

**MOLECULAR STUDIES ON A LECTIN**  
**FROM *CICER ARIETINUM* L.**

Thesis submitted to University of Pune

For the degree of

**DOCTOR OF PHILOSOPHY**

IN

**CHEMISTRY (BIOCHEMISTRY)**

By

**MADHURIMA S. WAKANKAR**

Under the guidance of

**DR. SUSHAMA M. GAIKWAD**

DIVISION OF BIOCHEMICAL SCIENCES

NATIONAL CHEMICAL LABORATORY

PUNE - 411 008 (INDIA)

**APRIL 2013**

*for*

*Aai, Baba, Tejaswini*

*Mummy, Papa, Suprit*

# CONTENTS

	<b>Page No.</b>
<b>ACKNOWLEDGEMENTS</b>	<b>i</b>
<b>CERTIFICATE</b>	<b>v</b>
<b>DECLARATION BY THE CANDIDATE</b>	<b>vi</b>
<b>ABBREVIATIONS</b>	<b>vii</b>
<b>ABSTRACT</b>	<b>ix</b>
<b>LIST OF PUBLICATIONS</b>	<b>xii</b>
<b>Chapter 1: Introduction</b>	
<b>A. Lectins</b>	<b>1-20</b>
1.1. History	1
1.2. Definition	3
1.3. Detection	4
1.4. Occurrence and Biological role of lectins	5
1.5. Classification of lectins	8
1.6. Classification of Plant Lectins	11
1.7. Lectin-sugar interactions	16
1.8. Most recently reported crystal structures of few plant lectins	17
1.9. Applications of lectins	18
<b>B. Protein folding</b>	<b>20-31</b>
1.10. The fundamental mechanism of protein folding	20
1.11. The protein folding problem	20
1.12. Models to explain protein folding	22
1.13. Free energy landscapes of Proteins	23
1.14. Biophysical techniques to study protein folding	24
1.15. Protein misfolding and aggregation	26
1.16. Molecular chaperones	28
1.17 Applications of protein folding in industry	29
1.18. Unsolved problems of protein physical	30

	science	
	<b>C. Bioinformatics</b>	<b>31</b>
	<b>D. Present investigation</b>	<b>32</b>
	References	33
<b>Chapter 2:</b>	<b>Cloning, expression and purification of a lectin from <i>Cicer arietinum</i> L.</b>	
	Summary	44
	2.1. Introduction	44
	2.2. Materials	48
	2.3. Methods	50
	2.4. Results and Discussion	57
	References	65
<b>Chapter 3:</b>	<b>Ligand binding and Solute quenching studies of the recombinant lectin from <i>Cicer arietinum</i> L</b>	
	Summary	68
	3.1. Introduction	68
	3.2. Materials	69
	3.3. Methods	70
	3.4. Results and Discussion	74
	References	85
<b>Chapter 4:</b>	<b>Conformational transitions of the recombinant lectin</b>	
	Summary	89
	4.1. Introduction	89
	4.2. Materials	96
	4.3. Methods	96
	4.4. Results and Discussion	98
	References	108
<b>Chapter 5:</b>	<b>Computational studies on the recombinant lectin</b>	
	Summary	113
	5.1. Introduction	113
	5.2. Methods	116
	5.3. Results and Discussion	118
	References	141
<b>Chapter 6:</b>	<b>Discussion</b>	
	Discussion	146
	Conclusions	151
	References	153

**“All have their worth and each contributes to the worth of the others.”**

**J. R.R. Tolkien**

## **Acknowledgements**

Interdependence is more important than independence. It becomes easier to achieve a goal, if there are people to help and support you. My stay at the National Chemical Laboratory for my doctoral study has been a great learning experience. Many people have directly and indirectly helped me in my research work and I will always remain grateful to them.

First and foremost, I am greatly indebted to my guide **Dr. Sushama M. Gaikwad** for her invaluable guidance and keen interest during the course of this investigation. I am very thankful to her for her unending support, encouragement, and valuable direction throughout my research work. She has given me the freedom to think, and work; and I shall cherish my learning experience with her. Without her dedication and persistence, it would have been impossible to complete my work successfully.

I am extremely grateful to **Late Dr. M. Islam Khan** for providing me an opportunity to pursue my doctoral study with him during the first two years. My initial work was conceived under his guidance. His scientific temperament, innovative approach, dedication towards research, and modest yet straightforward nature has inspired me the most. I will cherish invaluable memories of learning from him. The love and care shown by Mrs. Kausar Khan is fondly remembered.

I am also grateful to my senior Late Mr. Rohtas Rangera for providing initial starting material for my research work. Without his apt record keeping and accurate experimental hand, it would have been difficult to achieve the initial objectives of my research.

I am particularly thankful to Dr. Musti V. Krishnasastry from the National Centre for Cell Science, Pune, for providing me the

freedom to use the facilities of molecular biology in his laboratory. The entire cloning and expression exercise was designed and carried out under his guidance. He has been extremely helpful during the entire course of my investigation by giving useful suggestions from time to time. I will cherish the learning experience gained from him.

I am thankful to Dr. C.G. Suresh for his keen interest and valuable suggestions during my entire work, and also for being a member of my PhD committee.

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

I would also like to extend my sincere thanks to Dr. Moneesha Fernandez, Organic Chemistry Division, NCL, and Dr. H. N. Gopi, IISER, Pune for the permission to use the CD facility.

I would like to thank Dr. Alok Sen, Organic Chemistry Division, NCL, for collaborating on our work on insecticidal activity of the lectin.

I am extremely thankful to Tulika for her invaluable and timely help in putting two of my chapters in proper shape. She has taught me several new techniques and has become a good friend.

I am also grateful to Krunal for carrying out the docking simulations of my lectin. It would have been very difficult to complete this task without his help. He has also been very helpful in putting my manuscripts in order and teaching me new techniques.

Special thanks are due to my dear colleagues Dr. Shashidhar, Dr. Ansary, Dr. Avinash, Sonali, Sayli, Ekta and Priya for their continuous support and maintaining a pleasant atmosphere for working in the laboratory. Ansary, Avinash, Dr. Geetha, Poonam, and Dr. Varsha were my pillars of strength. I have wonderful memories of the time spent with Varsha, Sonali and Ganesh. Ameya, Ekta, Manas, Sayli and Tulika have been extremely helpful during some difficult times and I am grateful to them for their love and concern.

I express my deep gratitude for my seniors Dr. Nagraj, Dr. Atul, Dr. Sreekanth, Dr. Shashidhar, Dr. Sajid, Dr. Shabab, and Dr. Urvashi for their relentless help and support during my stay at NCL.

I am thankful to my seniors from NCCS, Pune: Dr. Aejaaz, Dr. Amita, Dr. Saumya, and Dr. Neesar for teaching me the basic techniques of molecular biology. Dr. Aejaaz has been instrumental in helping me achieve the successful expression of the lectin gene. The technical help provided by Mr. Anil Lotke and my juniors Ekansh, Santosh and Shipra is greatly appreciated.

I am also thankful to my friends in the division: Dr. Manasi, Manisha, Neha, Nishant, Ravi, Pallavi, Parth, Prashant, Priyabrata, Ranu, Ruby, Sana, Sarika, Dr. Shadab, and Vithhal who have helped me in many ways.

My old friends from Baroda: Dr. Dhvani, Dr. Falguni, Inshiya, Jasmi, Dr. Meera, Dr. Prakash, Dr. Ruchi and Dr. Saurabh have always been a great source of inspiration.

My deepest sense of gratitude goes to **Aai** and **Baba** for their love, patience, kindness and blessings. This thesis is because of their immense belief in me. I am highly indebted to them and my younger sister **Tejaswini**, who has been very supportive and encouraging.

I am equally thankful to **Mummy** and **Papa** for the love and care showered on me. They have been very supportive of my research work and have encouraged me to work harder. I am also thankful to my brothers-in-law, **Dr. Sukrit** and **Sumit** for their support and love.

No words can express my gratitude towards my best friend, my husband, **Dr. Suprit**. He has been a constant source of motivation and encouragement. The task of completion of the thesis seemed unsurmountable with each passing day, but his crucial support and help allowed me to wade through.

I thank the Director, National Chemical Laboratory, and the Head, Division of Biochemical Sciences, for permitting me to carry out

my PhD work in the Division of Biochemical Sciences, NCL and to submit this work in the form of the thesis. I am also thankful to the library staff, administrative and technical staff for their help during the course of my study. The financial assistance from the Council of Scientific and Industrial Research, India is also duly acknowledged.

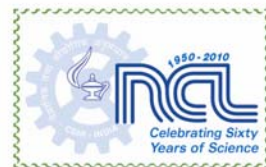
Finally, I thank God for providing me the strength and energy to pursue and complete my doctoral study.

**Madhurima S. Wakankar**





राष्ट्रीय रासायनिक प्रयोगशाला  
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)  
डॉ. होमी भाभा रोड, पुणे - 411 008. भारत  
**NATIONAL CHEMICAL LABORATORY**  
(Council of Scientific & Industrial Research)  
Dr. Homi Bhabha Road, Pune - 411008. India



## CERTIFICATE

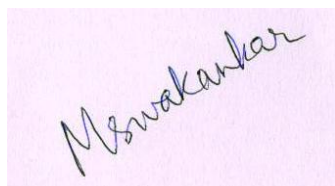
Certified that the work incorporated in the thesis entitled “**Molecular studies on a lectin from *Cicer arietinum* L.**” submitted by Ms. Madhurima S. Wakankar was carried out under my supervision. Such materials has been obtained from other sources has been duly acknowledged in the thesis.

Date: 23<sup>rd</sup> April 2013

Dr. Sushama M. Gaikwad  
**Research Guide**

## DECLARATION OF THE CANDIDATE

I declare that the thesis entitled “**Molecular studies on a lectin from *Cicer arietinum* L.**” submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me during the period from **26<sup>th</sup> July, 2007 to 8<sup>th</sup> October, 2012** under the guidance of **Dr. Sushama M. Gaikwad** and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or other institute of Higher learning. I further declare that the material obtained from other sources has been duly acknowledged in the thesis.



Signature of the Candidate

Date: 23<sup>rd</sup> April, 2013

Madhurima S. Wakankar

## LIST OF ABBREVIATIONS USED

ABA	<i>Agaricus bisporus</i> agglutinin
AFM	Atomic force microscopy
ANS	1-anilino-8-naphthalenesulfonate
Arg	arginine
Asn	asparagine
Asp	aspartic acid
BLAST	Basic Local Alignment Sequence Tool
°C	degree celsius
CD	circular dichroism
Con A	Concanavalin A
CTAB	cetyl trimethyl ammonium bromide
Da	dalton
DEAE	di ethyl amino ethyl
DEPC	diethylpyrocarbonate
dNTP	deoxy nucleotide triphosphate
DSL	<i>Datura stramonium</i> lectin
DNA, cDNA	deoxyribonucleic acid, complementary DNA
EcorL	<i>Erythrina corallodendron</i> lectin
EDTA	ethylene diamine tetra acetic acid
ERGIC-3	Endoplasmic reticulum-Golgi intermediate compartment protein 3
ESI-MS	electrospray ionization mass spectrometry
FPLC	fast performance liquid chromatography
FRET	Fluorescence Resonance Energy Transfer
FTIR	Fourier transform infrared spectroscopy
Gal	galactose
GalNAc	N-acetyl galactosamine
GdnHCl	guanidine hydrochloride
GlcNAc	N-acetyl glucosamine
Glu	glutamic acid
GNA	<i>Galanthus nivalis</i> agglutinin
His	histidine
HFIP	1,1,1,3,3,3-hexafluoroisopropanol
HPLC	high pressure liquid chromatography
HPSEC	high performance size exclusion chromatography
Ile	isoleucine
IPTG	isopropyl- $\beta$ -D-thiogalactoside
kDa	kilo dalton
LB	Luria Bertani Broth
Leu	leucine
Lys	lysine
MAA	<i>Maackia amurensis</i> agglutinin
MG	molten globule
MALDI	Matrix Assisted Laser Desorption Ionization
Man	mannose
MRE	Mean Residue Ellipticity
mRNA	messenger ribonucleic acid
NRMSD	normalized root mean square deviation

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
Pfu polymerase	<i>Pyrococcus furiosus</i> polymerase
PNA	peanut agglutinin
PSI-BLAST	position specific iterative BLAST
PVPP	poly vinyl poly pyrrolidone
rpm	revolutions per minute
RCA	<i>Ricinus communis</i> agglutinin
RIP	Ribosome-inactivating protein
SBA	soybean agglutinin
SDS	sodium dodecyl sulphate
Ser	serine
<i>Taq</i> polymerase	<i>Thermus aquaticus</i> polymerase
TFE	2, 2, 2-trifluoroethanol
Trp	tryptophan
UV	ultra violet

## ABSTRACT

Lectins and hemagglutinins are proteins, which have at least one non-catalytic domain that exhibits reversible binding to specific monosaccharides or oligosaccharides. They bind to the carbohydrate moieties on the surface of erythrocytes and agglutinate them, without altering the properties of the carbohydrates. Lectins with specific carbohydrate specificity have been purified from various plant tissues and other organisms. Plant lectins attract much attention because of their ease of isolation and their utility as reagents for detecting glycoconjugates in solution and on cell surfaces. The recognition of carbohydrate moieties by lectins has important implications in a number of biological processes such as in cell-cell interactions, signal transduction and in cell growth and differentiation. Among the lectins, legume lectins are well studied and a great deal of information including primary structure and crystal structure of many of them is available. Additionally, thermal and chemical unfolding processes of a number of legume lectins have been characterized. Computational studies have also been carried out on several lectins.

### **Chapter 1: Introduction**

This part comprises a literature survey of lectins with reference to their detection, occurrence, classification, isolation, and applications. It also gives an account of protein folding and bioinformatics.

### **Chapter 2: Cloning, expression and purification of a lectin from *Cicer arietinum* L. seeds**

The chick pea lectin gene comprising of 693 bp was isolated from total RNA using primers designed from the nucleotide sequence of the major seed albumin PA2 of *Pisum sativum*. The gene was cloned into the pGEM-T Easy vector by A-tailing. For expression, the gene was sub-cloned in the expression vector pET-28a(+) in between the NcoI and EcoRI sites. *E. coli* strain BL21-CodonPlus (DE3)-RIL was used for expression; induction was carried out with IPTG. The crude cell lysates from the induced cultures were used for the purification of the lectin. The lectin was purified to

homogeneity using a DEAE-cellulose and Superose 12 prep grade FPLC gel filtration column. The lectin, consisting of 462 amino acids, was found to be a homodimer of approximate molecular mass of 54,600 Da, and a subunit molecular mass of 27,000 Da. Around 5 to 8 mg of lectin could be purified per litre of *E. coli* broth with a specific activity of  $5 \times 10^3$  hemagglutinating units/mg.

### **Chapter 3: Ligand binding and Solute quenching studies of the recombinant lectin from *Cicer arietinum* L.**

The recombinant chick pea lectin did not recognize any simple mono- or oligosaccharides, but could bind only fetuin and its asialo-triantennary glycan. Thermodynamics of binding of rCAL with the asialo fetuin glycan indicated the process to be spontaneous and enthalpically driven. Presence of the hemopexin domain in the nucleotide sequence of the lectin led us to investigate the hemin binding property of rCAL. Homologous albumins from plant seeds have been shown to bind hemin, spermine and thiamine. Fluorescence spectroscopy studies indicated that the lectin showed high affinity towards hemin, and a slightly lower affinity for spermine and thiamine. Solute quenching studies revealed the microenvironment of the single tryptophan to be negatively charged. Denaturation of the lectin resulted in redistribution of charge density around the tryptophan even at lower concentrations of GdnHCl. During acrylamide quenching of rCAL fluorescence in the denatured state, the static component was more prominent than the collisional one

### **Chapter 4: Conformational transitions of the recombinant lectin**

Conformational characterization of the recombinant lectin was carried out using fluorescence spectroscopy and circular dichroism. Thermal unfolding of the lectin resulted in rapid structural rearrangements above 55 °C. Transient exposure of hydrophobic residues was observed that resulted in aggregation. The denaturation was irreversible. GdnHCl-mediated unfolding was multi-step and irreversible. Denaturation resulted in unfolding followed by dissociation of the dimer. The structure of the protein was drastically affected within one hour in acidic and alkaline buffers.

## **Chapter 5: Computational studies on the recombinant lectin**

Homology model of rCAL was built using the crystal structure of LS24 albumin from *Lathyrus sativus* (PDB ID:3LP9) as the template. Validation of the model with standard parameters indicated the reliability of the model. This model was used for the docking of hemin, spermine and thiamine using Auto Dock Vina 1.1.2. Binding energies for each ligand were calculated and the amino acid residues involved in the interaction were identified. The microenvironment of the tryptophan in the model was visualized in PyMOL. It showed that the tryptophan was exposed to the solvent and was surrounded by hydrophobic residues and a weak positive charge. This confirmed the results obtained from solute quenching studies. Aggregation-prone residues of the lectin were predicted using the BioLuminate1.0 tool of the Schrodinger suite, as well as web-based servers like the Aggrescan, AmylPred and the FoldAmyloid. Common amino acid residues were identified by all these softwares as being “hot-spots” of aggregation.

## **Chapter 6: Discussion**

In this chapter the properties of the lectin from *Cicer arietinum* are discussed.

## LIST OF PUBLICATIONS:

**1. Madhurima S. Wakankar**, Musti V. Krishnasastry, Tulika M. Jaokar, Krunal A. Patel, Sushama M. Gaikwad. **Solution and *in silico* studies on the recombinant lectin from *Cicer arietinum* seeds.** *International Journal of Biological Macromolecules*, **56** (2013) 149– 155.

**2. Madhurima S. Wakankar**, Krunal A. Patel, Musti V. Krishnasastry, Sushama M. Gaikwad. **Solution and *in silico* ligand binding studies of *Cicer arietinum* lectin.** Special issue on “Structure and Function of Proteins” of *Biochemistry and Physiology - Open Access* (Accepted for publication).



## **CHAPTER: 1**

---

### **INTRODUCTION**

---

## A. LECTINS

Lectins comprise the major class of carbohydrate-specific proteins, which bind mono- and oligosaccharides reversibly and with high specificity. They are devoid of catalytic activity (like the enzymes), and are not the products of an immune response (like antibodies). Lectins exist in most organisms, ranging from viruses and bacteria to plants and animals.

### 1.1. HISTORY

The lectins have the longest scientific history among all plant proteins. The commencement of the scientific discipline called “lectinology” dates back to 1888 when Stillmark observed that the toxicity of castor bean extracts was linked to the presence of a proteinaceous hemagglutinating factor called “ricin” [1]. His discovery was followed by similar toxins in other plants also.

A wealth of literature is available on lectins, reflecting the tremendous research being carried out in this field. More than 40 reviews are available on various aspects of lectins, some of the most important ones are mentioned here [2-15]. Currently, crystal structures of around 110 lectins have been reported in the PDB database (as accessed in March, 2013). Around 30 lectins are commercially available (in Sigma-Aldrich) for use in analytical and clinical industry; additionally twenty fluorescent lectin conjugates are also available. We have tried to compile the vast literature on lectins and give a brief account on most of the aspects of lectins in this chapter. Table 1.1 depicts the milestones of the long history of lectinology.

**Table 1.1 History of study of lectins (adapted from Singh and Parthasarathi, 2012 [15]).**

<b>Year</b>	<b>Milestone</b>
<b>1884</b>	Toxicity in <i>Abrus precatorius</i> seed extracts
<b>1886</b>	Toxicity in <i>Ricinus communis</i> seed extracts
<b>1888</b>	Hemagglutinating activity in <i>Ricinus communis</i> seed extracts Toxicity in <i>Croton tiglium</i> seed extracts
<b>1890</b>	Use of abrin and ricin in immunological research
<b>1891</b>	Hemagglutinating activity in <i>Abrus precatorius</i> seed extracts
<b>1902</b>	Reversibility of the hemagglutination by heat

<b>Year</b>	<b>Milestone</b>
<b>1902</b>	Inhibition of the hemagglutinating activity by non-immune serum
<b>1903</b>	Hemagglutinating activity in horseshoe crab
<b>1907</b>	Hemagglutinating activity in non-toxic plants
<b>1908</b>	Agglutination of leucocytes and kidney and liver cells by <i>Phaseolus vulgaris</i>
<b>1908</b>	Species specificity of plant hemagglutinins
<b>1909</b>	Inhibition of hemagglutinating activity by heat treated serum
<b>1909</b>	Inhibition of the hemagglutinating activity by mucin
<b>1912</b>	Hemagglutinins and germination
<b>1917</b>	Isolation and crystallization of Concanavalin A
<b>1926-27</b>	Applicability of lectins for blood typing
<b>1935</b>	Specificity of eel serum agglutinins
<b>1936</b>	Sugar specificity of Concanavalin A
<b>1947-49</b>	Blood group specificity of plant hemagglutinins
<b>1949</b>	Thermo-inactivation of <i>Phaseolus vulgaris</i> lectins
<b>1952</b>	Inhibition of lectins by simple sugars
<b>1954</b>	Introduction of the term 'lectin'
<b>1960</b>	Mitogenic stimulation of lymphocytes by <i>Phaseolus vulgaris</i> lectin
<b>1963</b>	Agglutination of malignant cells by lectins
<b>1964</b>	Parallel inactivation of hemagglutinating and antinutritional activity by heat
<b>1965</b>	Affinity chromatography for lectin purification
<b>1966</b>	Lectins from algae
<b>1970</b>	Use of Con A for affinity purification of glycoproteins
<b>1974</b>	Role of animal lectins in endocytosis of glycoproteins
<b>1977</b>	Role of bacterial lectins in infection

Year	Milestone
1981	Use of lectins in bone marrow transplantation
1984	Combined use of lectin and enzyme in clinical identification of micro organisms
1987	Control of root-knot nematodes by lectins
1989	Root lectin as a specificity determinant in the <i>Rhizobium</i> -legume symbiosis
1990	Con A expression in <i>Escherichia coli</i> cells
1992 to 1993	Identification of impaired synthesis of lectin (selectin) ligands by defective fucosylation as the cause for leukocyte adhesion deficiency type II
1995	Structural analysis of a lectin-ligand complex in solution by NMR Spectroscopy
1996 to 1998	Detection of differential conformer selection by plant and animal lectins
2000 to 2005	Lectins as tools to understand glycomics
2005 to 2010	Use of glycomics and glycan arrays Use of lectins as vaccines for cancer and HIV

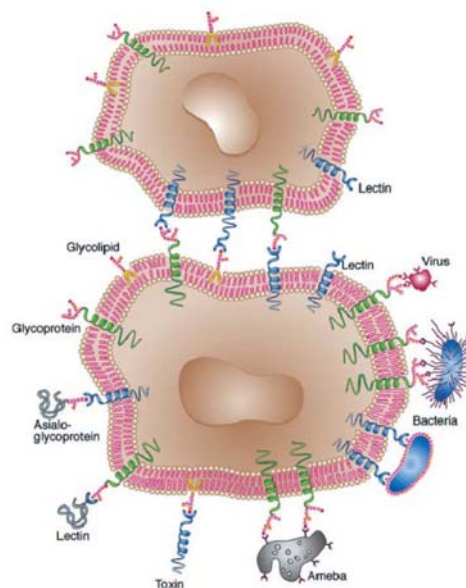
## 1.2. DEFINITION

The first definition of lectins was based primarily on the sugar specificity and inhibition of the agglutination reaction. According to this definition, lectins are carbohydrate binding proteins of non-immune origin which agglutinate cells and/or precipitate glycoconjugates [16]. The landmark discovery of selective agglutination of erythrocytes of a particular human blood group within the ABO system by some hemagglutinins led to the introduction of the term “lectin” which is derived from the Latin word “*lectus*” (the past tense of *legere*) which means “to select” or “to choose” [17]. This definition was restricted to multivalent carbohydrate binding proteins, and did not include those lectins that showed poor agglutination activity or the chimeric lectins that possessed (in addition to their sugar binding domain) a second domain with catalytic activity. To overcome this problem, a definition was introduced that was based on a single and simple criterion, namely the presence of at least one non-catalytic domain that binds reversibly to a specific carbohydrate. Accordingly, plant lectins were

defined as “all plant proteins that possess at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide” [6]. Currently, with advances in lectin research, they can be aptly defined as: “*Lectins represent a heterogeneous group of proteins ubiquitously present; having multiple subunits; possessing specific sugar specificities, and thus possessing carbohydrate binding sites but widely varying in their size, subunit structure, molecular organization, and in their sugar combining sites*”.

### 1.3. 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, glycosyl transferases, antibodies etc. Hemagglutination is routinely used to detect the presence of lectins in a biological source [18]. The erythrocytes used (human or animal) may be plain or treated with papain, trypsin, or neuraminidase [19]. Other types of materials like polysaccharides [20] have also been used for the detection of lectins. Techniques like affinity electrophoresis [18, 21] and enzyme multiplied assay [22] too have been employed.



**Figure 1.1: The interactions between lectin and carbohydrates on the cell surface.** Lectins serve as means of attachment of different kinds of cells as well as viruses to other cells via the surface carbohydrates of the latter. In some cases, cell surface lectins bind particular glycoproteins (e.g., asialoglycoproteins), whereas in other cases the carbohydrates of cell surface glycoproteins or glycolipids serve as sites of attachment for biologically active molecules that themselves are lectins (e.g. carbohydrate-specific bacterial and plant toxins, or

---

galectins; shown in green for virus, bacteria, amoeba, and in blue for toxin). (Based on a diagram from BioCarbAB Lund, Sweden, Sharon and Lis, 2004 [23]).

## 1.4. OCCURRENCE AND BIOLOGICAL ROLE OF LECTINS

### 1.4.1. Viral Lectins

One of the most thoroughly studied lectins, the Influenza virus hemagglutinin (type I membrane protein) [24] is a viral lectin. Influenza virus agglutinin is *N*-acetylneuraminic acid (sialic acid) specific, and hence binds to sialic acid containing carbohydrates on the surface of host cell and thus facilitates the entry of viral genome inside the host and causes its subsequent replication. The hemagglutinin has been crystallized and its structure has been determined (PDB ID: 1HTM) [25]. Crystal structures of several viral lectins are known till date [Swine H1N2 influenza virus hemagglutinin – PDB ID: 4F3Z; hemagglutinin from H7N7 influenza virus – PDB ID: 4DJ6; hemagglutinin from avian H1N1 influenza A virus – PDB ID: 3HTO; H5N1 influenza virus hemagglutinin – PDB ID: 2FKO etc.].

### 1.4.2. Bacterial Lectins

Many bacterial lectins have been isolated, characterized and described in detail. They can be divided into two classes: (1) Adhesins: Lectins are localized on the surface of the bacteria and aid in their adhesion to the surface leading to colonization, and (2) Bacterial toxins secreted in the external environment. The bacterial lectins generally prefer binding to glycolipids rather than glycoproteins and their binding specificity to a particular structure can be correlated with the particular tissue it colonizes in the host.

### 1.4.3. Protozoan lectins

Among the vast number of protozoans which infect humans as well as animals and cause infection, the well studied lectins are from *Entamoeba histolytica* [26], the causative agent of amoebic dysentery.

### 1.4.4. Fungal Lectins

The occurrence of lectins in fungi is wider than in higher plants and they have been isolated from conidia, mycelium, basidiomes, and fruiting bodies. *Aleuria aurantia* was the first fungal lectin, for which the crystal structure was solved by Wimmerova *et al.* in 2003 [27]. Other fungal lectins characterized include Phytohemagglutinin A and Phytohemagglutinin B (PHA and PHB) from *Agaricus bisporus*, *Ganoderma* lectin (*Ganoderma capense*); *Grifola* lectin (*Grifola frondosa*) [28] and many more have been reported. Mushroom lectins have been studied from

edible mushrooms as *Amanita phalloides*, *Lactarius deliciosus*, *Boletus edulis*, *Laetiporus sulphureus* [29] etc.

#### 1.4.5. Animal Lectins

In animals, lectins are secreted in extra cellular form as well as inside the cell membrane. Lectins have been implicated in mediating a variety of binding and recognition events in animal cells.

#### 1.4.6. Plant Lectins

Lectins in plants have been studied since the advent of lectinology. They comprise the largest group of lectins that have been characterized with respect to structural and functional aspects. They are mainly localized in the vegetative and storage tissue such as the seeds, tubers, rhizomes, leaves, stems, bark, fruits, bulbs and constitute the main protein component of that tissue.

The functions of lectins in microbes and animals are presented below in Table 1.2, while their functions in plants have been enumerated in Table 1.3.

**Table 1.2 Functions of lectins in microorganisms and animals**

(Adapted from Singh and Parthasarathi, 2012 [15]).

Lectin/ Source	Role
<b>Microorganisms</b>	
Amoeba	Infection
Bacteria	Infection
Influenza virus	Infection
<b>Animals</b>	
Calnexin, calreticulin, ERGIC-53 (endoplasmic reticulum)	Control of glycoprotein biosynthesis
Collectins	Innate immunity
Dectin-1	Innate immunity
Galectins	Regulation of cell growth and apoptosis Regulation of the cell cycle Modulation of cell-cell and cell-substratum interactions
Macrophage mannose	Innate immunity

Receptor	Clearance of sulphated glycoprotein hormones
Man-6-P receptors	Targeting of lysosomal enzymes
L-selectin	Lymphocyte homing
E- and P-selectins	Leukocyte trafficking to sites of inflammation
Siglecs (human basophils, eosinophils, mast cells)	Cell-cell interactions in the immune and neural system
Spermadhesin	Sperm-egg interaction

**Table 1.3 Functions of lectin activity in plants (Adapted from Gabius *et al.*, 2004 [8])**

Activity	Example of lectin
<b>External activities</b>	
Protection from fungal attack	<i>Hevea brasiliensis</i> (rubber tree), <i>Urtica dioica</i> (stinging nettle), <i>Solanum tuberosum</i> (potato)
Protection from herbivorous animals	<i>Phaseolus vulgaris</i> (French bean), <i>Ricinus communis</i> (castor bean), <i>Galanthus nivalis</i> (snowdrop), <i>Triticum vulgare</i> (wheat)
Involvement in establishing symbiosis between plants and bacteria	<i>Pisum sativum</i> (common pea), <i>Lotononis bainesii</i> (miles lotononis), <i>Arachis hypogaea</i> (peanut), <i>Triticum vulgare</i> (wheat), <i>Oryza sativa</i> (rice)
<b>Internal activities</b>	
Storage proteins	All lectins belong to this category
Ordered deposition of storage proteins and enzymes in protein bodies and mediation of Contact between storage proteins and	<i>Pisum sativum</i> (common pea), <i>Lens culinaris</i> (lentil), <i>Glycine max</i> (soybean), <i>Oryza sativa</i> (rice)



protein body membranes	
Modulation of enzymatic activities such as phosphatase activity	<i>Secale cereale</i> (rye), <i>Solanum tuberosum</i> (potato), <i>Pleurotus ostreatus</i> (oyster mushroom), <i>Glycine max</i> (soybean), <i>Dolichos biflorus</i> (horse gram)
Participation in growth regulation	<i>Medicago sativa</i> (alfalfa), <i>Cicer arietinum</i> (chick pea)
Adjustment to altered environmental Conditions	<i>Triticum aestivum</i> (winter wheat)

### 1.5. CLASSIFICATION OF LECTINS

Lectins belonging to the larger class of carbohydrate binding proteins are known to possess some common structural features and binding motifs. Hence, a better way of classifying lectins would be on the basis of homologies in their primary sequences and three-dimensional structures. Classification of lectins has been based on their molecular and overall structure.

#### 1.5.1. Based on molecular structure

On the basis of their molecular structure they can be classified as (a) Simple Lectins, (b) Mosaic (or multi domain), and (c) Macromolecular assemblies [30]. In each of this class of lectins, they can be further grouped into distinct families on the basis of similar sequence and resemblance of their structural characteristics.

**Table 1.4 Classification of lectins on the basis of their molecular structure**

Simple	Mosaic (multidomain)	Macromolecular assemblies
<ul style="list-style-type: none"> <li>• Comprises all known plant lectins and the galectins.</li> <li>• Contain lesser number of subunits, which may not be identical.</li> </ul>	<ul style="list-style-type: none"> <li>• Consists of lectins from diverse sources such as viral lectins, animal lectins of the C, P and I- types.</li> </ul>	<ul style="list-style-type: none"> <li>• Common in bacteria, and play a role in structural support for mobility of the organism.</li> </ul>

<ul style="list-style-type: none"> <li>• Subunits of MW &lt; 40 kDa.</li> <li>• May contain an additional domain besides the carbohydrate binding domain.</li> </ul>	<ul style="list-style-type: none"> <li>• Structurally composite molecules with several kinds of protein domains, one of which exhibits carbohydrate binding property.</li> <li>• Generally monovalent, but may behave as multivalent proteins.</li> </ul>	<ul style="list-style-type: none"> <li>• Filamentous organelles with helically arranged subunits (pilins) in a well-defined order.</li> <li>• Some contain hundreds of subunits, some of which bind carbohydrates.</li> <li>• Three-dimensional structures not solved yet.</li> </ul>
--	---	---

### Examples:

#### 1. Simple lectins

- Legume lectins:** ConA, SBA, PHA, EcorL
- Cereal lectins:** WGA, Rice lectins, Barley lectin
- Amaryllidaceae & related families:** Snowdrop (*Galanthus nivalis*) lectin (GNA)
- Moraceae:** Jacalin from *Artocarpus integrifolia*; Lectin from *Maclura pomifera*
- Euphorbiaceae:** Ricin (*Ricinus communis* agglutinin) (RCA)
- Animal lectins: (Galectins)** Galectin from electric organ of eel
- Animal lectins (Pentraxins) :** Serum Amyloid P component (SAP).

#### 2. Mosaic lectins

- Viral Hemagglutinins:** Influenza Virus Hemagglutinin; lectin from *Murine Polyoma Virus*.
- C-Type Lectins**
  - Endocytic Lectins : galactose /N-acetyl galactosamine specific lectin from rabbit hepatocytes (RHL); Avian hepatic lectin specific for N-acetylglucosamine; Fucose-specific receptor (lectin) found on the Kupffer cells of the liver.

ii) Collectins: MBP A (serum type MBP) and C (liver type MBP); Pulmonary surfactant apoproteins A and D; Collectin CL-43 from bovine serum and bovine conglutinin.

c) **Selectins**: consist of three groups - E-lectins, P-lectins and L-lectins.

d) **P-type lectins**: Two mannose 6-phosphate (Man-6-P) receptors belong to this family.

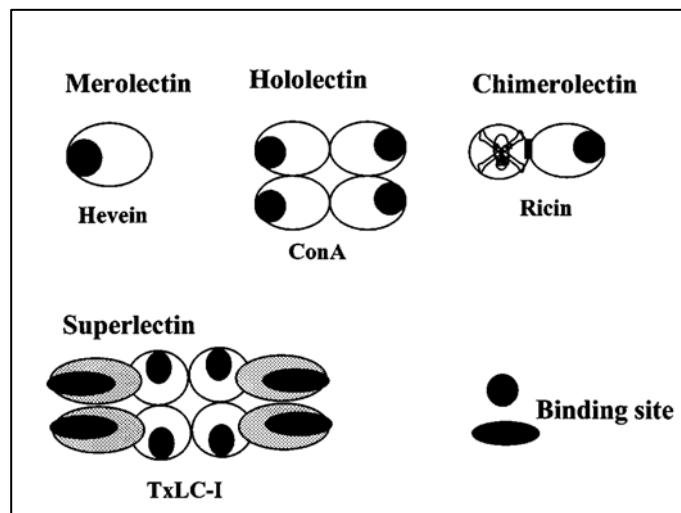
e) **I-type lectins**: Sialoadhesins include macrophage receptor CD22 present in B cells; CD33 present on early myeloid cells; Myelin-associated glycoprotein.

### 3. Macromolecular assemblies

Fimbriae of *E. coli*; *Klebsiella. pneumoniae*; *Pseudomonas aeruginosa* pilin.

#### 1.5.2. Based on the overall structure

On the basis of the overall structure of lectins, a distinction is made between so called hololectins and chimerolectins [6]. Hololectins consist solely of carbohydrate binding domains; in contrast, chimerolectins also contain an unrelated domain that acts independently of the lectin domain. If the hololectin is composed of a single carbohydrate binding domain, it is referred to as a merolectin, whereas hololectins with two different carbohydrate binding domains are also referred to as superlectins (Figure 1.2 and Table 1.5).



**Figure 1.2: Schematic representation of merolectins, hololectins, chimerolectins, and superlectins [adapted from Van Damme *et al.*, 1998].**

Table 1.5: Classification of lectins based on their overall structure

Class of lectin	Characteristics	Examples
<b>Merolectins</b>	Monovalent hence cannot precipitate glycoconjugates or agglutinate cells.	Hevein from latex of rubber tree ( <i>Hevea brasiliensis</i> )
<b>Hololectins</b>	Di- or multivalent hence can agglutinate cells and/or precipitate glycoconjugates. Exclusively contain (at least two) identical or very homologous carbohydrate-binding domains that recognize same or structurally related sugar(s).	Most plant lectins like Con A, soybean agglutinin, pea lectin, jacalin etc.
<b>Chimerolectins</b>	Fusion proteins containing one or more carbohydrate-binding domain(s) tandemly arrayed to an unrelated domain (with enzymatic or biological activity independent of carbohydrate recognition).	Ricin and abrin
<b>Superlectins</b>	Exclusively contain carbohydrate-binding domains that recognize structurally unrelated sugars.	Tulip lectin TxLC-1 with a Mannose and an unrelated GalNAc-binding domain

## 1.6. CLASSIFICATION OF PLANT LECTINS

On the basis of their monosaccharide specificity plant lectins were initially classified into five groups (Van Damme *et al.*, 1998): (i) Mannose/Glucose (Man/Glu) binding lectins, (ii) Galactose/*N*-acetyl Galactosamine (Gal/GalNAc) binding lectins, (iii) Glucose/*N*-acetyl Glucosamine (Glc/GlcNAc) binding lectins, (iv) Fucose specific binding lectins, and (v) *N*-acetyl neuraminic acid binding lectins. But lectins from the

same plant can show specificity for different sugars. Hence, a better method of classification was indispensable. On the basis of a comprehensive analysis of plant genome/transcriptome analyses, a system was devised in 1998, whereby the majority of all plant lectins known at the time were classified into seven families of structurally and evolutionarily related proteins [6]. Due to several recent developments in the identification of new plant lectins in the last decade, the classification required to be updated. At present, twelve different carbohydrate binding domains have been identified in plants [10]. According to biochemical and transcriptome data, all these domains are found in expressed proteins.

### 1.6.1. Overview of the structurally classified plant lectin families

#### 1. *Agaricus bisporus* agglutinin homolog

- Dimeric proteins with sub units comprising 140-142 amino acids.
- Monomers consist of a  $\beta$ -sandwich built from two bundles of  $\beta$ -sheets interconnected by a helix-loop-helix motif consisting of two short  $\alpha$ -helices.
- Recognize T-antigen (Gal $\beta$ 1,3GalNAc)  
Examples: *Marchantia polymorpha* ABA homolog (MarpoABA).

#### 2. Amaranthins

- Homodimers of 33 kDa protomers with 303 amino acids.
- Each homologous domain has a  $\beta$ -trefoil structure formed by six strands of anti parallel  $\beta$ -sheets capped by three  $\beta$ -hairpins into a short  $\beta$ -barrel. These two domains are linked by a short  $\alpha$ -helical  $3_{10}$  segment.
- Recognize T-antigen disaccharide Gal $\beta$ 1,3GalNAc.  
Example: *Amaranthus* species.

#### 3. Class V chitinase homologs with lectin activity

- Homodimer of 337 amino acids.
- Lack chitinase activity in spite of homology with class V chitinases.
- Contain the TIM-fold comprising of an inner crown of  $\beta$ -sheet strands surrounded by an outer crown of  $\alpha$ -helices, and an additional hairpin-shaped loop composed of three anti parallel strands of  $\beta$ -sheet that protrude from one edge of the TIM-barrel structure.
- Weak agglutinin but interact with high mannose N-glycans having the proximal pentasaccharide core structure.

Examples: *Robinia pseudoacacia* chitinase-related agglutinin.

#### 4. Cyanovirin family

- Named after a virucidal protein isolated from extracts of cyanobacterium *Nostoc ellipsosporum* (CV-N).
- Irreversibly inactivate human immunodeficiency virus (HIV) and simian immunodeficiency virus.
- Strong anti-HIV activity due to high affinity interactions with high-mannose N-glycans on viral surface glycoprotein gp120.
- CV-N consists of a single polypeptide with 101 amino acids with an internal duplication.
- Both repeats built up of a triple-stranded  $\beta$ -sheet and a  $\beta$ -hairpin.

Example: *Nostoc ellipsosporum*.

#### 5. EEA family

- *Euonymus europaeus* agglutinin (EEA) is a homodimeric protein with 152 amino acids.
  - Interaction with both blood group B oligosaccharides and high-mannose N-glycans.
- Examples: Lectins from *Euonymus europaeus* and other spindle trees.

#### 6. GNA family

- Earlier known as monocot-mannose binding lectins since exclusively found in monocots.
- Three four-stranded  $\beta$ -sheets organized around a pseudosymmetry axis form a typical  $\beta$ -prism fold.
- Recognize mannose weakly but exhibit a strong affinity toward oligomannosides and high-mannose N-glycans.

Examples: Tulip lectin (TxLC-I), *Galanthus nivalis* lectin (snow drop).

#### 7. Proteins with hevein domains

- Family of chitin-binding proteins with hevein domains.
- Diverse in terms of domain architecture of mature protomers and for the sometimes complex post-translational processing of primary translation products.
- Hevein is a single polypeptide of 43 amino acids that forms a typical structural motif consisting of three anti parallel strands of  $\beta$ -sheet associated to two short  $\alpha$ -helical segments. Overall fold stabilized by four disulphide bridges.
- Recognize tri-N-acetylchitotriose (NAG<sub>3</sub>) Hevein from latex of *Hevea brasiliensis*.

---

Examples: Class I and class IV chitinases with hevein domains, Potato lectin.

### 8. Jacalins

- Jacalin-related lectins (JRLs) divided in two classes: galactose-specific (GJRL) and mannose-specific (MJRL).
- GJRLs comprise jacalin and its homologs, built up of cleaved protomers, showing preference for galactose over mannose and located in vacuolar compartment.
- MJRLs are built up of uncleaved protomers, and exhibit exclusive specificity towards mannose and are located in the cytoplasmic/nuclear compartment.
- All jacalin domains, cleaved or not, contain the same  $\beta$ -prism fold formed by three Greek key motifs.

Examples: Jacalin from *Artocarpus integrifolia* and lectins from other *Artocarpus* species, *Maclura pomifera* agglutinin, Lectins from *Moraceae* family (*Artocarpus*, *Maclura*, *Morus*), *Calystegia sepium* lectin (hedge bindweed), *Castanea crenata* lectin and *Parkia platycephala* lectin.

### 9. Proteins with legume lectin domain

- Members with protomers of around 250 amino acids and highly diverse molecular structure depending on (i) number of subunits (one, two or four), (ii) absence/presence of proteolytic cleavage of protomer into smaller polypeptides, (iii) absence/presence of an interchain disulphide bond, and (iv) absence/presence of N-glycans.
- All protomers are built up of a flat seven-stranded  $\beta$ -sheet and a curved six-stranded  $\beta$ -sheet. Both  $\beta$ -sheets are interconnected by turns and loops to form a flattened dome-shaped  $\beta$ -sandwich structure, commonly referred to as the jelly-roll tertiary fold.
- The protomers contain a single carbohydrate binding site formed by four loops protruding at the top of the dome.
- They also harbour a hydrophobic cavity at one side of the  $\beta$ -sandwich, to accommodate hydrophobic ligands in some lectins.
- Metalloproteins with two divalent cations incorporated in their structure –  $\text{Ca}^{2+}$  linked to the Asp residue of the carbohydrate binding site, and a  $\text{Mn}^{2+}$  ion linked in the vicinity of carbohydrate binding site.

Examples: *Canavalia ensiformis*, *Phaseolus vulgaris*, *Glycine max*, *Arachis hypogea*, *Pisum sativum*

**10. LysM domains**

- The Lysin domain was first identified in LysM-domain containing receptor-like kinases from legumes involved in perception of rhizobial signals.
- The overall fold in this domain corresponds to a  $\beta$ - $\alpha$ - $\alpha$ - $\beta$  secondary structure in which two helices pack onto the same side of an anti parallel  $\beta$ -sheet. A shallow groove on surface of protein acts as a binding site.
- Interact with either chitin-oligosaccharides or structural analogs substituted with fatty acids.

Examples: *Arabidopsis thaliana*, *Pteris ryukyuensis*, *Selaginella moellendorffii*.

**11. Nictaba family (formerly *Cucurbitaceae* phloem lectins)**

- Chitin-binding lectins from phloem exudates of *Cucurbitaceae* family, containing an extra N-terminal sequence of around 65 residues and a short five amino acid residue C-terminal extension with two Cys residues.
- Subunits comprise of around 220-230 amino acids.

Examples: *Nicotiana tabacum* agglutinin (Nictaba).

