INVESTIGATION INTO THE STRUCTURES OF A LECTIN FROM *TRICHOSANTHES DIOICA* AND ONE FROM *ERYTHRINA INDICA* & BIOPHYSICAL CHARACTERIZATION OF ARACEAE LECTINS FROM *SAUROMATUM GUTTATU*M AND *ARISAEMA TORTUOSUM*

THESIS SUBMITTED TO

THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

BY

POORVA DHARKAR

DIVISION OF BIOCHEMICAL SCIENCES
NATIONAL CHEMICAL LABORATORY
PUNE 411 008
INDIA

April 2009

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Investigation into the structures of a lectin from *Trichosanthes dioica* and one from *Erythrina indica* & Biophysical characterization of araceae lectins from *Sauromatum guttatu*m and *Arisaema tortuosum*."** submitted by **Ms. Poorva Dharkar** was carried out under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

Date

Dr. C. G. Suresh
(Research Supervisor)
Scientist
Division of Biochemical Sciences
National Chemical Laboratory
Pune 411 008

DECLARATION

I hereby declare that the thesis entitled **"Investigation into the structures of a lectin from *Trichosanthes dioica* and one from *Erythrina indica* & Biophysical characterization of araceae lectins from *Sauromatum guttatum* and *Arisaema tortuosum*."** submitted by me to the University of Pune for the degree of Doctor of Philosophy is the record of original work carried out by me under the supervision of **Dr. C. G. Suresh** at the Division of Biochemical Sciences, National Chemical Laboratory, Pune 411008 and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or Institution.

I further declare that the material obtained from other sources has been duly acknowledged in the thesis.

Poorva Dharkar **Date**

Division of Biochemical Sciences

National Chemical Laboratory

Pune 411008

Dedicated to.....

Mamma

Papa-Amma

&

Aji

CONTENTS

Chapter 3: Crystallization and X-ray crystallographic studies on a lectin from Trichosanthes dioica.

Chapter 4: Biophysical and stability studies on two *Trichosanthes* lectins.

Chapter 5: Crystallization and X-ray crystallographic studies of a Galactose/N-acetyl Galactosamine specific lectin from *Erythrina indica*

Chapter 6: Comparative biophysical studies on two araceae lectins from Sauromatum guttatum & Arisaema tortosum by Fluorimetric methods and Circular dichroism (CD).

ACKNOWLEDGEMENTS

In my journey of achieving a doctorate, many people have been of tremendous help, support and encouragement. I take this opportunity to thank all of them, though I might forget to mention a few of them.

I consider myself very fortunate to get an opportunity to work under Dr. C.G. Suresh. From him I have not only learned the working etiquettes and fundamentals of crystallography in great details but have also learned the art of doing things patiently and to perfection. Throughout this long period of research he has been extremely supportive to me in innumerable ways. He has always allowed me to explore science in the way I wanted and has always extended his full support and given ideas wherever required. Sir and Indu madam have both been very affectionate and it has always given me a very cozy feeling, besides Shruti & Neethi have always given a very cheerful atmosphere. I fall out of words to express my thanks to all of them.

Our collaboration with Sushama madam has been scientifically very fruitful and it has also extended to friendly terms. I will always be grateful to her for whatever she has done for me throughout my research.

I take the opportunity to thank Dr. M.I. Khan for his constant guidance and for giving valuable suggestions which helped me to go ahead and try out different experiments. Again I thank him and Dr. Vidya Gupta for being my Ph.D. committee members, Dr. Jayant Pal from the Department of Biotechnology, University of Pune for being in my Ph.D assessment committee and giving his timely suggestions.

I express my heartfelt gratitude to Dr. Eleanor Dodson from YSBL, UK who helped me solve problems related to the structural work and promptly provided her most valued help in spite of her very busy schedule. I also thank Dr. Guy Dodson for his encouraging suggestions.

Our collaborators, Dr. SV Bhide and Dr. Rakesh Kestwal have provided me pure *Erythrina indica* lectin and I thank them sincerely. I also express my sincere thanks to Dr. Singh and Dr. Vikram Dhuna for providing purified SGA and ATL and giving a free hand to conduct experiments on those lectins.

I am grateful to Dr. Shama Barnabas for her time to time suggestions and making the atmosphere in our lab very cheerful with her esteemed presence. I am thankful to Dr. Shirgurkar for providing me with tamarind seeds.

I sincerely thank Dr. M. V. Badiger (Division of Polymer Sciences, NCL); Dr. MRN Murthy (MBU, IISc, Bangalore) for allowing me to use the Fluorescence and CD facility in their respective departments. I thank Gayathri and Sonali for being supportive during these experiments.

I take the opportunity to thank all the scientists and students of biochemical division, NCL for their help whenever required.

My labmates have always provided me with a very comfortable and positive feeling in lab, by always extending their help in all the possible ways. They have always provided a very cheerful atmosphere along with being great friends. I take the opportunity to thank Anuradha, Manish, Satya, Priya, Suresh Kumar, Uma, Urvashi, Nishant, Manas, Tulika, Payal and Srijith. Special thanks are due for Uma, with whom I have spend some very good time in hosteland who was a good company always; and to Suresh kumar for helping me in every possible way and being a close friend. I especially thank Manas, Payal and Tulika for their extensive efforts in correcting my thesis with a friendly and sincere attitude. I will always remember their help in this crucial period. I also thank the people who helped me even though they were associated with

our lab for short duration: Raamesh, Nitin, Ashish, Ratnarekha, Mithila and Sharad. I am grateful to Ashish for his help in conducting experiments.

I will always remember and I thank all those people staying in and around NCL for being good friends and helping me in need. They include my very close friends Abhishek, Sharmili, Feroz, Sambhaji, Sudarshan, Anand, Sridevi, Ashwini, Vinod and Anamika. I take the opportunity to thank Abhishek for providing me with his laptop in my last and most important phase of thesis writing and for being a great company always. Shambhaji and Sharmili have been very good and ever supportive friends.

Some other friends who have been very helpful are Atul Fairazi, Deepti, Ambrish, Shabab, Sachin, Ashwini Poopal, Sofia, Gayatri, Radhika, Gyanprakash, Aarohi, Anish, Ajit, , Sreekanth, Sajid, Ansari, Asad, Jayprakash, Ashutosh, Shashi, Atul Thakur, Nagraj, Avinash, Sathish, Ashwini, Gayathri, Vaijayanti, Radhika, Rashmi, Gowri and Shilpa.

My school friends Paridhi, Manish, Indrayani and Riddhi have always encouraged me and have given good refreshment to me in their wonderful company from time to time. I will always remember our good times together. I also thank a few friends Shivani, Tamali, Shankar, Priyanka, Partho, Anuja, Mukul, Suchi, Wanu, Prabhu, Swapna and Mandar who lifted my spirits up.

I acknowledge the help provided by Takalkar madam, Ph.D department, University of Pune for making the university process very smooth and convenient. I also thank Ms. Amrita, Mrs. Indira, Mr. Mari and Ms. Satyali for their help in administrative work. Mr. Karanjkar and Mr. Trehan are acknowledged for their technical assistance.

I acknowledge the Director, NCL; Head Biochemical Division for providing me with the working space together with all the facilities.

My family members have been my greatest strength and support to me all the time. I owe my doctorate to them. It is only because of the positive backup; understanding and encouragement provided by my father-in-law and father, the loving care and sweet support of both mothers that I could achieve this degree. It is hard to describe in words the role of my husband Nachiket at every step of my work; understanding all my problems, helping me in every possible way and making it possible to reach the end successfully. My siblings Avanti & Satyakam have been extremely sweet, understanding and supportive all the time and I owe a lot to them. I am grateful to all my very close relatives who have been with me throughout this period and who have also made my stay in Pune a memorable one.

ABSTRACT

Lectins belong to a group of proteins that bind free sugar or sugar residues of polysaccharides, glycoproteins, or glycolipids, which exist in free or bound form (as in cell membranes). Plant lectins have been attracting much attention because of their ease of isolation and their usefulness as reagents for detecting glycoconjugates in solution and on cell surfaces. The recognition of carbohydrate moieties by lectins has important implications to a number of biological processes such as in cell-cell interactions, signal transduction and in cell growth and differentiation. A large number of lectins have been reported from many plant species but correlation between structure, conformation and stability has been studied only for a few of them.

Cucurbitaceae lectins have been purified and characterized in some detail from plants *Momordica charantia* (bitter gourd), *Trichosanthes anguina* (snake gourd), *Trichosanthes kirilowii* (Chinese Snakegourd) and *Trichosanthe cucumerina.* Ribosome Inactivating Proteins (RIP's) belong to two different classes (RIP type-I & II) and are found ubiquitously in plant kingdom. The cucurbitaceae lectins from *Trichosanthes dioica* and *Trichosanthes anguina* have been assumed to belong to type-II RIP family with respect to their secondary structural elements, biophysical characteristics, carbohydrate specificity as well as subunit composition. In the study reported here these lectins were biophysically characterized with respect to thermal, chemical and pH stability. One of them, the lectin from *T. dioica* has been structurally characterized.

