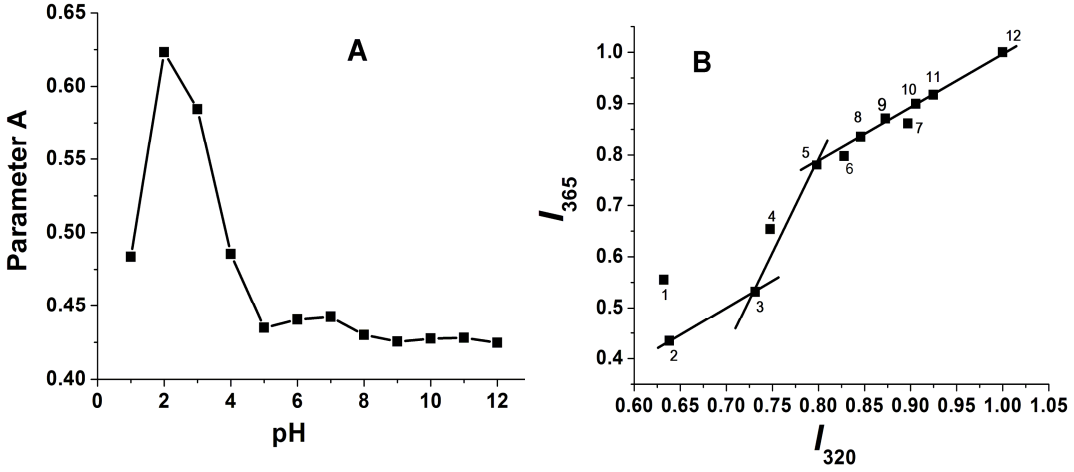

Figure 3.5. Parameter A (A) and Phase diagram (B) of ACL as a function of pH. The parameter A was plotted as the ratio of intrinsic fluorescence intensity at 320 nm (I_{320}) to that of 365 nm (I_{365}) vs pH of the protein. Phase diagram was constructed by monitoring the changes of I_{365} as a function of I_{320}. The lectin concentration used was 0.05 mg ml^{-1}.

Structural changes induced in the protein at extremes of the pH were analyzed by recording the CD spectra of protein at various pH in the range 1-10. The far-UV CD spectra revealed complete loss of secondary structure of ACL at pH 1, while at pH 3.0, 30 % loss in the secondary structure was seen (Fig. 3.6A). The protein was able to retain the major part of the structure at pH 3.0 as compared to the native one. ACL at pH 3.0 shows strong ANS binding due to exposed hydrophobic amino acid side chains on the surface as discussed in the previous section. This partially unfolded state of ACL at pH 3.0 showed slightly loose tertiary structure in the near UV CD spectra (Fig.3.6B) revealing existence of the lectin in molten globule conformation.

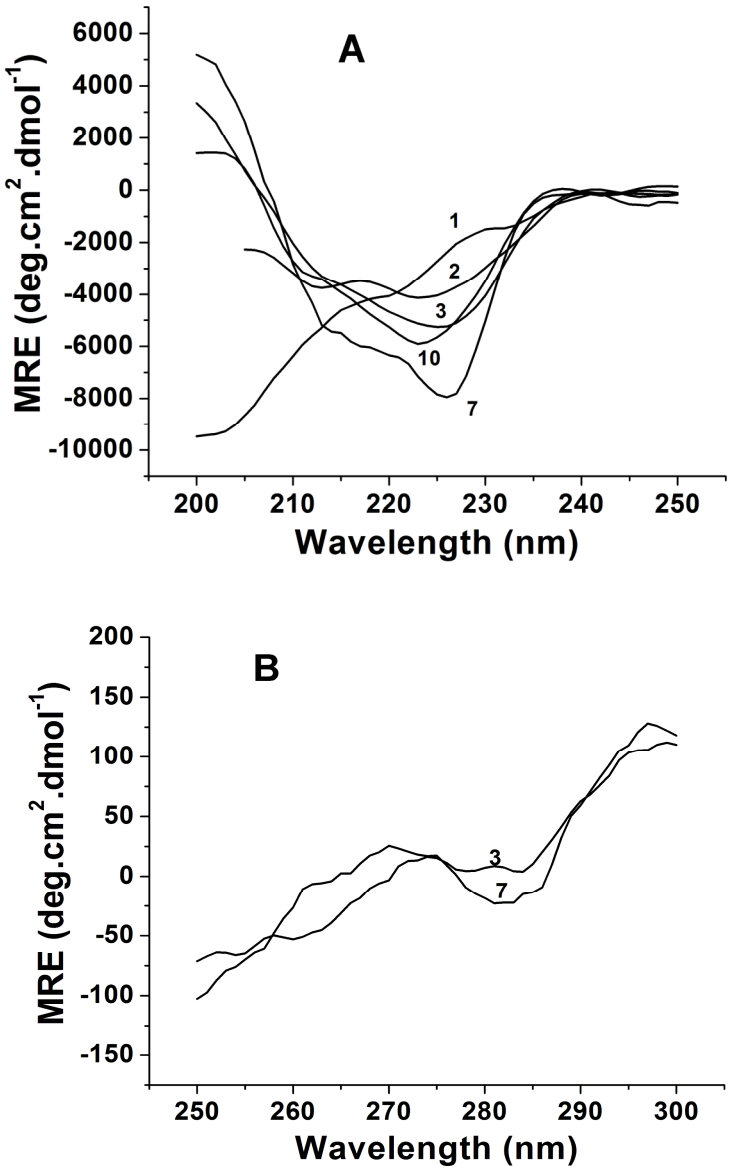

Figure 3.6. pH induced structural changes in ACL as seen in the far (A) and near UV CD (B) spectra of the protein recorded at respective pH. The numbers on the spectra corresponds to the pH of the sample.

3.5.3. Unfolding and refolding studies on ACL using GdnHCl

The shift in the λmax and change in the fluorescence intensity of ACL with increasing concentration of GdnHCl (0-6 M) are shown in fig. 3.7 A and B, respectively. The perturbation in the tryptophan environment of ACL in the presence of 0.25-2.5 M GdnHCl was reflected in the alternate red and blue shift in the λmax of fluorescence emission (Fig. 3.7A). In the presence of 0.25 M GdnHCl, fluorescence intensity at 365 nm increased 1.5 times while at 0.5 M GdnHCl it decreased to the original level (Fig. 3.7B) indicating the alteration in the micro-environment of the tryptophan due to breakage of weak interactions resulting in enhanced and reduced fluorescence emission. Significant perturbation in the fluorescence intensity at 320 nm, up to 2M GdnHCl indicated alteration in the micro-environment of hydrophobic conformer of the tryptophan. At 3M GdnHCl and above the fluorescence intensity remained stable. Gradual red shift in the λmax was observed which reached up to 356 nm with 34% decrease in the fluorescence intensity in presence of 6M GdnHCl indicating exposure of tryptophan to more polar environment (aqueous) due to unfolding of the protein (Fig. 3.7A and B).

Parameter A (ratio of I_{320} to I_{365}) gives the idea about shape and position of the tryptophan spectrum and the phase diagram plotted as the I_{320} vs I_{365} gives the information regarding intermediates existing during unfolding. Parameter A for ACL shows the formation of several intermediates of protein in the presence of 0.25- 2.5 M GdnHCl (Fig. 3.6C). However, after renaturation, refolding of the protein seems to be quite in order. Two linear portions were observed in the phase diagram for unfolding of ACL (Fig. 3.6D). The two lines intersect at 0.25 M GdnHCl indicating the presence of an intermediate. The intermediate was able to retain complete hemagglutination activity as

that of native lectin. No linearity was observed in the phase diagram for the refolding of ACL (Fig. 3.7D).

Far-UV CD spectra for ACL treated with different concentrations of GdnHCl were recorded. In the vicinity of 0.25 M GdnHCl, the decrease in the negative ellipticity of ACL (Fig. 3.8) indicated more pronounced secondary structure confirming the existence of an intermediate. Major transition in the secondary structure of the protein was observed between 2.5 and 3.0 M GdnHCl. Above 3M GdnHCl, the negative ellipticity decreased slowly. Maximum unfolding of the protein took place in the presence of 6M GdnHCl. The existence of an intermediate species between 2.0-2.5 coincides with that observed in the plot of I_{365} vs. concentration of GdnHCl (Fig.3.7B). Thus, both the fluorescence and CD data reveal the multistate unfolding of ACL in the presence of GdnHCl. The anomalous behavior of ACL in presence of 0.25 to 2.5 M GdnHCl i.e. instantaneous changes in the tryptophan microenvironment (alternate increase and decrease in the fluorescence intensity and red and blue shift in the λ max) and secondary structure (increase and decrease in the negative ellipticity) is an interesting feature of ACL.

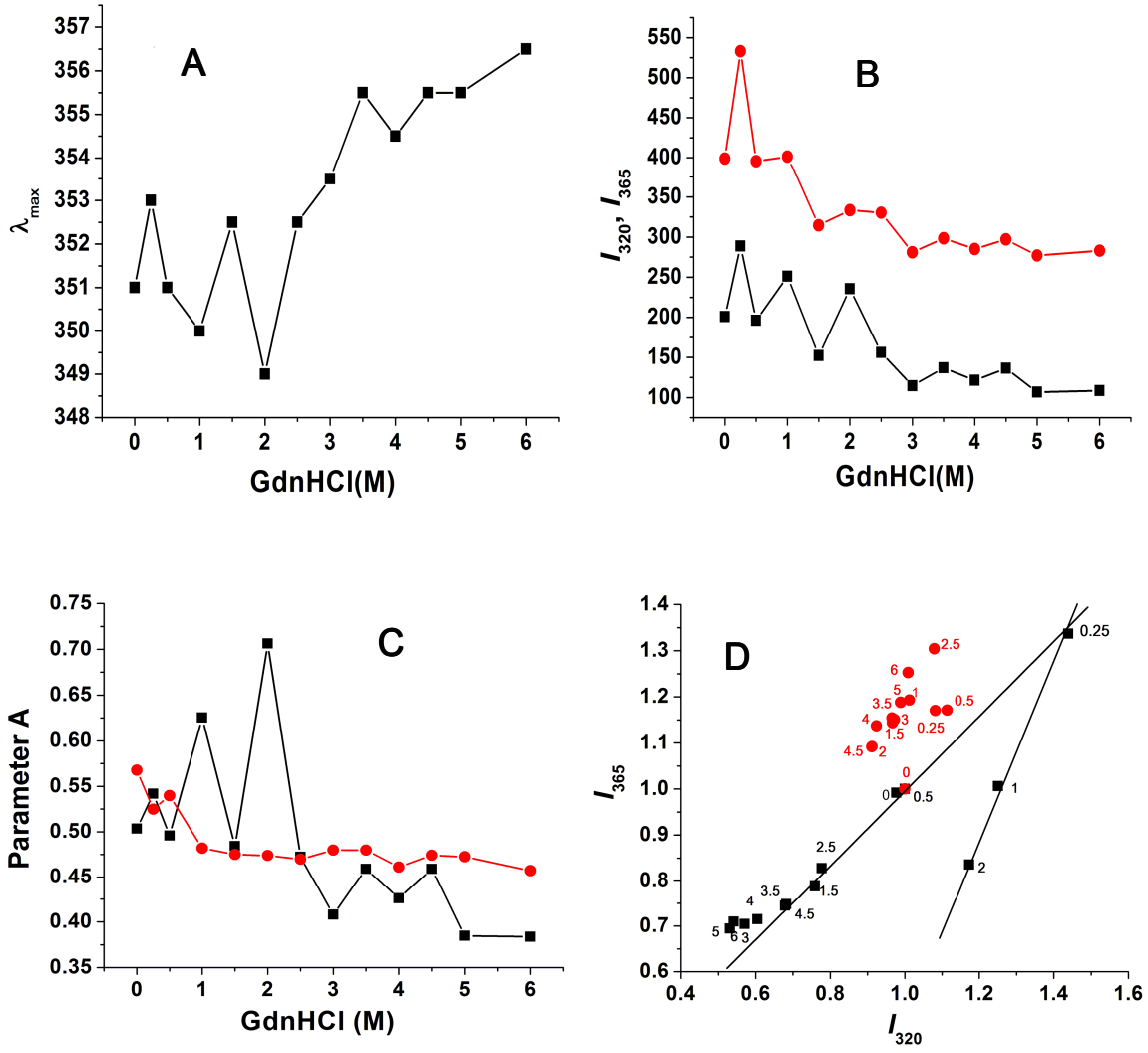

Figure 3.7. GdnHCl induced unfolding of ACL. Shift in λmax of unfolded ACL vs GdnHCl concentration (A) and Change in tryptophan fluorescence at 320 (black square) and 365 nm (red circle) with increasing concentration of GdnHCl (B). Parameter A (C) and Phase diagram (D) of ACL treated with GdnHCl. The parameter A was plotted as the ratio of intrinsic fluorescence intensity at 320 nm (I_{320}) to that of 365 nm (I_{365}) vs Gdn HCl concentration. Phase diagram was constructed by monitoring the changes of I_{365} as a function of I_{320} (black squares and red circles represent unfolding and refolding respectively).

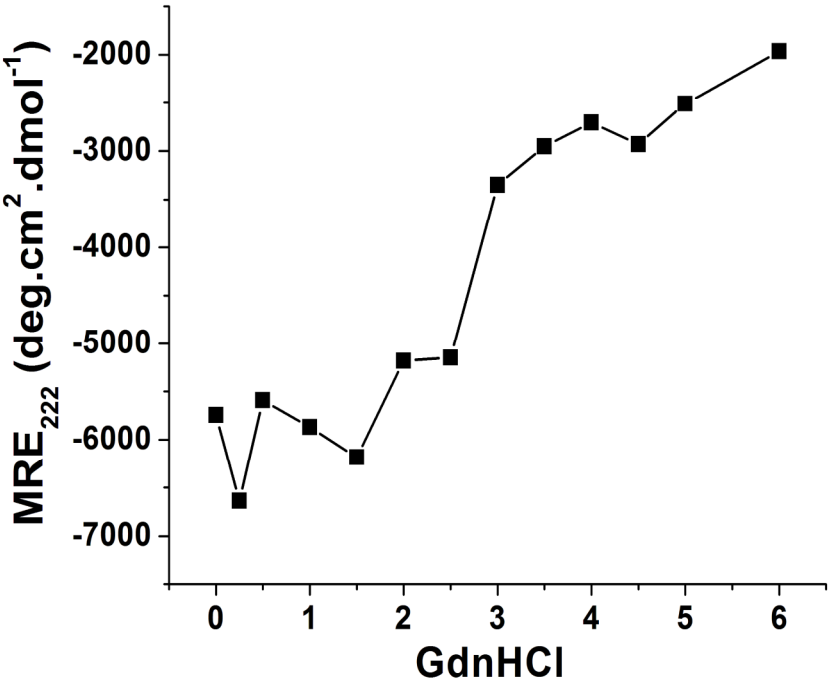

Figure 3.8. Effect of GdnHCl on MRE of ACL at 222 nm. The ellipticity values were taken from the far UV CD spectra of GdnHCl treated protein (0.1 mg ml^{-1}). The samples were incubated with GdnHCl for 5 h at 25 °C

3.5.4. Characterization of the Molten Globule of ACL

The acid induced molten globule form of ACL was found to possess rearranged secondary structure (1.5 % increase in helical content, 5 % more turns and 6 % loss of beta sheet content as listed in table.3.2)**.** The geometry at sugar binding site of the lectin remains intact as the protein retained 40 % of the hemagglutination activity even after incubation at pH 3.0 for 3hrs. The molten globule of ACL at pH 3.0 might be involved in the protein folding pathway as an active intermediate.

Stability of ACL molten globule

Effect of GdnHCl on ACL molten globule is shown in Fig.3.9A and B. At pH 3.0 decrease in the fluorescence intensity and blue shift in the λmax in presence of GdnHCl up to 1.0 M concentration was observed (Fig. 3.9 A). Increase in the fluorescence intensity and red shift in the λmax was observed at and above 1.5 M GdnHCl, the trend opposite to that observed at pH 7.0 (Fig. 3.7 A). The amino acid side chains in the microenvironment of tryptophan are protonated due to which the fluorescence is quenched at pH 3.0 which gets quenched further at lower concentration of GdnHCl. Thus the structural characteristics of the intermediate at pH 3.0 are significantly different from the native or the unfolded one. Although there is hardly any linearity found in the phase diagram, the parameter A shows the presence of a stable intermediate between 0.5-1.0 M GdnHCl concentration (Fig. 3.9B). Upon dilution of the chemical denaturant the protein at pH 3.0 almost refolded back to the original state.

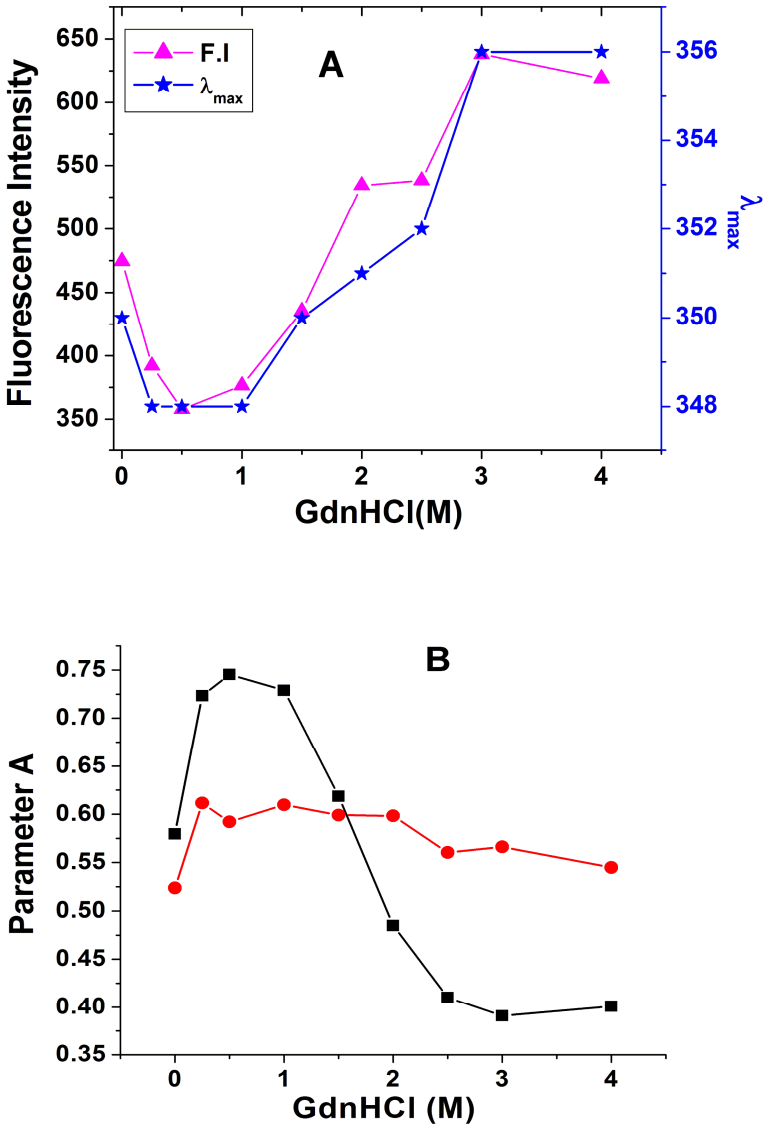

Figure 3.9. Unfolding of ACL molten globule at pH 3.0. (A) Effect of GdnHCl on fluorescence intensity (filled triangle) and λmax (blue stars) of ACL at pH 3.0 and (B) Parameter A calculated for GdnHCl mediated unfolding of ACL molten globule (black square and red circle represents unfolding and refolding respectively.)

3.5.5. Thermal stability of ACL

The hemagglutinating activity of ACL was measured at different temperatures as mentioned in section 3.4.1. ACL showed progressive increase in activity after incubating at different temperatures, exhibiting maximum activity for the sample incubated at 95 ^{o}C

(Fig. 3.10 A) probably due to the exposure of sugar binding site upon thermal unfolding. Structural transitions occurring in ACL at different temperatures were studied at pH 7.0 and 3.0 (Fig 3.10 B and C respectively). The secondary structure of ACL at pH 7.0 remained intact at temperature up to 45 ^{o}C while at pH 3.0 it was stable up to 55 ^{o}C as monitored by MRE 208 (Fig.3.10 D). The loss of structure at 65 ^{o}C was less at pH 3.0 as compared to that at pH 7.0. This evidence supports the existence of rigid and compact molten globule state of ACL at pH 3.0 showing the strong intra molecular interaction between the side chains of the protein.

Few of the lectins retain hemaggltuination activity even after incubation at temperatures above 75 ^{o}C, including *Trichosanthes dioica* lectin (Dharkar *et al*., 2006), lectin from wild edible mushroom *Agaricus arvensis (*Zhao *et al.,* 2011), mannose-binding lectin from *Dendrobium findleyanum* (showing activity even after boiling to 100 ^{o}C) (Sudmoon *et al.*, 2008), lectin from *Kaempferia parviflora* (Konkumnerd *et al.*, 2010) and *Moringa oleifera* lectin (MoL) from drumstick seeds (Katre *et al.*, 2008)

The thermal unfolding observed at pH 7.0 was found to be reversible in nature and ACL refolds back to the native conformation as reflected from the secondary structure contents listed in table.3.2. This can be correlated with the presence of hemagglutination activity even after incubation at 95 ^{o}C. The intact and stable active site structure was resumed upon cooling, possibly due to the small energy barrier required for the thermal unfolding. In case of the molten globule, once heated upto 95 ^{o}C and recooling back to room temperature resulted into the misfolded state of ACL (Fig. 3.10 C, spectra in green curve).

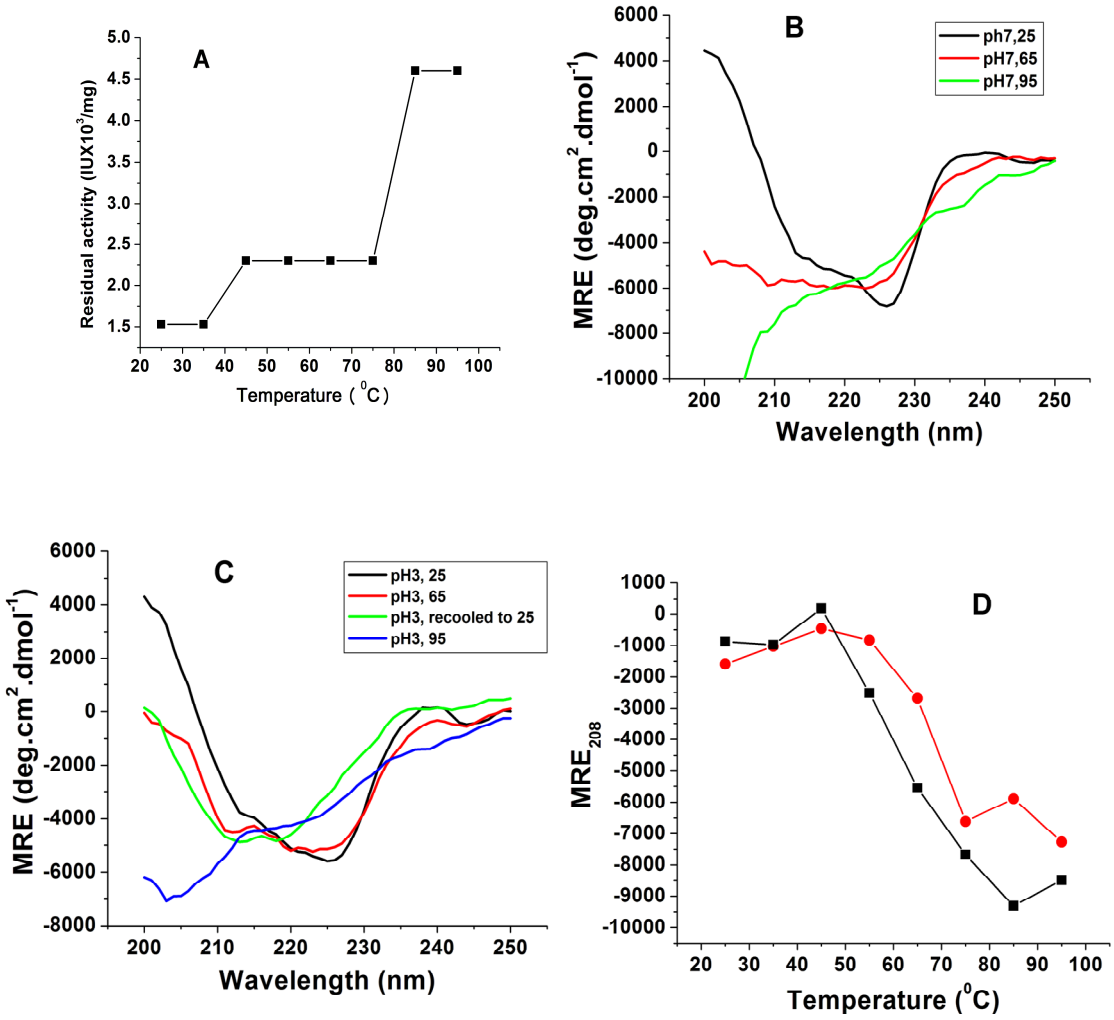

Figure 3.10. Thermal denaturation studies of ACL. Hemagglutination activity as a function of temperature (A). Thermal unfolding of ACL (0.1 mg ml^{-1}) at pH 7.0 (B) and pH 3.0 (C). MRE $_{208}$ as a function of temperature (filled square and filled circle for native and pH 3.0 incubated ACL, respectively) (D)

Table 3.2. Secondary structure analysis of ACL as a function of pH, temperature and chemical denaturants. The far-UV CD spectra of the ACL under these conditions were analyzed by CDSSTR programme available online with CDPRO suit. The hemagglutinating activity of ACL under native condition, at pH 3.0 and ACL heated at 95 ^{o}C and cooled are also shown.

ACL	Helix (%)	Sheet	Turn	Unordered	NRMSD	Residual activity
Native	4.9	45.2	18	31.4	0.07	100
pH 3.0	5.5	39.5	23.3	31.3	0.08	40
4M Gdn-HCl (pH7)	5.5	34.5	33.6	26.4	0.07	NA
95 ^{0}C (ph7.0)	5.3	36.3	25.2	31.0	0.09	NA
Recooled at 25 ^{0}C (pH7.0)	5.2	40.2	23.2	30.5	0.06	100

3.5.6. Fluorescence quenching studies on ACL

Solute quenching studies were carried out to study the microenvironment of tryptophan as the protein was found to lose the hemagglutinating activity on modification of this residue. Also, as a part of the characterization of molten globule, the quenching studies of ACL were carried out at pH 3.0. Analysis of the quenching data was done by the Stern–Volmer equation (3.2) as well as by the modified Stern–Volmer equation (3.3) (Lehrer, 1978).

$$F_0/F_C = 1 + K_{sv}[Q] \qquad (3.2)$$

$$F_0/\Delta F = f_a^{-1} + 1/[(K_a f_a)(Q)] \quad (3.3)$$

where F_0 and F_c are the fluorescence intensities (A.U) corrected for dilution, in the absence and presence of quencher respectively, $\Delta F = F_0 - F_c$ is the change in fluorescence intensity at any point in the quenching titration, (Q) is the resultant concentration of the quencher, K_{sv} is the Stern–Volmer constant of the protein for the given quencher, f_a is the fraction accessible to the quencher and K_a is the corresponding quenching constant for the accessible fraction. Equation (3.3) shows that the slope of a plot of $F_0/\Delta F$ versus $[Q]^{-1}$ (modified Stern–Volmer plot) gives the value of $(K_a f_a)^{-1}$ and its Y-intercept gives the value of f_a^{-1}.

The fluorescence quenching profile, Stern–Volmer and Modified Stern-Volmer plots for quenching with the various quenchers at pH 7.0 and 3.0 are shown in Fig. 3.11, 3.12 and 3.13. For ACL fluorescence quenching efficiency was highest with iodide. Stern–Volmer analysis for iodide quenched data showed downward curvature of the plot, which could be resolved into two linear components suggesting more than one population of tryptophan on the surface of the protein, one getting quenched before the other. The K_{sv} value for acrylamide is 7.5 M^{-1} while that for I^- is 10.94 M^{-1} (K_{sv1}) and 12.92 M^{-1} (K_{sv2}) (Table3.3). The higher value of K_{sv} for Iodide compared to that of acrylamide is unusual. Very low K_{sv} for CsCl indicated high density of positive charge around surface tryptophan. The modified Stern-Volmer analysis showed slight increase in the accessibility for acrylamide while the accessibility for I^- decreased slightly after denaturation of the protein. The higher K_{sv} and lower accessibility for I^- as compared to acrylamide could be due to the capacity of the later to penetrate into the interior of the protein.

The profile obtained from Stern-Volmer analysis with acrylamide quenching showed a positive curvature at pH 3.0 which indicated presence of both dynamic and static quenching components. The static quenching involves complex formation, while dynamic quenching involves collisions with acrylamide during the lifetime of tryptophan in excited state.

The accessibility of the fluorescence to CsCl was 19 % in the native state, 62 % in the denatured state and 46 % at pH 3.0 suggesting that the protein is in different conformational state at pH 3.0 and the negative charge density on the surface is significantly increased. At pH 3.0 acrylamide as well as KI exhibited enchanced quenching associated with a blue shift of 14 nm revealing that after quenching of the fluorescence of surface tryptophan conformer, the protein shows fluorescence from conformer in the hydrophobic interior (Fig. 3.11 B and 3.12 B). Significant loss in the native conformation of ACL and major change in the side chain arrangement of aromatic residues was observed.

Iodide having a large ionic radius and being negatively charged, probably binds to the positively charged amino acid residues of ACL on the surface at pH 3.0 which is reflected in 83 % accessibility to the fluorescence. The Ksv further increased at pH 3.0 (Table.3.3). There could also be a possibility of some nonspecific binding of iodide to the protein around the single tryptophan in the protein leading to affinity quenching of the fluorescence rather than collisional quenching.

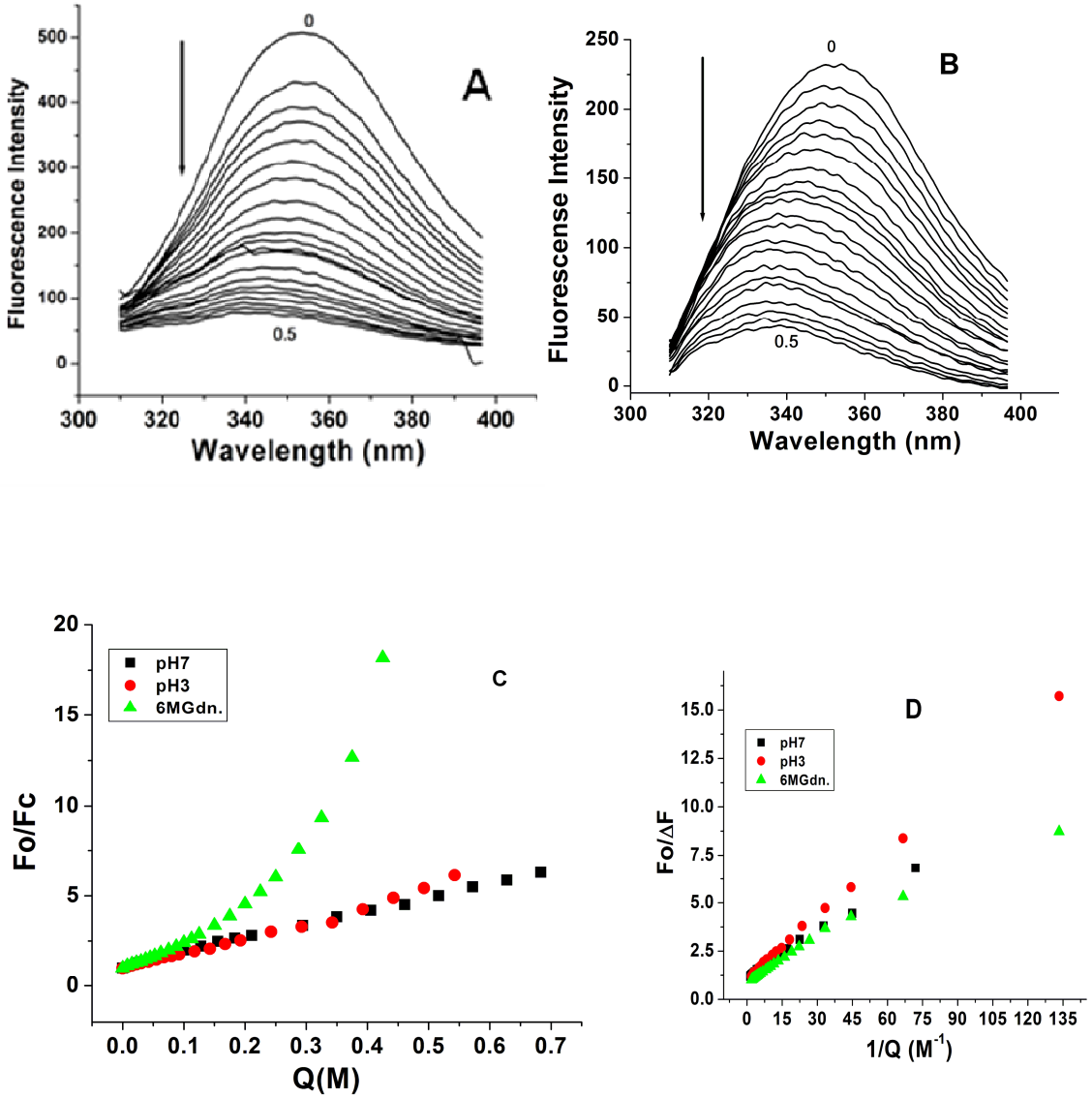

Figure 3.11. Quenching of the intrinsic fluorescence of ACL with acrylamide. pH 7.0 (A) and pH 3.0 (B) & Stern-Volmer plot (C) and modified Stern-Volmer plot (D) for both pH 7.0 and pH 3.0.

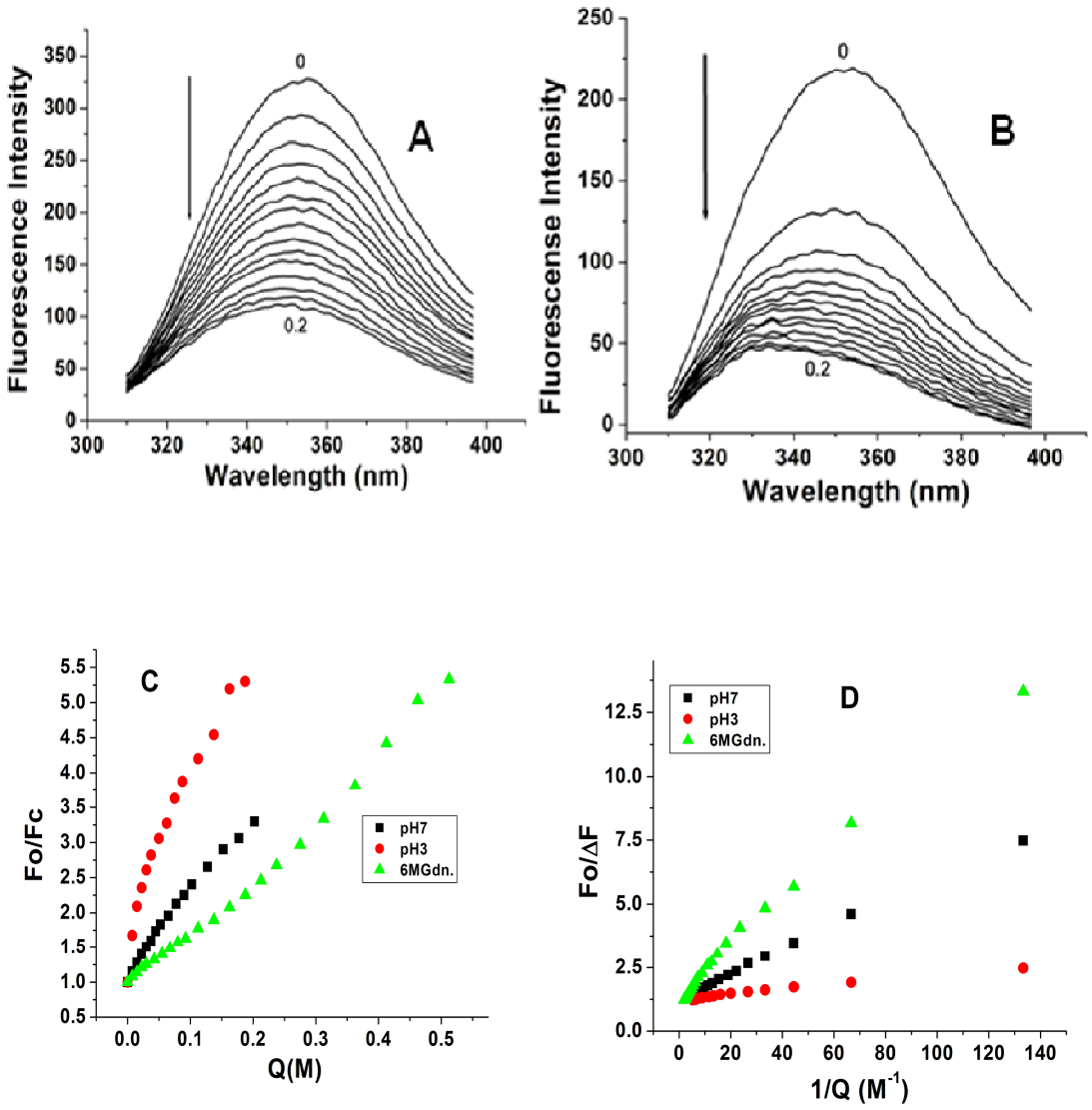

Figure 3.12. Quenching of the intrinsic fluorescence of ACL with Iodide. pH 7.0 (A) and pH 3.0 (B) & Stern-Volmer plot (C) and modified Stern-Volmer plot (D) for both pH 7.0 and pH 3.0.