**12. Ricin-B family**

- Basic carbohydrate binding unit consists of three tandemly arrayed sub domains of around 40 amino acids.
- Type 2 RIPs are typical chimeric proteins built up of an A-chain with a polynucleotide:adenosine glycosidase domain and a B-chain with a duplicated ricin-B domain.
- Bifunctional proteins with carbohydrate and enzymatic activity.
- Potent cytotoxins inside the cell.
- The lectinic B-chain consists of two tandemly arrayed homologous ricin-B domains; both comprise three homologous subdomains ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and an additional subdomain  $\lambda$ . Each domain is built up of three bundles of four short  $\beta$ -strands arranged into a  $\beta$ -trefoil fold around a threefold symmetry axis. Disulphide bridges stabilize the  $\beta$ -trefoil.
- Most plant lectins with ricin-B domains are considered galactose or GalNAc-specific.
- Few species of *Sambucus* recognize sialylated glycans.

Examples: *Ricinus communis* agglutinin, *Viscum* sp., *Sambucus* sp., *Abrus precatorius*, *Adenia* sp.

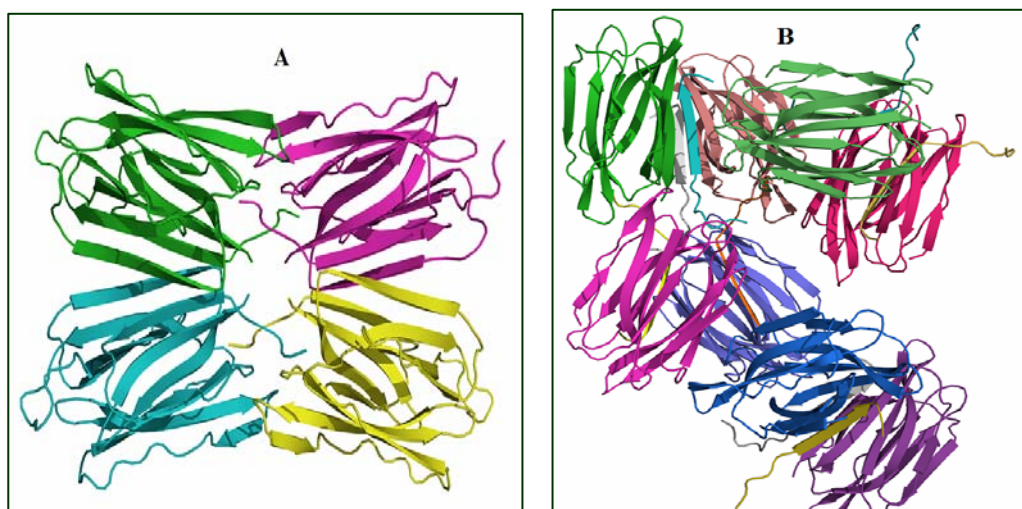


### 1.7. LECTIN-SUGAR INTERACTIONS

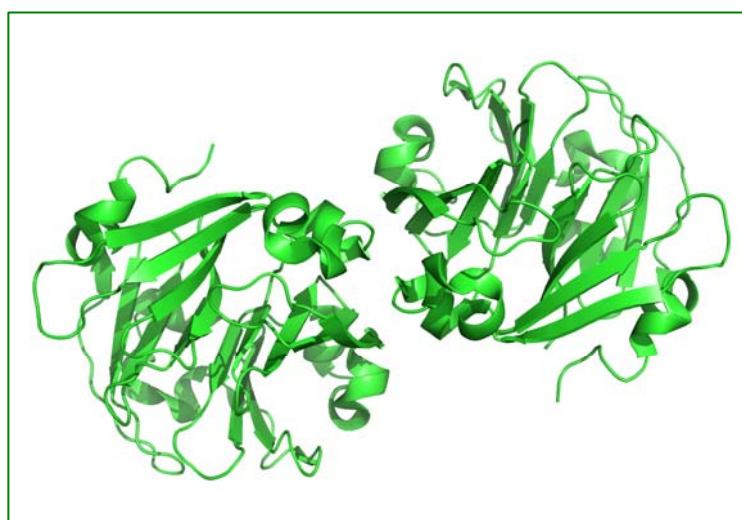
Many biological recognition events require the participation of lectins. This kind of recognition requires a strict selectivity. Hence, a stringent geometry is imposed upon both the ligand and the corresponding receptor. This, in turn, confers unique sugar-specificity upon the lectins [31, 32]. Carbohydrates can interact with lectins *via* hydrogen bonds, metal coordination bonds and van der Waal's and hydrophobic interactions. Selectivity is a result of specific hydrogen bonding and/or metal coordination bonds with important hydroxyls of the carbohydrates that act as both acceptors and donors of hydrogen bonds. Water molecules serve as bridges in these interactions. Further fine-tuning of the saccharide specificity of the lectin is achieved by steric exclusions that minimise unwanted recognition. The binding selectivity is increased diversely due to subsite binding and subunit multivalency, where possible [33, 34]. In the case of subsite binding, the primary binding site appears critical for carbohydrate recognition, but secondary binding sites contribute to enhanced affinity of the lectin towards specific oligosaccharides. For example, the legume lectins *Lathyrus ochrus* isolectin II (LOL II) and Con A are both Man/Glc specific lectins, but their oligosaccharide preferences are very different. LOL II has several-fold higher affinity for oligosaccharides that have additional  $\alpha$  (1–6)-linked fucose residues while Con A does not [for reviews see 35, 36]. In subunit multivalency, several subunits of the same lectin contribute to the binding by recognising different extensions of the carbohydrate or different chains of a branched oligosaccharide. This kind of binding is exhibited, among others, by the asialoglycoprotein receptor, the mannose binding protein (MBP) from the serum, the chicken hepatic lectin and the cholera toxin.

## 1.8. MOST RECENTLY REPORTED CRYSTAL STRUCTURES OF PLANT LECTINS

The crystal structures of two lectins from *Artocarpus integer* (Champedak) have been solved in 2013. These structures are given in Fig. 1.3.



**Figure 1.3:** (A) Mannose-binding lectin from Champedak (CMB) (*Artocarpus integer*) (PDB ID: 4AKD) (2.1 Å resolution) and (B) Galactose-binding lectin from Champedak (CGB) (*Artocarpus integer*) (PDB ID: 4AK4) (1.65 Å resolution). CGB recognizes and binds the T-cell antigenic determinant Gal $\beta$ 1-3GalNAc in IgA and C1 inhibitor molecules [36]. This binding has been utilized as a tool to study the carbohydrate structures of IgA1 from patients with IgA nephropathy.



**Figure 1.4:** Plant albumin from *Cicer arietinum* (PDB ID: 3V6N) (2.2 Å resolution) (Sharma U, Suresh CG, 2011) (Four  $\beta$ -propellar folds).

## 1.9. APPLICATIONS OF LECTINS

Lectins with well defined specificity are excellent tools for several analytical and preparative purposes. Plant lectins have innumerable applications, both direct and indirect, either in their native form or as soluble and immobilized derivatives. As these proteins have high affinity for carbohydrates, they have always been important as molecular tools in the identification, purification and stimulation of specific glycoproteins on human cells. The understanding of this molecular basis of plant lectin-carbohydrate interactions and of the intracellular signalling events can prove very promising for the design of novel drugs for the treatment of infectious, inflammatory and malignant diseases. It may also be of help for the structural and functional investigation of glycoconjugates and their changes during physiological and pathological processes. Some of the uses of plant lectins are enumerated below.

### Biochemistry

1. Purification of lectin-reactive glycoconjugates and characterization of glycans using lectin-affinity chromatography. Examples include the use of Con A lectin (for mannose) and lectins from *Lens culinaris* (for glucose and mannose), *Triticum vulgare* (for N-acetyl-D-galactosamine), *Ricinus communis* (for galactose) etc.
2. Glycome analysis (glycomics). Use of *Phaseolus vulgaris leucoagglutinin* (L-PHA), *P. vulgaris erythroagglutinin* (E-PHA), *Aleuria aurantia* (AAL), and *Datura stramonium* (DSL) in the glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer [37].
3. Quantification of lectin-reactive glycoconjugates in enzyme-linked lectin binding assays (ELLA).
4. Quantification of enzymatic activities (of glycosyltransferases/glycosidases) by specific detection of products of these reactions using lectins.
5. As microarrays for high throughput analysis of protein glycosylation. Some representative examples include lectins like Con A (specific for branched and terminal mannose and terminal GlcNAc residues, *Galanthus nivalis* (GNA) (for terminal  $\alpha$ -1,3 mannose), *Griffonia simplicifolia* I (GS-I) (for  $\alpha$ -galactose), *Griffonia simplicifolia* II (GS-II) (for terminal GlcNAc), *Maackia amurensis* (MAA) (for  $\alpha$ -2,3 sialic acid, *Glycine max* (SBA) (for terminal GalNAc), *Sambucus nigra* (SNA) (for  $\alpha$ -2,6 sialic acid), *Ulex europaeus* (UEA) (for  $\beta$ -fucose), and *Triticum vulgare* (WGA) (for  $\beta$ -GlcNAc, sialic acid, GalNAc) [38].

---

6. Assessment of the functionality of ligands on carbohydrate- presenting scaffold (e.g. glycodendrimers) (Example - Con A).

### **Cell biology**

1. Study of cell-cell interactions and signalling functions in the haemopoietic, immune and nervous systems. Example: Siglecs – sialic acid-binding immunoglobulin-like lectins.
2. Fluorescent tagged (FITC-labelled) lectins have been used to study cell-cell interactions in model marine biofilms. Examples include Concanavalin A (specific for glucose and mannose), *Bandeiraea simplicifolia* lectin (specific for galactose and *N*-acetylgalactose), *Tetraconolobus purpurea* lectin (specific for fucose), *Arachis hypogaea* lectin (specific galactosyl-*N*- galactose), succinyl concanavalin A (specific for glucose and mannose), and *Phaseolus vulgaris* lectin (specific for general oligosaccharides) [39].
3. Selection of cell variants (mutants, transfectants) with altered lectin binding properties to analyse glycosylation machinery and glycan functionality.
4. For profiling overall changes in the surface glycomes of mammalian and bacterial cells.
5. Fractionation of cell populations.
6. Monitoring cell proliferation and identification of the signalling pathways.
7. Model system to study of cell aggregation, adhesion, and migration.

### **Medicine**

1. Histo-blood group typing.
2. Quantification of aberrations of cell surface glycan presentation, e.g. in malignancy.
3. Cell marker for diagnostic purposes including marking infectious agents (viruses, bacteria, fungi, parasites).
4. Carriers for delivery of chemotherapeutic agents (e.g. ricin and abrin) and as therapeutics in anticancer treatment (e.g. European mistletoe (*Viscum album*) lectin). These lectins inhibit tumor growth, especially by causing cytotoxicity, apoptosis, and down-regulation of telomerase activity and inhibition of angiogenesis [40]. Examples: Ricin and abrin have been coupled to specific monoclonal antibodies for cancer therapy. Con A has been shown to inhibit melanoma cells in mice; PHA (phytohemagglutinin) is known to inhibit non-Hodgkin's lymphoma in mice; WGA (wheat germ agglutinin) has been shown to inhibit colon carcinoma in rat and colorectal cancer in humans.

5. As biomarkers of oral cavity tumours. Examples - peanut agglutinin (PNA) and *Ulex europeaus* agglutinin (UEA-1), while Con A and UEA-1 have been used as histochemical markers of neoplastic glandular tissue.

The subsequent sections in this chapter give a brief account on protein folding as well as bioinformatics since the present thesis deals with the conformational and *in silico* characterization of the lectin.

## **B. PROTEIN FOLDING**

### **1.10. THE FUNDAMENTAL MECHANISM OF PROTEIN FOLDING**

Protein molecules represent a remarkable relationship between structure and function at the molecular level. The surface pattern of proteins regarding their shape, charge and hydrophobicity, determines their diverse and highly specific function. This surface pattern is determined by the unique three-dimensional structure of the polypeptide chain, while the covalent structure is defined by the order in which the constituent amino acids are linked. Anfinsen's famous experiments in the 1960s have established the fact that folding and the resulting native structure of proteins are governed and determined by the amino acid sequence of a particular protein and its natural solvent environment [41].

The process by which a linear polypeptide chain transforms into a specific three-dimensional functional structure is known as protein folding. This mechanism of protein folding has until recently been covered in mystery. The process of protein folding begins from the unfolded state and proceeds to the native or misfolded states via various kinds of folding intermediates. Revealing the structural and dynamic properties of these states, particularly the unfolded states and folding intermediates, at an atomic level is crucial for understanding protein folding pathways [42].

### **1.11. THE PROTEIN-FOLDING PROBLEM**

The protein folding problem is the question of how a protein's amino acid sequence dictates its three-dimensional atomic structure. The protein-folding problem deals with three main questions [43]: (i) The physical folding code: How is the 3D native structure of a protein determined by the physicochemical properties that are encoded in its 1D amino-acid sequence? (ii) The folding mechanism: A polypeptide

chain has a vast number of possible conformations. How can proteins fold so fast? (iii) Predicting protein structures using computers: Can a computer algorithm be designed that can predict a protein's native structure from its amino acid sequence? Such an algorithm might overcome the time-consuming process of experimental protein structure determination and accelerate the discovery of protein structures and new drugs.

### **1.11.1. The Physical Code of Protein Folding**

What forces drive a protein to its three-dimensional folded structure? The following factors appear to contribute [44]: (i) Hydrogen bonds. (ii) van der Waals interactions. (iii) Backbone angle preferences. (iv) Electrostatic interactions. (v) Hydrophobic interactions. (vi) Chain entropy. Opposing the folding process is a large loss in chain entropy as the protein collapses into its compact native state from its many open denatured configurations.

### **1.11.2. The Mechanism of Protein Folding**

Despite the huge number of conformations accessible to it, how does a protein molecule fold to its one precisely defined native structure so quickly (microseconds, for some proteins)? How does the protein "know" what conformations not to search? To answer this question, a major experimental hunt to characterize the kinetics of protein folding and identify folding intermediates, was carried out. This, in turn, led to the development of new powerful experimental methods, including fluorescence labelling, mutational studies, hydrogen exchange, etc. [45] (see Table 1.6 for details).

### **1.11.3. Computing Protein Structures from Amino Acid Sequences**

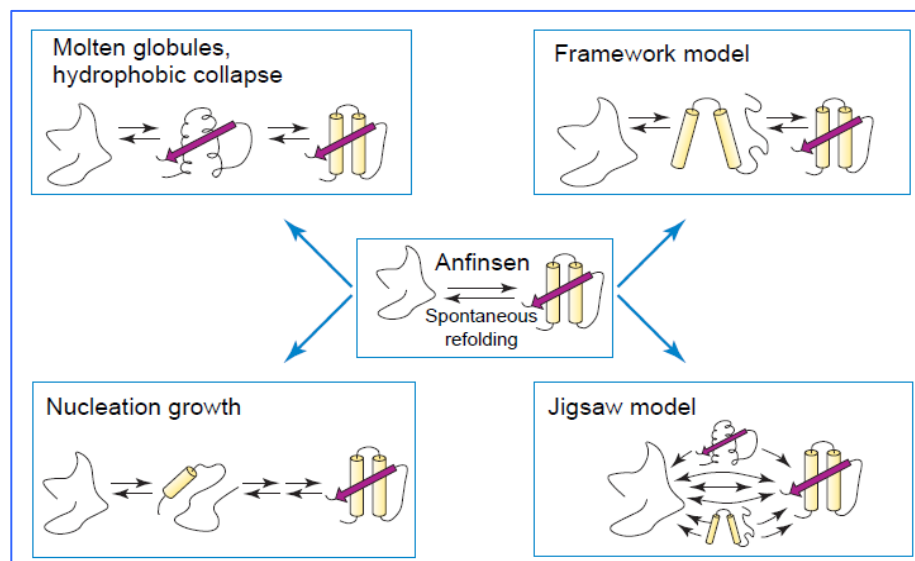
It has been a huge challenge to develop a computer algorithm that can predict the three-dimensional native structure of a protein from its amino acid sequence. The knowledge of native structures functions as a good starting point for understanding the biological mechanisms and for the discovery of new drugs that can inhibit or activate those proteins. But on the other hand, around 1000-fold more protein sequences are known compared to their structures and this gap is increasing because of developments in high-throughput sequencing methods. So it is important to devise methods that could accurately predict structures from sequences. Currently, all successful structure-prediction algorithms are based on the assumption that similar sequences lead to similar structures. These methods are heavily dependent on the Protein Structure Databank (PDB), which now contains more than 80,000 structures. However, many of these

structures are similar and the PDB contains only around 4000 structural families and 1200 folds [46].

### 1.12. MODELS TO EXPLAIN PROTEIN FOLDING

Many physical models have been proposed to explain pathways of protein folding. These include the nucleation growth model [47, 48], framework model [49, 50], diffusion-collision mechanism model [51, 52], hydrophobic collapse model [53, 54], and jigsaw model [55] (see Figure 1.5).

The recently developed free-energy folding funnel model [56] proposes that during the early stages of folding, a protein possesses a large ensemble of structures. Folding does not involve finding a single route to the native state. Instead, it involves the description of the topography of the free-energy landscape involving different statistical combinations; this in turn characterizes the dynamics of the ensemble. The process of folding is easy if the landscape resembles a multidimensional funnel that leads to the native structure through numerous pathways.



**Figure 1.5: Early models for mechanisms of folding.** Anfinsen's original experiments demonstrated that proteins fold spontaneously and reversibly into their native conformation. (1) The nucleation growth model proposed that residues adjacent in sequence form a nucleus from which the native structure then develops in a sequential manner. By contrast, (2) the framework model suggested that local elements of secondary structure form first and these then dock into the native tertiary structure of the protein, possibly by a (3) diffusion–collision mechanism. In the (4) hydrophobic collapse model, a protein buries its hydrophobic side chains from solvent

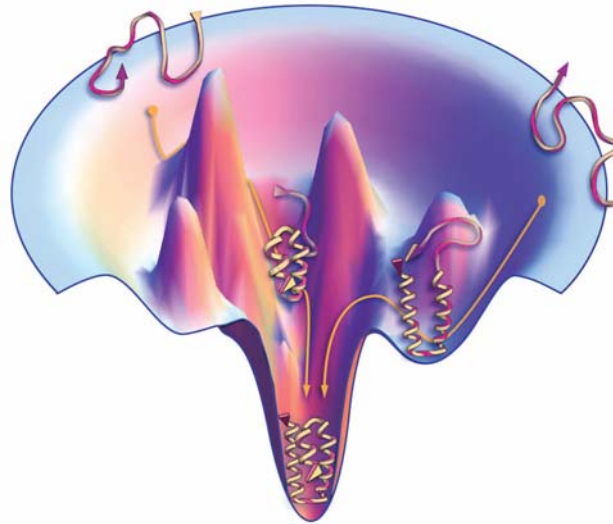
water early during folding, forming a collapsed intermediate or molten globule species, from which the native state develops by searching within this conformationally-restricted state. Finally, (5) the jigsaw model suggests that each protein molecule could fold by a different path. Although it is now clear that there is not a single sequential folding route, as was implicit in some of these early models, features of these models are relevant today in the context of energy landscape models of folding (Adapted from Radford, 2000 [57]).

Onuchic *et al.* first estimated the folding funnel for a fast-folding 60-residue helical protein [56, 58, 59]. The width and depth of the funnel represent the entropy and energy, respectively. The authors showed the flow of the molecule through the molten globule, folding bottleneck, or transition state ensemble and then through a glass transition region in which discrete pathways exist. The free-energy folding funnel model has been used to illustrate folding in the presence or absence of intermediates [60, 61], folding of transition states [62], folding in confinement [63, 64], folding and aggregation [65, 66], and chaperone-assisted folding [67].

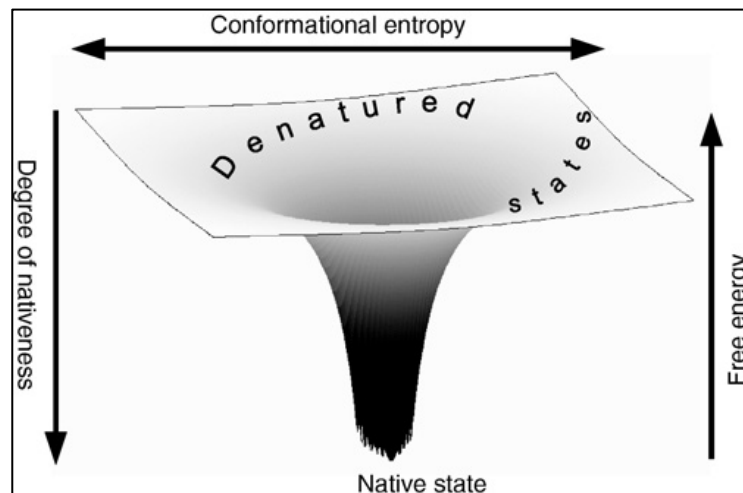
### **1.13. FREE-ENERGY LANDSCAPES OF PROTEINS**

Energy landscapes are often visualized as a surface in three dimensions. The vertical axis represents the free energy and the horizontal axes represent the conformational degrees of freedom available to the polypeptide chain. In a real protein, most of the interactions that can form between parts of the chain are mutually supportive and cooperatively lead to a low-energy structure which is therefore “minimally frustrated.” This “principle of minimal frustration” [68], understood from simplified models of proteins and the theory of spin glasses, led to the realization that the energy landscape of a real protein should be shaped like a funnel.





**Figure 1.6: Energy landscape.** Proteins have a funnel-shaped energy landscape with many high energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories (Adapted from Dill and MacCallum, 2012).



**Figure 1.7:** The width of the funnel represents the conformational freedom of the chain. The vertical axis represents the free energy; as free energy decreases, the nativeness of the chain increases. Denatured (unfolded) states are at the top of the funnel while the native state is the global minimum. There is some ruggedness (roughness) in the energy landscape near the native state (Adapted from Szilagyi *et al.*, 2007).

The top of the funnel, representing the non-native states, is wide (the conformational entropy is high), and it narrows as one gets closer to the bottom: near-native states represent more compact conformations, and therefore have low conformational entropy.

### 1.14. BIOPHYSICAL TECHNIQUES TO STUDY PROTEIN FOLDING

In the last decade a wide range of physical and chemical methods complemented the fundamental techniques to study protein folding. A brief summary of the advances of these methods is mentioned below (Table 1.6).

**Table 1.6 Experimental techniques to study of protein folding<sup>a</sup>**

Technique	Time scale <sup>b</sup>	Structural parameter probed	Refs
<b>Fluorescence</b> (i) Intrinsic (ii) ANS binding (iii) Substrate binding (iv) FRET (v) Anisotropy	ns-s	Environment of Trp and Tyr Exposure of hydrophobic surface area Formation of the active site Inter-residue distances Correlation time	69, 70
<b>Circular dichroism</b> (i) Far-UV (ii) Near-UV	ns-s	Secondary-structure formation Tertiary-structure formation	71, 72
<b>Protein engineering</b>	Depends on method of detection	Role of individual residues in stabilizing intermediates and transition states	73
<b>Small-angle X-ray scattering</b>	≥ ms	Dimension and shape of polypeptide	74
<b>Absorbance (near-UV)</b>	ns-s	Environment of aromatic residues or cofactors	69, 70
<b>FTIR</b>	ns-s	Secondary structure formation	75
<b>NMR</b> (i) Real time (ii) Dynamic NMR	ms-s 250 μs	Environment of individual residues Line shape analysis provides folding-unfolding rates close to equilibrium	76, 77
<b>Hydrogen exchange (HX)</b> (i) Native state (ii) Pulsed HX NMR (iii) Pulsed HX ESI MS	min-months ms-s ms-s	Global stability and metastable states Hydrogen-bond formation in specific residues Folding populations	78, 79
<b>Force spectroscopy (AFM or optical tweezers)</b>	s	Unfolding forces and unfolding-rate constants of single molecules	80

<sup>a</sup> The table was adapted from references Radford, 2000 and Brockwell, 2000.

<sup>b</sup> The timescale available depends on the method used to trigger folding. The fastest reactions can be monitored using optical methods and temperature perturbations (ns); ultra-rapid mixing methods permit measurements following the dilution from denaturant on the msec timescale, whereas stopped flow methods provide information on the ms timescale.

Abbreviations: AFM, atomic force microscopy; ANS, 1-anilino naphthalene sulphonic acid; ESI MS, electrospray ionization mass spectrometry; FRET, fluorescence resonance energy transfer; FTIR, fourier transform infra-red.

### 1.15. PROTEIN MISFOLDING AND AGGREGATION

Native states of proteins almost always represent thermodynamically the most stable conformation under physiological conditions [81]. All the information regarding the native structure is hidden in the amino acid sequence. However, correct folding is a challenge for proteins in a living cell and only a part of the proteins can assume their native structure spontaneously. In the crowded milieu of the cell, efficient protein folding and transport depend on the presence of a complex machinery of chaperones, chaperonins, and cofactors. The primary mission of this machinery is to prevent the aggregation of nascent polypeptide chains and proteins that unfold upon environmental stress [82].

#### **Intermolecular interactions among proteins are not always beneficial**

Proteins and peptides carry out a huge number of functions required for survival and maintenance of life with the help of complex and usually transient networks of intermolecular interactions. But establishing protein intermolecular contacts is not always beneficial for the fitness of the organism, as can be seen from the increasing evidence regarding the rising number of human pathologies, like the Alzheimer's and Parkinson's diseases, which are associated with the formation of abnormal interactions between adjacent protein molecules [83].

Different strategies can be employed during normal cell function to minimize the occurrence and persistence of these aberrant interactions, but they can not be avoided totally [84]. This is because the same generic physicochemical principles are involved both in the rapid formation of native interfaces as well as in the formation of non-native contacts [85-87]. Such aberrant interactions then result in misfolding or misbinding events in proteins or protein complexes, thus exposing previously hidden

regions that can now establish specific, but unwanted contacts; these contacts in most cases cause self-assembly and formation of large and insoluble structures.

More importantly, it has now become clear that such abnormal contacts are not encoded in the protein sequence, or at least favoured by it, in contrast to the contacts that allow the formation of native structures or complexes. Yet highly regular structures known as the amyloid fibrils are formed due to these anomalous interactions. In fact, the ability to adopt amyloid-like conformations might be a fundamental property of polypeptide chains [42]. It is these aggregates and the different intermediates that accumulate in the pathway that exert the cytotoxic effect [88-90]. Quite often the conformational diseases arising due to the formation of protein aggregates exhibit degenerative pathologies. This may have a major impact on the health of the elderly in the rapidly ageing and may cause a crisis in public health services in the near future. Therefore, the study of protein aggregation in protein chemistry has become imperative in biological and medical sciences, mainly because understanding what allows polypeptides to remain soluble in the crowded cell and what happens when this property is lost, is important to cure these pathologies.

### **Types of aggregates**

The abnormal association of misfolded proteins usually leads to the formation of aggregates. Small aggregates can remain soluble, but large protein aggregates precipitate out of solution under physiological conditions. From the morphological point of view we can differentiate essentially two types of aggregates: amorphous aggregates and amyloid fibrils. Amorphous aggregates display a granular appearance when imaged by electron microscopy and consist mostly of disordered polypeptide chains, even if they display certain regions enriched in  $\beta$ -sheet structures, which cement the macromolecular assembly [91, 92]. Amyloid fibrils are highly ordered and repetitive structures where all polypeptides adopt a common fold. Amyloid fibrils are thread-like protein aggregates with a core region formed by repetitive arrays of  $\beta$ -sheets oriented perpendicularly to the fibril axis forming a structure known as cross- $\beta$  because it displays two sets of characteristic X-ray diffraction signals, forming a “cross” pattern.

### **Causes of unwanted protein aggregation**

Unwanted protein aggregation can be caused due to several reasons.

- When the ubiquitin-proteasome protein degradation system is unable to eliminate misfolded, aggregation prone molecules, it causes the malfunctioning of the

cellular protein quality-control machinery, thus resulting in unwanted protein aggregation [93, 94].

- Unwanted protein aggregation can also result from insufficient functioning of the chaperone machinery [95] or when the normal cellular transport route of a protein is damaged.
- Inappropriate protease activity can produce amyloidogenic protein fragments that can, in turn, lead to aggregation. An increased expression level or pathological mutations that destabilize the protein structure can cause and promote intermediate amyloid conformation or aggregation [96].

**Table 1.7 Some protein misfolding diseases.**

<b>Diseases</b>	<b>Protein involved</b>	<b>Cause</b>
Cruetzfeldt-jacob	Prion protein	Toxic folding /aggregation
Alzheimer's	Beta-amyloid	Toxic folding/aggregation
Cystic fibrosis	CFTR	Misfolding
Cancer	Protein 53 (transcription factor)	Misfolding
Familial amyloid polyneuropathy	Apolipoprotein	Aggregation
Parkinson's	Alpha-synuclein	Amyloid fibril formation
Huntington's	Huntington	Amyloid fibril formation
Familial visceral amyloidosis	Lysozyme	Aggregation
Cataract	Crystallins	Aggregation
Marfan syndrome	Fibrillin	Misfolding
Scurvy	Collagen	Misfolding
Osteogenesis imperfect	Type 1 procollagen	Misassembly
Amyotrophic lateral sclerosis	Superoxide dismutase	Misfolding
Finnish type familial amyloidosis	Gesolin	Amyloid fibril formation
Kelandic cerebral angiopathy	Cystatin C	Amyloid fibril formation
Type II diabetes	Islet amyloid polypeptide	Amyloid fibril formation

### 1.16. MOLECULAR CHAPERONES

Molecular chaperones are facilitators and regulators of protein conformational change. Importantly, they do not provide information with regard to the final protein structure. Chaperones bind to and stabilize conformers of other proteins, and via cycles of regulated binding and release, they are able to facilitate the correct fate of their client protein and reduce protein aggregation. Through this mechanism, molecular chaperones play an essential role in many cellular processes. They have a principal role in protein folding, where they are involved in the *de novo* synthesis of polypeptides, transport across membranes and the refolding of proteins denatured by environmental stress.

### 1.17. APPLICATIONS OF PROTEIN FOLDING IN INDUSTRY

Extensive research is focused on understanding the structural basis of protein folding and stability and mechanistic role of early folding intermediates. Finding a solution to protein folding problem has many practical applications.

- **Predicting structure of protein**

Our understanding of protein folding has been greatly advanced through computational models and simulations. They provide information complementary to that obtained from experiments. In fact, there is a synergy between theory and experiment: theory provides testable models and experiments provide the means to test and validate the models. In particular, simulations can help identify or predict transition and intermediate states along the folding pathway, provide predictions of the rate of folding and in some cases, predict the final, folded structure.

- **Solving the protein aggregation problem**

Protein aggregation *in vivo* is a widespread phenomenon that arises from early folding intermediates through kinetic competition between proper folding and misfolding. A wide range of human diseases resulting from incorrect folding or altered conformation of proteins (see Table 1.7) have been found. Possible treatment of these diseases could exploit detailed knowledge of protein folding and the prevention of abnormal folding.

- ***De novo* designing**

An important field of research dependent upon our understanding of protein folding is *de novo* protein designing for constructing completely new positions with determined function. DeGrado *et al.* [97] have reviewed the different methods for *de novo* protein design, which usually consist of choosing a function of protein followed

by searching a protein's scaffold capable of supporting the reactive groups in desired geometry. It is then necessary to determine an amino acid sequence capable of folding into an adequate and stable three-dimensional structure. This can be done by either exploring genetic methods or by combinatorial and computational algorithms. Success in functionally-active designed polypeptide has been achieved in the field of catalysis, metal ion and heme binding, and introduction of cofactors. It is now evident that *de novo* protein design represents a growing field of research that will be useful both in testing the principles of protein folding and in offering the perspectives to design new proteins with practical applications for pharmaceuticals and diagnostics.

- **Treating protein misfolding**

The purpose of studying any human disease is to find ways to treat it. The story of protein folding has not yet led to treatments for the diseases involved, but this could happen within the next decade. The key lies in the discovery of a small molecule, a drug that can either stabilize the normally folded structure or disrupt the pathway that leads to a misfolded protein. Although, many molecular biologists and protein chemists believe this will be quite difficult, others are more optimistic.

### **1.18. UNSOLVED PROBLEMS OF PROTEIN PHYSICAL SCIENCE**

Is the protein-folding problem “solved” yet [3]? Protein folding is now a whole field of research—a large, growing, and diverse enterprise. A few old puzzles generate more new puzzles. For the field of protein physical science, the future is at least as convincing as the past. Some of the unsolved problems are:

- Little experimental knowledge of protein-folding energy landscapes.
- Lack of consistent prediction of the structures of proteins to high accuracy.
- Absence of a quantitative microscopic understanding of the folding routes or transition states for arbitrary amino acid sequences.
- Inability to predict a protein's propensity to aggregate, which is important for aging and folding diseases.
- Non-availability of algorithms that accurately give the binding affinities of drugs and small molecules to proteins.
- Inability to understand why a cellular proteome does not precipitate, because of the high density inside a cell.
- Little knowledge about how folding diseases happen, or how to intervene.

- Relatively little knowledge about the structure, function, and folding of membrane proteins [98, 99] in spite of their importance.
- Limited information about the ensembles and functions of intrinsically disordered proteins [100], even though nearly half of all eukaryotic proteins contain large disordered regions.

This is sometimes called the “protein nonfolding problem” or “unstructural biology.”

## C. BIOINFORMATICS

Bioinformatics is often defined as the application of computational techniques to understand and organise the information associated with biological macromolecules. The aims of bioinformatics are threefold. First, to begin with, bioinformatics organises data in a manner that allows researchers to access existing information and to submit new entries as they are produced, e.g. the Protein Data Bank for 3D macromolecular structures. The second aim is to develop tools and resources that aid in the analysis of data. The third aim is to use these tools to analyse the data and interpret the results in a biologically meaningful manner. Traditionally, biological studies examined individual systems in detail, and frequently compared them with a few that are related. In bioinformatics, global analyses of all the available data can be conducted with the aim of uncovering common principles that apply across many systems and highlight novel features.

One of the major objectives of computational biology has been to predict a protein’s three-dimensional native structure from its amino acid sequence. This could help to (a) accelerate drug discovery by replacing slow, expensive structural biology experiments with faster, cheaper computer simulations, and (b) annotate protein function from genome sequences. With the rapid growth of experimentally determined structures available in the Protein Databank (PDB), protein structure prediction has become as much a problem of inference and machine learning as it is of protein physics.

A large amount of carbohydrate-lectin functional and interaction data is being generated by major international glycomics efforts. As a result, computational strategies can be pursued to both augment the existing structural knowledge and to provide new insights into the structural aspects of carbohydrate recognition by lectins. In this regard, computational (*in silico*) techniques are particularly useful. Homology



modeling can be used to generate the structure of a lectin in the absence of an experimentally determined structure. Molecular docking can be used to investigate carbohydrate recognition by a lectin where that particular carbohydrate–lectin complex is unavailable. Molecular dynamics can be used to study the movement of the protein over time, and how this movement affects ligand binding.

#### **D. PRESENT INVESTIGATION**

Literature survey of lectins and their applications in various areas of biology has been presented in this chapter. The importance of studying protein folding and bioinformatics also has been discussed. Our group at the CSIR-National Chemical Laboratory (Pune, India) has been working in the area of protein chemistry, lectins and carbohydrate-protein interactions, for the last 23 years. Lectins from microbial, animal and plant sources have been purified and characterized for their biochemical and biophysical properties in our group. A lectin from *Cicer arietinum* (chick pea) seeds showing complex sugar specificity was purified and characterized to some extent. But it proved difficult to crystallize this lectin, despite several modifications in the crystallization procedures. Hence, molecular cloning of the lectin gene in *Escherichia coli* was initiated to carry out the structural analysis of the protein.

The present work describes the cloning and expression of the *Cicer arietinum* (chick pea) seed lectin in *Escherichia coli*. Conformational characterization of the recombinant lectin (rCAL) was performed using biophysical and bioinformatics tools. Binding properties of the recombinant lectin were studied using hemagglutination assay as well as fluorescence spectroscopy. The microenvironment of the tryptophan residue was investigated with solute quenching studies. The behaviour of the lectin under different denaturing conditions was studied with fluorescence spectroscopy and circular dichroism. Homology model of the lectin was constructed to evaluate/visualize the environment of the tryptophan in the lectin. Docking simulations of the ligands were carried out to identify their binding sites. Amino acid residues showing propensity for aggregation were identified from different web-based servers. The computational results were found to correlate well with the results obtained from experimental studies.

---

**REFERENCES**

1. Stillmark, H. (1888), Über Ricin ein giftiges Ferment aus den Samen von *Ricinus communis* L. Und einige anderen Euphorbiaceen, *Inaugural Dissertation Dorpat*, Tartu.
2. Liener, I.E. (1976), Phytohemagglutinins (Phytolectins), *Annual Review of Plant Physiology*, **27**, 291–319.
3. Etzler, M.E. (1985), Plant lectins: Molecular and biological aspects, *Annual Review of Plant Physiology*, **36**, 209–234.
4. Drickamer, K. and Taylor, M.E. (1993), Biology of animal lectins, *Annual Review of Cell Biology*, **9**, 237–264.
5. Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1998), Legume lectin structure, *Biochimica et Biophysica Acta*, **1383**, 9–36.
6. Van Damme, E.J.M., Peumans, W.J., Barre. A. and Rougé, P. (1998), Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles, *Critical Reviews in Plant Sciences*, **17**, 575–692.
7. Lis, H. and Sharon, N. (1998), Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition, *Chemical Reviews*, **98**, 637–674.
8. Gabius, H.J., Siebert, H.C., Andre, S., Jimenez-Barbero, J. and Rudiger, H. (2004), Chemical Biology of the Sugar Code, *ChemBioChem*, **5**, 740–764.
9. Dam, T.K. and Brewer, C.F. (2010), Lectins as pattern recognition molecules: The effects of epitope density in innate immunity, *Glycobiology*, **20**, 270–279.

- 
10. Van Damme, E.J.M., Lannoo, N. and Peumans, W.J. (2008), Plant Lectins, In: Kader, J.C. and Delseny, M. Editor(s), *Advances in Botanical Research*, Academic Press, **48**, 107–209.
  11. Liu, B., Bian, H.J. and Bao, J.K. (2010), Plant lectins: Potential antineoplastic drugs from bench to clinic, *Cancer Letters*, **287**, 1–12.
  12. Gabius, H.J., Andre', S., Jime'nez-Barbero, J., Romero, A. and Soli's, D. (2011), From lectin structure to functional glycomics: principles of the sugar code, *Trends in Biochemical Sciences*, **36**, 298–313.
  13. Ghazarian, H., Idoni, B. and Oppenheimer, S.B. (2011), A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics, *Acta Histochemica*, **113**, 236–247.
  14. Vandenborre, G., Smaghe, G., Van Damme, E.J.M. (2011) Plant lectins as defence proteins against phytophagous insects, *Phytochemistry*, **72**, 1538–1550.
  15. Singh, H. and Sarthi, S.P. (2012), Insight of Lectins – a review, *International Journal of Scientific and Engineering Research*, **3**, 1–9.
  16. Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T. and Sharon, N. (1980), What should be called a lectin? *Nature*, **285**, 66.
  17. Boyd, W.C. and Reguera, R.M. (1949), Studies on haemagglutinins present in seeds of some representatives of the family Leguminosae, *Journal of Immunology*, **62**, 333–339.
  18. Burger, M.M. (1974), Assays for agglutination with lectins, *Methods in Enzymology*, **32**, 615–621.
  19. Sharon, N. and Lis, H. (1972), Lectins: cell-agglutinating and sugar-specific proteins. *Science*, **177**, 949–959.

- 
20. Horejsí, V. and Kocourek, J. (1974), Studies on phytohemagglutinins XVIII. Affinity electrophoresis of phytohemagglutinins, *Biochimica et Biophysica Acta*, **336**, 338–343.
21. Ohta, M., Kawahara, N., Liu, M., Taketa, K., Kudo, T. and Taga, H. (1998), Developmental Alterations of  $\alpha$ -Fetoprotein Sugar Chain in Amniotic Fluids Analyzed by Lectin Affinity Electrophoresis, *Acta Medica Okayama*, **52**, 27–33.
22. 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, *Biochemical Journal*, **183**, 185–188.
23. Sharon, N. and Lis, H. (2004), History of lectins: from hemagglutinins to biological recognition molecules, *Glycobiology*, **14**, 53R–62R.
24. Sauter, N.K., Hanson, J.E., Glick, G.D., Brown, J.H., Crowther, R.L., Park, S.J., Skehel, J.J. and Wiley, D.C. (1992), Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography, *Biochemistry*, **31**, 9609–9621.
25. Bullough, P.A., Hughson, F.M., Skehel, J.J. and Wiley, D.C. (1994), Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature*, **371**, 37–43.
26. Sharma, M., Bhasin, D. and Vohra, H. (2008), Differential induction of immunoregulatory circuits of phagocytic cells by Gal/Gal NAc lectin from pathogenic and nonpathogenic *Entamoeba*, *Journal of Clinical Immunology*, **28**, 542–557.
27. Wimmerova, M., Mitchell, E., Sanchez, J.F., Gautier, C. and Imberty, A. (2003), Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin, *Journal of Biological Chemistry*, **278**, 27059–27067.

- 
28. Sueyoshi, S., Tsuji, T. and Osawa, T. (1985), Purification and characterization of four isolectins of mushroom (*Agaricus bisporus*), *Biological chemistry Hoppe-Seyler*, **366**, 213–221.
29. Kawagishi, H., Mitsunaga, S.I., Yamawaki, M., Ido, M., Shimada, A., Kinoshita, T., Murata, T., Usui, T., Kimura, A. and Chiba, S. (1997), A lectin from mycelia of the fungus *Ganoderma lucidum*, *Phytochemistry*, **44**, 7–10.
30. Sharon, N. and Lis, H. (1990), Legume lectins-a large family of homologous proteins, *FASEB Journal*, **4**, 3198–3208.
31. Elgavish, S. and Shaanan, B. (1997), Lectin-carbohydrate interactions: different folds, common recognition principles, *Trends in Biochemical Sciences*, **22**, 462-467.
32. Sharma, V. and Surolia, A. (1997), Analyses of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity, *Journal of Molecular Biology*, **267**, 433–445.
33. Bundle, D.R. and Young, N.M. (1992), Carbohydrate-protein interactions in antibodies and lectins, *Current Opinion in Structural Biology*, **2**, 666–673.
34. Rini, J.M. (1995), Lectin structure, *Annual Review of Biophysics and Biomolecular Structure*, 1995, **24**, 551.
35. Weis, W.I. and Drickamer, K. (1996), Structural basis of lectin-carbohydrate recognition, *Annual Review of Biochemistry*, **65**, 441–473.
36. Gabrielsen, M., Abdul-Rahman, P.S., Isaacs, N.W., Hashim, O.H. and Cogdell, R.J. (2010), Crystallization and initial X-ray diffraction analysis of a mannose-binding lectin from champedak, *Acta Crystallographica Section F: Structural Biology and crystallization Communications*, **66**, 592–594.

37. Abbott, K.L., Nairn, A.V., Hall, E.M., Horton, M.B., McDonald, J.F., Moremen, K.W., Dinulescu, D.M. and Pierce, M. (2008), Focused glycomic analysis of the *N*-linked glycan biosynthetic pathway in ovarian cancer, *Proteomics* **8**, 3210–3220.
38. Pilobello, K.T. and Maha, L.K. (2007), Deciphering the glycode: the complexity and analytical challenge of glycomics, *Current Opinion in Chemical Biology*, **11**, 300–305.
39. Wigglesworth-Cooksey, B. and Cooksey, K.E. (2005), Use of Fluorophore-Conjugated Lectins To Study Cell-Cell Interactions in Model Marine Biofilms, *Applied and Environmental Microbiology*, **71**, 428.
40. Choi, S.H., Lyu, S.Y. and Park, W.B. (2004), Mistletoe lectin induces apoptosis and telomerase inhibition in human A253 cancer cells through dephosphorylation of Akt, *Archives of Pharmacal Research*, **27**, 68–76.
41. Szilágyi, A., Kardos, J., Osváth, S., Barna, L. and Za'vodszy, P. (2007), *Protein Folding*, Springer-Verlag, Berlin.
42. Dobson, C.M. (2003), Protein folding and misfolding, *Nature*, **426**, 18–25.
43. Dill, K.A. and MacCallum, J.L. (2012), The Protein-Folding Problem, 50 Years on, *Science*, **338**, 1042–1046.
44. Dill, K.A. (1990), Dominant forces in protein folding, *Biochemistry*, **29**, 7133–7155.
45. Dill, K.A., Ozkan, S.B., Shell, M.S. and Weikl, T.R. (2008), The Protein Folding Problem, *Annual Review of Biophysics*, **37**, 289–316.
46. Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995), SCOP: a structural classification of proteins database for the investigation of sequences and structures, *Journal of Molecular Biology*, **247**, 536–540.