Another lectin that was taken up for structural studies is Galactose / N-acetyl Galactosamine specific lectin from *Erythrina indica.* Many Galactose / N-acetyl Galactosamine specific lectins have been isolated from various species of the genus *Erythrina* belonging to sub-family Fabaceae of Leguminoseae. They include lectins from *E. cristagalli* , *E. indica*, *E. costarensis*, *E .variegata*, *E. arborescens, E. lithosperma and E. suberosa, E. speciosa,* and *E. velutina*. Studies on lectin from the seeds of *Erythrina corallodendron* has been reported in the literature. The *Erythrina indica* seed lectin was obtained in pure form for structure determination through collaboration. Thus, lectins mainly from *Trichosanthes dioica*, and *Erythrina indica* were selected for structural investigation for their importance as being present in commonly used edible fruits or stems.

In the past few years, araceae has been identified as a lectin-rich plant family with lectins constituting 70-80% of storage proteins in their tubers. Although, extensive studies on three-dimensional structures of lectins and lectin-carbohydrate interactions have been carried out, till date no report of folding-unfolding as well as fluorescent spectroscopic characterization of monocot lectins belonging to araceae are available. Here we report biophysical characterization of the *Sauromatum guttatum* (Voodoo lily) *Arisaema tortosum* (Himalayan cobra lily) lectins by using steady-state and time resolved fluorescence and CD spectroscopy.

Organization of the chapters in this thesis is as follows:

Chapter 1: General Introduction

Chapter 2: Materials and Methods

Chapter 3: Crystallization and X-ray crystallographic studies of a lectin from *Trichosanthes dioica*.

Chapter 4: Biophysical and stability studies on two *Trichosanthes* lectins.

Chapter 5: Crystallization and X-ray crystallographic studies of a Galactose/N-acetyl Galactosamine specific lectin from *Erythrina indica*.

Chapter 6: Comparative biophysical studies on two araceae lectins from *Sauromatum guttatum* and *Arisaema tortosum*, by fluorimetric methods and circular dichroism (CD).

Chapter 7: Comparisons of plant lectin characteristics studied and conclusions.

Chapter 1: General Introduction

This introductory chapter of the thesis describes the studies on various lectins from plant, animal and microbial sources with respect to their biomedical application, their sugar specificity and three-dimensional structures. Besides this general overview, it also provides a brief outline of the three plant lectin families from which five lectins have been chosen for detailed study, through structural investigations as well as comparative biophysical studies described in the thesis.

Chapter 2: Materials & Methods

This chapter presents the details of the materials and various methods used for the purification of *Trichosanthes dioica* lectin to obtain good diffraction quality crystals. It also includes the methodology followed for crystallization, X-ray data collection, data processing, structure determination, structure refinement and analysis of the refined structure of both the *Trichosanthes dioica* and *Erythrina indica* lectins. The biophysical

experiments conducted for the comparison of *T. dioica* and *T. anguina* lectins as well as for lectins from *S. guttatum* and *A. tortosum* have been described.

Chapter 3: Crystallization and X-ray crystallographic studies of a lectin from *Trichosanthes dioica*

The lectin from *T.dioica* seeds has previously shown to be similar to type-II RIPs on the basis of its biochemical characteristics. Another lectin from the cucurbitaceae family, the *T.kirilowii* lectin (TKL-1) has been structurally characterized and it has also been placed in this plant lectin family. The structure of TDSL has been determined at 2.8Å, and its structure has been solved by molecular replacement using the coordinates of TKL-1 (PDB code: 1GGP). The structure has been refined using the same model, but better values of R_{factor} and R_{free} have not been obtained in the absence of sequence information. Hence a well refined structure could not be obtained. The structure shows close similarity with type-II RIPs such as abrin-a and ricin, but like TKL-1 it lacks the residue Tyr74 (of abrin-a) that is essential for adenine binding and thus the RIP activity. While the active site residues Glu164 and Arg167 are also observed in TDSL. All other active site residues and the invariant residues that have been described in all other type-II RIPs are also conserved in TDSL. On the basis of these structural features, TDSL has been categorized as type-II RIP.

Chapter 4: Biophysical and stability studies on two *Trichosanthes* lectins

In this chapter a comparative account of the biophysical characterization of lectins from *Trichosanthes dioica* and *Trichosanthes anguina* with respect to their thermal, chemical and pH stability has been described. The thermal denaturation studies of these two lectins indicate that both these cucurbitaceae lectins show high thermal stability,

although the disulphide linkages do not play any role in the stability of these proteins. Some RIPs have been shown to be highly thermostable. Since the *T.dioica* lectin resembles RIPs in molecular weight, subunit composition, size, and sugar specificity, it may belong to this family. As it has been pointed out that some *Trichosanthes* lectins have similarity with RIPS; here we show that like some RIPs, these lectins are resistant to wide range of chemical and pH denaturing conditions.

Chapter 5: Crystallization and X-ray crystallographic studies of a Galactose/N-acetyl Galactosamine specific lectin from *Erythrina indica*

This chapter describes the structural study of the lectin from the seeds of *E.indica* (coral tree). The lectin belongs to the large family of legume lectin. The structure has been determined at 2.5 Å and solved by molecular replacement using the coordinates of EcorL (PDB code: 1FYU). EiSL is a homodimeric lectin and *N*-linked glycosylation was observed at four sites in the structure *i.e.* 17 and 113 in both the chains. A heptasachharide at B17 and a hexasaccharide at B113 were traced properly in the electron density; while A-chain at both locations showed a trisachharide comprising of Nag and Fuc residues. Structurally EiSL varies from the very closely related lectins of the same genus EcorL and ECL, with respect to a few amino acid residues. More than 90% residues have been shown to be conserved. Similar folds, conformation of loops, type of coordination bonds at the metal binding sites and residues at the sugar combining sites have been observed in EiSL as in EcorL and ECL. A galactose molecule was bound at a position that is close to the glycosylation site in a-chain (A 116 Asp) but it has been observed to be facing away from the N-linked heptasaccharide.

Chapter 6: Comparative biophysical studies on two araceae lectins from *Sauromatum guttatum* and *Arisaema tortosum*, by fluorimetric methods and circular dichroism (CD).

This chapter presents a comparative study of the structural stability and dynamics of two araceous lectins, one from *Sauromatum guttatum* (Voodoo lily) (SGA), and another from *Arisaema tortosum* (Himalayan cobra lily) (ATL). Among the Araceae lectins those from *Arisaema tortosum* schott, *Arisaema consanguineum* Schott (ACA), *A. curvature* Kunth (ACmA), *Gonatanthus pumilus* (GPA), *Sauromatum guttatum* Schott (SGA) and Alocasia *cucullata* have been purified and characterized. All these lectins show mitogenic potential for human blood lymphocytes and have complex sugar specificity, and show hemagluttination inhibition by desialylated fetuin. They are thus distinctly different from other monocot lectins in carbohydrate binding specificity.

Both the lectins chosen for the study are non-mannose binding monocot lectins isolated from the tubers of the respective araceous plants. SGA and ATL consist of a mixture of isolectins differing in charge, similar to *Alocasia indica* lectin, WGA and those from Amaryllidaceae and Alliaceae. Recently, SGA as well as some lectins from Arisaema sp. have been found to show antiproliferative as well as mitogenic activity.

The two monocot lectins have been characterized with respect to their tryptophan environment and secondary structure changes against chemical, pH and thermal denaturation. Transitions in the tryptophan microenvironment and secondary structure changes were studied using steady state and time resolved fluorescence and CD spectroscopy. The lectins exist as tetramers with an estimated single tryptophan residue per monomer in a polar environment. Quenching with ionic quenchers showed

predominantly electropositive environment for tryptophan residues. Acrylamide had maximum quenching effect. A decrease in KI quenching due to lectin denaturation indicated redistribution of charges as a result of possible conformational changes. The two values for lifetimes of tryptophanyl population (1.2-1.4 and 6.3-6.4 ns) reduced substantially on quenching or denaturation. Similarly, both the lectins showed a drastic loss of secondary structure in 5M Gdn-HCl or 6M Urea or at pH 2.0 and below. For the first time araceous lectins, like some of the legume lectins, are shown to bind adenine. The presence of a compact structure at highly alkaline pH (10.0 - 12.0) was observed in CD spectra.

Chapter 7: Comparisons of plant lectin characteristics studied and conclusions

This chapter outlines the comparative analysis of the structures, the sugar-binding of lectins, as well as the biophysical characteristics of lectins that have been studied in this thesis, belonging to three structurally different plant lectin families.

Comparison of the sugar binding sites of Galactose/*N*-acetyl Galactosamine specific lectin TDSL and EiSL was done with lectins having similar sugar specificity. The structural features of the active site of *T. dioica* lectin was compared with binding sites of galactose specific lectins, a type-II RIP Abrin-a and a C-type animal lectin CEL-III from *Cucumaria echinata*. The active site geometry and residues involved in sugar binding for EiSL have been compared with that of Jacalin, mouse galectin-9, a tunicate lectin DCL-I from *Didemnum candidum* and C-type lectins.