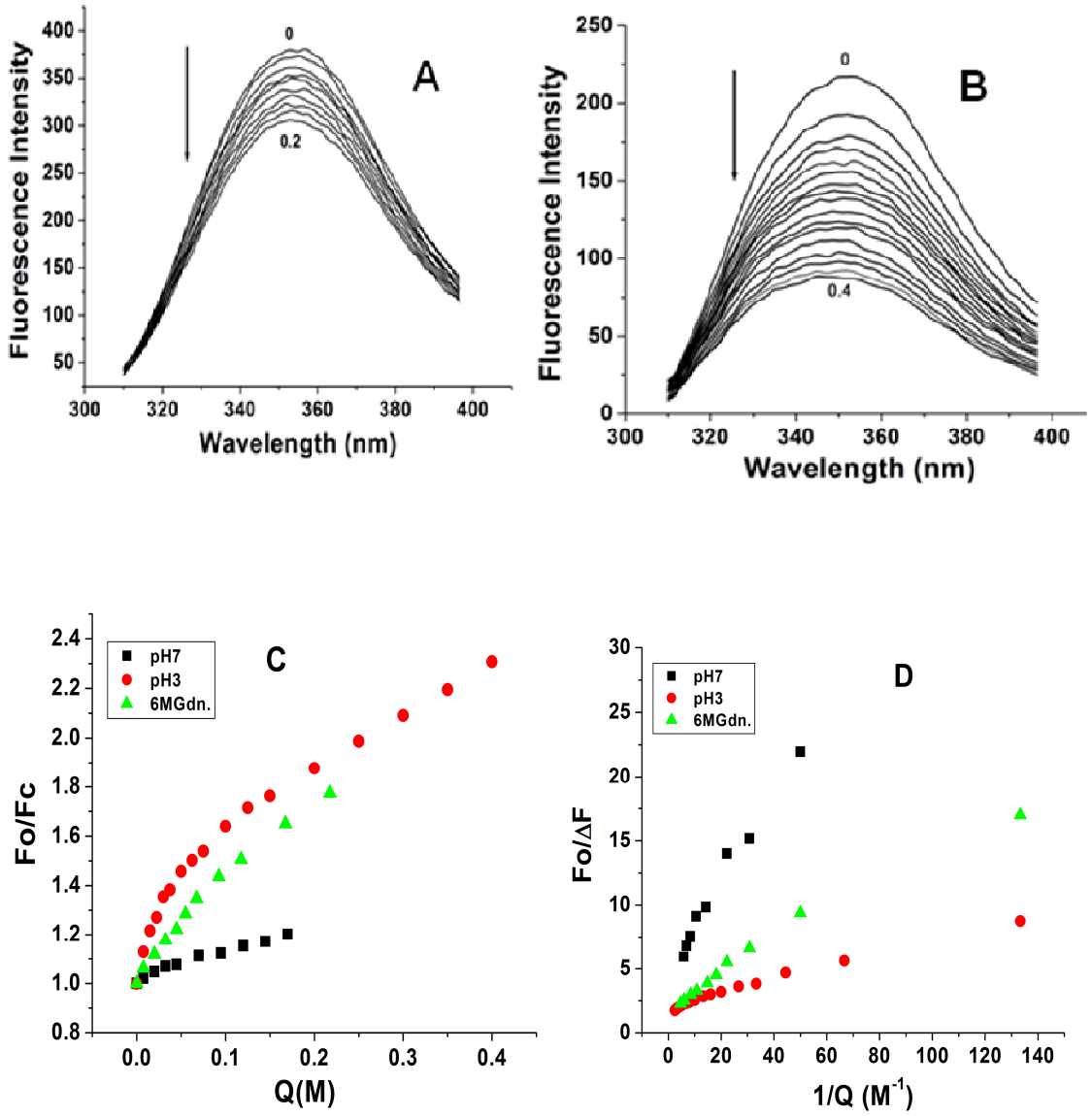

Figure 3.13. Quenching of the intrinsic fluorescence of ACL with CsCl. pH 7.0 (A) and pH 3.0 (B) & Stern-Volmer plot (C) and modified Stern-Volmer plot (D) for both pH 7.0 and pH 3.0.

Table 3.3. Summary of parameters obtained from Stern–Volmer and modified Stern–Volmer analysis of the intrinsic fluorescence quenching of ACL with different quenchers.

Quencher and Samples	$K_{sv1}(M^{-1})$	k_{q1} ($x10^9 M^{-1}S^{-1}$)	$K_{sv2}(M^{-1})$	$k_{q2}(x10^9 M^{-1}S^{-1})$	f_a	$K_a(M^{-1})$
Acrylamide						
Native	7.5	6.04	-	-	0.86	16.57
pH3.0	9.0	-	-	-	0.93	10.7
6M Gdn-HCl	-	-	-	-	0.91	18.16
KI						
Native	10.94	8.82	16	12.90	0.77	-
pH3.0	23.6	-	14.7	-	0.83	-
6M Gdn-HCl	8.27	4.82	-	-	0.73	15.22
CsCl						
Native	1.02	0.82	-	-	0.19	16.3
pH3.0	2.44	-	-	-	0.46	-
6M Gdn-HCl	1.63	0.95	-	-	0.62	10.8

3.5.7. Lifetime fluorescence studies on ACL

The lifetime measurement of the intrinsic fluorescence of ACL from the decay curve (Fig. 3.14) was carried out by fitting it to a multiexponential function ($\chi^2 < 1.095$) .The decay curve of native lectin consists of three components with the lifetimes, τ1 (0.28 ns) and τ2 (1.14ns) and τ3 (4.11 ns) indicating the presence of three conformers of the single tryptophan with average lifetime of 1.24 ns (Fig.3.14A) (Table 3.4). The ACL denatured with 6M Gdn-HCl showed increased average life time of 1.72 with τ1 (0.31 ns)

and τ2 (1.18 ns) and τ3 (3.11 ns) due to change in the environment of tryptophan after unfolding of protein.

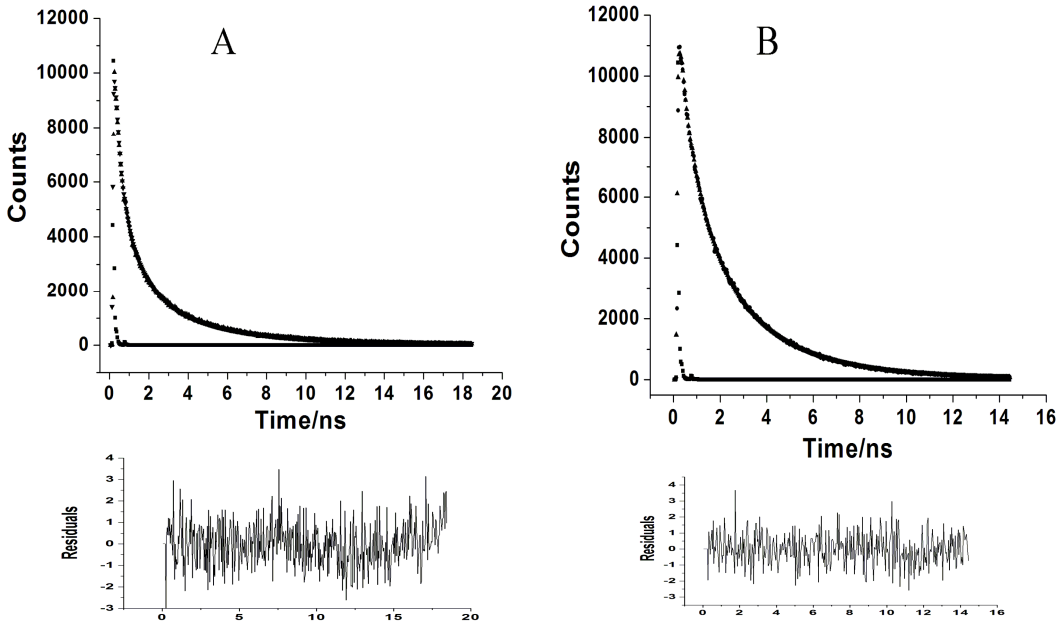

Figure 3.14. Life time measurement of ACL. Time resolved fluorescence decay profile of native ACL (0.3 mg ml^{-1}) (A) and 6M GdnHCl denatured ACL (B). The solid line corresponds to the nonlinear least square fit of the exponential data. The lower panel represents the residual.

From the life time measurements of the quenching of the intrinsic fluorescence of ACL with 0.5 M acrylamide, the decay curve revealed existence of four conformers of tryptophan. There was four times decrease in τ1 (0.05 ns) and τ2 (0.25 ns) and τ3 (0.89 ns) in presence of 0.5 M acrylamide. τ1, τ 2, τ 3 and the fourth conformer of τ4 (1.95 ns) showed average τ of 0.37 ns (Table. 3.4).

Table 3.4 Parameters obtained from lifetime fluorescence measurement of ACL

Samples	**τ 1 (ns)**	**τ 2 (ns)**	**τ 3 (ns)**	**τ 4 (ns)**	**<τ> (ns)**	**χ^2**
Native	0.278	1.142	4.108	-	1.241	1.095
Native +0.5M Acrylamide	0.045	0.250	0.886	1.951	0.372	0.967
Native+0.5M KI	0.068	0.406	1.049	5.182	0.429	0.928
Native+0.5M CsCl	0.136	0.517	1.963	4.512	1.032	1.041
Denatured 6M GdnHCl,	0.308	1.176	3.112	-	1.715	1.004
Denatured+0.5 M Acrylamide,	0.185	0.529	1.293	-	0.472	1.026
Denatured+0.5 M KI	0.041	0.451	1.202	9.745	0.498	0.954
Denatured+0.5 M CsCl	0.220	1.123	2.498	-	1.320	1.046

Similar effect on the decay curve was observed upon quenching with KI where τ4 was detected and the average life time was calculated as 0.73 ns. Although in steady state fluorescence quenching, K_{sv} for iodide was greater than that for acrylamide, time resolved measurement gave higher average lifetime for iodide quenched ACL than that for acrylamide quenched protein (0.37 ns). This could be due to some non-specific binding of iodide with the protein due to the higher density of positive charge around

surface tryptophan. The tryptophan lifetime was found to be longest for the Cs+ quenched protein with average τ of 1.03 ns and 1.32 ns in the native and denatured state, respectively.

3.6. Conclusions

Despite several crystallization attempts, ACL could not be crystallized presumably due to the presence of isolectins differing in their charge. Biophysical studies to investigate the structure-function aspect of ACL reveal that the tryptophan microenvironment gets perturbed in the presence of low concentration of the chemical denaturant. There exists an active intermediate in the vicinity of 0.25 M GdnHCl with altered tryptophan microenvironment and pronounced secondary structure. At pH 3.0, the protein shows compact secondary and slightly disrupted tertiary structure with hydrophobic amino acid side chains exposed on the surface. This altered conformation of ACL, identified as a molten globule was able to retain the hemagglutination activity suggesting that the architecture of the sugar binding site either remains intact or can regain the original geometry. Molten globule states have been characterized increasingly in the past decade for the diverse range of proteins including lectins; however, to the best of our knowledge only Peanut lectin in its molten globule state has been reported to retain the carbohydrate binding capacity (Reddy *et al.,* 1999). Molten globule states of proteins show varying degree of structural organization which is dependent on the nature of proteins. Existence of premolten globule (Ptitsyn, 1995), molten-coil, and highly ordered molten globule states are documented earlier (Redfield *et al.*, 1994, Ferrer *et al.*, 1995, Morozova *et al.*, 1995). The structure of ACL molten globule could be considered to be the type of highly ordered one occurring late during the stages of protein folding as

reported for IL-4 (Redfield *et al*., 1994). Thermal unfolding studies revealed the acid induced molten globule to be rigid in nature while its chemical denaturation profile was significantly different than that of ACL at pH 7.0. The formation of this molten globule is a reversible reaction, as seen upon titration to neutral pH it refolds back to the native state. The pH dependent unfolding of ACL can be explained as a three state process such as N $\Leftrightarrow$ MG $\Rightarrow$ U. This will further pave the way for understanding the protein folding mechanism in those lectins which show specificity for complex sugars.

Chapter 4

Biophysical studies of a hemagglutinin from *Moringa oleifera* using NMR, circular dichroism (CD) and fluorescence spectroscopy

4.1. Summary

The third lectin under the present study showing specificity for complex sugar is MoL, a member of moringaceae family. Structural transitions in *Moringa oleifera* lectin (MoL) in presence of reducing agent were monitored using intrinsic and extrinsic fluorescence, CD and NMR spectroscopy. Enhanced exposure of the hydrophobic amino acids on the surface of the protein was observed by hydrophobic dye binding of the protein in presence of β-mercaptoethanol. ^{1}H NMR spectrum of MoL showed fully exposed phenylalanine and tyrosine residues in the presence of 5 mM β-mercaptoethanol. ^{1}H NMR studies of MoL also showed exposed hydrophobic amino acid patches on treatment with BME. Upon reduction of three disulphide bonds the protein undergoes transition to an intermediate state with partial loss of secondary structure and slight loosening of tertiary interactions, leading to increased hydrophobicity. The native protein showed unusually high thermostability at pH 7.0 as no transition was observed in DSC upto 130 ^{o}C, while at pH 2.0 and pH 10.0, the Tm detected was 89 and 86.7 °C, respectively.

4.2. Introduction

Moringa oleifera is a multipurpose tree belonging to family Moringaceae, which is a single genus family of shrubs and trees. Moringaceae, although very similar to the Capparidaceae, form a family of their own including one genus (Moringa) and 14 species of which *Moringa oleifera* has gained considerable importance as a flocculant, drug and food in nearly all tropic and subtropic regions (Gassenschmidt *et al.*,1995). Native to the Northern India, it is now cultivated in most of the tropical regions for its use as food and medicine (Fuglie, 2001). It is small, deciduous, perennial tree of 2.5-10 m in height, the timber of which is of low quality (Fahey, 2005). It is a fast growing and draught resistant plant and is known to be non-toxic to humans and animals (Maikokera and Kwaambwa, 2007).

The immature green pods, called as drumsticks are commonly consumed in India as a delicious vegetable. Mature seeds (Fig.4.1) are edible upon roasting like peas or nuts. Flowers and leaves are used as green leafy vegetable to cook and are highly nutritious (Nesamani, 1999).

Figure 4. 1. Mature dried seeds of *Moringa oleifera* (drumstick)

Most part of the *Moringa oleifera* tree such as bark, sap, roots, leaves, seeds, fruits (immature pods), oil and flowers act as cardiac and circulatory stimulants, possess antitumor, antipyretic, antiepileptic, antiinflammatory, antiulcer, antispasmodic, diuretic, antihypertensive, cholesterol lowering, antioxidant, antidiabetic, hepatoprotective, antibacterial and antifungal activities, and are being employed for the treatment of different ailments in the indigenous system of medicine, particularly in South Asia (Anwar *et al.*, 2007). The roots contain an alkaloid spirochin, which has been identified as a fatal nerve paralyzing agent (Morton, 1991). The root extract also acts as an anti-inflammatory agent (Ezeamuzle *et al.*, 1996). The leaf extract has been reported to be hepatoprotective and hypotensive (Pari and Kumar, 2002; Faizi *et al.*, 1995) as well as having antitumor activities (Murakami *et al.*, 1998). The *Moringa oleifera* leaf extract shows significant antiproliferative and apoptotic potential on human tumor (KB) cell line and has been claimed as therapeutic target for cancer (Sreelatha *et al.*, 2011). The aqueous extract from leaves of *M. oleifera* has been claimed as an ethnomedicine for the treatment of diabetes mellitus (Jaiswal *et al*., 2009). The ethanolic extract of seeds of *Moringa oleifera* possesses anti-arthritic and anti-inflammatory activities and their inhibitory action on systemic and local anaphylaxis are well investigated (Mahajan and Mehta, 2007).

The seeds contain 38-40% edible oil, also called Ben oil, which can be used in cooking, cosmetics, and lubrication. It has physical and chemical properties equivalent to that of olive oil and contains a large quantity of tocopherols and > 80% unsaturated fatty acid content (Tsaknis *et al.,* 1999; Mohammed *et al.*, 2003). Apart from this, the seeds show strong coagulative and antimicrobial properties (Eilert *et al.*, 1981). In Taiwan and

China, these seeds are used to treat athlete's foot and *Tinea* infections (Chuang *et al.*, 2007). The *Moringa oleifera* seed flour is traditionally utilized as a coagulant in water treatment (Gassenschmidt *et al.*, 1995). Use of these seeds for softening the hard water has also been reported (Muyibi and Evison, 1995).

The isolation, purification and structure-function studies of *Moringa oleifera* lectin (MoL) has been reported from our laboratory (Katre *et al.*, 2008a and 2008b). The lectin resembles the same fluocculating protein from *Moringa oleifera* described previously (Gassenschmidt *et al.*, 1995; Ndabigengesere *et al.*, 1995). MoL is a single tryptophan containing 14 kDa dimeric lectin (MoL), showing specificity for complex sugars (Katre *et al.*, 2008a). MoL was found to be highly basic in nature due to high arginine content. The protein is highly stable and active in extremes of pH. MoL has three cysteine residues per monomer of the protein (Katre UV, 2007). The tertiary structure of MoL was stabilised by the presence of three disulphide bonds making the protein a highly thermostable one. In solution MoL was found to exist as a mixture of isolectins differing in charge which could be a reason why crystallization experiments have not yet succeeded.

Proteins undergo transition to their partially folded conformation by alteration in the native conditions induced by changing pH, temperature or due to the denaturants. A molten globule is a partially folded state of protein having native like compact secondary structure but disordered tertiary interactions (Ptitsyn, 1995). Thus, the lack of fixed tertiary structure in the molten globule conformation results in a heterogeneous population of species at a given time point. The ^{1}H NMR spectrum of the partially folded species differ substantially from those of both the native and fully unfolded states

showing the intermediate level of order. Studies on partially folded states provide significant understanding of the protein folding problem as the molten globule states are well characterized in the initial as well as in the final stages of the folding of Lysozyme (Morozova *et al.,* 1995) and IL-4 (Redfield *et. al*., 1994). The present chapter is concerned with the characterization of a partially unfolded intermediate of MoL, resulted from the breakage of disulphide bond, based on the solution studies using NMR and other spectroscopic tools.

4.3. Materials and methods

4.3.1. Materials

Moringa oleifera seeds were purchased from the local market. DEAE cellulose and CM-Sephadex, ANS, β-mercaptoethanol, D_2O and Guanidine hydrochloride (GdnHCl) were procured from Sigma (USA). Sodium chloride, sodium hydroxide, potassium dihydrogen phosphate, dipotassium hydrogen phosphate and glycine were purchased from SRL, India. All other chemicals were of analytical grade and purchased from local suppliers.

4.3.2. Purification of Mol from Drumstick seeds

The MoL was purified from the crude seed extract of mature drumstick by 90 % ammonium sulfate precipitation followed by two successive ion exchange chromatography steps of DEAE cellulose and CM- sephadex according to the method of Katre *et al.,* (2008a). Purified MoL was checked for hemagglutination activity using 3 % rabbit RBC's suspension in PBS.

4.3.3. Circular dichroism (CD) spectroscopy

CD spectra of MoL samples were recorded at room temperature using a Jasco J - 715 spectropolarimeter at 25 ^{0}C in the range of wavelengths 200–250 nm at a scan speed of 100 nm min^{-1} with a response time of 1 s and slit width 1 nm. A rectangular quartz cell of 1 mm path length was used.

To analyse the effect of reducing agents on the secondary structure of MoL (0.08 mg ml^{-1}), β-ME was used in varying concentration of 2-15 mM at pH 2.0, 7.0 and 10.0. Sample in each condition was incubated for 20 min at room temperature after adding β-ME and the spectra were recorded. To monitor the tertiary structure of the protein at pH 2.0, 7.0 and 10.0 in the presence of β-ME, CD spectra in the near UV region (250-300nm) were recorded using a cuvette of path length 1 cm. The lectin concentration used was 0.8 mg ml^{-1}. Each spectrum was the average of 5 scans. Buffer scans recorded under the same conditions were subtracted from the lectin spectra for further analysis.

4.3.4. Fluorescence spectroscopy

The intrinsic fluorescence of the protein was measured using a Perkin Elmer Life Sciences LS50 fluorescence spectrophotometer. The lectin solution (~0.02 mg ml^{-1}) was excited at 295 nm and the emission spectra were recorded between 300-400 nm at different concentration of β-ME (2, 5, 10 and 15 mM). Both the excitation and emission spectra were obtained by setting the slit-width at 7 nm, and speed 100 nm min^{-1}. The background emission due to the signal produced by either buffer solution, or buffer containing the appropriate quantity of reducing agent was subtracted prior to analysis.

4.3.5. Binding of Hydrophobic dye ANS to MoL

The binding of the hydrophobic dye 8-Anilino-1-naphthalene sulfonic acid (ANS) to MoL upon unfolding with β-ME was studied by recording the emission spectra in the range 430–550 nm with excitation at 375 nm using steady-state spectrofluorimeter. The protein was incubated at pH 2.0, 7.0 and 10.0 in the respective buffers for 5 hrs prior to addition of β-ME. 5 µl of 25 mM ANS was mixed with 2 ml of protein (0.02 mg ml^{-1}) solution treated with β-ME . Buffer spectrum in the presence of β-ME and ANS in each of the condition was subtracted from the spectrum of the protein.

4.3.6. Differential scanning calorimetry (DSC)

DSC is a powerful technique to characterize temperature-induced conformational changes in proteins and other biological macromolecules. Solution differential scanning calorimetry (DSC) measures the difference in heat energy uptake (C_p endotherm) between a sample solution and reference (buffer/solvent) with increase in temperature. A typical experiment comprises one (or more) scans of the sample solution, together with separate control using buffer alone to establish the instrumental baseline. A differential scanning calorimeter basically consists of two cells: the reference cell and the sample cell. To carry DSC on proteins the sample cell contains diluted protein solution and the reference cell has the pure solvent (buffer). Both cells are simultaneously heated at a constant scan rate and the difference in the heat capacity between the two cells is recorded by the instrument. In order to check the reversibility of the processes, DSC experiments on protein denaturation usually include a “reheating run”; that is, a second scan carried out after cooling the protein solution (inside the calorimetric cell) to room temperature. Typically, a process is considered calorimetrically reversible if 85–90% of the endotherm is recovered in the reheating run. Thermodynamic parameters

characterizing the unfolding transition(s) such as transition temperature (T_m), calorimetric enthalpy (ΔH_{cal}), van't Hoff enthalpy (ΔH_{vH}), and the changes in excess heat capacity (ΔCp) can be obtained from the DSC measurements. ΔH_{cal}, is equal to the area under the curve, and ΔH_{vH} is calculated by the standard formula given below (Xie *et al.*, 1991)

$$\Delta H_{vH} = 4\, RT_m^2\, C_{pmax} / \Delta H_{cal} \qquad (4.1)$$

Where, C_{pmax} is the maximum of the excess heat capacity function, T_m, is the transition temperature defined as the temperature location of C_{pmax} and R is the gas constant.

DSC measurements of MoL samples were made on a MicroCal VP-DSC differential scanning calorimeter (MicroCal LLC, Northampton, MA, USA) equipped with two fixed cells, a reference cell and a sample cell at Molecular Biophysics Unit, IISC, Bangalore, India. DSC experiments were carried out as a function of pH and protein concentration. The scan rate of 30 K/hr was kept constant for all the measurements. Samples were incubated for 5 hrs at room temperature before recording the thermograms. Buffer and protein solutions were degassed before loading. All the data were analyzed by using the Origin DSC software provided by the manufacturer.

4.3.7. NMR studies on MoL

A detailed understanding of the function of a biological macromolecule requires knowledge of its three-dimensional structure. Most atomic-resolution structures of biological macromolecules have been solved either by X-ray diffraction of single crystals or by nuclear magnetic resonance (NMR) in solution. In order to get the structural details

since the several crystallization experiments did not yield crystals of MoL, detailed NMR investigations were carried out in native and unfolded states.

4.3.7.1. The principle of NMR spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is an important analytical technique for determination of molecular structure and dynamics in solution and solid state. It provides a molecular level understanding of the structure and dynamics in systems as diverse as biomolecules, to simple organic molecules. In the area of solution state biomolecular NMR spectroscopy, recent developments in NMR instrumentation and methodology together with improved isotope labelling schemes in proteins have brought about significant advances in their three-dimensional structure determination.

The basis of all NMR measurements is a quantum mechanical property of the nucleus: the spin. The nuclei of interest in NMR of biological macromolecules carry a spin of ½ which allows only two different spin states often referred to as spin up (+ ½) and spin down (- ½) (Wider, 2000). In some atoms (^{12}C, ^{16}O, ^{32}S) these spins are paired and cancel each other out so that the nucleus of the atom has no overall spin. However in many atoms (^{1}H, ^{13}C, ^{31}P, ^{15}N, ^{19}F etc.) the nucleus does possess an overall spin. The overall spin, I, is important. Quantum mechanics tells us that a nucleus of spin I will have 2I + 1 possible orientations. A nucleus with spin 1/2 will have 2 possible orientations. In the absence of an external magnetic field, these orientations are of equal energy. If a magnetic field is applied, then the energy levels split. Each level is given a magnetic quantum number, m with values I, (I-1), (I-2)…. -I. Transitions between these energy levels ($\Delta m = \pm 1$) can be achieved by applying electromagnetic radiation in the region of radio frequencies (RF).

Associated with the spin is a magnetic moment which for a spin can be interpreted as a magnetic dipole. These very small atomic dipoles can only orient parallel or antiparallel to an external magnetic field. The two possible orientations correspond to slightly different energies and spins are allowed to jump from one orientation to the other, absorbing or emitting the energy difference in the form of electromagnetic radiation. Only the very small difference between the number of parallel and anti-parallel spins contributes to the NMR signal. This fact explains why NMR is an insensitive technique compared to other optical absorption spectroscopy.

4.3.7.2. The NMR setup

NMR experiments are performed on a NMR spectrometer consisting of two components: a high field superconductive magnet which produces an extremely homogeneous, strong static magnetic field and a console that generates and controls short bursts (pulses) of high power RF energy that is used to excite the sample in the probe. The NMR console also receives and detects the very weak signals coming back from the probe as a response to the pulse. Probe is the name given to that part of the spectrometer that accepts the sample, sends RF energy into the sample, and detects the signal coming from the sample. For an NMR experiment a glass tube of 5 mm diameter containing the solution of the molecule of interest is placed in the room temperature bore of the superconductive magnet.

4.3.7.3. The NMR Parameters

Three most important NMR parameters used for structural characterization are Chemical shift (δ), coupling constant (J) and relaxation times (T_1 and T_2). Chemical shift provides information about the chemical environment of various NMR active nucleus

present in a molecule while the coupling constant J shows how various nuclei are linked through chemical bonds in the molecule. Besides, the J value carries the signature of dihedral angles of the interacting spins. The relaxation times T1 and T2 depends on the dynamics of the molecule and T1 also depends on the proximity of the dipoles and hence related to the three dimensional structure through a phenomenon called Nuclear Overhauser Effects (NOE). Through-space correlations provide the basis for geometric information required to determine the three-dimensional structure of a molecule. The NOE reflects the transfer of magnetization between spins interacting via their associated dipoles. This technique allows the measurement of the inter-proton distances, because:

$$NOE_{ij} \propto r_{ij}^{-6} \qquad (4.2)$$

where NOE_{ij} is the intensity enhancement observed between nucleus i and j and r_{ij} is the distance between the two nuclei. Due to the dependence on the inverse sixth power of the distance between the nuclei the NOE intensity falls off rapidly with increasing distance. In general NOEs between protons separated by more than 0.5 nm are usually not observed. NMR experiments which measure the NOE are often referred to as NOESY experiments where NOESY stands for NOE Spectroscopy.

4.3.7.4. Applications of NMR in biology

NMR spectroscopy employs measurement of chemical shifts which are easily accessible and richly informative. For proteins in particular, chemical shifts can reveal exquisitely detailed information about backbone dihedral angles, side-chain χ angles, hydrogen bond interactions, local electric fields, proximity and orientation of aromatic rings, ionization states, oxidation states, backbone dynamics, ring-flip rates, and even internuclear O–H distances (Szilagyi, 1995). Although NMR has some limitations with

respect to the size of the molecules and the resolution of the structures, it provides considerably more than mere structural information, i.e. it goes beyond a static picture of the three-dimensional structure and gives functionally relevant information on molecular dynamics, thermodynamics and kinetic aspects to characterize weak and strong interactions between biomacromolecules and small ligands. Transient intermolecular interactions are crucial for phenomena such as regulation of protein expression, enzyme activation/repression, signaling, etc. NMR is particularly well suited to study weak and transient interactions, as it allows researchers to investigate the systems of interest in solution, which is often the physiological state, or, for membrane systems, in bi-layers or micelles, which simulate the membrane environment. Protein-protein interactions, when weak, cannot be studied with any other technique at the atomic level. In this respect the results of interaction studies by NMR can provide unique information for structural analysis of the interactome maps.

Although the chemical shift is primarily determined by the covalent structure of the amino acid residue, it can also be significantly affected by the interactions with the solvent. Therefore, the exclusion of the solvent water from the interior of a globular protein causes the chemical shifts of the core residues to be different from those of the water-exposed amino acid residues, so that even NMR lines originating from multiple residues of the same amino acid type can be distinguished. This "conformation-dependent chemical shift dispersion" was found to be sufficiently large to enable ^{1}H NMR studies of protein denaturation.

In proteins which are isotope labeled with ^{15}N and ^{13}C, J couplings between ^{1}H, ^{15}N and ^{13}C allow through-bond correlations across the peptide bond. NMR experiments

which correlate different nuclei via J coupling are often referred to as COSY experiments (Correlation Spectroscopy) (Aue *et al.,* 1976; Wider, 1998; Wider *et al.*, 1984). An important feature of COSY experiments is that they can transfer magnetization between different types of nuclei. This property opens the possibility to start an experiment with one type of nucleus and to transfer the magnetization to another or several other types of nuclei in the course of the experiment. Such magnetization transfers are of great practical importance since widely different sensitivities are obtained with different types of nuclei especially with biological macromolecules.

The NMR method for protein structure determination relies on a dense network of distance constraints derived from NOEs between nearby hydrogen atoms in the protein (Wüthrich, 1986). NOEs connect pairs of hydrogen atoms separated by less than 0.5 nm. In contrast to COSY-type experiments the nuclei involved in the NOE correlation can belong to amino acid residues that may be far apart along the protein sequence but close in space which gives valuable information in structure determination. Recent strategies employed for protein structural characterization are based on uniformly ^{13}C and ^{15}N labeled proteins using a series of triple resonance experiments involving ^{1}H, ^{13}C and ^{15}N nuclei.

To record the NMR spectrum of MoL in native and unfolded state all the measurements were carried out on a Bruker AV 500 spectrometer operating at 500.13 MHz, 125.75 MHz, respectively for ^{1}H and ^{13}C at the Central NMR facility provided at NCL, Pune. A sample with 40 mg ml^{-1} of the protein dissolved in 10 % D_2O in a standard 5mm NMR tube was prepared and the ^{1}H, COSY, TOCSY, ^{1}H decoupled ^{13}C, ^{13}C DEPT, ^{13}C-^{1}H HSQC, ^{13}C-^{1}H HMBC were performed. The one dimensional ^{13}C experiments

were performed on 5mm BBFO probe at ambient temperature (~ 28 ^{o}C). All the 2D experiments were conducted either on a 5mm broad band inverse gradient or BBFO probe. Gradient spectroscopic techniques were employed for all the 2D experiments. 20000 and 10000 transients were collected for ^{13}C and DEPT135 spectra. To record the ^{1}H and ^{13} C NMR spectrum of MoL in in presence of reducing agent 5 mM β-ME (freshly prepared) was added to the MoL in the NMR tube and experiments were carried out as stated above.

The COSY and the HMBC spectra were collected in a magnitude mode while a phase sensitive (States-TPPI) mode was used for HSQC, NOESY measurements. A mixing time of 1 sec and 300 m sec was employed for NOESY and ROESY experiments, respectively. The numbers of scans used for each t1 increment for other 2D experiment were as follows: 64 (^{13}C HSQC,), 200 (^{13}C HMBC), 32 (COSY), 24 (TOCSY). The ^{13}C HMBC data were optimized for a long range coupling constant of 6 Hz. A pulse sequence employing a double low pass filter was found to give better results for ^{13}C HMBC due to spread in 1JC-H values (160 -135 Hz). The HMBC spectra were acquired without proton decoupling during detection. The 90º pulse lengths for ^{1}H and ^{13}C were 13.5 and 10 μ sec, respectively. Appropriate window functions viz. sine squared bell with no phase shift for all magnitude mode and phase shifted (ssb = 2) sine squared bell for phase sensitive mode were used for data processing. In general a 2Kx2K data matrix size was used for the 2D experiments. The ^{1}H and ^{13}C chemical shifts were indirectly referred to TMS.

4.4. Results and discussion

Purity and homogeneity of MoL preparation was confirmed by 15 % SDS-PAGE (Fig. 4.2) where MoL showed single band corresponding to a subunit mass of 7 kDa

however the protein exists as a dimer in solution as confirmed by size- exclusion chromatography. To start with a few of the initial experiments reported previously (Katre *et al*., 2008a) were repeated to ensure the reproducibility.

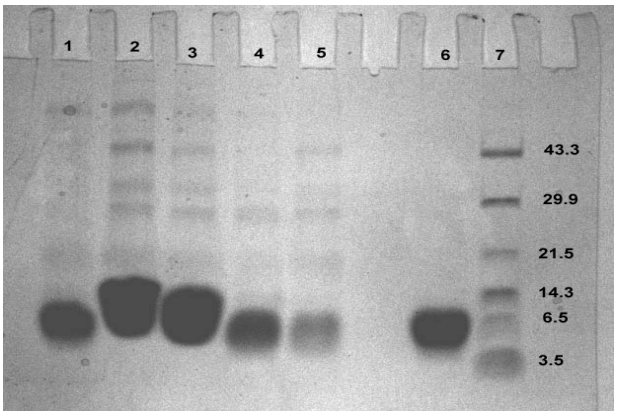

Figure 4.2. Purification profile of MoL. Lane 1- Crude extract, Lane 2- Ammonium sulfate precipitate, lane 3, 4 and 5- DEAE cellulose unadsorbed fractions, lane 6- CM sephadex elute, lane 7- Molecular weight marker (low range, GENEI, Bangalore).

4.4.1. Effect of reducing agents

We have previously reported the effect of reducing agents on MoL incubated with BME at pH 7.0 and 10 (Katre *et al*., 2008a). Far UV profiles of MoL showed partial loss of the secondary structure at pH 7.0 and drastic loss at pH 10 in the presence of BME. The protein possesses three disulfide bonds as reported by Katre *et al*, (2008a). To start with, the previous experiments were repeated to verify that the sample preparation matches with the already reported one (Fig.4.3 A). About 40 % loss of secondary structure of MoL at pH 7.0 in the presence of reducing agent was observed and the effect was more prominent at pH 10 where MoL is significantly unfolded. Further

characterization of the protein was taken up. The near UV CD spectra indicated considerable loss of tertiary interactions at pH10 (Fig. 4.3B), while only minor perturbance was observed at pH 7.0 in the presence of 5 mM BME.

Intrinsic fluorescence spectra of MoL in presence of 5 mM BME showed 12 % quenching of the fluorescence intensity indicating a slight change in the microenvironment of surface exposed tryptophan (Fig. 4.3C). The hydrophobic amino acids present on the surface of native MoL get more exposed at pH 10.0 and pH 2.0 (Katre *et al*., 2008b). In the present work, hydrophobic dye binding studies of MoL at pH7.0 in the presence of BME showed relative increase in fluorescence intensity combined with the blue shift to 480 nm (Fig.4.3 D). This could be due to the rearrangement of the protein structure consequent to reduction of the disulfide bonds.

Unfolding of MoL by BME was found to be irreversible as the dilution of the reagent did not lead to the folded structure. MoL remained in the unfolded state indicating that the disulphide bridges are crucial for the native fold of the protein.

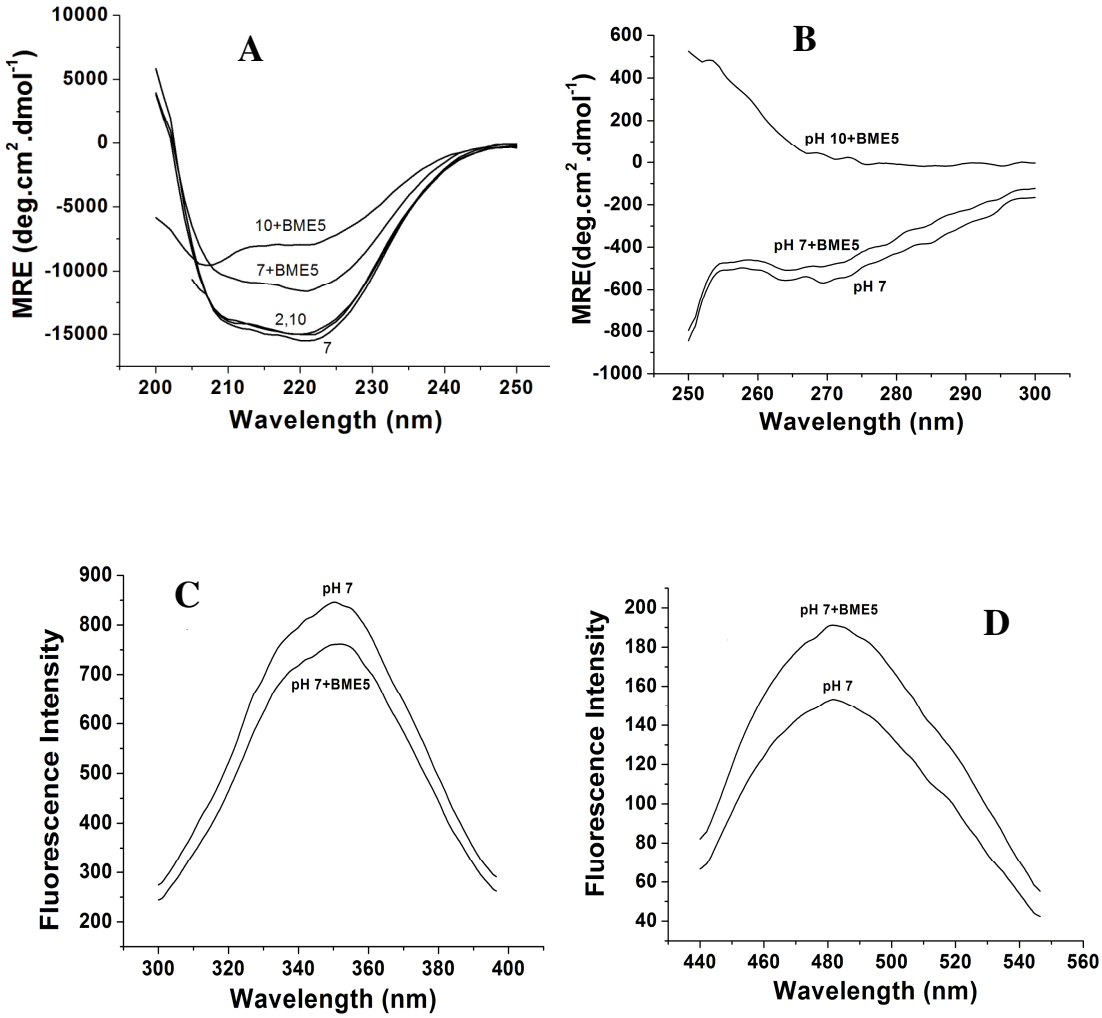

Figure 4.3. Effect of β-ME on structure and intrinsic fluorescence of MoL. Far-UV (0.1 mg ml^{-1}) (**A**) and Near-UV (1 mg ml^{-1}) (**B**) CD spectra of MoL at pH 2.0, 7 and 10 in the presence of 5 mM βME (The numbers on the spectra correspond to the pH of the protein) Intrinsic fluorescence (0.02 mg/ml)(**C**) and ANS binding to MoL (**D**) at pH 7.0 and in the presence of β-ME. (The numbers on the spectra correspond to the pH of the protein)

4.4.2. NMR studies on MoL

4.4.2.1. ^{1}H and ^{13}CNMR studies

The ^{1}H NMR spectrum of purified MoL (40 mg ml^{-1}) was recorded in 90:10 H2O:D2O mixtures at pH 7.0. As expected the ^{1}H NMR spectrum was dominated by the strong water peak (Fig.4.4) and hence required water suppression by a suitable technique. Suppression of the strong water signal is necessary in order to overcome the dynamic range problems. The ^{1}H spectrum obtained after presaturation technique is presented in fig 4.5. The signals from the protein molecules can clearly be borne out from this proton NMR spectrum. Well dispersed NMR spectra of the MoL in the native state showed the characteristic of properly folded protein.