- 
47. Lifson, S. and Roig, A. (1961), On the Theory of Helix—Coil Transition in Polypeptides, *Journal of Chemical Physics*, **34**, 1963–1974 .
48. Zimm, B.H. and Bragg, J.K. (1959), Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chain, *Journal of Chemical Physics*, **31**, 526-525.
49. Baldwin, R.L. (1989), How does protein folding get started? *Trends in Biochemical Sciences*, **14**, 291–294.
50. Ptitsyn, O.B. (1991), How does protein synthesis give rise to the 3-D structure? *FEBS Letters*, **285**, 176–181.
51. Karplus, M. and Weaver, D. L. (1976), Protein-folding dynamics. *Nature*, **260**, 404–406.
52. Karplus, M. and Weaver, D. L. (1976), Protein folding dynamics: the diffusion-collision model and experimental data, *Protein Science*, **3**, 650–668.
53. Dill, K.A. (1985), Theory for the Folding and Stability of Globular Proteins. *Biochemistry*, **24**, 1501–1509.
54. Levitt, M. and Warshel, A. (1975), Computer simulation of protein folding, *Nature*, **253**, 694–698.
55. Dill, K.A., Fiebig, K.M. and Chan, H.S. (1993), Cooperativity in protein-folding kinetics, *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 1942–1946.
56. Wolynes, P.G., Onuchic, J. N. and Thirumalai, D. (1995), Navigating the folding routes, *Science*, **267**, 1619–1620
57. Radford, S.E (2000), Protein folding: progress made and promises ahead, *Trends in Biochemical Sciences*, **25**, 611–618.

- 
58. Nymeyer, H., Garcia, A. E. and Onuchic, J. N. (1998), Folding funnels and frustration in off-lattice minimalist protein landscapes, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 5921–5928.
59. Dou, X.H. and Wang, J.H. (2008), Folding free energy landscape of the decapeptide *chignolin*, *Modern Physics Letters B*, **22**, 3087–3098.
60. Andrews, B.T., Gosavi, S., Finke, J.M., Onuchic, J.N. and Jennings, P. A. (2008), The dual-basin landscape in GFP folding, *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 12283–12288.
61. Bai, Y.W. (2006), Energy barriers, cooperativity, and hidden intermediates in the folding of small proteins, *Biochemical and Biophysical Research Communications*, **340**, 976.
62. Ozkan, S.B., Bahar, I. and Dill, K. A. (2001), Transition states and the meaning of F-values in protein folding kinetics, *Nature Structural and Molecular Biology*, **8**, 765–769.
63. Mittal, J. and Best, R. B. (2008), Thermodynamics and kinetics of protein folding under confinement, *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 20233–20238.
64. Altschuler, G. M. and Willison, K. R. and Soc, J. R. (2008), Development of free-energy-based models for chaperonin containing TCP-1 mediated folding of actin, *Interface*, **5**, 1391–1408.
65. Kent, A., Jha, A.K., Fitzgerald, J.E. and Freed, K.F. (2008), Benchmarking implicit solvent folding simulations of the amyloid beta(10-35) fragment, *J. Phys. Chem. B.*, **112**, 6175–6186.
66. Cobb, N. J., Apetri, A.C. and Surewicz, W.K. (2008), Prion Protein Amyloid Formation under Native-like Conditions Involves Refolding of the C-terminal  $\alpha$ -Helical Domain, *J. Biol. Chem.*, **283**, 34704–34711.



- 
67. Lu, D.N., Liu, Z. and Wu, J.Z. (2007), Molecular dynamics for surfactant-assisted protein refolding, *J. Chem. Phys.*, **126**, 064906 (1–13).
68. Bryngelson, J.D. and Wolynes, P.G. (1987), Spin glasses and the statistical mechanics of protein folding, *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 7524–7528.
69. Eaton, W.A., Muñoz, V., Hagen, S.J., Jas G.S., Lapidus, L.J., Henry E.R. and Hofrichter, J. (2000), Fast kinetics and mechanisms in protein folding, *Annual Review of Biophysics and Biomolecular Structure*, **29**, 327–359.
70. Roder, H. and Shastry, M.C.R. (1999), Methods for exploring early events in protein folding, *Current Opinion in Structural Biology*, **9**, 620–626.
71. Goldbeck R.A., Kim-Shapiro, D.B. and Kliger, D.S. (1997), Fast natural and magnetic circular dichroism spectroscopy, *Annual Review of Physical Chemistry*, **48**, 453–479.
72. Kiyama, S., Takahashi, S., Ishimori, K. and Morishima, I. (2000), Stepwise formation of  $\alpha$ -helices during cytochrome *c* folding, *Nature Structural and Molecular Biology*, **7**, 514–520.
73. Fersht, A. (1998), In: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, pp. 540–614, W.H. Freeman.
74. Pollack, L., Tate, M.W., Darnton, N.C., Knight, J.B., Gruner, S.M., Eaton, W.A. and Austin, R.H. (1999), Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering, *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10115–10117.
75. Troullier, A., Reinstädler, D., Dupont, Y., Naumann, D. and Forge, Vet. (2000), Transient non-native secondary structures during the refolding of alpha-lactalbumin detected by infrared spectroscopy, *Nature Structural and Molecular Biology*, **7**, 78–86.

- 
76. Dobson, C.M. and Hore, P.J. (1998), Kinetic studies of protein folding using NMR spectroscopy, *Nature Structural and Molecular Biology*, **5**(Suppl), 504–507.
77. Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L. and Oas, T.G. (1997) The energy landscape of a fast-folding protein mapped by Ala-Gly substitutions, *Nature Structural and Molecular Biology*, **4**, 305–310.
78. Englander, S.W. (2000), Protein folding intermediates and pathways studies by hydrogen exchange, *Annual Review of Biophysics and Biomolecular Structure*, **29**, 213–238.
79. Miranker, A., Robinson, C.V., Radford, S.E. and Dobson, C.M. (1996), Investigation of protein folding by mass spectrometry, *FASEB Journal*, **10**, 93–101.
80. Fisher, T.E., Oberhauser, A.F., Carrion-Vazquez, M., Marszalek, P.E. and Fernandez, J.M. (1999), The study of protein mechanics with the atomic force microscope, *Trends in Biochemical Sciences*, **24**, 379–384.
81. Vendruscolo, M., Zurdo, J., MacPhee, C.E. and Dobson, C.M. (2003), Protein folding and misfolding: A paradigm of self-assembly and regulation in complex biological systems, *Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences*, **361**, 1205–1222.
82. Frydman, J. (2001), Folding of newly translated proteins in vivo: The role of molecular chaperones, *Annual Review of Biochemistry*, **70**, 603–647.
83. Luheshi, L.M. and Dobson, C.M. (2009), Bridging the gap: from protein misfolding to protein mis-folding diseases, *FEBS Letters*, **583**, 2581–2586.
84. Monsellier, E. and Chiti, F. (2007), Prevention of amyloid-like aggregation as a driving force of protein evolution, *EMBO Reports*, **8**, 737–742.

- 
85. Pechmann, S., Levy, E.D., Tartaglia, G.G., Vendruscolo, M. (2009), Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins, *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 10159–10164.
86. Castillo, V. and Ventura, S. (2009), Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases, *PLoS Computational Biology*, **5**, e1000476.
87. Pastore, A. and Temussi, P.A. (2012), The two faces of Janus: functional interactions and protein aggregation, *Current Opinion in Structural Biology*, **22**, 30–37.
88. Selkoe, D.J. (2003), Folding proteins in fatal ways, *Nature*, **426**, 900–904.
89. Wolfe, K.J., Cyr D.M. (2011), Amyloid in neurodegenerative diseases: friend or foe? *Seminars in Cell & Developmental Biology*, **22**, 476–481.
90. Stefani, M. (2010), Protein aggregation diseases: toxicity of soluble prefibrillar aggregates and their clinical significance, *Methods in Molecular Biology*, **648**, 25–41.
91. Morell, M., Bravo, R., Espargaro, A., Sisquella, X., Aviles, F.X., Fernandez-Busquets, X., Ventura, S. (2008), Inclusion bodies: specificity in their aggregation process and amyloid-like structure, *Biochimica et Biophysica Acta*, **1783**, 1815–1825.
92. Wang, L., Maji, S.K., Sawaya, M.R., Eisenberg, D. and Riek, R. (2008), Bacterial inclusion bodies contain amyloid-like structure, *PLoS Biology*, **6**, e195.
93. Ross, C.A., Pickart, C.M., (2004), The ubiquitin-proteasome pathway in Parkinson's disease and other neurodegenerative diseases, *Trends in Cell Biology*, **14**, 703–711.

- 
94. Mandel, S., Grunblatt, E., Riederer, P., Amariglio, N. and Jacob-Hirsch, J. (2005), Gene expression profiling of sporadic Parkinson's disease substantia nigra pars compacta reveals impairment of ubiquitin-proteasome subunits, SKP1A, aldehyde dehydrogenase, and chaperone HSC-70, *Annals of the New York Academy of Sciences*, **1053**, 356–375.
95. Lee, S. and Tsai, F.T.F. (2005), Molecular chaperones in protein quality control, *Journal of Biochemistry and Molecular Biology*, **38**, 259–265.
96. Chiti, F. and Dobson, C.M. (2006), Protein misfolding, functional amyloid and human disease, *Annual Review of Biochemistry*, **75**, 333–366.
97. DeGrado, W.F., Wasserman, Z.R. and Lear, J.D. (1989), Protein design, a minimalist approach, *Science*, **243**, 622–628.
98. Arinaminpathy, Y., Khurana, E., Engelman, D.M. and Gerstein, M.B. (2009), Computational analysis of membrane proteins: the largest class of drug targets, *Drug Discovery Today*, **14**, 1130–1135.
99. MacCallum, J.L. and Tieleman, D.P. (2011), Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions, *Trends in Biochemical Sciences*, **36**, 653–662.
100. Uversky, V.N. and Dunker, A.K. (2010), Understanding protein non-folding, *Biochimica et Biophysica Acta*, **1804**, 1231.

**CHAPTER: 2**

---

**CLONING, EXPRESSION AND  
PURIFICATION OF A LECTIN FROM  
*CICER ARIETINUM* L.**

---

## SUMMARY

BLAST analysis of the N-terminal amino acid sequence of the wild type lectin from *Cicer arietinum* showed maximum identity (99 %) with the major seed albumin from *Pisum sativum* seeds. Hence, its sequence was used to design primers for the isolation of the lectin gene from *Cicer arietinum* seeds.

Total RNA was isolated from 36 days old *Cicer arietinum* (chick pea) seeds. After enrichment of the mRNA, double stranded cDNA was isolated using primers designed from the nucleotide sequence of the PA2 albumin from *Pisum sativum*. An amplicon of 700 bp was obtained corresponding to the chick pea lectin gene (693 bp). TA cloning was employed for its ligation with the pGEM-T Easy vector. The sequence of the lectin gene was submitted to GenBank (Accession number JX298876). For expression of the lectin gene, the amplicon was sub-cloned into the pET-28a(+) vector within the NcoI and EcoRI sites. The *E. coli* BL21-CodonPlus (DE3)-RIL strain was used for producing the lectin gene after induction with 1 mM IPTG. The crude cell lysates from the induced cultures were used for the purification of the recombinant lectin. The lectin was purified to homogeneity using an ion-exchange (DEAE-Cellulose) and a gel filtration (Superose 12 prep grade, FPLC) column. The lectin was found to be a homodimer (54,600 Da) of subunit molecular mass 27,000 Da. It showed hemagglutination only against pronase-treated rabbit erythrocytes. The yield of the lectin was around 5 to 8 mg/litre of culture broth with a specific activity of  $5 \times 10^3$  hemagglutination units/mg.

## 2.1. INTRODUCTION

Genetic engineering is a versatile technique that facilitates the production of proteins tailor made for an intended purpose. The simplest approach for constructing new proteins has been to modify the protein backbone in a desired fashion of investigation. In fact, this approach may also be favoured by nature because these modifications add many different functional properties on the same polypeptide chain.

Plant lectins represent a well known class of multivalent carbohydrate binding proteins which mediate various biological processes. The rapid development of genetic engineering has boosted plant lectin research in the last two decades by providing access to the complete cDNAs of a large number of these proteins. Cloning and expression of lectins serve the following purposes: (a) to establish their primary structures; (b) to study their genetics, evolution and biosynthesis; (c) to examine the

role of different amino acid residues of the lectins in carbohydrate recognition and subunit assembly; (d) to produce lectins with novel specificities, and (e) to study the function of the lectins in plants.

Several plant lectins have been expressed in heterologous systems, ranging from bacteria (predominantly *Escherichia coli*) and yeasts (often *Saccharomyces cerevisiae*) to plants (e.g. tobacco) and animals (e.g. monkey kidney) [1]. These recombinant plant lectins may have different sugar specificity or affinity to that of native protein. This is mainly because in plant, the lectins undergo co- or post- translational processing reactions, some of which are most unusual. For example, mature pea and *Vicia faba* bean lectins get cleaved into two chains, whereas the primary sequence of Con A is rearranged by domain swapping so that it becomes circularly homologous with that of the other legume lectins. Lectins expressed in bacteria are devoid of their carbohydrate, whereas those expressed in yeast or other cells may carry different glycans [2, 3]. Furthermore, in case of prokaryotic expression system recombinant plant lectins cannot be processed because of the lack of required modification machinery as in case of their native counterparts.

Although, the native protein can be used to delineate some of the properties, understanding at molecular level is greatly hindered due to the inability to modify the backbone of native lectins. The diversity in specificity of lectins might have arisen due to various hidden factors such as multimeric, multichain and partial glycosylation. It is not clear to date as which property i.e. multimeric form/multi chain/post translational modification contributes to the sugar specificity and/or affinity of each lectin. Over the period of last three decades, more than hundred lectins have been produced in recombinant form for several purposes. In the following table, we have presented few latest reports of the recombinant lectins.

Table 2.1 describes the examples of few lectins that have been cloned, expressed (and purified) in bacterial or heterologous hosts using novel techniques. The wild type and recombinant forms of few of these lectins show similar carbohydrate-binding property. For some of them, the recombinant form is more potent than the wild type one.

Table 2.1: Lectins cloned using novel techniques

Sr. No.	Lectin (source) and its reference	Characteristic feature
1.	CRAMOLL 1 from <i>Cratylia mollis</i> seeds [4]	<ul style="list-style-type: none"> <li>• A chemically synthesized DNA encoding mature amino acid sequence of CRAMOLL 1 used for cloning and expression of recombinant CRAMOLL.</li> <li>• Both plant and recombinant lectins identical with respect to sugar specificity and agglutination against rabbit erythrocytes and <i>Trypanosoma cruzi</i> epimastigote cells</li> </ul>
2.	Volkensin from <i>Adenia volkensis</i> [5]	<ul style="list-style-type: none"> <li>• Separate cloning and expression of recombinant volkensin A and B chains.</li> <li>• Heterodimer associated by co-folding strategy of folded monomers in presence of sugar.</li> </ul>
3.	Breadfruit seed ( <i>Artocarpus incise</i> ) lectin [6]	<ul style="list-style-type: none"> <li>• An experimental factorial design employed to maximise the soluble expression.</li> <li>• Recombinant and wild type lectins similar in sugar binding.</li> </ul>
4.	Mannose-Binding Lectin (CaMBL1) from pepper leaves [7]	<ul style="list-style-type: none"> <li>• Over expression in <i>Arabidopsis thaliana</i> confers enhanced resistance to <i>Pseudomonas syringae</i> pv tomato and <i>Alternaria brassicicola</i> infection.</li> <li>• Transient expression induces activation of defence-related genes and induction of cell death in pepper.</li> </ul>
5.	<i>Pinellia</i> bulb lectin [8]	<ul style="list-style-type: none"> <li>• Lectin gene expressed in transgenic tobacco plants to study its insecticidal property.</li> <li>• Insect bioassays demonstrated increased mortality to tobacco aphids (<i>Myzus nicotianae</i>) compared to wild type.</li> </ul>
6.	<i>Oryza sativa</i> lectin [9]	<ul style="list-style-type: none"> <li>• Methyl jasmonate induces expression of lectin, hence may be involved in rice defence.</li> </ul>

### Classification of *Cicer arietinum*



**Kingdom – Plantae – Plants**

**Sub kingdom – Tracheobionta – Vascular plants**

**Superdivision – Spermatophyta – Seed plants**

**Division – Magnoliophyta – Flowering plants**

**Class – Magnoliopsida – Dicotyledons**

**Sub class – Rosidae**

**Order – Fabales**

**Family- Fabaceae – Pea family**

**Genus – Cicer L. - cicer**

**Species – *Cicer arietinum* L. chick pea**

Chick pea (*Cicer arietinum*) is a very versatile legume and forms an important part of the diet in the Indian sub-continent as well as many Middle-Eastern countries. Originated in the Middle East (southeastern Turkey), it is now cultivated in the tropical, sub-tropical and temperate regions. Being world's second most widely grown legume, its cultivation is of particular importance to food security in the developing world where, owing to its capacity for symbiotic nitrogen fixation, chick pea seeds are a primary source of human dietary protein.

Chickpea seeds are very nutritive with high protein content (25-29 %) [10] and hence are considered an important part of the vegetarian diet. Apart from this, they are also a good source of soluble and insoluble fiber, carbohydrates, vitamins and minerals. Soluble fibres 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 amount of fats most of which are polyunsaturated, which is also helpful in preventing cardiovascular diseases.

There is a previous report of the isolation of a lectin from this plant [11]. They report the presence of an albumin from the chick pea seeds with agglutinating activity against papainized erythrocytes. No further studies have been done on this lectin. Pedroche *et al.* [12] have also reported the isolation of the pa2 albumin from chickpea seeds with blood cell agglutination properties. Esteban *et al* [13] have reported the structural features of a chickpea cDNA encoding a putative vegetative lectin and its sequence. Qureshi *et al.* [14] have reported the isolation of a lectin gene from a cDNA library prepared from the seeds of chickpea. This lectin (CpSL) is shown to belong to the mannose-specific family of lectins. But none of these lectins resemble the lectin described in the current study.

Purification and preliminary characterization of a wild type lectin from the seeds of chick pea has already been reported from our group [15]. This lectin was shown to agglutinate pronase-treated erythrocytes alone. The hemagglutination activity was inhibited only by complex sugars, and not by simple mono- or oligosaccharides.

The N- terminal sequence of this lectin was found to be T/G/K - N/V - F/G - Y/G - K/Y - I/V - N/K - A/S - A/T. When a BLAST analysis was carried out using this sequence, it showed 90% identity with the sequence of the major seed albumin (PA2) from *Pisum sativum* (NCBI accession no. P08688).

No conformational studies have been reported for this lectin. Moreover, initial trials of crystallization had failed for this lectin, despite attempting different procedures. Hence, the cloning of this lectin gene was initiated. Primers designed from the sequence of the PA2 albumin from *P. sativum* were utilized to isolate the lectin gene from chick pea seeds.

## 2.2. MATERIALS

**2.2.1 *Cicer arietinum* seeds:** Cultivar BDN 9-3 of chick pea seeds was procured from the Badnapur Agriculture University, Jalna, India.

### 2.2.2. Bacterial Strains:

- (a) *E. coli* JM109: Promega, USA
- (b) *E. coli* BL21 (DE3): Novagen, USA
- (d) *E. coli* BL21 CodonPlus (DE3)-RIL: Stratagene, USA.

### 2.2.3. Plasmids:

- (a) pGEM-T Easy: Promega, USA
- (b) pET-28a(+): Novagen, USA

### 2.2.4. PCR Primers:

All primers were obtained from Geno Mechanix, USA

RF1 - Chick pea forward primer

5' ATGACGAAAACAGGTTACATTAAT G 3'

RF2 - Chick pea forward primer

5' GGTTACATTAATGCTGCTTTTCG 3'

RF3 - Chick pea forward primer

5' GAAGCTTATCTTTTCATCAATG 3'

RR1 - Chick pea reverse primer

5' TTAATTCTCAAGAGGTATTATGCC 3'

RR2 - Chick pea reverse primer.

5' CCAGCAAACCGTCACAGCCTCAC 3'

MIK10 - Chick pea forward primer with the NcoI site

5'-TCTTCGccatggCAAAAACAGGTTACATTAAT-3'

MIK12 - Chick pea reverse primer with the EcoRI site

5'-TATATAgattcTTAATTCTCAAGAGGTATTATGCCACG-3'

T7 Promoter primer

5' TAATACGACTCACTATAGGG 3'

SP6 Promoter primer

5' CATACGATTTAGGTGACACTATAG 3'

### 2.2.5. Antibiotics:

(a) Ampicillin: Sigma, USA (25 µg/ml in liquid broth and solid agar for selection)

(b) Kanamycin: Sigma, USA (30 µg/ml in liquid broth and solid agar for selection)

(c) Chloramphenicol: Sigma, USA (34 µg/ml in liquid broth and solid agar for selection)

### 2.2.6. Media (for bacterial growth):

(a) Luria Bertani (LB) Medium: Bactotryptone, yeast extract and agar were purchased from HiMedia, India.

Media was prepared according to protocols described in Molecular Cloning, A laboratory manual [16].

### 2.2.7. Restriction enzymes and other DNA Modifying enzymes:

All the restriction enzymes and DNA modifying enzymes were obtained from New England Biolabs, USA. Restriction enzymes included NcoI, EcoRI, and HindIII. DNA modifying enzymes were Pfu DNA polymerase, Taq DNA polymerase, and T4 DNA Ligase (MBI Fermentas, USA). The respective 10 X buffers were supplied by the manufacturers.

### 2.2.8. Reagents for SDS-PAGE and protein estimation:

Tris-(hydroxymethyl) aminomethane, SDS and Brilliant Blue-R250 were purchased from Sigma, USA. Glycine was purchased from Qualigens, India. Folin reagent for protein estimation was purchased from Bio-Rad.

### 2.2.9. Reagents for protein purification:

(j) Anion Exchange Matrices: DEAE-cellulose (Sigma, USA).

(ii) FPLC Gel filtration matrix: Superose 12 prep grade (Pharmacia Biotechnology).

All other chemicals and reagents used were of analytical grade and obtained from Sigma, USA or Merck, Germany.

**2.2.10. Blood for hemagglutination assay:** Rabbit blood was obtained from the animal house of the National Institute of Virology, Pune, India

## 2.3. METHODS

### [A] Cloning of the chick pea lectin gene

#### 2.3.1. Total RNA Extraction from chick pea seeds

Total RNA was extracted from chickpea seeds (36 days old) as follows:

##### Solutions and Reagents:

- Extraction buffer (EB): The extraction buffer consisted of 2 % CTAB, 2 % PVPP, 100 mM Tris-HCl (pH – 7.4), 2 M NaCl, 25 mM EDTA and 4 % (v/v)  $\beta$ -mercaptoethanol (added just before use).
- Tris – saturated phenol (pH – 7.4)
- Phenol: chloroform: isoamyl alcohol (25:24:1, v/v/v) and Chloroform : Isoamyl alcohol (24:1, v/v)
- 3 M Sodium acetate (pH 5.2 and pH 5.6)
- Ethanol (absolute and 70 %, v/v)
- 8 M Lithium chloride (LiCl)

All the solutions were treated with 0.1 % (v/v) DEPC; Tris buffers were also prepared in DEPC-treated water.

RNA extraction was carried out according to standard protocol [19]. Briefly, one gram of tissue was ground to fine powder in 10 ml of pre-warmed EB buffer. After incubation at 65 °C for 10 min, the mixture was extracted with equal volumes of phenol: chloroform: isoamyl alcohol (25:24:1, v/v/v) and chloroform: isoamyl alcohol (24:1, v/v). The RNA was precipitated from the aqueous phase, washed with 3 M sodium acetate (pH 5.2 and 5.6), 70% ethanol and 8 M LiCl. It was finally dissolved in 500  $\mu$ l of DEPC-treated Milli Q.

#### 2.3.2. cDNA synthesis

The total RNA isolated from the seeds was used for the preparation of mRNA using the Oligotex mRNA synthesis kit from QIAGEN [17]. The enriched mRNA was then used for the synthesis of the 1<sup>st</sup> strand of cDNA according to Promega's Reverse Transcription System (Promega, USA) [18]. The cDNA 1<sup>st</sup> strand obtained was used as a template for the synthesis of the second strand. Three different forward (RF1, RF2

and RF3) and two reverse primers (RR1 and RR2) were designed from the nucleotide sequence of the major seed albumin PA2 from *Pisum sativum*.

The Polymerase Chain Reaction (PCR) mix consisted of the cDNA template, 7 picomoles each of the forward and reverse primers, 0.2 mM each of the four dNTPs and Taq DNA polymerase (3 U/ $\mu$ l), in addition to Taq buffer and Milli Q water. Six different PCR reactions were set up with these five primers. The reaction mixture was subjected to initial denaturation (95 °C, 5 min), 35 three-step cycles of denaturation (95 °C, 1 min), annealing (51 °C, 30 sec) and extension (72 °C, 1 min) and final extension (72 °C, 6 min). Amplified product (5  $\mu$ l) was detected on 1 % agarose gel for a product of 650 bp.

### 2.3.3. Cloning into pGEM-T Easy vector

#### 2.3.3.1. A-tailing of the amplicon

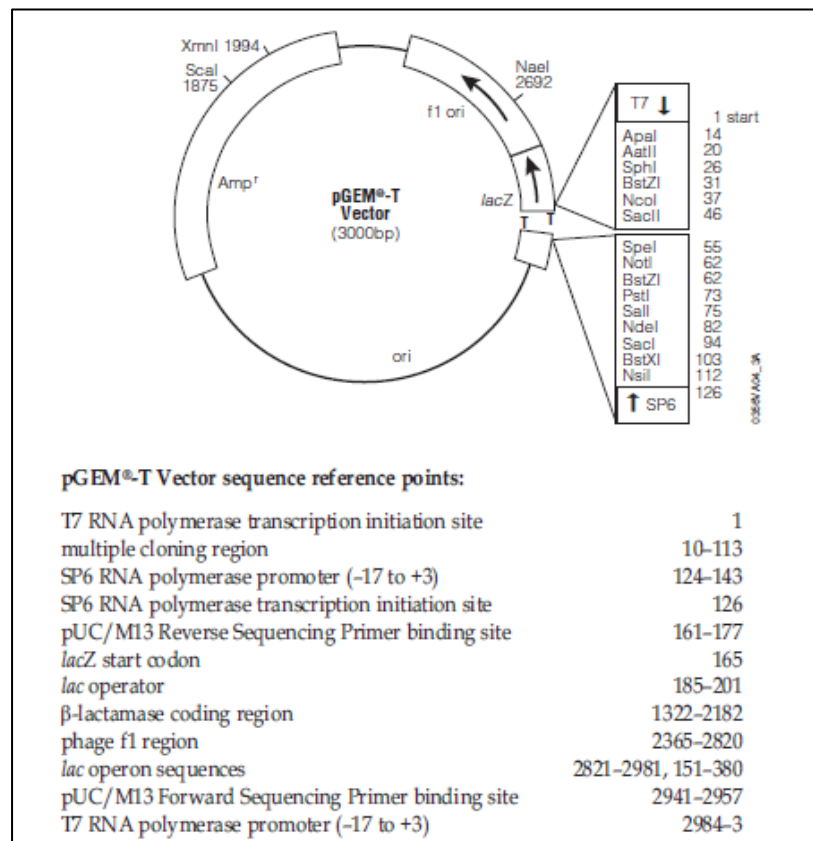


Figure 2.1: pGEM-T Easy vector map and sequence reference points.

Addition of a 3' terminal 'A' overhang onto the amplicons generated by the Taq polymerase facilitates their PCR-based cloning into the pGEM-T Easy vector (Promega T-vector series). The amplicon obtained after PCR amplification was first eluted from

the gel using the QIAGEN gel elution kit. The protocol followed was as mentioned in the Promega manual [19].

### 2.3.3.2 Ligation with pGEM-T Easy vector

The A-tailed fragments were ligated with Promega's pGEM-T Easy vector using T4 DNA ligase according to standard procedure. The ligation reaction was incubated overnight at 4 °C.

### 2.3.3.3. Competent cell preparation and transformation

Competent *E. coli* JM109 cells were prepared essentially according to the Hanahan method [16]. Briefly, 10 µl of JM109 competent *E. coli* cells were added into 5 ml of Luria Bertani (LB) broth containing 100 µl of 0.5 M MgSO<sub>4</sub> and 50 µl of 1 M MgCl<sub>2</sub>. This culture was grown overnight at 200 rpm and 37 °C. 0.5 ml of the overnight grown culture (pre-culture) was added into 100 ml LB broth containing 1 ml of 0.5M MgSO<sub>4</sub> and 0.5 ml of 1 M MgCl<sub>2</sub>. It was allowed to grow at 37 °C and 200 rpm till the OD at 600 nm reached less than 0.5 (0.3 to 0.5). The flask was kept on ice for 30 min. The culture was centrifuged at 4000 rpm for 10 min at 4 °C. 5 ml of ice-cold Tfb I buffer solution (10 mM potassium acetate, 100 mM potassium chloride, 10 mM calcium chloride, 45 mM manganese chloride and 10% glycerol) was added slowly to dissolve the pellet completely. After incubation on ice for one hour, it was centrifuged at 3000 rpm for 10min at 4 °C. The aliquots of cell suspension were stored at -80 °C till further use.

### 2.3.3.4. Transformation of competent *E. coli* cells with plasmid

Competent cells were mixed with plasmid DNA (1 to 10 ng) and incubated on ice for 30 minutes. The cells were subjected to heat shock at 42 °C for 90 seconds and then allowed to recover in 1 ml of LB broth at 37 °C for 1 hour with constant shaking. The transformed cells were plated on appropriate selection plate (LB agar containing Ampicillin) and incubated for 12-16 hours at 37 °C.

### 2.3.3.5. Plasmid isolation from bacterial cells

A single colony from the overnight grown LB-Amp plates was inoculated in 10 ml LB broth containing 10 µl Ampicillin (stock 25 mg/ml) and the culture was allowed to grow overnight under shaking conditions at 200 rpm and 37 °C.

Plasmid was prepared from bacteria according to the alkaline SDS lysis method [16, 20]. Briefly, log phase bacterial cultures were centrifuged and the pellet was resuspended in chilled miniprep lysis buffer followed by incubation for 5 min at RT. After addition of double volume of alkaline SDS solution, the cells were further

incubated on ice of not more than 5 min. The lysis buffer was neutralized by addition of half volume 5 M potassium acetate and thorough vortexing. The plasmid was then extracted twice into aqueous phase using phenol / chloroform / isoamyl alcohol (25:24:1, v/v/v) and then once in chloroform / isoamyl alcohol (24:1, v/v). Plasmid was precipitated using absolute ethanol at -20 °C for 30 minutes. Plasmid pellet was recovered by high speed centrifugation at 14,000 rpm for 30 min and washed in 70 % chilled ethanol. The plasmid pellet was air dried and dissolved in 200 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). After treating the plasmid with DNase free RNase (2 µl of 10 mg/ml stock) for 30 min at 37 °C, the plasmid was again extracted by phenol-chloroform-isoamyl alcohol method and precipitation with 0.5 volume of ammonium acetate and two volumes of ethanol. Plasmid DNA pellet obtained after centrifugation was washed with 70 % ethanol, air dried and dissolved in Milli Q water.

The plasmid was quantitatively checked by determining the ratio of absorbances at 260 and 280. The yield of plasmid was calculated using the formula  $1 \text{ (O.D.) } A_{260} = 50 \text{ } \mu\text{g/ ml DNA}$ . The plasmid DNA was visualized on a 0.8 % agarose gel run at 70 V/min. The plasmid was sequenced at the DNA Sequencing Facility of the National Centre for Cell Science, Pune, India.

#### **2.3.3.6. Sub-cloning into expression vector pET-28a(+)**

To achieve ligation with an expression vector, it was essential that compatible restriction sites be introduced in the plasmid as well as in the expression vector. Hence, the plasmid was amplified using a new set of primers namely MIK10 and MIK12, which contain sites for NcoI and EcoRI at their 5' and 3' ends respectively. The PCR amplification was set up as mentioned earlier, with the exception that the annealing temperature was set at 58 °C and the extension was carried out for 1.5 min.

For sub-cloning, the pET-28a(+) vector as well as the insert (amplicon obtained with primers MIK10 and MIK12) were double digested with the restriction enzymes NcoI and EcoRI. The vector was digested either step-wise with EcoRI and NcoI or simultaneously with these two enzymes. The insert underwent simultaneous double-digestion with these two enzymes. The reaction mixture was incubated at 37 °C for one hour and then visualized on a 0.8 % low melting point agarose gel. The bands were excised from the gel and purified using the QIAgen gel purification kit.





subjected to PCR with primers MIK 10 and MIK 12. To confirm the validity of the amplicon, it was sequenced and also digested with the enzyme HindIII. The plasmid DNA was then transformed into competent BL21-CodonPlus (DE3)-RIL cells.

### **[B] Expression of the chick pea lectin gene**

The BL21-CodonPlus (DE3)-RIL cells were used for protein expression [21]. These cells enable efficient high level expression of heterologous proteins in *E. coli*.

Competent BL21-CodonPlus (DE3)-RIL cells were prepared by the Hanahan method as mentioned before. The only difference was in the selection of the antibiotic resistance marker. Here, Kanamycin was used since the pET-28a(+)vector carries a Kanamycin resistance gene. Chloramphenicol was the selection marker for the CodonPlus cells. The transformation mixture obtained above was plated on LB agar plates containing 30 µg/µl Kanamycin and 34 µg/µl Chloramphenicol and incubated overnight at 37 °C.

#### **2.3.3.7. Induction of expression**

A single colony from the transformed plate was inoculated in 10 ml LB broth containing 30 µg/µl of Kanamycin and 34 µg/µl of Chloramphenicol and allowed to grow overnight at 200 rpm and 37 °C. 1 % of the overnight inoculum was added into 100 ml LB broth containing 30 µg/µl of Kanamycin and 34 µg/µl of Chloramphenicol and allowed to grow till the absorbance at 600 nm reached approximately 0.3 – 0.35 (< 0.4). An aliquot of uninduced culture was removed at this point. To the remaining culture, IPTG (from a 1 M stock) was added to a final concentration of 1 mM. Both the uninduced and induced cultures were grown for a further period of 4 to 6 hours for induction of the lectin gene. The culture was then centrifuged at 4000 rpm for 10 min at 4 °C. The supernatant was discarded and the pellet was stored at -20 °C.

#### **2.3.3.8. Preparation of crude cell lysates**

The pellets from uninduced and induced cultures were dissolved in 50 mM Tris-HCl, pH 8.0 (containing 150 mM NaCl) and kept on ice. The solution was then sonicated using a probe ultrasonicator for 2 minutes with a 10 sec 'ON' and 10 sec 'OFF' cycle and amplitude of 50 %. The sonicated solution was then centrifuged at 10,000 rpm for 20 min at 4 °C. The pellet and supernatant samples were separately stored at -20 °C.

### 2.3.3.9. SDS Poly Acrylamide Gel Electrophoresis:

Electrophoresis was essentially carried out according to Laemmli's protocol with minor modifications [22]. Samples were prepared using 5 X Laemmli sample buffer and boiling for 10 minutes. Proteins were resolved on 12% resolving gel and 5 % stacking gel at a constant voltage of 100 V.

Coomassie Brilliant Blue staining and destaining: The gels were stained in 0.2% Coomassie Brilliant Blue R250 solution with constant shaking for 2 hours followed by destaining in methanol/ water/ acetic acid (2:7:1, v/v/v) until the background became clear and distinct bands appeared on the gel.

### 2.3.3.10. Haemagglutination assay

The haemagglutination assay was performed to confirm the activity of the proteins (induced by IPTG) in the crude cell lysate. Haemagglutination tests were performed in standard U- shaped micro titre well plates by the two-fold serial dilution method. Rabbit erythrocytes were washed thoroughly with PBS (phosphate buffer 20 mM, pH 7.2, NaCl 150 mM) and a 3 % suspension of these erythrocytes was prepared in PBS. Pronase enzyme (10 mg/ml) was added to the 3 % suspension and incubated at 37 °C for one hour. The erythrocytes were further washed thrice with PBS and used for the assay. A 50 µl aliquot of the erythrocyte suspension was mixed with 50 µl of serially diluted lectin. Agglutination was examined visually after incubation for one hour at room temperature.

## [C] Purification of the Recombinant lectin

The crude cell lysates showing activity were dialysed extensively against 20 mM Phosphate buffer, pH 7.2 and loaded on a column of DEAE-cellulose.

### 2.3.3.11. DEAE-cellulose Chromatography:

The lectin was partially purified by anion exchange chromatography on DEAE-cellulose. The above extract (approximately 50 mg) was loaded on a DEAE-cellulose column (packed volume of 30 ml) pre-equilibrated with 20 mM phosphate buffer, pH 7.2 at a constant flow rate of 10 ml/hour. The column was washed with the same buffer and the washings were collected as fractions of 3 ml until  $A_{280}$  fell below 0.05; they were then checked for hemagglutination activity.

Fractions of the column showing hemagglutination activity were pooled and concentrated before loading on an FPLC gel filtration column.

### 2.3.3.12. Gel filtration on FPLC using a Superose 12 prep grade column

Superose prep grade is a cross-linked, agarose-based medium optimized for high performance gel filtration for the preparative purification of biomolecules. Superose 12 prep grade medium is used for the optimal separation of proteins from 1,000 to  $3 \times 10^5$  Da in the pH range of 3 to 12.

500  $\mu$ l of the sample (in 20 mM phosphate buffer, pH 7.2, 150 mM NaCl) was loaded on the column pre-equilibrated with 20 mM phosphate buffer, pH 7.2 (containing 150 mM NaCl). The column had a bed volume of 30 ml and 30 fractions of 1 ml each were collected at a flow rate of 10 ml/hour. The fractions were checked for their hemagglutination activity against pronase-treated rabbit erythrocytes; those showing activity were loaded on a 12 % SDS PAGE to check their purity.

### 2.3.4. Molecular mass determination by HPLC

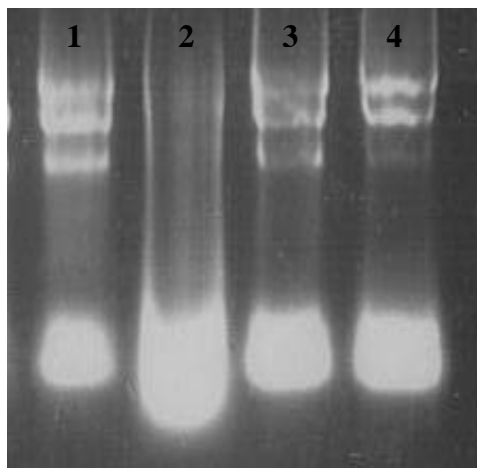
Molecular mass of rCAL was determined using WATERS HPLC unit and WATERS gel filtration Protein Pak™ 300SW (7.5 x 300 mm column). 25 mM phosphate buffer, pH 7.2 (with 150 mM NaCl) served as the mobile phase. 100  $\mu$ l protein sample (600  $\mu$ g/ml) was injected into the column. Flow rate was maintained at 0.5 ml/min and the elution profile was monitored at 280 nm. Standard proteins used were (a)  $\beta$ -amylase (200,000 Da), (b) bovine serum albumin (66,000 Da), (c) ovalbumin (43,000 Da) and (d) soybean trypsin inhibitor (22,000 Da). Molecular mass of rCAL was calculated from the plot of  $V_e/V_o$  vs log [molecular mass] of the standard proteins.

## 2.4. RESULTS AND DISCUSSION

### [A] Cloning of the chickpea lectin gene

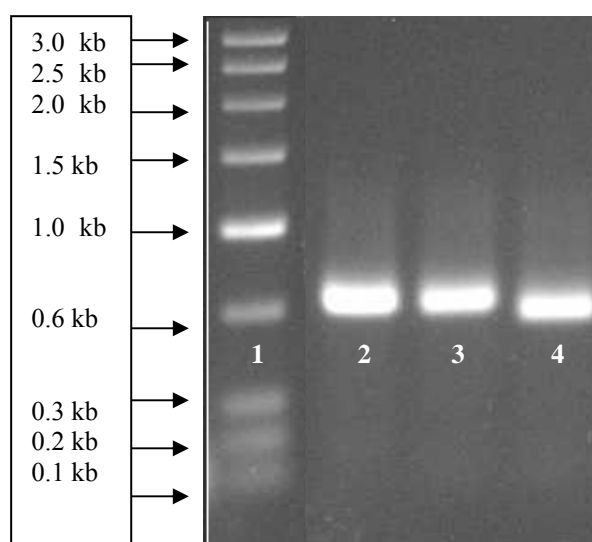
#### 2.4.1. Total RNA Extraction:

Three out of four RNA preparations yielded good quality, intact RNA without any genomic DNA contamination (Figure 2.3). These three preparations were pooled and used for mRNA purification using the QIAGEN Oligotex mRNA synthesis system.



**Figure 2.3:** Agarose gel electrophoresis of total RNA isolated from the chick pea seeds (lanes 1, 2, 3 and 4).

Good quality mRNA and first strand of cDNA could be prepared from the Oligotex system and the reverse transcriptase system respectively. Three amplicons obtained after PCR with the six primer-combinations (using first strand as the template) yielded a single band in the size range of 600-650 bp after the agarose gel electrophoresis (Figure 2.4).



**Figure 2.4:** Gel electrophoresis of the amplified products (cDNA). Lane 1: Low range DNA Ruler, Lane 2: sample 1; Lane 3: sample 2; and Lane 4: sample 3.

Thus, a blunt-ended amplicon (generated due to the action of Taq polymerase enzyme) representing the chick pea lectin gene was obtained.

### 2.4.2. Cloning into pGEM-T easy vector:

A-tailing ensured the ligation of the lectin gene with the pGEM-T easy vector. The ligation product was subjected to PCR amplification using the primers RF1 and RF2 as well as the universal forward and reverse pUC/ M13 primers. The universal primers amplify the insert region in the vector and hence the sequence of the insert can be obtained. This also confirmed the correct direction of insertion of the amplicon into the vector.

### 2.4.3. Sub-cloning into expression vector pET-28a(+)

The sequence of the plasmid DNA isolated from the ligation product (pGEM-T Easy + inserted CAL gene) is as follows:

5'

ATGACGAAAACAGGTTACATTAATGCTGCATTTTCGTTTCATCTAGGAACAAT  
 GAAGCTTACTTATTCATCAATGATAAGTACGTGCTATTGGATTATGCTCCAG  
 GAACTAGCAATGATAAAGTTTTGTATGGGCCAAGTCTTGTTCGTGATGGAT  
 ATAAATCACTTGCTAAGACAATATTTGGAACATATGGAATAGATTGTTTCCTT  
 CGATACTGAATACAACGAGGCATTCATATTTTATGAAAACCTTTGTGCTCGC  
 ATAGACTATGCTCCACATTCCGACAAAGACAAAATAATTCAGGTCCTAAG  
 AAAATCGCTGACATGTTTCCTTTTTTCAAAGGAACCGTGTTTGAAAATGGA  
 ATAGATGCTGCATTTAGATCAACTAAGGGAAAAGAAGTTTACTTATTCAA  
 GAAGACAAGTATGCTCGTATAGATTATGGCACAAATCGTCTCGTTCAAAT  
 ATCAAGTACATTAGTGATGGGTTTCCTTGCTTACGTGGCACAATATTTGAAT  
 ATGGAATGGATTCAGCTTTTGTCTCATAAGACAAATGAAGCTTATCTTTT  
 CAAAGGAGAATATTATGCACGTATCAATTTTACACCTGGATCAACAAATGA  
 TACCATTATGGGTGGTGTGAAGAAACTCTTGACTATTGGCCATCTCTTCGT  
 GGCATAATTCCTCTTGAGAATTAA 3'

The sequence has been deposited in the NCBI (GenBank) database under the accession number of JX298876.

BLAST analysis of this sequence showed maximum identity with the following organisms.

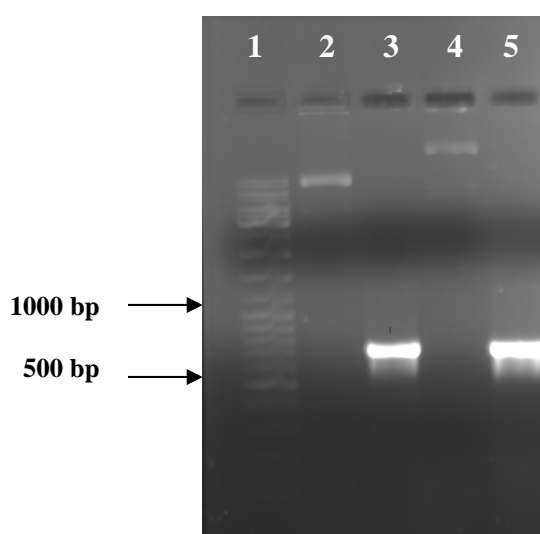
1. [gb|M17147.1|PEAALBMAJ](#) *P.sativum* (pea) major seed albumin mRNA, complete cds. clone pPS15-21 (100 % identity) (Query coverage 99%)
2. Pea (*P.sativum*) albumin 2 (PA2) mRNA, complete cds (83 % identity) (Accession M13791.1)

The DNA sequence of the gene was fed in the ExPASy Translate tool software and the following amino acid sequence was obtained in the 5'3' reading frame without the presence of any stop codon.