The three-dimensional structure of *E. indica* lectin was similar to that of lectins from *E. cristagalli* and *E. corallodendron* in terms of the N-linked heptasacchride,

conserved water molecules and residues in the sugar binding site. Minor structural differences have been highlighted and compared to show the higher stability of this lectin as compared to EcorL.

We have for the first time carried out structural investigations on the two araceaous lectins from *S. guttatum* and *A. tortuosum,* to probe the tryptophan environment and secondary structure of the members of this monocot family. For both these lectins the far-UV CD spectra indicate a predominantly β sheet structure and type-III β-turns. Adenine-binding, a feature of legume lectins has been observed in members of this family for the first time. The two lectins belonging to araceae and differing from other mannose specific lectins show similarity with respect to their tryptophan environments as well as structural stability with respect to temperature, pH as well as presence of chemical denaturants.

LIST OF ABBREVIATIONS

(List of abbreviations and standard notations used in the text)

ACA	*Arisaema consanguineum* Schott agglutinin
ACmA	*A. curvature* Kunth agglutinin
AIDS	Aquired immune deficiency syndrome
AMoRe	Automated Molecular Replacement
ANS	8-anilino-1-napthalene sulfonic acid
AS	Ammonium sulfate ($(NH_4)_2SO_4$)
ATL	*Arisaema tortuosum* lectin
AU	Absorption Unit
BL	Binding loop
BME	2-mercaptoethanol (β-mercaptoethanol)
BPB	Bromo-Phenol Blue
BSA	Bovine serum albumin
CBD	Carbohydrate Binding Domain
Cc	Correlation coefficients
CCD	Charge-coupled device
CCP4	Colloborative Computational Project No. 4
CD	Circular Dichroism
CPB	Citrate-phosphate buffer
CRD	Carbohydrate Recognition Domain
Cs+	Cesium ion
ConA	Concanavalin A
CsCl	Cesium Chloride
DCL	*Didemnum candidum* lectin
DDW	Double distilled water
DEAE	Diethylaminoethyl
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DTNB	5,5'-Dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol
ECL	*Erythrina cristagalli* lectin
EcorL	*Erythrina corallodendron* lectin
EGF	epidermal growth factor
EiSL	*Erythrina indica* seed lectin
ELLA	Enzyme-linked lectin binding
ERGIC-53	ER-Golgi intermediate compartment
FFT	Fast Fourier Transform
FOM	Figure of merit
Gal	Galactose
Gal-1	Galectin -1
GalNAc	N-acetyl-D-galactosamine
GdnHCl	Guanidium hydrochloride
Glc	Glucose

GlcNAc	N-acetyl-D-glucosamine
GNA	*Galanthus nivalis* agglutinin
GPA	*Gonatanthus pumilus* D. Don
GS-IB4	*Griffonia simplicifolia* lectin
HBP	Hepatic binding protein
HCl	Hydrochloric acid
HmRIP	Himalayan mistletoe ribosome inactivating protein
HPA	*Helix pomatia* agglutinin
HPLC	High performance liquid chromatography
HXT	1,2,6- Hexanetriol
IEF	Isoelectric focusing
IL	Interleukin
IP	Image plate
IU	International Unit
K Da	Kilodaltons
K_a	Association constant
kcal/mol	Kilo Calories per mole
KI	Potassium Iodide
K_q	Bimolecular quenching constant
K_{sv}	Stern-Volmer quenching constant
L	Liter
LacNAc	N-acetyl-D-lactosamine
Man	Mannose
Man-6-P	Mannose-6-Phosphate
MBP	Mannose binding Protein
MIR	Multiple isomorphous replacement
MPA	*Macula pomifera* agglutinin
MPD	2-methyl-2,4-pentanediol
MR	Molecular replacement
MW	Molecular weight
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NBS	*N*-Bromosuccinimide
NCBI	National Center for Biotechnology Information
NCS	Non-crystallographic symmetry
nm	nanometers
NMR	Nuclear Magnetic Resonance
ns	nanoseconds
OD	Optical Density
OD_{280}	Optical density at 280 nm
PAGE	Polyacrylamide gel electrophoresis
PB	Phosphate buffer
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
PHA	Phytohaemagglutinin

PNA	Peanut agglutinin
PWM	Pokeweed Mitogen
r.m.s.d	Root Mean Square Deviation
RBC	Red blood cells
RCA	*Ricinus communis* agglutinin
RIP	Ribosome inactivating protein
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RSA	*Rhizoctonia solani* agglutinin
SAP	Serum Amyloid P component
SBA	Soybean (*Glycine max*)seed lectin
SCAfet	Fetuin-binding lectin from *Scilla campanulata* bulbs
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SGA	*Sauromatum guttatum* lectin
TDSL	*Trichosanthes dioica* seed lectin
T_m	Melting temperature
TF	Translation Function
Tris	Tris-hydroxylmethyl amino methane
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultra violet
v	volume
VPI	Viral Protein
w	Weight
WGA	Wheat germ agglutinin

α	alpha
β	beta
γ	gamma
µmol	Micro mole
µg	microgram
µl	microliter
µm	micrometer
µM	micromolar
Å	Angstrom
C	Degree centrigtrade
Da	Dalton
G	Gram
H/h	hour
K	degree Kelvin
Ksv	Stern-Volmer constant
M	meter
M	Molar

Mm	milli meter
Nm	nanometer
U	Unit
Vm	Mattew's number
λ	Lambda/Wave length
∑	sigma/Sum
X	Chi
Ψ	Psi
Φ	Phi

Chapter 1

INTRODUCTION

INTRODUCTION

1.1 Genesis of the thesis

Interaction between protein and carbohydrate form the crux of many important biological processes. A significant class of carbohydrate binding proteins is lectins, which are ubiquitously present in organisms and play an important role in mediating biological processes that require recognition of a carbohydrate. Plant lectins are a large family among lectins and the work on this thesis is focused on three such plant lectins families: the legume lectins; type-II RIPs and monocot mannose binding lectins. The study in this thesis includes the structural study of two Gal/Gal-*N*-acetylgalactosamine specific lectins: *T.dioica* seed lectin from cucurbitaceae and belonging to the type-II RIPs; *E.indica* seed lectins leguminoseae and belonging to legume lectins. Biophysical characterization of type-II RIP family lectins from *T.dioica* and *T.anguina* and Araceaous lectins from *S.guttatum* & *A.tortuosum* belonging to the mannose binding lectin family. These lectins have been characterized with respect to the aspects mentioned above and their biophysical and structural features are compared with those of already reported lectins; conclusions have been drawn that might be useful for further investigations.

1.2 Lectins

Lectins are a heterogenous group of proteins, present ubiquitously in nature, which act as mediators of cellular recognition in a variety of systems (Lis & Sharon, 1998). Understanding the molecular basis of lectin-carbohydrate interactions is of interest for the design of novel drugs for treatment of infectious, inflammatory and malignant diseases as Cancer and AIDS. It may also be of help for the development of new application of lectins for preparation as well as structural

and functional investigation of glycoconjugates, for examining changes that occur on cell surfaces during physiological and pathological processes, from cell differentiation to cancer.

There is a widespread occurrence of carbohydrate binding proteins in nature. The pioneering work carried out by Watkins and Morgan (1952) gave a new direction to modern glycobiology as they identified sugars present on cell surfaces and which have potential application as markers of specific types of cells. Recent development in protein chemistry has focused more in the identification and targeting of these markers on the surface of pathogens as well as the affected cells, as a potential method for treatment of diseases for human welfare. Progress in the area of glycobiology has revealed that lectins and other carbohydrate binding proteins specifically recognize and bind to oligosaccharides present on the surface of cells, and thus play a prominent role in mediating cellular recognition events. Thus the lectin–carbohydrate interaction being very specific and widespread has an enormous potential for exploitation of these molecules in the development of drugs. Lectins have thus been a major area of research since the past century and research in the last decade has focussed more on their application as important biological molecules. Current research is focused on identification of new lectin molecules, as well as for getting a deeper understanding of their functional aspect, characterizing the structure of lectins, their carbohydrate binding specificity, conformational/functional properties of their carbohydrate- binding 'pocket', mechanism of association with specific ligand molecules, and their biological roles.

1.2.1 Discovery, Definition & Applied aspect

Lectinology includes the study of lectins considering two different aspects. First being the

general study of lectins, their structural and functional characterization, secondly towards the application of lectins as tools to study the function of glycoconjugates for human welfare.

The Beginnning of Lectinology

The word Lectin has been derived from the latin word *lectus* (the past tense of *legere*) which means to select or to chose was introduced in 1954 by William C. Boyd (Boyd & Shapleigh, 1954). The history of discovery of lectins dates back to some 100 years, when these were discovered in plants.