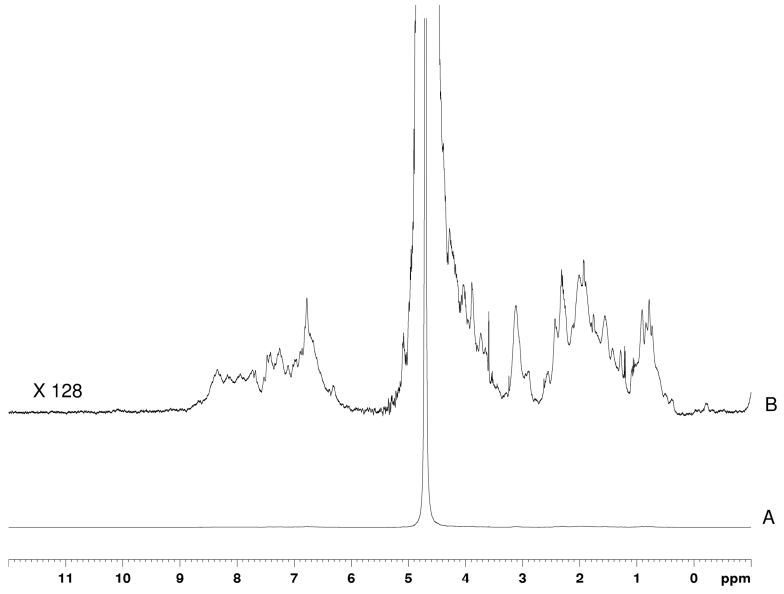

Figure 4.4. 500 MHz ^{1}H NMR spectrum of MoL at pH 7.0 (A) only the water peak is visible. The amplified spectrum is shown in B.

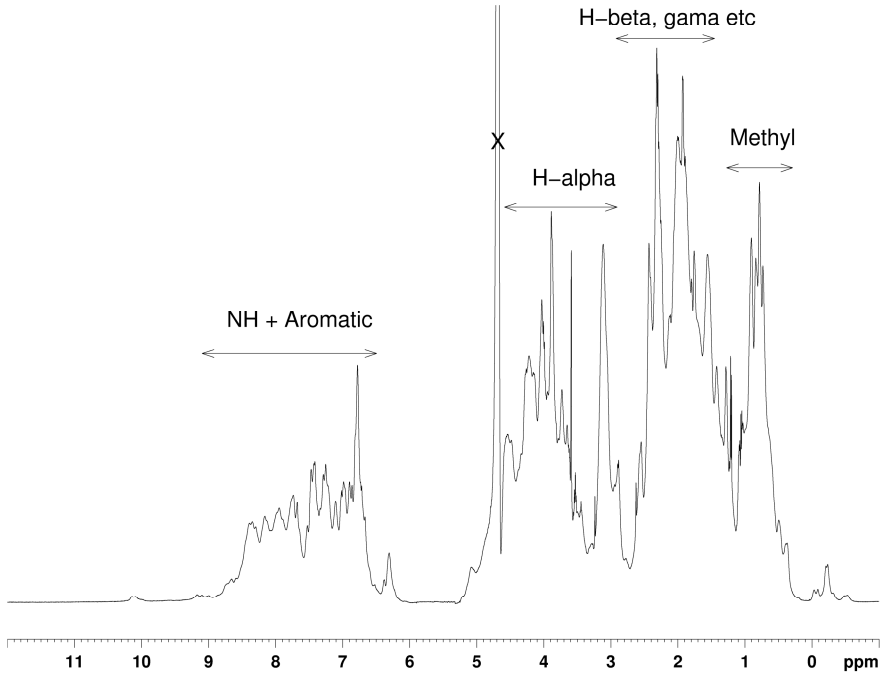

Figure 4.5. Well dispersed ^{1}H NMR spectrum of MoL at pH 7 as seen after water suppression. The methyl, aliphatic, aromatic and NH protons peaks are indicated by double headed arrows on the spectrum.

The peaks at 1 to -1 ppm region indicate the presence of methyl protons. α and β protons cover the region between 1-5 ppm. Of these α protons are more down fielded and it should be generally in the region of 4 -5 ppm. The intense peaks at the region 6-7.5 indicate the presence of aromatic protons and NH protons (Fig. 4.5). N-H protons in unstructured (unfolded or random coil) protein occur in the region 7.5-8.5 ppm. In a protein with a well defined folded conformation, the hydrogen bonding leads to down field shift. So in folded proteins the N-H peaks are visible in the region 8-11 ppm. The

presence of peaks below zero ppm is also an indication for the presence of folding as observed in case of MoL. These are the peaks of methyl protons, which are more shielded probably by the anisotropic effect of aromatic ring. Due to folding some of the alkyl protons may come in the shielded cone of the ring current and they are responsible for this low δ value. It is very difficult to get detailed information from the ^{1}H NMR spectrum of MoL. The weak peaks around 10 ppm is likely to be of the single tryptophan in the protein.

The ^{13}C spectrum of MoL is shown in fig 4.6 which shows a comparison of the ^{1}H broad band decoupled and DEPT (Distortion less Enhancement by Polarization Transfer) 135 spectra.

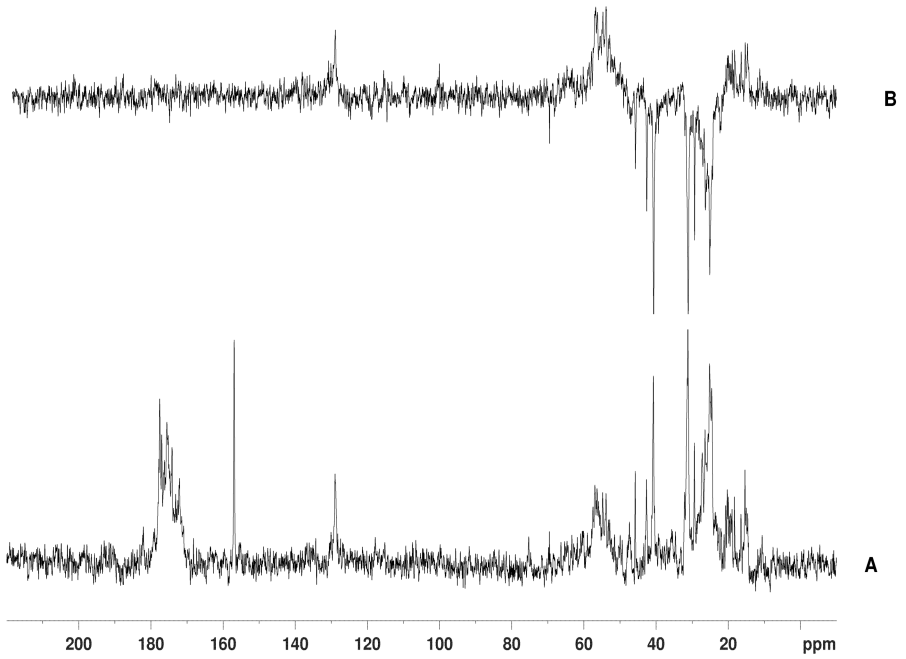

Figure 4.6. ^{13}C NMR spectrum of MoL at pH 7.0. The lower trace (A) is the ^{1}H decoupled spectrum while the upper trace is the DEPT135 spectrum (B).

The later spectrum is used to differentiate the different types of carbons present in the protein. Here, the CH_2 carbons appear with a negative phase while the CH and CH3 carbons appear with a positive phase. Quaternary carbons (carbons without any attached protons) such as the amide carbonyls, acid carbonyls etc. will not appear. The signals in the region 170 185 ppm are of various cabonyls while a sharp signal at 157 ppm is typical of the guanidine moiety of Arg. The large intensity present, to a first approximation, is indicative of high Arg content in MoL. The evidence for the presence of aromatic amino acids is obtained from the signals in the region 120-135 ppm. All the other aliphatic carbons appear in the 70 to 10 ppm. In general, the alpha carbons resonates in the region 60-50 δ, the beta carbons and other CH2 carbons in the region 50-25 δ and the methyl carbons are most shielded (25-10 δ). No further NMR studies were performed on this system due to lack of sensitivity and broadness of the peaks at the available magnetic field strength (11.7 T).

Since MoL is known to contain three Cys residues, the formation of disulphide linkages cannot be ruled out. In order to get structural information about the reduced form of MoL, we have recorded the ^{1}H NMR spectrum of MoL in presence of 5mM BME (Fig. 4.7). This will also help us to study the dynamics of the flexible regions of MoL. A comparison of the ^{1}H NMR spectrum of the native protein and the BME treated one are presented in Fig.4.8. The occurrence of structural changes is very much evident from this. The major structural changes were observed in the aromatic and the methyl regions as indicated by narrowing of the resonances and minor changes in ^{1}H chemical shifts (Fig.4.9 & 4.10) It is interesting to note at this point that the doublet signal at

~1.31δ and 1.15 δ, presumably arising from Ala and Thr, are found to be narrow in the native protein (Fig.4.11)

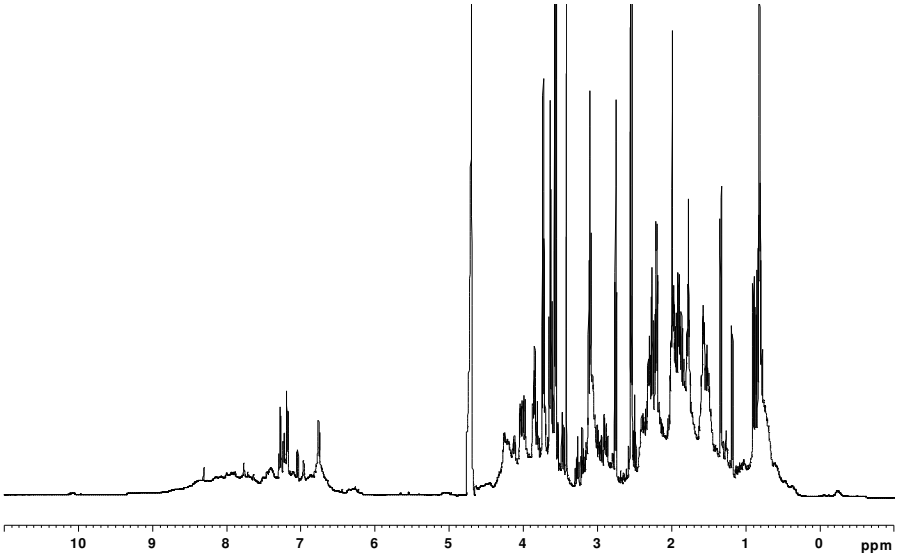

Figure 4.7. ^{1}H NMR spectrum of MoL at pH 7 in presence of 5 mM β-ME.

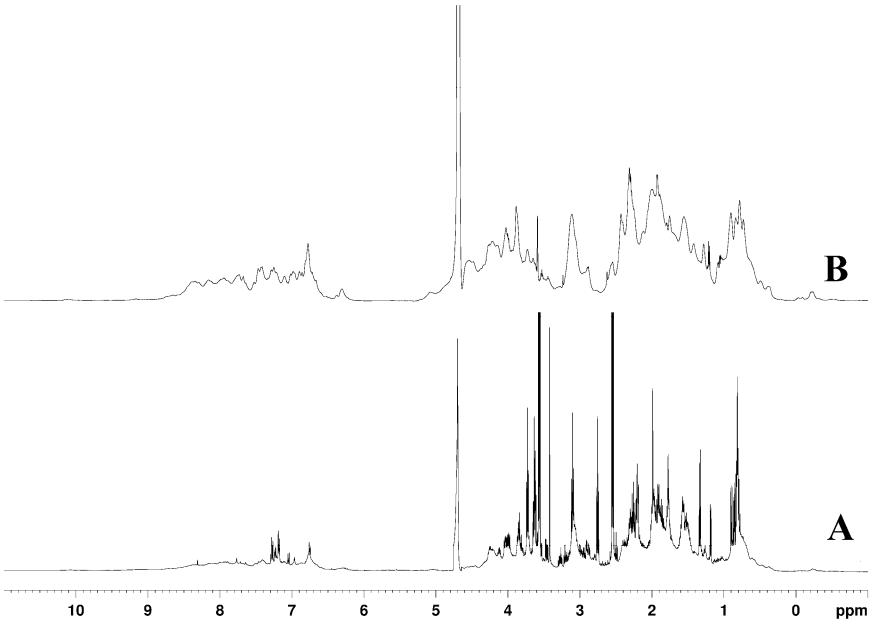

Figure 4.8. Comparison of the water suppressed ^{1}H NMR spectrum of MoL at pH 7.0 in presence of 5 mM β-ME (A) and the native protein (B). The sharpening of methyl, aliphatic protons can be clearly visible in (A).

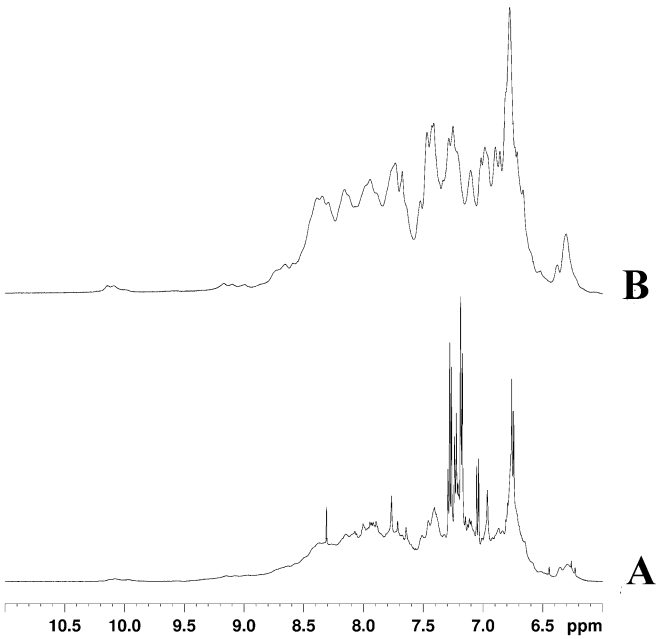

Figure 4.9. Comparison of the water suppressed 1H NMR spectrum of the NH and aromatic region of MoL at pH 7 in presence of 5 mM β-ME (A) and the native protein (B). The sharpening of aromatic protons can be clearly seen in the spectrum (A).

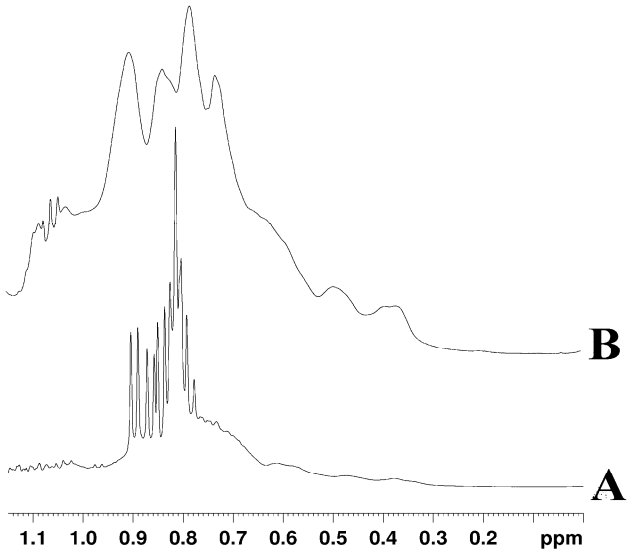

Figure 4.10 Comparison of the water suppressed ^{1}H NMR spectrum of the methyl region of MoL at pH 7.0 in presence of 5 mM β-ME (A) and the native protein (B). The sharpening of various methyl protons can be clearly seen in (A)

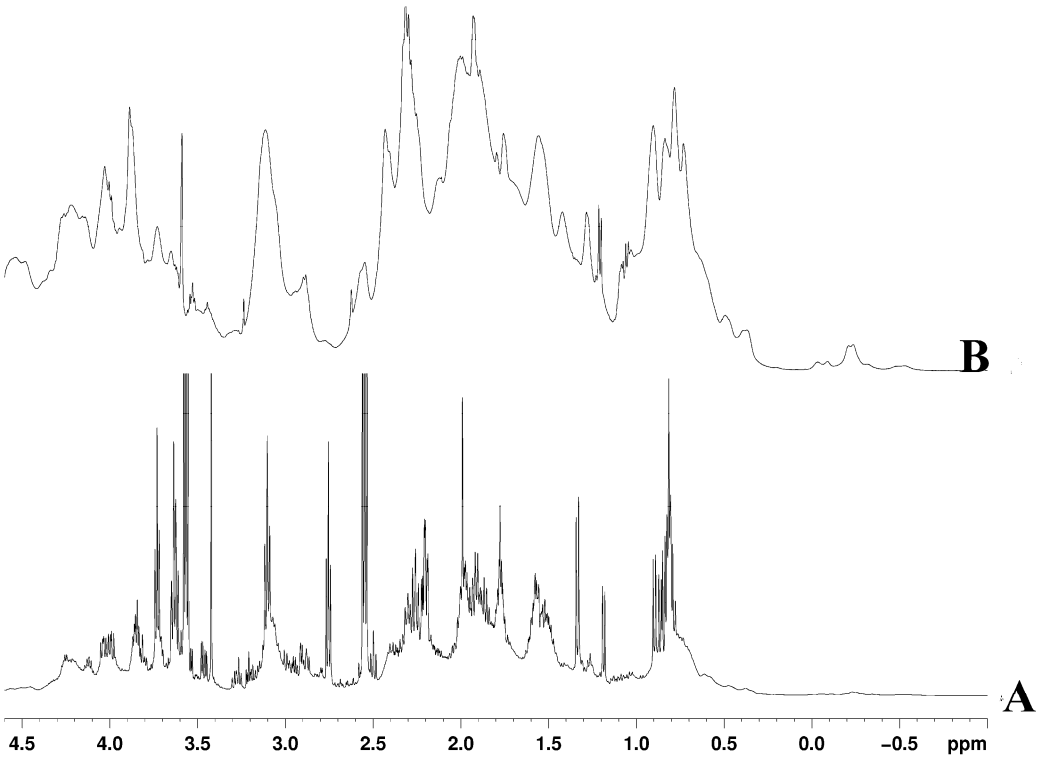

Figure 4.11. Comparison of the water suppressed ^{1}H NMR spectrum of the aliphatic region of MoL at pH 7.0 in presence of 5 mM β-ME (A) and the native protein (B) The sharpening of the various methyl and aliphatic resonances can be clearly seen from the spectrum.

The narrowing of the ^{1}H signals can be correlated with the enhanced exposure of hydrophobic amino acid residues in presence of BME. It is very evident that the aromatic amino acids (phenyalanine and tyrosine) and the amino acids with methyl side chains (Leu, Ile and Val) are exposed upon treatment with BME (Fig. 4.9 and 4.10). Similarly more resolved peaks were observed in the alpha, beta and gamma protons regions of the protein upon breakage of disulphide bridges compared to the native MoL (Fig. 4.11). Similar behaviour could also be seen in the ^{13}C spectrum of the protein and some of the

characteristic signals from the Phe and Tyr could easily be assigned. Addition of more amounts of BME (10mM) did not bring out any further change.

It is observed that in presence of β-ME the protein sample remains stable for a long time. The temporal behaviour of the sample, monitored for about 210 days is shown in fig.4.12. Two strong peaks (characteristic triplets) around 2.65 and 3.67 ppm correspond to the unreacted BME present in the sample recorded immediately after mixing. These signals slowly disappear as they get oxidized to its disulphide form which resonates at 2.85 and 3.83 ppm. In addition, narrowing of the signals is also observed with time.

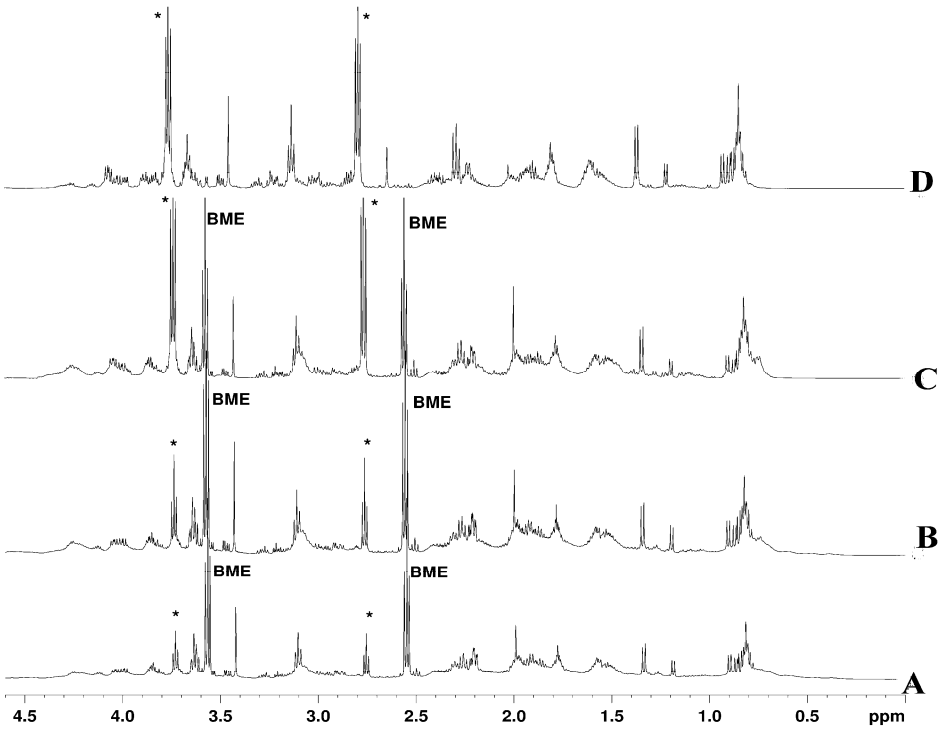

Figure 4.12. Comparison of the water suppressed ^{1}H NMR spectrum of MoL at pH 7.0 in presence of 5 mM β-ME as a function of time. (A) Immediately after mixing, (B) on the second day, (C) after 140 days and (D) after 210 days. The signals for BME is labelled. The signals with * belongs to the oxidized form of

4.4.4.2. COSY, TOCSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC correlation studies

Attempts has been made to assign signals of various amino acid residues in the protein sample containing 10 mM BME by a combination of standard 1D (^{1}H and ^{13}C) and 2D (COSY, TOCSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC) experiments.

The COSY spectrum correlates the proton signals that are coupled by the geminal (2-bond) and vicinal (3-bond) couplings. In this 2D spectrum, which is a square matrix, both the axes has proton chemical shifts and the presence of coupling is manifested as cross peaks. From this the chemical shifts of the coupled partners can be obtained. The diagonal peak is nothing but the 1D spectrum. This is a powerful technique which can be used for identification of the protons present in a particular amino acid residue. COSY spectrum of aliphatic regions of MoL is shown in fig.4.13.

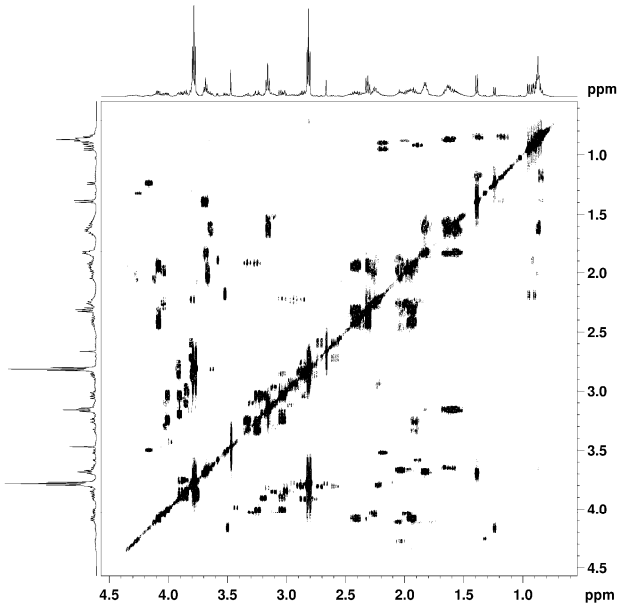

Figure 4.13. COSY spectrum of MoL at pH 7 in presence of 10

For example, from the part of the COSY spectrum of the methyl regions (Fig. 4.14) Ile, Leu and Val residues could easily be identified. Similar approach can also be employed for identification of some of the amino acid residues such as Ala, Thr etc.

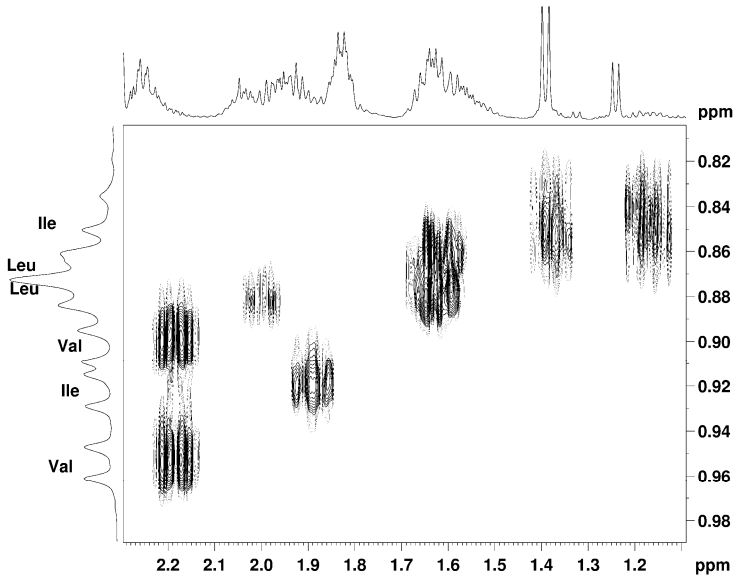

Figure 4.14. COSY spectrum of the methyl region of MoL at pH 7 in presence of 10 mM β-ME.

Another correlation experiment that is commonly employed for structural characterization is TOCSY. This connects the entire spin system in a coupled network. Hence, by knowing the assignment of one of the signals in the network, the assignments of other resonances can be made. The TOCSY spectrum of the aliphatic and methyl regions is shown in fig. 4.15 & 4.16. More number of cross peaks can be seen in the TOCSY spectrum compared to the corresponding COSY spectrum. Thus, form the TOCSY spectrum of the methyl region the assignments of the corresponding alpha protons are also made.

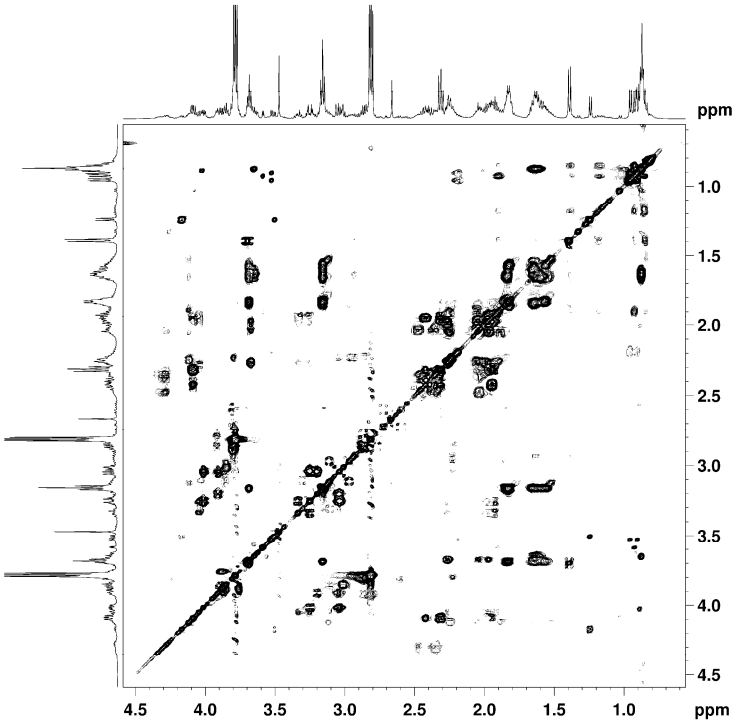

Figure 4.15. TOCSY spectrum of of MoL at pH 7.0 in presence of 10 mM β-ME

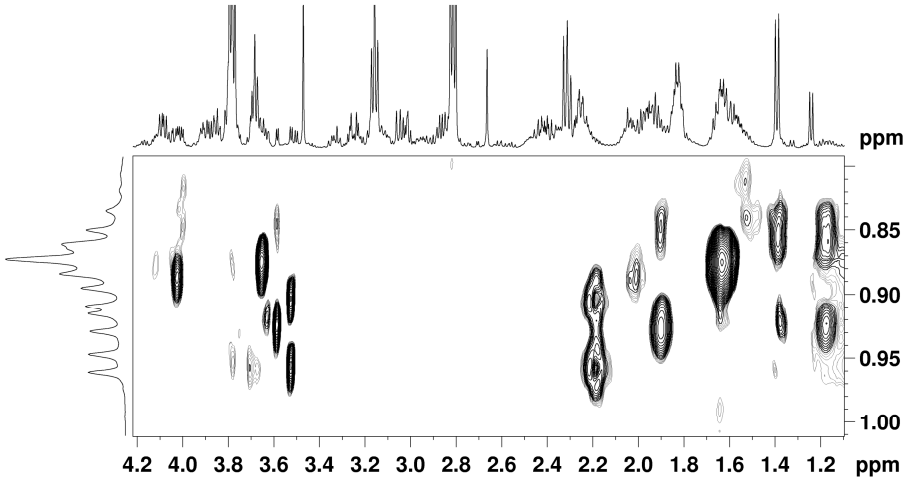

Figure 4.16. TOCSY spectrum of the methyl region of MoL at pH 7.0 in presence of 10 mM β-ME.

A complete assignment of residues could not be obtained with these two techniques due to overlap of proton signals. Hence ^{1}H-^{13}C hetero nuclear experiments (HSQC and HMBC) were conducted on the sample containing 10 mM BME. The ^{1}H-^{13}C HSQC (Hetero-nuclear Single Quantum Coherence) experiment connects the carbons with its attached protons. Hence, from the ^{1}H assignment, one can assign the ^{13}C signals or vice versa. ^{1}H-^{13}C HMBC (Hetero-nuclear Multiple bond Correlation) correlates the ^{13}C chemical shift with the protons which are two or three bonds away from it. These techniques make use of the one bond and 2/3 bond scalar coupling between ^{13}C and ^{1}H nuclei. The HSQC spectrum of the aliphatic region, methyl region and the aromatic region of the BME containing MoL is shown in fig. 4.17- 4.19.

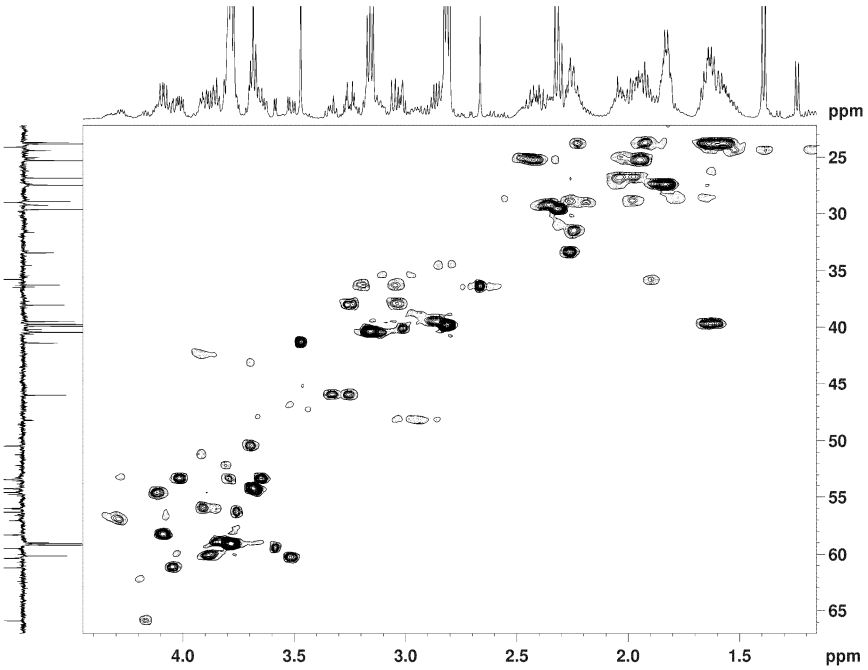

Figure 4.17. HSQC spectrum of the aliphatic region of MoL at pH 7.0 in presence of 10 mM β-ME

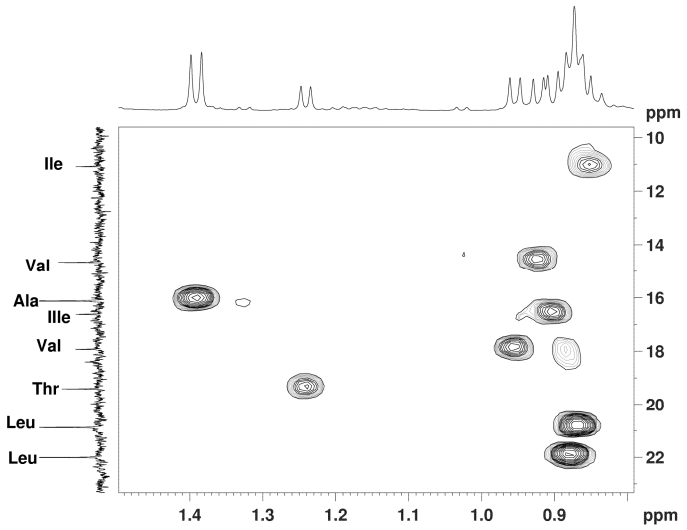

Figure 4.18. HSQC spectrum of the methyl region of MoL at pH 7.0 in presence of 10 mM β-ME

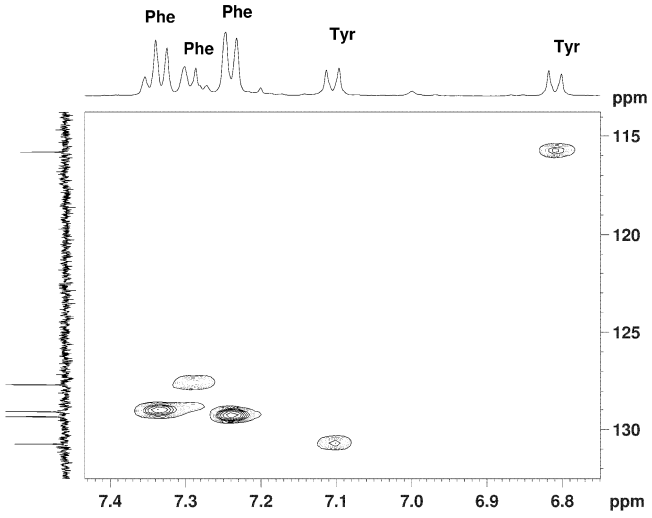

Figure 4.19. HSQC spectrum of the aromatic region of MoL at pH 7.0 in presence of 10 mM β-ME

Assignments of directly bonded ^{13}C and ^{1}H resonances are only possible by the HSQC technique. The assignments of carbons without any attached proton (eg. carbonyl

carbons) are done by the HMBC correlation experiment. Here, cross peaks between the carbons and protons that are separated by 2 or 3 bonds are obtained. The 2D spectrum obtained here is more complex than the HSQC case due to large number of correlations seen. The complete HMBC spectrum of MoL containing 10 mM BME is shown in fig. 4.20.

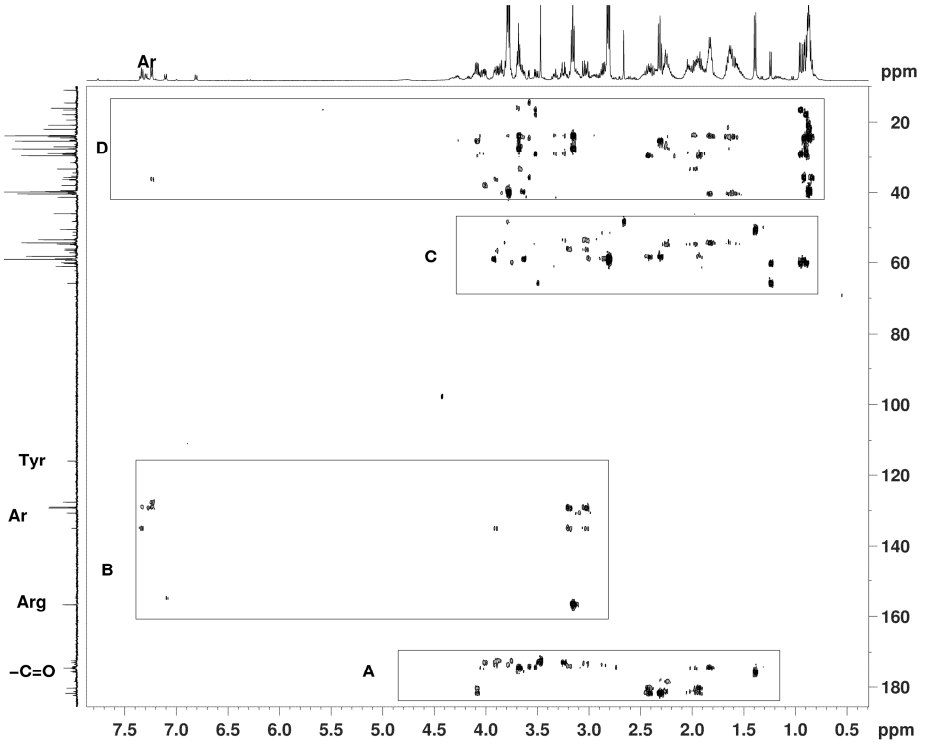

Figure 4.20. ^{1}H-^{13}C HMBC spectrum of MoL at pH 7.0 in presence of 10 mM β-ME. Different regions of interest are marked in boxes: (A) carbonyl to aliphatic proton correlation, (B) Aromatic carbons to proton correlation (C) alpha carbons to protons and (D) beta, gamma etc. carbons to protons.

Some of the correlations derived from the carbonyl regions are depicted in fig.4.21. The carbonyl peaks appearing below 180 ppm are most likely due to the

carboxyl carbons from the side chains. Some of the assignments are straight forward. For example, it is very easy to differentiate the resonances due to Asp and Glu as the COOH group of former one show only the correlations to its inequivalent βCH2 protons while the latter shows an additional correlation to its γ CH2 protons also.