5'3' Frame 1

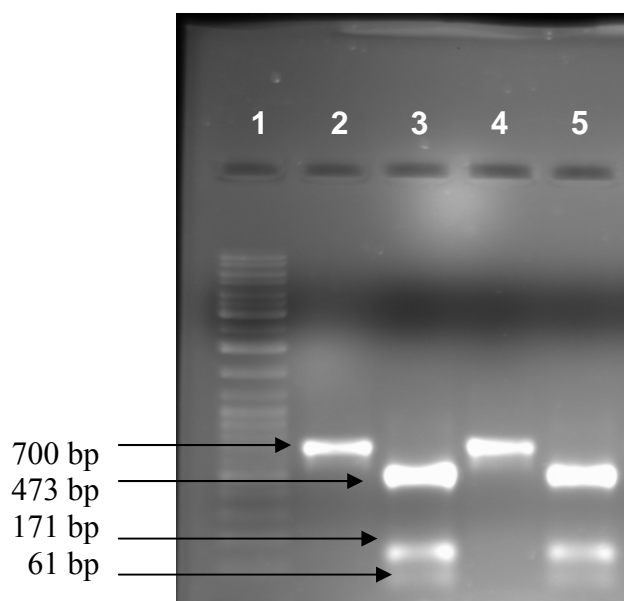
**Met** T K T G Y I N A A F R S S R N N E A Y L F I N D K Y V L L D Y A P G T S N  
 D K V L Y G P S L V R D G Y K S L A K T I F G T Y G I D C S F D T E Y N E A F  
 I F Y E N L C A R I D Y A P H S D K D K I I S G P K K I A D **Met** F P F F K G T V  
 F E N G I D A A F R S T K G K E V Y L F K E D K Y A R I D Y G T N R L V Q N I  
 K Y I S D G F P C L R G T I F E Y G **Met** D S A F A S H K T N E A Y L F K G E Y  
 Y A R I N F T P G S T N D T I **Met** G G V K K T L D Y W P S L R G I I P L E N  
**Stop**

Amplification of this gene with primers MIK10 and MIK12 yielded a 700 bp amplicon.



**Figure 2.5: 0.8 % agarose gel showing the plasmids (pGEM-T Easy and pET-28a(+)) and their amplicons obtained using primers MIK 10 and MIK 12.** Lane 1: O' GeneRuler DNA Ladder Mix from Fermentas; Lane 2: plasmid (pGEM-T vector with CAL gene); Lane 3: amplicon of sample in lane 2 with primers MIK 10 and MIK 12 (700 bp); Lane 4: plasmid (pET-28a(+)) vector with CAL gene); and Lane 5: amplicon of sample in lane 4 with primers MIK 10 and MIK 12 (700 bp).

Digestion of this amplicon with HindIII enzyme resulted in the formation of three bands - those corresponding to 61 bp, 171 bp and 473 bp. This enzyme has three restriction sites in the chick pea lectin gene – at positions 61, 232 and 705, hence the bands obtained corresponded to the sites of the enzyme in the lectin gene. This confirmed that the insert was indeed the chick pea lectin gene (Figure 2.6).



**Figure 2.6: Agarose gel showing the amplicons before and after Hind III digestion.**

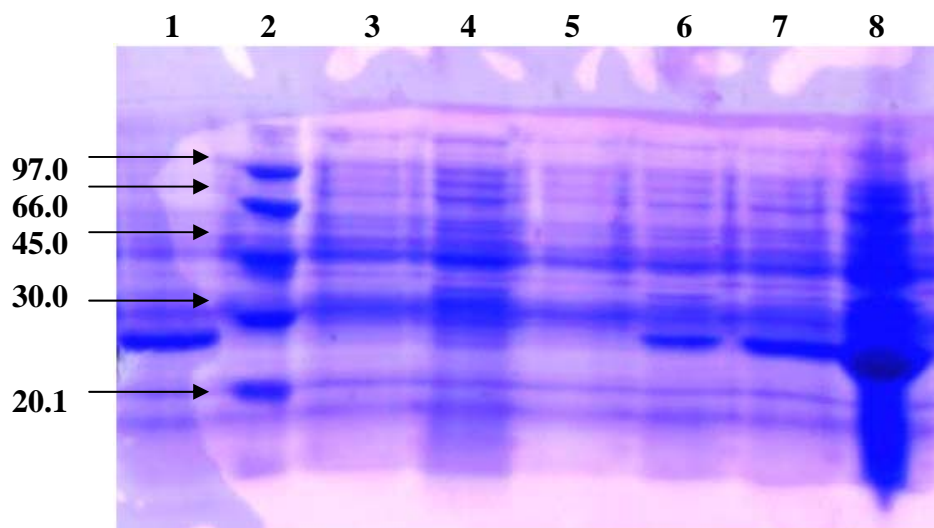
Lane 1: O' GeneRuler DNA Ladder; Lane 2: undigested amplicon 1 (pGEM-T with CAL gene) with T7 and SP6 promoter primers; Lane 3: HindIII digested amplicon 1 (showing bands of 473 bp, 171 bp and 61 bp); Lane 4: undigested amplicon 2 (pET-28a(+)) with CAL gene) with primers MIK10 and MIK12; Lane 5: HindIII digested amplicon 2 (showing bands of 473 bp, 171 bp and 61 bp).

The insert thus had been sub-cloned into the expression vector pET-28a(+).

### [B] Expression of the chick pea lectin gene

During heterologous expression of eukaryotic genes in *E. coli*, there is a possibility that lack of certain tRNAs may limit the translation of the foreign protein. Supplementation of the expression plasmid with such limiting tRNAs can prevent the formation of truncated proteins. BL21-CodonPlus strains are engineered to contain extra copies of genes that encode the tRNAs that most frequently limit translation of heterologous proteins in *E. coli*. Availability of such tRNAs allows high-level expression of heterologous recombinant genes in BL21-CodonPlus cells that are poorly expressed in conventional BL21 strains. It is a lysogenic strain containing the T7 RNA polymerase gene under the control of the inducible *lacUV5* promoter along with a ColE1-compatible, pACYC-based plasmid containing extra copies of the *argU*, *ileY*, and *leuW* tRNA genes. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively.

Uninduced and induced cultures of pET vector (lanes 3 and 4 in Figure 2.7) were loaded to rule out the possibility of the induced band being of that of the enzyme Chloramphenicol acetyltransferase that has a molecular mass of 25,000 Da. A number of proteins were found to be expressed in the crude cell lysates of the uninduced and induced cell lysates (Figure 2.7).



**Figure 2.7: SDS-PAGE showing the proteins expressed in uninduced and induced crude cell lysates.** Lane 1: Wild type chick pea lectin; Lane 2: Low Molecular Weight protein standard (Amersham) showing bands corresponding to 97.0 kDa, 66.0 kDa, 45.0 kDa, 30.0 kDa and 20.1 kDa); Lane 3: **uninduced** culture of pET vector alone, without CAL gene (in BL21 DE3 CodonPlus cells); Lane 4: **induced** culture of pET vector alone, without CAL gene (in BL21 DE3 CodonPlus cells); Lane 5: **uninduced** culture 1 (pET with CAL gene in BL21 DE3 cells); Lane 6: **induced** culture 1 (pET with CAL gene in BL21 DE3 cells); Lane 7: **uninduced** culture 2 (pET with CAL gene in BL21 DE3 CodonPlus cells); and Lane 8: **induced** culture 2 (pET with CAL gene in BL21 DE3 CodonPlus cells).

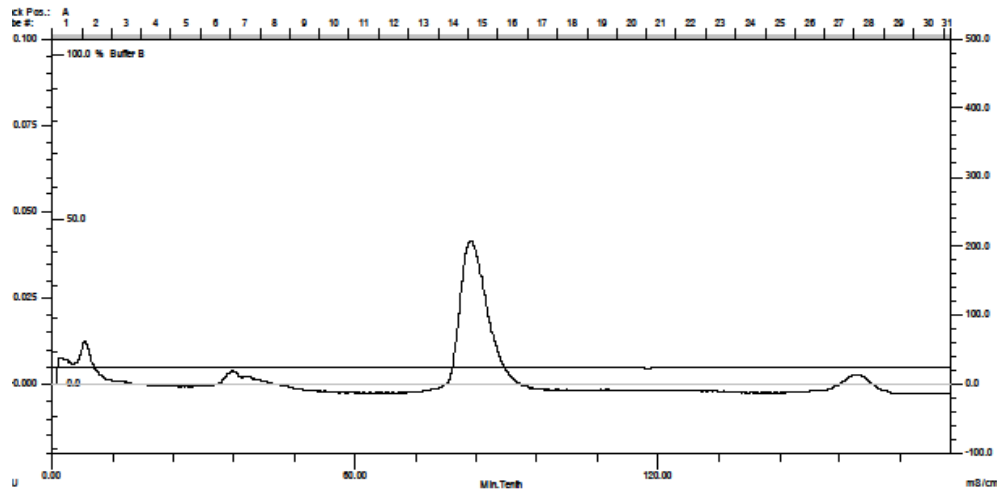
An induced protein with molecular weight less than 30,000 Da (approximately 27,000 Da) was observed in the induced samples but absent in the uninduced ones. The induction was more pronounced in the crude lysates in which the BL21-CodonPlus (DE3)-RIL cells were used as the host strain, compared to the BL21 DE3 cells. There occurred some expression of rCAL in the uninduced form of these cells also, due to leaky control of the T7 promoter. Even this extent of expression was more pronounced than that in the BL21 DE3 cells.

The lectin activity (against pronase-treated rabbit erythrocytes) could be detected in the soluble fraction of the induced cell lysates.

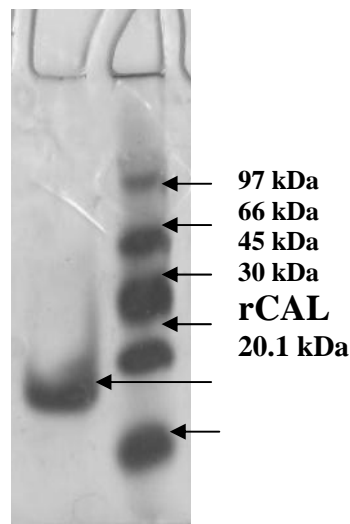


### [C] Purification of the recombinant lectin

The lectin fraction could be detected in the unbound fraction of the DEAE-cellulose column. On the gel filtration column, the lectin eluted in fractions numbered 15 and 16 (Figure 2.8). These fractions also showed 6-well hemagglutination activity and a single band on a 12 % SDS PAGE (Figure 2.9).

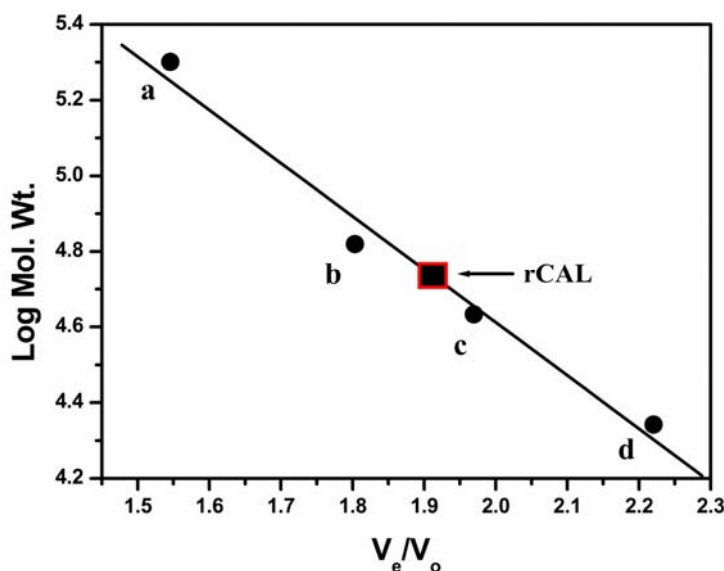


**Figure 2.8: FPLC profile of recombinant CAL on a Superose 12 prep grade gel filtration column.**



**Figure 2.9: 12% SDS PAGE for recombinant lectin.** Lane 1: FPLC pooled fractions 15 and 16 and Lane 2: Low molecular weight marker (Amersham) showing bands corresponding 97.0 kDa, 66.0 kDa, 45.0 kDa, 30.0 kDa and 20.1 kDa.

### 2.3.6. Molecular mass determination of purified lectin



**Figure 2.10: Molecular mass determination of rCAL by gel filtration chromatography on WATERS HPLC unit and WATERS gel filtration Protein Pak™ 300SW (7.5 mm × 300 mm column) column.** Standard proteins used were (a)  $\beta$ -amylase (200,000 Da), (b) bovine serum albumin (66,000 Da), (c) ovalbumin (43,000 Da) and (d) soybean trypsin inhibitor (22,000 Da). rCAL lectin (approximately 54,600 Da).

The lectin showed a molecular mass around 54,600 Da as determined by gel filtration (Figure 2.10), while in the denaturing gel, it appeared as a single band of approximate molecular mass of 27,000 Da indicating the lectin to be a homodimer. The molecular mass on the basis of number of amino acids (231 in each subunit) is around 52,608 Da. Around 5 to 8 mg of purified lectin could be obtained from one litre of *E. coli* culture. The specific activity of the purified lectin was obtained as  $5 \times 10^3$  hemagglutination units/mg.

The BL21-CodonPlus (DE3)-RIL cells have previously been used for the heterologous expression of several proteins including *Plasmodium falciparum*  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase [23] and of the major peanut allergen Ara h 2 [24]. We could successfully express the chick pea lectin gene in *E. coli* using these cells.

---

**REFERENCES**

1. Streicher, H. and Sharon, N. (2003), Recombinant Plant Lectins and Their Mutants Pages, *Methods in Enzymology*, 363, 47–77.
2. Pusztai, A. (1991), Plant Lectins, University Press, Cambridge.
3. Rudiger, H. (1997), Glycosciences. In Gabius H.-J. and Gabius S., editor. Chapman and Hall, Weinheim.
4. Varejão, N., Almeida, M.D.S, De Cicco, N.N.T., Atella, G.C, Coelho, L.C.B.B., Correia, M.T.S. and Foguel, D. (2010), Heterologous expression and purification of a biologically active legume lectin from *Cratylia mollis* seeds (CRAMOLL 1), *Biochimica et Biophysica Acta-Proteins and Proteomics*, 1804, 1917–1924.
5. Chambery, A., Severino, V., Stirpe, F. and Parente, A. (2007), Cloning and expression of the B chain of volkensin, type 2 ribosome inactivating protein from *Adenia volkensis* harms: Co-folding with the A chain for heterodimer reconstitution, *Protein Expression and Purification*, 51, 209–215.
6. Oliveira, C., Costa, S., Teixeira, J.A. and Domingues, L. (2009), cDNA Cloning and Functional Expression of the  $\alpha$ -D-Galactose-Binding Lectin Frutalin in *Escherichia coli*, *Molecular Biotechnology*, **43**, 212–220.
7. Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., Ma, Z. and Wang, Y. (2011), A jacalin-related lectin-like gene in wheat is a component of the plant defence system, *Journal of Experimental Botany*, **62**, 5471–5483.
8. Wu, Z.M., Yan, H.B., Pan, W.L., Jiang, B., Liu, J.G., Geng, B.J., Sun, Y.T., Wang, Y.H. and Dong, W.Q. (2012), Transform of an ectopically expressed bulb lectin gene from *Pinellia pedatisecta* into tobacco plants conferring resistance to aphids (*Myzus nicotianae*), *Australian Journal of Crop Science*, **6**, 904–911

9. Jiang, J.F., Han, Y., Xing, L.J., Xu, Y.Y., Xu, J.H., Chong, K. (2006), Cloning and expression of a novel cDNA encoding a mannose-specific jacalin-related lectin from *Oryza sativa*, *Toxicon*, **47**, 133–139.
10. Hulse, J.H. (1991), Nature, composition and utilization of grain legumes, In: *Uses of tropical Legumes: Proceedings of a Consultants' Meeting*, 11-27, ICRISAT Center. ICRISAT, Patancheru, Andhra Pradesh, India.
11. Kolberg, J., Michaelsen, T.E. and Sletten, K. (1983), Properties of a lectin purified from the seeds of *Cicer arietinum*, *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, **364**, 655–664.
12. Pedroche, J., Yust, M.M., Lqari, H., Megias, C., Calle, J.G., Alaiz, M., Millan, F. and Vioque, J. (2005), Chickpea pa2 albumin binds hemin. *Plant Science*, **168**, 1109-1114.
13. Esteban, R., Dopico, B., Muñoz, F. J., Romo, S. and Labrador, E. (2002), A seedling specific vegetative lectin gene is related to development in *Cicer arietinum*. *Physiologia Plantarum*, **114**, 619–626.
14. Qureshi, I.A., Srivastava, P.S. and Koundal, K.R. (2007), Isolation and characterization of a lectin gene from seeds of chickpea (*Cicer arietinum* L.), *DNA Sequence*, **18**, 196–202.
15. 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 Crystallographica*, **61**, 141–143.
16. Sambrook, J. and Russell, D.W. (2001), *Molecular Cloning: A Laboratory Manual*, Volume 1, 2, 3, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.
17. Oligotex handbook (2002), *Oligotex handbook for Purification of poly A+ RNA from total RNA*, ([www.qiagen.com](http://www.qiagen.com)).

18. Promega Technical Manual (2012), *ImProm-II™ Reverse Transcription System promega*, ([www.promega.com/tbs](http://www.promega.com/tbs)).
19. Promega Technical Manual (2012), A-tailing with pGEM-T Easy vector, ([www.promega.com](http://www.promega.com)).
20. Birnboim, H.C. and Doly, J. (1979), A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Research*, **7**, 1513–1523
21. Jerspeth, M., Breister, L. and Greener, A. (1998), *Strategies*, **11**, 3–4.
22. Laemmli, U.K. (1970), Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*, **227**, 680–685.
23. Lack G., Homberger-Zizzari E., Folkers G., Scapozza L., and Perozzo R. (2006), Recombinant Expression and Biochemical Characterization of the Unique Elongating  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase Involved in Fatty Acid Biosynthesis of *Plasmodium falciparum* Using Natural and Artificial Substrates, *The Journal of Biological Chemistry*, **281**, 9538–9546.
24. Lehmann K., Hoffmann S., Neudecker P., Suhr M., Becker W., and Rosch P. (2003), High-yield expression in *Escherichia coli*, purification, and characterization of properly folded major peanut allergen Ara h 2, *Protein Expression and Purification*, **31**, 250–259.

## **CHAPTER: 3**

---

# **LIGAND BINDING AND SOLUTE QUENCHING STUDIES OF THE RECOMBINANT LECTIN**

---

## SUMMARY

The recombinant chick pea lectin (rCAL) did not recognize any simple mono- or disaccharides, nor any glycoproteins other than fetuin and its asialo triantennary glycan. The binding with asialo fetuin glycan was spontaneous and enthalpically driven as revealed by thermodynamic parameters. The nucleotide sequence of rCAL showed the presence of the hemopexin fold. Binding of rCAL was checked with hemin, spermine, and thiamine using fluorescence spectroscopy. Hemin could strongly bind rCAL confirming the functional presence of the hemopexin domains. rCAL also showed binding to spermine and thiamine. Solute quenching studies showed that the microenvironment of the tryptophan in the lectin was negatively charged. Upon complete denaturation of the protein, the charge density around the tryptophan gets redistributed. The upward curvature obtained for acrylamide quenching was resolved with lifetime fluorescence measurements. The contribution of the static component ( $K_s$ ) was more prominent than that of the collisional component ( $K_{sv}$ ).

## 3.1. INTRODUCTION

Recognition is a key event in the biological functionality of the lectins. The unique and highly specific sugar-binding property of plant lectins have made these molecules useful probes for glycan detection in carbohydrates and glycoproteins for many years, providing the basis for either biochemical analysis or for development of diagnostic tools in histology, blotting and biosensor applications [1]. Lectins are used as “glycan deciphers” to interpret the carbohydrate structure in living organs and cells. The overall biological activities of the lectins are manifestation of their specificities.

It is essential to understand the mechanism of ligand binding to lectins, in order to facilitate their use as an analytical tool and for better understanding of lectin interaction with cell bound carbohydrates. Determination of association constants with a series of ligands provides considerable insight into the spatial features of the combining site of a lectin. Complementary thermodynamic data offer information on the forces involved in the binding and explain affinity differences encountered.

Since hemagglutination-inhibition studies provide semi-quantitative information, fluorescence spectroscopy has been used to obtain more quantitative information about a lectin's binding process. The advantage of using fluorimetry in studies of carbohydrate-protein interactions is that the binding can be studied at

equilibrium without physical separation of the bound complex from the free ligand and the protein [2].

The fluorescence of the indole side chains of the tryptophan (trp) residues in a protein is extremely sensitive to its environment, and hence changes which affect the tryptophan environment can cause changes in fluorescence properties of the protein [3]. Such changes in the intrinsic fluorescence of proteins have been used to obtain information regarding structure, specificity and conformation of the proteins [4]. The accessibility of tryptophan in a protein molecule can be measured by use of small molecular weight quenchers that perturb fluorescence [5]. Fluorescence quenching by these molecules proceeds mainly via physical contacts of the quenchers with the fluorophores and hence is directly dependent on the extent to which they can approach the fluorophores in the protein. This, in turn, makes the process extremely dependent on the nature of the solvent, the degree of exposure or burial of the trp residues as well as the nature of the quenchers themselves [6]. In cases, where trp residues are at the ligand-binding site, this technique has been used to study the changes resulting from ligand binding. Lifetime studies of fluorescence decay have similarly been used to obtain information regarding the trp environment and its interaction with various quenchers [7].

The present chapter describes the functional characterization of the recombinant *Cicer arietinum* lectin (rCAL) with respect to its carbohydrate and ligand binding properties. The binding studies of rCAL with the asialo triantennary glycan from fetuin and with hemin, spermine and thiamine were carried out using fluorescence spectroscopy. Thermodynamic parameters for saccharide binding have also been determined. The microenvironment of the tryptophan (accessibility of trp residue) has been investigated using solute quenching studies.

### 3.2. MATERIALS

Cultivar BDN 9-3 of chick pea seeds was obtained from Badnapur Agricultural University, Jalna, India. Pronase-E, carboxypeptidase, aminopeptidase and neuraminidase enzymes and hemin, spermine and thiamine were obtained from Sigma-Aldrich, India. Glucose, galactose, mannose, rhamnose, xylose, fucose, raffinose, trehalose, lactose, glucosamine, mannosamine, galactosamine, Methyl- $\alpha$ -glucose, methyl- $\beta$ -mannose, methyl- $\alpha$ -galactose, N-acetyl-mannosamine, N-acetyl-galactosamine, N-acetyl-D-glucosamine, D-mannitol, D-glucuronic acid, D-



polygalacturonic acid, D-glucose pentaacetate, bovine submaxillary mucin, and fetuin were obtained from Sigma-Aldrich, India. Acrylamide, cesium chloride and potassium iodide as well as all other general chemicals and buffers were procured from Sigma-Aldrich, India. Rabbit blood for the hemagglutination assay was obtained from the animal house of the National Institute of Virology, Pune, India.

### 3.3. METHODS

#### 3.3.1. Protein Preparation

The recombinant *Cicer arietinum* lectin was purified to homogeneity from the induced cell lysates of *E. coli* as mentioned in Chapter 2. This purified lectin, rCAL, was the subject of the present study.

#### 3.3.2. Preparation of asialo triantennary glycan

Triantennary-N-glycan was prepared from fetuin. One gram of the glycoprotein was dissolved in 100 ml of 20 mM Tris-HCl, pH 7.2 (containing 150 mM NaCl, 0.5 % w/v sodium azide), pH 7.2 and digested by adding 50 mg of pronase-E powder at 37 °C for 72 h; 20 mg of pronase-E was added after every 24 h. The digest was lyophilized, dissolved in 5 ml of 100 mM acetic acid, centrifuged (10,000 rpm, 20 min), and the supernatant was collected. The pellet was re-extracted five times in 1 ml of 100 mM acetic acid. Two ml of clear supernatant was chromatographed on Sephadex G-25 column (1.5 x 100 cm), pre-equilibrated with 20 mM acetic acid, and eluted with the same buffer at a flow rate of 20 ml/hour. The fractions (2 ml) were collected and those showing the presence of sugar were pooled and further digested by carboxypeptidase (10 U at pH 7.0 and 25 °C for 24 h) and aminopeptidase (10 U at pH 8.5 and 25 °C for 24 h). The residual peptides were removed by chromatography on a Dowex-50 column (1.5 x 4 cm) in 20 mM acetic acid [8, 9]. The glycopeptides were desialated by incubating the sample with 5 U of neuraminidase in 20 mM Tris-HCl buffer, pH 7.2 at 37 °C for 4 h; the enzyme and sialic acid were removed by successive chromatography on Sephadex G-25 (1.5 x 10 cm) and Dowex-50 as described above. Thus, the N-linked asialo-triantennary glycan was prepared from fetuin.

Protein concentrations were determined according to the method of Lowry *et al* [10] using bovine serum albumin (BSA) as the standard.

#### 3.3.3. Neutral sugar estimation

The sugar solutions (400 µl) were incubated with 400 µl of 5 % (w/v) phenol for 10 min at room temperature. Two ml of sulphuric acid was then added and the

mixture was allowed to cool for 20 min at room temperature. The colour developed was then measured spectrophotometrically at 490 nm. The concentration of the sugar was estimated using D-mannose as the standard [11].

#### **3.3.4. Erythrocyte preparation**

Rabbit erythrocytes were washed 5 to 6 times with 20 mM potassium phosphate buffer pH 7.2 containing 150 mM NaCl. A 3% (v/v) suspension of the erythrocytes in the above buffer was treated with the enzyme pronase (10 mg/ml) at 37 °C for 1 h, washed 3 times with the same buffer and used for further studies.

#### **3.3.5. Hemagglutination assays**

Hemagglutination assays were performed in standard microtitre plates by the two-fold serial dilution method. A 50 µl aliquot of the erythrocytes suspension was mixed with 50 µl of serially diluted lectin and agglutination was examined visually after incubation for one hour. A unit of hemagglutination activity (U) is expressed as the reciprocal of the highest dilution (titre) of the lectin that showed complete agglutination. The specific activity of the lectin is defined as the number of hemagglutination units/mg of the protein.

#### **3.3.6. Hemagglutination inhibition assays**

Hemagglutination inhibition assays were performed similarly, except that serially diluted sugar solutions (25 µl from 0.5 M stock) were pre-incubated for 15 min at 27 °C with 25 µl of the lectin (600 ng). Erythrocyte suspension (50 µl) was then added, mixed and the plates were examined visually after one hour.

#### **3.3.7. Fluorescence measurements**

Steady-state intrinsic fluorescence measurements were performed on a Perkin Elmer LS-50B luminescence spectrofluorimeter at 25 °C. Samples were excited at 295 nm (1.0 cm cell path length) and the emission spectra were recorded in the range of 310 nm to 400 nm. Slit widths of 7 nm each were set for excitation and emission monochromators and the spectra were recorded at 100 nm/min. The baseline was corrected by subtracting the contribution of the buffer. rCAL sample (150 µg/ml) in 20 mM phosphate buffer, pH 7.2 was placed in a quartz cuvette maintained at desired temperature ( $\pm 0.1$  °C) by means of a Julabo circulating water bath. The sugar/ligand solution was added in 10-12 aliquots (5 to 10 µl each). A 1.8 mM stock concentration of the asialo triantennary glycan from fetuin was used. The ligands were used in the concentration of 10 µM (for hemin) and 5 mM (for spermine and thiamine). Each spectrum was an average of 5 accumulations. The fluorescence intensity at 343 nm

( $\lambda_{\max}$  of the lectin) was considered for further analysis. Corrections were also made to compensate the dilution effect upon addition of sugar/ligand to the lectin. At the highest concentration of the sugar/ligand to lectin, volume change was less than 5 % of the solution in the cuvette. Each data point was an average of three independent sets of experiments with SD less than 2 %.

### 3.3.8. Ligand-binding data analysis

The association constants were calculated according to the method described by Chipman *et al.* [12]. The abscissa intercept of the plot of  $\log [C]_f$  against  $\log \{(\Delta F)/(F_c - F_\infty)\}$ , where  $[C]_f$  is the free ligand concentration, yielded  $pK_a$  value for lectin-ligand interaction according to the relationship

$$\log [F_o - F_c / F_c - F_\infty] = \log K_a + \log \{[C]_t - [P]_t (\Delta F / \Delta F_\infty)\} \quad (1)$$

where  $F_c$  is the fluorescence intensity of the lectin at any point during the titration,  $[P]_t$  is the total protein concentration,  $\Delta F_\infty$  is the change in fluorescence intensity at saturation binding,  $[C]_t$  is the total ligand concentration, and  $[C]_f$  is the free ligand concentration, given by,

$$[C]_f = \{[C]_t - [P]_t (\Delta F / \Delta F_\infty)\} \quad (2)$$

Free energy changes of association ( $\Delta G$ ) were determined by the equation,

$$\Delta G = -RT \ln K_a \quad (3)$$

Temperature dependence of the association constants was used to determine the thermodynamic parameters. Changes in enthalpy ( $\Delta H$ ) were determined from the Van't Hoff plots by using the equation,

$$\ln K_a = (-\Delta H / RT) + \Delta S / R \quad (4)$$

where  $\Delta H$  is enthalpy change,  $R$  is gas constant,  $\Delta S$  is entropy change and  $T$  is the absolute temperature. The entropy change was obtained from the equation,

$$\Delta G = \Delta H - T\Delta S \quad (5)$$

### 3.3.9. Solute quenching studies

#### 3.3.9.1. Steady state fluorescence spectroscopy

Fluorescence quenching experiments were carried out by the addition of a small aliquot of acrylamide, potassium iodide (KI) or cesium chloride (CsCl) stock solution (5 M) to the protein solution (180  $\mu\text{g/ml}$ ) at 25 °C and the fluorescence intensities were determined. Iodide stock solution contained 0.2 M sodium thiosulfate to prevent formation of tri-iodide ( $I^3$ ). For quenching studies with denatured protein, rCAL was

incubated with 6 M GdnHCl overnight (16 hours) at 25 °C. Fluorescence intensities were corrected for volume changes before further analysis of quenching data.

The steady-state fluorescence quenching data obtained with all the quenchers used in this study were analyzed by Stern–Volmer (Eq. 6) and modified Stern–Volmer (Eq. 7) equations in order to obtain quantitative quenching parameters [5, 13].

$$F_0/F_c = 1 + K_{sv} [Q] \quad (6)$$

$$F_0/\Delta F = f_a^{-1} + 1/[K_a f_a(Q)] \quad (7)$$

where  $F_0$  and  $F_c$  are the relative fluorescence intensities in the absence and presence of the quencher, respectively,  $[Q]$  is the quencher concentration,  $K_{sv}$  is Stern–Volmer quenching constant,  $\Delta F = F_0 - F_c$  is the change in fluorescence intensity at any point in the quenching titration,  $K_a$  is the quenching constant and  $f_a$  is the fraction of the total fluorophores accessible to the quencher. Equation (7) shows that the slope of a plot of  $F_0/\Delta F$  versus  $1/Q$  (modified Stern–Volmer plot) gives the value of  $(K_a f_a)^{-1}$  and its Y-intercept gives the value of  $f_a^{-1}$ .

### 3.3.9.2. Time resolved fluorescence spectroscopy

Lifetime measurements were carried out on an FLS-920 single photon counting spectrofluorimeter supplied by Edinburgh Instruments. A laser pico second pulsed light emitting diode (model EPLED-295) was used as the excitation source and a Synchronization photomultiplier was used to detect the fluorescence. The diluted Ludox solution was used for measuring Instrument Response Function (IRF). The sample (200 µg/ml) was excited at 295 nm and emission was recorded at 343 nm. Slit widths of 10 nm each were used for the excitation and emission monochromators. The resultant decay curves were analyzed by a Reconvolution fitting program supplied by Edinburgh Instruments.

The static and dynamic quenching components obtained for the denatured lectin with acrylamide quenching were further resolved by fluorescence lifetime measurements using the following equations

$$\tau_0/\tau = (1 + K_{sv} [Q]) \quad (8)$$

where  $\tau_0$  is the average lifetime in the absence of the quencher and  $\tau$  is the lifetime in the presence of the quencher at a concentration  $[Q]$  and

$$F_0/F_c = (1 + K_{sv} [Q]) (1 + K_s [Q]) \quad (9)$$

where  $K_{sv}$  is the Stern-Volmer (dynamic) quenching constant,  $K_s$  is the static quenching constant and  $[Q]$  is the quencher concentration. The dynamic quenching

constant reflects the degree to which the quencher achieves the encounter distance of the fluorophore and can be determined by the fluorescence lifetime measurements according to the equation (9).

### 3.4. RESULTS AND DISCUSSION

#### 3.4.1. Hemagglutination and hemagglutination inhibition assay

The lectin showed hemagglutination activity only against pronase-treated rabbit erythrocytes. The hemagglutination titre was obtained to be  $5 \times 10^3$  hemagglutination units per mg of protein. Simple sugars like glucose, mannose, galactose, rhamnose, xylose, fucose, raffinose, trehalose, lactose, glucosamine, mannosamine, galactosamine, Methyl- $\alpha$ -glucose, methyl- $\beta$ -mannose, methyl- $\alpha$ -galactose, N-acetyl-mannosamine, N-acetyl-galactosamine, N-acetyl-D-glucosamine, D-mannitol, D-glucuronic acid, D-polygalacturonic acid, D-glucose pentaacetate etc failed to inhibit the hemagglutinating activity of the lectin. Hemagglutination by rCAL was inhibited by glycoproteins viz. fetuin, asialofetuin, and its asialo-triantennary glycopeptides. The minimum inhibitory concentration for fetuin and its asialo glycan was found to be 32  $\mu\text{g}$ . Lectins recognizing only complex glycans have been reported earlier. For example, the lectin from wild cobra lily, *Arisaema flavum* [14] shows inhibition only with asialo fetuin with an inhibitory concentration of 250  $\mu\text{g}/\text{ml}$ .

#### 3.4.2. Ligand binding with Asialo triantennary glycan from Fetuin

Fluorescence titration of rCAL with the N-linked asialo triantennary glycan prepared from fetuin (Figure 3.1(A)) resulted in quenching of the lectin fluorescence (inset, Fig 3.1(B)). The value of  $F_\infty$  was derived from the plot shown in Figure 3.1(B). The binding constant  $K_a$  of  $1.01 \times 10^5 \text{ M}^{-1}$  at 25  $^\circ\text{C}$  was determined from the plot of  $\log [\Delta F / (F_c - F_\infty)]$ , versus  $\log [C]_f$  (Figure 3.1(C)).

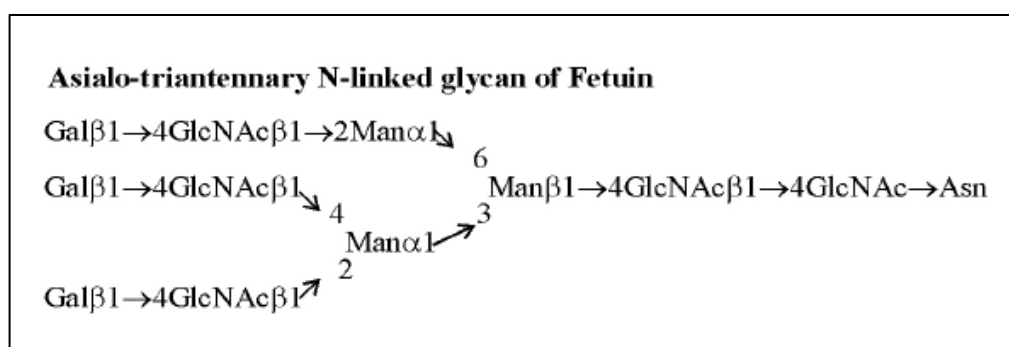
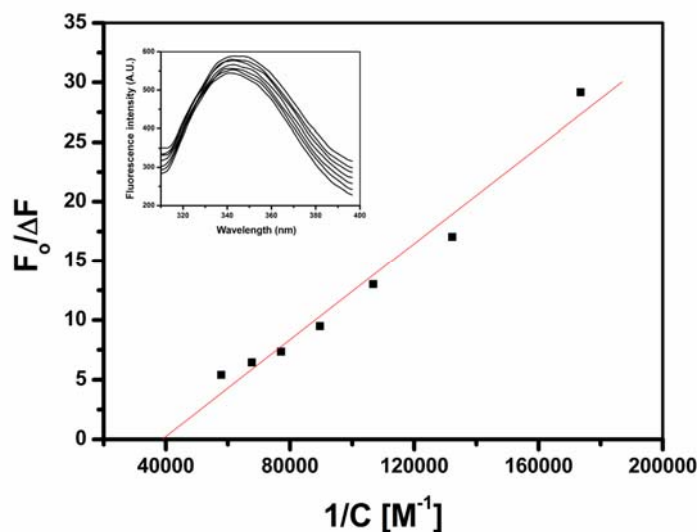
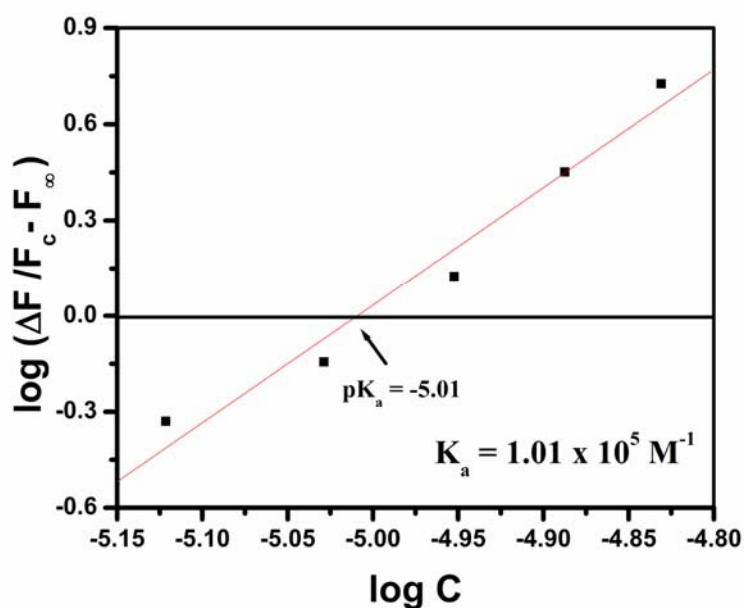


Figure 3.1(A): Structure of the asialotriantennary glycan from fetuin



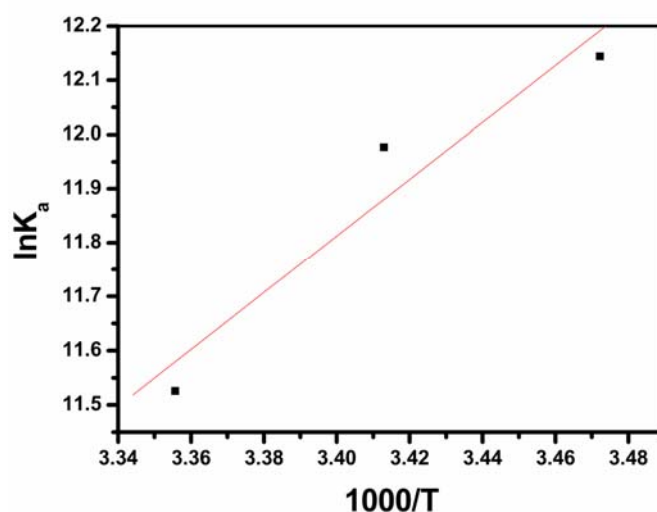
**Figure 3.1(B):** The fluorescence change at saturation binding ( $F_{\infty}$ ) is obtained from the Y-intercept of the plot of  $F_0/\Delta F$  versus  $1/C$ . Inset: Quenching of fluorescence of rCAL by the glycan.



**Figure 3.1(C):** Plot of  $\log \{ \Delta F / (F_c - F_{\infty}) \}$  versus  $\log [C]$ . The X-intercept of the plot gives  $pK_a$  value for the interaction between rCAL and the glycan.

Binding was also checked at 15 °C ( $K_a$ ,  $1.88 \times 10^5 \text{ M}^{-1}$ ) and 20 °C ( $K_a$ ,  $1.59 \times 10^5 \text{ M}^{-1}$ ). A decrease in the association constant ( $K_a$ ) was observed with increase in

temperature. Thermodynamic parameters were calculated for the glycan binding. Van't Hoff plots were linear ( $r > 0.9$ ) in the range of temperatures studied (Figure 3.1(D)).



**Figure 3.1(D):** Van't Hoff plot for the association drawn according to the regression equation ( $r > 0.9$ ,  $N = 3$ ).

The change in Gibb's free energy ( $\Delta G$ ) was  $-28.56 \text{ kJmol}^{-1}$  indicating binding to be spontaneous; the negative value of change in enthalpy ( $\Delta H$ ) of  $-43.65 \text{ kJmol}^{-1}$  demonstrated the exothermic nature of the reaction. A negative change in entropy ( $\Delta S$ ) ( $-50.65 \text{ Jmol}^{-1}\text{K}^{-1}$ ) indicated the formation of more H-bonds between the glycan and the lectin. Few reports are available on lectins showing inhibition against complex glycans, like the lectin IV from *Griffonia simplicifolia* and PHA-L from *P. vulgaris* [15]. These belong to the group termed "complex" with specificity towards N-glycans. A stronger inhibition of rCAL activity by asialo fetuin than that with fetuin was observed. The interference due to sialic acid may be abolished by the action of neuraminidase, resulting in better interaction with rCAL (due to the exposure of the Gal $\beta$ 1 $\rightarrow$ 4GlcNAc residue). A similar observation has been made for the lectin from *Fusarium solani*, by Khan *et al* [16] which interacts better with the asialo glycans of fetuin and fibrinogen than with simple galactose residue or its derivatives. The binding in this case is also enthalpically driven.

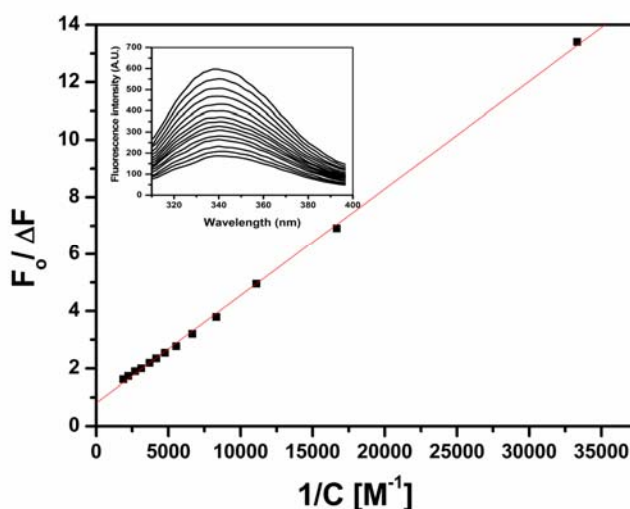
Lectin-carbohydrate interactions are generally characterized by a low affinity for monovalent ligand. In general, affinity in the milli molar range is observed for lectins binding to monosaccharides. When longer oligosaccharides act as ligands, corresponding to an extended binding site on the lectin surface, increased affinity up to

micromolar range can be observed. Branched complex glycans, due to the clustering effect, offer multiple binding sites leading to several fold increase in the affinity [17].

### 3.4.3. Ligand binding with hemin, spermine and thiamine

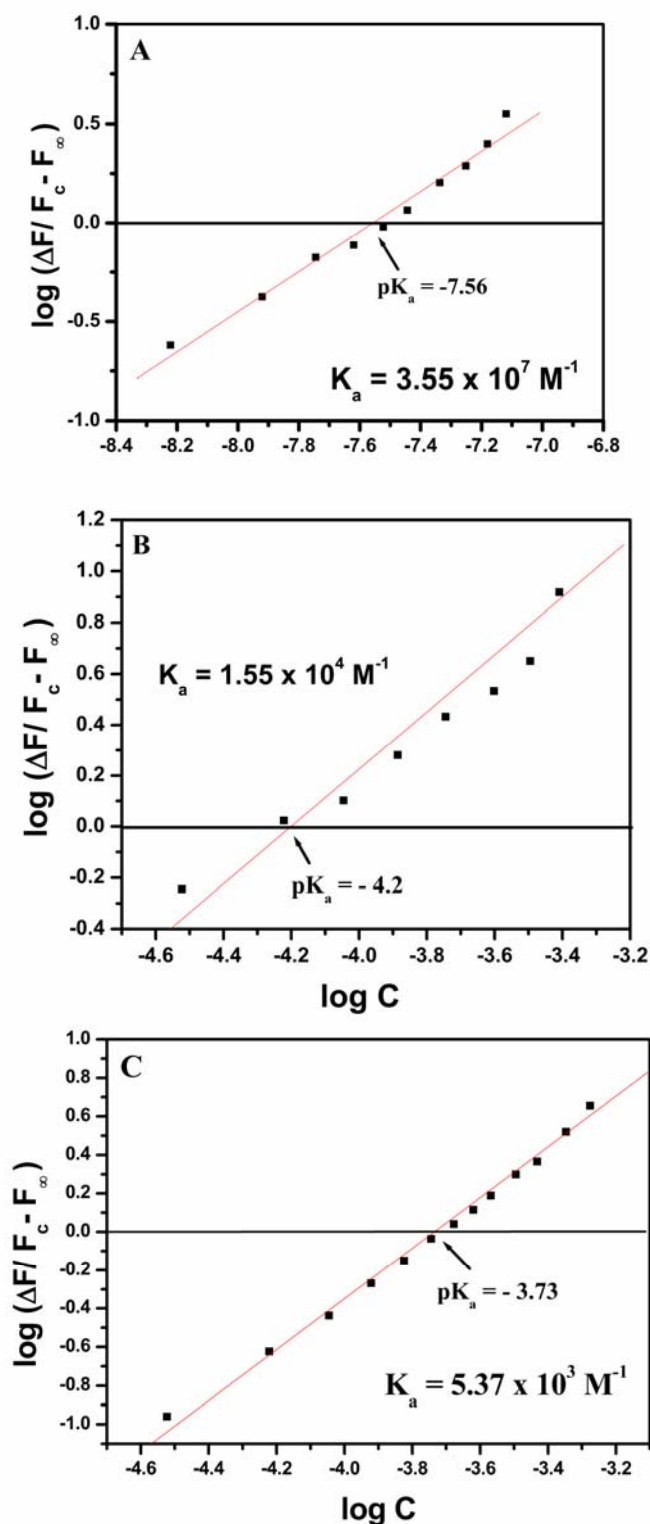
As a first step towards investigating its biological role, binding of rCAL with hemin, spermine and thiamine was examined using fluorescence spectroscopy.