- First evidence for the presence of lectins came approximately 100 years back with the finding that a substance capable of agglutinating erythrocytes was present in plants.
- Stillmark (1888) in his doctoral work isolated the proteinaceous agglutinating factors, and referred to them as hemagluttinins or phytoagglutinins, as they were isolated from plants. He specifically worked on one such hemagluttinin from the seeds of *Ricinus communis* (Castor Bean) which was highly toxic, and he later on termed it as Ricin.
- H. Hellin (1891) demonstrated the presence of another toxic hemagluttinin in jequirity bean (*Abrus precatorius*) soon after Stillmark's discovery of Ricin.
- Paul Ehrlich (1891), an imminent immunologist, working at the Royal Institute of Experimental Therapy (Frankfurt) showed that these agglutinins, ricin and abrin could be used for immunological purposes. He had used both these lectins as antigens then, and later in 1890's he demonstrated on the basis of these lectins, the fundamental immunological characteristics such as specificity of antigenic response, how humoral response is passed on from parents to offspring and immunological memory.

- James Sumner working in the Cornell University in 1918 isolated a hemagluttinin from jack bean (*Canavalia ensiformis*) which was named as Concanavalin A (ConA).
- Sugar specificity of lectins: It was shown for the first time in ConA by Sumner and Howell (1936) that hemagluttination activity is inhibited by sugars such as sucrose. They also showed that ConA agglutinates erythrocytes and yeast cells.
- Discovery of Mitogenicity of lectins: The fact that lectins can stimulate lymphocytes to undergo mitosis was established first in Phytohemaglutinin (agluttinin from red kidney beans *Phaseolus vulgaris*) by Peter C. Nowell in 1960. Soon after it was demonstrated in ConA that mitogenic stimulation is due to the binding of lectins to sugars on cell surface.
- Malignant cells are agglutinated by lectins: It was shown first by J.C. Aub *et al.* (1965) in case of WGA, and later on by Burger & Martin (1972) in case of ConA and by N. Sharon using SBA (1966) that lectins possess the ability to distinguish cancerous cells from normal ones.
- Stockert *et al.* (1974) described the first lectin from mammals called as the ‘‘mammalian hepatic lectin’’, an agglutinin from rabbit liver that agglutinated human erythrocytes.
- Lectin purification by affinity chromatography: The pace of lectin purification was improved, after the isolation of conA on immobilized dextran as reported for the first time by Agrawal & Goldstein (1965).
- Although the first published record of lectin activity was by Stillmark (1888), Weir Mitchell (1860) was the first to observe aggutinination activity in rattle snake venom, which was thus the first observed record of lectin activity and also the first animal lectin activity observed.
- But the detailed work on the snake venom lectins was done much later. The first snake venom lectin isolated in purified form was from Bothrops atrox, a lectin called as thrombolectin (Hume & Weidemann, 1980). Snake venom lectins are useful as reagents to investigate platelet function.

- Alfred Gottschalk (1965) showed that in case of viral infection by the influenza virus, the viral hemagglutininin recognizes the host cell receptors and thus causes the virus to bind to the host cells.
- Research in 1970 implicated an important biological role for lectins as cell recognition molecules, being present inside cells, on the surface of cell and in physiological fluids (Sharon *et al.*, 1974).

Definition of Lectins

The term phytohemagluttinin or hemaglutinin was given to that class of proteins that were discovered in the last decade of 19th century, and were able to agglutinate erythrocytes. Ricin and Abrin belonged to this class of proteins. Later on they were shown to be specific to sugars. They were shown to be present mostly in plants, to some extent in animals, microbes, bacteria, fungi and viruses also. These widespread groups of proteins that could agglutinate erythrocytes and some other cells such as yeast, and which could bind to sugars have been termed as **Lectins** by Boyd as described earlier. After a few years of lectin related research, the definition was modified as: Lectins are multivalent carbohydrate-binding proteins or glycoproteins except for enzymes and antibodies.

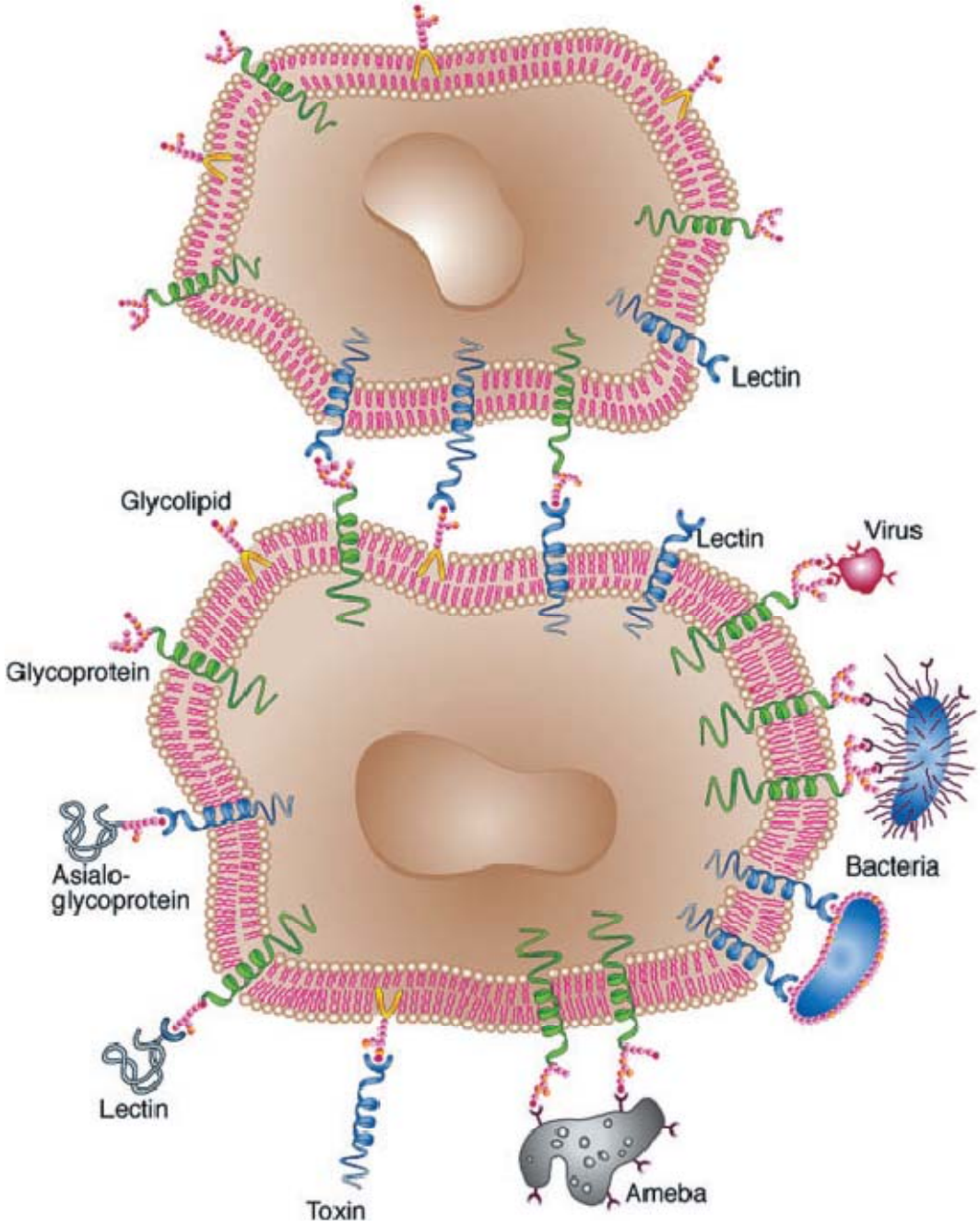
Currently, with advances in lectin research they could be aptly defined as: Lectins represent a heterogeneous group of proteins ubiquitously present; having multiple subunits; possessing specific sugar specificities, and thus possessing carbohydrate binding sites but widely varying in their size, subunit structure, molecular organization, and in their sugar combining sites. Lectins have many biological properties in common; they represent a diversified group of proteins with respect to size, composition and structure (Sharon *et al.,* 1974).

Functions & Applied aspect of Lectins

The most important question related to lectins that remain unanswered till date is their biological role. Lectins belonging to the same family (as the animal lectins or plant lectins) possess diverse features, and this interfamily divergence is responsible for the diverse roles they play. Even lectins having homologous sequences as those of the legume lectin family differ in their functions as they possess different sugar specificities, and their appearance in the plant part and tissue is also quite distinct.

In plants, lectins are located mostly in the storage tissues such as seeds and tubers where they function as reserve proteins. Besides that lectins could play important roles in plants such as in defense against pathogens, cell organization, embryo morphogenesis, phagocytosis, cell wall elongation, pollen recognition and much importantly in plant-microorganism symbiosis.

The animal lectins may be either extracellular or localized in the plasma membrane and they mediate a range of functions including cell adhesion, cell signalling, glycoprotein clearance and pathogen recognition. In invertebrates, lectins function mainly for providing innate immunity for defense against microbial infections (Vasta *et al.*, 2004). (http://www.imperial.ac.uk/research/animallectins/default.html)



<u>Figure 1.1</u>. The interactions between lectin and carbohydrates on the cell surface. Lectins serve as means of attachment of different kinds of cell 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). (Based on an original diagram from BioCarbAB Lund, Sweden).