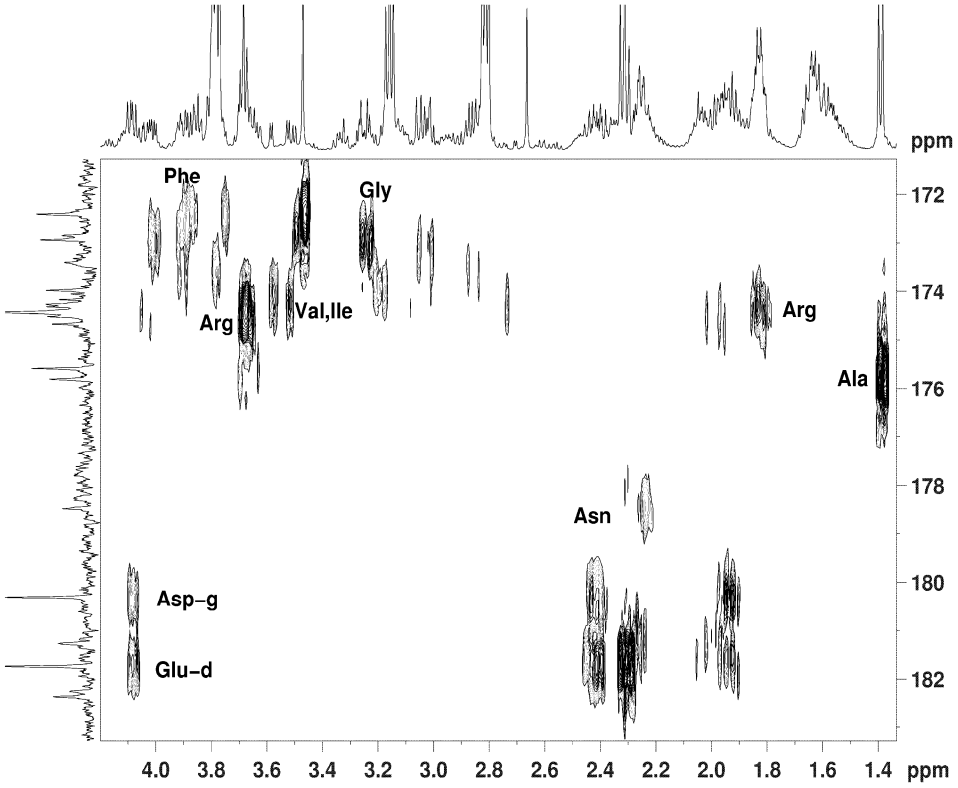

Figure 4.21. ^{1}H-^{13}C HMBC spectrum of the carbonyl region of MoL at pH 7.0 in presence of 10 mM β-ME. This connects the carbonyl carbons to the alpha and beta protons of the same residue. Some of the assignments are indicated.

Similarly, assignments of the protons of the aromatic residues (Phe and Tyr) and the Arg could also be made (Fig. 4.22). The guanidine carbon of the Arg residue appearing at ~157 ppm shows HMBC correlation to δ CH2 protons at ~3.25 ppm which

showed COSY cross peak to the γ CH2 protons at 1.95 and 1.90 ppm which in turn were connected to the β protons and the α proton by COSY/TOCSY connectivities as shown schematically in fig.4.23. Once the protons are identified, the ^{13}C Chemical shifts are extracted from the ^{1}H-^{13}C HSQC spectrum.

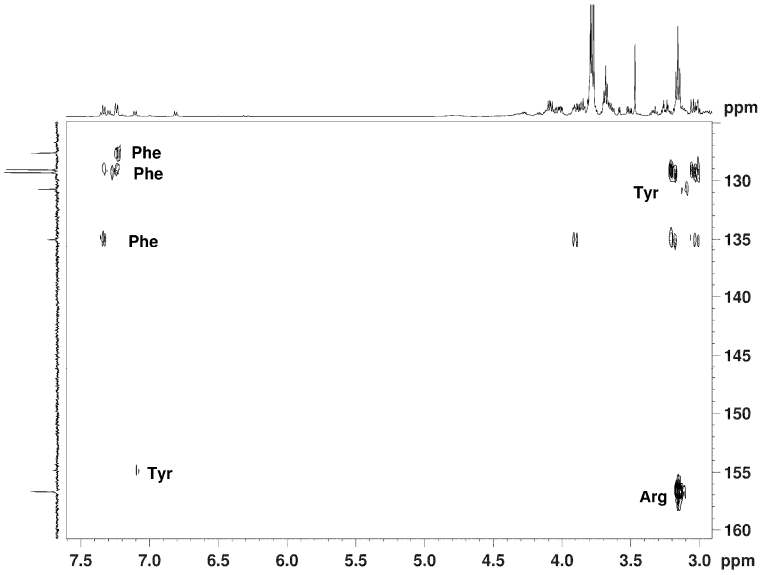

Figure 4.22. ^{1}H-$1^{3}C$ HMBC spectrum of the aromatic region of MoL at pH 7.0 in presence of 10 mM β-ME. This connects the carbons in the aromatic region to the alpha and beta protons of the same residue. Some of the assignments are indicated.

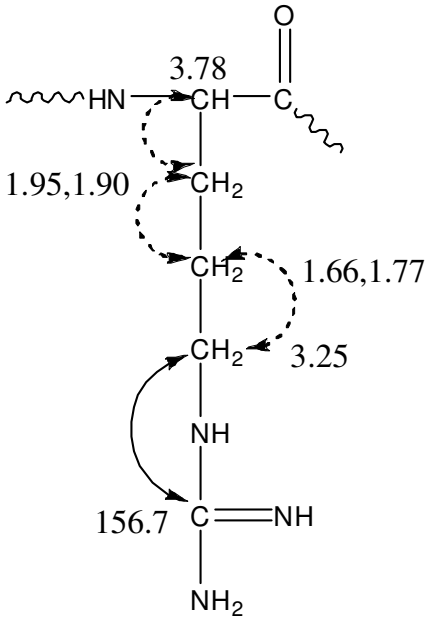

Figure. 4.23. Strategy used for the identification of individual amino acid residue as shown for Arginine.

The assignment of protons of Arginine residues are shown in fig. 4.24. Similar strategies have been used for identification of various residues present in MoL. The residues that could be assigned unambiguously are presented in Table 4.1 and 4.2.

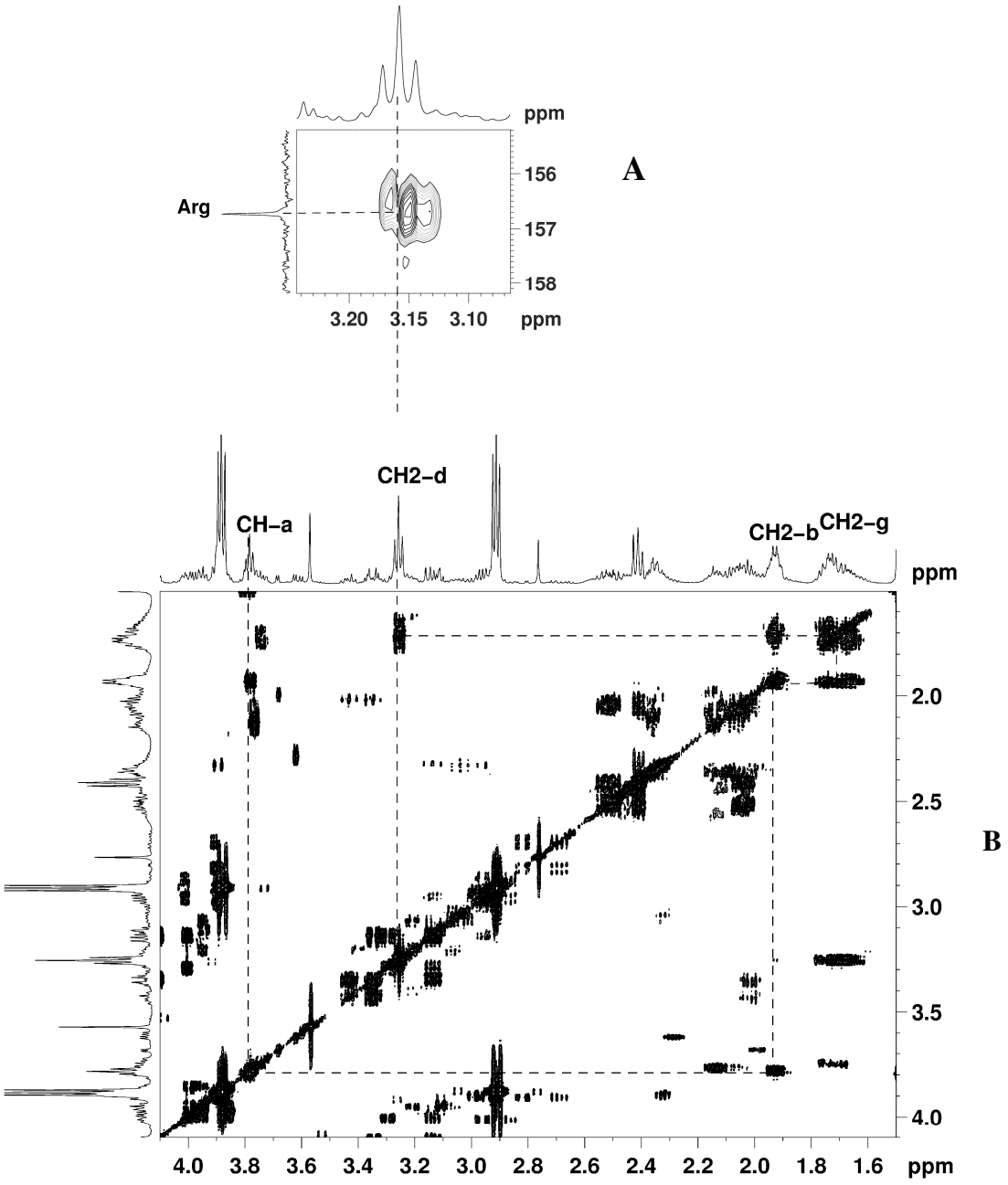

Figure 4.24. Assignment of Arg residue of the aromatic region of MoL at pH 7.0 in presence of 10 mM β-ME from ^{1}H-^{13}C HMBC (A) and COSY (B) spectra. Assignments are indicated in the Fig.

Table. 4.1. Proton chemical shifts for amino acids residues in MoL.

Type	Hα	Hβ	Other
Gly	3.42, 3.36	-	-
Ala	3.78	1.49	-
Val	3.62	2.29	1.05, 1.0 (γ)γ′
Ser	-	3.9, 4.0	-
Thr	3.59	4.12	1.34 (γ)
Cys	-	-	-
Asp	4.20	2.52, 2.05	-
Asn	4.01	2.34, 2.06	-
Glu	4.10	2.41	2.53, 2.05 (γ)
Gln	-	-	-
Ile	3.68	1.98	1.0 (methyl) 1.26, 1.48(γ), 0.94 (δ) (methyl)
Leu	3.75	1.70-1.63	1.63 (γ), γ′ 0.98-0.9 (δ)
Lys	-	-	-
Arg	3.78	1.95-1.90	1.66 ,1.77 (γ), 3.25 (δ)
Met	-	-	-
Pro	4.12	2.40-2.30	1.95-2.0 (γ), 3.42-3.32 (δ)
Phe	3.94	3.22, 3.08	7.34 (o), 7.44 (m), 7.39 (p)
Tyr	3.89	3.14, 3.02	7.20 (o), 6.91 (m)
Trp	-	-	-
His	-	-	-

Table 4.2. Carbon chemical shifts for amino acids residues in MoL.

Type	C=O	Cα	Cβ	Other
Gly	-	41.5	-	-
Ala	175.5	50.5	-	-
Val	174	60.4	29.0	17, 14 (γ)
Ser	-	-	60.2	-
Thr	-	60.3	65.9	19.4 (γ)
Cys	-	-	-	-
Asp	-	54.3	25.0	180.2 (COOH)
Asn	-	53.3	33.8	178.5 (CONH2)
Glu	-	58.3	25.3	29.7 (γ), 181.7 (COOH)
Gln	-	-	-	-
Ile	174.4	59.5	35.8	16.6 (methyl), 124.4 (γ), 11.1 (methyl)
Leu	175.8	53.4	39.7	24.1 (γ), 21, 22 (δ)
Lys	-	-	-	-
Arg	174.1	50.5	27.5	23.9 γ (γ), 40.5 (δ)σ, 157.7(guanidine)
Met	-	-	-	-
Pro	-	61.2	29.3	25.2 (γ), γ 46.2 σ (δ)
Phe	174.1	60.5	38.2	129.3(o), 129.1(m), 127.7(p), 135.1
Tyr	-	56.2	35.6	130.8(o), 115.8(m),126.7, 154.9
Trp	-	-	-	-
His	-	-	-	-

The ^{1}H NMR spectrum of MoL containing 10 mM BME shows a number of signals coming from the various amide NH groups. Hence it was not very easy to get finer details of the weaker aromatic protons. Hence, the sample was lyophilized with D_2O a few times to exchange all the exchangeable protons. The ^{1}H NMR spectrum thus obtained is compared with the unexchanged one (Fig. 4.25). These spectra look very similar except for the absence of exchangeable protons in D_2O.

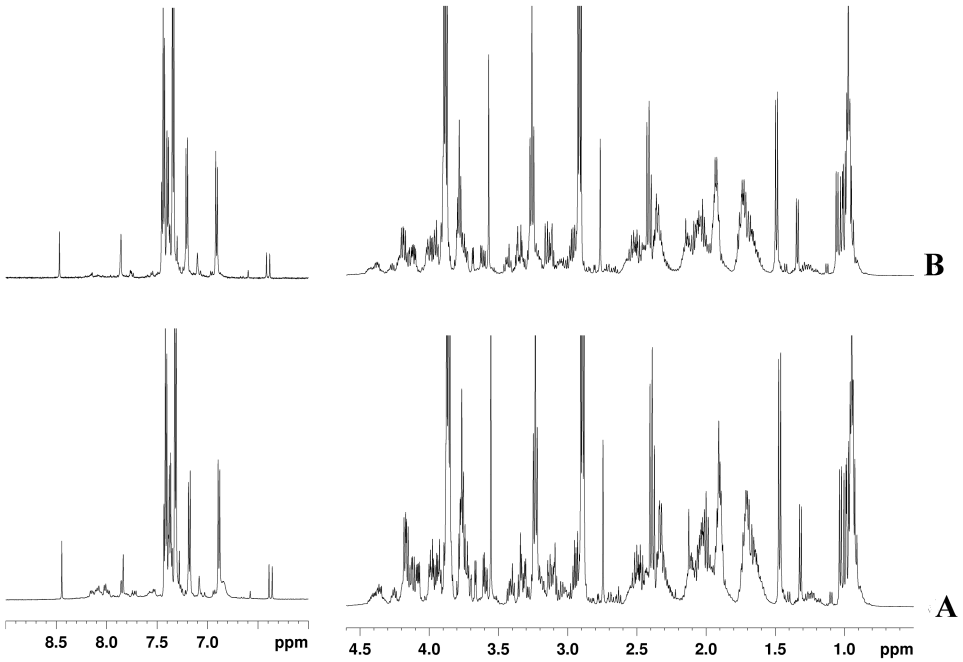

Figure 4.25. Comparison of the ^{1}H spectrum of MoL at pH 7.0 in presence of 10 mM β-ME in 90% H2O + 10% D2O (A) and in 99 % D2O (B). The disappearances of NH peaks are noticeable

A rough estimate of the various residues can also be obtained from the ^{1}H spectrum of MoL containing BME in D_2O. The integrated values of various ^{1}H signals

are given in figs.4.26 & 4.27. Here, the ^{1}H signal of the Tyr residue at 6.91δ (2, protons) is taken as standard.

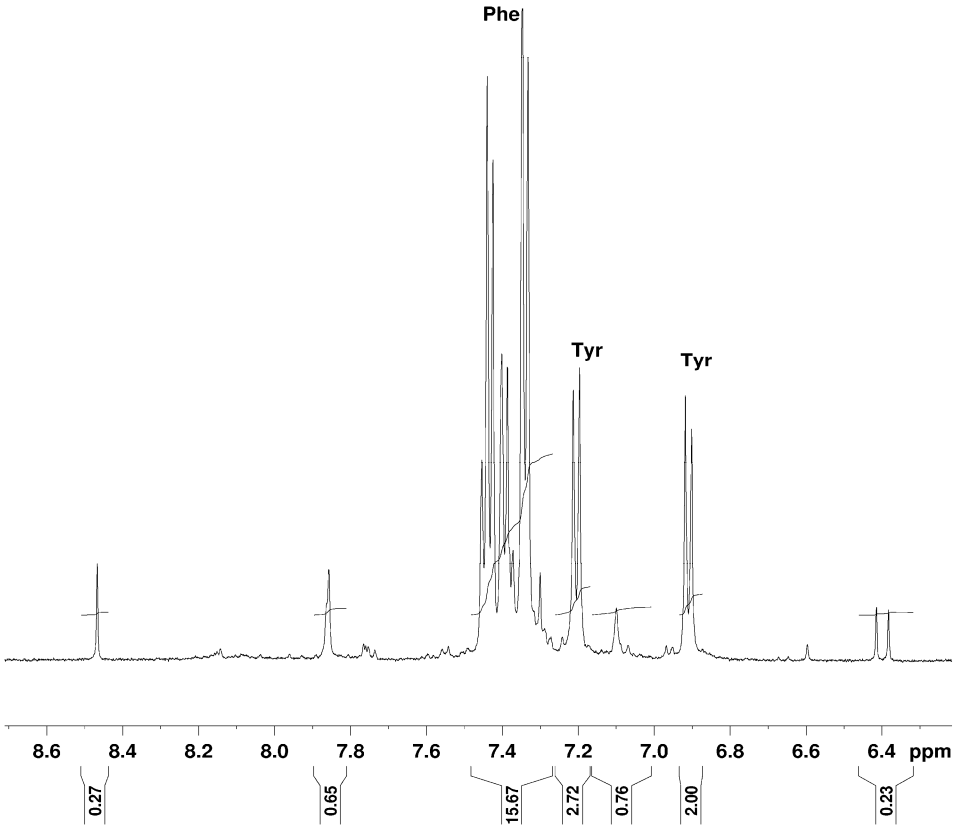

Figure 4.26. ^{1}H spectrum of the aromatic MoL at pH 7.0 in presence of 10 mM β-ME and D_2O. Integration of various protons is indicated.

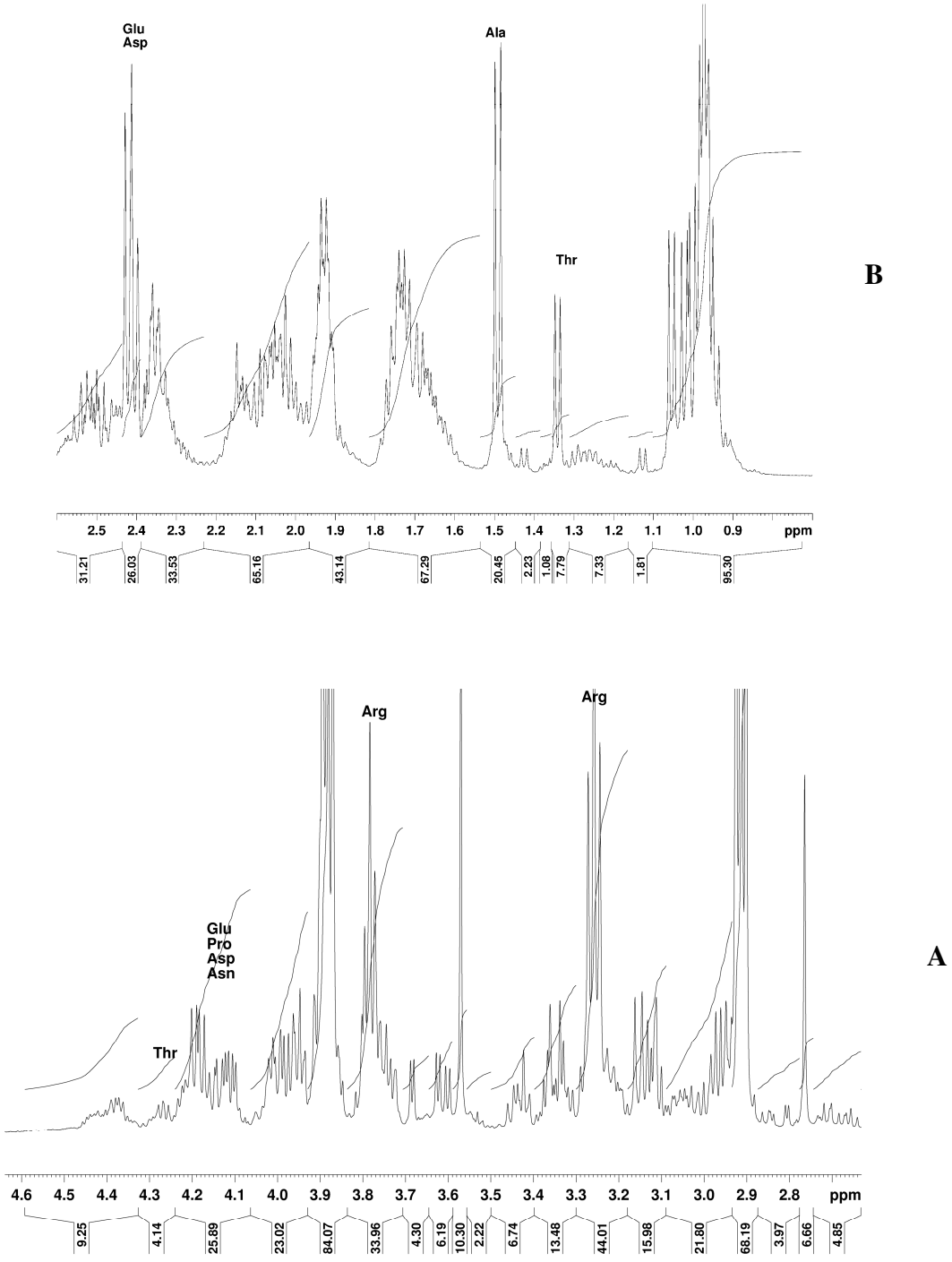

Figure 4.27. ^{1}H spectrum of aliphatic regions of MoL at pH 7.0 in presence of 10 mM β-ME & D_2O. Integration of various protons is indicated. (A) 4.6 to 2.6 ppm region and (B) 2.6 to 0.7 ppm region.

From the area of the various signals it is very clear that MoL is very rich in Arg, Glu/Asp/Asn and leu/Ile/Val. Table 4.3 shows some of the estimations calculated from the ^{1}H spectrum. The estimated values of some these residues were found to be more than that was reported prviously (Katre UV, 2007) as shown in table 4.4 for comparison. One of the reasons for this is overlap of signals from other residues in the ^{1}H spectrum.

Table. 4.3. Estimation of amino acid composition of MoL inferred from ^{1}H NMR spectrum.

Amino acid	Number of residues per monomer of MoL	Chemical shift
Tyr	1	6.91
Phe	2	7.46 to7.32
Arg	~15	3.25
Glu/Asp/Asn	~13	2.42
Ala	~6	1.4
Leu/Val/Ile	~15	1 to 0.9
Val	3	-
Ile	3-4	-
Leu	8-9	-

Table 4.4. Amino acid composition of MoL (Reproduced from Katre UV, Ph.D thesis, 2007).

Amino acid	No. of residues per monomer
Glutamic acid and Glutamine	15
Arginine	11
Proline	6
Glycine	6
Alanine	5
Leucine	5
Threonine	4
Valine	3
Isoleucine	3
Cysteine	3
Aspartic acid and Asparagine	2
Serine	2
Phenylalanine	2
Histidine	2
Tyrosine	1
Methionine	1
Tryptophan	1
Lysine	0
Total	**72**

The α, β, γ and δ protons of arginine in MoL were located at 3.78, 1.95-1.90 1.66-1.77 and 3.25 ppm, respectively. Out of these the signal at 3.25 ppm (2 protons) has relatively less overlap and give an integral value of ~30 which corresponds to 15 residues. There were at least six alanines as located in the ^{1}H NMR spectrum (corresponding to β-ME at ~ 1.4 ppm (integral value of ~20). However according to the

earlier report it is five only. A doublet peak around 1.34 ppm coming from the γ CH3 with molar ratio of 7.77 indicated presence of at least two threonines in the protein. The Hydrphophic amino acids of MoL showing multiple signals in the region 1.0 to 0.9 ppm showed an integral value of ~90 and thus summing up together for 15 residues (Ile, Val, Leu) (each of these residue contains two methyl groups and thus the integral value is divided by 6). However, to our surprise the signature of the single tryptophan in the ^{1}H NMR spectrum of MoL could not be ascertained. The weak peak around 10 ppm could be seen in the native spectrum of MoL coming from the NH of tryptophan (Fig. 4.5) but that could not be traced in the spectrum of MoL upon treatment with β-ME. Presence of very weak coupled spins corresponding to the expected positions for Trp (~7.6 to 7.7 ppm) in the aromatic region could be detected in the TOCSY and COSY spectrum (Fig.4.28). However the signal intensity seems to be very weak. The observation further suggests that the chemical environment of tryptophan upon unfolding has changed as also seen in the CD and fluorescence data of MoL treated with β-ME.

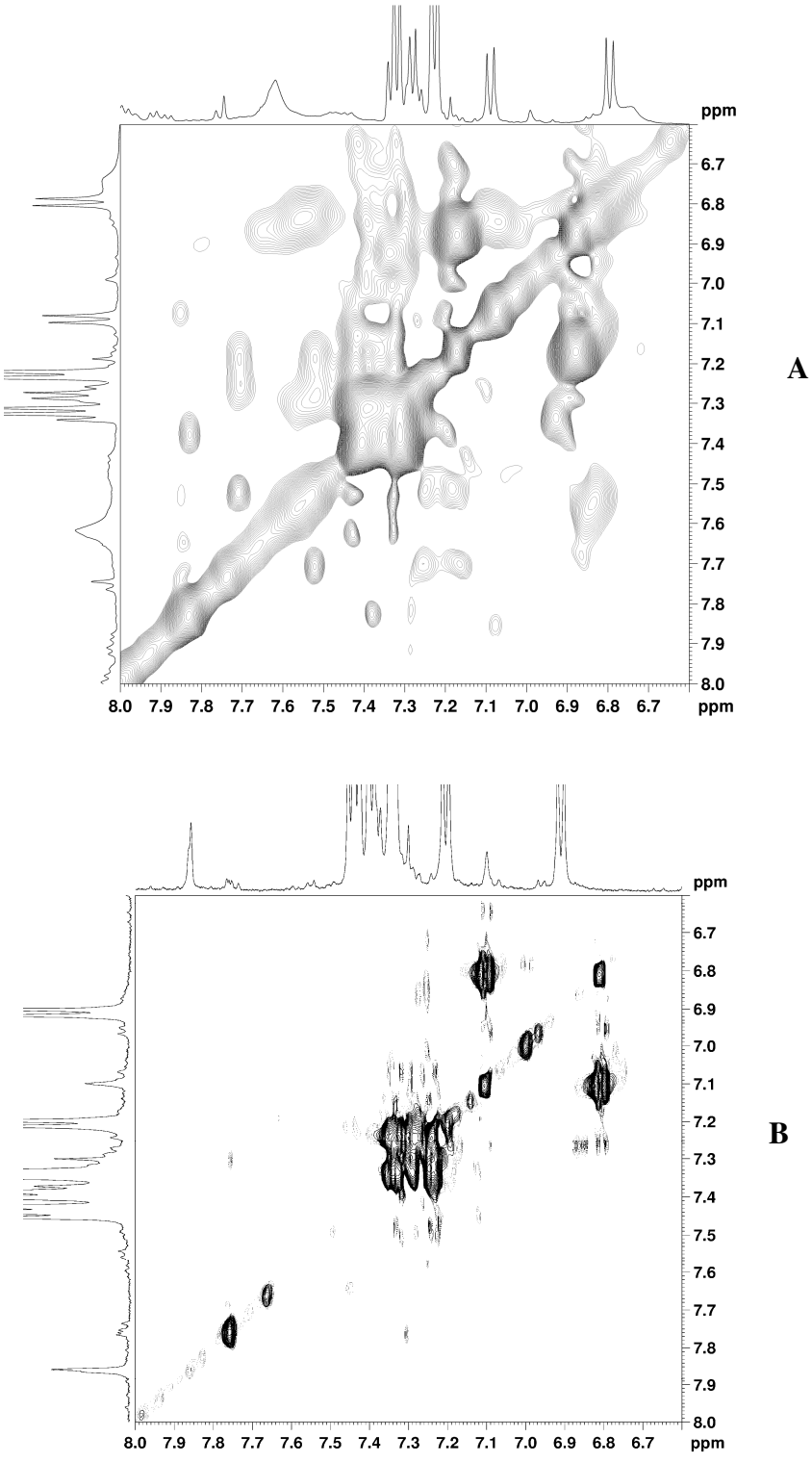

Figure 4.28. TOCSY spectrum of the aromatic region of MoL at pH 7.0 in presence of 10 mM β-ME in H2O + D2O (A) and in D2O (B). Weak cross peaks seen for the peak at ~7.8 to 7.7 ppm may be coming from Trp residue.

4.4.3. Thermal Stability of MoL

MoL is a thermostable protein which shows the hemagglutinating activity even after incubation at 85°C for 30 min and major retention of the secondary structure up to 90°C (Katre *et al.*, 2008a & b). Differential scanning calorimetric measurements on MoL were carried out to study the conformational changes and energetics of pH dependent thermal denaturation. The heat capacity function of MoL measured at pH 2.0 and 10.0 is shown in fig. 4.29 A and B. The transition was irreversible as demonstrated by the repeated scans of the same sample. MoL showed unusual stability at pH 7.0 as no transition was observed in the DSC measurements upto temperature 130 ^{o}C. At pH 2.0 and pH 10.0, MoL showed a high thermal midpoint of transition of 89 and 86 ^{o}C with ΔH_{cal} value of 5.66 and 1.4 Kcal, respectively. The ratio of ΔH_{VH}/ ΔH_{cal} for thermal unfolding of MoL at pH 2.0 and 10.0 is 0.36 and 35.68, respectively, indicating the anomalous behavior of the protein. The sharpening of DSC peak at pH10 (Fig.4.29B) is suggestive of irreversible/aggregation effects which can happen in the cases when $\Delta H_{cal} < \Delta H_{VH}$ (Cooper A, 1993). Furthermore, at pH 10.0 the irreversible DSC transitions of MoL closely resembles a two-state irreversible model, which assumes that only the native and the final irreversible-denatured states are significantly populated during the DSC scan and that the conversion from the native to the final state is described by a first-order rate equation (Sanchez-Ruiz 1992). Such kind of models are assumed to be limiting case of more complex situations and which are obtained when the irreversible alteration is very fast and the population of states other than native and final becomes negligible. If a DSC transition follows the two-state irreversible model, equilibrium thermodynamics analysis is not possible and information about the equilibrium denaturation mechanism cannot be

obtained (Ibarra-Molero and Sanchez-Ruiz, 2006) In DSC measurements the transition state of MoL was found to be concentration dependent as at the lower concentration i.e. at 0.5-3 mg/ml no specific transition was observed. However, when the protein concentration of 5 mg/ml was used a sharp transition was obtained at pH 2 and 10 (Fig.4.29 A and B).

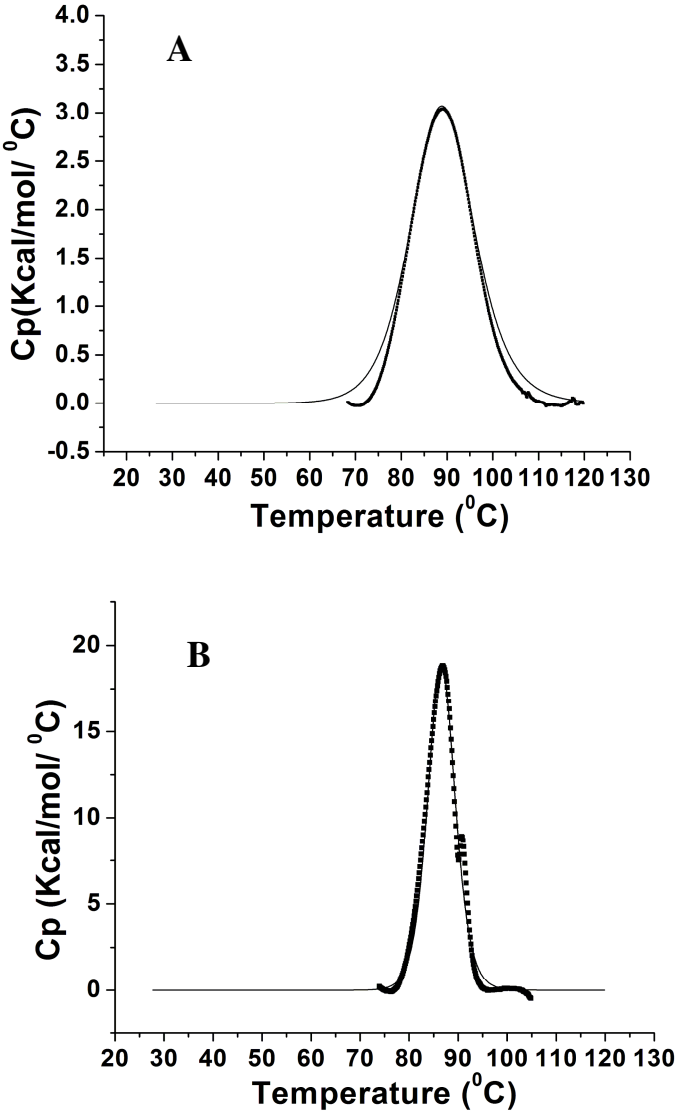

Figure 4.29. DSC studies on MoL at pH 2.0 (A) and 10.0 (B). The protein concentration used was 5 mg/ml. The data points are shown as dotted squares and the solid line represents the best fit of data to two state model.

4.5. Conclusions

To summarize, the spectroscopic studies on MoL indicated that reduction of disulphide linkages in MoL not only inactivated the hemagglutinin but also caused significant conformational changes with special reference to aromatic, hydrophobic and methyl peaks as also observed in the NMR spectrum, leading to the partially unfolded state. The breakage of disulphide bridges of MoL results into partially unfolded state with distinct structural property and enhanced stability as also reported for the α-lactalbumin molten globule (Redfield *et. al.* 1999). The signature of the single tryptophan in MoL could not be obtained upon treatment with β-ME indicating the changed chemical environment surrounding the residue. From the NMR studies on MoL, 15 arginines could be identified in the protein which was slightly more than identified earlier (11). Upon treatment β-ME MoL showed clear sharpening of the signals coming from hydrophobic amino acid residues. Altogether, 15 of them (Ile, Val, Leu) was identified from the ^{1}H NMR spectrum. MoL is unusually thermostable as no transition was observed in the DSC scan under native condition. Thermal unfolding of MoL at pH 10.0 could be fitted to a two state model with irreversible mode of transitions as also reflected by unusually high value of $\Delta H_{VH}/ \Delta H_{cal}$.

Chapter 5

Structural study of single chain recombinant jacalin with reduced sugar affinity

5.1. Summary

Jacalin is a two chain tetrameric glycoprotein with a molecular weight of 66 kDa comprising of a heavy chain (α) of 133 amino acid residues and a light chain (β) of 20 amino acid residues. Jacalin is the lectin in which the β-prism I fold, characteristic of the Moraceae plant lectin family, was observed for the first time. The crystal structure of native Jacalin at 2.4 Å has shown that the protein posses a three-fold symmetric β-prism fold made up of three 4-stranded antiparallel β-sheets. The single chain recombinant Jacalin (rjacalin) was prepared to study the role of post translational modification in generating carbohydrate specificity elicited by the two chain native Jacalin (nJacalin). The purified recombinant protein at a concentration of 30 mg/ml was crystallized using the hanging-drop vapour-diffusion method. The crystals of rjacalin belonged to the space group C2 having unit cell dimensions a = 118.9, b = 42.3, c = 73.6 Å, β = 122.3°. The diffraction data were 94.1 % complete at 2.0 Å resolution with R_{merge} = 0.072. The structure of recombinant jacalin was determined using the molecular replacement module AMORE of CCP4 suit and coordinates of native Jacalin (PDB ID-1JAC) used as model. The structure of rjacalin turned out to be similar to the njacalin. Interestingly, the rjacalin structure has no sugar bound and no electron density for the linker peptide.

5.2. Introduction

Jackfruit (genus *Artocarpus*) is a plant belonging to moraceae family which grows abundantly in tropical climates, its fruit and seeds are edible. The seeds of jackfruit contain mixture of lectins which agglutinates erythrocytes from several species. Two such lectins isolated from the seeds of *Artocarpus integrifolia* (jackfruit) differing in sugar specificity are: Jacalin (Chatterjee *et al*., 1979) and Artocarpin (Suresh *et al*., 1997). Jacalin was the first lectin belonging to Moraceae family to be isolated (Kumar *et al.*, 1982), sequenced and crystallized. The three-dimensional structure of Jacalin showed the three-fold symmetric β-prism I fold observed for the first time in plant lectins (Sankaranarayanan *et al.*, 1996). Until recently, it was believed that Jacalin-related lectins (JRLs) were found only in a few genera of the family Moraceae to which the jackfruit plant (*Artocarpus integrifolia*) belongs. However, in the last few years JRLs are reported from several taxonomically unrelated plant families as well, namely, Moraceae (Jacalin, MPA, *Artocarpus hirsuta* lectin, Artocarpin), Convolvulaceae (calsepa, conarva), Asteraceae (heltuba), Gramineae (barley and wheat lectins) and Musaceae (banana lectin)

The crystal structures of several other lectins in this family were determined subsequently namely, artocarpin from *Artocarpus integrifolia* (Pratap *et al*., 2002), hirsuta lectin from *Artocarpus hirsuta* (Rao *et al*., 2004*)*, heltuba from *Helianthus tuberosus* ((Bourne *et al*., 1999), MPA from *Maclura pomifera* (Lee *et al*., 1998) and calsepa from *Calystedia sepium* (Bourne *et al.*, 2004). At present 12 crystal structures of jacalin are available in PDB, crystals grown under different conditions. The crystals belong to different space groups and the molecules in them are complexed with a variety of sugar ligands (table 5.1). Table 5.1 also lists all the structures reported in PDB for the

JRLs belonging to different plant families in their apo as well as sugar bound forms. The three-dimensional structures of all these JRLs show the characteristic β-prism I fold, which is defined by a 3-fold symmetric β-prism made of three four stranded β-sheets, which are arranged like the three faces of a prism.

The strands of β-prism fold are almost parallel to the 3-fold axis. Each monomer has a single carbohydrate binding site. Most of the lectins of this family form tetramers with 222 symmetry as their functional biological units and hence are tetravalent. An exception to the tetrameric assembly was observed in case of *Helianthus tuberosus* lectin (Heltuba) where eight monomers are assembled as a donut-shaped octamer with eight solvent-exposed carbohydrate-binding sites (Bourne *et al.*, 1999) and Calsepa (lectin from *Calystegia sepium*) which exhibits a novel dimeric assembly that mimics the canonical 12-stranded β-sandwich dimer typically found in legume lectins (Bourne *et al.*, 2004).

Table.5.1. JRLs of known structure belonging to different plant families and specificity groups.