Titration of rCAL with hemin at 25 °C resulted in gradual quenching of the fluorescence (Figure 3.2, inset). The slope of the plot of  $\log [\Delta F/F_c - F_\infty]$ , versus  $\log [C]_f$  was near unity, hence one binding site per monomer of rCAL was predicted. The lectin bound hemin with an association constant,  $K_a = 3.55 \times 10^7 \text{ M}^{-1}$  (Figure 3.3A), indicating high affinity of binding. A similar quenching profile was observed when rCAL was titrated with 5 mM each of Spermine and Thiamine (Figures 3.3B and 3.3C) at 25 °C. The association constants  $K_a$  obtained for spermine and thiamine binding were  $1.55 \times 10^4 \text{ M}^{-1}$  and  $5.37 \times 10^3 \text{ M}^{-1}$ , respectively (Table 3.1), indicating significant affinity for the lectin.



**Figure 3.2:** The fluorescence change at saturation binding ( $F_\infty$ ) is obtained from the Y-intercept of the plot of  $F_0/\Delta F$  versus  $1/C$ . Inset: Representative quenching of fluorescence of rCAL by thiamine.





**Figure 3.3: Determination of the association constant for the interaction of rCAL with (A) hemin, (B) spermine, and (C) thiamine from the plot of  $\log \{ \Delta F / (F_c - F_\infty) \}$  versus  $\log [C]_f$ . The X-intercept of the plot gives  $pK_a$  value for the interaction. The stock concentrations of the ligands A, B, C were  $10 \mu\text{M}$ ,  $5 \text{mM}$  and  $5 \text{mM}$ , respectively.**

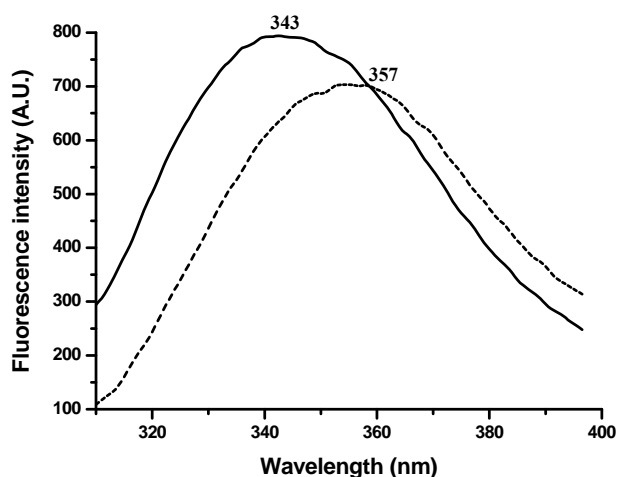
**Table 3.1 Summary of ligand binding to rCAL with fluorescence spectroscopy**

Ligand	$K_a$ ( $M^{-1}$ )	$\Delta G^*$ ( $kJmol^{-1}$ )
Hemin	$3.55 \times 10^7$	- 43.0727
Spermine	$1.55 \times 10^4$	- 23.90509
Thiamin	$5.37 \times 10^3$	- 21.27883

$$*\Delta G = -RT \ln(K_a)$$

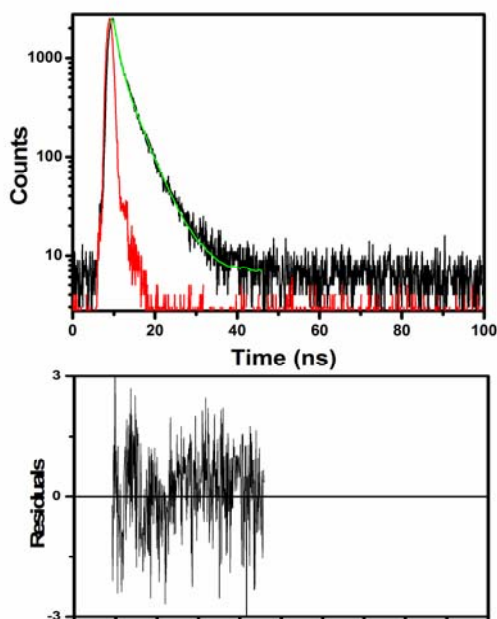
### 3.4.4. Steady-state and Lifetime fluorescence measurements of rCAL

The gene sequence revealed the presence of a single tryptophan (trp) in rCAL. The native protein showed intrinsic fluorescence with maximum emission ( $\lambda_{max}$ ) at 343 nm, suggesting the presence of the single trp in a partially hydrophobic environment (Figure 3.4(A)).



**Figure 3.4(A): Intrinsic fluorescence spectrum of rCAL (150  $\mu g/ml$ )** (straight line: native rCAL, dashed line: rCAL denatured with 6M GdnHCl).

The lifetime of intrinsic emission decay of rCAL was studied in nanosecond domain (Figure 3.4(B)). When fitted into a bi-exponential curve ( $\chi^2 < 1.26$ ), it could be described by two decay components  $\tau_1$  and  $\tau_2$  with decay times of 1.04 ns (45.40 %) and 4.14 ns (54.60 %) respectively. The average  $\tau$  was found to be 2.65 ns. The single trp thus existed as two different conformers at a given time: one with the shorter lifetime lies on the surface of the protein and its fluorescence decays faster, while the longer conformer lies in the interior and decays slowly. The  $\lambda_{max}$  at 343 nm is hence the cumulative intrinsic fluorescence of rCAL.



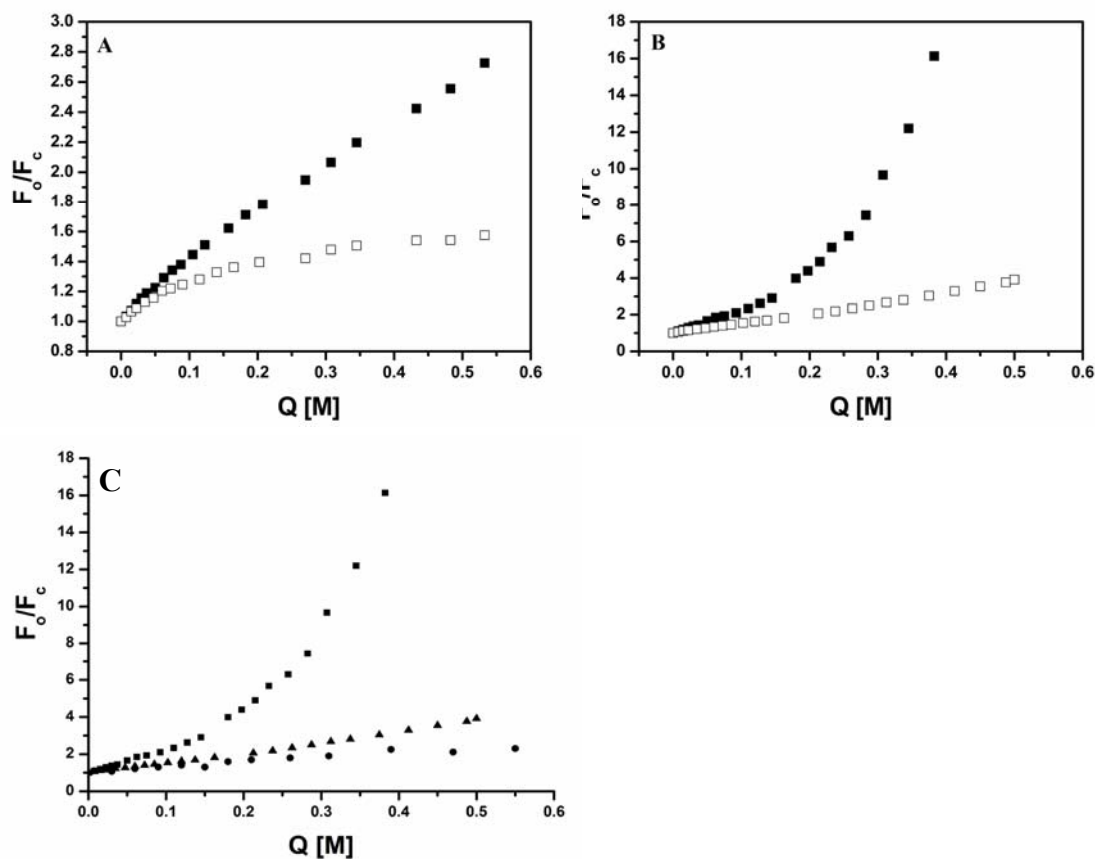
**Figure 3.4(B): Time-resolved fluorescence decay profile of native rCAL (300 µg/ml).** The red lines correspond to the instrument response, the black lines correspond to the experimental data and the green lines correspond to the nonlinear biexponential fit of the experimental data to a biexponential function. The lower panel represents the residuals.

#### 3.4.5. Solute Quenching studies

Acrylamide, being neutral by nature, was found to be the most efficient quencher for intrinsic fluorescence of rCAL in the native state, with a Stern Volmer ( $K_{sv}$ ) constant of  $3.17 \text{ M}^{-1}$  (Figure 3.5(A)). 73% of the trp fluorescence was accessible to it. Among the ionic quenchers, fluorescence could be quenched only by CsCl and not by KI, indicating a negative charge density around the trp. The biphasic curvature obtained for CsCl quenching (Figure 3.5(A)) indicated a heterogeneous ionic environment around the trp with the presence of two conformers – one getting quenched earlier than the other ( $K_{sv1} = 3.3 \text{ M}^{-1}$  and  $K_{sv2} = 1.02 \text{ M}^{-1}$ ).

The ionic solute quenching profile was reversed for the denatured protein. The lectin denatured with 6 M GdnHCl could be quenched only by KI and not by CsCl, indicating a reorientation of charge density around the trp from negative to positive.  $K_{sv}$  of  $5.59 \text{ M}^{-1}$  and 70 % accessibility was obtained for KI (Figure 3.5(B)). The trp environment was examined in the partially unfolded state. In the vicinity of 1.75 M

GdnHCl, rCAL was partially unfolded with a  $\lambda_{\max}$  of 348 nm (Chapter 4, Figure 4.6 (A)). Even at this stage, the fluorescence quenching with CsCl was very low ( $K_{sv} = 0.35 \text{ M}^{-1}$ ), while that with KI was significant ( $K_{sv} = 2.97 \text{ M}^{-1}$ ) (Figure 3.5(C)).



**Figure 3.5: Stern-Volmer plots for quenching of rCAL.** (A) Native lectin with acrylamide (filled squares) and CsCl (open squares). (B) 6 M GdnHCl-denatured lectin with acrylamide (filled squares) and KI (open squares). (C) 6 M GdnHCl-denatured lectin with acrylamide (filled squares); 1.75 M-denatured lectin with KI (filled circles) and 6 M-denatured lectin with KI (filled triangles). After fitting the data, the R value in each case was 0.99 (except filled squares in (B) and (C)). Protein concentration used was  $180 \mu\text{g/ml}$ .

Accessibility of acrylamide to fluorescence increased to 100% upon denaturation of the lectin with 6 M GdnHCl. An upward curvature was obtained in the Stern Volmer's Plot, indicating the involvement of both collisional and static components (Figure 3.5(B)). The static mechanism is a consequence of complex formation, while the dynamic mechanism involves collisions with acrylamide during the lifetime of the excited trp [18].

The average lifetimes were calculated using the following equations [19, 20]:

$$\tau = \sum_i \alpha_i \tau_i / \sum_i \alpha_i \quad \text{Eq. 10}$$

$$\langle \tau \rangle = \sum_i \alpha_i \tau_i^2 / \sum_i \alpha_i \tau_i \quad \text{Eq. 11}$$

For  $i = 1, 2, 3, \dots$ , and where  $\tau$  and  $\langle \tau \rangle$  are the average life times obtained by two different approaches and  $\alpha$  is the weighting factor (see Table 3.2).

**Table 3.2: The lifetimes of fluorescence decay of denatured rCAL and the corresponding pre-exponential factors along with calculated average lifetimes for acrylamide quenching.**

Q[M]	$\tau_1$ (ns)	$\alpha_1$	$\tau_2$ (ns)	$\alpha_2$	$\tau_3$ (ns)	$\alpha_3$	$\tau$ (ns)	$\langle \tau \rangle$ (ns)	$\chi^2$
0	0.030	0.076	2.61	0.043	6.15	0.003	1.86	2.55	1.21
0.015	1.13	0.073	3.48	0.036	21.83	0.000	1.67	2.11	1.21
0.03	0.79	0.076	2.87	0.046	7.68	0.002	1.56	2.38	1.21
0.055	0.63	0.077	2.10	0.051	4.81	0.008	1.43	2.27	1.18
0.08	0.72	0.084	2.19	0.045	6.19	0.003	1.35	2.11	1.20
0.105	0.71	0.086	2.02	0.044	5.00	0.004	1.27	1.90	1.30
0.130	0.58	0.098	1.76	0.049	5.01	0.004	1.08	1.66	1.12
0.165	0.64	0.101	1.72	0.043	4.83	0.004	1.07	0.98	1.07
0.200	0.77	0.132	2.75	0.018	13.10	0.000	1.01	1.42	0.96
0.235	0.61	0.129	1.81	0.031	5.78	0.002	0.90	1.48	1.15
0.270	0.72	0.147	2.87	0.014	10.75	0.000	0.91	1.31	1.13
0.305	0.49	0.156	1.72	0.029	5.80	0.002	0.74	1.38	1.07
0.355	0.44	0.176	1.48	0.032	4.96	0.003	0.66	1.27	1.04
0.405	0.44	0.160	1.23	0.039	4.26	0.004	0.67	1.2	1.01
0.455	0.47	0.194	1.91	0.020	6.66	0.001	0.63	1.18	1.03
0.505	0.36	0.188	1.03	0.047	4.23	0.005	0.57	1.19	1.06

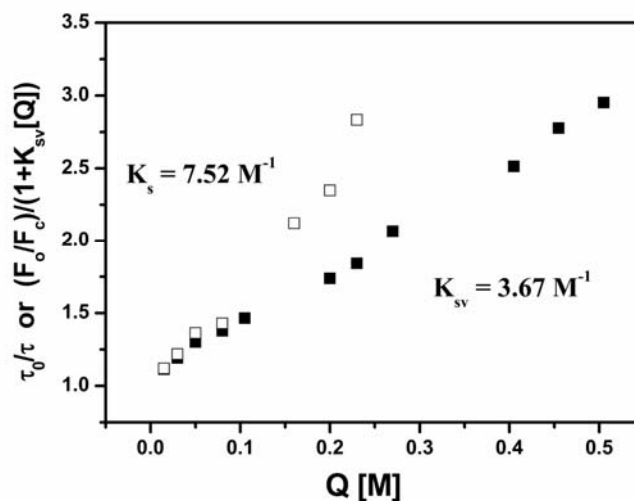
Using the average lifetimes obtained from the analysis of time resolved fluorescence data (Table 3.2), the value of  $K_{sv}$  for the lectin was obtained as  $3.67 \text{ M}^{-1}$  (Figure 3.6) (from earlier equations 8 and 9).

$$\tau_0/\tau = (1+K_{sv} [Q])$$

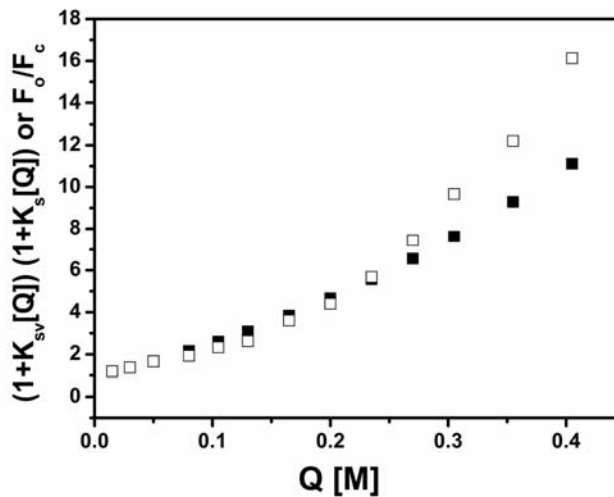
$$F_0/F_c = (1+K_{sv} [Q]) (1+K_s [Q])$$

Putting this value in Eq. (9) and plotting a graph of  $(F_0/F_c)/(1+K_{sv}[Q])$  against  $[Q]$ , the value of the static quenching constant ( $K_s$ ) was found to be  $7.52 \text{ M}^{-1}$  (Figure 3.6). The static component hence contributes more to the quenching by acrylamide than the dynamic component. The bimolecular quenching constant,  $k_q$  (calculated as  $k_q = K_{sv}/\tau$ ), was found to be  $1.39 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$ . According to Lehrer *et al* [13], this value of  $k_q$  corresponds to a negatively charged indole-3-acetate moiety; it correlates with our trp microenvironment which is negatively charged in the native state. Lectins showing lack of positive charge density around the tryptophan residues in their native state not been reported yet.

Incorporating the values of  $K_{sv}$  and  $K_s$  in the expression  $(1+K_{sv}[Q])(1+K_s[Q])$ , the values obtained were plotted against  $[Q]$ . It was observed that the values of  $F_0/F_c$  and  $(1+K_{sv}[Q])(1+K_s[Q])$  match very well (Figure 3.7).



**Figure 3.6: Acrylamide quenching of denatured rCAL.** Calculation of the collisional ( $K_{sv}$ ) (filled squares) and static ( $K_s$ ) (open squares) quenching constants for the upward curvature obtained with acrylamide for rCAL denatured with 6M GdnHCl.



**Figure 3.7: Coincidence plot for rCAL.** The plot of  $F_0/F_c$  and  $(1+K_{sv}[Q])(1+K_s[Q])$  against  $[Q]$  corresponding to quenching of denatured rCAL with acrylamide (open squares:  $F_0/F_c$ ; closed squares  $(1+K_{sv}[Q])(1+K_s[Q])$ ).

The summary of quenching parameters obtained for rCAL is given in Table 3.3.

**Table 3.3: Summary of the fluorescence quenching parameters for rCAL**

Quencher	$K_{sv} (M^{-1})/ K_s (M^{-1})$	$f_a$
<b>Acrylamide</b>		
Native	3.17	0.73
6 M GdnHCl- treated	3.67 <sup>*</sup> /7.52 <sup>*</sup>	1.05
<b>Cesium chloride</b>		
Native	3.3, 1.02	0.5
1.75 M GdnHCl-treated	No fluorescence quenching observed	
6 M GdnHCl- treated	No fluorescence quenching observed	
<b>Potassium iodide</b>		
Native	No fluorescence quenching observed	
1.75 M GdnHCl-treated	3.02	0.73
6 M GdnHCl- treated	5.59	0.70

\* Derived from time resolved fluorescence.

Several lectins expressing complex sugar specificity have been reported. Examples include lectins from *Moringa oleifera* [21], *Arisaema curvatum* [22], *Sauromatum venosum* [23], *Acacia constricta* [24], *Scilla campanulata* [25], *Ficus cunia* [26], *Salvia sclerea* [27], and *Glechoma hederocsea* [28] etc.

rCAL is an albumin by nature, hence binding properties of homologous plant seed albumins for other ligands were checked for rCAL. It was found that rCAL showed similar binding properties. Similar studies of binding with hemin, spermine or thiamine have not been reported for other lectins, though there are reports of lectins known to bind hydrophobic ligands like adenine [29, 30].

---

**REFERENCES**

1. Sharon, N. and Lis, H. (1990), Legume lectins—a large family of homologous proteins, *FASEB Journal*, **4**, 3198–3208.
2. Lee, Y.C. (1997), Fluorescence Spectrometry in Studies of Carbohydrate-Protein Interactions, *Journal of Biochemistry*, **121**, 818–825.
3. Lakowicz, J.R. (1983), *Principles of Fluorescence Spectroscopy*, Plenum Press, New York.
4. Gaikwad, S.M., Gurjar, M.M. and Khan M.I. (2002), *Artocarpus hirsuta* lectin. Differential modes of chemical and thermal denaturation, *European Journal of Biochemistry*, **269**, 1413–1417.
5. Eftink, M.R. and Ghiron, C.A. (1984), Indole fluorescence quenching studies on proteins and model systems: use of the inefficient quencher succinimide, *Biochemistry*, **23**, 3891–3889.
6. Kronman, M.J. and Holmes, L.G. (1971), The fluorescence of native denatured reduced-denatured proteins, *Photochemistry and Photobiology*, **14**, 113–134.
7. Alexander Ross, J.B., Schmidt, C.J. and Brand, L. (1981), Time-resolved fluorescence of the two tryptophans in horse liver alcohol dehydrogenase, *Biochemistry*, **20**, 4369–4377.
8. Lis, H., Sharon, N. and Katchalski, E. (1996), Soybean hemagglutinin a plant glycoprotein: Isolation of glycopeptides, *Journal of Biological Chemistry*, **241**, 684–689.
9. Townsend, R.R., Hardy, M.R., Wong, T.C. and Lee, Y.C. (1986), Binding of N-linked bovine fetuin glycopeptides to isolated rabbit hepatocytes: Gal/GalNAc hepatic lectin discrimination between Gal- $\beta$ -(1, 4) GlcNAc and Gal- $\beta$ -(1, 3) GlcNAc in triantennary structure, *Biochemistry*, **25**, 5716–5725.



10. Lowry, O.H., Rosenbrough, N.H., Lewis, F.A. and Randall, R.J. (1951), Protein measurement with the folin reagent, *Journal of Biological Chemistry*, **193**, 265–275.
11. Dubois, M., Gills, K.K., Hamilton, J.K., Rebers, P.A. and Smoth, F. (1956), Colorimetric method for determination of sugars and related substances, *Analytical Chemistry*, **28**, 350–356.
12. Chipman, D.M., Grisaro, V. and Sharon, N. (1967), The binding of oligosaccharides containing N-acetylglucosamine and N-acetylmuramic acid to lysozyme. The specificity of binding subsites, *Journal of Biological Chemistry*, **242**, 4388–4394.
13. Lehrer, S.S. and Leavir, P.C. (1978), *Methods in Enzymology*, **49**, 222–236.
14. Singh, J., Singh, J. and Kambhoj, S.S. (2004), A novel mitogenic and antiproliferative lectin from a wild cobra lily, *Arisaema flavum*, *Biochem. Biophys. Res. Commun.*, **318**, 1057–1065.
15. Guzman-Partida, A.M., Robles-Burueno, M.R., Ortega-Nieblas, M. and Vazquez-Moreno, M. (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.
16. Khan, F., Ahmad, A. and Khan, M.I. (2007), Interaction of *Fusarium Solani* Lectin with Monosaccharides and Oligosaccharides: A Fluorometric Study, *Photochemistry and Photobiology*, **83**, 966–970.
17. 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.
18. 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.

19. Inokuti, M. and Hirayama, F. (1965), Influence of energy transfer by the exchange mechanism on donor luminescence, *Journal of Chemical Physics*, **43**, 1978–1989.
20. Grinvald, A. and Steinberg, I.Z. (1974), On the analysis of fluorescence decay kinetics by the method of least-squares, *Analytical Biochemistry*, **59**, 583–598.
21. Katre, U.V., Suresh, C.G., Khan, M.I. and Gaikwad, S.M. (2008), Structure–activity relationship of a hemagglutinin from *Moringa oleifera* seeds, *International Journal of Biological Macromolecules*, **42**, 203–207.
22. 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.
23. Singh, B.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*), *Biochimica et Biophysica Acta*, **1723**, 163–174.
24. 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.
25. Wright, L.M., Van Damme, E.J.M., Barre, A., Allen, A.K., Van Leuven, F., Reynolds, C.D., Rouge, P. and Peumans, W.J. (1999), Isolation, characterization, molecular cloning and molecular modelling of two lectins of different specificities from bluebell (*Scilla campanulata*) bulbs, *Biochemical Journal*, **340**, 299–308.
26. Ray, S., Ahmed, H., Basu, S. and Chatterjee, B.P. (1993), Purification, characterization, and carbohydrate specificity of the lectin of *Ficus cunia*. *Carbohydrate Research*, **242**, 247–263.

- 
27. Piller, V., Piller, F. and Cartron, J.P. (1986), Isolation and characterization of an *N*-acetyl galactosamine specific lectin from *Salvia sclarea* seeds, *Journal of Biological Chemistry*, **261**, 14069–14075.
28. Singh, T., Wu, J.H., Peumans, W.J., Rougé, P., Van Damme, E.J.M., Alvarez, R. A., Blixt, O. and Wu, A.M. (2006), Carbohydrate specificity of an insecticidal lectin isolated from the leaves of *Glechoma hederacea* (ground ivy) towards mammalian glycoconjugates, *Biochemical Journal*, **393**, 331–341.
29. 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.
30. Kavitha, M., Sultan, N.A.M. and Swamy, M.J. (2009), Fluorescence studies on the interaction of hydrophobic ligands with *Momordica charantia* (bitter gourd) seed lectin, *Journal of Photochemistry and Photobiology B: Biology*, **94**, 59–64.

## **CHAPTER: 4**

---

# **CONFORMATIONAL TRANSITIONS OF THE RECOMBINANT LECTIN**

---

## SUMMARY

Conformational transitions of the recombinant *Cicer arietinum* lectin were studied using fluorescence and circular dichroism spectroscopy. Thermal denaturation of rCAL caused rapid secondary structural rearrangements above 50 °C and transient exposure of hydrophobic residues at 55 °C, leading to aggregation. The denaturation was irreversible. GdnHCl-mediated unfolding of rCAL indicated a progressive unfolding of the protein chain with increasing concentration of the reagent. The unfolding involved several intermediates, also it was irreversible. Treatment of rCAL with GdnHCl resulted in unfolding followed by dissociation of the dimer. The protein structure was drastically affected within one hour in acidic as well as alkaline buffers.

## 4.1. INTRODUCTION

Specific biological functions of proteins emerge directly from their unique and highly individualistic three-dimensional structure, which is attained in a very short time after their synthesis. The three dimensional structure assumed by a protein can, in general, be considered to be thermodynamically the most stable conformation adopted by the polypeptide chain. This stable structure of a protein is held together by non-covalent interactions *viz.* hydrogen bonds, ionic interactions, hydrophobic interactions, van der Waals' forces and covalently by disulfide linkages. Conditions, which disturb these stabilizing forces, affect the native conformation of the protein by changing its physical properties and biological activity. A polypeptide can also adopt a less rigid or more flexible conformation, different from its functional native form, in response to changes in its environment. Thus, proteins can be considered only marginally stable because of the functional requirement for their inherent dynamic state and due to delicate balancing of interactions involved in stabilizing or destabilizing particular structure [1-4].

Determination of the conformational stability of a protein is critical for understanding the physical interactions that stabilize the protein [5]. The folding pathway of proteins, sometimes, involves intermediate states and hence it is important to study such partially folded conformations, to understand the principles governing protein folding/unfolding (6, 7). Such stable intermediates have been identified and characterized for several proteins using modern sensitive techniques *viz.* spectroscopy and NMR (8). By recording changes in intrinsic tryptophan fluorescence and the secondary and tertiary structural features of protein in response to tailored changes in

surroundings, one can establish the presence of interesting structural intermediates relevant to structure-function relationship of the protein.

It has been known for many years that proteins can be unfolded in aqueous solution by high concentrations of certain reagents such as guanidine hydrochloride or urea or at higher temperatures. Denaturation with these chemicals is one of the primary ways of measuring the conformational stability of the proteins [9]. The roles and magnitudes of specific electrostatic interactions in determining the stability of a protein can be studied by measuring the dependence of the stability on pH. pH is known to influence the stability of a protein by altering the net charge of the protein. The aim of this study is to understand the conformational stability of rCAL as a function of temperature, pH and chemical denaturants using intrinsic fluorescence and circular dichroism.

The unfolding of oligomeric proteins requires the disruption of additional molecular interactions over those of monomeric proteins, since both inter- as well as intra-subunit interactions would make distinct and differential contributions to their overall structure and stability. In this regard, legume lectins serve as attractive models for studying the folding process of oligomeric proteins [10]. Several plant lectins have been characterized for their unfolding behaviour in presence of denaturing agents like temperature, urea or guanidine hydrochloride [see Table 4.1 for details].

**Table 4.1 Representative examples of plant lectins studied for unfolding**

<b>Sr. No.</b>	<b>Name and source of lectin and its refs</b>	<b>Main observations</b>
1.	Pea lectin[11]	<ul style="list-style-type: none"> <li>• Urea-induced unfolding is two-state process.</li> <li>• Protein undergoes cold denaturation that becomes distinct above 3.8 M urea.</li> <li>• Refolding upon thermal- and urea-induced unfolding is reversible.</li> </ul>
2.	Pea lectin[12]	<ul style="list-style-type: none"> <li>• Secondary structure resistant to conformational changes up to pH 2.5.</li> <li>• Addition of 80 % TFE retained the residual <math>\beta</math>-sheet structure, but with a loss in tertiary structure.</li> </ul>

		<ul style="list-style-type: none"> <li>• Transition from <math>\beta</math>-sheet structure to <math>\alpha</math>-helical structure began in presence of 12 % HFIP, completed in 30 %.</li> <li>• Occurrence of a common intermediate in the unfolding induced by these two fluoroalcohols (differing in their mode of stabilization of structure).</li> <li>• TFE was found not to induce <math>\alpha</math>-helical structure, but HFIP induced a structure rich in <math>\alpha</math>-helical contacts</li> </ul>
3.	<b>Pea lectin (PSL) [13]</b>	<ul style="list-style-type: none"> <li>• Dimeric PSL formed of two chains – a long <math>\beta</math>-chain and a short <math>\alpha</math>-chain.</li> <li>• GdnHCl-induced unfolding opens up the fragments to reveal a <math>\beta</math>-fragment as intermediate, with molten globule like characteristics.</li> <li>• Larger fragment of PSL may behave as monomeric or single domain protein undergoing unfolding through intermediate(s).</li> </ul>
4.	<b>Winged bean acidic lectin from <i>Psophocarpus tetragonolobus</i> [14]</b>	<ul style="list-style-type: none"> <li>• Thermal denaturation leads to dissociation of dimer into its monomers at denaturation temperature.</li> <li>• Glycosylation leads to less stable (lower denaturation temperature) structure compared to other legume lectins.</li> <li>• Inter sub unit interface is less extensive compared to glycosylated lectins like Con A, pea and lentil lectins, hence less thermally stable.</li> </ul>
5.	<b>Peanut agglutinin (tetrameric) [15]</b>	<ul style="list-style-type: none"> <li>• Thermal unfolding is reversible, with three states.</li> <li>• Tetramers unfold to folded monomers, which then unfold into monomers (folded monomers</li> </ul>

		form the intermediates).
6.	<b>Peanut agglutinin [16]</b>	<ul style="list-style-type: none"> <li>• GdnHCl-induced denaturation is completely reversible.</li> <li>• Biphasic profile, with a non-two-state unfolding process.</li> <li>• Intermediate formed during unfolding has reduced tertiary structure, hence is a molten globule and is more compact.</li> <li>• Intermediate retains its carbohydrate activity considerably.</li> </ul>
7.	<b>Peanut agglutinin [17]</b>	<ul style="list-style-type: none"> <li>• At pH 2.5, 15 % TFE induces a molten globule like structure.</li> <li>• Increasing TFE concentration leads to increase in <math>\alpha</math>-helical content and compactness of protein.</li> <li>• Compact PNA at higher TFE concentration is structurally different than the native protein.</li> <li>• TFE at neutral pH does not induce molten globule like state.</li> </ul>
8.	<b>Peanut agglutinin (PNA) [18]</b>	<ul style="list-style-type: none"> <li>• Urea-induced denaturation of tetrameric PNA is three-state, involving molten globule as the intermediate.</li> <li>• Refolding involves rapid appearance of intermediate.</li> <li>• Tetramerization contributes significantly to stabilization of oligomers.</li> </ul>
9.	<b>Leucoagglutinin from <i>Phaseolus vulgaris</i> [19]</b>	<ul style="list-style-type: none"> <li>• Protein is dimeric at pH 2.5 and tetrameric at pH 7.2.</li> <li>• Thermal denaturation at neutral pH is irreversible.</li> <li>• Extremely thermostable protein, with transition temperature of around 82 °C and above 100 °C</li> </ul>



		<p>for pH 2.5 and pH 7.2 respectively.</p> <ul style="list-style-type: none"> <li>• Protein remains in compact-folded state, even at pH 2.3, denaturation begins at 60 °C.</li> </ul>
10.	<b>Leucoagglutinin (PHA-L) from <i>Phaseolus vulgaris</i> [20]</b>	<ul style="list-style-type: none"> <li>• Homotetrameric protein refolds at pH 2.5 with the formation of a dimeric intermediate.</li> <li>• Denaturation kinetics at pH 2.5 followed a single exponential decay pattern; rate of denaturation independent of protein concentration.</li> <li>• Renaturation kinetics was dependent on protein concentration.</li> </ul>
11.	<b><i>Artocarpus hirsuta</i> lectin [21]</b>	<ul style="list-style-type: none"> <li>• Unfolding partially reversible in presence of GdnHCl</li> <li>• Protein dissociates reversibly into partially unfolded dimer and then irreversibly into unfolded inactive monomer in presence of GdnHCl.</li> <li>• Thermal denaturation irreversible, lectin loses hemagglutinating activity rapidly above 45 °C.</li> <li>• Insoluble aggregates during thermal denaturation leads to irreversible denaturation.</li> </ul>
12.	<b>Concanavalin A [22]</b>	<ul style="list-style-type: none"> <li>• Urea or GdnHCl-induced denaturation exhibits three-state mechanism, involving a structured monomer between native tetrameric and unfolded monomeric states.</li> <li>• Structural stability maintained by sub unit association.</li> </ul>
13.	<b>Concanavalin A [23]</b>	<ul style="list-style-type: none"> <li>• Addition of 30 % PEG retains the secondary structure compared to that in 70 % PEG; tertiary structure also retained.</li> <li>• Tryptophan environment changed in presence of PEG.</li> <li>• Compact ‘molten globule’ formed in presence</li> </ul>

		<p>of 30 % PEG.</p> <ul style="list-style-type: none"> <li>• Considerable amount of carbohydrate binding and specificity retained in 30 % PEG.</li> <li>• GdnHCl-induced denaturation was a single-step, two-state transition</li> </ul>
14.	<b>Lentil lectin [24]</b>	<ul style="list-style-type: none"> <li>• Structure lost at pH 0.8 and more so in presence of GdnHCl compared to native protein at pH 7.0.</li> <li>• Presence of intermediate at low pH.</li> <li>• Acid-unfolded structure stabilized by addition of fluoroalcohols TFE and HFIP by inducing <math>\alpha</math>-helical contacts.</li> <li>• This non-native structure (generated by TFE and HFIP) regained more activity than the native protein when treated with bromelain.</li> </ul>
15.	<b><i>Dolichos lablab</i> seed lectin [25]</b>	<ul style="list-style-type: none"> <li>• Thermally stable protein; unfolds at 75 °C.</li> <li>• Ligand (methyl-<math>\beta</math>-D-galactose) binding leads to stabilization of secondary structure.</li> </ul>
16.	<b>Soybean agglutinin (SBA) [26]</b>	<ul style="list-style-type: none"> <li>• Refolded SBA after urea denaturation exhibits similar quaternary structure as that of native lectin.</li> <li>• Dimerization of SBA dimers occurs faster than the dimerization of SBA monomers.</li> </ul>
17.	<b><i>Erythrina indica</i> lectin [27]</b>	<ul style="list-style-type: none"> <li>• Monophasic urea-induced unfolding transition from native dimer to unfolded monomers.</li> <li>• Rate of unfolding increases several fold with increase in temperature.</li> </ul>
18.	<b><i>Moringa oleifera</i> lectin [28]</b>	<ul style="list-style-type: none"> <li>• Highly thermostable lectin.</li> <li>• Hemagglutinating activity and secondary structure not affected at extreme temperature and pH.</li> </ul>

		<ul style="list-style-type: none"> <li>• Secondary structure drastically affected in presence of dithiothreitol at and above pH 7.0, indicating role of disulphide linkages in maintaining active conformation of lectin.</li> </ul>
19.	<i>Trichosanthes dioica</i> seed lectin [29]	<ul style="list-style-type: none"> <li>• Thermal unfolding is three-state.</li> <li>• Ligand binding stabilizes native conformation of protein.</li> <li>• Protein more stable at acidic pH.</li> <li>• Lectin structure stable over wide range of pH.</li> <li>• GdnHCl-mediated unfolding is three-state, involving an intermediate.</li> </ul>
20.	<i>Clitoria ternatea</i> seed agglutinin [30]	<ul style="list-style-type: none"> <li>• Compact molten-globule (MG) conformation at pH 1.2.</li> <li>• Two-step non cooperative thermal denaturation of MG compared to cooperative single-step transition of native protein.</li> <li>• 72% carbohydrate binding activity retained by MG.</li> </ul>
21.	<i>Sauromatum guttatum</i> and <i>Arisaema tortuosum</i> lectins [31]	<ul style="list-style-type: none"> <li>• Drastic loss in secondary structure of both the lectins at pH 2 and below.</li> <li>• Detection of a compact structure between pH 10 – 12</li> </ul>
22.	<i>Arisaema curvatum</i> lectin [32]	<ul style="list-style-type: none"> <li>• In 0.25 M GdnHCl protein exhibits pronounced secondary structure, hemagglutinating activity and altered tryptophan environment.</li> <li>• Detection of acid-induced molten globule with higher thermostability at pH 3.</li> <li>• Hemagglutinating activity retained even at 95 °C.</li> </ul>

## 4.2. MATERIALS

Cultivar BDN 9-3 of chick pea seeds (*Cicer arietinum* L.) was obtained from Badnapur Agricultural University, Jalna, India. Acrylamide, cesium chloride, potassium iodide and guanidine hydrochloride and all other chemicals and reagents as well as buffers used for unfolding studies were from Sigma-Aldrich, USA. The buffers used at different pH were Glycine-HCl (pH 2 – 3), sodium acetate (pH 4 – 5), sodium phosphate (pH 6 – 7), Tris-HCl (pH 8 – 9) and glycine-NaOH (pH 10 – 12). Guanidine hydrochloride was freshly prepared at pH 7.2 and filtered through a 0.22 micron syringe filter before use.

## 4.3. METHODS

### 4.3.1. Protein Preparation

The recombinant lectin from *Cicer arietinum* seeds purified as mentioned in Chapter 2. This purified lectin, rCAL, was the subject of the present study. Protein concentrations were determined according to the method of Lowry *et al* [33] using bovine serum albumin (BSA) as the standard.

### 4.3.2. Spectroscopic measurements

**a) Fluorescence studies:** Steady-state intrinsic fluorescence measurements were performed on Perkin Elmer LS 50B luminescence spectrometer at 25 °C. Protein solution of 150 µg/ml was excited at 295 nm (1.0 cm cell path length) and emission was recorded from 310 to 400 nm. Slit widths of 7 nm each were set for excitation and emission monochromators and the spectra were recorded at 100 nm/min. To eliminate the background emission, the signal produced by either buffer solution or buffer containing the desired quantity of GdnHCl or different buffers was subtracted.

Rayleigh (light) scattering was measured at 400 nm with the excitation and emission slit widths set at 2.5 and 10 nm, respectively, to follow the aggregation of the protein at different temperatures.

**b) Hydrophobic dye binding:** The intermediate states during thermal unfolding and GdnHCl-mediated unfolding and refolding of rCAL were analyzed by binding with the hydrophobic dye (8-anilino-1-naphthalenesulphonate, ANS). The final ANS concentration used was 50 µM, excitation wavelength was 375 nm and total fluorescence emission was monitored between 400 and 550 nm. Reference spectrum of ANS in either buffer or buffer containing desired concentration of GdnHCl was subtracted from the spectrum of the sample.

c) **CD measurements:** The CD spectra of rCAL were recorded on a J-715 Spectropolarimeter (Jasco, Tokyo, Japan) at 25 °C in a quartz cuvette. Each CD spectrum was accumulated from five scans at 100 nm/min with a 1 nm slit width and a time constant of 1 s for a nominal resolution of 0.5 nm. Far UV CD spectra of rCAL (200 µg/ml) were collected in the wavelength range of 190 to 250 nm (for thermal unfolding) and 210 to 250 nm (for GdnHCl-mediated unfolding) using a cell of path length 0.1 cm for monitoring the secondary structure. All spectra were corrected for buffer contributions and observed values were converted to mean residue ellipticity (MRE) in deg cm<sup>2</sup> dmol<sup>-1</sup> defined as

$$\text{MRE} = M \theta_{\lambda} / 10 d c r$$

where M is the molecular weight of the protein,  $\theta_{\lambda}$  is CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ml and r is the average number of amino acid residues in the protein. Secondary structure elements were calculated using the CDPro software [34]. Near UV CD spectra of rCAL (600 µg/ml) were collected in the wavelength range of 250 to 300 nm using a cell of path length 1.0 cm.

#### 4.3.3. Thermal denaturation

Unfolding as a function of temperature was monitored by fluorescence and circular dichroism. Protein samples were incubated at temperatures ranging from 25 to 80 °C for ten minutes each. Protein concentration used was 150 µg/ml for fluorescence studies and 200 µg/ml for CD studies.

##### 4.3.3.1. Reversibility of thermal unfolding

To check reversibility of thermal unfolding of rCAL, first the protein was gradually heated from 25 °C up to 50 °C and then allowed to cool to 25 °C. The fluorescence scans were recorded at each temperature.

#### 4.3.4. GdnHCl induced Unfolding and Refolding.

Unfolding as a function of GdnHCl was monitored by fluorescence and circular dichroism. Protein samples (150 µg/ml) were incubated in 0–6 M GdnHCl solution at pH 7.2 for 6 h to attain equilibrium. For renaturation, a concentrated sample (600 µg/ml) was denatured in 6 M GdnHCl and then diluted ten times with refolding buffers of 0 to 6 M GdnHCl concentrations. The reaction was maintained for 3 hours before recording the spectra. Respective blanks of GdnHCl were subtracted before analysing the data. Protein concentration of 200 µg/ml was used to study unfolding by CD.

#### 4.3.4.1. Reversibility of GdnHCl-induced unfolding:

To check the reversibility of GdnHCl-induced unfolding, in a second set of experiments, separate aliquots of rCAL were denatured in varying concentrations of GdnHCl, from 0 to 6 M (for 2 hours) and the samples were then diluted ten times. The samples were allowed to renature for 1 hour before recording their fluorescence intensity.

#### 4.3.5. Decomposition of fluorescence spectra

The decomposition of trp fluorescence spectra was carried out using PFAST program (<http://pfast.phys.uri.edu/pfast/>) based on the SIMS and PHREQ algorithm as described in [35].

#### 4.3.6. Parameter A analysis

Parameter A, the ratio of the intrinsic fluorescence intensity at 320 nm to that at 365 nm ( $I_{320}/I_{365}$ ), is an attribute of fluorescence spectral shape and position [36]. This analysis was carried out to detect conformational changes of rCAL during GdnHCl-induced unfolding and refolding.

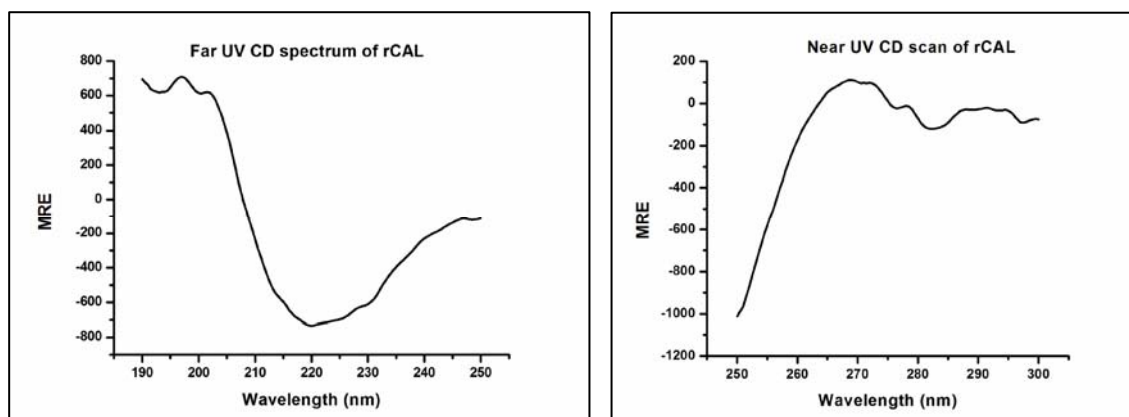
#### 4.3.7. Size exclusion chromatography

Dissociation of oligomeric rCAL in presence of GdnHCl was monitored using WATERS HPLC unit and WATERS gel filtration Protein Pak™ 300SW (7.5 x 300 mm column). 25 mM Phosphate buffer (with 0.15M NaCl), pH 7.2 containing respective concentrations of GdnHCl served as the mobile phase. 100  $\mu$ l protein sample (600  $\mu$ g/ml) incubated in 0 to 4M GdnHCl for four hours was injected into the column. Flow rate was maintained at 0.5 ml/min and the elution profile was monitored at 280 nm.