Microbial lectins such as those found on the surface of *E.histolytica, Vibrio cholera*, Influenza virus and some fungi bind to the host membrane by lectin-carbohydrate interaction and thus help in penetration of the organism in the host cell and play an important role in infection. Some bacterial lectins present in the pili or fimbrae aid in adhesion, colonization whereas other bacterial lectins function as toxins. Some information on the functions of lectins in microbes, plants and animals is presented below in Table 1.1

Table1.1 Functions of lectins in all classes of organisms (adapted from Gabius *et al.*, 2004)

Lectin source/ Name	**Role they play**
Microbial lectins	
Amoeba	Aid in infecting the host
Bacteria	Aid in infecting the host for adhesion to membrane ; toxins for protection
Influenza virus	Aid in infecting the host
Plants Protection of the plant against attacks by bacteria, fungi, protozoa; mediate interaction between stored proteins and proteins in membranes; In legumes they aid in the symbiosis with nitrogen-fixing bacteria.	
Animals	

Calnexin, calreticulin, ERGIC-53	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 receptor	Innate immunity; clearance of sulfated 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	Cell-cell interactions in the immune and neural system
Spermadhesin	Sperm-egg interaction

Lectins are the tools to read sugar-encoded messages (Gabius *et al.*, 2004), as they possess the ability to act as recognition molecules inside cells, on cell surfaces, and in physiological fluids. They thus find use in the investigation of protein-carbohydrate interactions, because each one has specificity towards defined carbohydrate structure. They have been implicated in cell-to-cell recognition and signalling, blood group typing, in immune recognition process, and various other biological processes, such as viral, bacterial, mycoplasmal and parasitic infections, fertilization, cancer metastasis, growth and differentiation.

Certain legume lectins and some lectins isolated from animals such as the selectins and galectins have been exploited in many ways for medical purposes. They are implicated for use in development of anti-viral drugs, multivalent hydrophobic carbohydrates for anti-adhesion therapy of microbial diseases; highly effective inhibitors of the selectins are used for treatment of leucocyte-mediated pathogenic conditions, such as asthma, septic shock, stroke and myocardial infarction; inhibitors of the galectins and other lectins involved in metastasis; lectin-replacement therapy for lectin deficiency defects; and application of lectins for facile and improved disease diagnosis (Sharon, 2008). Some major applications of lectins are given below.

Summary of the major applications of lectins

- Cell identification and separation.

- Detection, isolation, and structural studies of glycoproteins.
- Investigation of carbohydrates on cells and subcellular organelles; histochemistry and cytochemistry.
- Mapping of neuronal pathways.
- Mitogenic stimulation of lymphocytes.
- Purging of bone marrow for transplantation.
- Selection of lectin-resistant mutants.
- Studies of glycoprotein biosynthesis.

1.2.2 Classification of Lectins

Lectins belonging to the larger class of carbohydrate binding proteins are known to possess some common structural features and binding motifs. They are easy to purify and have been obtained in pure form by methods such as affinity chromatography, on immobilized ligands and also by recombinant DNA techniques. Lectins are usually classified based on their saccharide specificity, although lectins in the same class may have different sugar binding patterns. Hence a better way of classifying lectins would be on the basis of homologies in their primary sequences and three-dimensional structures.

1.2.2.1 Classification of lectins based on their molecular structure

On the basis of their molecular structure they can be classified as (a) Simple Lectins, (b) Mosaic (or multidomain), and (c) Macromolecular assemblies (Sharon & Lis, 1990). 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 1.2 Classification of lectins on the basis of their molecular structure, characteristics and examples from each class.

I. SIMPLE LECTINS

- Contain lesser number of subunits, which may not be identical.
- Subunits are small of MW < 40 kDa.
- May contain an additional domain besides the carbohydrate binding domain.
- Comprises all known plant lectins and also the galectins.

Lectin Subfamily	Characteristics	Examples
a) Legume lectins	• Consist of two or four subunits of 25-30 kDa, each with one Ca^{2+} and one Mn^{2+} required for activity. • Tertiary structures are similar and hence superimposable; quaternary structures are different. • Possess diverse carbohydrate specificities. • Have a single combining site per subunit with low *K*a for monosaccharides (10^3 M^{-1}). • Many three-dimensional structures have been reported for the native lectin and also in complex with ligands.	ConA (Edelman *et al.*, 1972), SBA (Dessen *et al.*, 1995), PHA (Hamelryck *et al*., 1996), EcorL (Elgavish & Shaanan, 1998)

b) Cereal lectins	• Consist of two subunits. • Remarkably different from the legume lectins. • Exceptionally rich in cysteine. • Presence of multiple binding sites. • Secondary structure lacks the common structural elements such as α-helix and β-sheet. • Only WGA has been characterized in detail in this family.	WGA (Aub *et al*., 1965); Rice lectins (Tabary *et al.*, 1984); Barley lectin (Wright *et al.*, 1993) ;
c) Amaryllidaceae & related families	• Highly conserved primary sequences, exhibiting 80-90% homology. • Isolated from the tubers of plants belonging to amaryllis, orchid, and garlic families. • Bind to mannose. • Monomers are small (~12 kDa). • No metal requirement for activity or sugar binding. • The affinity for the monosaccharide ligands is very less (Ka < 10^2 M^{-1}).	Snowdrop (*Galanthus nivalis*)lectin (GNA) (Shibuya, 1988) in complex with methyl R-mannoside and with an octasaccharide PDB entry : 1NIV
d) Moraceae	• Jacalin is one of the well characterized lectin in this family. • It has the characteristic *β*-prism fold. • It is galctose specific, tetrameric molecule. • Each subunit consists of a heavy	Jacalin (Kumar *et al.*, 1982; Jeyaprakash *et al.*, 2003) from *Artocarpus integrifolia* (PDB entry : 1JAC,1UGW)*;* lectin from *Maclura pomifera* (Weimar *et al.*, 2000)

	α chain and a light β-chain. • Subunit structure consists of three antiparallel β sheets connected by loops , and arranged such that it forms the faces of a triangular prism , hence the name β-prism fold	
e) Euphorbiaceae	• Their structure is complex but still they are classified as simple lectins. • Heterodimer with two subunits A & B linked by disulphide linkages. B chain is stabilized by intrachain disulphide bonds. • A chain is cytotoxic having the RNA n-glycosidase activity, while B chain contains carbohydrate binding sites.	Ricin (Rutenber & Robertus, 1991) ; *Ricinus communis* agglutinin (RCA). (Nicolson & Blaustein, 1972; Gabdoulkhakov *et al.*, 2003)
f) Animal lectins (Galectins)	• This is a family of soluble, α-galactoside specific lectins that bind preferentially to lactose and *N*-acetyllactosamine. • These are animal lectins predominantly in mammals, but also in other vertebrates and some sponges. • Structure contains the S-carbohydrate recognition domain, or S-CRD which is a highly homologous domain. • Three-dimensional structure exhibits the jelly roll fold like the legume lectins.	Galectin from electric organ of eel (Levi & Teichberg, 1981)

g) Animal lectins (Pentraxins)	• This is a family of oligomeric plasma proteins. • The subunits in this family are arranged in a pentameric pattern, hence the name. • Binding of the ligand requires Ca^{++}. • Similar to the legume lectin fold, but very less sequence similarity with the legume lectins and also differ with respect to the carbohydrate binding site.	Serum Amyloid P component (SAP) (Hohenester, 1997)

II. MOSAIC (MULTIDOMAIN) Lectins

- Consists of lectins from diverse sources such as viral lectins, animal lectins of the C, P and I- types.
- Structurally these are composite molecules consisting of several kinds of protein domains, one of which contains carbohydrate binding domain. Generally monovalent, but may behave as multivalent proteins.