JRLs	Plant family	Source	Sugar specificity	PDB-ID (For native and structure in complex with sugars)
Jacalin *(Artocarpus Integrifolia)*	Moraceae	Seed	Gal/GalNAc	1KU8, 1JAC, 1KUJ, 1M26, 1UGW, 1UH0, 1UGX, 1UH1, 1UGY,1WS4, 1WS5, 1PXD
Hirsuta (*Artocarpus Hirsuta*)	Moraceae	Seed	Gal/GalNAc	1TOQ, 1TP8
MPA (*Maclura pomifera* agglutinin)	Moraceae	Seed	Gal/GalNAc	1JOT
Artocarpin *(Artocarpus Integrifolia)*	Moraceae	Seed	Mannose and derivatives	1J4S, 1J4T, 1J4U,1VBO,1VBP
Heltuba *(Helianthus tuberosus)*	Asteraceae	Tuber	Mannose and derivatives	1C3K, 1C3M, 1C3N
Calsepa *(Calystegia sepium)*	Convolvulaceae	Rhizome	Mannose and derivatives	1OUW
Morniga M *(Morus nigra)*	Moraceae	Bark	Mannose	1XXQ, 1XXR
Banana Lectin *(Musa paradisiacal* and *Musa acuminate)*	Musaceae	Pulp	Mannose and its derivatives	2BMY, 2BMZ, 2BN0, 1X1V, 3MIT, 3MIU, 3MIV
Parkia lectin *(Parkia platycephala)*	Fabaceae	Seed	Mannose	1ZGR, 1ZGS

Recent structural studies have pointed to a marked evolutionary plasticity within the family of the JRLs. Although all members of this still expanding lectin family share high sequence similarity and are built up of subunits with a similar overall architecture, they differ from each other with respect to their carbohydrate-binding specificities, molecular structure of the protomers, and subcellular location. Accordingly, the JRLs are now subdivided into two subfamilies, the galactose- and mannose-specific JRLs (Barre *et al.,* 2001). The recent identification of a JRL in a true fern further contributes to the widespread distribution of this lectin family in the plant kingdom (Tateno *et al.*, 2003). The galactose-specific jacalin-related lectins (gJRLs) are a small homogeneous group of galactose/T-antigen-binding agglutinins occurring exclusively in the Moraceae plant family. Jacalin, the first member identified in this family, follows the secretory pathway and gets accumulated in storage protein vacuoles (Peumans *et al.*, 2000). All currently known gJRLs closely resemble jacalin that can be considered as a prototype of this subfamily. Jacalin is synthesized as a pre-prolectin, consisting of a 21 amino acid signal sequence, 39 amino acid pro peptide, 20 amino acid β-chain, a linker peptide "Thr-Ser-Ser-Asn" and 133 amino acid α-chain, which has the sugar binding site. In mature jacalin, the signal sequence (21 aa) and pro-peptide (39 aa) are removed, probably through post- or co-translational processing as the mechanism is not yet apparent. The four amino acid linker peptide "T-S-S-N", which connects both β- and α- chains, is excised to generate two chains of the active form of jacalin. In addition to the processing, jacalin is partially N-glycosylated, probably at Asn 74 as supported by 1H-NMR data (Capon, 1990). This is the most likely site of glycosylation as the other two

potential sites at amino acid positions 16 and 35 were not glycosylated as per sequencing studies (Ruffet *et al*. 1992).

The mannose-specific jacalin-related lectins (mJRLs) are a growing group of lectins that occur in a wide range of species from different taxonomic groups and display an exclusive carbohydrate specificity toward mannose (Bourne *et al.*, 2004). Unlike jacalin, native mJRLs are built up of two, four, or eight protomers consisting of a single uncleaved polypeptide chain. They are synthesized and located in the cytoplasm and do not undergo co- or post-translational proteolytic modifications (Peumans *et al.*, 2000).

Jacalin is the single major protein representing more than 50% of the proteins of jackfruit crude seed extract (Kabir, 1998). Jacalin is a tetrameric glycoprotein containing approximately 7-10% carbohydrate (Aucouturier *et al*., 1987; Ahmed and Chatterjee, 1989; kabir, 1993; Ruffet *et al*. 1992), and exhibits specificity towards Galactose and its α- linked derivatives. It is reported to have a molecular mass of 65-66 kDa as determined by: (a) analytical ultracentrifugation studies (Ruffet *et al*. 1992), (b) gel filtration in presence of Me α-Gal and (c) non-denaturing PAGE at pH 4.3 (Banerjee *et al.*, 1991). Four cDNA clones sharing 94% sequence homology have been isolated from jackfruit seed cDNA library (Yang and Czapla, 1993). Jacalin is a tetrameric, two chain glycoprotein with a molecular weight of 66 kDa. As already mentioned the mature jacalin polypeptide consists of a heavy chain (α) of 133 amino acid residues and a light chain (β) of 20 amino acid residues. From the reported crystal structure of native jacalin at 2.4 Å resolution (Sankarnarayanan *et al.*, 1996) it has been shown that the protein possesses a three-fold symmetric β-prism fold. The jacalin subunit has approximate internal threefold

symmetry, but it binds only one sugar molecule. It is not clear till date which property (i.e., multimeric form/multi chain/posttranslational modifications) contribute to the sugar specificity and /or the affinity exhibited by the lectin. It has been already suggested that the newly generated amino terminus of the α-chain of jacalin due to posttranslational cleavage at glycine residue may be contributing to higher affinity for galactose (Sankarnarayanan *et al*., 1996). However, jacalin remains the most thoroughly studied β-prism I fold lectin in terms of structural investigations and thermodynamic measurements and has attracted much attention because of the several interesting biological property exhibited by the lectin (Jeyaprakash *et al.*, 2005; Goel *et al.*, 2004; Gupta *et al.*, 1992; Mahanta *et al.*, 1990). It selectively binds IgA1 and other glycoproteins such as carcinoma-related mucins. It induces mitogenic response selectively in CD41 T lymphocytes and also blocks infection by the human immunodeficiency virus type 1 (HIV-1) in T lymphoid cell lines (Pineau *et al.*, 1989,1990; Kabir, 1998; Bunn-Moreno and Campos-Neto 1981) Thus, jacalin also provides a valuable tool for AIDS research. Given the ability of lectins to specifically recognize the glycocode (carbohydrate code) on different cell surfaces and distinguish between diseased and normal tissues, these additional sites may be viewed as potential drug carrying sites that could be exploited for targeted delivery to sites of choice (Komath *et al*., 2006).

To investigate the binding site interactions and the role of posttranslational modifications which contributes towards generating specificity and recognition for galactose recombinant jacalin (rjacalin) was prepared as a single chain where the alpha and beta chain are linked by a tetra-peptide "TSSN" unlike in the njacalin which has two chains. While investigating its hemagglutination and hemagglutination inhibition profile

as compared to that of njacalin it was interesting to note that although rjacalin does not undergo any proteolytic processing in an *E. coli* environment, it is able to recognize galactose and methyl-α galactose. However, rjacalin has 100-fold less affinity for methyl-galactose (Ka: 2.48×10^2) in comparison to njacalin (Ka: 1.58×10^4) (Sahasrabudhe *et al*., 2004). It was also observed that rjacalin does not bind to mannose. To answer these differences in sugar binding and recognition profile of rjacalin, detailed structural studies by X-ray crystallography was started in our laboratory. Also, the rjacalin being unglycosylated and single chain will serve as an interesting model to study the sugar binding sites; interaction contributing towards sugar recognition and affinity compared to the njacalin as well as to that of single chain Artocarpin from the same family which is mannose specific.

5.3. Materials and Methods

5.3.1. Materials

Ammonium acetate, sodium acetate and sodium chloride were purchased from Merck (India). PEG 4000 and 8000 were procured from Sigma (USA). Resins for column chromatography including Mono Q and Uno-S were obtained from Biorad laboratories (USA). Rabbit erythrocytes were obtained from animal house maintained at NCCS, Pune, India. All the other chemicals used were of analytical grade and purchased from local suppliers.

5.3.2. Expression of recombinant jacalin clone

The gene of jacalin was cloned as a single chain (the α and β-chain linked by T-S-S-N loop) in the pT_7Nc vector between Nco I and EcoR I restriction sites, unlike in the native jacalin, to make a single chain protein in recombinant jacalin (Sahasrabudhe *et al*., 2004). For structural studies the rjacalin clone was received and expressed in *E.coli* BL21DE3 cells. Purification strategy employed by Sahasrabudhe *et al*., 2004 was slightly modified after several trials to increase the yield of the rjacalin upto 1 mg per litre of culture in the soluble form.

Briefly, 100 ng of plasmid DNA containing jacalin insert was transferred to *E.coli.* BL21DE3 cells, cells were heat shocked at 42°C and fresh LB was added to the cells and kept for expression for 45mins to1hr. 100 µl of transformed cells BL21DE3 were plated on LB-Amp Agar. Plates were checked for expressed colonies. Single colony of BL21DE3 (*E.coli* strain transformed with pT_7Nc) was inoculated in 10 ml LB-ampicillin and grown overnight at 37°C (200 rpm) and spinned at 3500 rpm, 10 mins, 25 °C. Pellet was resuspended in 5 ml LB and transferred to1litre LB-amp (30 µg/ml) and further grown at 28°C at 200 rpm till $O.D_{600}$ = 0.6-0.7. Cells were harvested by spinning at 4000 rpm, 4°C for 15mins. Pellet was redissolved in 20 ml lysis buffer (10 mM Tris, 5 mM EDTA, 200 mM NaCl) and was kept on ice for 30 mins, sonicated (10 pulses, amplitude-50 for 10 mins) and was spin at 10,000 rpm at 4 °C for 30 mins. Although more than 50 % protein was present in inclusion bodies only the supernatant was used for purification as it was difficult to refold the protein from inclusion bodies in active form. Prior to processing for purification, protein expression was checked by running 15 % SDS-PAGE.

5.3.3. Purification of recombinant jacalin

Purification of rjacalin was achieved by two successive steps of ion exchange column chromatography using anion exchanger resin mono Q sepharose and cation exchanger UNO-S-sepharose on a FPLC system. Final purification protocol was optimized for mono-Q step followed by a uno-S column step with continuous linear gradient of NaCl where rjacalin eluted out as 1st single peak in the NaCl concentration range of 90-150 mM. Elutes were checked for hemagglutination activity using 3 % rabbit erythrocytes and active fractions were resolved on 15 % SDS-PAGE. The active fractions showing single band were pooled and dialyzed against acetate buffer pH 5.2 and concentrated to 30 mg/ml using centrivap concentrator in freezed state for crystallization trials. Amount of protein present in purified fractions was estimated by Bradford method (Bradford, 1976).

5.3.4. Crystallization

Initial crystallization trials were conducted using Crystal Screen 1 and 2 supplied by Hampton research and also by using locally designed conditions for setting up experiments using hanging-drop vapor-diffusion method (Ducruix & Giege´, 1992; McPherson, 1982). The purified protein at a concentration of 30 mg/ml crystallized in condition numbers 10 and 28 of Crystal Screen 1 (Hampton Research) using the hanging-drop vapor-diffusion method with 2 µl of the precipitant solution mixed with 2 µl protein solution. Growth of crystals was further improved by varying the precipitant and its concentration. A single crystal of appropriate size was mounted in thin glass capillary (1.5mm) and used for data collection at room temperature (295 K).

5.3.5. Data collection, structure determination and validation

The diffraction data were collected on an R-Axis IV^{++} image plate using Cu-Kα radiation from Rigaku rotating anode X-ray generator. Crystal-to-detector distance was adjusted to 150 mm and each oscillation frame was recorded with 0.5^0 oscillation width. The diffraction data were processed and scaled using the DENZO and SCALEPACK modules of the HKL package (Otwinowski & Minor, 1997). The diffraction data was 94.1% complete at 2.0 Å resolution with R_{merge} = 0.07. The three-dimensional structure was determined using the molecular-replacement method implemented in the AMORE (Navaza, 1994) program from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Coordinates of native jacalin (PDB-ID:1JAC) were used for modelling.

Data in the resolution range 56-2.0 Å were used in the refinement of rjacalin structure. For the calculation of the free R factor 5 % of the reflections were kept aside. The initial phases obtained from molecular replacement were improved in the subsequent refinement and model building using the rjacalin sequence which resulted in the improvement of both the R factor and Rfree. After each run of the refinement program REFMAC, electron-density maps were calculated and the model was fitted onto a 2Fo-Fc map using the AUTOFIT module of QUANTA. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the geometry of the refined model. The non-crystallographic symmetry (NCS) present in the structures was used as a restraint in the initial stages of refinement. The X-SOLVATE module of QUANTA was used interactively to add solvent molecules initially at 5 sigma level followed by cautious additions of a few more at 3 sigma level. The solvent molecules in general were added

carefully considering the polar interactions and short contacts. The atomic coordinates and the structure factors of the refined structure of rjacalin have been deposited in the PDB. The assigned PDB code for the atomic coordinates is 3P8S.

5.4. Results and Discussion

Purified rjacalin showed the subunit molecular mass of 17 kDa on 15 % SDS-PAGE (Fig. 5.1 A) and was also confirmed to be a single chain (where the 20 aa β-chain and 133 aa α-chain are linked by the peptide "TSSN") as compared with njacalin shown in fig.5.1.B.

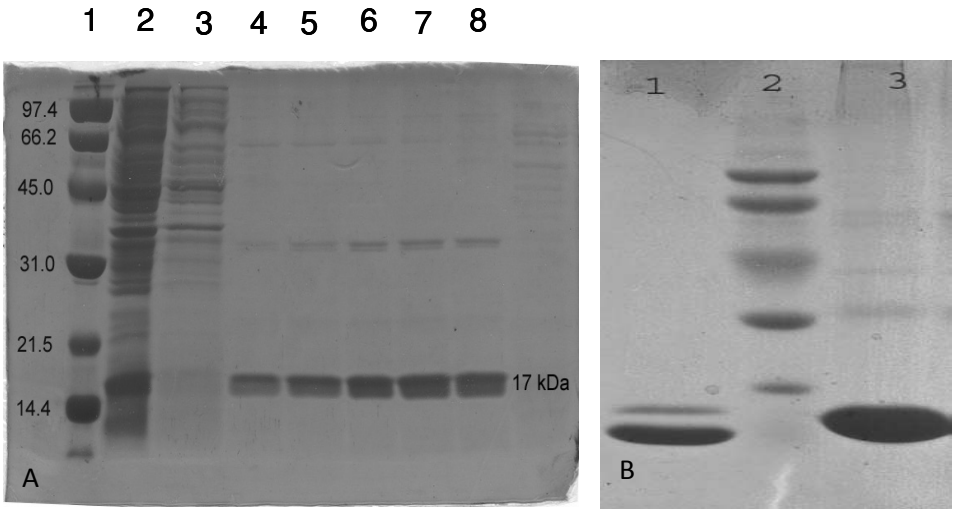

Figure 5.1. Purification profile of recombinant jacalin (A); Lane 1:Mol.wt. marker (Biorad) **Lane 2**:Sup after sonication, **Lane 3**: Flow through **Lane 4-8:** rjacalin peak fractions and **(B) Comparision of native vs recombinant jacalin; Lane-1**: Two chain njacalin, **Lane 2**: Molecular weight markers, **Lane 3**: rjacalin

Crystals of rjacalin grown in condition no. 10 (composition 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 4.6 and 30 % w/v PEG 4000) of the crystal screen 1(Hampton research, USA) diffracted to 2.0 Å resolution. The quality and size of the crystals were improved by substituting 10 % and 15 % w/v PEG 8000 for PEG 4000. The crystals appeared and grew to full size in 6 days. The crystals of rjacalin belonged to the space group C2 having unit cell dimensions a = 118.9, b = 42.3, c = 73.6 Å and β = 122.3°. The crystals diffracted to 2.0 Å resolution at home source at room temperature (295K). The estimated Mathews coefficient for rjacalin crystals was 2.3 corresponding to 46 % of the solvent content. Good quality crystals of rjacalin could be obtained when grown in presence of sugar. However, despite several co-crystallization attempts to complex the rjacalin with methyl-α galactose no sugar bound structure could be obtained.

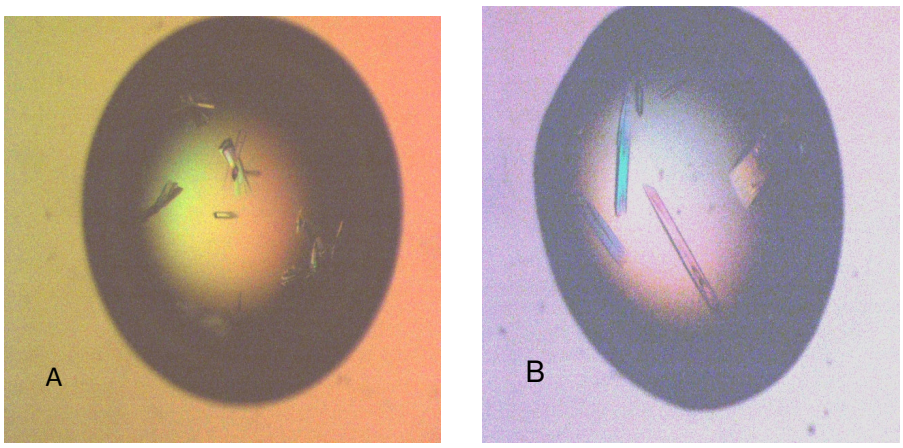

Figure 5.2. Crystals of recombinant jacalin grown with (A) PEG 4000 and (B) PEG 8000**.**

The data collection and refinement statistics of the rjacalin crystal used for structure calculation are listed in table 5.2. After several cycles of refinement, the Cc and

Rfactor values improved for the rjacalin model, and remained almost unchanged upon further refinement. In the final refined model of rjacalin, 2308 protein atoms and 88 solvent molecules could be located. PROCHECK showed that 86.7 % of residues were in the allowed regions of Ramachandran map. However, the electron density corresponding to the "TSSN" connecting the α- and β-chain was missing.

Table 5.2. Data collection and refinement statistics (Values in parentheses are for the last shell)

	Recombinant Jacalin
Space group	C2
Unit cell parameters (Å)	
A	118.9
B	42.3
C	73.6
β °	122.3
Z	4
Resolution range (Å)	50 - 2.0 (2.03 - 2.0)
No. of observations	68183
Completeness(%)	94.1(63.4)
R_{sym} %	7.2 (13.5)
Average I/σ(I)	11.64 (6.71)
Mathews coefficient	2.3
Solvent content (%)	46
No. of unique reflections	19935
Crystallographic R_{factor}	0.20
No. of reflections used for refinement	18912
R_{free}	0.26
No. of reflections for R_{free}	1023
No. of protein atoms	2304
No. of water molecules	88
Ramachandran Plot **Residues in the allowed regions (%)**	86.7
Residues in additionally allowed regions (%)	13.3
Overall average G factor	-0.11

The overall structure and the fold of recombinant jacalin appeared similar to that of native jacalin. However, in the asymmetric unit of rjacalin a dimer was present while in case of the njacalin there are four molecules in asymmetric unit containing two dimers belonging to two half tetramers. Biological assembly of rjacalin is predominated by tetramers like that of njacalin (Fig.5.3 A and B). The three β-sheets forming the faces of the prism fold into three Greek key motifs arranged around a pseudo-3-fold axis. The three Greek key motifs (GK1, 2 and 3) in the β-prism I fold of rjacalin are similar to that of njacalin. However, since rjacalin is single chain where the beta and alpha chain are linked together, all the three greek keys are part of Chain A in rjacalin monomer where the first (GK1) contains 46 residues (Val10-Asn24, Gly25-Phe33 and Leu136-Leu157), the second (GK2) contains 52 residues (Thr34-Pro85) and the third (GK3) contains 46 residues (Glu87-Ile132). The overlap of rjacalin on the methyl-α-galactose bound njacalin structure showed no change in the sugar binding site (Fig.5.3C). The root-mean square deviation between native and rjacalin is 0.31 Å for the c-alpha atoms. Moreover, the apo-jacalin (PDB ID-1KU8) which also has two chains and thus different from rjacalin showed rmsd of 0.28 Å for c-alpha atoms which is still lower as there is no sugar/ligand bound to the lectin as that of rjacalin.

Artocarpin (MW-65 kDa), the second lectin from jackfruit seeds, unlike njacalin, is non-glycosylated and specific for mannose at the monosaccharide level (Suresh *et al*., 1997). Artocarpin is single chain and differs from njacalin in carbohydrate specificity and biological action, although their overall structures are identical. Despite the low sequence identity of 58.1 % between rjacalin and Artocarpin their structures were significantly closer showing a root-mean square deviation of only 0.79 Å (Fig.5.5).

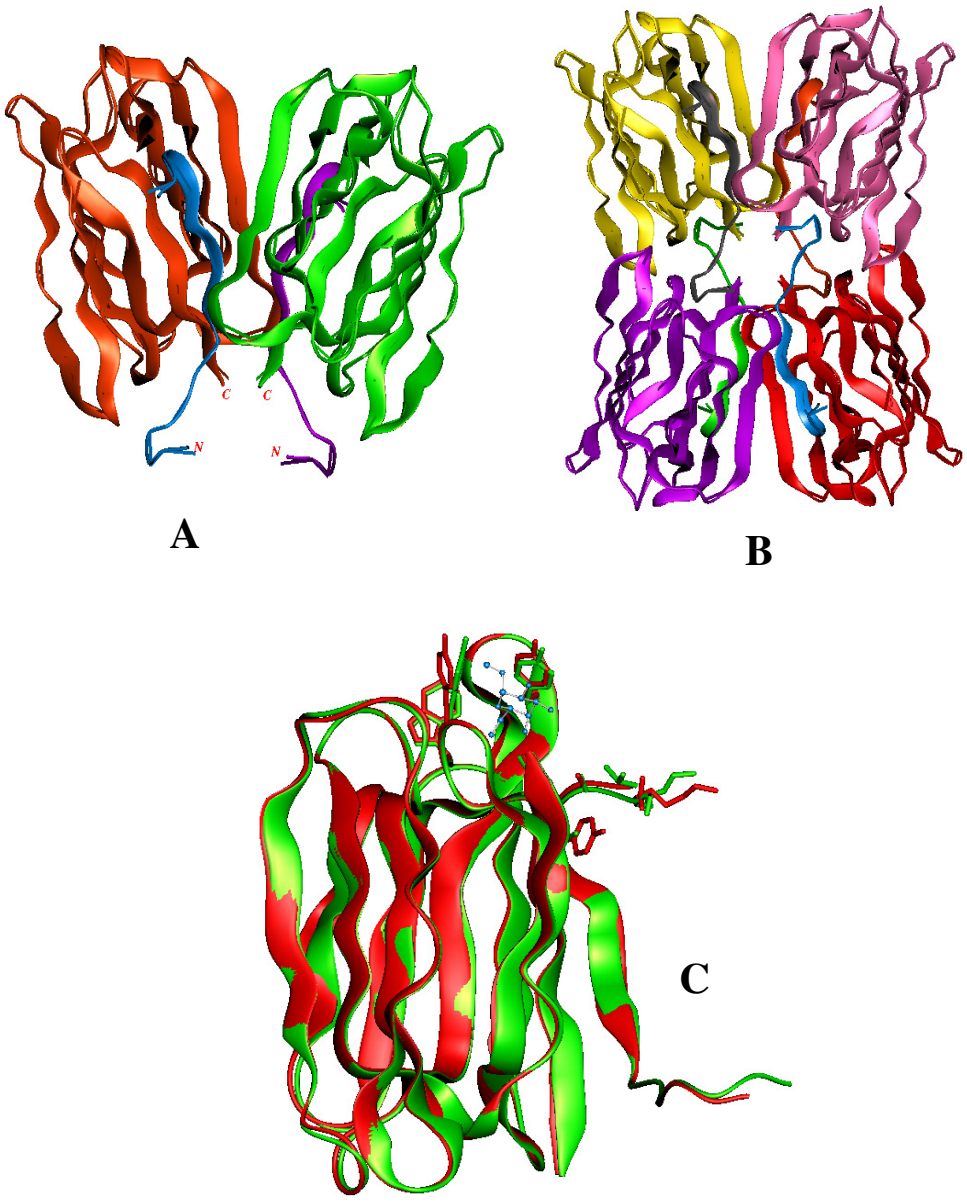

Figure 5.3. (A) Three-dimensional structure of Recombinant Jacalin dimer in the asymmetric unit

(B) Jacalin tetramer constructed by including the symmetry related dimer

(C) Overlap of native jacalin (red, PDB ID – 1JAC) with recombinant jacalin (green). Methyl-α-galactose bound to Jacalin is shown as ball and stick (in blue) (The rmsd value based on c-alpha atoms is 0.31Å).

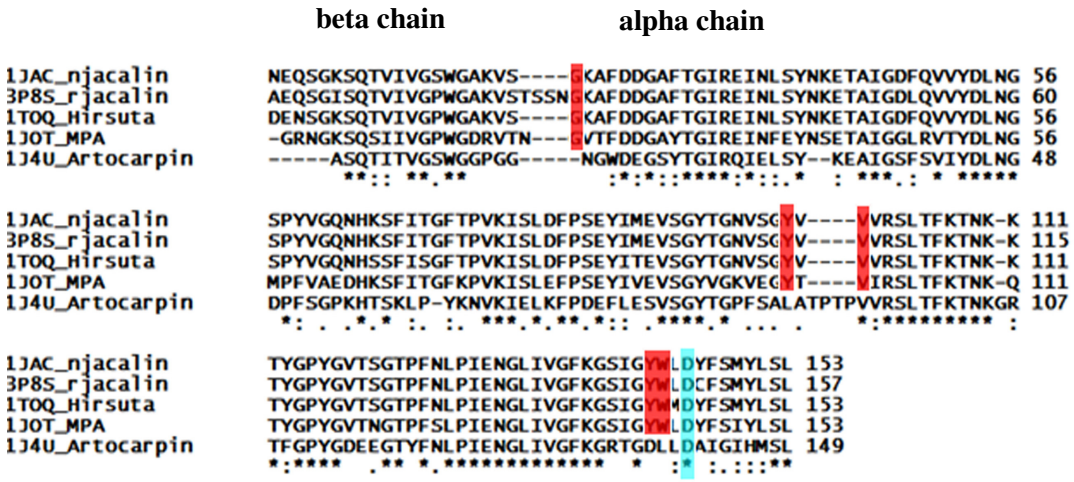

Figure 5. 4. Multiple sequence alignment (CLUSTAL W) of rjacalin with JRLs. The native Jacalin (PDB ID-1JAC), hirsuta lectin (PDB ID-1TOQ) and MPA (PDB ID-1JOT) are galactose specific two chain (containing β and α) lectins while single chain artocarpin (PDB ID 1J4U) is mannose specific. The symbols *, : and . corresponds to the conserved residues, substitution by a similar type of amino acid and substitution by a non-similar type of amino acid residues, respectively. The blocks highlighted in red and cyan shows conserved residues involved in sugar binding.

Figure 5.5. Overlap of single chain Artocarpin- mannose complex (blue, PDB ID – 1J4U) with rjacalin (green). The bound mannose is shown in stick representation (cyan) (The rmsd value for the overlap of c-alpha atoms is 0.79 Å).

As discussed in the introduction section rjacalin differs from njacalin being a single chain where β and α chains are joined by tetrapeptide "TSSN" the structure of rjacalin is identical with that of njacalin whereas there is a 100 folds reduction in the affinity of rjacalin towards galactose. Thus the similar structure and almost similar hemagglutination and hemagglutination inhibition profile of rjacalin with njacalin raises an interesting question; how does rjacalin, without processing, recognize Galactose and methyl α-galactose but not Mannose or Glucose?

The carbohydrate binding site of njacalin is formed by the loops that connect the inner strands of Greek keys 2 and 3. Residues that take part directly in the binding of the sugar molecule are Tyr78 and Val80 of Greek key GK3, and residues Gly1, Gly121, Tyr122, Trp123 and Asp125 of GK1. The atoms of these residues form the α-chain, are located within 4.0 Å of sugar binding site. These residues are well conserved for homologous Hirsuta lectin and MPA. The side chain of Tyr 78 stacks against the B-face of methyl α-galactose and makes van der Waals contact with four sugar carbon atoms (C1, C3, C4 and C5) (Sankarnarayan *et al.*, 1996). A close view of the sugar binding pocket of njacalin, apo-jacalin and rjacalin suggests that the Tyr 122 (Tyr 146 in rjacalin) and Tyr 78 (Tyr 102 in rjacalin) are very close to the sugar which might be creating clashes hindering the firm binding in case of rjacalin (Fig. 5.6). This shift in the spatial positioning of the two important tyrosines may lead to the decrease in the area of active site pocket available for sugar binding.

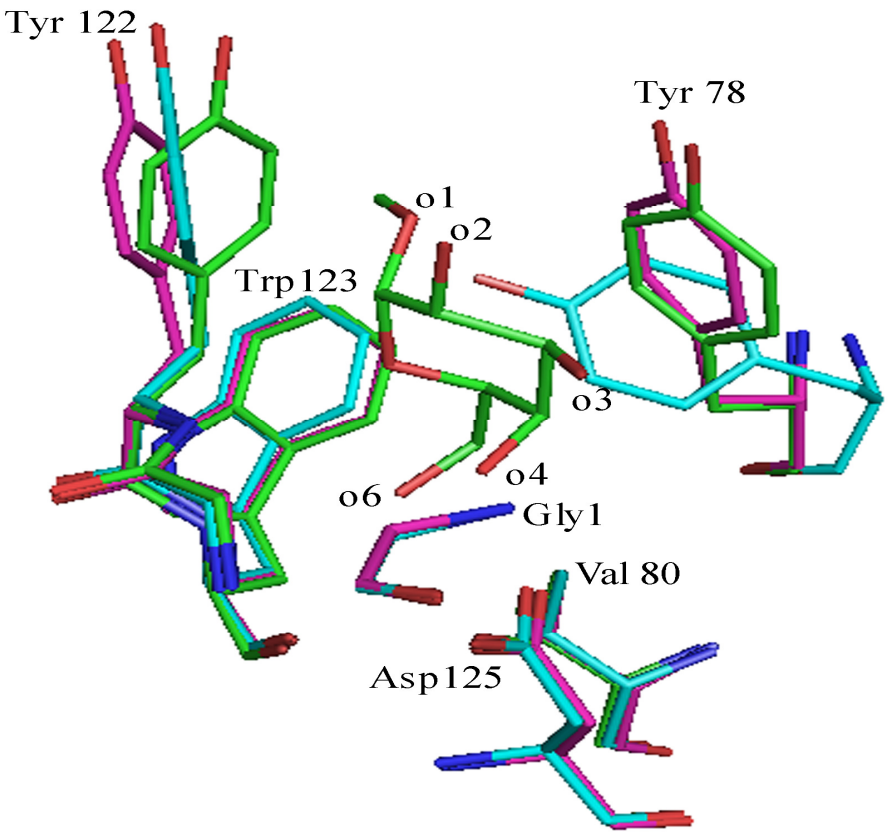

Figure 5.6. Interactions at the sugar binding site of Jacalin. Superposition of rjacalin (green) on njacalin (magenta) and apo-jacalin (cyan). Coordinates of methyl alpha galactose were taken from njacalin structure shown in green sticks in the centre. The diagram was prepared using PYMOL (Delano, 2002).

This observation was further supplemented by the results obtained from comparing active site pocket of rjacalin with njacalin (1JAC) and apo-jacalin (1KU8) using the CASTp server (Dundas *et al.,* 2006) available online. The active site pocket area and volume analysis of rjacalin in comparison with njacalin and apo-jacalin revealed significant reduction in the area and volume of the sugar binding pocket in rjacalin. The total surface area of the binding pocket in njacalin is 102.1 $Å^2$ which reduced to 231.2 $Å^2$ in the rjacalin associated with almost 2-fold reduction in volume of the binding site

pocket (Table. 5.3). To rule out the possibility that this size reduction could be due to absence of ligand; the active site pocket area of apo-jacalin is also analyzed. CASTp analysis of the apo-jacalin showed almost five fold increase in surface area and volume of binding pocket in the absence of bound sugar. The results obtained are summarized in table 5.3.

Table. 5.3. Comparison of the active site pocket of rjacalin with njacalin and apo-jacalin using CASTp server available online.

CASTp analysis	**njacalin (1JAC)**	**Apo-jacalin (1KU8)**	**rjacalin (3P8S)**
Active site surface area (Å^2)	102.1	29.5	231.2
Active site Volume (Å^3)	143	16.7	277.8
Residues involved	Gly1, Phe 47, Tyr 78, Val 80, Gly121, Tyr122, Trp 123, Asp125	Tyr 78, Val 80, Tyr 122, Trp 123, Asp 125	Ala 17, Val 19, Gly 25, Phe 71, Tyr 102, Val 103, Val 104, Ser 143, Gly 145, Tyr 146, Trp 147, Asp149, Cys 150

The CASTp analysis further suggests that in case of rjacalin due to the presence of "TSSN" loop the residues corresponding to beta chain of njacalin such as Ala 17 and Val 19 are also coming closer to the active site pocket. It has been proposed that the size and conformation of carbohydrate-binding site of jacalin makes the lectin a polyspecific one. As a result of proteolytic cleavage of protomer, the size of carbohydrate binding sites of jacalin and MPA is more extended than that of their mannose-specific

homologues with uncleaved protomers. This increased size explains the apparent lack of monosaccharide-binding capacity of carbohydrate binding sites of jacalin (Astoul *et. al.*, 2002). As said earlier due to highly dynamic nature of the four residue loop, electron density corresponding to the loop region is missing from the map although the rjacalin is single chain. Thus, rjacalin which is a single chain lectin containing the highly flexible TSSN loop is not able to bind galactose firmly, though it is able to recognize the sugar and showing hemagglutination inhibition. The possible explanation for this could be: (a) Probably due to the presence of the loop, amino terminus of Gly 1 in njacalin which plays important role in carbohydrate recognition is now (Gly 25 in rjacalin) not free to interact with sugar. (b) There could also be a possibility that highly dynamic loop at the entrance of active site pocket might be blocking entry of sugar as could be inferred from the reduction in interactive surface area and volume of active site pocket and (c) the single chain rjacalin, which has intact polypeptide chain due to the presence of loop, might have lost extended binding site as reflected in the changes in sugar binding and recognition mode of rjacalin compared to njacalin (Sahasrabudhe *et al*., 2004).

The strategies used by lectins to generate carbohydrate specificity include the extensive use of water bridges, post-translational modification and oligomerisation (Vijayan and Chandra, 1999). The crystal structures and detailed modelling studies reported earlier (Pratap *et al*., 2002) indicates two main differences between the carbohydrate binding sites of artocarpin and njacalin generating specificity for mannose and galactose respectively. Firstly, unlike njacalin, artocarpin does not contain a new N-terminus generated by posttranslational proteolysis which is also the case with rjacalin, still the later is able to bind galactose at the primary binding site with reduced affinity.

Secondly, there is no aromatic residue in the binding site of artocarpin whereas there are four in jacalin (Phe47, Tyr78, Tyr122 and Trp123) which are involved in stacking interactions with the carbohydrate. A comparison with similar lectins of known structures or sequences, suggests that, in general, stacking interactions with aromatic residues are important for the binding of galactose while such interactions are usually absent in the carbohydrate binding sites of mannose-specific lectins of the β-prism I fold (Pratap *et al*., 2002). Besides that, another newly identified feature that demarcate sugar specificity between galactose and mannose are: the hydrophobic lid formed by the three loop regions corresponding to residues 47–50, 75–78, and 121–125 in njacalin named as binding loop 1 (BL1), BL2, and BL3 and the presence of an influencing loop (IL) which can also influence the conformation of some of these residues and hence the specificity. This IL loop corresponds to residues 20–23 in jacalin and is found to be longer in the galactose-binding lectins than mannose-binding ones. In njacalin the residues from this loop interact with residues of BL1, BL2, and BL3 (Raval *et al*., 2004)

Simple modelling by the replacement of an axial hydroxyl by an equatorial hydroxyl at C4 (the configurations at C2 is of no consequence), showed that njacalin can easily accommodate a mannose in the binding site. Likewise, artocarpin can accommodate a galactose as well. No steric clash results when the configuration at C4 is changed. The change results only in the abolition of the hydrogen bond of O4 with Asp125 or 141 OD2 (Pratap *et al.*, 2002). In the case of njacalin, the change leads to the disruption of one more and crucially important, hydrogen bond of O4 with the terminal amino group generated by posttranslational modification. Thus, a preference of njacalin for galactose over mannose is understandable however; the observed degree of the

preference is unexplainable. In view of the above facts the absence of a second free N-terminus in case of rjacalin is not just sufficient to generate affinity towards mannose.

We have shown that the recombinant jacalin is a single chain by comparing the SDS-PAGE profile of rjacalin and njacalin (Fig.5.1B). In rjacalin the α and β chains are linked together by "T-S-S-N" loop whose electron density could not be observed in the electron density map presumably due to the high flexibility of the loop region. Mutations detected in the recombinant jacalin are not in the sugar binding site and hence the reduction in affinity could be due to either the absence of a second free N- terminus of α-chain or due to high dynamics of the "TSSN" loop. The highly dynamic loop may not be giving chance for entry of the sugar and subsequent firm binding. A possibility that the galactose binding is facilitated by the side chain of Asn of the tetrapeptide (T-S-S-N) in the place of the N- terminus of Gly1 is also proposed (Sahasrabudhe *et al*., 2004). This possibility is ruled out due to highly dynamic linker peptide. Hence, to conclude some selected point mutations that stabilize the loop region of rjacalin or mutating the N-terminal Glycine to Alanine and crystallographic studies of these mutants will provide a clear picture of sugar binding and recognition mode. Studies are underway in our laboratory to provide more evidence and further detailed insights into the mechanism of alteration of specificity/ affinity in the njacalin and the reduction in the affinity observed for rjacalin.

Chapter 6

Purification and preliminary X-ray crystallographic analysis of a trypsin inhibitor protein (CPTI) from *Cicer arietinum*.

6.1. Summary

Apart from CAL, a newly identified Kunitz type trypsin inhibitor protein (CPTI) was also purified from chickpea seeds and preliminary X-ray characterization of the CPTI was carried out. The CPTI showed a molecular mass of 18 kDa on SDS-PAGE. The IC50 value of CPTI determined was 2.5 µg against trypsin. The inhibitory activity of CPTI is 114 TIU/mg of protein which is quite high compared to those of other known Kunitz type trypsin inhibitors from legumes. CPTI crystallized in three different orthorhombic crystal forms, $P2_12_12A$, B and $P2_12_12_1$. The crystals of form $P2_12_12A$ with unit cell parameters a=37.2, b=41.2, c=104.6 Å diffracted up to 2.0 Å resolution at home source and 1.4 Å in beamline BM14, ESRF. The CPTI showed no homology to the known trypsin inhibitors of kunitz family so, the molecular replacement method could not yield a successful model. Preparation of heavy atom derivatives of CPTI crystals were carried out. Data collection statistics for crystals grown in presence of various heavy metal salts are reported in this chapter. The Matthews coefficient calculated was 2.37 $Å^3$ Da^{-1}, corresponding to a solvent content of 42 % for orthorhombic crystal form $P2_12_12A$. The diffraction quality of the other two crystal forms ($P2_12_12B$ and $P2_12_12_1$) was comparatively poor.