### 4.4. RESULTS AND DISCUSSION

#### 4.4.1. Secondary and tertiary structure of rCAL

The far UV CD spectrum showed a minimum around 218-220nm, which is characteristic of  $\beta$ -sheet containing proteins (Figure 4.1 (A)). The spectrum was analyzed using the CONTINLL program of the CDPro software [34] to calculate the secondary structure elements. The lectin showed the presence of 3.7% helix, 41.1% sheets, 21.3% turns and 33.9% unordered structure with an NRMSD value of 0.129. The near UV CD spectrum (Figure 4.1 (B)) showed a negative minimum at 284nm and maxima at 255nm and 295nm, indicating an ordered tertiary structure.

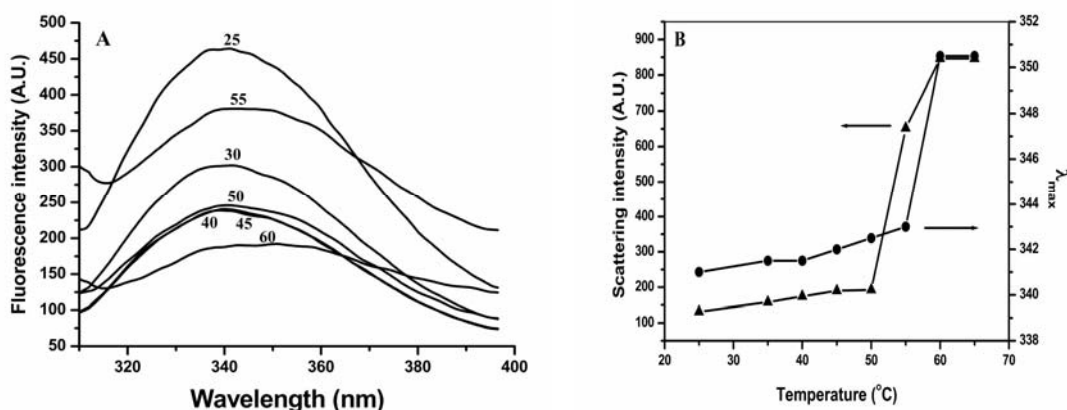


**Figure 4.1: Secondary structure of rCAL.** (A) Far UV CD spectrum of rCAL and (B) Near UV CD spectrum of rCAL. Protein concentration used was 200  $\mu\text{g/ml}$  for far UV and 600  $\mu\text{g/ml}$  for near UV CD.

#### 4.4.2. Thermal denaturation of rCAL

The fluorescence intensity decreased upon increasing temperature from 25 to 45  $^{\circ}\text{C}$ ; a marginal increase was obtained at 50  $^{\circ}\text{C}$ , and a sharp increment in the intensity could be detected at 55  $^{\circ}\text{C}$  (Figure 4.2 (A)).

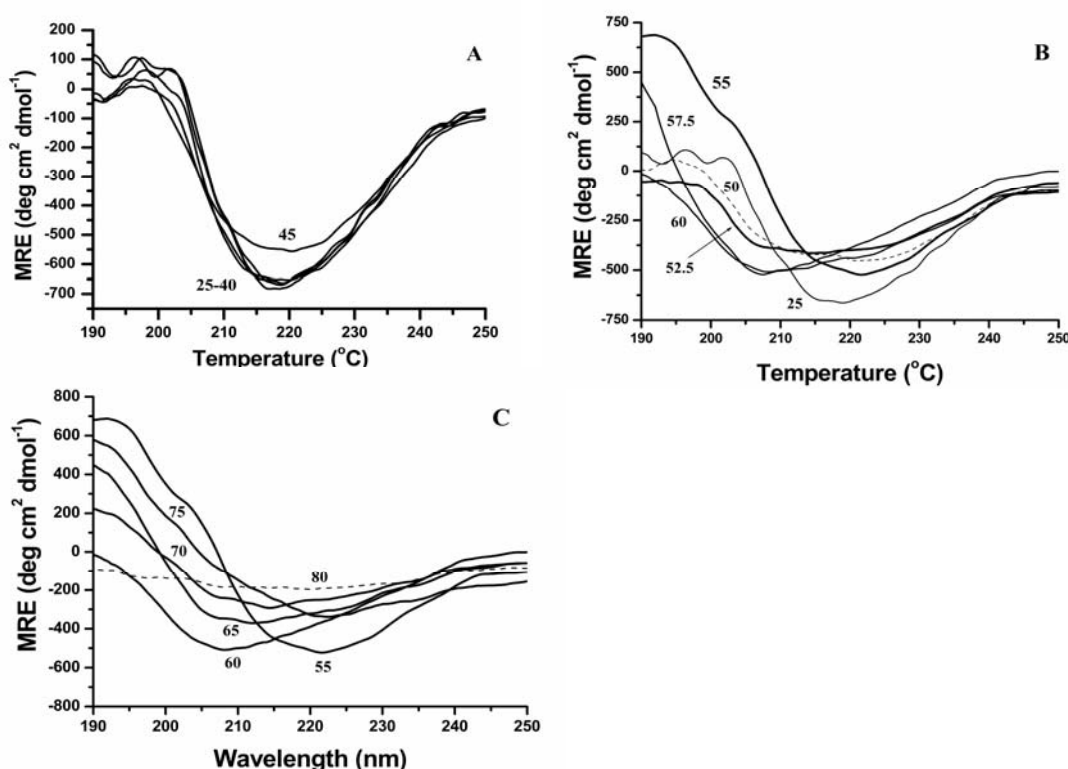
A slight red shift in the  $\lambda_{\text{max}}$  of intrinsic fluorescence was observed for rCAL upon increasing the temperature from 25 to 55  $^{\circ}\text{C}$  (Fig. 4.2 (B)). The light scattering intensity, measured simultaneously, increased suddenly at 55  $^{\circ}\text{C}$  and reached its maximum at 60  $^{\circ}\text{C}$ . This indicated aggregation of the protein involving major structural change.



**Figure 4.2: Thermal unfolding of rCAL.** (A) Fluorescence intensity of rCAL with increasing temperature. (B) Plot of  $\lambda_{\text{max}}$  (filled circles) and light scattering (filled triangles) of the chick pea lectin (150  $\mu\text{g/ml}$ ) at different temperatures.

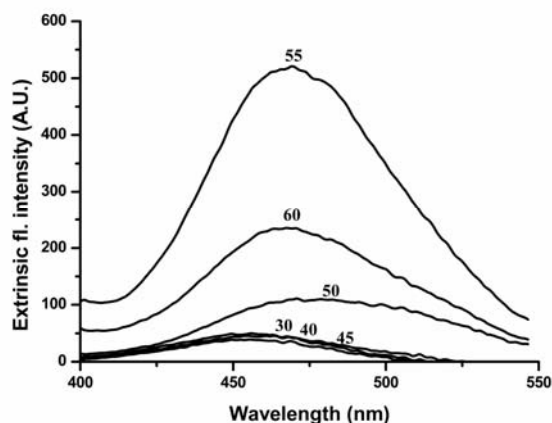
The minimum in the far UV CD spectrum of rCAL remained stable at 218-220 nm from 25 to 45 °C (Figure 4.3(A)). A distinct alteration in the secondary structure took place at 50, 52.5 and 55 °C; at 50 and 52.5 °C, the minima shifted to 215 nm (Figure 4.3(B)). Interestingly, at 55 °C, the minimum was again centred at 220-222 nm and a marked increase in positive ellipticity within 190-200 nm was observed as compared to the native protein. At 55 °C, the protein began to aggregate and this may have led to the formation of intermolecular  $\beta$ -sheets as reported by Uversky *et al* [37]. Thus, the protein shows a flexible thermo-labile nature with subtle changes occurring around 55 °C. The structure showed flexibility even between 60 to 75 °C and was completely lost at 80 °C (Figure 4.3 (C)).

An increase in the fluorescence intensity of the hydrophobic dye, ANS, accompanied by a blue shift to 475 nm, was observed at 55 °C (Figure 4.4). Partial unfolding of the protein at 55 °C might have resulted in exposure of hydrophobic sites and simultaneous formation of new intermolecular interactions. This might contribute to formation of aggregates via non-covalent interactions, as described by Vetri *et al* [38] for BSA.



**Figure 4.3: Thermal unfolding of rCAL.** Far UV Mean Residue Ellipticity (MRE) spectra for temperatures (A) 25 to 45 °C, (B) 50 to 60 °C and (C) 55 to 80 °C. 200 µg/ml protein was used.

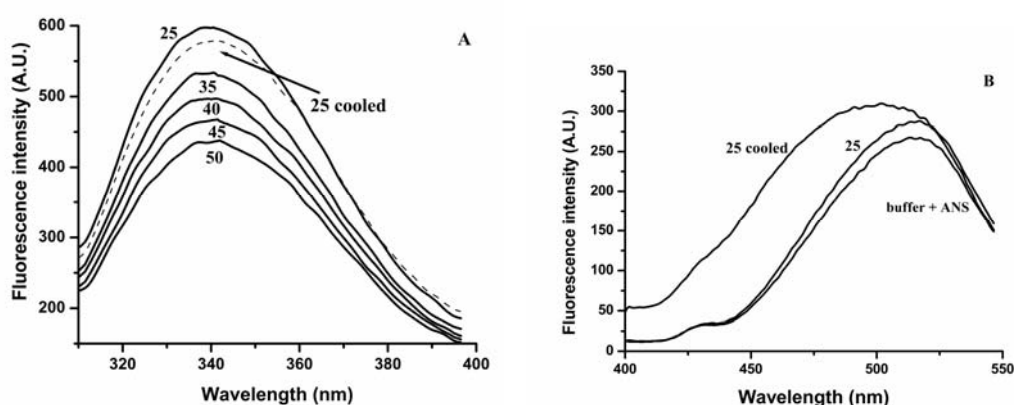




**Figure 4.4: ANS binding of the lectin during thermal unfolding of rCAL.** The protein sample was incubated at the indicated temperature separately and ANS was added. All the scans were recorded after incubation for ten minutes at indicated temperature. 150  $\mu\text{g/ml}$  protein was used.

#### 4.4.2.1. Reversibility of thermal unfolding

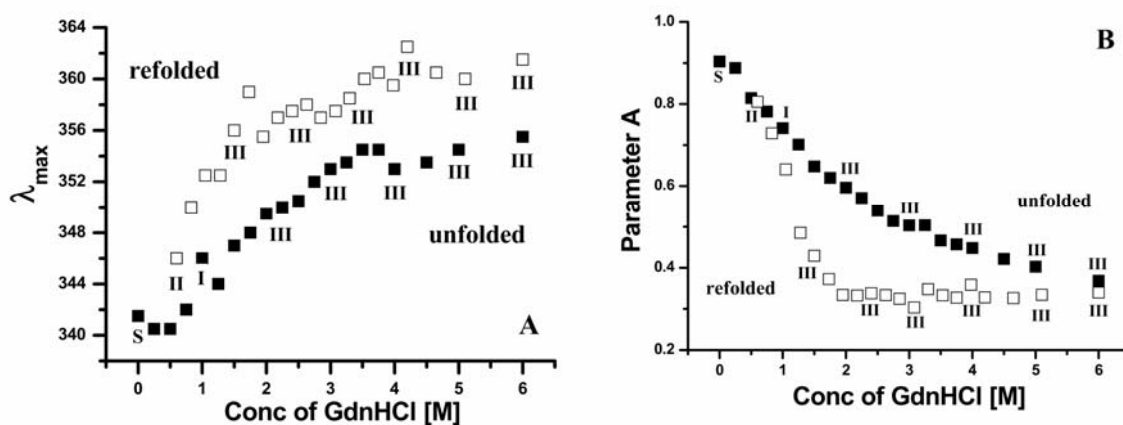
Since rCAL began to aggregate at 55  $^{\circ}\text{C}$  and above, it was interesting to check if the thermal denaturation was reversible when the protein was heated only up to 50  $^{\circ}\text{C}$  and then cooled back to 25  $^{\circ}\text{C}$ . Here, the fluorescence intensity and  $\lambda_{\text{max}}$  were almost regained after cooling to 25  $^{\circ}\text{C}$  ( $\lambda_{\text{max}}$  of renatured sample remained same as that of native protein) (Figure 4.5). The renatured sample showed marginal ANS binding, indicating slight exposure of hydrophobic residues upon cooling from 50 to 25  $^{\circ}\text{C}$ .



**Figure 4.5: Reversibility of thermal unfolding of rCAL.** (A) Fluorescence intensity vs temperature. (B) ANS binding of renatured rCAL. 150  $\mu\text{g/ml}$  protein was used.

#### 4.4.3. Guanidine hydrochloride-induced unfolding and refolding of rCAL

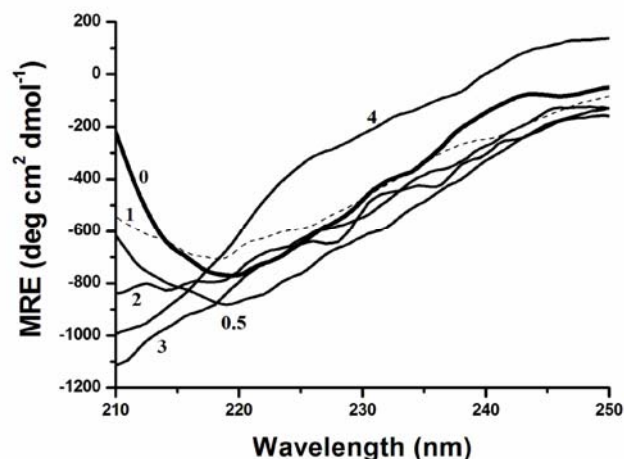
A gradual unfolding of the lectin was observed upon addition of increasing concentrations of GdnHCl, as indicated by a red shift in the  $\lambda_{\max}$  (Figure 4.6(A)) from 341 to 356 nm in the vicinity of 6 M denaturant. Parameter A analysis ( $I_{320}/I_{365}$ ) indicated that unfolding of rCAL was multi step, involving several intermediates (Figure 4.6 (B)). Trp conformers of class S and class I was predominant up to 1 M GdnHCl, indicating a hydrophobic environment for the residue. Treatment with 2 M and higher concentration of GdnHCl caused significant unfolding of the protein, with appearance of 100 % of class III conformer, representing complete exposure of trp to solvent.



**Figure 4.6: GdnHCl-mediated unfolding and refolding of rCAL at 25 °C.** (A) The graph for  $\lambda_{\max}$  vs GdnHCl concentrations. (B) Parameter A analysis for the unfolding (6 h) (150  $\mu\text{g/ml}$ ) and refolding (3 h) (600  $\mu\text{g/ml}$ ), showing the trp conformer classes.

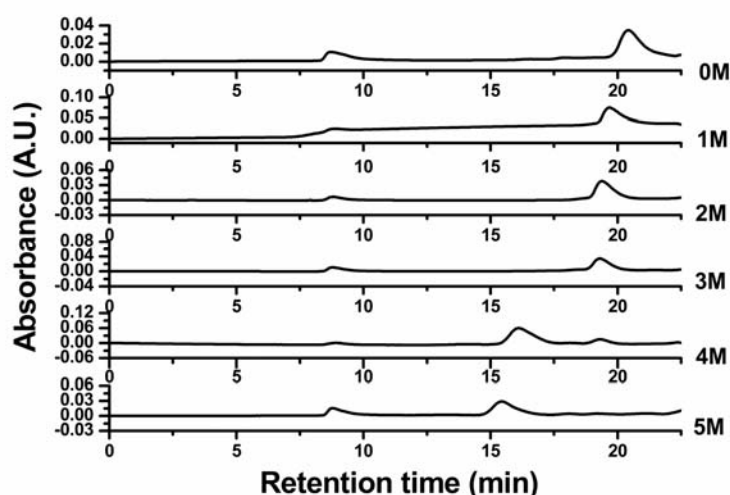
Unfolding of rCAL was observed to be irreversible, the  $\lambda_{\max}$  showed further red shift after providing renaturing conditions to the samples (Figure 4.6 (A)). Formation of more disordered structure takes place when allowed to renature. Class III trp conformer was detected by decomposition analysis of all the above samples. The sample allowed to renature in low concentration of GdnHCl (0.6 M) appeared to be partially reversible as indicated by the appearance of class II conformer ( $\lambda_{\max} = 346$  nm).

A gradual but significant loss in the secondary structure as shown by altered far UV CD spectra (decrease in the positive ellipticity in the range of 190-205 nm) was observed in samples treated with GdnHCl after 6 hours (Figure 4.7).



**Figure 4.7** GdnHCl-mediated unfolding and refolding of rCAL at 25 °C. Far UV MRE spectra of rCAL (200 µg/ml) incubated in 0 to 4 M GdnHCl for 6 hours.

To investigate the dissociation profile of dimeric rCAL during denaturation, HPLC size exclusion chromatography (HPSEC) was performed in the vicinity of 0 to 5M GdnHCl (Figure 4.8). The native rCAL of molecular mass of 53 kDa eluted as a single peak centred at 20.38 min. The retention time decreased slightly with increasing concentrations from 1 to 3M GdnHCl (19.67 min, 19.38 and 19.28 min respectively). However, in presence of 4 M GdnHCl, the above peak (19.28 min) reduced drastically and a new peak was detected at 16.08 min. Further, a single peak appearing at 15.41 min was detected in the presence of 5M GdnHCl. The minor peak corresponding to the retention time of 8.8 min was that of void volume of the column.



**Figure 4.8:** GdnHCl-mediated unfolding and refolding of rCAL at 25 °C. HPSEC profile of rCAL (600 µg/ml) incubated in 0 to 5 M GdnHCl for 4 hours.

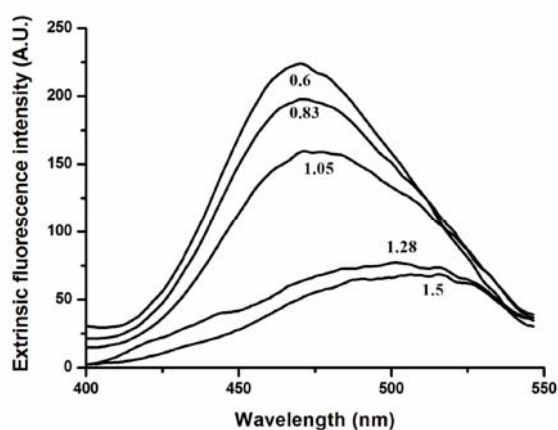
The native dimeric protein began to unfold in the presence of 1M GdnHCl (Figure 4.7 (A)). If dissociation had been the first step, structured monomer would have eluted with higher retention time than that of the native protein. The unfolded (but not yet dissociated) dimer seemed to populate in the presence of 2 and 3 M GdnHCl. Further treatment with 4 and 5M GdnHCl could lead to dissociation of the unfolded dimer to more unstructured monomer, with concomitant loss in the structure and significant reduction in the elution volume. This could be possible if the unfolded monomer is adopting an extensively random coiled conformation compared to the native protein, as has been observed for the Con A lectin [17] and *Erythrina indica* lectin [26].

Taking into account the data from fluorescence, CD and HPSEC (summarized in Table 4.2), we propose the following scheme to delineate the GdnHCl-mediated unfolding of rCAL



where N indicates native rCAL and U denotes unfolded dimer, that dissociates into two unfolded monomers (U).

No ANS binding was observed during the unfolding process of rCAL (data not shown). However, under renaturing conditions, protein in 0.6 M to 1.05 M GdnHCl did bind ANS (Figure 4.9), indicating the exposure of hydrophobic residues in an intermediate species formed during refolding in these low concentrations of dilution buffers.



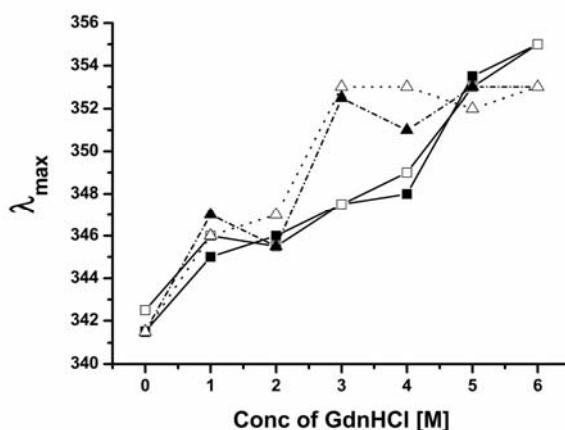
**Figure 4.9: ANS binding of rCAL (600 µg/ml) after 3 hours of renaturation.** The far UV CD spectra of the lectin after renaturing in different concentrations of GdnHCl could

not be recorded, since the lectin has an inherent tendency to aggregate at higher concentrations (more than 700  $\mu\text{g/ml}$ ).

**Table 4.2 Comparison of GdnHCl-mediated unfolding of rCAL with different probes**

Probe	GdnHCl (M)					
	Native	1	2	3	4	5
<b>Intrinsic Fluorescence</b> $\lambda_{\text{max}}$ (nm)	342	346	350	353	354	355
<b>Far UV CD</b> $\theta$ (MRE <sub>210</sub> )	-221	-542	-840	-1115	-993	--
<b>HPSEC</b> Retention time (min)	20.38	19.67	19.38	19.28	16.08 (slight at 19.28)	15.41 (slight at 8.8)

#### 4.4.3.1. Reversibility of GdnHCl after unfolding up to 2M

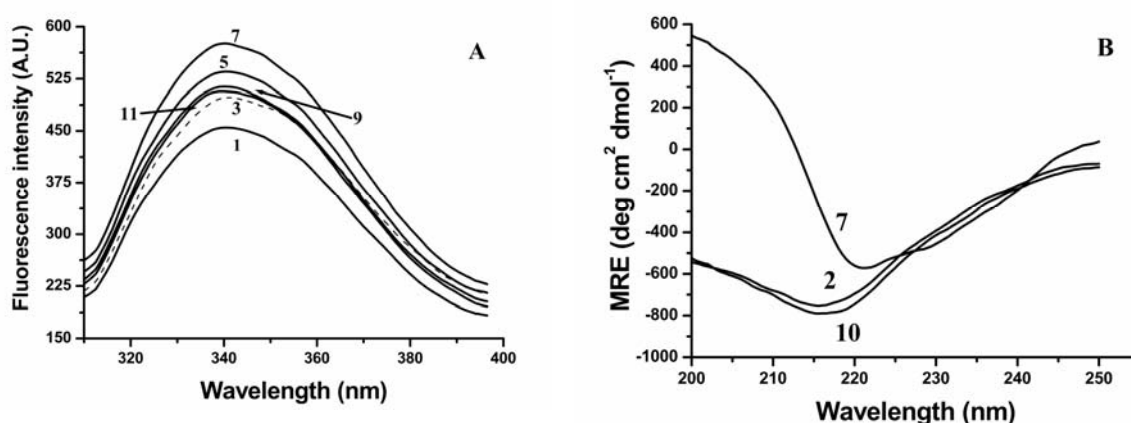


**Figure 4.10: Reversibility of rCAL upon denaturation with different concentrations of GdnHCl.** The first set of experiments is represented by samples unfolded in 6 M GdnHCl for 4 hours (unfolded 4 h, open squares) and then allowed to renature for 4 hours in dilution buffers (refolded 4 hours, open triangles). The second set of experiments involved unfolding of samples in varying concentrations of GdnHCl for 2 hours (unfolded 2h, closed squares) and then allowed to renature for 1 hour (refolded 1h, closed triangles).

To check for reversibility of GdnHCl-induced unfolding, separate aliquots of rCAL were denatured in varying concentrations of GdnHCl, from 0 to 6 M (for 2 hours) and the samples were then diluted ten times to check for renaturation. The denaturation was reversible only for samples treated with 2 M GdnHCl (Figure 4.10). Above 2 M, the denaturation was irreversible as obtained before. As seen in Figure 4.7 earlier, the 2 M-denatured sample had begun to lose its secondary structure and it showed the presence of unfolded monomers as seen from Figure 4.8. Hence, this could be the reason why above 2 M denaturant concentration, the protein shows irreversible denaturation and is unable to achieve the conformation of the native protein.

#### 4.4.4. Effect of pH on rCAL

A decrease in fluorescence intensity was observed after incubation of rCAL in buffers of different pH; maximum decrease was obtained on the acidic side.



**Figure 4.11: Effect of pH on rCAL.** (A) Fluorescence intensity at different pH. (B) Secondary structure of rCAL after incubation in buffers of pH 2, 7 and 10 for one hour. 200  $\mu\text{g/ml}$  protein was used for fluorescence and 600  $\mu\text{g/ml}$  protein was used for CD.

No ANS binding was observed for samples treated with different buffers of different pH. Hence, indicating no exposure of hydrophobic residues (data not shown). The secondary structure of rCAL was lost after incubation in acidic (pH 2) and alkaline (pH 10) even after one hour of incubation (Figure 4.11 (B)). This indicates that the protein structure is labile in acidic and alkaline pH.

The process of aggregation results from conformational changes of the secondary structure as well as due to the partial unfolding of the tertiary structure. Recent studies by Militello and group [39] have shown that a partial opening of the

protein chain constitutes the first step of aggregation. This, in turn, leads to exposure of hydrophobic sites or free SH groups, hence resulting in aggregate formation [38]. In the case of rCAL also, partial opening of the chain at 50 °C caused exposure of hydrophobic residues (as seen by ANS binding at 50 and 55 °C), further leading to aggregation.

In favour of higher  $\beta$ -sheet structure content, the conformational changes in the secondary structure of many proteins may promote the formation of intermolecular bonds; this can also be one of the first steps for the formation of ordered aggregates in the form of amyloid fibrils. Among such proteins, albumin is well known for its ability to self assemble in the form of aggregates under particular conditions [38, 40]. rCAL is basically an albumin by nature, hence it shows an inherent tendency to aggregate. Another albumin from lentil (*Lens culinaris*) seeds, albumin-2 was also shown to unfold at 40 °C [41]. Further heating till 60 °C resulted in aggregation and precipitation of the protein. In contrast to rCAL, the *Dolichos lablab* lectin was found to be highly thermo-stable and unfolded at temperatures above 75 °C [23]. The Leucoagglutinin from *Phaseolus vulgaris* is also an extremely thermo-stable lectin, unfolding above 100 °C [15].

**REFERENCES**

1. Nicholson, E.M. and Scholtz, J.M. (1996), Conformational Stability of the *Escherichia coli* HPr Protein: Test of the Linear Extrapolation Method and a Thermodynamic Characterization of Cold Denaturation, *Biochemistry*, **35**, 11369–11378.
2. Agashe, V.R. and Udgaonkar, J.B. (1995), Thermodynamics of Denaturation of Barstar: Evidence for Cold Denaturation and Evaluation of the Interaction with Guanidine Hydrochloride, *Biochemistry*, **34**, 3286–3299.
3. Neet, K.E. and Timm, D.E. (1994), Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation, *Protein Science*, **3**, 2167–2174.
4. Johnson, C.R., Morin, P.E., Arrowsmith, C.H. and Freire, E. (1995), Thermodynamic analysis of the structural stability of the tetrameric oligomerization domain of p53 tumor suppressor, *Biochemistry*, **34**, 5309–5316.
5. Bowie, J.U. and Sauer, R.T. (1989), Equilibrium dissociation and unfolding of the arc repressor dimer, *Biochemistry*, **28**, 7139–7143.
6. Dill, K.A. and Shortle, D. (1991), Denatured states of proteins, *Annual Review of Biochemistry*, **60**, 795–825.
7. Dobson, C.M. (1992), Unfolded proteins, compact states and molten globules, *Current Opinion in Structural Biology*, **2**, 6–12.
8. Kuwajima, K. (1989), The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure, *Proteins*, **6**, 87–103.
9. Pace, C.N. (1986), Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods in Enzymology*, **131**, 266–280.



10. Alber, T. and Matthews, B.W. (1987), Structure and thermal stability of phage T4 lysozyme, *Methods in Enzymology*, **154**, 511–533.
11. Ahmad, N., Srinivas, V.R., Reddy, G.B. and Surolia, A. (1998), Thermodynamic Characterization of the Conformational Stability of the Homodimeric Protein, Pea Lectin, *Biochemistry*, **37**, 16765–16772.
12. Naseem, F. and Khan, R.H. (2005), Characterization of a common intermediate of pea lectin in the folding pathway induced by TFE and HFIP, *Biochimica et Biophysica Acta*, **1723**, 192–200.
13. Sen, D. and Mandal, D.K. (2011), Pea lectin unfolding reveals a unique molten globule fragment chain, *Biochimie*, **93**, 409–417.
14. Srinivas, V.R., Singha, N.C., Schwarz, F.P. and Surolia, A. (1998), Differential scanning calorimetric studies of the glycoprotein, winged bean acidic lectin, isolated from the seeds of *Psophocarpus tetragonolobus*, *FEBS Letters*, **425**, 57–60.
15. Reddy, G.B., Bharadwaj, S. and Surolia, A. (1999), Thermal Stability and Mode of Oligomerization of the Tetrameric Peanut Agglutinin: A Differential Scanning Calorimetry Study, *Biochemistry*, **38**, 4464–4470.
16. Reddy, G.B., Srinivas, V.R., Ahmad, N. and Surolia, A. (1999), Molten Globule-like State of Peanut Lectin Monomer Retains Its Carbohydrate Specificity: Implications In Protein Folding and Legume Lectine Oligomerization, *Journal of Biological Chemistry*, **274**, 4500–4503.
17. Dev, S., Khan, R.H. and Surolia, A. (2006), 2,2,2-Trifluoroethanol-Induced structural change of peanut agglutinin at different pH: A comparative account, *IUBMB Life*, **58**, 473–479.
18. Dev, S., Devi, N.K., Sinha, S. and Surolia, A. (2006), Thermodynamic Analysis of Three State Denaturation of Peanut Agglutinin, *IUBMB Life*, **58**, 549–555.

19. Biswas, S. and Kayastha, A.M. (2002), Thermal Stability of *Phaseolus vulgaris* Leucoagglutinin: a Differential Scanning Calorimetry Study, *Journal of Biochemistry and Molecular Biology*, **35**, 472–475.
20. Biswas, S. and Kayastha, A.M. (2004), Unfolding and refolding of Leucoagglutinin (PHA-L), an oligomeric lectin from kidney beans (*Phaseolus vulgaris*), *Biochimica et Biophysica Acta*, **1674**, 40–49.
21. Gaikwad, S.M., Gurjar, M.M. and Khan, M.I. (2002), *Artocarpus hirsuta* lectin Differential modes of chemical and thermal denaturation, *Eur. J. Biochem.*, **269**, 1413–1417.
22. Chatterjee, A., Mandal, D.K., (2003), Denaturant-induced equilibrium unfolding of *concanavalin A*. is expressed by a three-state mechanism and provides an estimate of its protein stability, *Biochimica et Biophysica Acta*, **1648**, 174–183.
23. Naeem, A., Khan, A. and Khan, R.H. (2005), Partially folded intermediate state of *concanavalin A*. retains its carbohydrate specificity, *Biochemical and Biophysical Research Communications*, **331**, 1284–1294
24. Naseem, F., Khan, R.H., (2004), Fluoroalcohol-induced stabilization of the a-helical intermediates of lentil lectin: implication for non-hierarchical lectin folding, *Archives of Biochemistry and Biophysics*, **431**, 215–223.
25. Sultan, N.A.M., Rao, R.N., Nadimpalli, S.K. and Swamy, M.J. (2006), Tryptophan environment, secondary structure and thermal unfolding of the galactose-specific seed lectin from *Dolichos lablab*: Fluorescence and circular dichroism spectroscopic studies, *Biochimica et Biophysica Acta*, **1760**, 1001–1008.
26. Chatterjee, M. and Mandal, D.K., (2003), Kinetic Analysis of Subunit Oligomerization of the Legume Lectin Soybean Agglutinin, *Biochemistry*, **42**, 12217–12222.

27. Ghosh, S. and Mandal, D.K. (2006), Kinetic stability plays a dominant role in the denaturant-induced unfolding of *Erythrina indica* lectin, *Biochimica et Biophysica Acta*, **1764**, 1021–1028.
28. Katre, U.V., Suresh, C.G., Khan, M.I. and Gaikwad, S.M. (2008), Structure–activity relationship of a hemagglutinin from *Moringa oleifera* seeds. *International Journal of Biological Macromolecules*, **42**, 203–207.
29. Kavitha, M. and Swamy, M.J. (2009), Spectroscopic and differential scanning calorimetric studies on the unfolding of *Trichosanthes dioica* seed lectin. Similar modes of thermal and chemical denaturation, *Glycoconjugate Journal*, **26**, 1075–1084.
30. Naeem, A., Saleemuddin, M. and Khan, R.H. (2009), Compact Acid-Induced State of *Clitoria ternatea* Agglutinin Retains Its Biological Activity, *Biochemistry (Moscow)*, **74**, 1088–1095.
31. 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.
32. Sharma, U., Gaikwad, S.M., Suresh, C.G., Dhuna, V., Khan, M.I., Singh, J. and Kamboj, S.S. (2011), Conformational Transitions in *Ariesaema curvatum* Lectin: Characterization of an Acid Induced Active Molten Globule. *Journal of Fluorescence*, **21**, 753–763.
33. Lowry, O.H., Rosenbrough, N.H., Farr, L.A. and Randall, R.J. (1951), Protein measurement with the folin reagent, *Journal of Biological Chemistry*, **193**, 265–275.
34. Sreerama, N., Venyaminov, S.Y. and Woody, R.W. (1999), Estimation of the number of helical and strand segments in proteins using circular dichroism spectroscopy, *Protein Science*, 370–380.

35. Burstein, E.A., Abornev, S.M. and Reshetnyak, Y.M. (2001), Decomposition of protein tryptophan fluorescence spectra into log-normal components. I. Decomposition algorithms, *Biophysical Journal*, **81**, 1699–1709.
36. Turoverov, K.K., Haitlina, S.Y. and Pinaev, G.P. (1976), Ultra-violet fluorescence of actin. Determination of actin content in actin preparation, *FEBS Letters*, **62**, 4–6.
37. Uvesrsky, V.N., Li, J. and Fink, A.L. (2001), Evidence for a partially folded intermediate in  $\alpha$ -synuclein fibril formation, *Journal of Biological Chemistry*, **276**, 10737–10744.
38. Vetri, V., Librizzi, F., Leone, M. and Militello, V. (2007), Thermal aggregation of bovine serum albumin at different pH: comparison with human serum albumin, *European Biophysics Journal*, **36**, 717–725.
39. Rondeau, P., Navarra, G., Cacciabauda, F., Leone, M., Bourdon, E. and Militello, V. (2010), Thermal aggregation of glycated bovine serum albumin, *Biochimica et Biophysica Acta*, **1804**, 789–798.
40. Rondeau, P., Armenta, S., Caillens, H., Chesne, S. and Bourdon, E. (2007), Assessment of temperature effects on beta-aggregation of native and glycated albumin by FTIR spectroscopy and PAGE: relations between structural changes and antioxidant properties, *Archives of Biochemistry and Biophysics*, **460**, 141–150.
41. Scarafoni, A., Gualtieri, E., Barbiroli, A., Carpen, A., Negri, A. and Duranti, M. (2011), Biochemical and Functional Characterization of an Albumin Protein Belonging to the Hemopexin Super family from *Lens culinaris* Seeds, *Journal of Agricultural and Food Chemistry*, **59**, 9637–9644.

**CHAPTER: 5**

---

**COMPUTATIONAL STUDIES ON THE  
RECOMBINANT LECTIN**

---

## SUMMARY

The crystal structure of LS-24 albumin from *Lathyrus sativus* (PDB ID: 3LP9; showing 85 % identity to rCAL sequence) was used to build a homology model for rCAL using the MODELLER 9v10 software. Validation of the model with standard parameters proved the authenticity of the model. The model generated was used for docking of hemin, spermine and thiamine using Autodock Vina 1.1.2. Binding energies were calculated for each ligand and the amino acids involved in the interaction of these ligands with the rCAL homology model were identified. In order to identify the microenvironment of the single tryptophan present in the lectin, the model was visualized in PyMOL ([www.pymol.org](http://www.pymol.org)). The trp was found to be relatively exposed to the solvent in the model. The trp was found to be surrounded by hydrophobic residues, and showed a distinct lack of positively charged residues and the presence of weak negative charge. This corroborated the result of solute quenching obtained experimentally. Aggregation-prone residues of the lectin were predicted using the BioLuminate tool of the Schrodinger suite, as well as online servers like the Aggrescan, AmylPred and the FoldAmyloid. Common amino acid residues were identified by all these softwares as being “hot-spots” for aggregation.

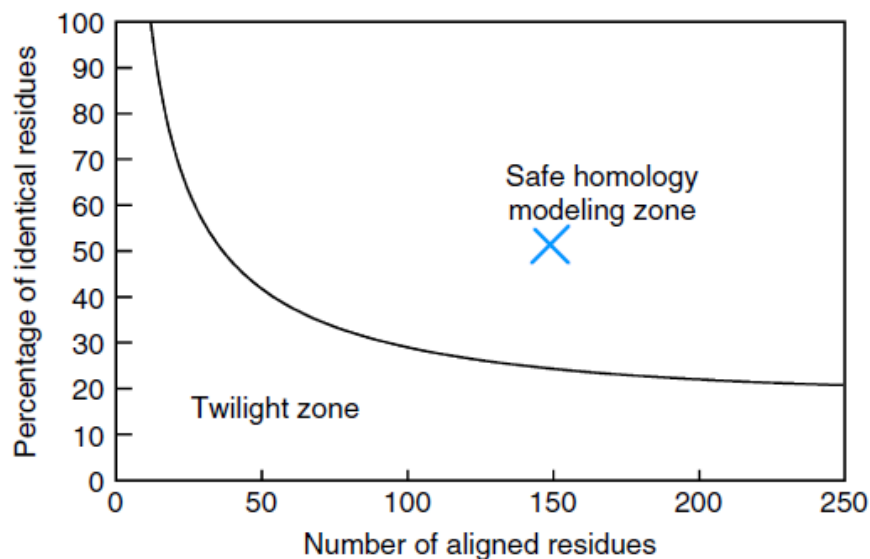
## 5.1. INTRODUCTION

Bioinformatics concerns the development of new software and tools for the analysis of genomic and molecular biological data. Computational methodologies have become a crucial component of many drug discovery programmes, from hit identification to lead optimization and beyond [1], and approaches such as ligand [2] or structure based virtual screening [3] techniques are widely used in many drug discovery efforts. In recent years, the drug discovery process has benefited significantly from computational studies on protein-carbohydrate interactions. Lectins have been successfully used in pharmaceutical applications as a system for drug targeting [4] and drug delivery [5]. For example, the *Bacillus anthracis* tetrasaccharide has been promoted as a promising lead for an anthrax vaccine [6].

The main aim of protein modelling is to predict a structure from its sequence with an accuracy that is comparable to the best results achieved experimentally [7]. This will help in the use of rapidly generated *in silico* protein models in all those cases where today only experimental structures provide a solid basis: structure-based drug

design, analysis of protein function and interactions, and rational design of proteins with increased stability or novel functions. In addition, protein modeling is the only way to obtain structural information if experimental techniques fail. Many proteins are simply too large for NMR analysis and cannot be crystallized for X-ray diffraction.

Homology modelling is the easiest approach to three-dimensional structure prediction. It is based on two major observations: (1) The structure of a protein is uniquely determined by its amino acid sequence [8]. Knowing the sequence should, at least in theory, suffice to obtain the structure. (2) During evolution, the structure is more stable and changes much slower than the associated sequence, so that similar sequences adopt practically identical structures and distantly related sequences still fold into similar structures [9, 10]. A precise limit for this rule is shown in Figure 5.1. As long as the length of two sequences and the percentage of identical residues fall in the region marked as “safe,” the two sequences are practically guaranteed to adopt a similar structure.



**Figure 5.1: The two zones of sequence alignments.** Two sequences are practically guaranteed to fold into the same structure if their length and percentage sequence identity fall into the region marked as “safe”. An example of two sequences with 150 amino acids, 50% of which are identical, is shown (grey cross) (Adapted from Kreiger *et al.*, 2003).

In practice, homology modeling is a multistep process that can be summarized in seven steps: (1) Template recognition and initial alignment, (2) Alignment correction, (3) Backbone generation, (4) Loop modeling, (5) Side-chain modeling, (6) Model optimization, and (7) Model validation.

Docking of small molecules to protein binding sites is an important methodology that was pioneered during the early 1980s [11], and remains a highly active area of research. The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site. Generally speaking, molecular docking comprises the process of generating a model of a complex based on the known 3D structures of its components, free or complexed with other species [12]. In terms of protein–ligand docking methods, the docking problem can be rationalized as the search for the precise ligand conformations and orientations (commonly referred to as ‘posing’) within a given targeted protein when the structure of the protein is known or can be estimated. The binding affinity prediction problem addresses the question of how well the ligands bind to the protein (scoring).

In the last decade, protein aggregation has moved beyond being a mostly ignored area of protein chemistry to become a key topic in medical sciences [13], due to the development of many debilitating human disorders including the amyloidoses and several neurodegenerative diseases [14] arising from the presence of insoluble deposits in human tissues. Data have begun to accumulate suggesting that the composition and the primary structure of a polypeptide determine to a large extent its propensity to aggregate and that small changes may have a huge impact on solubility. The ability to predict the aggregation propensity of a protein from its sequence would be of much value, for example, in the control of unwanted protein deposition events through specific sequence targeted therapeutics or in the discovery of more soluble variants of proteins of biotechnological interest.

Computational tools are increasingly being applied to solve the protein aggregation problem, providing insight into amyloid structures and aggregation mechanisms. The nature of protein aggregation (limited structural order, insolubility in water, and involvement of cell membrane) renders its experimental study extremely difficult. The difficulties have inspired new experimental innovations, and attracted computationally intensive approaches to complement experimental investigations.



Computational simulations, coupled with experimental investigations have increased our understanding of the complicated picture of the process of protein aggregation.

In this chapter, we describe the homology model construction of rCAL using the crystal structure of LS24 albumin from *Lathyrus sativus* as the template. The model was used for docking study of hemin, spermine and thiamine. This helped in the identification of the residues involved in binding of these ligands with the lectin. The model was also visualized in PyMOL to investigate the microenvironment of the single trp; the results correlated well with the solute quenching results obtained earlier (Chapter 3). Aggregation-prone residues were predicted using different servers, and common residues were identified.

## 5.2. METHODS

### 5.2.1. Protein Preparation

The recombinant *Cicer arietinum* lectin was purified to homogeneity from the induced cell lysates of *E. coli* as mentioned in Chapter 2. This purified lectin, rCAL, was the subject of the present study.

### 5.2.2. Sequence Alignment

Nucleotide sequence of the recombinant lectin from *Cicer arietinum* (rCAL) (NCBI accession number JX298876) was translated into protein sequence by the Translate tool of ExPASy ([www.expasy.org](http://www.expasy.org)). Secondary structure was predicted via the PDBSum [15] and VADAR servers [16]. PSI-BLAST with the available protein sequences in the PDB ([www.rcsb.pdb.org](http://www.rcsb.pdb.org)) was carried out to identify homologous protein sequences. The sequence and 3D structure of the homologous proteins were extracted from the PDB database. Multiple sequence alignments of the target and template sequences were carried out using Clustal W program [17] with default parameters.

### 5.2.3. Homology model construction and validation

To build the three-dimensional model of rCAL, protein sequences of the closest homologues with known crystal structures were selected from the PDB. MODELLER 9v10 program was used to build the 3D models of rCAL [18]. Out of the 50 models generated, the model with the lowest DOPE score was selected for validation using the

Ramachandran plot from PROCHECK [19], ProSAII [20] and ERRAT [21] programs. The final model was further refined through energy minimization by applying OPLS\_2005 force field (Schrödinger, Inc.) [22]. The RMSD between model and template was calculated by the Carbon alpha fitting method from PyMOL program [23].

#### 5.2.4. Docking studies of rCAL model with ligands

The AutoDock Vina 1.1.2 [24] docking package was used for ligand flexible docking simulations using the default settings. The structures of the ligands (hemin, spermine and thiamine) were obtained from the PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) database. The pdbqt files of ligands and receptor (rCAL) were generated using the MGLTools [25]. The pdbqt file is a modified protein data bank file format containing the information of atomic charges, atom type definitions and, for ligands, topological information. The grid box was defined from the binding sequences reported for hemin and spermine in the crystal structure of LS-24 (PDB ID: 3LP9). The size of the docking grid was kept as 32 Å x 32 Å x 32 Å and the grid spacing set at 1 Å. The interactions of the ligands with protein were visualized in PyMOL 1.3 ([www.pymol.org](http://www.pymol.org)) and LIGPLOT v 4.5.3 [26]. The binding energy was obtained for each ligand. The lowest binding energy conformation was considered as the most favourable docking pose.

#### 5.2.5. Prediction of aggregation-prone regions

The rCAL modelled structure was fed in the BioLuminate tool from the Schrödinger suite to identify the hot spots responsible for aggregation [27]. The amino acid sequence of rCAL was also fed into web-based tools like the Aggrescan [28], AmylPred [29], FoldAmyloid [30] etc. to predict the propensity of aggregation.

#### 5.2.6. Prediction of transmembrane domain

The presence of transmembrane domains in the lectin was predicted by the PSIPRED tool [31].