Lectin Subfamily	**Characteristics**	**Examples**
a)Viral Hemagglutinins	• *Influenza Virus Hemagglutinin* from this family is a multidomain protein. • Each subunit has two polypeptides HAI & HAII linked by disulphides. • The viral protein VPI from the *Murine Polyoma Virus* is arranged as pentamers on the capsid of the virus. • Each subunit structure resembles the lectin fold having two	*Influenza Virus Hemagglutinin* (Sauter *et al.*, 1992) *(*PDB entry HGF); Lectin from *Murine Polyoma Virus* (Stehle & Harrison, 1997)

	antiparallel β-sheets which are interconnected by several loops.	
b) C-Type Lectins	• This family is characterized by the presence of carbohydrate recognition domain (C-CRD) consisting of 115-130 amino acids, of which 14 are invariant and 18 are highly conserved. • Requires Ca^{++} for activity hence named C-type lectins. • They function to provide the sugar recognition activity in a variety of cell surfaces, initiate biological processes such as adhesion, endocytosis and complement fixation. • They can be categorized in three types <u>(a)Endocytic Lectins</u>: • These are mammalian proteins. First mammalian lectin to be studied was the HBP of this class. These are type II transmembrane proteins. <u>(b)Collectins:</u> • These are soluble proteins and MBP from this class are the most thoroughly studied lectins. • Structurally they consist of an amino terminal cysteine-rich domain, followed by a number of collagen-like repeats, an α-helical neck region and a carboxy terminal (CRD). <u>(c)Selectins</u>: • The group consists of highly asymmetrical membrane bound proteins and can be divided into	*Endocytic Lectins :* galactose/*N*-acetylgalactosamine specific lectin from rabbit hepatocytes (RHL), also known as hepatic asialoglycoprotein receptor (or hepatic binding protein, HBP)(Lowe *et al.* , 1983), Avian hepatic lectin specific for *N*-acetylglucosamine, galactosespecific lectin on peritoneal macrophages (Hudgin *et al.*, 1974), Fucose-specific receptor (lectin) found on the Kupffer cells of the liver, Mannose-specific macrophage surface lectin *Collectins:* MBP A (serum type MBP, first C-type lectin structure to be elucidated) and C (liver type MBP), pulmonary surfactant apoproteins A and

	three sub-groups, the E-lectins, P-lectins and L-lectins. • These proteins mediate selective contact between cells. • They contain an additional epidermal growth factor (EGF)-like domain. • They bind oligosaccharides with high affinities in the range of 10^4-10^5.	D, Collectin CL-43 from bovine serum and bovine conglutinin (Dec & Wernicki, 2006) *Selectins*: consist of three groups, E-lectins, P-lectins and L-lectins. (Nimrichter *et al.*, 2008)
d) P-type lectins	• This family consists of two closely related lectins, the mannose 6-phosphate receptors and the type I transmembrane glycoproteins. • Structurally the extra cellular domain consists of homologous repeating units and contains either one or two high-affinity binding sites.	Two mannose 6-phosphate (Man-6-P) receptors belong to this family (Olson *et al.*, 1999)
e) I-type lectins	• These are the members of the immunoglobulin superfamily and they contain variable numbers of extracellular immunoglobulin-like domains. • Most well characterized lectins of this family are the Sialoadhesins, which are a family of sialic acid specific type I membrane glycoproteins. • A conserved Arg residue was found to play a key role in the affinity of the I-type lectins to glycoconjugates containing sialic acid.	Sialoadhesins include : macrophage receptor, CD22 present in B cells only(lymphocyte surface antigen) (Abdu-Allah *et al*., 2008); CD33 present on early myeloid cells ; myelin-associated glycoprotein (Cao, 2007)

III. MACROMOLECULAR ASSEMBLIES

- Common in bacteria, and play a role in structural support for mobility of the organism.
- They are filamentous organelles consisting of helically arranged subunits (pilins) assembled in a well-defined order.
- Some of them are made up of hundreds of subunits and some of the subunit have carbohydrate combining site for binding and sugar specificity.
- Three-dimensional structures of this family of lectins have not been solved yet.
- Example is fimbriae of *E. coli* (Li *et al.*, 2009)**,** *K. pneumonia* (Duncan *et al.*, 2005), *Pseudomonas aeruginosa* pilin (Craig *et al.*, 2003)

1.2.2.2 Classification based on organisms they are found in

Animal Lectins

Only after the discovery of animal lectins, lectins were proven to have cell recognition property. The first of such lectins to be discovered was the galactose-specific liver lectin by Gilbert Ashwell (Hudgin & Ashwell, 1974), which recognizes and binds asialo-glycoproteins and is responsible for their uptake by the liver and eventual degradation. This was followed by isolation of lectin from electric eel (*Electrophorus electricus*) and chicken muscle, the first members of the family of the soluble β-galactose-specific lectins.

In animals, lectins are secreted extracellular as well as inside the cell membrane. Lectins have been implicated in mediating variety of binding and recognition events in animal cells. Intracellular animal lectins are involved in quality control; sorting of glycoproteins in the secretory pathway, in specific processes in the cytoplasmic and/or nuclear compartment, while surface lectins act as receptors and play a role in the extracellular trafficking of glycoproteins, cell adhesion, cell–cell interactions and innate immunity. Extracellular lectins are involved in the organization of the extracellular matrix; as signal molecules in immunity and cell growth and

development. Mammalian cells contain a variety of glycoconjugates on their surfaces. Evidently, lectin-mediated protein–carbohydrate interactions are essential for animals.

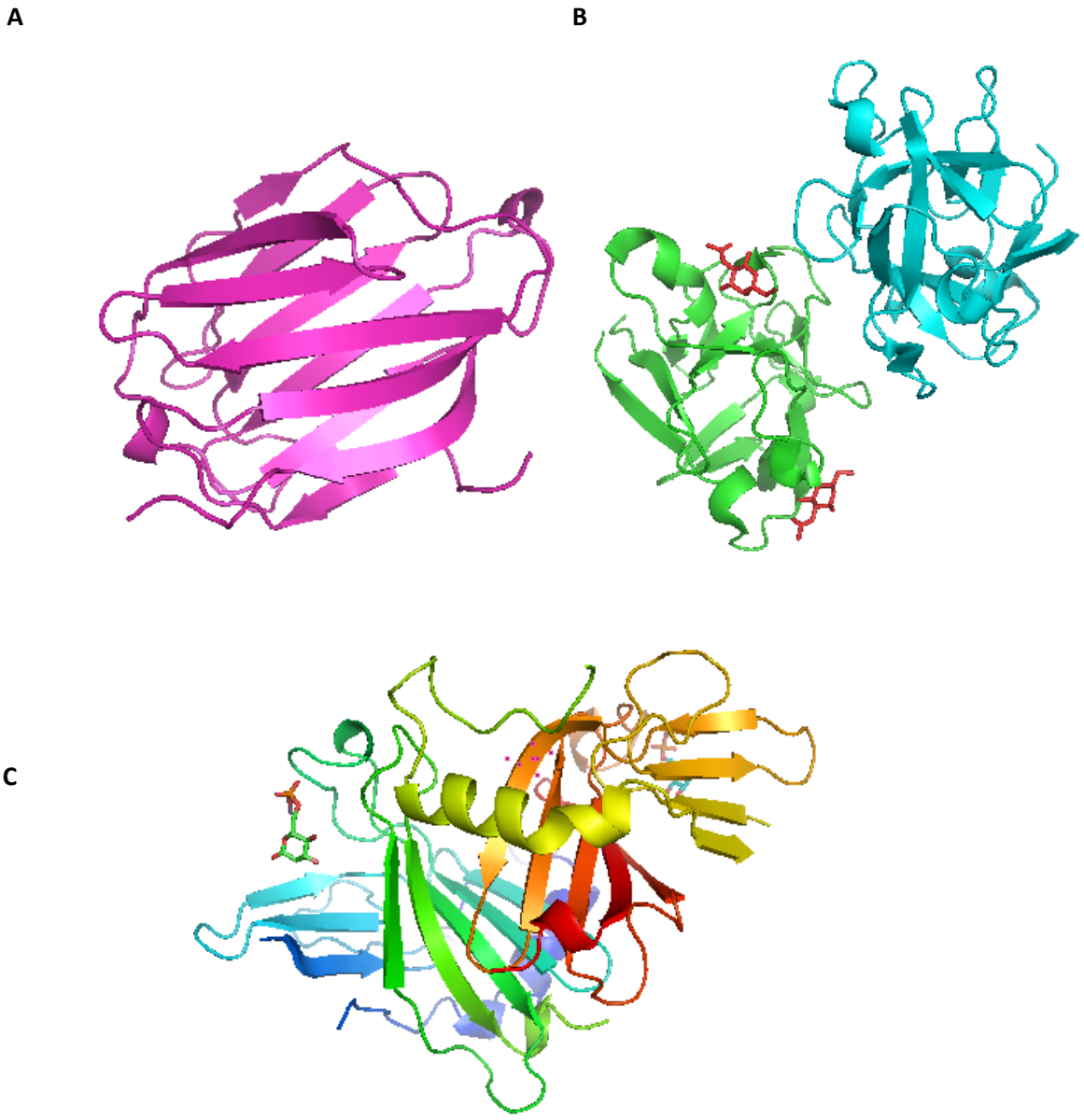


Fig 1.2 Crystal structures of some animal lectins.

(A) Crystal structure of charcoat-layden protein showing a β-sheet rich galectin CRD (PDB code: 1LCL).

(B) Crystal structure of earthworm lectin (*Lumbricus terrestris*) in complex with GalNAc which is shown as a stick model. The structure consists of two homologous domains A & B, of which the B-chain shows bound galactose, and it also shows homology to Ricin (B-chain) having a β-trefoil structure. (PDB code: 2ZQO).

(C) Crystal structure of the extracytoplasmic domain of the Bovine Mannose-6-phosphate receptor shown in various colors. It has bound ligand which is shown as stick model (PDB code: 1M6P).

Like the lectins from other lower organisms, animal lectins also use the property of multivalency for specific and high affinity binding to the target cells. Animal lectins are complex, multidomain proteins. Their sugar binding activity results from a particular domain called Carbohydrate Recognition Domain (CRD). The different functional aspects of lectins in animals are summarized in Table 1.3.