6.2. Introduction

Plant seeds are essential for human and animal food supply. Plant seeds provide approximately 80 % of food calories in human diet. Cereals, dry beans, oilseeds (including certain legumes), nuts, tree nuts, fruit and vegetable seeds, and other edible seeds are valuable sources of proteins and other nutrients for humans and animals. Depending on the seed source, growing conditions (geographic, seasonal variations, and climatic conditions), cultivar/hybrid, and the type of seed, the amount of protein in plant seeds varies considerably. A systematic study of seed proteins was initiated by Osborne (Osborne, 1924) who classified these proteins into four classes: albumins, globulins, prolamins (hordein) and glutelins according to their solubility in water, salt solutions, aqueous alcohol and basic and acid solutions respectively. Modern classification of plant proteins was proposed by Shewry & Casey (1999) (Shewry and Casey, 1999) based on protein function and molecular/biochemical relationships. The functional proteins present in seeds are further classified into three main classes. They are: Storage proteins which store nitrogen, carbon and sulfur; Structural and metabolic proteins that are essential for the growth and structure of the seed and defense or protective proteins involved in plant defense. In response to an external attack, plants generate various inter and intra cellular signals to activate genes for the induction of various substances such as antibiotics, alkaloids, terpenes, and proteins (Ryan, 1990). The defense proteins (peptides) that are involved in the defense against pathogens and invading organisms include ribosome-inactivating proteins, lectins, protease inhibitors and antifungal peptides and proteins (Kim *et al.*, 2009). In the last two decades pioneering work has been carried out to understand the mechanism of plant defense associated with the study of lectins and

protease inhibitors (Bowles, 1990; Ryan, 1990; Chrispeels and Raikhel, 1991; Valueva and Mosolov, 1999). This thesis has dealt with the structure-function studies of plant lectins described in the previous chapters. Apart from lectins, protease inhibitors have also been implicated in plant defense. Our laboratory has been involved in structural studies of bowman-birk inhibitor from plants (Rao and Suresh, 2007). As an extension to these studies we have also studied a Kunitz type trypsin inhibitor isolated from the seeds of chickpea which belongs to the family of legumes (Urvashi and Suresh, 2011). Protein protease inhibitors from plants potently inhibited the growth of a variety of pathogenic bacterial and fungal strains and therefore are excellent candidates for their use as lead compounds for the development of novel antimicrobial agents.

Proteinase inhibitors (PIs) are the most studied class of plant defense proteins. PIs in plants are also considered as anti-nutritional factors and implicated in various physiological functions such as regulation of proteolysis, deposition of storage proteins as well as defense molecules against plant pests and pathogens (Koiwa *et al.*, 1997; Ryan, 1990). They are usually present in the seeds and storage tissues, but are also expressed in the aerial parts of the plant upon insect attack. Their activity on gut proteases attenuates amino acid assimilation and slows down the growth of feeding insects. Owing to the direct role of PIs in plant defense, several transgenic plants have been produced expressing specific inhibitors and tested for enhanced resistance against phytophagous insects (Jouanin *et al.*, 1998). Most PIs interact with their target proteases, resulting in the formation of the respective protease-inhibitor complex, blocking, altering or preventing access to the enzyme active site. Some insects overcome the effect of PI ingestion by up-regulating the synthesis of new proteases and thus are insensitive to PIs (Jongsma *et al.*,

1995). In this respect, structural studies on PIs and protease-PI complexes are important for a better understanding of the molecular mechanism of protease resistance which can be utilized in designing new PIs active against insensitive proteases.

PIs are widely present in organisms. They have been isolated and characterized from a large number of organisms, including plants, animals and microorganisms (Valueva and Mosolov, 2004; Christeller and Laing, 2005; Supuran *et al.*, 2002; Mosolov and Valueva, 2005). PIs occur widely in the plant kingdom particularly in Leguminosae, Gramineae and Solanaceae families. Plant PIs (PPIs) are small proteins mainly occurring in storage tissues, such as tubers and seeds, but they have also been found in the aerial parts of plants (De Leo *et al.*, 2002). PIs are essential for regulating the activity of their corresponding proteases and play key regulatory roles in many biological processes. For a few PIs, functions other than blocking protease action have also been found, such as growth factor activities, receptor clearance signaling or involvement in carcinogenesis (Qi *et al.*, 2005). A number of inherited diseases such as emphysema, and epilepsy also result from abnormalities in PIs (Habib and Khalid, 2007).

6.2.1. Classification of plant PIs (PPIs)

The protein proteinase inhibitors are divided into families according to the class of proteolytic enzymes inhibited, and show extensive sequential and structural homology among the members, and similar locations of the disulfide bridges and reactive sites (Laskowsky and Kato, 1980). In the past, plant PIs (PPIs) are primarily classified into four groups depending on their inhibitory action on mechanistic classes of serine, cysteine, aspartic and metallo-proteases (Laskowski, and Qasim, 2000). Recently the

database for PPIs (http://bighost.area.ba.cnr.it/PLANT-PIs) contains information about 495 inhibitors with several isoinhibitors identified in 129 different plants (De Leo *et al.*, 2002). The database describes nine families of PPIs based on sequence similarities. They are: Bowman-Birk serine proteinase inhibitor (BBI), Cereal trypsin/α-amylase inhibitor (BRI), Cysteine proteinase inhibitor (CYS), Metalocarboxypeptidase inhibitor (MCI), Mustard trypsin inhibitor (MSI), Potato Type I inhibitor (PI1), Potato Type II proteinase inhibitor (PI2), Serpin (SPI), Soybean trypsin inhibitor (Kunitz) (KNI) and Squash inhibitor (SQI). With the exception of cysteine protease inhibitor and metallocarboxypeptidase inhibitor family, all the reported families of PIs contain inhibitor of serine proteases the most diffused and studied PIs. Amongst them the Bowman-Birk family and the Kunitz family PIs are the most extensively studied ones.

6.2.2. Physiological role of PIs in plants

Protease inhibitors in general are known to play an important role in nature by regulating the proteolytic activity of their target proteases (Bode and Huber, 1992). The PIs prevent many unwanted proteolysis such as, 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. 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 (Bode and Huber, 2000; Ryan, 1990; Valueva and Mosolov, 1999; Wilson, 1981).

6.2.3. Molecular structure of PPIs

A large number of PIs have been isolated and characterized (Valueva and Mosolov, 1999). Mostly they are low molecular weight proteins (4-20 kDa), soluble in water and their polypeptide chains are non-glycosylated. Protease inhibitors show enormous diversity of function by regulating the proteolytic activity of their target proteinases (Leung *et al.*, 2000), especially serine proteinase inhibitors. The serine PIs have been isolated and characterized extensively from various leguminous plants (Giri *et al.*, 2003). Two types of serine PIs have been characterized by biochemical methods: Kunitz type having molecular weight of 20 kDa with two disulphide bridges (Odani and Ikenaka, 1973) and Bowman–Birk type showing molecular weight of 8 kDa and seven disulphide bridges (Birk, 1985). The inhibitor we report here belongs to Kunitz soybean trypsin inhibitor (STI) family, the members of which have a molecular mass of about 18-20 kDa and contain two disulfide bridges and a single reactive site for trypsin.

The crystal structures of four representative members of Kunitz (STI) family, showing inhibition to trypsin (*Erythrina caffra* trypsin inhibitor (ETI) and soybean trypsin inhibitor (STI) and chymotrypsin (winged bean chymotrypsin inhibitor (WCI) and *Erythrina variegata* chymotrypsin inhibitor (ECI)) have been reported (Onesti *et al.*, 1991; Song *et al.*, 1998; Dattagupta *et al.*, 1996; Iwanaga *et al.*, 1999).

The STI molecule consists of 12 antiparallel β-strands separated by irregular loops. Six of the strands form an antiparallel β-barrel and the top of the barrel is capped with the other six strands, symmetrically arranged in three pairs around the barrel axis (Fig. 6.1). The bottom of the barrel is closed by the loop containing the reactive site and

by the long N-terminal loop, wrapped on itself. The C-terminus also lies at the same end of the barrel (Meester *et al.*, 1998). The β-strands are named based on the topological equivalence between strands belonging to different subdomains as A1, A2, A3, A4, B1 . . . C4. The reactive-site loop containing the scissile bond Arg63-Ile64 is located at the bottom of the molecule between strands A4 and B1, which belong to the β-barrel. The side chain of Asn13 plays a key role in making makes hydrogen bonds with both the main-chain carbonyl O atoms of Ile 64 and Tyr 62 and the amide N atom of Ser 60, on either side of the scissile bond (shown by arrow in yellow loop, Fig. 6.1). Asn13 is a highly conserved residue in the sequences of the inhibitors belonging to the STI family and is required for relegation (Khamrui *et al.*, 2010). This topology has been described as β-trefoil fold and its structural determinants have been analyzed (Murzin *et al.,* 1992; Swindells and Thornton, 1993). This fold was first identified in ricin, a lectin from *Ricinus communis* belonging to type II RIP in 1991 (Rutenber *et al.*, 1991). The STIs share the same β-trefoil fold along with interleukin 1 and heparin-binding growth factors (Murzin *et al.*, 1992). Despite the structural similarity, STI shows no interleukin-1 bioactivity, presumably as a result of their primary sequence disparities. The active inhibitory site containing the scissile bond located in the loop between beta-strands 4 and 5 was found to be conserved in STI and ETI. The STIs belong to a superfamily that also contains the interleukin-1 proteins, heparin binding growth factors (HBGF) and histactophilin, all of which have very similar structures, but share no sequence similarity.

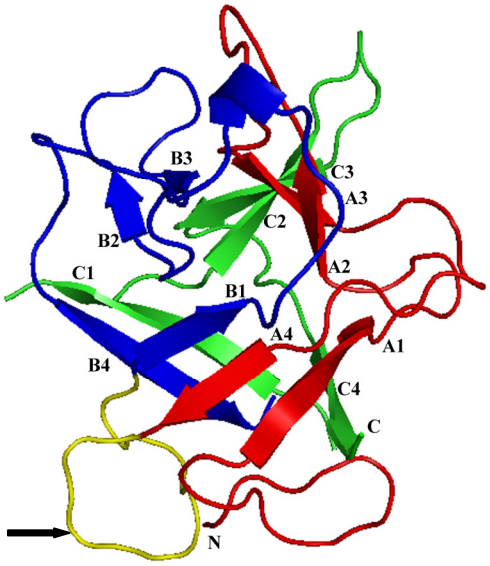

Fig. 6.1. A ribbon representation of the β-trefoil fold of the STIs (PDB ID-1BA7). The pseudo threefold axis is vertical and corresponds to the axis of the barrel. The three subdomains A, B, C are shown in red, blue and green, respectively. The reactive-site loop, between strands A4 and B1, is shown in yellow with the scissile bond (Arg 62-Ile 63) indicated by an arrow.

This chapter describes a newly identified trypsin inhibitor (CPTI) of molecular mass 18 kDa from the seeds of chickpea (*Cicer arietinum*). The CPTI showed efficient inhibitory action against chymotrypsin as well. A Kunitz type inhibitor of molecular mass 20 kDa was previously reported from chickpea (Srinivasan *et al.*, 2005) which showed inhibitory activity exclusively for trypsin and HGP (*Helicoverpa armigera* gut proteinases). A recently characterized trypsin inhibitor from *Cocculus hirsutus*, belonging to Menispermeaceae family causing inhibition of bovine trypsin and *Helicoverpa* gut proteases showed a molecular mass of 18 kDa (Bhattacharjee *et al.*, 2005), Ye and Ng

(Ye and Ng, 2002) have also reported an 18 kDa chickpea protein (CLAP) resembling to cyclophilin-like proteins and possessing antifungal and anti-HIV-1 reverse transcriptase activities.

A recently reported leguminous trypsin-chymotrypsin inhibitor Limenin with molecular mass of 18 kDa from *Phaseolus limensis* showed antibacterial, antifungal and anti-proliferative activity (Wang and Rao 2010). The CPTI inhibitor reported here showed similar molecular mass (18 kDa) like the Limenin from *Phaseolus limensis* and trypsin and chymotrypsin inhibition. All the above mentioned Kunitz type inhibitors have been characterized for their biological role; however, no studies are available related to the structure-function aspects of these 18 kDa Kunitz type inhibitors from legumes. This is the first reported crystallization attempt of a newly identified 18 kDa Kunitz type trypsin inhibitor from chickpea.

6.3. Materials and methods

6.3.1. Materials

Tris(hydroxymethyl)aminomethane (Trizma base), DEAE Sephadex, Sp-sephadex, , NaCl, Polyethylene glycols (PEGs), Ammonium sulfate, Bovine ß-trypsin, Porcine a-chymotrypsin, BApNA were purchased from Sigma Aldrich ltd. (USA). Multiwell trays for crystallization were from Corning (USA). All other chemicals used were of analytical grade and obtained from local suppliers. The matured chickpea seeds of cv PUSA-256 were obtained from PUSA agriculture university, PUSA, Bihar, India.

6.3.2. Purification of CPTI

50 gram dry seeds of *C. arietinum* cv. PUSA-256 were soaked in 250 ml of 10 mM Tris-HCl, pH 7.2 containing 150 mM NaCl. The soaked seeds were homogenized in a mixing blender and the suspension was stirred overnight, filtered through a muslin cloth to remove coarse particles and centrifuged at 11 000 rpm for 20 min at 277 K. The supernatant was filtered through Whatman paper-I to remove the fatty suspension. The filtrate was subjected to acid precipitation at pH 4.5 and subsequently centrifuged at 11 000 rpm for 20 min, at 277 K. The supernatant was again filtered through Whatman paper-I to remove lipidic suspension. The pH of the filtrate was adjusted back to 7.2 and centrifuged. The supernatant was subjected to 80 % saturated ammonium sulfate (AS) precipitation. The precipitate was dissolved in 20 mM Tris-HCl buffer pH 8.0, dialysed against same buffer and centrifuged. The clear supernatant was loaded onto a DEAE-cellulose column pre-equilibrated with 20 mM Tris-HCl buffer pH 8.0. The protein was obtained in the unadsorbed fractions. These fractions were resolved on a 12 % SDS-PAGE and also checked for activity. Fractions containing the 18 kDa CPTI were pooled together and dialysed against 20 mM acetate buffer, pH 5.0 and loaded onto an SP-Sephadex column (pre-equilibrated with 20 mM acetate buffer pH 5.0). Elution of the bound protein was carried out using 0.1-0.5 M NaCl gradient in stepwise manner where the CPTI elutes out between 0.2-0.3 M NaCl. Elutes were resolved on 12 % SDS-PAGE for checking the homogeneity. The purified and active fractions of CPTI were stored at -20 ^{o}C for further use.

6.3.3. Proteinase inhibitory activity of CPTI

Bovine trypsin and trypsin like activity were estimated using enzyme-specific chromogenic substrate, BApNA as reported elsewhere (Giri *et al.,* 1998). Trypsin activity was first calibrated by end point titration of varying amounts of proteinase (5, 10 and 25 µg of trypsin). For BApNA assay, 150 µl of diluted enzyme solution was added to 1 ml of 1 mM substrate solution (BApNA in 0.1M Tris buffer pH 7.8) and incubated at 310 K for 10 min. The reaction was terminated by the addition of 200 µl of 30% acetic acid and absorbance was checked at 410 nm for the BApNA.

Chymotrypsin inhibition activity was measured by azocaseinolytic assay (Brock *et al.,* 1982). For carrying out azocaseinolytic assay 60 µl of diluted enzyme was added to 200 µl of 1% azocasein (in 0.2 M glycine–NaOH, pH 10.0) and incubated at 37 °C for 30 min. The reaction was terminated by the addition of 300 µl of 5% trichloroacetic acid. After centrifugation at 14,230×g for 10 min at 4^0C, an equal volume of 1 M NaOH was added to the supernatant and the absorbance was measured at 450 nm.

One proteinase unit was defined as the amount of enzyme that increased the absorbance by 1.0 OD under the given assay conditions. For inhibitor assay a suitable amount of inhibitor and enzyme were pre-incubated at room temperature for 20 min and the residual enzyme activity was assayed as above. One PI unit is defined as the amount of inhibitor required for inhibiting one proteinase activity unit.

6.3.4. Crystallization

Initial crystallization trials were conducted using the commercially available sparse matrix crystallization screens supplied by Hampton research (USA) and Molecular

dimensions Ltd (UK) by hanging-drop vapor-diffusion method (McPherson, 1982). The purified and active CPTI crystallized from a 25 mg ml^{-1} solution. The crystals appeared and grew to full size in 3 days; these were used for data collection. To prepare heavy atom derivative of the CPTI, co-crystallization as well as quick soaking approach were tried. For co-crystallization freshly prepared heavy metal salt in the range of 1-5 mM was mixed with protein solution prior to setting up crystallization experiments. Alternatively, pre-formed crystals of CPTI were soaked in the mother liquor containing heavy metal salts at a maximum concentration the crystals survived, where the heavy atom concentration crystals could withstand was 10-1000 mM. The time of soak was varied from 10 seconds to several hours. These crystals were flash-freezed in liquid nitrogen for low temperature (-180 ^{o}C) data collection using Xstream. Salts of various heavy metals, namely $HgCl_2$, Mercuric acetate, $Pb(NO_3)_2$, KI, NaI, p-Hydroxy mercury benzoate (PCMB), dichloro(ethylenediamine)platinum (II), Sodium molybdate were used in the crystallization and soaking experiments. To avoid the formation of ice crystals at low temperature various cryo-protectants were tried out such as PEG's 200, 300, 400, 600, 1K, 2k, 4k, 5k, 6k, and 8K in the range of 20-40 %. In addition to PEG's Glycerol (10-25 %), Ethylene glycol (10-30 %), various combination of Paraffin and Silicon oil (1:1, 1:2, etc.), MPD (20-35 %) were also tried out. Amongst them, 26-30 % MPD turned out to be a better cryo-protectant of choice which offered appropriate cryo-protection with minimum mosaicity for CPTI crystals grown in the above condition.

6.3.5. X-ray data collection and analysis

Initial characterization of CPTI crystals were carried out at room temperature (RT, 295 K) using Cu-Kα radiation from a Rigaku rotating anode X-ray generator and a

R-Axis IV^{++} image plate detector. Crystal-to-detector distance was adjusted to 150 mm and each oscillation frame was recorded with a 0.5^0 oscillation. Initial data collected at 2 Å were processed using *HKL* suit (Otwinowski and Minor, 1997). The high resolution data (1.4 Å) of CPTI crystal belonging to orthorhombic form $P2_12_12A$ were collected at beam line BM14, ESRF, Grenoble. Diffraction data were recorded on a MAR CCD detector with a diameter of 130 mm. The crystal-to-detector distance was 100 mm with 1^o oscillations per frame and exposure time of 10 seconds. Data processing and scaling were accomplished with *MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1993) of the *CCP4* suite (CCP4, 1994). Data for MAD phasing were collected on BM14 beamline at ESRF, Grenoble, France for the native and Iodine derivative of CPTI at wavelengths 1.77 and 1.54 Ǻ for sulfur and iodine, respectively while the high resolution data were collected at 0.95 Ǻ.

6.4. Results and discussion

Total yield of pure CPTI after final purification step was 50 mg/100 gram of dried chickpea seeds. The purified and homogeneous CPTI elutes out from the SP-sephadex coulumn, with 200-300 mM NaCl gradient. The purified fractions of CPTI were resolved on 12 % SDS-PAGE, where the CPTI showed a single band corresponding to a molecular mass of 18 kDa (Fig.6.2). An attempted N-terminal sequencing analysis of CPTI did not work and hence gave no clue to the resemblance of this protein with any known trypsin inhibitors from legumes or other families.

Inhibitory action of CPTI was determined against trypsin as well as for chymotrypsin. CPTI showed strong inhibitory activity against trypsin compared to

chymotrypsin. The CPTI showed IC50 value of 2.5 µg against trypsin while it was 3.0 µg for chymotrypsin inhibition (fig.6.3A and B). The minimum inhibitory concentration estimated was 0.8 µg/ml and 0.75 µg/ml for trypsin and chymotrypsin, respectively. The corresponding inhibitory action of CPTI estimated was 114 TIU/mg of protein and 77 CIU/mg of protein. This inhibitory capacity of CPTI was five times more compared to the previously reported kunitz inhibitor from chickpea possessing molecular mass of 20 kDa (Srinivasan *et al.*, 2005).

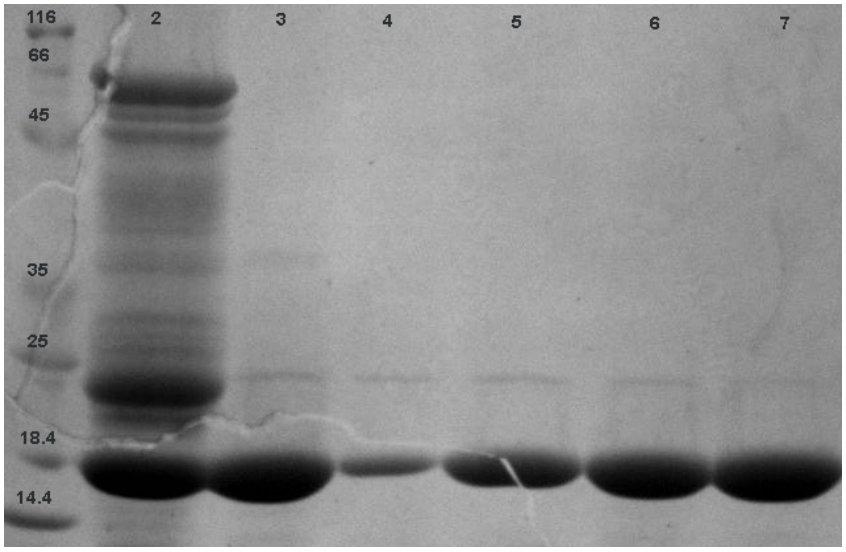

Figure 6.2. Purification profile of CPTI in SDS-PAGE (12 %) showing a single band of molecular mass 18 kDa. Lanes from left to right: Lane 1 -Molecular weight markers (fermentas), Lane 2-DEAE unadsorbed fraction, Lane 3-SP-sephadex elutes with 200 mM NaCl, Lane 4-7-SP-sephadex elutes with 300 mM NaCl.

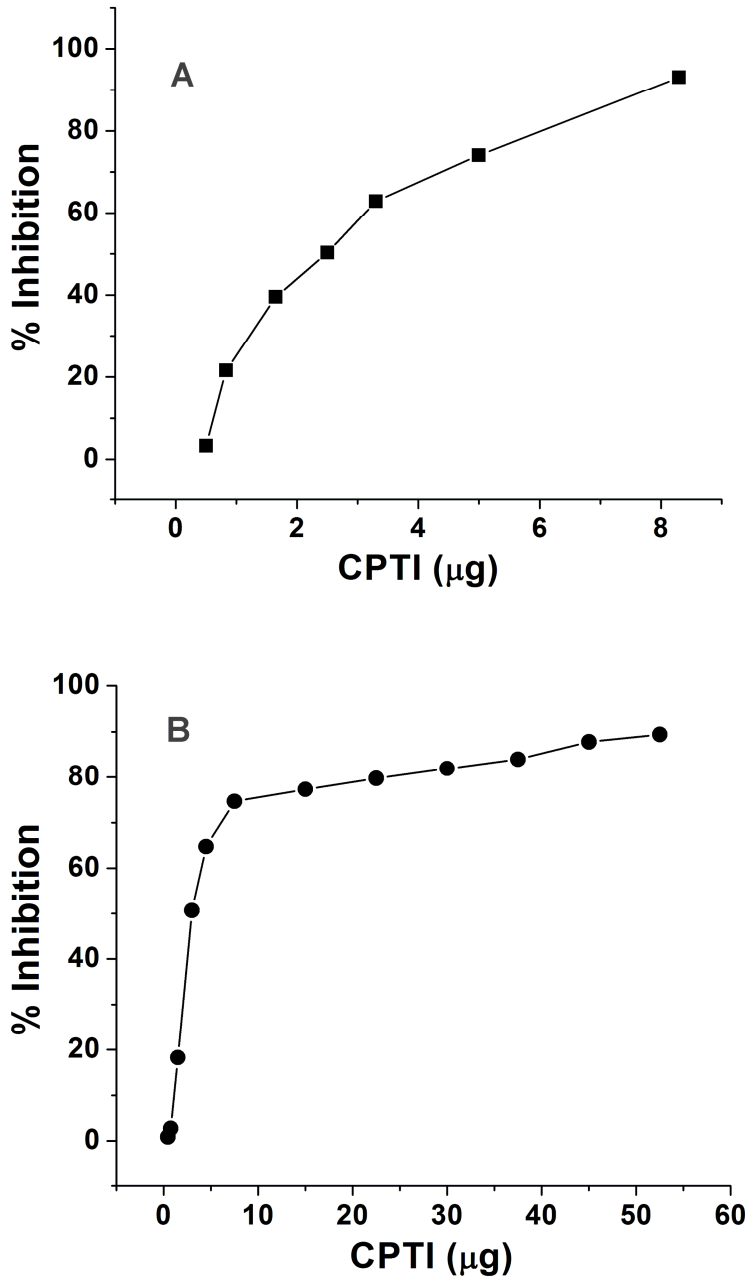

Figure 6.3. Inhibitory activity of chickpea trypsin inhibitor (CPTI) against trypsin (A) and chymotrypsin (B). (Filled square and filled circle represent trypsin and chymotrypsin, respectively)**.**

Crystallization experiments of purified CPTI gave three different orthorhombic crystal forms in three different conditions of JCSG screen. Preliminary X-ray

characterization of all the three crystal forms were carried out at home source. The crystals of orthorhombic form $P2_12_12A$ having unit cell dimensions a=37.2, b=41.2, c=104.6 Å were obtained from condition 18 (0.1 M citrate-phosphate buffer pH 4.2, 40 % ethanol, 5 % w/v PEG 1K) of the JCSG I screen (Fig.6.4 A). The calculated Matthews coefficient was 2.11 $Å^3$ Da^{-1} and the corresponding solvent content 42 % (Table.6.1). The crystals grew to maximum size in 3-4 days, diffracted up to 2.0 Å at home source and could be easily reproduced. Further studies were carried out with this crystal form only. CPTI crystals of P21212A form were soaked in the presence of several heavy metal salts as described above in crystallization section with varying times of soak and data were collected at low temperature in home source to check the presence of bound heavy metal if/any. The data collection and processing statistics of these crystals are listed in table 6.1.

Crystals belonging to the orthorhombic form $P2_12_12_1$ with unit cell a=41.1, b=50.1, c=75.1 Å (Fig.6.4B), were obtained from condition 12 (0.1 M imidazole pH-8.0, 10 % w/v PEG 8K) of the JCSG II screen. The crystals in this form used to appear immediately after overnight incubation. They diffracted only up to 3.0 Å in home source at room temperature. For the preparation of heavy atom derivative of this form first we tried co-crystallization. We could obtain few crystals grown in the presence of 2 and 3 mM KI. Diffraction data were collected for few of them and processed. However, co-crystallization could not succeed beyond 3 mM KI, therefore soaking experiments were tried out for better derivatization. During soaking it was observed that these crystals could not tolerate > 3 mM concentration of KI. The crystals of this form could not withstand heavy atom soaking and diffraction quality further worsened when freezed at low temperature. Many a times these crystals get dissolved immediately upon soaking

with heavy metal salts. Due to several difficulties encountered in getting good diffraction from these crystals further studies could not be planned with orthorhombic form $P2_12_12_1$ of CPTI crystals. Data collection and processing parameters for this form are shown in Table 6.2. Thin needle crystals of the orthorhombic form $P2_12_12B$ obtained from condition 9 (0.2 M ammonium chloride, 20 % PEG 3350) of JCSG screen I with unit cell dimensions a= 44.4, b= 75.7 and c= 133.7 Å diffracted only up to 3.5 Å (Fig.6.4C). No complete data could be collected due to radiation damage of these crystals.

Although models suitable for structure determination by molecular replacement could not be identified due to lack of sequence information, a few structures of other Kunitz inhibitors were used for phasing such as kunitz inhibitor from soybean STI (PDB ID- 1BA7), Kunitz inhibitor from *Erythrina caffra*, ETI (PDB ID-1TIE) and the trypsin inhibitor from the seeds of *Delonix regia*, DrTI (PDB ID-1R8N). The solutions for molecular replacement calculation using *PHASER* (McCoy, 2007) gave poor Z-score below four and thus could not proceed further. Efforts were made to get suitable heavy atom derivatives for structure determination using MIR, MAD or SAD methods. To increase the redundancy of the data to be utilized in these phasing methods data were collected at high intensity tunable synchrotron source, BM14 at ESRF, Grenoble, France. Data collection and processing statistics for the native and derivative crystals of orthorhombic form $P2_12_12A$ at BM14 beamline is shown in Table 6.3.

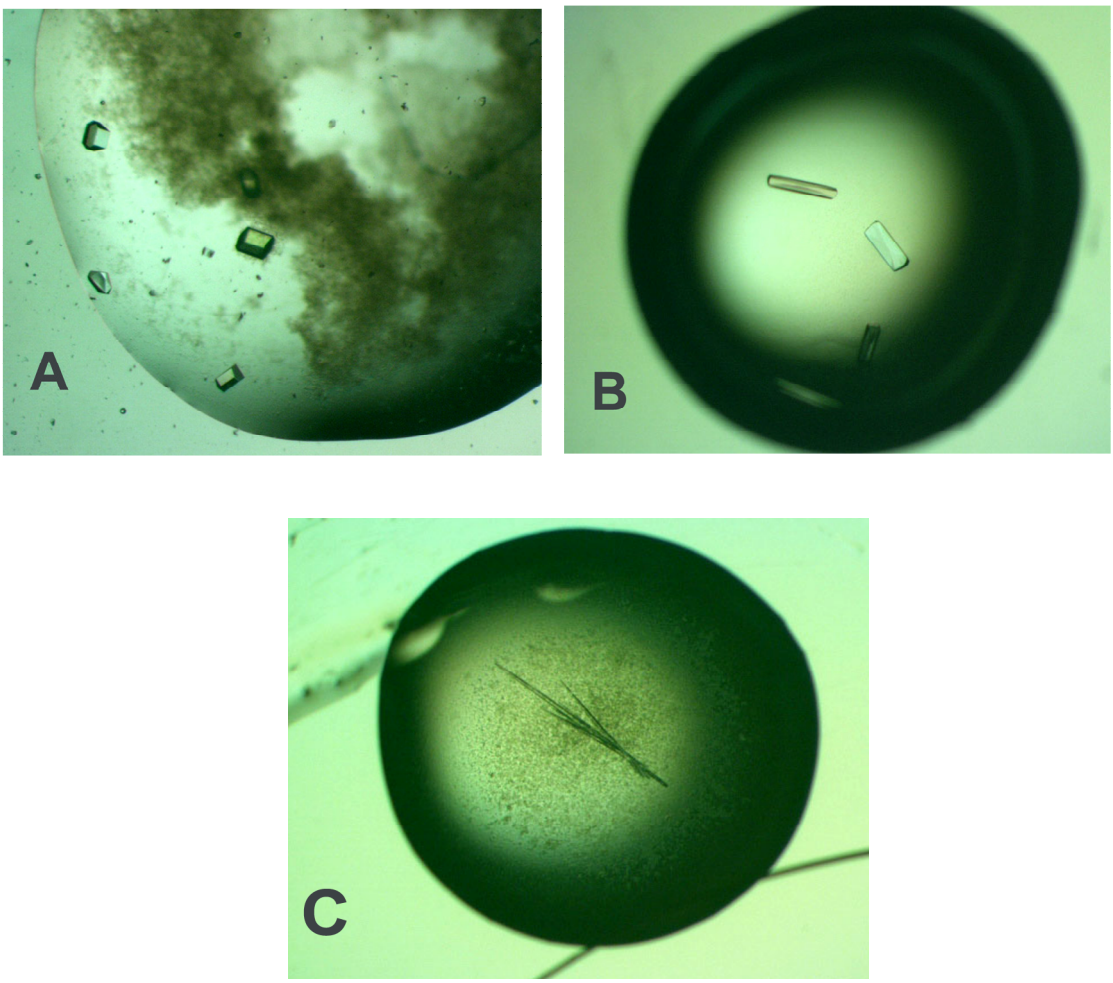

Figure 6.4. The crystals belonging to the three orthorhombic forms of CPTI. Orthorhombic form $P2_12_12A$ with unit cell a=37.2, b=41.2, c=104.6 Å (A), orthorhombic form $P2_12_12_1$ with unit cell a=41.1, b=50.1, c=75.1 Å (B) and orthorhombic form $P2_12_12B$ having unit cell a= 44.4, b= 75.7 and c= 133.7 Å (C).

Table 6.1. Data collection statistics for native CPTI crystals and the crystals soaked with heavy atom salts in the space group $P2_12_12A$ at home source.

Heavy atom salt used	Native	$HgCl_2$	Pt (Cis-PtII Diammine dichloride)	Iodine (KI)
X-ray source (ʎ)	CuKα (1.5418)	CuKα (1.5418)	CuKα (1.5418)	CuKα (1.5418)
Temp ^{0}C	-140	-140	-140	-140
Crystal-to-detector Distance mm	200	200	200	200
No. of frames	1-265	1-107	1-240	1-360
Unit cell *a* Å	36.6	36.6	36.8	36.6
Unit cell *b* Å	102.8	102.9	103.5	103.2
Unit cell *c* Å	40.6	40.6	40.8	40.7
Unit cell volume Å^3	153030	153666	155800	154193
Resolution range (last shell) Å	80-2.45 (2.49-2.45)	80-2.45 (2.49-2.45)	80-2.45 (2.49-2.45)	80-2.40 (2.44-2.40)
Reflections total	60,154	23,926	53,986	88,195
Reflections unique	6,025	10,904	11,090	11,494
Multiplicity	8.61	2.19	4.57	6.98
Completeness %	99.8 (100)	99.8 (99.5)	100 (100)	98.6 (94.3)
R_{merge}	4.1 (6.6)	3.6 (5.9)	4.4 (7.5)	5.9 (7.7)
Matthews Coefficient Å^3/Da	2.12	2.12	2.16	2.14
Solvent content %	42	42	43	42

Table 6.2. X-ray data collection and processing statistics for crystals of CPTI (orthorhombic form B) belonging to space group $P2_12_12_1$ at RT.

CPTI crystals	Native	Iodine Co-crystallization with 3mM KI
X-ray source (Å)	CuKα (1.5418)	CuKα (1.5418)
Temp ^oC	22	22
Crystal-to-detector Distance(mm)	150	150
Mosaicity	0.40	0.35
No. of frames	1-200	1-214
Unit cell a Å	41.1	41.1
Unit cell b Å	50.1	50.1
Unit cell c Å	75	75
Unit cell volume $Å^3$	155156	154770
Resolution range (last shell) Å	50-3.00 (3.05-3.00)	50-3.00 (3.05-3.00)
Reflections, total	25,178	27,159
Reflections, unique	3,400	3,291
Multiplicity	7.25	7.97
Completeness %	99.9 (98.0)	97.9 (100)
R_{merge}	11.9 (29.8)	13.9 (33.4)
Matthews Coefficient $Å^3$/Da	2.11	2.10
Solvent content %	42	42

Table 6.3. Data collection and processing statistics of CPTI crystals in orthorhombic form A (space group $P2_12_12A$) soaked in iodine solution.

Iodine salt used (M)	Native($P2_12_12A$)	NaI (500mM)	NaI (500mM)
Beamline	BM14, ESRF	BM14, ESRF	BM14, ESRF
X-ray wavelength (Å)	0.95	1.5418	0.95
Temp ^oC	-180	-180	-180
Crystal-to-detector Distance mm	187.6	109.4	109.4
Mosaicity	0.55	0.45	0.39
No. of frames	720	720	720
Unit cell a Å	36.56	36.49	36.48
Unit cell b Å	102.49	102.55	102.50
Unit cell c Å	40.55	40.44	40.43
Unit cell volume $Å^3$	151942	151328	151175
Resolution range (last shell) Å	34.43-1.60 (1.69-1.60)	31.75-1.98 (2.08-1.98)	20.50-1.30 (1.37-1.30)
Reflections, total	5,18,340	1,51,602	5,35,016
Reflections, unique	20,675	10,540	38,120
Multiplicity (last shell)	25.1	14.4 (13.9)	14 (13.4)
Completeness %	99.1 (93.9)	94.9 (90.8)	99.8 (99.4)
R_{merge} (Last shell)	0.062 (0.387)	0.033 (0.078)	0.048 (0.403)
Matthews Coefficient $Å^3$/Da	2.11	2.10	2.10
Solvent content %	41	41	41

From the datasets collected in the synchrotron source only a low occupancy site could be characterized for an iodine derivative. Attempts to phase the protein using this dataset were not successful. In continuation to this, studies are underway to determine the crystal structure of CPTI by complexation with proteases such as trypsin/chymotrypsin. Since these attempts are elaborate and thus bound to take time therefore, the structure of CPTI could not be reported in this thesis.

Chapter 7

Conclusions

Structural folds of plant lectins have been reviewed in detail by several groups (Vijayan and Chandra, 1999; Wright, 1997; Rini, 1995). The most commonly observed fold in plant lectins is the "Jelly-roll motif" belonging to legume lectins". This fold and the common features of carbohydrate recognition in legume lectins have been described in chapter 1.The three-fold symmetric β-trefoil fold has been observed mostly in type II-RIP's and in amaranthin as an exception. The other two three-fold symmetric lectin folds are β-prism II and β-prism I, observed exclusively in monocot mannose binding lectins and JRLs respectively. The overall subunit structure and the principles of sugar recognition are nearly identical within a particular lectin family. However, in case of legume and monocot mannose-binding lectin families variations have been observed in the mode of sugar recognition and new lectins have been discovered showing specificity for complex sugars and glycoproteins.

In the present investigations structure-function studies of three plant lectins showing specificity for complex sugars have been described. The first two lectins under the study are CAL and ACL, belonging to legume and monocot mannose-binding lectin families respectively. The third hemagglutinin studied is MoL from moraceaae family. All the three hemagglutinins show hemagglutination inhibition with complex glycoproteins such as fetuin, throglobulin, fibrinogen and their corresponding glycopepetides and not with any simple sugars.

The three-dimensional structure of CAL consists of four bladed β-propellers, which is unique amongst the known plant lectin folds. Since the sugar binding site of CAL is not known chemical modification studies were carried out to identify the putative sugar binding site which revealed the possibility of only one sugar binding site per

monomer of CAL. The lectins containing β-propeller motif are few. They have been reported from sources other than plant such as bacteria (six bladed *Ralstonia solanacearum* lectin, RSL), fungi (six-bladed *Aleuria aurantia* lectin, AAL) and eukaryotes (five bladed Tachylectin-2 form horse shoe crab). The RSL and AAL show specificity for fucose while tachylectin recognizes GlcNAc. The structure of these lectins at the sugar binding site were compared with individual blades of CAL, but the results were ambiguous and no further information could be obtained regarding the sugar binding site of CAL. Large deviations in C-alpha carbons were observed in all the cases with rmsd >3.0 Å. The β-propeller fold containing proteins are ubiquitous and they show diversity at functional, phylogenetic as well as at the level of their sequences (Paoli, 2001) thus imposing difficulty in protein structure and function assessment by sequence comparisons. The only structure available for the lectins showing specificity for complex sugars is for fetuin binding lectin SCAfet from the bulbs of *Scilla campanulata* (Wright *et al.*, 2000). Like other monocot lectins the structure of SCAfet showed β-prism II fold and no sugar bound structure is available. Comparison showed no similarity between the folds and sequences of SCAfet and CAL. Though CAL shows high hemagglutinating activity, the sequence and structure of CAL resembles more to the pa2 albumin from legumes (Gaur *et al.*, 2010 and 2011) than the known legume lectins. Plants might have utilized the pa2 albumin in defense as a lectin, apart from the other usual functions attributed to the albumins.