### 5.3. RESULTS AND DISCUSSION

#### 5.3.1. Sequence analysis of rCAL

The nucleotide sequence of rCAL was translated to obtain its amino acid sequence (see below). The lectin consists of two monomers of 231 amino acids each.

##### (A) Nucleotide sequence (NCBI accession no. JX298876)

5'- ATGGCAAAAACAGGTTACATTAATGCTGCTTTTCGTTTCATCTCAGAATA  
 ATGAAGCTTATCTTTTCATCAATGATAAGTACGTGTTGTTAGATTATGCTCC  
 AGGAACCAGCAACGATAAGGTTTTGTATGGACCAACTCCTGTTTCGTGATGG  
 TTTCAAATCACTTAATCAAACCGTATTTGGAAGCTACGGAGTTGATTGTAGC  
 TTCGACACCGATAATGATGAAGCTTTCATCTTTTATGAAAAGTTTTGTGCTC  
 TCATAGACTATGCTCCACATTCACAAAGACAAAATCATCTTAGGTCCTA  
 AGAAAATTGCTGACATGTTTCCTTTTTTCGAAGGAACTGTGTTTGAAAACGG  
 GATAGACGCGGCATACAGATCAACTAGGGGCAAAGAAGTTTACTTATTCAA  
 AGGAGATCAGTATGCTCGTATAGACTATGAAACCAACAGTATGGTTAATAA  
 AGAAATCAAAGCATTTCGCAACGGATTTTCCTTGTTTCCGTAACACAATCTTT  
 GAGAGTGGAACGGATGCAGCGTTTGCTTCTCACAAGACAAATGAAGTTTAC  
 TTTTCAAAGGTGACTACTATGCACGTGTTACTGTTACACCAGGCGCAACA  
 GATGATCAAATTATGGATGGTGTGAGGAAAACCTTTGACTATTGGCCATCT  
 CTTCGTGGCATAATACCTCTTGAGAATTAA - 3'

##### (B) Amino acid sequence in the 5'-3' frame

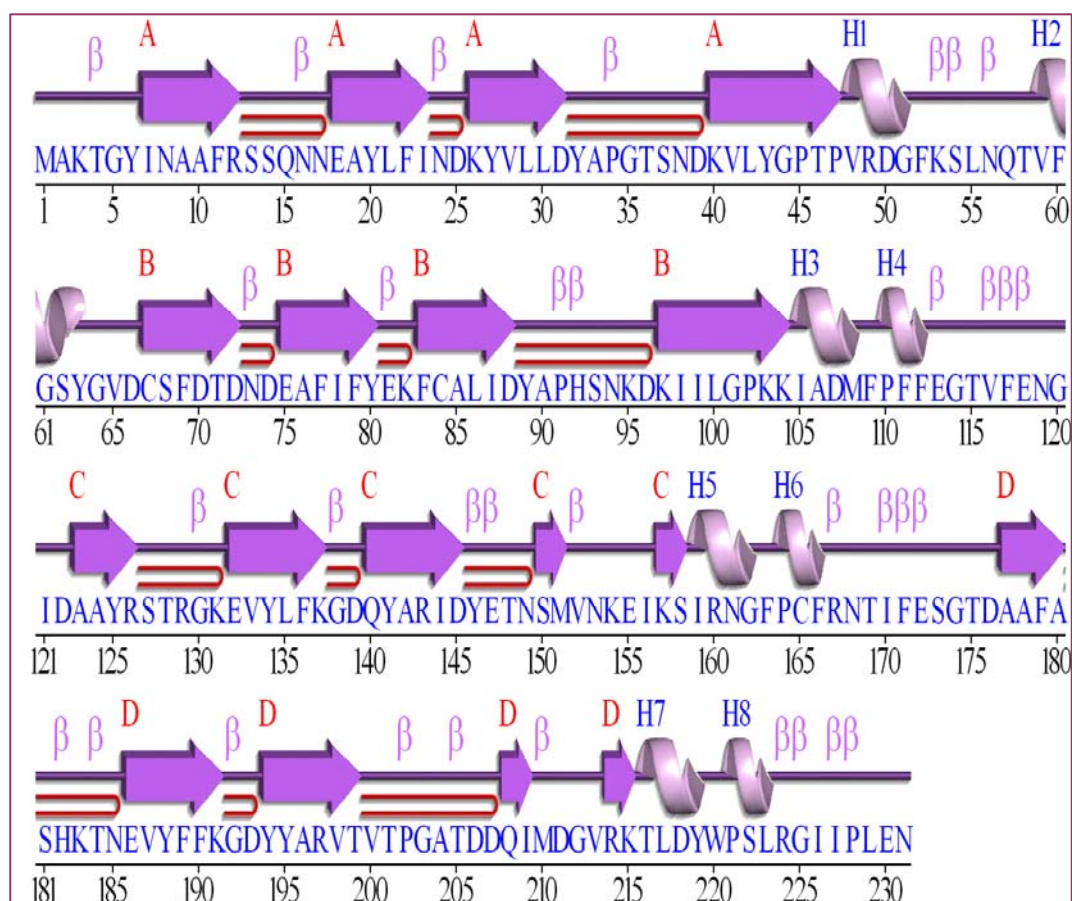
**Met** T K T G Y I N A A F R S S R N N E A Y L F I N D K Y V L L D Y A P G T S N  
 D K V L Y G P S L V R D G Y K S L A K T I F G T Y G I D C S F D T E Y N E A F  
 I F Y E N L C A R I D Y A P H S D K D K I I S G P K K I A D **Met** F P F F K G T V  
 F E N G I D A A F R S T K G K E V Y L F K E D K Y A R I D Y G T N R L V Q N I  
 K Y I S D G F P C L R G T I F E Y G **Met** D S A F A S H K T N E A Y L F K G E Y  
 Y A R I N F T P G S T N D T I **Met** G G V K K T L D Y W P S L R G I I P L E N  
**Stop**

### 5.3.2. Secondary structure analysis of rCAL

The secondary structure predictions of rCAL from the PDBSum and VADAR server correlated considerably well with those calculated from the experimental far UV CD spectrum (Chapter 4, Figure 4.1) (see Table 5.1 below). rCAL shows a relatively higher percentage of beta sheets (around 40 %).

**Table 5.1 (A) Secondary structure elements obtained from different programs.**

Program	% Helix	% Sheet	% Turn	% Unordered
CD Pro CONTINNL	3.7	41.1	21.3	33.9
PDBSum	9.1	39.8	--	51.1
VADAR	6	44	33	17

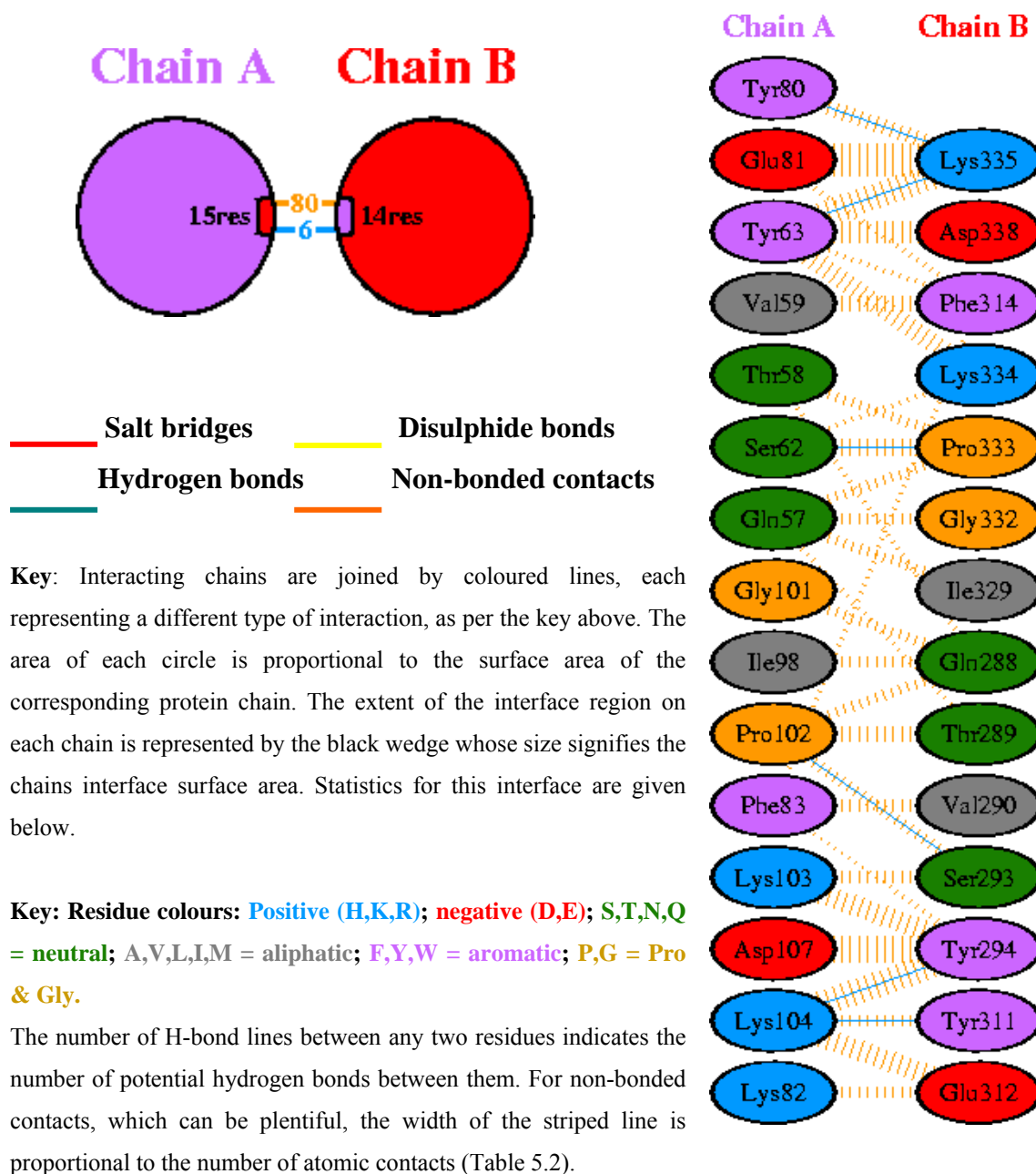


Key: Helices labelled as H1, H2, etc in lilac colour; strands by their sheets A, B, etc in red.

**Figure 5.2: Diagrammatic representation of the secondary structure of rCAL as seen in PDBSum.**

The secondary structure of rCAL consists of 4 beta sheets, 12 beta hairpins, 10 beta bulges, 18 strands, 8 helices, 3 helix-helix interactions and 34 beta turns.

Since rCAL is a dimer, it was interesting to identify the interface interactions taking place between the two monomers (Figure 5.3).



**Figure 5.3: Schematic diagram of interactions between the protein chains.**

**Table 5.2: Interface statistics between the two protein chains of rCAL.**

	<b>Chain A</b>	<b>Chain B</b>
<b>No. of interface residues</b>	15	14
<b>Interface area (<math>\text{\AA}^2</math>)</b>	657	666
<b>No. of salt bridges</b>	--	
<b>No. of disulphide bonds</b>	--	
<b>No. of hydrogen bonds</b>	6	
<b>No. of non-bonded contacts</b>	80	

Fifteen residues from chain A and fourteen residues from chain B were found to interact with each other to form the dimer, with the help of six hydrogen bonds and 80 non-bonded contacts.

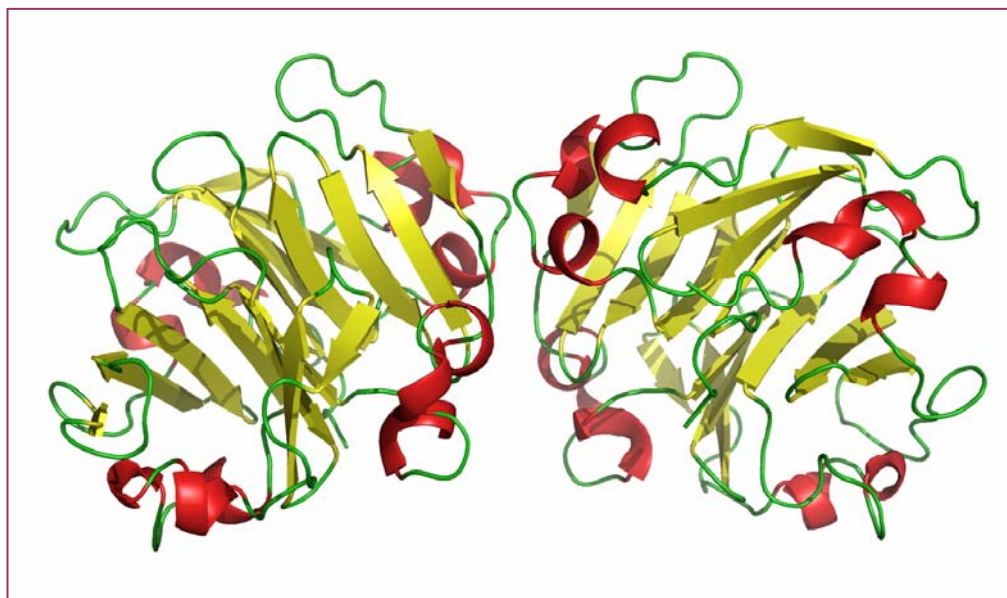
### 5.3.3. Multiple sequence alignment of rCAL

PSI-BLAST of the rCAL sequence was carried out from the PDB site ([www.rcsb.org](http://www.rcsb.org)) and three protein sequences showing maximum sequence similarity with rCAL were identified (Figure 5.4). 85% identity was obtained with the LS-24 albumin from *Lathyrus sativus* (PDB ID: 3LP9), 81% identity with the plant albumin from *Cicer arietinum* (PDB ID: 3V6N) and 51% identity was obtained with the CP4 hemopexin-fold protein (PDB ID: 3OYO) from *Vigna unguiculata*. Multiple sequence alignment using the Clustal X suite of rCAL with the sequences of these three proteins led to the identification of common residues among these four proteins.



chain non-hydrogen atoms. The entire dimeric template was used to model the dimeric rCAL.

Figure 5.5 and 5.6 are the cartoon representations of the dimer and monomer in the rCAL homology model respectively. Figure 5.7 shows the superimposition of the template and the rCAL homology model.

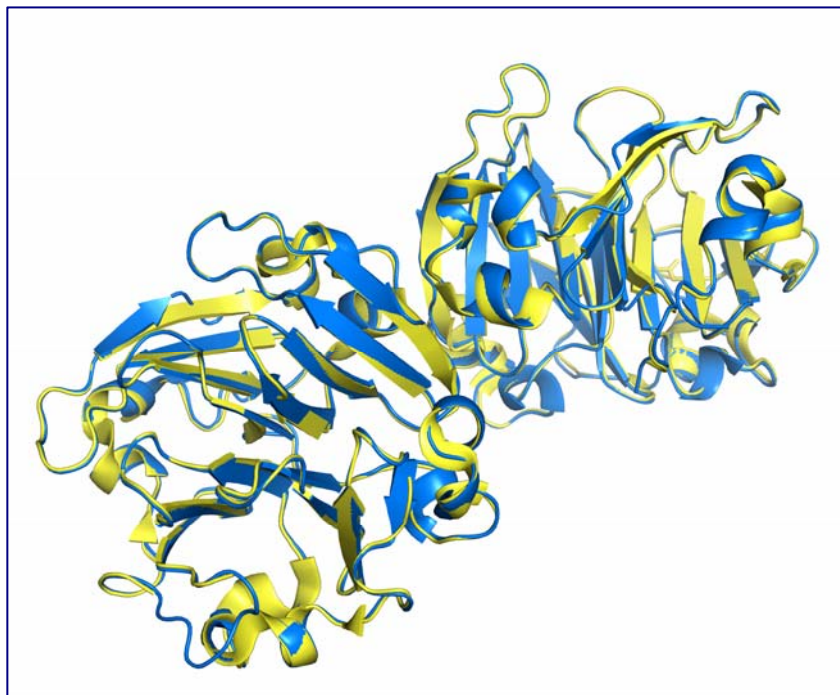


**Figure 5.5:** Cartoon representation of the dimer in the homology model of rCAL. The sheets are shown in yellow, helices in red and random coils in green colour.



**Figure 5.6:** Cartoon representation of the monomer in the homology model of rCAL. The sheets are shown in yellow, helices in red and random coils in green colour.





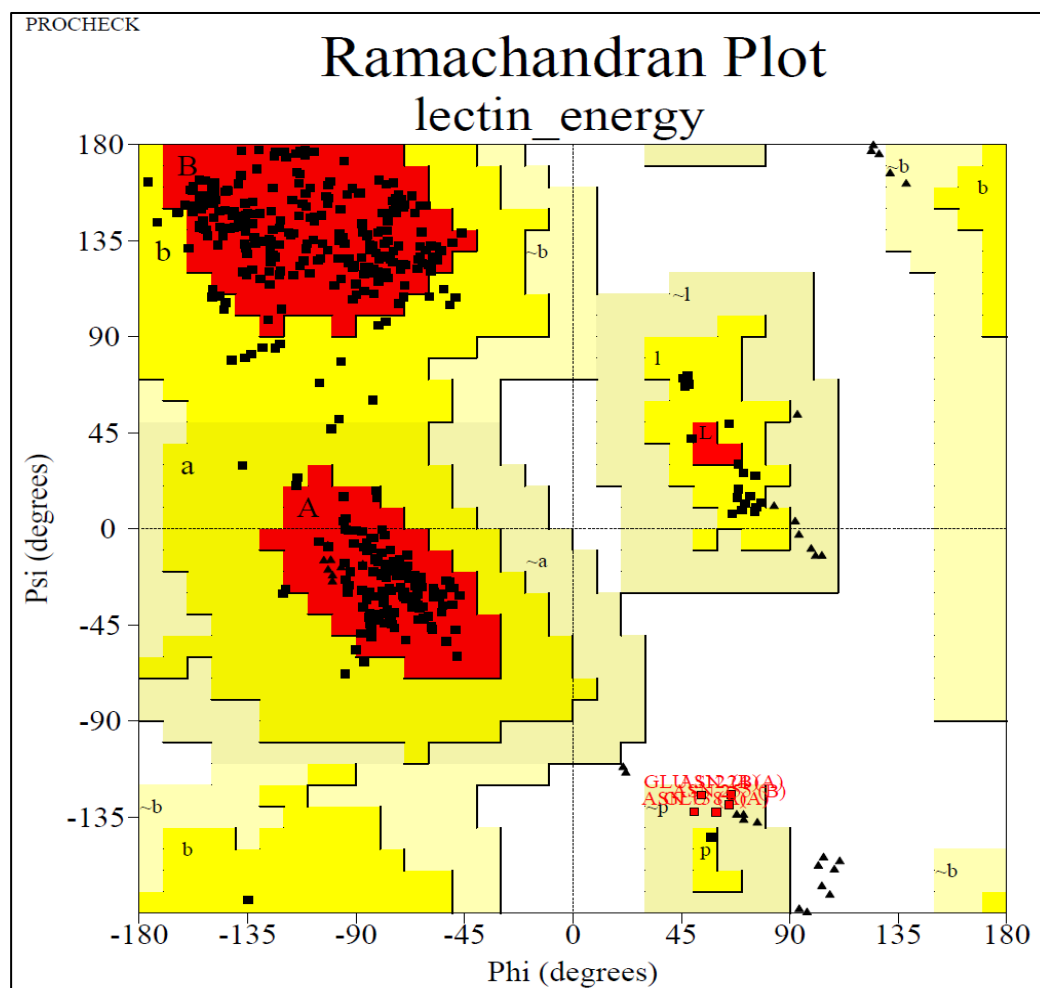
**Figure 5.7:** Superimposition of the homology model of rCAL (blue) on the template (yellow) (3LP9). RMSD value for the superimposition was 0.277.

### 5.3.5. Validation of the protein structure model

Stereochemical quality of the polypeptide backbone and side chains of the rCAL model was evaluated using the Ramachandran plot obtained from PROCHECK. It showed that 86.4 % of the residues are located in the most favoured region, 12.4 % in additionally allowed region and 1.2 % in generously allowed region of the Ramachandran plot (Figure 5.8). No residues were found in the disallowed regions.

The ERRAT program verifies the quality of the model. This program plots error values as a function of position in the sequence by sliding a nine residue window. The error function was based on the statistics of non-bonded atom-atom interactions in the template structure. The overall quality factor of modelled rCAL in ERRAT analysis was 89.391, expressed as the percentage of the protein for which the calculated error value falls below the 95 % rejection limit (Figure 5.9). On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value. Regions of the structure that can be rejected at the 95% confidence level are gray and regions that can be rejected at the 99% level are shown in dark black. The Verify 3D score was 90 %.

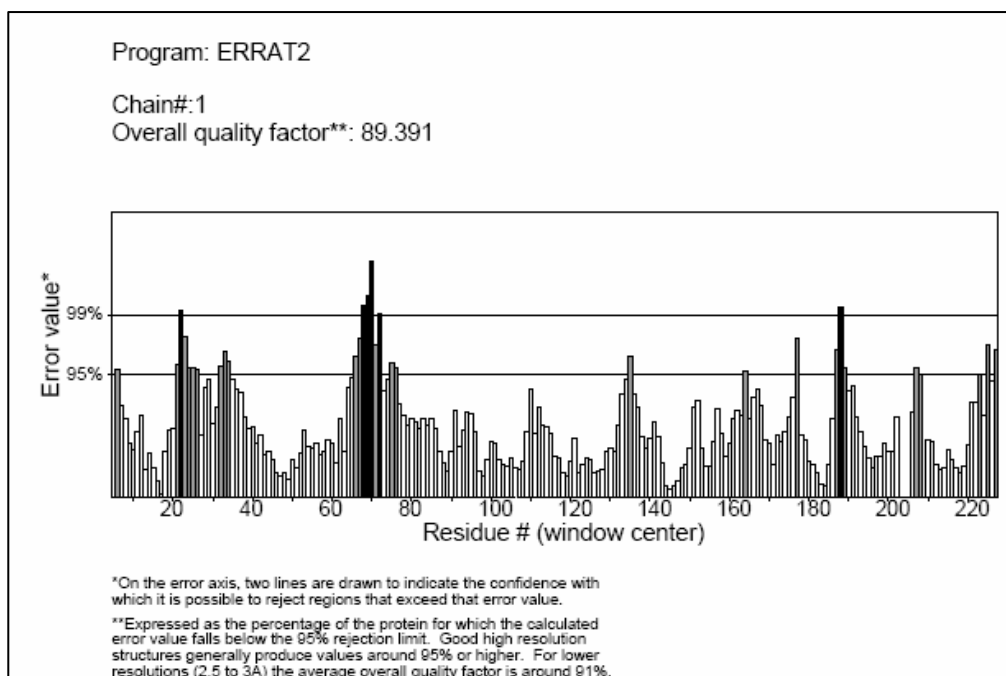
The ProSA tool was used to check the overall model protein structure for potential errors. The ProSAIL program (Protein Structure Analysis) is an established tool, which is frequently employed in structure prediction, refinement and validation of experimental protein structures. It generates Z score of model, which is a measure of compatibility between its sequence and structure. The model Z score should be comparable to the Z scores obtained from the template. ProSAIL analysis showed that protein folding energy of our modelled structure was in good agreement with that of the template. Z-plot analysis (ProSA) of the modelled protein measures compatibility between its sequence and structure. Z score value obtained for the rCAL model (-5.17) indicated its location within the space of protein related to X-ray (dark black dot in Figure 5.10). This value was quite comparable with template 3LP9 (-5.86) indicating that the obtained model was reliable.



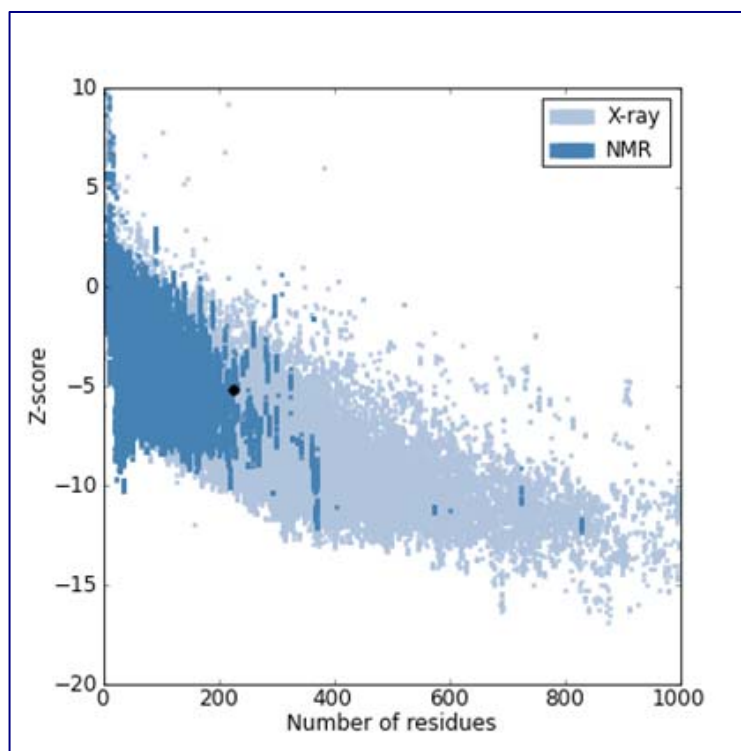
Plot statistics		
Residues in most favoured regions [A,B,L]	349	86.4%
Residues in additional allowed regions [a,b,l,p]	50	12.4%
Residues in generously allowed regions [~a,~b,~l,~p]	5	1.2%
Residues in disallowed regions	0	0.0%
-----		
Number of non-glycine and non-proline residues	404	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	20	
-----		
Total number of residues	462	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

**Figure 5.8: Ramachandran plot and its statistics for the validation of the homology model of rCAL.**



**Figure 5.9: ERRAT score for the validation of the homology model of rCAL**



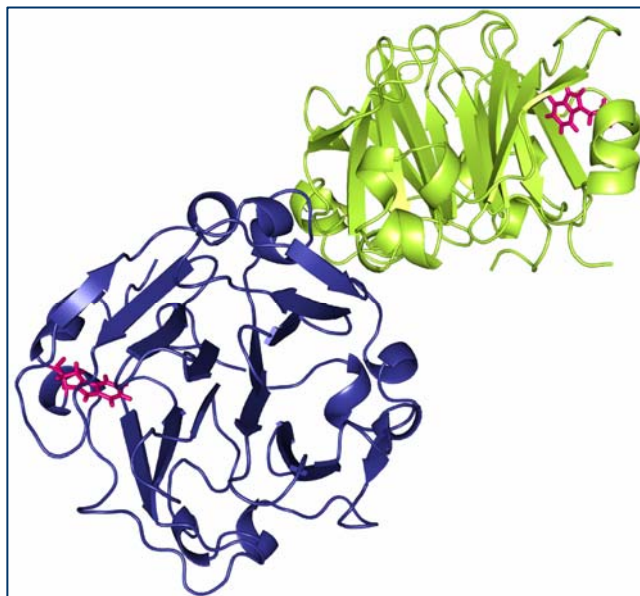
**Figure 5.10: ProSAn analysis for the validation of the homology model of rCAL.** The Z-Score for the model was - 5.17 and the Z-Score of the template was (3LP9) was - 5.86.

Thus, validation results suggested that the predicted model was a reliable 3D structure of rCAL. Validation data for the template 3LP9 was used as the baseline to assess the respective models.

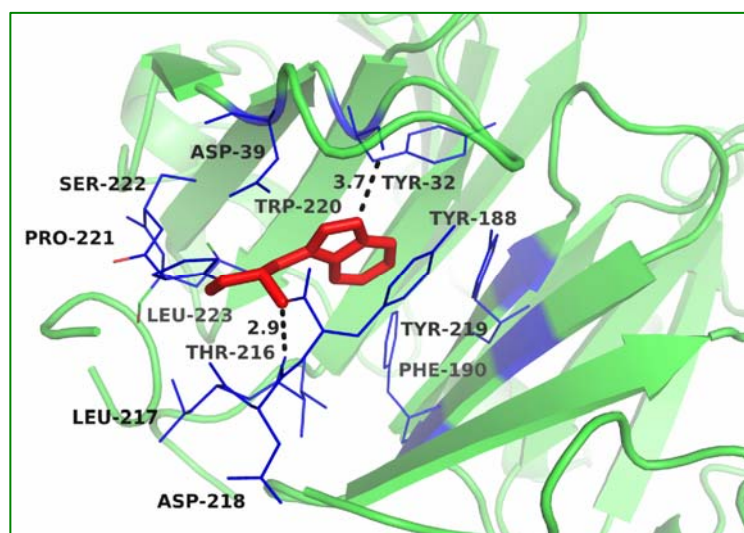
### 5.3.6. Tryptophan microenvironment in the model

The homology model of rCAL shows that the trp lies on the surface and is exposed to the solvent (Figure 5.11). The microenvironment of trp in the homology model of rCAL was visualized using PyMOL. The residues within 4 Å of the trp (W220) were identified (Figure 5.12). The residue was found to be surrounded by hydrophobic amino acids (L217, L223, T216, Y32, Y188, and Y219) and formed hydrogen bonds with Y32 and T216. A distinct lack of positively charged residues and the presence of two aspartic acid residues (D39 and D218) around W220 were observed. This justifies the observation that KI was unable to quench the fluorescence of rCAL in its native state, and also the fact that trp is in a heterogeneous environment, as deduced from the biphasic curvature obtained with CsCl in native state. The trp in

the model seems to be on the surface (Figure 5.11), but due to the presence of hydrophobic amino acids in its microenvironment, an effective emission maximum at 343 nm is obtained.



**Figure 5.11: Homology model of rCAL.** Cartoon representation of rCAL homology model constructed by the MODELER 9v10 software using 3LP9 (LS-24 albumin from *Lathyrus sativus*) as the template. The single trp in each monomer is shown in red as stick representation; the monomer on the left is oriented for clear projection of the trp.

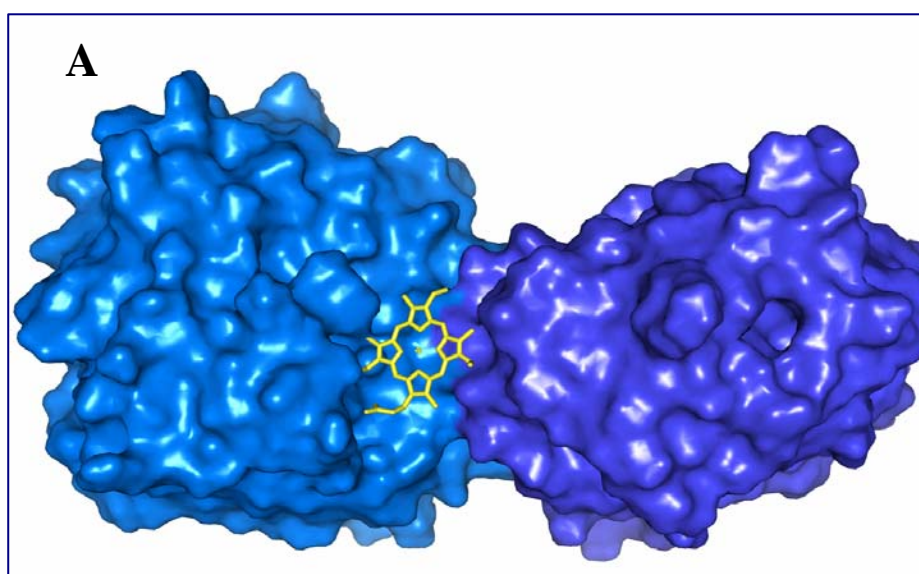


**Figure 5.12: Homology model of rCAL.** Cartoon representation of the trp microenvironment in the monomer of rCAL homology model. Trp 220 shown in red as stick representation; residues within 4 Å radius are shown as blue lines. Hydrogen bonds are shown as black dashes between W220 and T216 and Y32.

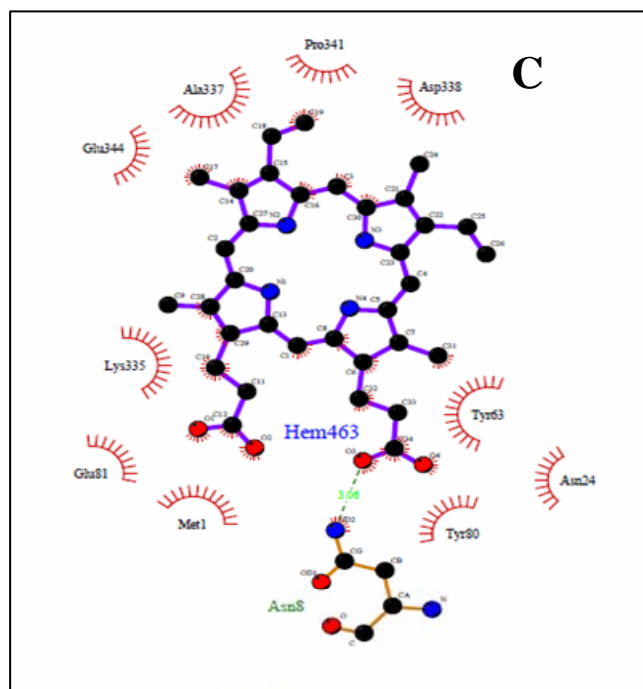
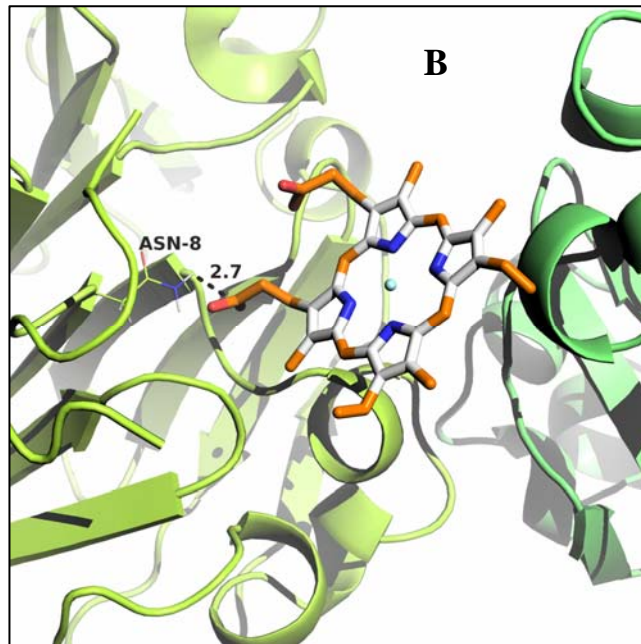
### 5.3.6. Docking with ligands

AutoDock is a suite of automated docking tools designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. It achieves significant improvements in the average accuracy of the binding mode predictions, while also being up to two orders of magnitude faster than AutoDock. It does not require choosing atom types and pre-calculating grid maps for them. Instead, it calculates the grids internally, for the atom types that are needed, and it does this virtually instantly.

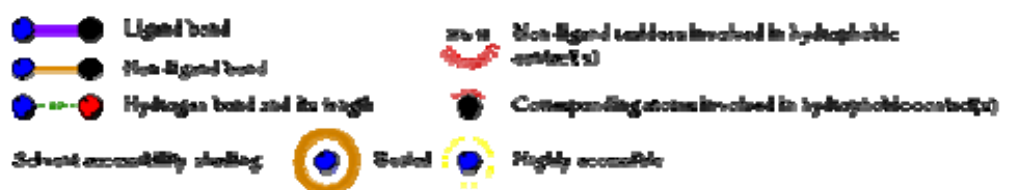
The probable binding site was deduced for hemin on one of the two monomers in the rCAL dimer model (Figure 5.13 (A)). A single amino acid (ASN-8) could be identified for the polar interaction between rCAL and hemin (Figure 5.13 (B)). Hydrophobic interactions occurring between hemin and rCAL were deduced from the LIGPLOT. Residues involved were MET-1, ASN-24, TYR-63, TYR-80, GLU-81 (all from the A chain), LYS-335 (B chain - 104), ALA-337 (B chain - 106), ASP-338 (B chain - 107), PRO-341 (B chain - 110) and GLU-344 (B chain - 113) (Figure 5.13 (C)). The most favourable docked pose of hemin had a binding energy of  $-31.35 \text{ kJmol}^{-1}$ .







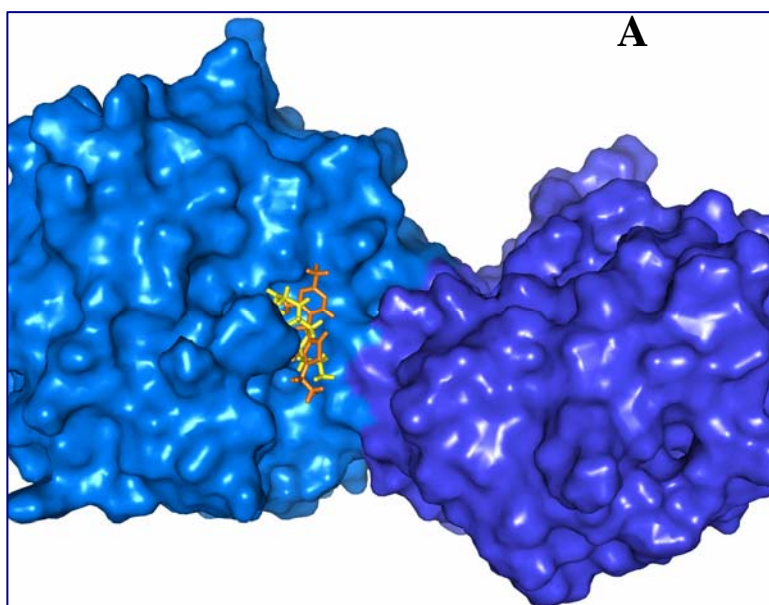
### Key



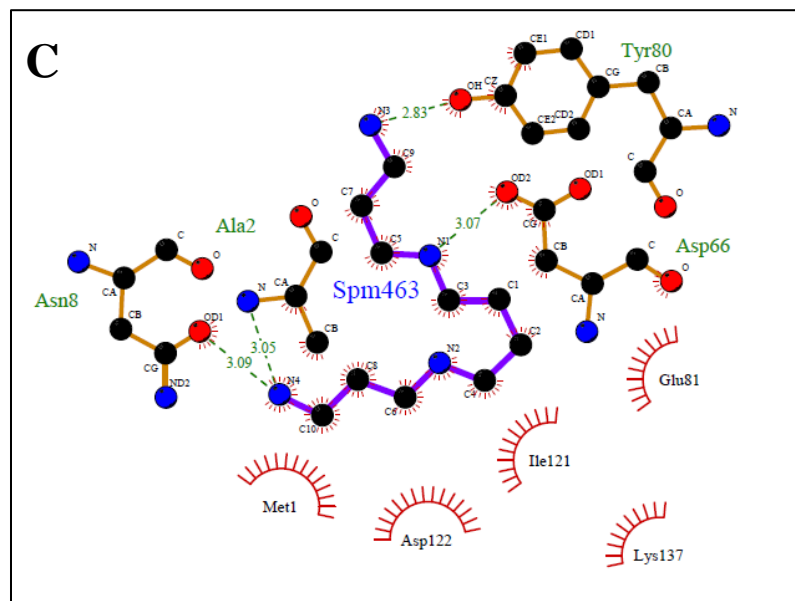
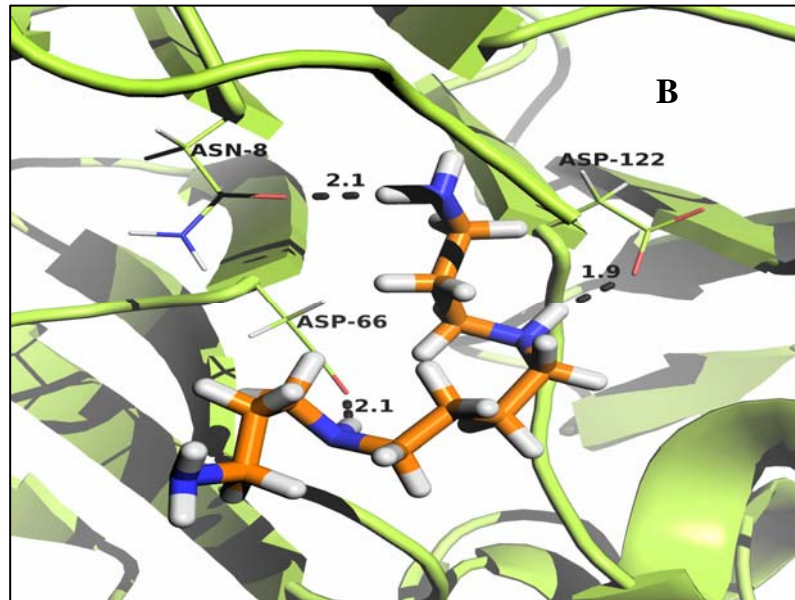
Key for Figure 5.13 (C): Ligand bonds are shown in thick purple lines, while non-ligand bonds (i.e., belonging to those rCAL residues to which the hemin is hydrogen-bonded) are shown with thin orange bonds. Hydrogen bonds are shown by green dashed lines with the bond length printed in the middle. Hydrophobic contacts between rCAL and hemin are indicated by brick-red spoked arcs. Other atoms involved in hydrophobic contacts have spokes coming out of the individual atoms.

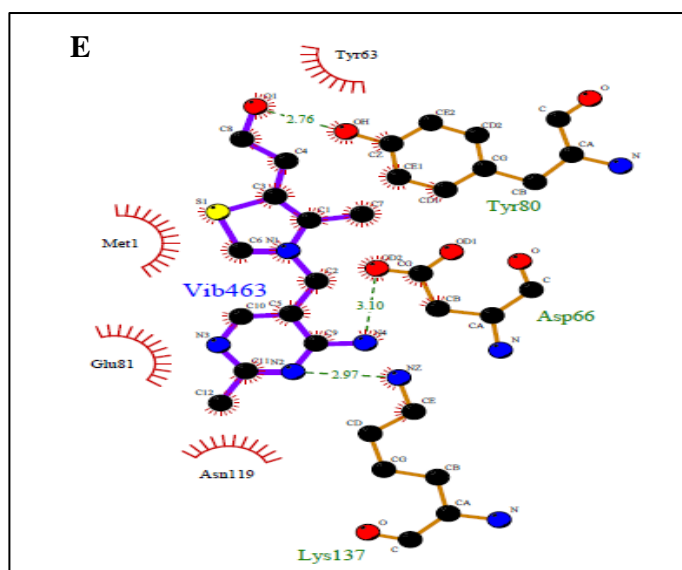
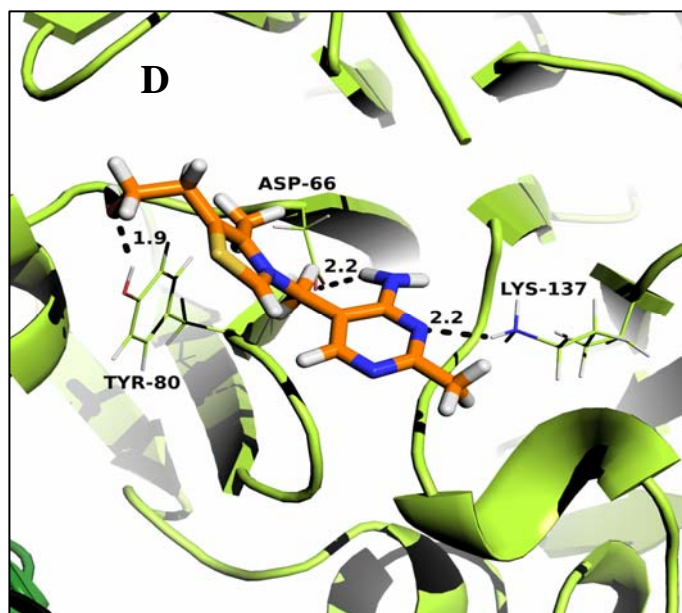
**Figure 5.13: Docking of Hemin on the rCAL homology model.** (A) Surface view of the docked hemin. Monomer in marine blue colour on the left represents the A chain while the monomer in tv\_blue colour on the right represents the B chain. (B) The hydrogen bond (bond length, 2.7 Å) between hemin and ASN-8 residue of the model is represented by black dash. The monomer on the left in light green colour represents the A chain, while the monomer in dark green colour on the right represents the B chain. Oxygen atom of the side chain of ASN-8 is represented in red colour. For the hemin ligand, nitrogen atoms are shown in blue colour, hydrogen atoms in white and carbon atoms are shown in tv\_orange colour. (C) Interactions between hemin and the residues on the rCAL model generated by the LIGPLOT.

For spermine and thiamine also, the binding sites on the model were present only on one of the two monomers and located close to each other (Figure 5.14 (A)). Spermine interacted with amino acids ASN-8 (2.1 Å), ASP-66 (2.1 Å) and ASP-122 (1.9 Å) of rCAL by formation of hydrogen bonds (bond lengths given in brackets) (Figure 5.14 (B)). Thiamine formed hydrogen bonds with ASP-66 (2.2 Å), TYR-80 (1.9 Å) and LYS-137 (2.2 Å) of rCAL (Figure 5.14 (C)). Binding energies of  $-15.88 \text{ kJ mol}^{-1}$  and  $-22.15 \text{ kJmol}^{-1}$  were obtained for spermine and thiamine, respectively.