Table 1.3: The functions of animal lectins (adapted from Gabius *et al.*, 2004)

Activity	**Example of lectin**
Intracellular routing of glycoconjugates and vesicles	P-type lectins ERGIC-53, VIP-36
Molecular chaperones during glycoprotein synthesis	Calnexin, calreticulin
Mediation of endocytosis	Asialoglycoprotein receptors, macrophage mannose receptor
Cellular growth regulation and induction of apoptosis/anoikis	Galectins, sarcolectin, cytokines, C-type lectins, amphoterin-like protein, hyaluronic- acid-binding proteins, cerebellar soluble lectin
	Geodia cydonium galectin, other galectins, interleukin-2, cytokines(e.g. IL-2:IL-2R and

Extracellular molecular bridging	CD3 of T-cell receptors), cerebellar soluble lectin
Cell–cell interactions for homing and trafficking	Selectins, CD22, CD31, CD44 and other C-type lectins (e.g. DC-SIGN), galectins, I-type lectins (e.g. siglecs, N-CAM, P0, or L1)
Cell–matrix interactions	Galectins, heparin- and hyaluronic acid binding lectins such as hyalectans/lecticans, calreticulin
Scavenging of cellular debris; anti-inflammatory action	Galectin-9
Hormones in hepatic endothelial cells	P-type lectins
Recognition of foreign glycans(β-1,3-glucans, LPS)	CR3 (CD11b/CD18), dectin-1, Limulus coagulation factors
Recognition of foreign or aberrant glycosignatures on cells (including endocytosis or initiation of opsonization or complement activation)	Collectins, L-ficolin, C-type macrophage and dendritic cell receptors, a/q-defensins, pentraxins (CRP, limulin), tachylectins
Targeting of enzymatic activity in multimodular proteins	Acrosin, laforin, Limulus coagulation factor C
Intra- and intermolecular modulation of enzyme activities in vitro	Porcine pancreatic a-amylase, galectin-1/a2-6-sialyltransferase
Cell migration and routing	Selectins and other C-type lectins, I-type lectins, galectins, hyaluronic-acid-binding proteins (RHAMM, CD44,hyalectans/lecticans)
Matrix network assembly	Proteoglycan core proteins (C-type CRD and G1 domain of hyalectans/lecticans), galectins (e.g. galectin-3/hensin), nonintegrin 67-kDa elastin/laminin-binding

On the basis of their structural domains, animal lectins have been classified into seven distinct families- i) Calnexin ii) L-Type lectins iii) P-Type lectins iv) C-Type lectins v) Galectins vi) I-Type lectins and the vii) R-Type lectins. (Fig. 1.3)

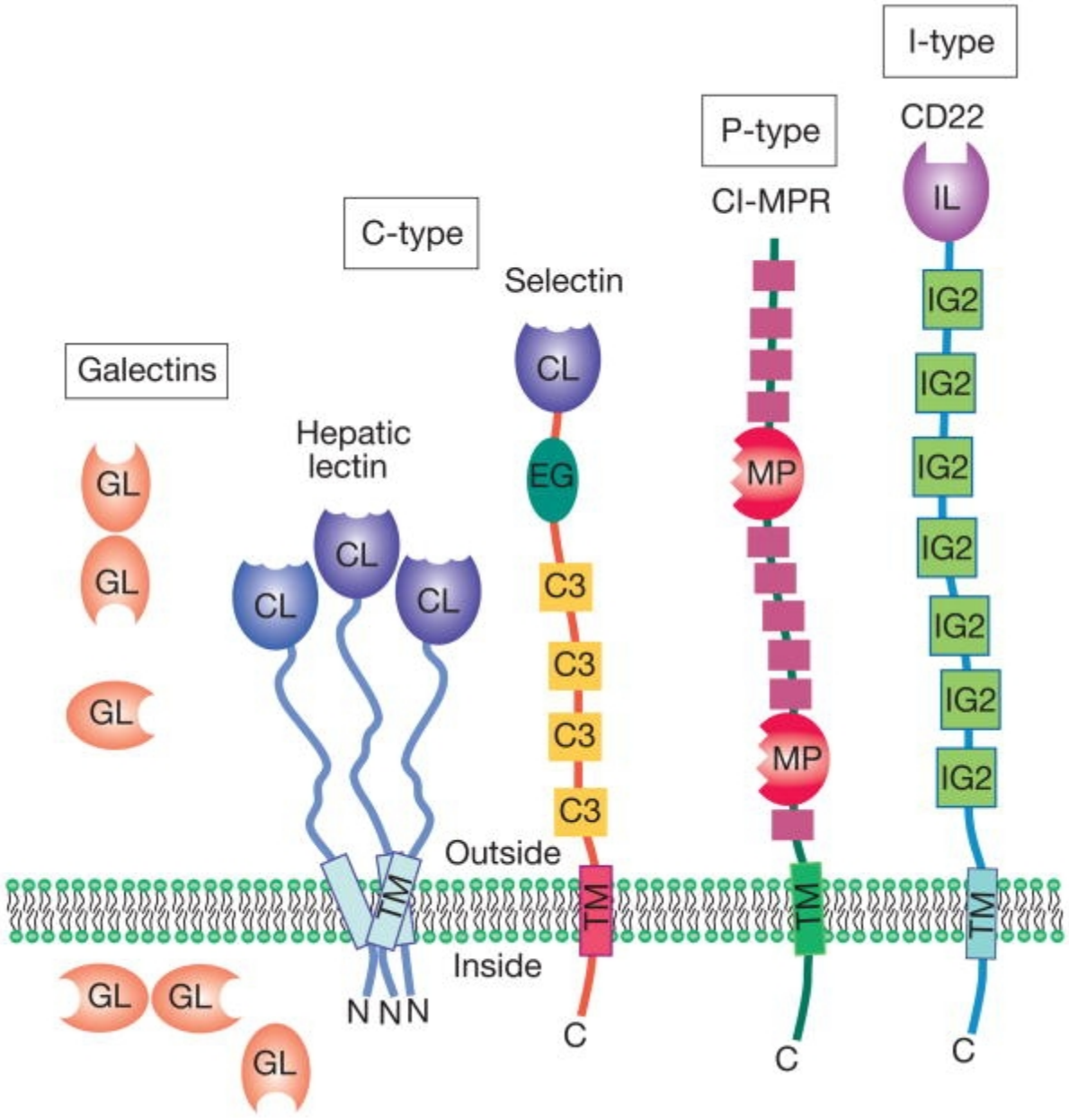


Fig. 1.3 Shows generic structures of several of the classes of animal lectins, the C-type, P-type, I-type lectins and the galectins

The different classes, with the functions, ligands they bind, and the examples of each class are presented in Table 1.4 below.

Table 1.4 The various classes of animal lectin, with their CBDs, function and examples of lectin structures

Lectin Group	**Ligands**	**Structure of CRD**	**Examples of function**	**Some Structures available in PDB**
Calnexin (Schrag *et al.*, 2001)	Glc_1Man_9 oligosaccharides	Unknown	Protein sorting in the endoplasmic reticulum	i)Calnexin 1JHN ; ii)Calreticulin P-domain : 1HHN
L-Type (Satoh *et al.*,2007)	Various sugars	β-sandwich	Protein sorting in the endoplasmic reticulum	i)VIP36 structures : 2DUO,2DUP,2DUQ ; CRD of ERGIC-53: 1R1Z
P-Type (Roberts *et al.*, 1998)	Mannose 6-phosphate	Unique β-rich structure	Protein sorting post golgi	i)Bovine MBP receptor :1M6P (Fig. 1.2 C); ii)Other MBP receptors : 2RLB,2RL8,3CY4
C-Type	Various	Unique mixed	Cell adhesion (Selectins); Glycoprotein	i) CEL-I from *Cucumaria echinata*:

(Sugawara *et al.*, 2004)	sugars	α/β structure	clearance; Innate immunity (collectins)	1WMY; ii)Tunicate lectin with D-Galactose : 1TLG
Galectins (Leonidas *et al.*, 1995)	β-galactosides	β-sandwich	Glycan crosslinking in the extracellular matrix	i)Charcot-Leyden Crystal Protein: 1LCL; ii)Human GRP CRD: 3B9C (Fig. 1.2 A)
I-Type (Attrill, 2006)	Sialic acid	Immunoglubulin superfamily	Cell Adhesion (siglecs)	i)Siglec-7 Native : 1O7S ii)Siglec-7 complexed : 2G5R,2HLR; iii)N-terminal of sialoadhesin complexed :1ODA,1OD7, 1OD9;
R-Type (Suzuki *et al.*, 2008)	Various sugars	β-trefoil	Enzyme targeting; Glycoprotein hormone turnover	i)Earthworm R-type lectin in complex with GalNAc: 2ZQO (Fig 1.2 B)

Microbial Lectins

Lectins have been widely studied in the microbial kingdom and these are surface lectins and aid in the adhesion of the organism to the host cell prior to infection. The adhesion of the organism to cell walls of the host epithelium is mediated by lectin-carbohydrate interactions. Such lectins have been reported from viruses, bacteria, protozoans, fungi and even slime moulds. This group of lectins bind to terminal sugar residues with high specificity like the human

influenza virus binds primarily to cells containing Siaα2-6Gal linkages, whereas other animal and bird influenza viruses preferentially bind to Siaα2-3Gal termini.