The second lectin showing specificity towards complex glycoproteins is ACL with molecular mass of 13kDa. Since, ACL could not be crystallized the structure-activity studies were carried out using biophysical and biochemical methods. The single

tryptophan in the protein might be important for sugar recognition as upon modification of this residue hemagglutination activity of ACL was lost. The Chemical unfolding studies with GdnHCl revealed the presence of an active intermediate at 0.25 M GdnHCL. At pH 3.0 the protein undergoes transition to a highly ordered molten globule state which retains 40 % hemagglutinating activity suggesting its involvement in protein folding pathway (Urvashi *et al.*, 2011). The ACL molten globule was also resistant to thermal and chemical denaturation compared to the native lectin.

The third lectin with complex sugar specificity is MoL, also a single tryptophan containing protein of 14kDa. The lectin showed unusual stability in solution but could not be crystallized. The structure of MoL was also found to be intact at higher temperature upto 85 ^{o}C owing to the presence of three disulphide bridges per dimer. In the presence of reducing agents like β-ME the protein unfolds and undergoes conformational changes leading to enhanced exposure of hydrophobic amino acids side chains on the surface of protein as revealed by NMR and fluorimetric studies.

Although the three lectins (CAL, ACL and MoL) showing specificity for complex sugars studied here are single tryptophan containing proteins, the residues involved in sugar recognition may not be same. In CAL, the tryptophan is not exposed to the surface of the protein as revealed from the structure and that may be the reason for the unchanged hemagglutination activity upon treatment with NBS for chemical modification. Unlike CAL, the modification of tryptophan in ACL led to complete loss of hemagglutination activity confirming the involvement of this residue in sugar binding. In case of MoL also modification of tryptophan did not affect the activity. Comparing the structural stability of these lectins, by inducing environmental changes by varying pH and temperature as

well as by treating with chemical denaturants it has shown that the structure of MoL is resistant to thermal and chemical unfolding to a significant extent compared to ACL. However, both ACL and MoL were found to be active in a wide range of pH. Biophysical studies on CAL are being carried out by another research group and are inconclusive hence, no attempts were made to repeat them here though it might have been informative to compare the structural stability and activity variation of this lectin whose structure is determined with ACL and MoL showing similar complex sugars specificity.

Amongst the plant lectins belonging to β-prism I fold, showing specificity for galactose and its derivatives jacalin is unique, where for the first time it was observed that the proteolytic processing of the mature polypeptide generates free amino terminus at Gly1 which imparts sugar specificity to the lectin (Sankarnarayanan *et al.*, 1996). To study the role of posttranslational modification in generating sugar specificity we have determined the structure of rjacalin (cloned in *E. coli.*) where the polypeptide chain is continuous (β and α-chain are linked by tetrapetide "TSSN", Sahasrabudhe *et al.*, 2004) and thus the amino terminus of Gly1 of α-chain is not free for binding the sugar. Single chained and unglycosylated rjacalin is still able to recognize galactose but with highly reduced affinity. A galactose bound structure of rjacalin could not be obtained because the highly dynamic TSSN loop at the entrance to the sugar binding site prevents the firm binding of sugar. Our analysis also showed that the loop brings the residues from the beta chain closer to the sugar binding site, leading to the decrease in total surface area and volume of the binding pocket and loss of the extended binding site in the lectin which is reflected in the differences in sugar recognition and binding property of rjacalin compared to njacalin.

Apart from the CAL, a trypsin inhibitor protein (CPTI) of molecular mass 18kDa is also purified from chickpea seeds for the first time. The CPTI belonged to the Kunitz soyabean type inhibitor family and showed strong inhibitory action against trypsin and chymotrypsin. CPTI revealed no similarity with the known Kunitz inhibitor from chickpea and other proteins. CPTI crystallized in three different orthorhombic crystal forms and high resolution data were collected for orthorhombic form $P2_12_12$ A (Urvashi and Suresh, 2011). Owing to the lack of sequence and a related structure molecular replacement method was not successful to determine the structure of CPTI. In continuation, attempts are underway to phase the reflections of the CPTI using MAD/SAD method or by complexing CPTI with trypsin/chymotrypsin.

References

- Adams P D, Afonine P V, Bunkóczi G, Chen V B, Davis I W, Echols N, Headd J J, Hung L W, Kapral G J, Grosse-Kunstleve R W, McCoy A J, Moriarty N W, Oeffner R, Read R J, Richardson D C, Richardson J S, Terwilliger T C and Zwart P H, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* D **66**, 213-22.
- Agrawal K M and Bahl O P, (1968) Glycosidases of *Phaseolus vulgaris*. II. Isolation and general properties. *J. Biol. Chem.* **243**, 103–111.
- Ahmed H and Chatterjee B P, (1989) Further Characterization and Immunochemical Studies on the Carbohydrate Specificity of Jackfruit (*Artocarpus integrifolia*) Lectin. *J. Biol. Chem.* **264**, 9365-9372.
- Albani J R, (2004) Structure and dynamics of macromolecules: absorption and fluorescence studies. Elsevier Science.
- Allen A K, Bolwell G P, Brown D S, Sidebottom C and Slabas A R, (1996) Potato lectin: a three-domain glycoprotein with novel hydroxyproline-containing sequences and sequence similarities to wheat-germ agglutinin. *Int. J. Biochem. Cell. Biol.* **28**, 1285-1291.
- Anwar F, Latif S, Ashraf M and Gilani A H, (2007) *Moringa oleifera*: A Food Plant with Multiple Medicinal Uses. *Phytother. Res*. **21**, 17–25.
- Ashraf M T and Khan R H, (2003) Mitogenic lectins. *Med. Sci. Monit*. **9,** 265-269.
- Astoul C H S , Peumans W J , Van Damme E J M, Barre A , Bourne Y and Rouge P, (2002) The size, shape and specificity of the sugar-binding site of the jacalin-related lectins is profoundly affected by the proteolytic cleavage of the subunits. *Biochem. J.* **367**, 817-824.
- Aucouturier P, Mihaesco E, Mihaesco C and Preud'homme J L, (1987) Characterization of jacalin, the human IgA and IgD binding lectin from jackfruit. *Mol. Immunol*. **24**, 503-511.
- Audette G F, Vandonselaar M and Delbaere LT, (2000) The 2.2 A resolution structure of the O(H) blood-group-specific lectin I from *Ulex europaeus. J. Mol. Biol.* **304,** 423-433.
- Aue W P, Bartholdi E and Ernst R R, (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229–2246.

- Baker S C, Saunders N F W, Willis A C, Ferguson S J, Hajdu J and Fülöp V, (1997) Cytochrome cd1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β-propeller folds. *J. Mol. Biol*. **269**, 440-455.
- Balzarini J, Neyts J, Schols D, Hosoya M, Van Damme E, Peumans W and De Clercq E, (1992) The mannose-specific plant lectins from *Cymbidium* hybrid and *Epipactis helleborine* and the (*N*-Acetylglucosamine)n-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication in vitro. *Antiviral Res.* **18**, 191-207.
- Balzarini J, Schols D, Neyts J, Van Damme E, Peumans W and De Clercq E, (1991) α-(1–3)- and α-(1–6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob. Agents. Chemother.* **35**, 410-416.
- Banerjee R, Dhanaraj V, Mahanta SK, Surolia A and Vijayan M, (1991) Preparation and X-ray characterization of four new crystal forms of jacalin, a lectin from *Artocarpus integrifolia. J. Mol. Biol*. **221**, 773–776.
- Banerjee R, Mande S C, Ganesh V, Das K, Dhanaraj V, Mahanta S K, Suguna K, Surolia A and Vijayan M, (1994) Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc. Natl. Acad. Sci. USA.* **91**, 227–231.
- Barbieri L, Battelli M G and Stirpe F, (1993) Ribosome-inactivating proteins from plants. *Biochim. Biophys. Acta* **1154**, 237-282.
- Barondes S H, (1988) Bifunctional properties of lectins: lectins redefined. *Trends. Biochem. Sci.* **13**, 480–482.
- Barre A, Bourne Y, Van Damme E J M, Peumans W J and Rouge P, (2001) Mannose-binding plant lectins: Different structural scaffolds for a common sugar-recognition process. *Biochimie* (Paris). **83**, 645–651.
- Baumann C M, Strosberg A D and Rudiger H, (1982) Purification and Characterization of a Mannose/Glucose-Specific Lectin from *Vicia cracca. Eur. J. Biochem. **122**,* 105-3 10.
- Beauchamp J C and Isaacs N W, (1999) Methods for X-ray diffraction analysis of macromolecular structures. *Curr. Opin. Chem. Biol.* **3**, 525-529.

- Beintema J J and Peumans W J, (1992) The primary structure of stinging nettle (*Urtica dioica*) agglutinin: a two-domain member of the hevein family. *FEBS Lett*. **299**, 131–134.
- Beintema J J, (1994) Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett.* **350**, 159-163.
- Beisel H-G, Kawabata S-i, Iwanaga S, Huber R and Bode W, (1999) Tachylectin-2: crystal structure of a specific GlcNAc/ GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *The EMBO Journal*. **18**, 2313–2322.
- Bhattacharjee C, Manjunath N H and Prasad D T, (2009) Purification of a Trypsin Inhibitor from *Cocculus hirsutus* and Identification of Its Biological Activity. *J. Crop Sci. Biotech.* **12**, 253- 260.
- Birk Y, (1985) Trypsin- and chymotrypsin-inhibitor from soybeans. *Int. J. Pept. Protein Res*. **25**, 113–131.
- Bode W and Huber R, (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433-451.
- Bode W and Huber R, (1994) Proteinase-protein inhibitor interactions. *Fibrinolysis*. **8**,161-171.
- Bompard-Gilles C, Rousseau P, Rougé P and Payan F, (1996) Substrate mimicry in the active center of a mammalian α-amylase: structural analysis of an enzyme–inhibitor complex. *Structure* **4,** 1441-1452.
- Bouckaert J, Hamelryck T, Wyns L and Loris R, (1999) Novel structures of plant lectins and their complexes with carbohydrates. *Curr. Opin. Stru. Biol*. **9**, 572-577.
- Bouckaert J, Loris R, Poortmans F and Wyns L, (1995) The crystallographic structure of metal free concanavalin A at 2.5 Å resolution. *Proteins Struct. Funct. Genet.* **23,** 510-514.
- Bourne Y, Mazurier J, Legrand D, Rougé P, Montreuil J, Spik G and Cambillau C, (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, Roig-Zamboni V, Barre A, Peumans W J, Astoul C H, Van Damme E J M and Rouge P, (2004) The Crystal Structure of the *Calystegia sepium* Agglutinin Reveals a Novel Quaternary Arrangement of Lectin Subunits with a β-Prism Fold. *J. Biol. Chem.* **279**, 527-533.

- Bourne Y, Roig-Zamboni V, Barre A, Peumans W J, Astoul C H, Van Damme E J M and Rouge P, (2004) The Crystal Structure of the *Calystegia sepium* Agglutinin Reveals a Novel Quaternary Arrangement of Lectin Subunits with a β-Prism Fold. *J. Biol. Chem.* **279,** 527-533.
- Bourne Y, Zamboni V, Barre A, Peumans W J, Van Damme E J M and 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.
- Boyd W C and Reguera R M, (1949) Studies on haemagglutinins present in seeds of some representatives of the family Leguminoseae. *J. Immunol.* **62**, 333-339.
- Boyd W C and Shapleigh E, (1954). Specific precipitating activity of plant agglutinins (lectins). *Science* **119**, 419.
- Bradford M M, (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* **72**, 248–254.
- Brahamachari S K, Bhatnagar R S and Ananthanarayanan, (1982) Proline-Containing &Turns in Peptides and Proteins. 11. Physicochemical Studies on Tripeptides with the Pro-Gly Sequence. *Biopolymers* **21**, 1107-1125.
- Brewin N J and Kardailsky I V, (1997) Legume lectins and nodulation by *Rhizobium. Trends Plant Sci.* **2**, 92-98.
- Brock R M, Forsberg C W and Buchanan-Smith J G, (1982) Proteolytic activity of rumen microorganisms and effect of proteinase inhibitors. *Appl. Environ. Microbiol*. **44,** 561–569.
- Broekaert W F and Peumans W J, (1986) Lectin release from seeds of *Datura stramonium* and interference of the *Datura stramonium* lectin with bacterial motility. **In***: Lectins, Biology, Biochemistry, Clinical Biochemistry*. Vol. 5., pp. 57–66, Bøg-Hansen TC, Van Driessche E, Eds, W. De Gruyter, Berlin.
- Broekaert W F, Marien W, Terras F R G, De Belle M F C, Proost P, Van Damme J, Dillen L, Claeys M, Rees S B, Vanderleyden J and Cammuet B P A, (1992) Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. *Biochemistry* **31**, 4308-4314.

- 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, (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Cryst.* D **49**, 24-36.
- Bunn-Moreno MM and Campos-Neto A, (1981) Lectin(s) extracted from seeds of Artocarpus integrifolia (jackfruit): potent and selective stimulator (s) of distinct human T and B cell functions. *J. Immunol.* **127**, 427–429.
- Burger M M, (1974) Assays for agglutination with lectins. *Methods Enzymol.* **32,** 615-621.
- Burstein E A, Abornev S M and Reshetnyak Y K, (2001) Decomposition of protein tryptophan fluorescence spectra into log-normal components. I.Decomposition algorithms. *Biophys. J.* **81**, 1699–1709.
- Burstein E A, Vedenkina N S and Ivkova M N, (1973) Fluorescence and the Location of Tryptophan Residues in Protein Molecules. *Photochem. Photobiol*. **18**, 263–279.
- Bushmarina N A, Kuznetsova I M, Biktashev A G, Turoverov K K and Uversky V N, (2001) Partially Folded Conformations in the Folding Pathway of Bovine Carbonic Anhydrase II: Fluorescence Spectroscopic Analysis. *Chem. Bio. Chem.* **2**, 813-821.
- Capon C, Piller F, Wieruszeski J M, Leroy Y and Fournet B, (1990) Structural analysis of the carbohydrate chain isolated from Jacalin lectin. *Carbohydr. Res.* **121**, 1–1.
- Chandra N R, Ramachandraiah G, Bachhawat K, Dam T K, Surolia A and Vijayan M, (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. *J. Mol. Biol*. **285**, 1157-1168.
- Chatterjee B P, Vaith P, Chatterjee S, Karduck D and Uhlenbruck G, (1979) Comparative studies of new marker lectins for alkali-labile and alkali-stable carbohydrate chains in glycoproteins. *Int. J. Biochem.* **10,** 321-327.
- Chayen N E, (1998) Comparative studies of protein crystallization by vapor diffusion and microbatch. *Acta Cryst*. **D54**, 8–15.
- Chopra R N, Nayar S L and Chopra I C, (1986) Glossary of Indian Medicinal Plants (Including the Supplement). Council of Scientific and Industrial Research, New Delhi. Very terse details of

medicinal uses of plants with a wide range of references and details of research into the plants chemistry. Not for the casual reader.

- Chrispeels M J and Raikhel N V, (1991) Lectins, lectin genes, and their role in plant defense. *Plant Cell* **3**, 1-9.
- Christeller J and Liang W, (2005) Plant serine protease inhibitors. *Protein and Peptide Letters.* **12**, 439-447.
- Christiane Bies, Claus-Michael Lehr and Woodley J F, (2004) Lectin - mediated drug targeting: history and applications. *Advanced Drug Delivery Reviews.* **56**, 425-235.
- Clemente A, Olı´as R and Olı´as J M, (2000) Purification and characterization of broad bean lipoxygenase isoenzymes. *J. Agric. Food Chem.* **48**, 1070–1075.
- Collaborative Computational Project, Number 4 (1994) *Acta Cryst* D, **50**, 760-763.
- Collinge D B, Kragh K M, Mikkelsen J D, Nielsen K K, Rasmussen U and Vad K, (1993) Plant chitinases. *Plant J.* **3**, 31-40.
- Cooper A, (1999) Thermodynamic analysis of biomolecular interactions. *Curr. Opin. Chem. Biol.* **3**, 557-563.
- Corbeau P, Pasquali J L and Devaux C, (1995) Jacalin, a lectin interacting with O-linked sugars and mediating protection of CD4+ cells against HIV-I, binds to the external envelope glycoprotein gp120. *Immunol. Lett*. **47**, 141-143.
- Crowther R A and Blow D M, (1967) A method of positioning a known molecule in an unknown crystal structure. *Acta Cryst.* **23**, 544-548.
- Croy R R D, Hoque M S, Gatehouse J A and Boulter D, (1984) The major albumin proteins from pea (*Pisum sativum* L.) purification and some properties. *Biochem. J.* **218**, 795–803.
- Dattagupta J K, Podder A, Chakrabarti C, Sen U, Dutta S K, Singh M, (1996) Structure of a Kunitz-type chymotrypsin from winged bean seeds at 2.95 A resolution. *Acta Crystallogr* D **52**, 521–528.
- Dauter Z, (1997) Data collection strategy. *Methods Enzymol*. **276**, 326 - 344.
- 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.
- DeLano W L, (2002) The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

- Delbaere LT J, Vandonselaar M, Prasad L, Quail J W, Wilson K S and Dauter Z, (1993) Structure of the lectin IV of *Griffonia simplificolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* **230**, 950-965.
- Deng H, Chen G, Yang W and Yang J J, (2006) Predicting Calcium-Binding Sites in Proteins—A Graph Theory and Geometry Approach. *PROTEINS: Structure, Function, and Bioinformatics* **64**, 34–42
- Dharkar P D, Anuradha P, Gaikwad S M and Suresh C G, (2006) Crystallization and preliminary characterization of a highly thermostable lectin from *Trichosanthes dioica* and comparison with other *Trichosanthes* lectins *Acta Cryst.* F**62**, 205-209.
- Dharker P N , Gaikwad S M , Suresh C G, Dhuna V, Khan M I, Singh J and Kamboj S S, (2009) Comparative Studies of Two Araceous Lectins by Steady State and Time-Resolved Fluorescence and CD Spectroscopy. *Journal of Fluorescence.* **19**, 239-248.
- Dhuna V, Bains J S, Kamboj S S, Singh J and Saxena S, (2005) Purification and characterization of a lectin from *Arisaema tortuosum* Schott having in-vitro anticancer activity against human cancer cell lines. *J. Biochem. Mol. Biol.* **38**, 526-532.
- Dhuna V, Kamboj S S, Kaur A, Saxena A K, Bhide S V, Shanmugavel and Singh J, (2007) Characterization of a lectin from *Gonatanthus pumilus* D. Don having anti-proliferative effect against human cancer cell lines. *Protein. Pept. Lett.* **14**, 71-78.
- Di Sabato G, Hall J M and Thompson L, (1987) T cell mitogens and polyclonal B cell activators. *Methods Enzymol.* **150**, 3-17.
- Diaz C, Melchers L S, Hooykaas P J J, Lugtenberg B J J and Kijne J W, (1989) Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. *Nature* **338**, 579-581.
- Drenth J and Haas C, (1992) Protein crystals and their stability. *J. Cryst. Growth*, **122**, 107-109.
- Drickamer K, (1995) Multiplicity of lectin carbohydrate interactions. *Nat. Struct. Biol.* **2,** 437-439.
- Ducruix A and Giege´ R, (1992) Crystallization of Nucleic Acids and Proteins, pp. 130–135. Oxford University Press.

- Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y and Liang J, (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*. **34**, Web Server issue, doi:10.1093/nar/gkl282
- Edelman G M, Cunningham B A, Reeke G N Jr, Becker J W, Waxdal M J and Wang J L, (1972) The covalent and three-dimensional structure of concanavalin A. *P Natl Acad Sci USA*. **69**, 2580-2584.
- Elfstrand M, (1898) Über blutkörperchenagglutinierende Eiweisse. **In**; *Görberdorfer Veröffentlichungen a. Band I.* pp. 1–159. Kobert R, Ed, Enke, Stuttgart, Germany.
- Elgavish S and Shaanan B, (1998) Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides. *J. Mol. Biol*. **277,** 917-932.
- Elgavish S, and Shaanan B, (1997) Lectin-carbohydrate interactions, different folds, common recognition principles. *Trends in Biochemical Sciences*, **22**, 462-467.
- Emsley P and Cowtan K, (2004) COOT: model-building tools for molecular graphics. *Acta Cryst.* D **60**, 2126-2132.
- Endo Y, Mitsui K, Motizuki M and 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.
- Engh R A and Huber R, (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* A **47**, 392-400.
- Evans P R, (1993) Proceedings of the CCP4 Study Weekend. Data Collection and Processing, edited by L. Sawyer, N. Isaacs & S. Bailey, 114-122. Warrington: Daresbury Laboratory.
- Ezeamuzle I C, Ambadederomo A W, Shode F O and Ekwebelem S C, (1996) Antiinflammatory effects of *Moringa oleifera* root extract. *Int. J. Pharmacogn.* **34**, 207-212.
- Fahey J W, (2005) *Moringa oleifera*: A review of the medical evidence for its nutritional, therapeutic, and prophylactic properties. Part 1. Trees for Life Journal, 1:5.

- Faizi S, Siddiqui B S, Saleem R, Siddiqui S, Aftab K and Gilani A H, (1995) Fully acetylated carbonate and hypotensive thiocarbamate glycosides from *Moringa oleifera*. *Phytochemistry*. **38**, 957–963.
- Favero J, Corbeau P, Nicolas M, Benkirane M, Trave G, Dixon J F, Aucouturier P, Rasheed S, Parker J W, Liautard J P *et al*., (1993) Inhibition of human immunodeficiency virus infection by the lectin Jacalin and by a derived peptide showing a sequence similarity with gp120. *Eur. J. Immunol.* **23,** 179-185.
- Ferrer M, Barany G, and Woodward C, (1995) Partially folded, molten globule and molten coil states of bovine pancreatic trypsin inhibitor. *Nature Struct. Biol*. **2**, 211–217.
- Fülöp V and Jones D T, (1999) β-Propellers: structural rigidity and functional diversity. *Current Opinion in Structural Biology* **9**, 715–721.
- Fuglie L J, (Ed.) (2001) The Miracle Tree- The multiple attributes of *Moringa*. Technical Centre for Agricultural and Rural Cooperation (CTA)/Church World Service (CWS), New York.
- Gaastra W and Svennerholm A M, (1996) Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC) *Trends. Microbiol*. **4**, 444-452.
- Gabius H J, (1991) Detection and functions of mammalian lectins with emphasis on membrane lectins. *Biochim. Biophys. Acta.* **1071**, 1-18.
- Gabor F, Bogner E, Weissenboec A and Wirth M, (2004) The lectin-cell interaction and its implications to intestinal lectin-mediated drug delivery. *Adv. Drug. Deliv. Rev.* **56**, 459-480.
- Gage D J, (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev*. **68**, 280-300.
- Gao X, Tao W, Lu W, Zhang Q, Zhang Y, Jiang X and Fu S, (2006) Lectin - conjugated PEG-PLA nanoparticles: preparation and brain delivery after intranasal administration. *Biomaterials.* **27,** 3482-3490.
- Gassenschmidt U, Jany K D, Tauscher B and Niebergall H, (1995). Isolation and characterization of a flocculating protein from *Moringa oleifera* Lam. *Biochim. Biophys. Acta.* **1243**, 477-481.

- Gaur V, Chanana V, A Jain and Salunke D M, (2011) The structure of a haemopexin-fold protein from cow pea (*Vigna unguiculata*) suggests functional diversity of haemopexins in plants. *Acta Cryst.* F**67**, 193–200.
- Gaur V, Qureshi I A, Singh A, Chanana V and Salunke D M, (2010) Crystal Structure and Functional Insights of Hemopexin Fold Protein from Grass Pea. *Plant Physiology.* **152**, 1842–1850.
- Geigé R and McPheson A, (2001) Crystallization: General Methods. In International Tables for Crystallography, Volume F, edited by Rossmann. M. G. & Arnold, E. pp. 81-93.
- Gerfen C R and Sawchenko P E, (1985) A method for anterograde axonal tracing of chemically specified circuits in the central nervous system: combined *Phaseolus vulgaris*-leucoagglutinin (PHA-L) tract tracing and immunohistochemistry. *Brain. Res.* **343**, 144-150.
- Ghosh M, Bachhawat B K and Surolia A, (1979) A Rapid and Sensitive Assay for Detection of Nanogram Quantities of Castor-Bean (*Ricinus communis*) Lectins. *Biochem. J.* **183,** 185-188.
- Giri A P, Harsulkar A M, Deshpande V V, Sainani M N, Gupta V S and Ranjekar P K., (1998) Chickpea defensive proteinase inhibitors can be inactivated by podborer gut proteinases. *Plant Physiology.* **116**, 393–401.
- Giri A P, Harsulkar A M, Ku M S B, Gupta V S, Deshpande V V, Ranjekar P K and Franceschi V R, (2003) Identification of potent inhibitors of *Helicoverpa armigera* gut proteinases from winged bean seeds. *Phytochemistry,* **63**, 523–532.
- Goel M, Anuradha P, Kaur K J, Maiya B G, Swamy M J and Salunke D M, (2004) Porphyrin binding to jacalin is facilitated by the inherent plasticity of the carbohydrate-binding site: novel mode of lectin-ligand interaction. *Acta Crystallogr.* D **60**, 281–288.
- Gohlke U, Gomis-Rfith F-X, Crabbe T, Murphy G, Docherty A J P and Bode W, (1996) The C-terminal (haemopexin-like) domain structure of human gelatinase A (MMP2): structural implications for its function. *FEBS Letters.* **378**, 126-130.
- Gold A M and Fahrney D, (1964) Sulfonyl Fluorides as Inhibitors of Esterases. 11. Formation and Reactions of Phenylmethanesulfonyl α-Chymotrypsin. *Biochemistry* **3**, 783-791.
- Goldstein I J and Hayes C E, (1978) The Lectins: Carbohydrate-Binding Proteins of Plants and Animals. *Advances in Carbohydrate Chemistry and Biochemistry* **35**, 127-340.

- Goldstein I J and Poretz R D, (1986) Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. **In:** *The Lectins, Properties, Functions, and Applications in Biology and Medicine.* pp. 33–247. Liener IE, Sharon N, Goldstein IJ, Eds., Academic Press, Orlando, FL.
- Goldstein I J, Hughes R C, Monsigny M, Osawa T and Sharon N, (1980) What should be called a lectin? *Nature* **285**, 66.
- Gupta D, Rao N V, Puri K D, Matta K L, Surolia A, (1992) Thermodynamic and kinetic studies on the mechanism of binding of methylumbelliferyl glycosides to jacalin. *J. Biol. Chem.* **267**, 8909–8918.
- Gurjar M M, Khan M I and Gaikwad S M, (1998) α-Galactoside binding lectin from *Artocarpus hirsuta*: characterization of the sugar specificity and the binding site. *Biochim. Biophys. Acta.* **1381,** 256–264.
- Guzmán-Partida A M, Robles-Burgueño M R, Ortega-Nieblas M and Vázquez-Moreno I, (2004) Purification and characterization of complex carbohydrate specific isolectins from wild legume seeds: *Acacia constricta* is (vinorama) highly homologous to *Phaseolus vulgaris* lectins. *Biochimie.* **86**, 335-342.
- Habeeb A F S A, (1966) Determination of free amino groups in proteins by trinitrobenzenesulfonic acid. *Anal Biochem* **14**, 328-336.
- Habib H and Khalid M F, (2007) Plant protease inhibitors: a defense strategy in plants. *Biotechnology and Molecular Biology Review.* **2**, 68-85.
- Hagiwara K, Collet-Cassart D, Kobayashi K and Vaerman J P, (1988) Jacalin: isolation, characterization, and influence of various factors on its interaction with human IgA1, as assessed by precipitation and latex agglutination. *Mol. Immunol*, **24**, 69-83.
- Hamelryck T W, Dao-Thi M H, Poortmans F, Chrispeels M J, Wyns L and Loris R, (1996a) The crystallographic structure of phytohemagglutinin L. *J. Biol. Chem.* **271**, 20479-20485.
- Hamelryck T W, Poortmans F, Goossens A, Angenon G, Van Montagu M, Wyns L, and Loris R, (1996b) Crystal structure of arcelin-5, a lectin-like defense protein from *Phaseolus vulgaris. J. Biol. Chem*. **271**, 32796–32802.

- Hardman K D, Ainsworth C F, (1972) Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**, 4910–4919.
- Harley S M and Beevers H, (1986) Lectins in castor bean seedlings. *Plant Physiol.* **80**, 1–6.
- He H W, Zhang J, Zhou H M and Yan Y B, (2005) Conformational Change in the C-Terminal Domain Is Responsible for the Initiation of Creatine Kinase Thermal Aggregation. *Biophys. J.* **89**, 2650–2658.
- Hedge R, Maiti T K and 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.
- Hellin H, (1891) *Der Eiwesskorpe Abrin; Sieve Wirkung auf Blut. Dissertation* Dorpat.
- Hester G, Kaku H, Goldstein I J and Wright C S, (1995) Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat Struct. Biol*. **2**, 472–479.
- Hiemstra P S, Gorter A, Stuurman M E, Van Es L A and Daha M R, (1987) The IgA-binding lectin jacalin induces complement activation by inhibition of C-1-inactivator function. *Scand. J. Immunol.* **26**, 111-117.
- Hirsch A M, Brill L M, Lim P O, Scambray J, and Van Rhijn P, (1995) Steps toward defining the role of lectins in nodule development in legumes. *Symbiosis.* **19**, 155-174.
- Horejsí V and Kocourek J, *(*1974) Studies on phytohemagglutinins XVIII. Affinity electrophoresis of phytohemagglutinins. *Biochim. Biophys. Acta* **336***,* 338-343.
- Hortin G L and Trimpe B L, (1990) Lectin affinity chromatography of proteins bearing O-linked oligosaccharides: application of jacalin-agarose. *Anal. Biochem*. **188**, 271-277.
- http://www.cermav.cnrs.fr/lectines/
- Huesing J E, Murdock L and Shade R E, (1991) Effect of wheat germ isolectins on development of cowpea weevil. *Phytochemistry* **30**, 785-788.
- Hunter C A, Packer M J and Zonta C, (2005) From structure to chemical shift and vice-versa. *Prog. Nucl. Magn. Reson. Spectrosc.* **47**, 27–39.

- Ibarra-Molero B, Sanchez-Ruiz J M, (2006) Differential Scanning Calorimetry of Proteins: an Overview and Some Recent Developments. Springer Series in Biophysics. J.L.R. Arrondo and A. Alonso Advanced Techniques in Biophysics © Springer-Verlag Berlin Heidelberg, pp-27-48.
- Imberty A, Gautier C, Lescar J, Perez S, Wyns L and Loris R, (2000) An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **275,** 17541-17548.
- Imberty A, Mitchell E P and Wimmerova M, (2005) Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Current Opinion in Structural Biology*. **15**, 525–534.
- Iwanaga S, Nagata R, Miyamoto A, Kouzuma Y, Yamasaki N and Kimura M, (1999) Conformation of the primary binding loop folded through an intramolecular interaction contributes to the strong chymotrypsin inhibitory activity of the chymotrypsin inhibitor from *Erythrina variegata* seeds. *J Biochem.* **126,** 162–167.
- Jaiswal D, Kumar Rai P, Kumar A, Mehta S and Watal G, (2009) Effect of *Moringa oleifera* Lam. leaves aqueous extract therapy on hyperglycemic rats. *J. Ethnopharmacol.* **123**, 392-396.
- Jeyaprakash A A, Geetha Rani P, Banuprakash Reddy G, Banumathi S, Betzel C, Sekar K, Surolia A and Vijayan M, (2002) Crystal structure of the jacalin-T-antigen complex and a comparative study of lectin-T-antigen complexes. *J. Mol. Biol*. **321**, 637–645. Jeyaprakash A A, Katiyar S, Swaminathan C P, Sekar K, Surolia A and Vijayan M, (2003) Structural basis of the carbohydrate specificities of jacalin: an X-ray and modeling study. *J. Mol. Biol.* **332**, 217–228.
- Jeyaprakash A A, Jayashree G, Mahanta S K, Swaminathan C P, Sekar K, Surolia A and Vijayan M, (2005) Structural basis for the energetics of jacalin-sugar interactions: promiscuity versus specificity. *J. Mol. Biol*. **347**, 181–188.
- Jongsma M A, Bakker P L, Peters J, Bosch D and Stiekema W J, (1995) Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. *Proc. Natl. Acad. Sci.* (USA) **92**, 8041-8045.
- Jouanin L, Bonade-Bottino M, Girard C, Morrot G and Giband M, (1998) Transgenic plants for insect resistance. *Plant Science*. **131**, 1-11.

- Kabir S, (1993) Simultaneous isolation of intestinal IgA and IgG from rabbits infected intraduodenally with *Vibrio cholerae* 01 by combined lectin affinity chromatography involving jacalin and protein A. *Comp. Immunol. Microbiol. Infect. Dis.* **16**, 153-161.
- Kabir S, (1998) Jacalin: a jackfruit (*Artocarpus heterophyllus*) seed-derived lectin of versatile applications in immunobiological research. *J. Immunol. Methods*. **212**, 193–211.
- Kamemura K, Furuichi Y, Umekawa H and Takahashi T, (1993) Purification and characterization of novel lectins from Great Northern bean, *Phaseolus vulgaris* L. *BBA-Gen. Subjects.* **1158**, 181-188.
- Kaneda Y, Whittier R F, Yamanaka H, Carredano E, Gotoh M, Sota H, Hasegawa Y and Shinohara Y, (2002) The high specificities of *Phaseolus vulgaris* erythro- and leukoagglutinating lectins for bisecting GlcNAc or β1-6-linked branch structures, respectively, are attributable to loop B. *J. Biol. Chem.* **277**, 16928-16935.
- Katre U V and Suresh C G, (2009) Features of homotetrameric molecular association in protein crystals. *Acta Cryst*. D**65**, 1-10.
- Katre U V, Gaikwad S M, Bhagyawant S S, Deshpande U D, Khan M I and Suresh C G (2005) Crystallization and preliminary X-ray characterization of a lectin from *Cicer arietinum* (chickpea). *Acta Cryst.* **F61**, 141-143.
- Katre U V, Ph.D thesis, (2007) Structural studies on two hemagglutinins from *Cicer arietinum* and *Moringa oleifera,* and a study of polymorphism in the crystals of plant lectins*,* University of Pune, Maharastra, India.
- Katre U V, Suresh C G, Khan M I and Gaikwad S M, (2008a) Structure–activity relationship of a hemagglutinin from *Moringa oleifera* seeds. *International Journal of Biological Macromolecules* **42**, 203-207.
- Katre U V, Suresh C G, Khan M I and Gaikwad S M, (2008b) Steady State and Time-Resolved Fluorescence Studies of a Hemagglutinin from *Moringa oleifera*. *J. Fluoresc*. **18**, 479–485
- Kaur A, Kamboj S S, Singh J, Saxena A K and Dhuna V, (2005a) Isolation of a novel *N*-acetyl-D-lactosamine specific lectin from *Alocasia cucullata* (Schott.) *Biotechnol. Lett*. **27**, 1815-1820.

- Kaur A, Singh J, Kamboj S S, Sexana A K, Pandita R M and Shamnugavel M, (2005b) Isolation of an *N*-acetyl-D-glucosamine specific lectin from the rhizomes of *Arundo donax* with antiproliferative activity. *Phytochemistry* **66**, 1933-1940.
- Kaur M, Singh K, Rup P J, Kamboj S S, Saxena A K, Sharma M, Bhagat M, Sood S K and Singh J, (2006b) A tuber lectin from *Arisaema jacquemontii* Blume with anti-insect and anti-proliferative properties. *J. Biochem. Mol. Biol.* **39**, 432-440.
- Kaur M, Singh K, Rup P J, Saxena A K, Khan RH, Ashraf M T, Kamboj S S and Singh J, (2006a) A tuber lectin from *Arisaema helleborifolium* Schott with anti-insect activity against melon fruit fly, *Bactrocera cucurbitae* (Coquillett) and anti-cancer effect on human cancer cell lines. *Arch. Biochem. Biophys.* **445**, 156-165.
- Kaur N, Dhuna V, Kamboj S S, Agrewala J N and Singh J, (2006c) A novel antiproliferative and antifungal lectin from *Amaranthus viridis* Linn seeds. *Protein. Pept. Lett.* **13**, 897-905.
- Khamrui S, Majumder S, Dasgupta J, Dattagupta J K and Sen U, (2010) Identification of a novel set of scaffolding residues that are instrumental for the inhibitory property of Kunitz (STI) inhibitor. *Protein science*. **19**, 593—602.
- Kieliszewski M J, Showalter A M and Leykam J F, (1994) Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). *Plant J.* **5**, 849-861.
- Kim J-Y, Park S-C, Hwang I, Cheong H, Nah J-W, Hahm K-S and Park Y, (2009) Protease Inhibitors from Plants with Antimicrobial Activity. *Int. J. Mol. Sci.*, **10**, 2860-2872.
- Kochhar S, Gartenmann K, Juillerat M A, (2000) Primary structure of the abundant seed albumin of *Theobroma cacao* by mass spectrometry. *J. Agric. Food Chem.* **48**, 5593–5599.
- Kocourek J and Horejsi V, (1983). A note on the recent discussion on definition of the term 'lectin'. **In**: *Lectins: Biology, Biochemistry and Clinical Biochemistry*. **3**:3–6. Bøg- Ha nsen, T C and Spengler, G A, Eds., Walter de Gruyter, Berlin, Germany.
- Koiwa H, Bressan R A, Hasegawa P M, (1997) Regulation of protease inhibitors and plant defense. *Trends. Plant. Sci.* **2**, 379–384.

- Kolberg J, Michaelsen T E and Sletten K (1983) Properties of a lectin purified from the seeds of *Cicer arietinum. Hoppe-Seyler's Z Physiol. Chem.* **364**, 655-664.
- Komath S S, Kavitha M and Swamy M J, (2006) Beyond carbohydrate binding: new directions in plant lectin research. *Org. Biomol. Chem.* **4**, 973-988.
- Komath S S, Nadimpalli S K and Swamy M J, (1996) Purification in high yield and characterisation of the galactose-specific lectin from the seeds of snake gourd (*Trichosanthes anguina*). *Biochem. Mol. Biol. Int.* **39**, 243-252.
- Kondoh H, Kobayashi K, Hagiwara K and Kajii T, (1986) Jacalin, a jackfruit lectin, precipitates IgA1 but not IgA2 subclass on gel diffusion reaction. *J. Immunol. Methods.* **88**, 171-173.
- Konkumnerd W, Karnchanatatb A and Sangvanichc P, (2010) A thermostable lectin from the rhizomes of *Kaempferia parviflora. J. Sci. Food Agric.* **90**, 1920–1925.
- Kostla´nova´ N, Mitchell E P, Lortat-Jacob H , Oscarson S, Lahmann M , Gilboa-Garber N, Chambat G, Wimmerova M and Imberty A, (2005) The Fucose-binding Lectin from *Ralstonia solanacearum.* A new type of β-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J. Biol. Chem.* **280**, 27839–27849.
- Krissinel E and Henrick K, (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-794.
- Kristina E, Veen M T v, Murray J M and Aguirre G D, (1991) Rod and cone specific domains in the interphotoreceptor matrix. *The Journal of Comparative Neurology*. **308,** 371-380.
- Kumar A M, Timm D E, Neet K E, Owen W G, Peumans W J and Rao A J, (1993) Characterization of the lectin from the bulbs of *Eranthis hyemalis* (winter aconite) as an inhibitor of protein synthesis. *J. Biol. Chem.* **268**, 25176–25183.
- Kumar G S, Appukuttan P S and Basu D, (1982) α-D-Galactose- specific lectin from jackfruit (*Artocarpus integra*) seed. *J. Biosci.* **4**, 257.
- Ladokhin A S, (2006) Fluorescence Spectroscopy in Peptide and Protein Analysis: onlinelibrary.wiley.com/doi/10.1002/9780470027318.a1611/full.