**Figure 5.14: Docking of Spermine and Thiamine on the rCAL homology model.** (A) Surface view of the docked spermine (yellow) and thiamine (orange). Monomer in marine blue colour on the left represents the A chain while the monomer on the right in tv\_blue colour represents the B chain. (B) The hydrogen bonds between spermine and ASN-8 (2.1 Å), ASP-66 (2.1 Å) and ASP-122 (1.9 Å) of the model are shown as black dash (bond lengths in brackets). Oxygen atoms of the side chains of ASN-8, ASP-66 and ASP-122 are represented in red colour, while the interacting hydrogen atoms from the spermine are shown in white colour. For the spermine ligand, nitrogen atoms are shown in blue colour, hydrogen atoms in white and

carbon atoms are shown in tv\_orange colour. (C) Interactions between spermine and the residues on the rCAL model generated by the LIGPLOT. (D) The hydrogen bonds between thiamine and ASP-66 (2.2 Å), TYR-80 (1.9 Å) and LYS-137 (2.2 Å) are shown as black dash (bond lengths in brackets). Oxygen atoms of the side chains of ASP-66, TYR-80 and LYS-137 are represented in red colour, while the interacting hydrogen atoms from thiamine ligand are shown in white. For the thiamine ligand, nitrogen atoms are shown in blue colour, hydrogen atoms in white, carbon atoms are shown in tv\_orange colour while the sulphur atom is shown in yelloworange. (E) Interactions between thiamine and the residues on the rCAL model generated by the LIGPLOT.

Interestingly, both hemin and spermine were found to interact with ASN-8 of the protein. Gaur *et al.* [32] have shown that hemin and spermine bind to the LS-24 albumin in a mutually exclusive manner. Spermine is bound to the dimeric LS-24 albumin (3LP9 is its crystal structure), but hemin binding results in dissociation of the subunits accompanied by loss in the ability to bind spermine. This binding property might be true for rCAL also, wherein a common residue is involved in the binding of both hemin and spermine.

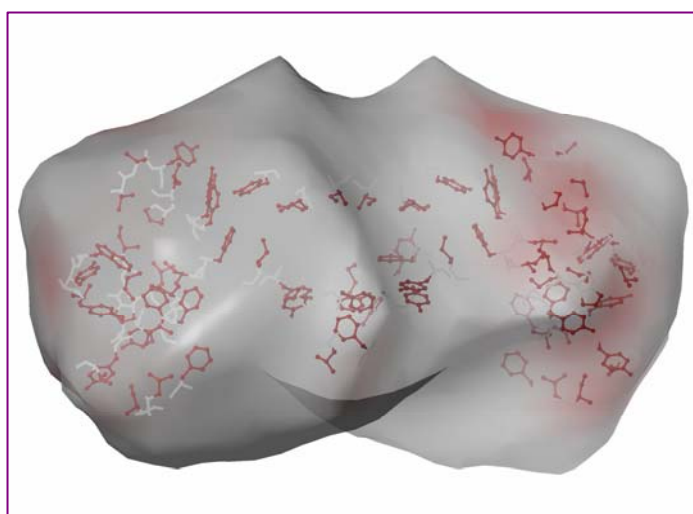
In the crystal structure of LS-24, binding of spermine involves the interaction of GLU-80, ASN-81, and SER-118 [32]. In the rCAL model, spermine binding involves ASN and two ASP residues. As can be seen from Figure 5.14 (B), the oxygen atoms from the side chains of ASN-8, ASP-66, and ASP-122 are involved in hydrogen bonding with the hydrogens of the spermine ligand. Hence, the replacement of Serine in the template to Aspartate does not affect the binding of spermine with the model. The binding pocket formed for spermine in the model of rCAL is different from that formed in the crystal structure of LS-24. The rCAL sequence has LYS residue in the 82<sup>nd</sup> position. In the model, the side chains of this LYS-82 residue from both monomers protrude in such a way that they occupy a position in the intersection (in between) of the two monomers (this is the site where spermine is shown to bind in the crystal structure of LS-24). Hence, this site is unavailable for the binding of spermine in the model.

### 5.3.7. Prediction of Aggregation

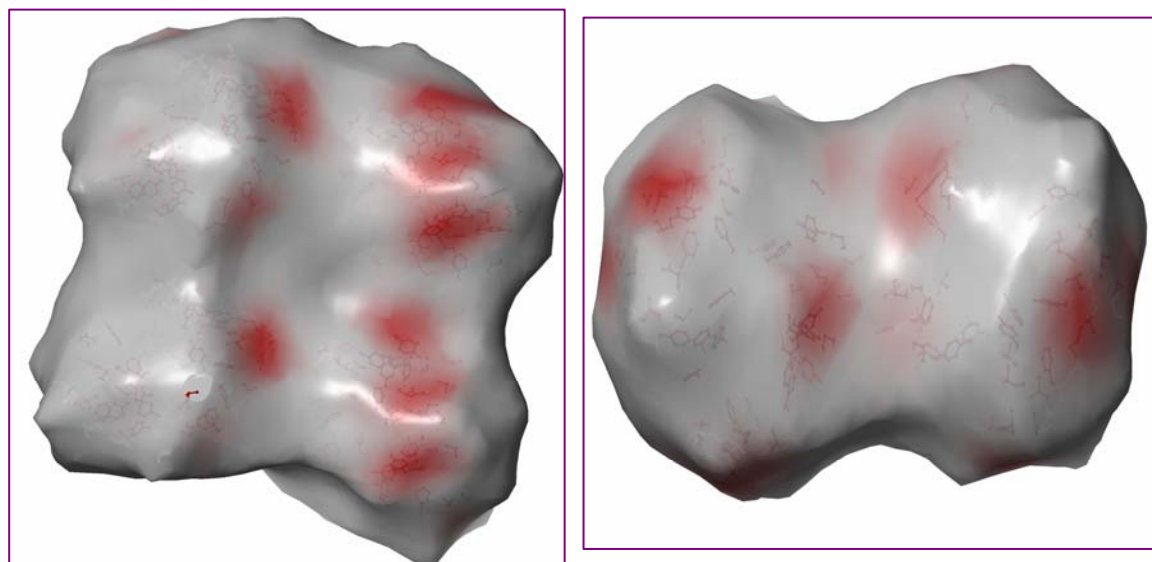
There are several computational approaches for detecting aggregation-prone regions and predicting polypeptide propensities for amyloid fibril formation. The

aggregation of rCAL upon thermal unfolding and at concentrations higher than 600  $\mu\text{g/ml}$  led us to investigate the reason for aggregation of this lectin. It is well known that the propensity for aggregation of a protein depends on its amino acid sequence. Hence, the rCAL modelled structure as well its protein sequence was fed into various web-based tools to predict the propensity of aggregation of its constituent amino acids.

### 5.3.7.1. BioLuminate1.0-Schrodinger



**Figure 5.15:** Aggregation-prone patches on the rCAL dimer as identified from the BioLuminate tool.



**Figure 5.16:** Identification of aggregation-prone patches on the structures of (A) LS-24 albumin and (B) CP4 hemopexin-fold protein from the BioLuminate tool.

The BioLuminate1.0 tool could identify several patches (as seen in red in the monomer on the right) as being prone towards aggregation. Similar patches were also found on the back side of the monomer on the left (Figure 5.15).

Similar red patches indicating the propensity towards aggregation were obtained in the crystal structures of the LS-24 albumin from *Lathyrus sativus* and the CP4 hemopexin-fold protein from *Vigna unguiculata* (Figure 5.16). This indicated that the propensity towards aggregation was conserved in the amino acids of these three proteins, which show a considerable sequence and structure similarity.

### 5.3.7.2. Aggrescan ([http:// bioinf.uab.es/aggrescan/](http://bioinf.uab.es/aggrescan/))

AGGRESKAN is based on an aggregation-propensity scale for natural amino acids derived from *in vivo* experiments and on the assumption that short and specific sequence stretches modulate protein aggregation. It relies on experimental or theoretical calculations of individual amino acid aggregation propensities and on the use of these values to scan protein sequences. For each polypeptide sequence input, AGGRESKAN calculates and reports: i) an aggregation-propensity value for each residue in the sequence and a graphical representation of the profile for the entire polypeptide; ii) the areas of profile peaks over a precalculated threshold and a graphical representation of peak-area values; iii) putative aggregation "hot spots", identified from the polypeptide's aggregation profile.

Eleven hot spots were identified as being prone to aggregation. These involved the residues: 5-10, 19-33, 55-67, 75-89, 105-109, 111-115, 132-138, 166-170, 186-194, 196-201, and 222-227.

### 5.3.7.3. Amyl Pred (<http://biophysics.biol.uoa.gr/AMYLPRED/>)

AMYLPRED2 is a web tool that employs a consensus of different methods that have been found or specifically developed to predict features related to the formation of amyloid fibrils. The consensus of these methods is defined as the hit overlap of at least  $n/2$  (rounded down) out of  $n$  selected methods (i.e. 5 out of 11 methods, if the user chooses to use all available methods). This is the primary output of the program. The tool predicts probable 'amyloidogenic determinants' peptides for a given amino acid sequence of a peptide or protein. Consensus prediction of residues 19 – 25, 27 – 31, 75 – 89, 108 – 111, 164 – 168, 187 – 192, and 194 – 199 as being prone to aggregation.

#### 5.3.7.4. FoldAmyloid (<http://antares.protres.ru/fold-amyloid/>)

This tool utilizes two characteristics (expected probability of hydrogen bonds formation and expected packing density of residues) to detect amyloidogenic regions in a protein sequence. The regions with high expected probability of the formation of backbone–backbone hydrogen bonds as well as regions with high expected packing density are mostly responsible for the formation of amyloid fibrils.

The residues predicted to be amyloidogenic are: 19-23, 27-31, 76-82, 84-88, 164-171, 195-199, and 218-222.

#### 5.3.7.5. WALTZ (<http://waltz.switchlab.org>)

Waltz was developed based on research with the aid of existing amyloid-prediction algorithms, but it has several distinctive features. With Waltz, numerous unknown amyloid aggregates, including aggregates that are hard to predict such as those forming prions, have been identified. Eighty percent of the data provided by Waltz have been validated with technologies such as electron microscopy and infrared spectroscopy. It is extremely important that data obtained with the aid of computerized tools be validated. None of the existing algorithms have a validation rate this high.

Residues likely to form amyloid regions are 19-25, 75-89, and 185-196. Table 5.3 summarizes the residues identified as being aggregation-prone from all the four servers.

**Table 5.3 Prediction of aggregation-prone residues from different servers**

Server	Aggrescan	AmylPred	FoldAmyloid	Waltz
<b>Residues predicted to aggregate</b>	5-10			
	<b>19-33</b>	<b>19 – 25</b>	<b>19-23</b>	<b>19-25</b>
		27 – 31	27-31	
	55-67			
	<b>75-89</b>	<b>75 – 89</b>	<b>76-82</b>	<b>75-89</b>
			<b>84-88</b>	
	105-109	108 – 111		
	111-115			
	132-138			
	166-170	164 – 168	164-171	
	186-194	187 – 192		185-196.
	196-201	194 – 199	195-199	
222-227		218-222.		

The patches of amino acids from 19 to 25 and 75 to 88 were identified by all the four servers. Thus, it can be seen that several common residues were identified that showed propensity towards aggregation by different web-based tools that utilized different criteria to identify the aggregation-prone or amylogenic residues.

### 5.3.8. Prediction of transmembrane domains

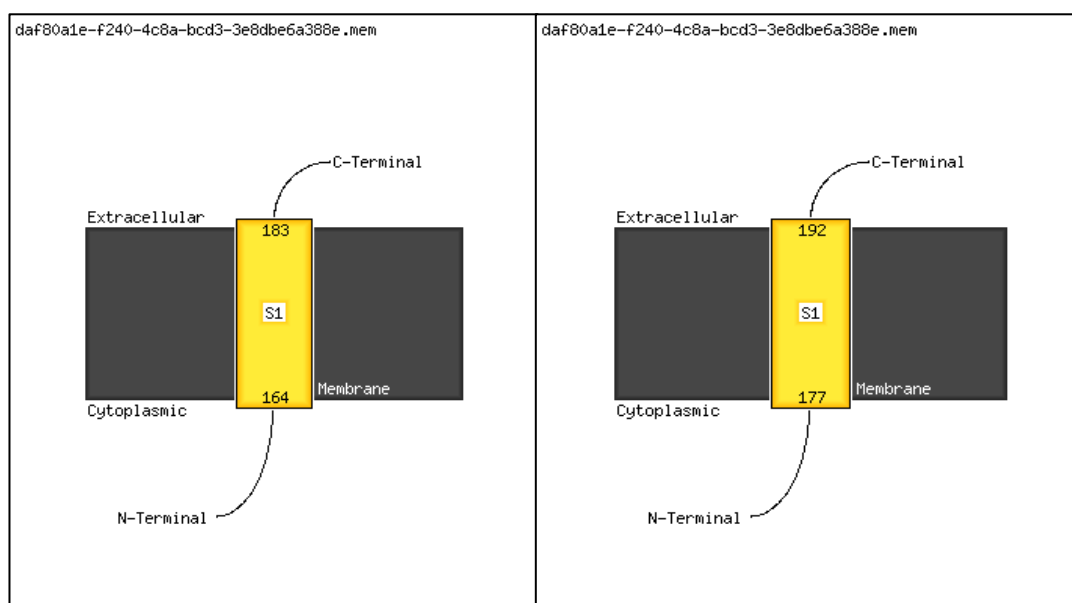
PSIPRED is a simple and accurate secondary structure prediction method, incorporating two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST). Using a very stringent cross validation method to evaluate the method's performance, PSIPRED 3.2 achieves an average Q<sub>3</sub> score of 81.6%.

#### 5.3.8.1. MEMSATSV: Transmembrane helix prediction

MEMSATSV is highly accurate predictor of transmembrane helix topology. It is capable to discriminating signal peptides and identifying the cytosolic and extracellular loops.

#### 5.3.8.2. MEMSAT3: Transmembrane Topology Prediction

MEMSAT V3 is a widely used all-helical membrane protein prediction method MEMSAT. The method was benchmarked on a test set of transmembrane proteins of known topology. From sequence data MEMSAT was estimated to have an accuracy of over 78% at predicting the structure of all-helical transmembrane proteins and the location of their constituent helical elements within a membrane.



**Figure 5.17:** Cartoon depicting the transmembrane topology of the rCAL lectin as obtained from the (A) MEMSAT3VM and (B) MEMSAT3 prediction method.

A single transmembrane domain comprising of amino acid residues (164 to 183, or 177 to 192) was identified in the rCAL sequence (common residues included 177 to 183) (Figure 5.17). As already mentioned in Chapter 3, the rCAL lectin does not recognize any simple sugars but only agglutinates fetuin and its asialo glycan. It might be possible that rCAL binds to the erythrocyte membranes with the help of this transmembrane domain and hence results in agglutination of the erythrocytes. This might explain the lack of sugar specificity obtained for rCAL.

### 5.3.9. Protein stability from the ratio of arginine to lysine residues

Earlier studies have indicated a correlation between protein stability and the number of arginines on the protein surface [33]. Thermophilic proteins have, on average, a higher content of arginine and a greater amount of reducing lysines on the protein surface [34, 35]. The thermotolerance of proteins is increased upon Lys to Arg mutations, because of stronger hydrogen bonding of the large guanidinium group of arginine with neighbouring polar groups [33]. Due to its large side chains, arginine may be important in both local and long-range interactions [34]. Additionally, they may also prove advantageous due to their involvement in ion pair networks, shielding of the hydrophobic hydrocarbon chain from water, more rigidity, and ability to remain charged at higher pHs.

Table 5.4 gives a comparative account of the Arg: Lys ratios of some thermolabile lectins/ albumins including rCAL with thermostable lectins. The ratios have been calculated manually by us from the available amino acid sequences.

From the table we can see that the lectins/ albumins reported to be thermolabile; from biophysical experiments, show a lower ratio of Arg:Lys residues (lectins 1 to 7). In contrast, the *Trichosanthes kirilowii* lectin as well as the *Ricinus communis* lectin are reported to be extremely thermostable and the Arg:Lys ratio for these two lectins was found to be higher than one. There is no fixed criterion yet known that determine the stability of a protein. Several factors are known to contribute towards protein stability. And the ratio of Arg:Lys is one among many such factors that seem to play a role in determining stability of a protein.



Table 5.4: Arg:Lys ratio of few lectins and albumins

	Lectin	Arg: Lys ratio
1.	rCAL from <i>Cicer arietinum</i> (recombinant)	0.625
2.	nCAL from <i>Cicer arietinum</i> (wild type)	0.526
3.	LS-24 albumin from <i>Lathyrus sativus</i>	0.474
4.	CP4 protein from <i>Vigna unguiculata</i>	0.583
5.	<i>Artocarpus hirsuta</i> lectin [36]	0.222
6.	Jacalin (wild type) from <i>Artocarpus integrifolia</i> [37]	0.222
7.	Jacalin (recombinant) from <i>Artocarpus integrifolia</i>	0.235
8.	<i>Trichosanthes kirilowii</i> lectin [38]	1.231
9.	<i>Ricinus communis</i> lectin (Ricin) [38]	3.67

**REFERENCES**

1. Langer, T. and Hoffmann, R.D. (2001), Virtual screening: an effective tool for lead structure discovery, *Current Pharmaceutical Design*, **7**, 509–527.
2. Bajorath, J. (2002), Integration of virtual and high-throughput screening, *Nature Reviews Drug Discovery*, **1**, 882–894.
3. Gohlke, H. and Klebe, G. (2002), Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, *Angewandte Chemie International Edition*, **41**, 2644–2676.
4. Wirth, M., Fuchs, A., Wolf, M., Ertl, B. and Gabor, F. (1998), Lectin mediated drug targeting: Preparation, binding characteristics, and antiproliferative activity of wheat germ agglutinin conjugated doxorubicin on Caco-2 cells, *Pharmaceutical Research*, **15**, 1031-1037.
5. Clark, M.A., Hirst, B.H. and Jepson, M.A. (2005), Lectin-mediated mucosal delivery of drugs and microparticles, *Advanced Drug Delivery Reviews*, **43**, 207–223.
6. Werz, D. and Seeberger P (2005), Total synthesis of antigen *Bacillus anthracis* tetrasaccharide: Creation of an anthrax vaccine candidate, *Angewandte Chemie International Edition*, **44**, 6315–6318.
7. Krieger, E., Nabuurs, S.B. and Vriend, G. (2003), Homology Modelling, In: *Structural Bioinformatics*, Eds: Bourne, P.E. and Weissig, H., Wiley-Liss Inc., 507–521.
8. Epstein, C.J., Goldberger, R.F. and Anfinsen, C.B. (1963), The genetic control of tertiary protein structure. Model systems, *Cold Spring Harbor Symposia on Quantitative Biology*, **28**, 439–449.
9. Chothia, C. and Lesk, A.M. (1986), The relation between the divergence of sequence and structure in proteins, *EMBO Journal*, **5**, 823–836.

10. Sander, C. and Schneider, R. (1991), Database of homology-derived protein structures and the structural meaning of sequence alignment, *Proteins*, **9**, 56–68.
11. Kuntz, I.D., Blaney, J.M., Oatley, S. J., Langridge, R. and Ferrin, T.E. (1982), A geometric approach to macromolecule–ligand interactions, *Journal of Molecular Biology*, **161**, 269–288.
12. Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J. (2004), Docking And Scoring In Virtual Screening For Drug Discovery: Methods And Applications, *Nature Reviews Drug Discovery*, **3**, 935–949.
13. Smith, A. (2003), Protein Misfolding, *Nature*, **426**, 883–883.
14. Rochet, J.C., Lansbury Jr., P.T. (2000), Amyloid fibrillogenesis: themes and variations, *Current Opinion in Structural Biology*, **10**, 60–68.
15. Laskowski, R.A. (2009), PDBSum new things, *Nucleic Acids Research*, **37**, D355-D359.
16. Wishart, D.S., Willard, L., Richards, F.M. and Sykers, B.D. (2012), VADAR: A Comprehensive Program Suite for Protein Structural Analysis, (<http://redpoll.pharmacy.ualberta.ca/vadar>).
17. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007), Clustal W and Clustal X version 2.0., *Bioinformatics*, **23**, 2947–294.
18. Sali, A. and Blundell, T.L. (1993), Comparative protein modelling by satisfaction of spatial restraints, *Journal of Molecular Biology*, **234**, 779–815.

19. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996), AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *Journal of Biomolecular NMR*, **8**, 477–486.
20. Widerstein, M. and Sippl, M.J. (2007), ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Research*, **35**, W407–W410.
21. Colovos, C. and Yeates, T.O. (1993), Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Science*, **2**, 1511–1519.
22. *Schrödinger Suite 2011 Protein Preparation Wizard*, Epik version 2.3, Schrödinger LLC, New York.
23. PyMOL, (2012), *The PyMOL Molecular Graphics System*, Version 1.2r3pre, Schrödinger LLC, New York, ([www.pymol.org](http://www.pymol.org)).
24. Trott, O. and Olson, A.J. (2010), AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *Journal of Computational Chemistry*, **31**, 455–461.
25. Sanner, M.F. (1991), Python: a programming language for software integration and development, *Journal of Molecular Graphics and Modelling*, **17**, 57–61.
26. Wallace, A.C., Laskowski, R.A. and Thornton, J.M. (1996), LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Engineering*, **8**, 127–134.
27. *BioLuminate 1.0 (2012) User Manual*, Schrödinger, LLC.
28. Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X. and Ventura, S. (2007), AGGRESKAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides, *BMC Bioinformatics*, **8**, 65–82.

29. Tsolis, A.C., Papandreou, N.C., Iconomidou, V.A. and Hamodrakas, S.J. (2013), A Consensus Method for the Prediction of "Aggregation-Prone" Peptides in Globular Proteins, *PLoS ONE*, **8**, e54175.
30. Garbuzynskiy, S.O., Lobanov, M.Y. and Galzitskaya, O.V. (2010), FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence, *Bioinformatics*, **26**, 326–332.
31. McGuffin, L.J., Bryson, K. and Jones, D.T., (2000), The PSIPRED protein structure prediction server, *Bioinformatics*, **16**, 404–405.
32. 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.
33. Turunen, O., Vuorio, M., Fenel, F. and Leisola, M. (2002), Engineering of multiple arginines into the Ser/Thr surface of *Trichoderma reesei* endo-1,4- $\beta$ -xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH, *Protein Engineering*, **15**, 141–145.
34. Argos, P., Rossman, M.G., Grau, U.M., Zuber, H., Frank, G. and Tratschin, J.D. (1979), Thermal stability and protein structure, *Biochemistry*, **18**, 5698–5703.
35. Vieille, C., Epting, K.L., Kelly, R.M. and Zeikus, J.G. (2001), Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase, *European Journal of Biochemistry*, **268**, 6291–6301.
36. Gaikwad, S.M., Gurjar, M.M. and Khan, M.I. (2002), Artocarpus hirsuta lectin Differential modes of chemical and thermal denaturation, *European Journal of Biochemistry*, **269**, 1413–1417.

37. Gaikwad, S.M. and Khan, M.I. (2003), pH-dependent aggregation of oligomeric *Artocarpus hirsuta* lectin on thermal denaturation, *Biochemical and Biophysical Research Communications*, **311**, 254–257.

38. 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 Crystallographica Section F Structural Biology and Crystallization Communications*, **62**, 205–209.

## **CHAPTER: 6**

---

## **DISCUSSION**

---

For over a century, the areas of nucleic acids, proteins and lipids have captured the attention of investigators worldwide. Carbohydrates have received increased attention only more recently through the expanding field of glycobiology, probably because they are very complex. The intriguing ability of glycans to encode an immense repertoire of biological information is due to the fact that they are not encoded by the genome [1]. The genome codes for enzymes that act on glycans such as glycosyl transferases and glycosidases. The glycosylation patterns of glycolipids and glycoproteins are determined by the combined activity levels of these enzymes in the endoplasmic reticulum (ER) and the Golgi apparatus [2-4].

It is difficult to discuss carbohydrates without reference to lectins. Lectins are a diverse group of carbohydrate binding proteins that are found within and associated with organisms from all kingdoms of life. They are defined as proteins that preferentially recognize and bind carbohydrate complexes protruding from glycolipids and glycoproteins [5-8]. The interaction of lectins with particular carbohydrates can be as specific as the interaction between those of antigen and antibody or substrate and enzyme [9]. Lectins bind not only to oligosaccharides on cells but also to free-floating glycans including monosaccharides. Lectin–monosaccharide interactions, however, are relatively weak with dissociation constants often of the order of micro molar to milli molar range [10, 11].

The high affinity and specificity of lectins for glycoconjugates has found many applications in biological and biomedical research [12]. Their contribution in cell identification and separation [13], detection, isolation, and structural studies of glycoproteins [14], investigation of carbohydrates on cells and subcellular organelles [15, 16], mapping of neuronal pathways mitogenic stimulation of lymphocytes, purging of bone marrow for transplantation [17] and studies of glycoprotein biosynthesis is remarkable [18].

Chickpea (*Cicer arietinum* L) seeds have been recognised as a valuable and economic source of vegetable proteins in the human diet of numerous developing countries. Chickpea is one of the important legume crops with production ranking third in the world, and first in the Mediterranean basin, among pulse crops [19]. The potential for increased use of chickpea is related to its relatively low cost, relatively high protein content (25.3–28.9%), high protein digestibility (76–78%) [20-21] and other desirable functionalities [22-23]. Chickpea proteins are usually classified into two



major 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. However, albumins play an essential role in seeds because they include most of the enzymatic and metabolic proteins. In addition, albumins are of nutritional importance because of their high content of lysine and sulphur amino acids.

A wild type lectin from the seeds of *Cicer arietinum* (chick pea) was purified using ion-exchange chromatography [24]. The N- terminal sequence of this wild type lectin showed 99 % homology with the nucleotide sequence of the PA2 albumin from *Pisum sativum*. Hence, the sequence of this PA2 was utilized to design primers for isolation of the lectin gene from *Cicer arietinum*. The gene consisting of 693 bp was successfully cloned in *E. coli*. For expression of the lectin, the *E. coli* strain BL21-CodonPlus (DE3)-RIL was used. The CodonPlus (DE3)-RIL cells are additionally supplemented with rare codons (Arg, Ile and Leu) known to limit the translation of eukaryotic proteins in *E. coli*. The advantage of using this strain was that the expression in the uninduced CodonPlus cells was more pronounced than that in the induced plain BL21 DE3 cells. Maximum expression of the lectin was achieved in the soluble fraction. The recombinant lectin was purified using ion-exchange and gel filtration chromatography. The purified lectin, consisting of 462 amino acids, was found to be a homodimer of approximate molecular mass 53,000 Da. It showed hemagglutinating activity against pronase-treated rabbit erythrocytes. This lectin was the subject of further biochemical and biophysical characterization.

The hemagglutinating activity of the recombinant lectin was checked against several simple mono- and oligosaccharides. But the lectin failed to inhibit any of these. It showed hemagglutination inhibition only against complex glycoproteins like fetuin and its asialo triantennary glycan. The thermodynamics of binding of the asialo triantennary fetuin glycan to the lectin was carried out by fluorescence spectroscopy. The binding was found to be spontaneous and enthalpically driven. The lack of recognition towards simple sugars led us to investigate the possibility of binding of the lectin to non-carbohydrate ligands. The nucleotide sequence of the recombinant lectin revealed the presence of the hemopexin fold, indicating the likelihood of heme binding to the lectin. The lectin is an albumin by nature; hence homologous albumins from other pea seeds were identified. The albumins from grass pea (*Lathyrus sativus*) and

cow pea (*Vigna unguiculata*) were found to bind hemin, spermine and thiamine. Hence, the binding of the chick pea lectin was assessed with hemin, spermine and thiamine through fluorescence spectroscopy. Hemin showed strong binding for the lectin, thus confirming the functional presence of the hemopexin domains. Spermine and thiamine also showed lectin binding in the milli molar range.

Plant hemopexins have been shown to be involved in spermine biosynthesis [25]. Albumins from pea and chick pea bind thiamine; hence they have been thought to be involved in storage of the vitamin during germination [26]. Both spermine and thiamine have been shown to play important roles in reducing the oxidative stress during different abiotic stress conditions [27, 28]. Taken together, these studies lead to the speculation of a possible role of rCAL in plant defence against various stress conditions.

The gene sequence revealed the presence of a single tryptophan (trp) in rCAL. The native protein showed intrinsic fluorescence with maximum emission ( $\lambda_{\max}$ ) at 343 nm, suggesting the presence of the single trp in a partially hydrophobic environment. Decomposition analysis of the trp fluorescence spectrum by the PFAST program revealed the presence of 39% of class S (buried trp residue or a conformer with slight flexibility of the microenvironment) and 61% of class II conformer (containing structured water molecules near to the indole ring of trp). The lifetime of intrinsic emission decay of rCAL when studied in nanosecond domain could be described by two lifetimes. The single trp thus existed as two different conformers at a given time: one with the shorter lifetime lies on the surface of the protein. And its fluorescence decays faster, while the longer conformer lies in the interior and decays slowly. The  $\lambda_{\max}$  at 343 nm is hence the cumulative intrinsic fluorescence of rCAL.

Studies of indole fluorescence quenching by added solutes have provided valuable information regarding the structure and dynamics of proteins in solution [37-38]. The quenching reaction involves physical contact between the quencher and an excited indole ring, and can be kinetically described in terms of a collisional and a static component. The microenvironment of the single tryptophan in rCAL was evaluated with solute quenching studies. The tryptophan residue was found to be surrounded by negatively charged amino acids, since its fluorescence could be quenched only by cesium chloride, and not by potassium iodide. Upon complete denaturation with 6 M guanidine hydrochloride, there occurs a redistribution of charges

from negative to positive. This redistribution occurs even in the presence of 1.75 M guanidine hydrochloride. The upward curvature obtained during quenching of the denatured lectin with acrylamide was resolved with lifetime fluorescence. The quenching involves the contribution from both static as well as collisional components; the static component being more prominent than the collisional one.

Stability of a protein can be assessed by evaluating its behaviour in the presence of denaturing agents like urea and guanidine hydrochloride, or at higher temperatures etc. Conformational transitions of the recombinant chick pea lectin were studied using fluorescence and circular dichroism spectroscopy. The conformational characterization of rCAL revealed the thermo-labile and denaturant sensitive nature of the protein. rCAL showed flexibility in its secondary structure above 45 °C. Transient exposure of hydrophobic residues at this temperature resulted in insoluble aggregation. The protein being albumin by nature was more susceptible to thermal aggregation. Progressive unfolding of the rCAL dimer was obtained with increasing concentrations of guanidine-hydrochloride. The unfolding process involved several intermediates and was irreversible. At lower denaturant concentration, the dimeric lectin began to unfold; this was followed by the dissociation of the dimer into two unfolded monomers with increasing denaturant concentration. When the protein was allowed to renature, a further loss in structure was observed indicating misfolding. A similar mode of thermal and chemical denaturation was hence observed for rCAL wherein denaturation under both these conditions was found to be irreversible. Incubation of rCAL in acidic as well as alkaline buffers resulted in the loss of secondary structure within one hour.

PSI-BLAST of the amino acid sequence of rCAL showed 85 % homology with the crystal structure of LS-24 albumin from *Lathyrus sativus* (PDB ID: 3LP9). This crystal structure was used as a template for the homology model construction of rCAL using MODELLER. The validation parameters of the model and template matched very well, indicating the reliability of the model. The model was used for docking simulations in order to identify the binding sites for hemin, spermine and thiamine. A common residue ASN-8 was identified as being crucial for the binding of both hemin and spermine. This is in agreement with the observation that the LS-24 albumin binds hemin and spermine in a mutually exclusive manner. Hemin is bound in between the two dimers; spermine binding results in dissociation of the two monomers. The

residues involved in the binding of hemin and spermine in both rCAL and LS-24 are conserved in nature.

The homology model was also utilized to visualize the microenvironment around the single tryptophan. In the model, the trp appears to be exposed to the solvent. The trp residue was found to be surrounded by hydrophobic amino acids and two aspartate residues. No positively charged residues could be identified within 4 Å of the trp. The cumulative fluorescence maximum obtained at 343 nm is hence a result of the presence of these hydrophobic residues. These results correlate very well with those obtained from solute quenching studies, wherein the native protein does not get quenched by potassium iodide. The downward curvature obtained with cesium chloride for the native protein indicates a heterogeneous population of amino acids around the trp, as was seen from the model.

The lectin showed aggregation at concentrations above 600 µg/ml and also at temperatures above 55 °C. It is a well known fact that the amino acid sequence of a protein determines its propensity towards aggregation. Hence, the amino acid sequence of rCAL was fed in several web-based servers to predict the residues or patches responsible for aggregation. The BioLuminate tool from the Schrödinger suite could identify several common patches in the rCAL model as well as in the crystal structures of the homologous albumins, viz. LS-24 from *Lathyrus sativus* and CP4 protein from *Vigna unguiculata*. Various web-based servers also predicted ‘hot-spots’ of aggregation in the rCAL sequence. A single transmembrane domain could be identified in the lectin. This might be the position of anchoring between the lectin and the erythrocyte membrane, leading to agglutination. Since no simple sugar could be identified specifically by the lectin, this can be a possible explanation for the phenomena of agglutination displayed by rCAL. Moreover, it has now become clear that more and more plant lectins are found to recognize complex carbohydrate structures found on animal cells. This might be the reason for agglutination observed with only fetuin and its asialo triantennary glycan.

---

## CONCLUSIONS

To summarize, the main findings from this work are enumerated below:

1. The recombinant chick pea seed lectin, rCAL, could be successfully cloned and expressed in *Escherichia coli* in an active form.
2. The purified lectin, consisting of 462 amino acids, was found to be a homodimer of approximate molecular mass 53,000 Da; sub unit molecular mass was around 27,000 Da.
3. Around 5 to 8 mg of purified lectin could be obtained from one litre of culture with a specific activity of  $5 \times 10^3$  hemagglutination units/mg.
4. No simple sugar specificity could be identified for the lectin; hemagglutination inhibited only by fetuin and its asialo triantennary glycan.
5. Binding of the asialo triantennary fetuin glycan was a spontaneous and enthalpically driven process.
6. Binding of hemin, spermine and thiamine to rCAL postulates physiological role in plant defence against stress conditions.
7. Single tryptophan in the lectin is present in a slightly polar environment and exists as two different conformers at a given time.
8. Solute quenching results reveal the trp environment of the native lectin to be slightly negatively charged. The charge density around the trp alters upon treatment with low concentration of GdnHCl.
9. Quenching of denatured rCAL by acrylamide involves the contribution from both static as well as collisional components.
10. Exposure of hydrophobic residues and increased formation of intermolecular  $\beta$ -sheets, leading to insoluble thermal aggregation at 55 °C.
11. Irreversible unfolding in the presence of GdnHCl; dissociation simultaneously accompanied by unfolding, leading to soluble aggregation of the protein.
12. Renaturation of the protein in low concentration of GdnHCl buffers led to the formation of a miscompacted species with exposed hydrophobic residues.
13. Docking of hemin, spermine and thiamine on the homology model of rCAL identified the residues involved in binding. Asn-8 was found to be crucial for binding.

14. Visualization of the trp microenvironment in the model detected the presence of hydrophobic residues and weak negative charge around the trp; complete lack of positive charge density around the trp correlates with the solute quenching results.
15. Amino acid residues exhibiting propensity towards aggregation were predicted.

**REFERENCES**

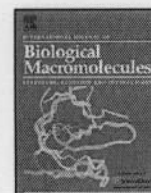
1. Ghazarian, H., Idoni, D. and Oppenheimer, S.B. (2011), A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics, *Acta Histochemica*, **113**, 236–247.
2. Sharon, N. (1980), Carbohydrates, *Scientific American*, **243**, 90–116.
3. Mody, R., Joshi, S. and Chaney, W. (1995), Use of lectins as diagnostic and therapeutic tools for cancer, *Journal of Pharmacological and Toxicological Methods*, **33**, 1–10.
4. Zheng, M., Fang, H., Tsuruoka, T., Tsuji, T., Sasaki, T. and Hakomori, S. (1993), Regulatory role of GM3 gangliosides in alpha 5 beta 1 integrin receptor for fibronectin-mediated adhesion of FUA169 cells, *Journal of Biological Chemistry*, **268**, 2217–2222.
5. Gorelik, E., Galili, U. and Raz, A. (2001), On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis, *Cancer Metastasis Reviews.*, **20**, 245–77.
6. Bies, C., Lehr, C.M. and Woodley, J.F. (2004), Lectin-mediated drug targeting: history and applications, *Advanced Drug Delivery Reviews*, **56**, 425–35.
7. Minko, T. (2004), Drug targeting to the colon with lectins and eoglycoconjugates, *Advanced Drug Delivery Reviews*, **56**, 491–509.
8. Bouckaert, J., Berglund, J., Schembri, M., DeGenst, E., Cools, L. and Wuhler, M. (2005), Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesion, *Molecular Microbiology*, **55**, 441–55.
9. Rabinovich, G.A. (1999), Galectins: an evolutionarily conserved family of animal lectins with multifunctional properties; a trip from the gene to clinical therapy, *Cell Death and Differentiation*, **6**, 711–21.

10. Singh, H. and Sarathi, S.P. (2012), Insight of Lectins- A review, *International Journal of Scientific and Engineering Research*, **3**, 1–4.
11. Kobayshi, M., Fitz, L., Ryan, M., Hewick, R.M., Clarck, S.C., Chan, S., Loudon, R., Sherman, F., Perussiaj, I.B. and Trinchierit, G. (1989), Identification and purification of Natural Killer cell stimulatory factor (NKSF), A cytokine with multiple biologic effects on human lumphocytes, *Journal of Experimental Medicine*, **170**, 827–845.
12. Alroy, J., Orgad, U., Ucci, A.A. and Pereira, M.E. (1984), Identification of glycoprotein storage diseases by lectins: a new diagnostic method, *Journal of Histochemistry and Cytochemistry*, **32**, 1280–1284.
13. Lis, H. and Sharon, N. (1989), Lectins as molecules and tools, *Annual Review of Biochemistry*, **55**, 35-67.
14. Sharon, N. and Lis, H. (1989), Lectins as cell recognition molecules, *Science*, **246**, 227-224.
15. Yoshihara, Y., Mizuno, T., Nakahira, N., Kawasaki, M, Watanabe, Y., Kagamiyama, H., Jishage, K.I., Ueda, O., Suzuki, H., Tabuchi, K., Sawamoto, K., Okano, H., Noda, T. and Mori, K. (1999), A Genetic Approach to Visualization of Multisynaptic Neural Pathways Using Plant Lectin Transgene, *Neuron*, **22**, 33–41.
16. Gary, R., Gunther, J., Wang, and Edelma, G.M. (1974), The Kinetics Of cellular commitment during stimulation of the lumphocytes by lectins, *Journal of Cell Biology*, **62**, 366–377.
- 17.. Rhodes, E.G. (1998), Combined lectin/monoclonal antibody purging of bone marrow for use in conjunction with autologous bone marrow transplantation in the treatment of multiple myeloma, *Methods in Molecular Medicine*, **9**, 351–361.

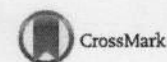


18. Wang, Y., Wu, S.L., and Hancock, W.S. (2006), Monitoring of glycoprotein products in cell culture lysates using lectin affinity chromatography and capillary HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry (LTQ/FTMS), *Biotechnology Progress*, **22**, 873–880.
19. Singh, K.B. and Ocampo, B. (1997), Exploitation of wild Cicer species for yield improvement in chickpea, *Theoretical and Applied Genetics*, **95**, 418–423.
20. Hulse, J.H. (1991), Nature, composition and utilization of grain legumes, In: *Uses of tropical legumes: Proceedings of a Consultants' Meeting*, 11–27, ICRISAT Center, ICRISAT, Patancheru, Andhra Pradesh, India.
21. Duke, J. A. (1981), *Handbook of legumes of world economic importance*, Food and Agriculture Organization of the United Nations, 52–57, Plenum Press, New York.
22. Sánchez-Vioque, R., Clemente, A., Vioque, J., Bautista, J. and Millán, F. (1999), Protein isolate from chickpea (*Cicer arietinum* L.): Chemical composition, functional properties and protein characterization, *Food Chemistry*, **64**, 237–243.
23. Tiwari, U., Gunasekaran, M., Jaganmohan, R., Alagusundaram, K. and Tiwari, B.K. (2009), Quality characteristic and shelf life studies of deep-fried snack prepared from rice brokens and legumes by product, *Food and Bioprocess Technology*,
24. 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 Crystallographica F.*, **61**, 141–143.
25. Vigeolas, H., Chinoy, C., Zuther, E., Blessington, B., Geigenberger, P. and Domoney, C. (2008), Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds, *Plant Physiology*, **146**, 74–82.

26. Adamek-Swierczynska, S. and Kozik, A. (2002), Multiple thiamine-binding proteins of legume seeds: thiamine-binding vicillin of *Vicia faba* versus thiamine-binding albumin of *Pisum sativum*, *Plant Physiology and Biochemistry*, **40**, 735-741.
27. Gill, S.S. and Tuteja, N. (2010), Polyamines and abiotic stress tolerance in plants, *Plant Signaling and Behavior*, **5**, 26–33.
28. Tunc-Ozdemir, M., Miller, G., Song, L., Kim, J., Sodek, A., Koussevitzky, S., Misra, A.N., Mittler, R. and Shintani, D. (2009), Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis, *Plant Physiology*, **151**, 421–432.



## Solution and *in silico* studies on the recombinant lectin from *Cicer arietinum* seeds



Madhurima S. Wakankar<sup>a</sup>, Musti V. Krishnasastri<sup>c</sup>, Tulika M. Jaokar<sup>a</sup>, Krunal A. Patel<sup>b</sup>,  
Sushama M. Gaikwad<sup>a,\*</sup>

<sup>a</sup> Division of Biochemical Sciences, National Chemical Laboratory, Pune, Maharashtra, 411 008, India

<sup>b</sup> Division of Plant Tissue Culture, National Chemical Laboratory, Pune, Maharashtra, 411 008, India

<sup>c</sup> National Centre for Cell Science, Pune, Maharashtra, 411 007, India

### ARTICLE INFO

#### Article history:

Received 25 December 2012

Received in revised form 6 February 2013

Accepted 7 February 2013

Available online 26 February 2013

#### Keywords:

Cicer lectin

BL21-CodonPlus (DE3)-RIL cells

Steady-state and time resolved

fluorescence

Unfolding

Aggregation

Homology model

### ABSTRACT

The *Cicer arietinum* seed lectin was cloned and expressed in *Escherichia coli* and purified in active form. Conformational characterization of the recombinant lectin (rCAL) was performed using biophysical and bioinformatics tools. Thermal denaturation of rCAL caused rapid secondary structural rearrangements above 50 °C and transient exposure of hydrophobic residues at 55 °C, leading to aggregation. Treatment of rCAL with GdnHCl resulted in unfolding followed by dissociation of the dimer. The single tryptophan in rCAL present on the surface of the protein is surrounded by hydrophobic and acidic amino acids and exists as different conformers. The experimental observations correlated well with the structural information revealed from the homology model of rCAL.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Lectins are a class of oligomeric proteins that bind carbohydrates specifically and reversibly and have found many applications in detection, isolation, and characterization of glycoconjugates, drug targeting, or cell sorting [1]. Lectins also have made tremendous contribution in increasing our understanding of protein chemistry. In spite of the abundance of equilibrium denaturation studies on the folding/unfolding of monomeric globular proteins [2], similar studies on oligomeric proteins are less common [3]. Inter-subunit interactions significantly impact the stability of oligomeric proteins, whereas intra-molecular interactions determine the stability of monomeric proteins. Legume lectins serve as attractive models for studying the folding process of oligomeric proteins [4]. Several legume lectins have been characterized for their unfolding behavior in presence of denaturing agents like temperature, urea or guanidine hydrochloride [5,6].

A lectin from the seeds of *Cicer arietinum* (chick pea) showing complex sugar specificity was purified and characterized to some extent in our laboratory [7]. The N-terminal amino acid sequence of this lectin showed 90% identity with the first 25 amino acids of the major seed albumin (PA2) from *Pisum sativum*. In this paper, we describe the cloning and expression of *C. arietinum* seed lectin in *Escherichia coli* using primers designed from the sequence of PA2 albumin. Purified recombinant lectin (rCAL) showed hemagglutinating activity with pronase-treated rabbit erythrocytes similar to that of wild type seed lectin. To our knowledge, this is the first report on conformational studies of the recombinant chick pea seed lectin and subsequent correlation of the experimental results with the topology of the lectin model using bioinformatics tools.

## 2. Materials and methods

### 2.1. Materials

Cultivar BDN9-3 of chick pea seeds (*C. arietinum* L.) was obtained from Badnapur Agricultural University, Jalna, India. Restriction enzymes NcoI and EcoRI were from New England Biolabs, USA. pET 28a<sup>+</sup> vector was from Novagen. *E. coli* BL21-CodonPlus (DE3)-RIL cells were obtained from Stratagene, USA. Kanamycin, Chloramphenicol, IPTG, DEAE-Cellulose matrix, Pronase and all the

Abbreviations: rCAL, recombinant *Cicer arietinum* lectin; GdnHCl, guanidine hydrochloride; ANS, 8-anilino-1-naphthalenesulphonate.

\* Corresponding author. Tel.: +91 2025902241; fax: +912025902648.

E-mail address: [sm.gaikwad@ncl.res.in](mailto:sm.gaikwad@ncl.res.in) (S.M. Gaikwad).