- Lafont V, Dornand J, Covassin L, Liautard, J P and Favero J, (1996) The lectin jacalin triggers CD4-mediated lymphocyte signaling by binding CD4 through a protein-protein interaction. *J. Leukoc. Biol*. **59**, 691-696.
- Lafont V, Dornand J, d'Angeac A D, Monier S, Alcover A and Favero J, (1994) Jacalin, a lectin that inhibits in vitro HIV-1 infection, induces intracellular calcium increase via CD4 in cells lacking the CD3/TcR complex. *J. Leukoc. Biol*. **56,** 521-524.
- Lakowicz E M and Weber G, (1973) Quenching of protein fluorescence by oxygen. Detection of structural fluctuations in proteins on the nanosecond time scale. *Biochemistry* **12**, 4171–4179.
- Lamb F I, Roberts L M, and Lord J M, (1985) Nucleotide sequence of cloned cDNA coding for preproricin. *Eur. J. Biochem.* **148**, 265–270.
- Landsteiner K and Raubitschek H (1907) Beobachtungen über Hämolyse und Hämagglutination. *Zentralbl Bakteriol Parasitenk Infektionskr Hyg Ab. 1: Orig* **45**, 660–667.
- Lange C and Rudolph R, (2009) Suppression of protein aggregation by L-arginine. *Curr. Pharm. Biotechnol.* **10**, 408-414.
- Laskowski M Jr and Qasim M A, (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochem. Biophys. Acta.* **1477**, 324–337.
- Laskowski R A, MacArthur M W, Moss D S and Thoronton J M, (1993) PROCHECK: A program to check the stereo-chemical quality of protein structures. *J. Appl. Crystllogr.* **26**, 283-291.
- Laskowski R A, McArthur M W, Moss D S and Thornton J M, (1993) PROCHECK: a program to check the stereo-chemical quality of protein structures. *J. App. Cryst.* **26**, 283-291.
- Laskowsky M Jr and Kato I, (1980) Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593–626.
- Lee X, Thompson A, Zhang Z, Ton-that H, Biesterfeldt J, Ogata C, Xu L, Johnston R A and 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.

- Lehrer S S and Leavis P C, (1978) Solute quenching of protein fluorescence. *Methods Enzymol.* **49**, 222–236.
- Leslie A G W, (1992) Jnt CCP4/ESF±EAMCB Newsl. Protein Crystallogr., 26.
- Leung D, Abbenante G, Fairlie D P, (2000) Protease Inhibitors: Current Status and Future Prospects. *J. Med. Chem.* **43**, 305-341.
- LeVine D, Kaplan M J and Greenaway P J, (1972) The purification and characterization of wheat-germ agglutinin. *Biochem. J.* **129**: 847–856.
- Li J, Brick P, O'Hare M V C, Skarzynski T, Lloyd L F, Curry V A, Clark I M, Bigg H F, Hazleman B L, Cawston T E and Blow D M, (1995) Structure off full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed β-propeller. *Structure* **3**, 541–549.
- Liener I E, Sharon N and Goldstein I J eds (1986) *The Lectins: Properties, Functions and Applications in Biology and Medicine*, Academic Press, Orlando, FLA.
- Lis H and Sharon N, (1998) Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **98**, 637-674.
- Lord J M, Roberts L M and Robertus J D, (1994) Ricin: Structure, mode of action, and some current applications. *FASEB. J.* **8**, 201-208.
- Loris R, De Greve H, Dao-Thi M -H, Messens J, Imberty A and Wyns L, (2000) Structural basis of carbohydrate recognition by Lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate binding site. *J. Mol. Bio.* **301**, 987-1002.
- Loris R, Hamelryck T, Bouckaert J and Wyns L, (1998) Legume lectin structure. *Biochim. Biophys. Acta.* **1383**, 9-36.
- Luo Y, Xu X, Liu J, Li J, Sun Y, Liu Z, Liu J, Van Damme E, Balzarini J and Bao J, (2007) A Novel Mannose-binding Tuber Lectin from *Typhonium divaricatum (L.)* Decne (family Araceae) with Antiviral Activity Against HSV-II and Anti-proliferative Effect on Human Cancer Cell Lines. *Journal of Biochemistry and Molecular Biology* **40**, 358-367.

- Mahajan S G and Mehta A A, (2007).Inhibitory Action of Ethanolic Extract of Seeds of *Moringa oleifera* Lam. On Systemic and Local Anaphylaxis. *J. Immunotoxicol.* **4**, 287-294.
- Mahanta S K, Sastry M V and Surolia A, (1990) Topography of the combining region of a Thomsen-Friedenreich-antigen-specific lectin Jacalin (*Artocarpus integrifolia* agglutinin). A thermodynamic and circulardichroism spectroscopic study. *Biochem. J*. **265**, 831–840.
- Maikokera R and Kwaambwa H M, (2007) Interfacial properties and fluorescence of a coagulating protein extracted from *Moringa oleifera* seeds and its interaction with sodium dodecyl sulphate. *Colloid Surface. B* **55**:173-178.
- Malhotra R S, Pundir R P S and Slinkard AE, (1987) Genetic resources of chickpea. 67-81. In: M. C. Saxena and K.B. Singh (Eds), The Chickpea. C. A. B. International Cambrian News Ltd, Aberystwyth, UK.
- Manandhar N P, (2002) Plants and People of Nepal Timber Press. Oregon. ISBN 0-88192-527-6.
- Matthews B W, (1968) The Solvent Content of Protein Crystals. *J. Mol. Biol.* **33**, 491-497.
- McCoy A J, (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Cryst.* D**63**, 32-41.
- McCoy A J, Grosse-Kunstleve R W, Storoni L C and Read R J, (2005) Likelihood enhanced fast translation functions. *Acta Cryst.* D **61**, 458-464.
- McPherson A, (1982) The preparation and analysis of protein crystals. (John Wiley and Sons, New York).
- Meester P de, Brick P, Lloyd L F, Blow D M and Onesti S, (1998) Structure of the Kunitz-Type Soybean Trypsin Inhibitor (STI): Implication for the Interactions between Members of the STI Family and Tissue-Plasminogen Activator. *Acta. Cryst.* D **54**, 589-597.
- Montfort W, Villafranca J E, Monzingo A F, Ernst S, Katzin B, Rutenber E, Xuong N H, Hamlin R and Robertus J D, (1987) The three-dimensional structure of ricin at 2.8 Å. *J. Biol. Chem.* **262**, 5398–5403.
- Morozova L A, Haynie D T, Arico-Muendel C, Van Dael H and Dobson C M, (1995) Structural basis of the stability of a lysozyme molten globule. *Nature Struct. Biol.* **2**, 871–875.

- Morton J F, (1991) The horseradish tree, *Moringa pterygosperma* (Moringaceae) - a boon to arid lands? *Econ. Bot.* **45**, 318-333.
- Mosolov V V and Valueva T A, (2005) Proteinase inhibitors and their function in plants: a review. *Applied Biochemistry and Microbiology.* **41**, 227-246.
- Murakami A, Kitazono Y, Jiwajinda S, Koshimizu K and Ohigashi H, (1998) Niaziminin, a thiocarbamate from the leaves of *Moringa oleifera*, holds a strict structural requirement for inhibition of tumorpromoter- induced Epstein–Barr virus activation. *Planta Med.* **64**, 319-323.
- Murdock L L, Huesing J E, Nielsen S S, Pratt R C and Shade R E, (1990) Biological effects of plant lectins on the cowpea weevil. *Phytochemistry* **29**, 85-89.
- Murshudov G N, Dodson E J and Vagin A A, (1996) Application of maximum likelihood methods for macromolecular refinement. Proceedings of the CCP4 Study Weekend (Macromolecular Refinement), 93-104.
- Murshudov G N, Vagin A A and Dodson J, (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta, Cryst.* D **53**, 240-255.
- Murzin A G, Lesk A M and Chothia C, (1992) β-trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors, interleukins-1β, and 1α and fibroblast growth factors. *J. Mol. Biol.* **223**, 531–543.
- Nagata Y and Burger M M, (1972) Wheat germ agglutinin. Isolation and crystallization. *J. Biol.Chem.* **247**: 2248–2250.
- Nakamura S, Tanaka K and Murakawa S, (1960) Specific Protein of Legumes which reacts with Animal Proteins. *Nature* **188,** 144-145.
- Navaza J, (1994) AMoRe: an automated package for molecular replacement. *Acta. Cryst.* A **50**, 157-163.
- Ndabigengesere A, Narasiah K S and Talbot B G, (1995) Active agents and mechanism of coagulation of turbid waters using *Moringa oleifera. Water. Res.* **29**, 703-710.
- Nesamani S, (1999) Medicinal Plants (vol. I). State Institute of Languages, Thiruvananthapuram, Kerala, India. 425.

- Odani S and Ikenaka T, (1973) The amino acid sequences of two soybean double headed proteinase inhibitors and evolutionary consideration on the legume proteinase inhibitors. *J. Biochem.* **74**, 697-715
- Ofek I and Doyle R J, (1994) Bacterial Adhesion to Cells and Tissues; Chapman and Hall: London. 578.
- Ofek I and Sharon N, (1990) Adhesins as lectins: specificity and role in infection. *Curr. Top. Microbiol. Immunol.* **152**, 91-113.
- Ohta M, Kawahara N, Liu M, Taketa K, Kudo T and Taga H, (1998) Developmental Alterations of α-Fetoprotein Sugar Chain in Amniotic Fluids Analyzed by Lectin Affinity Electrophoresis. *Acta. Med. Okayama.* **52,** 27-33.
- O'Keefe D and Ashman L, (1982) Peanut agglutinin: a marker for normal and leukaemic cells of the monocyte lineage. *Clin. Exp. Immunol.* **48**, 329-338.
- Onesti S, Brick P and Blow D M, (1991) Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.* **217**, 153–176.
- Osborne T B, (1924) In: The vegetable proteins., Longmans, Green & Co. eds., Vol. **2**. London, pp. 154.
- Otwinowski Z and Minor W, (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 307-326.
- Ovadi J, Libor S and Elodi P, (1967) Spectrophotometric determination of histidine in protein with diethylpyrocarbonate. *Acta Biochem. Biophys.* (Budapest) **2**, 455-458.
- Padma P, Komath S S, Nadimpalli S K and Swamy M J, (1999) Purification in high yield and characterisation of a new galactose-specific lectin from the seeds of *Trichosanthes cucumerina*. *Phytochemistry* **50**, 363-371.
- Paoli M, (2001) Protein folds propelled by diversity. *Progress in Biophysics & Molecular Biology* **76**, 103–130.
- Paoli M, Anderson B F, Baker H M, Morgan W T, Smith A and Baker E N, (1999) Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two -propeller domains. *Nature Structural & Molecular Biology* **6**, 926 – 931.

- Pari L and Kumar N A, (2002) Hepatoprotective activity of *Moringa oleifera* on antitubercular drug-induced liver damage in rats. *J. Med. Food.* **5**, 171-177.
- Pedroche J, Yust M M, Lqari H, Megı´as C, Giro´n-Calle J, Alaiz M, Milla´n F and Vioque J, (2005) Chickpea pa2 albumin binds hemin. *Plant Science* **168,** 1109–1114.
- Peumans W J and Van Damme E J M, (1995) Lectins as plant defense proteins. *Plant. Physiol.* **109**, 347-352.
- Peumans W J and 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, De Ley M and Broekaert W F, (1984) An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. *FEBS Lett*. **177**, 99-103.
- Peumans W J, Hause B and Van Damme E J M, (2000) The galactose-binding and mannose-binding jacalin-related lectins are located in different sub-cellular compartments *FEBS Lett.* **477**, 186–192.
- Peumans W J, Kellens J T, Allen A K and Van Damme E J M, (1991) Isolation and characterization of a seed lectin from elderberry (*Sambucus nigra* L.) and its relationship to the bark lectins. *Carbohydr. Res.* **213**, 7–17.
- Pidcock E and Geoffrey R M, (2001) Structural characteristics of protein binding sites for calcium and lanthanide ions. *J. Biol. Inorg. Chem.* **6**, 479-489.
- Pillai K R, Remani P, Kannan S, Sujathan K, Mathew B, Vijayakumar T, Nair M K and Menon V P, (1996) Lectin histochemistry of oral premalignant and malignant lesions: correlation of JFL and PNA binding pattern with tumour progression. *Eur. J. Cancer B. Oral. Oncol*. **32B**, 32-37.
- Piller V, Piller F and Cartron J-P, (1986) Isolation and characterization of an *N*-acetyl galactosamine specific lectin from *Salvia sclarea* seeds. *J. Biol. Chem.* **261**, 14069-14075.
- Pineau N, Aucouturier P, Brugier J C and Preud'homme J L, (1990) Jacalin: a lectin mitogenic for human CD4 T lymphocytes. *Clin Exp Immunol*. **80**, 420–425.
- Pineau N, Brugier J C, Breux J P, Becq-Giraudon B, Descamps J M, Aucouturier, P and Preud'homme J L, (1989) Stimulation of peripheral blood lymphocytes of HIV-infected patients by jacalin, a lectin mitogenic for human CD4+ lymphocytes. *AIDS*. **3**, 659-663.

- Poulsen F M, (2002) A brief introduction to NMR spectroscopy of proteins.
- Powell H R, (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Cryst.* D **55**, 1690-1695.
- Pratap J V, Jeyaprakash A A, Geetha P R, Sekar K, Surolia A and Vijayan M, (2002) Crystal Structures of Artocarpin, a Moraceae Lectin with Mannose Specificity, and its Complex with Methyl-a-D-mannose: Implications to the Generation of Carbohydrate Specificity. *J. Mol. Biol.* **317**, 237-247.
- Ptitsyn O B, (1995) Molten Globule and Protein Folding. *Adv. Protein. Chem.* **47**, 83–229.
- Qi R F, Song Z and Chi C, (2005) Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application: *Acta Biochemica. Et. Biophysica*. **37**, 283-292.
- Qureshi I A, Prasanta D, Srivastava P S and Koundal K R, (2006) Purification and characterization of an N-acetyl-D-galactosamine-specific lectin from seeds of chickpea (*Cicer arietinum* L.). *Phytochem. Anal.* **17**, 350–356.
- Raikhel N V and Wilkins T A, (1987) Isolation and characterization of a cDNA clone encoding wheat germ agglutinin. *Proc. Natl. Acad. Sci. USA* **84**, 6745–6749.
- Raikhel N V, Lee H I and Borekaert W F, (1993) Structure and function of chitin binding proteins. *Ann. Rev. Plant. Phys.* **44**, 591-615.
- Ramachandran G N and Sasisekharan V, (1968) Conformation of polypeptides and proteins. *Adv. Prot. Chem.* **23**, 283–438.
- Rao K N and Suresh C G, (2007) Bowman–Birk protease inhibitor from the seeds of *Vigna unguiculata* forms a highly stable dimeric structure. *Biochimica. et Biophysica. Acta.* **1774**, 1264–1273.
- Rao K N, Suresh C G, Katre U V, Gaikwad S M and Khan M I, (2004) Two orthorhombic crystal structures of a galactose-specific lectin from Artocarpus hirsuta in complex with methyl-alpha-D-galactose. *Acta Crystallogr.* D **60,** 1404-1412.
- Raval S, Gowda S B, Singh D D and Chandra N R (2004). A database analysis of jacalin-like lectins: sequence–structure–function relationships. *Glycobiology* **14**, 1247-1263.

- Ravishankar R, Ravindran M, Suguna K, Surolia A and Vijayan M, (1997) Crystal structure of the peanut lectin–T-antigen complex. Carbohydrate specificity generated by water bridges. *Curr. Sci.* **72**, 855-861.
- Ray S, Ahmed H, Basu S and Chatterjee B P, (1993) Purification, characterization, and carbohydrate specificity of the lectin of *Ficus cunia*. *Carbohydr. Res.* **242**, 247-263.
- Ray S, and Chatterjee B P, (1995) Saracin: a lectin from *Saraca indica* seed integument recognizes complex carbohydrates. *Phytochemistry* **40**, 643-649.
- Reddy G B, Srinivas V R, Ahmad N and Surolia A, (1999) Molten Globule-like State of Peanut Lectin Monomer Retains Its Carbohydrate Specificity. *Journal Bio. Chem.* **274**, 4500–4503.
- Redfield C, Smith R A G and Dobson C M, (1994) Structural characterization of a highly ordered molten globule at low pH. *Nature Struct. Biol.* **1**, 23–39.
- Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S., and Dobson, C. M, (1999) α-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat. Struct. Biol.* **6**, 948–952.
- Reisner Y, (1987) *Progress in Bone Marrow Transplantation* (Gale, R.P. & Champlin, R., eds), pp. 175-183, UCLA symposium on Molecular and Cellular Biology, New Series, **53**, Alan, R. Liss, New York.
- Remani P, Augustine J, Vijayan K K, Ankathil R, Vasudevan D M, Nair M K and Vijayakumar T, (1989) Jack fruit lectin binding pattern in benign and malignant lesions of the breast. *In Vivo*, **3**, 275-278.
- Renkonen K O, (1948) Studies on hemagglutinins present in seeds of some representatives of Leguminoseae. *Ann. Med. Exp. Biol. Fenn.* **26**, 66-72.
- Reshetnyak Y K, Koshevnik Y and Burstein E A, (2001) Decomposition of Protein Tryptophan Fluorescence Spectra into Log-Normal Components. III. Correlation between Fluorescence and Microenvironment Parameters of Individual Tryptophan Residues. *Biophysical. Journal.* **81**, 1735–1758.
- Rinderle S J, Goldstein I J and Remsen E E, (1990) Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds. *Biochemistry* **29**, 10555-10561.

- Rinderle S J, Goldstein I J, Matta K L and Ratcliffe R M, (1989) Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus caudatus*, that recognizes the T- (or cryptic T)- antigen. *J. Biol. Chem.* **264**, 16123-16131.
- Rini J M, (1995) Lectin Structure. *Annu. Rev. Biophys. Biomol. Struct.*.**24**, 551-77.
- Riordan J F, Wacker W E C and Vallee B L, (1965) *N*-Acetylimidazole: A Reagent for Determination of "Free" Tyrosyl Residues of Proteins. *Biochemistry.* **4**, 1758-1765.
- Roberts D D and Goldstein I J, (1983a) Binding of hydrophobic ligands to plant lectins: Titration with arylaminonaphthalenesulfonates. *Arch. Biochem. Biophys.* **224**, 479-484.
- Roberts D D and Goldstein I J, (1983b) Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J Biol. Chem.* **258**, 13820-13824.
- Rossmann M G and Blow D M, (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.
- Rüdiger H and Gabius H-J, (2001) Plant lectins: Occurrence, biochemistry, functions and applications. *Glycoconjugate J.* **18**, 589-613.
- Ruffet E, Paquet N, Frutiger S, Hughes G J and Jaton J C, (1992) Structural and electron-microscopic studies of jacalin from jackfruit (*Artocarpus integrifolia*) show that this lectin is a 65 kDa tetramer. *Biochem J.* **286**, 131-134.
- Rutenber E and Robertus J D, (1991) Structure of Ricin B-chain at 2.5 Å resolution. *Proteins* **10**, 260–269.
- Rutenber E, Katzin B J, Collins E J, Mlsna D, Ready M P and Robertus J D, (1991). Crystallographic refinement of Ricin to 2.5 Å. *Proteins* **10**, 240-250.
- Ryan C A, (1990) PROTEASE INHIBITORS IN PLANTS: Genes for Improving Defenses against Insects and Pathogens *Annu. Rev. Phytolpathol.* **28**, 425–449.
- Saha R P, Bahadur R P, Pal A, Mandal S and Chakrabarti P, (2006) ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Structural Biology*. **6**:11, doi:10.1186/1472-6807-6-11.

- Sahasrabuddhe A A, Gaikwad S M, Krishnasastry M V and Khan M I, (2004) Studies on recombinant single chain Jacalin lectin reveal reduced affinity for saccharides despite normal folding like native Jacalin. *Protein Sci.* **13**, 3264-3273.
- Sanchez-Ruiz J M, (1992) A theoretical analysis of Lumry–Eyring models in differential scanning calorimetry. *Biophys. J.* **61**, 921–935.
- Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A and Vijayan M, (1996) A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold. *Nat. Struct. Biol.* **3**, 596-603.
- Schlereth A, Standhardt D, Mock H P and Mu¨ ntz K, (2001) Stored cysteine proteinases start globulin mobilization in protein bodies of embryonic axes and cotyledons during vetch (*Vicia sativa* L.) seed germination. *Planta* **212**, 718–727.
- Shaanan B, Lis H and Sharon N, (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862-866.
- Shangary S, Jatinder S, Sukhdev S K, Kulwant K K and Rajinder S S, (1995) Purification and properties of four monocot lectins from the family Araceae. *Phytochemistry* **40**, 449-455.
- Sharma V and 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 and Lis H (1989a) in *Lectins* (Chapman and Hall, London).
- Sharon N and Lis H, (1972) Lectins: cell-agglutinating and sugar-specific proteins. *Science* **177,** 949-959.
- Sharon N and Lis H, (1990) Legume lectins –a large family of homologous proteins. The *FASEB. J.* **4**, 3198-3208.
- Sharon N and Lis H, (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **14,** 53-62.
- Sharon N, (1993) Lectin-carbohydrate complexes of plants and animals: An atomic view. *Trends Biochem. Sci.* **18**, 221-226.
- Sharon N, (2005) A life with lectins. *Cell. Mol. Life. Sci.* **62,** 1057-1062.
- Sharon, N and Lis, H (1989) Lectins as cell recognition molecules. *Science* **246,** 227-234.

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

- Shiraki K, Kudou M, Fujiwara S, Imanaka T and Takagi M, (2002) Biophysical Effect of Amino Acids on the Prevention of Protein Aggregation. *J. Biochem.* **132**, 591-595.

- Siddiqui M Z, Sharma A K and Kumar S, (1996) Solution conformation of tuftsin. *Int. J. Biol. Macromol.* **19**, 99–102

- Singh Bains J, Singh J, Kamboj S S, Nijjar K K, Agrewala J N, Kumar V, Kumar A and Saxena A K, (2005) Mitogenic and anti-proliferative activity of a lectin from the tubers of Voodoo lily (*Sauromatum venosum*). *Biochim. Biophys. Acta.* **1723**, 163-174.

- Singh J, Kamboj S S, Sandhu R S, Shangary S and Kamboj K K, (1993) Purification and characterization of a tuber lectin from *Alocasia indica*. *Phytochemistry.* **33**, 979–983.

- Singh J, Singh J and Kamboj S S, (2004) A novel mitogenic and antiproliferative lectin from a wild cobra lily, *Arisaema flavum. Biochem. Bioph. Re.s Co.* **318**, 1057–1065.

- Singh T, Wu J H, Peumans W J, Rougé P, Van Damme E J M, Alvarez R A, Blixt O, Wu A M, (2006) Carbohydrate specificity of an insecticidal lectin isolated from the leaves of *Glechoma hederacea* (ground ivy) towards mammalian glycoconjugates. *Biochem. J.* **393**, 331-341.

- Singha B, Adhya M, Chatterjee B P, (2007) Multivalent II [β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc] and Tα [β-D-Gal*p*-(1→3)-α-D-Gal*p*NAc] specific Moraceae family plant lectin from the seeds of *Ficus bengalensis* fruits. *Carbohydr. Res.* **342**, 1034-1043.

- Sinha U and Brewer J M, (1985) A spectrophotometric method for quantitation of carboxyl group modification of proteins using Woodward's Reagent K. *Anal. Biochem.* **151**, 327-333.

- Slifkin M and Doyle R J, (1990) Lectins and their application to clinical microbiology. *Clin. Microbiol. Rev*. **3**, 197–218.

- Song H K and Suh S W, (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J Mol. Biol.* **275**, 347–363.

- Spande T F and Witkop B, (1967) Determination of the tryptophan content of proteins with N-bromosuccinimide. *Methods Enzymol.* **11**, 498–506.
- Sreelatha S, Jeyachitra A and Padma P R, (2011) Antiproliferation and induction of apoptosis by *Moringa oleifera* leaf extract on human cancer cells. *Food. Chem. Toxicol.* doi:10.1016/j.fct.2011.03.006
- Sreerama N, Venyaminov S Y, Woody R W, (1999) Estimation of the number of helical and strand segments in proteins using circular dichroism spectroscopy, *Protein Sci.* **8**, 370–380.
- Srinivasan A, Giri A P, Harsulkar A M, Gatehouse J A and Gupta V S, (2005) A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae. *Plant Molecular Biology*. **57**, 359–374.
- Stillmark. H, (1888) Uber Ricin ein giftiges Fermentaus den Samen von Ricinus communis. L. und einige anderen Euphorbiaceen. Inagural Disseration Dorpat.
- Storoni LC, McCoy A J and Read R J, (2004) Likelihood-enhanced fast rotation functions. *Acta Cryst.* **D** 60, 432-438.
- Su Jing-Tan, Kim Sung-Hye and Yong-Bin Y, (2007) Dissecting the Pretransitional Conformational Changes in Aminoacylase I Thermal Denaturation. *Biophys. J.* **9,** 578–587.
- Sudmoon R, Sattayasai N, Bunyatratchata W, Chaveerach A and Nuchadomrong S, (2008) Thermostable mannose-binding lectin from *Dendrobium findleyanum* with activities dependent on sulfhydryl content Acta *Biochim. Biophys. Sin.* **20**, 811-818.
- Sujathan K, Kannan S, Remani P, Pillai K R, Chandralekha B, Amma N S and Nair M K, (1996) Differential expression of jackfruit–lectin-specific glycoconjugates in metastatic adenocarcinoma and reactive mesothelial cells-a diagnostic aid in effusion cytology. *J. Cancer Res. Clin. Oncol*. **122**, 433-436.
- Sultan N A M, Kenoth R and Swamy M J, (2004) Purification, physicochemical characterization, saccharide specificity, and chemical modification of a Gal/GalNAc specific lectin from the seeds of *Trichosanthes dioica*. *Arch. Biochem. Biophys*. **432**, 212-221.

- Sumner J B and Howell S F, (1936) The identification of the hemagglutinin of the Jack bean with concanavalin A. *J. Bacteriol.* **32**, 227–237.
- Supuran C T, Scozzafava A and Clare B W, (2002) Bacterial protease inhibitors. *Med. Res. Rev.* **22**, 329-72.
- Suresh A S, Rani P G, Pratap J V, Sankaranarayanan R, Surolia A and Vijayan M, (1997). *Acta Crystallog*. D**53**, 469–471.
- Swindells M B and Thornton J M, (1993) A study of structural determinants in the interleukin-1 fold. *Protein Eng.* **6**, 711-715.
- Szilagyi L, (1995) Chemical shifts in proteins come of age. *Prog. Nucl. Magn. Reson. Spectrosc.* **27**, 325–443.
- Tahirov T H, Lu T H, Liaw Y C, Chen Y L and Lin J Y (1995) Crystal structure of Abrin-A at 2.14 Å. *J. Mol. Biol.* **250**, 354-367.
- Takahashi K, (1968) The reaction of phenylglyoxal with arginine residues in proteins. *J. Biol. Chem.* **243**, 6171-6179.
- Tamma S M, Oyaizu N, McCloskey T W, Kalyanaraman V S and Pahwa S, (1996) HIV-1 gp 120 blocks jacalin-induced proliferative response in CD4+ T cells: jacalin as a useful surrogate marker for qualitative and quantitative deficiency of CD4+ T cells in HIV-1 infection. *Clin. Immunol. Immunopathol.* **80**, 290-297.
- Tateno H, Winter H C, Petryniak J and Goldstein I J, (2003) Purification, Characterization, Molecular Cloning, and Expression of Novel Members of Jacalin-related Lectins from Rhizomes of the True Fern *Phlebodium aureum* (L) J. Smith (Polypodiaceae). *J. Biol. Chem.* **278**, 10891–10899.
- Teixeira C R, Cavassani K A, Gomes R B, Teixeira M J. Roque- Barreira M C, Cavada B S, da Silva J S, Barral A and Barral-Netto M, (2006) Potential of KM+ lectin in immunization against *Leishmania amazonensis* infection. *Vaccine* **24,** 3001-3008.
- To W Y, Leung J C and Lai K N, (1995) Identification and characterization of human serum alpha2-HS glycoprotein as a jacalin-bound protein. *Biochim. Biophys. Acta*. **1249**, 58-64.
- Transue T R, Smith A K, Mo H, Goldstein I J and Saper M A, (1997) Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* lectin. *Nat. Struct. Biol.* **10**, 779-783.

- Turoverov K K, Haitlina S Yu and Pinaev G P, (1976) Ultra-Violet fluorescence of Actin. Determination of native actin content in actin preparations. *FEBS Lett.* **62**, 4-6.
- Urvashi S and Suresh C G, (2011) Purification, crystallization and X-ray characterization of a Kunitz type trypsin inhibitor protein from the seeds of chickpea (*Cicer arietinum*). accepted in *Acta cryst.* F (In press).
- Urvashi S, Gaikwad S M, Suresh C G, Dhuna V, Singh J and Kamboj S S, (2011) Conformational Transitions in *Ariesaema curvatum* Lectin: Characterization of an Acid Induced Active Molten Globule. *J. Fluoresc.* DOI 10.1007/s10895-010-0766-2.
- Valueva T A and Mosolov V V, (1999). Protein inhibitors of proteinases in seeds: 1. Classification, distribution, structure, and properties, 2. Physiological functions. *Russian J. Plant physiol.* **46**, 362-387.
- Valueva T A, Mosolov V V, (2004) Role of Inhibitors of Proteolytic Enzymes in Plant Defense against Phytopathogenic Microorganisms. *Biochemistry* **69**, 1305-1309.
- Van Damme E J M, Allen A K, and Peumans W J, (1987) Isolation and characterization of a lectin with exclusive specificity toward mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett.* **215**, 140–144.
- Van Damme E J M, Balzarini J, Smeets K, Van Leuven F and Peumans W J, (1994a) The monomeric and dimeric mannose-binding proteins from the Orchidaceae species *Listera ovata* and *Epipactis helleborine*: sequence homologies and differences in biological activities. *Glycoconj. J.* **11**, 321-332.
- Van Damme E J M, Brike F, Winter H C , Leuven F V, Goldstein I J and Peumans W J, (1996) Molecular cloning of two different mannose-binding lectins from tulip bulbs. *Eur. J. Biochem.* **236**, 419-427.
- Van Damme E J M, Goossens K, Smeets K, Van Leuven F, Erhaert P and Peumans W J, (1995) The major tuber storage protein of araceae species is a lectin. Characterization and molecular cloning of the lectin from *Arum maculatum*, L. *Plant Physiol.* **107**, 1147–1158.

- Van Damme E J M, Peumans W J, Barre A, Rougé P (1998) Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17**, 645–662.
- Van Damme E J M, Smeets K and Peumans W J, (1995) The mannose-binding monocot lectins and their genes. **In**: *Lectins: Biomedical Perspectives.* pp. 59-80. Pusztai A and Bardocz S, Eds, Taylor and Francis, London, UK.
- Van Damme E J M, Smeets K, Torrekens S, Van Leuven F and Peumans W J, (1994b) Characterization and molecular cloning of the mannose-binding lectins from three Orchidaceae species: *Listera ovata*, *Epipactis helleborine* and *Cymbidium* hybrid. *Eur. J. Biochem.* **221**, 769-777.
- van Eijsden R, Diaz C L, de Pater B S, and Kijne J W, (1995) Sugar-binding activity of pea (Pisum sativum) lectin is essential for heterologous infection of transgenic white clover hairy roots by *Rhizobium leguminosarum* biovar viciae. *Plant Mol. Biol*. **29**, 431-439.
- Van Parijs J, Broekaert W F, Goldstein I J and Peumans W J, (1991) Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. *Planta* **183**, 258-262.
- Vega N and Pérez G, (2006) Isolation and characterisation of a *Salvia bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry* **67**, 347-355.
- Vijayakumar T, Augustine J, Mathew L, Aleykutty M A, Nair M B, Remani P and Nair M K (1992) Tissue binding pattern of plant lectins in benign and malignant lesions of thyroid. *J. Exp. Pathol.* **6**, 11-23.
- Vijayan M and Chandra N, (1999) Lectins. *Curr. Opin. Struc. Biol.* **9**, 707-714.
- Vioque J, Clemente A, Sanchez-Vioque R, Pedroche J, Bautista J and Millan F, (1998) Comparative study of chickpea and pea pa2 albumins. *J. Agric. Food Chem.**46***, 3609–3613.
- Vioque J, Sa´nchez-Vioque R, Clemente A, Pedroche J, Bautista J and Milla´n F, (1999) Purification and Partial Characterization of Chickpea 2S Albumin. *J. Agric. Food Chem.* **47**, 1405-1409.
- Vodkin L O, Rhodes P R and Goldberg R B, (1983) A lectin gene insertion has the structural features of a transposable element. *Cell.* **34**, 1023–1031.

- Waljuno K, Scholma R A, Beintema J, Mariono A, Hahn A M, (1975) Amino acid sequence of hevein. *Proc. Int. Rubber. Conf, Kuala Lumpur* **2**, 518–531.
- Wang M-B, Boulter D and Gatehouse J A, (1994) Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of *Cucurbita pepo*. *Plant. Mol. Biol.* **24**, 159-170.
- Wang S, Rao P A, (2010) A leguminous trypsin-chymotrypsin inhibitor Limenin with antifungal activity from *Phaseolus limensis*. *Eur Food Res Technol.* **231**, 331–338.
- Weis W I and Drickamer K, (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65,** 441-473.
- Weiss M S and Hilgenfeld R, (1997) On the use of the merging R factor as a quality indicator for X-ray data. *J App. Cryst.* **30**, 203-205.
- Weiss M, (2001) Global indicators of X-ray data quality. *J App. Cryst.* **34**, 130-135.
- Wider G, (1998) Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. Progr. NMR Spectrosc. 32:193–275.
- Wider G, (2000) Structure Determination of Biological Macromolecules in Solution Using NMR spectroscopy. *BioTechniques* **29**, 1278–1294.
- Wider G, Macura S, Kumar A, Ernst R R and Wüthrich K, (1984) Homonuclear Two-Dimensional 1H NMR of Proteins. *Experimental Procedures. J. Magn. Reson.* **56**, 207–234.
- 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.
- Wimmerova M, Mitchell E, Sanchez J-F, Gautier C and Imberty A, (2003) Crystal Structure of Fungal Lectin. Six bladed β-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin. *J. biol. Chem*. **278**, 27059–27067.
- Wong J H, Wong C C T and Ng T B (2006) Purification and characterization of a galactose-specific lectin with mitogenic activity from pinto beans. *Biochim. Biophys. Acta.* **1760,** 808-813.
- Wright C S and Hester G, (1996) The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentose: evidence for two unique binding modes. *Structure* **4**, 1339-1352.

- Wright C S and Jaeger J, (1993) Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J. Mol. Biol*. **232**, 620–638.
- Wright C S, (1977) The crystal structure of wheat germ agglutinin at 2.2 Å resolution. *J. Mol. Biol.* **111**, 439–457.
- Wright C S, (1987) Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 Å resolution. *J. Mol. Biol.* **194**, 501–529.
- 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. *Current Opinion in Structural Biology* **7**, 631-636.
- Wright C S, Kaku H, Goldstein I J, (1990) Crystallization and preliminary X-ray diffraction results of snowdrop (*Galanthus nivalis*) lectin. *J. Biol. Chem*. **265**, 1676-1677.
- Wright L M, Reynolds C D, Rizkallah P J, Allen A K, Van Damme E J M, Donovan M J and Peumans W J (2000) Structural characterisation of the native fetuin-binding protein *Scilla campanulata* agglutinin: a novel two-domain lectin. *FEBS Lett*. **468**, 19-22.
- Wright L M, Van Damme E J M, Barre A, Allen A K, Van Leuven F, Reynolds C D, Rouge P, Peumans W J, (1999) Isolation, characterization, molecular cloning and molecular modelling of two lectins of different specificities from bluebell (*Scilla campanulata*) bulbs. *Biochem. J*. **340**, 299-308.
- Wright L M, Wood S D, Reynolds C D, Rizkallah P J, Peumans W J, Van Damme E J M, (1996) Purification, crystallization and preliminary X-ray analysis of a mannose-binding lectin from bluebell (*Scilla campanulata*) bulbs. *Acta Crystallogr* D**52**, 1021–1023.
- Wu J A, Herp A, Chow L P and Lin J Y, (2001) Carbohydrate specificty of a toxic lectin, abrinA, from the seeds of *Abrus precatorius* (jequirity bean). Life Sci. **69**, 2027-2038.
- Wüthrich K, (1986) NMR of Proteins and Nucleic Acids. Wiley: New York.
- Xie D, Bhakuni V, and Freire E, (1991) Calorimetric Determination of the Energetics of the Molten Globule Intermediate in Protein Folding: Apo-α-lactalbumin. *Biochemistry*. **30**, 10673-10678.

- Yang H and Czapla T H, (1993) Isolation and Characterization of cDNA Clones Encoding Jacalin Isolectin. *J. Biol. Chem.* **268**, 5905-5910.
- Ye X Y and Ng T B, (2002) Isolation of a new cyclophilin-like protein from chickpeas with mitogenic, antifungal and anti-HIV-1 reverse transcriptase activities *Life Sciences.* **70**, 1129–1138.
- Yi S M, Harson R E, Zabner J and Welsh M J, (2001) Lectin binding and endocytosis at the apical surface of human airway epithelia. *Gene Ther*. **8**, 1826-1832.
- Zabel P L, Noujaim A A, Shysh A and Bray J, (1983) Radioiodinated Peanut lectin: a potential radiopharmaceutical for immunodetection of carcinoma expressing T-antigen. *Eur. J. Nucl. Med.* **8**, 250-254.
- Zebda N, Bailley M, Brown S, Dore J F and Berthier-Vergnes O, (1994) Expression of PNA binding sites on specific glycoproteins by human melanoma cells is associated with a high metastatic potential. *J. Cell. Biochem.* **54**, 161-173.
- Zhao J K, Zhao Y C, Li S H, Wang H X and Ng T B, (2011) Isolation and characterization of a novel thermostable lectin from the wild edible mushroom. *Agaricus arvensis. Journal of Basic Microbiology* **51,** 1-8.