Viral Lectins (Hemaglutinins): One of the most thoroughly studied lectins, the Influenza virus hemagglutinin (type I membrane protein) (Sauter, 1992) 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 hemaglutinin has been crystallized and its structure has been determined (PDB code: 1HTM) (Bullough *et al.*, 1994). The lectin subunit consists of two polypeptides HA_1 and HA_2 which are linked by disulphide linkages. The subunits associate into trimers noncovalently, which are located on the viral membrane and help in the binding of the organism to the host membrane via lectin-carbohydrate interaction. Crystal structures of several viral lectins are known till date. The detailed knowledge of the sialic acid-hemagglutinin interaction from the available structure can provide the basis for the design of antiviral drugs that would block viral attachment to cells.

Bacterial Lectins (Adhesins and Toxins): Many bacterial lectins have also been isolated, characterized and described in detail. They can be divided into two classes:

1) Adhesins: Lectins that are localized on the surface of the bacteria and aid in bacterial adhesion to the surface and hence their colonization

2) Bacterial toxins secreted in the external environment.

The bacterial lectins generally prefer binding to glycolipids than glycoproteins and their binding specificity to a particular structure can be correlated with particular tissue it colonizes in the host. It has been shown that blocking of bacterial lectins can prevent infection *in vivo*. For example *Clostridium*, *E. coli*, and *Lactobacillus* (Duncan *et al.*, 2005) species that recognize

Galα1-4Gal only colonize the large intestine and not the small intestine where Galα1-4Galβ1-4Glcβ-ceramide is expressed in the epithelial cells. A class of bacterial lectins that helps in cellular movement is called as the fimbriae or pili present on the cell surface and extending in the environment. Multivalent interactions of the lectin in the fimbriae and the host glycan are known (Sharon, 1987). Example is heparan sulfate–binding adhesions.

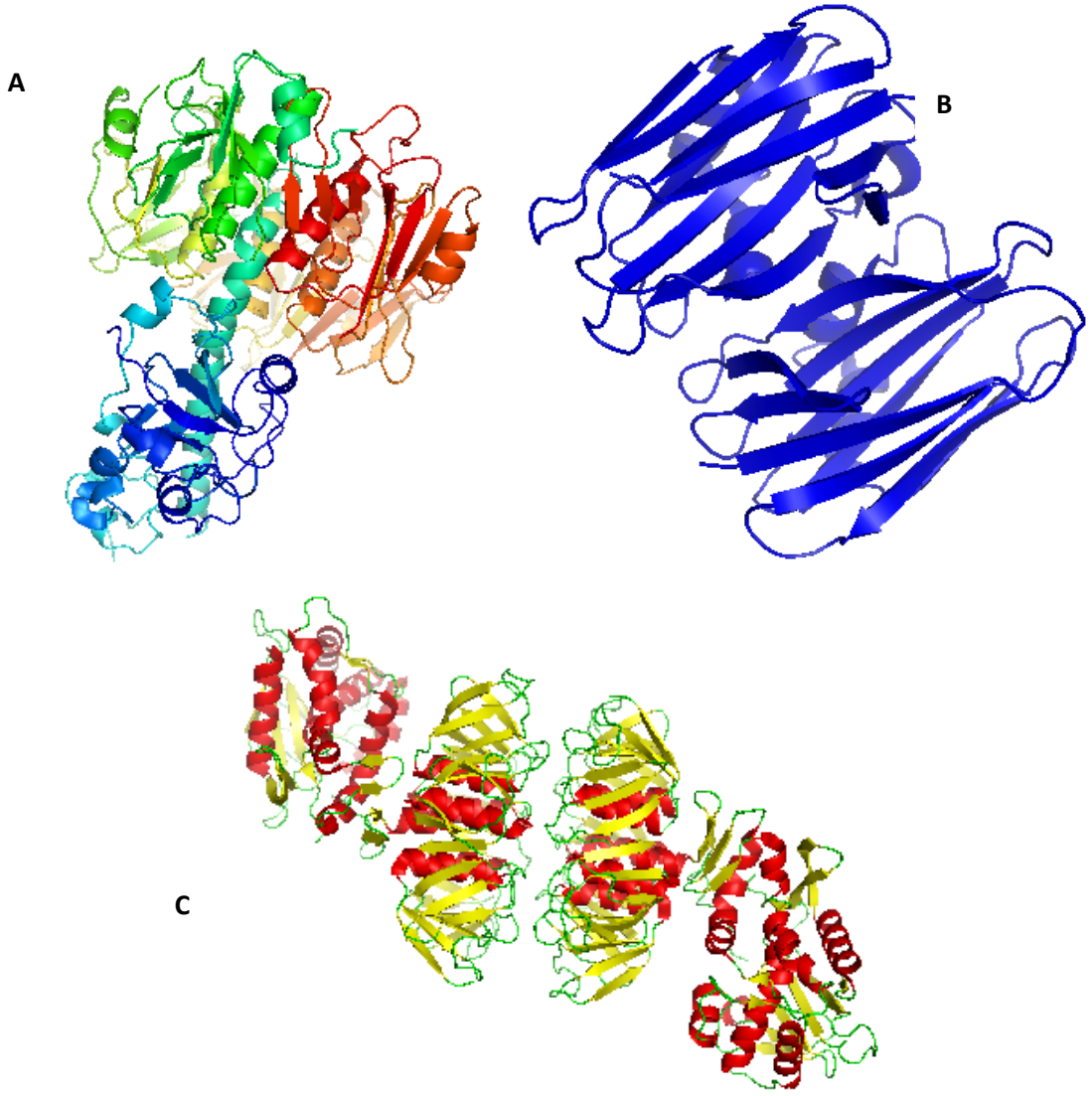


Fig 1.4 Crystal structures of some microbial lectins and a fungal lectin.

(A) Crystal structure of cholera toxin (PDB code: 1XTC), shown in various colours. (PDB code: 1R4Q). Each monomer has 2 chains; A-chain is wedge-shaped placed on the plane of the pentameric B-chain.

(B) The Crystal structure of β-sheet rich lectin from *Agaricus bisporus*. All the secondary structural elements are uniformely coloured in blue. (PDB code: 2Y2T)

(C) Crystal structure of shiga toxin from *E.coli.* The two B- chain pentameric subunits are seen facing each other in the center of the structure. The α-helix, β-sheets and the various loops are represented in red, yellow and green colours respectively.

A few heat sensitive bacterial toxins have been structurally characterized, such as the cholera toxin from *Vibrio cholera* (PDB code: 1XTC) (Zhang *et al.*, 1995), shiga toxin from *Shigella dysenteria* (PDB code: 1R4Q) (Fraser *et al.*, 2004), pertussis toxin from *Bordetella pertussis* (PDB code: 1BCP) (Hazes *et al.*, 1996), and enterotoxin from *E. coli* (PDB codes: 1ETL, 1ETM) (Sato *et al.*, 1994). The structure of an antiviral lectin Scytovirin (SVN) from a cynobacteria *Scytonema varium* has also been reported. (Moulaei *et al.*, 2007)

Protozoan lectins: Among the vast number of protozoan which infect humans as well as animals and cause infection, the well studied lectins are from *E.histolytica* (Sharma *et al.*, 2008), the causative agent of amoebic dysentery. Two lectins have been isolated from the protozoan.

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. Other fungal lectins characterized include PHA and PHB from *Agaricus bisporus, Ganoderma* lectin (*Ganoderma capense); Grifola* lectin (*Grifola frondosa*) (Sueyoshi *et al.*, 1985; Kawagishi *et al.*, 1990; Kawagishi *et al.*, 1997) and many more have been reported.

Mushroom lectins have been studied from edible mushrooms as *Amanita phalloides, Lactarius deliciosus, Boletus edulis, Laetiporus sulphureus* (Kobert, 1906; Guillot & Konska, 1997) etc.

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. Lectins have been characterized from a wide variety of plants, belonging to different geographical conditions. Plant lectins are defined as plant proteins that possess at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide. They may contain a domain other than the carbohydrate binding domain with a different structure and biological activity, and the main function of the lectin is generally determined by this separate non lectin domain. 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. Plant lectins generally are suggested to play role as defense agents against different kinds of predators such as viruses, fungi, invertebrates and higher animals. Another important function is the establishment of symbiosis between nitrogen fixing bacteria and leguminous plants.

Lectins isolated from plants are observed to belong to certain taxonomic groups such as legumes, solanaceae family and some monocot families. Study of the structures of plant lectins determined by x-ray crystallography has shown that a characteristic feature of this plant lectin family is their extreme structural diversity and thus they do not belong to a single superfamily. Biochemical, structural and molecular analyses has provided sufficient information on the basis of which they are classified into only seven families of structurally and evolutionarily related proteins (Van Damme *et al.*, 1998). Out of these seven families, there are four major families and three minor ones as given below in Table 1.5.

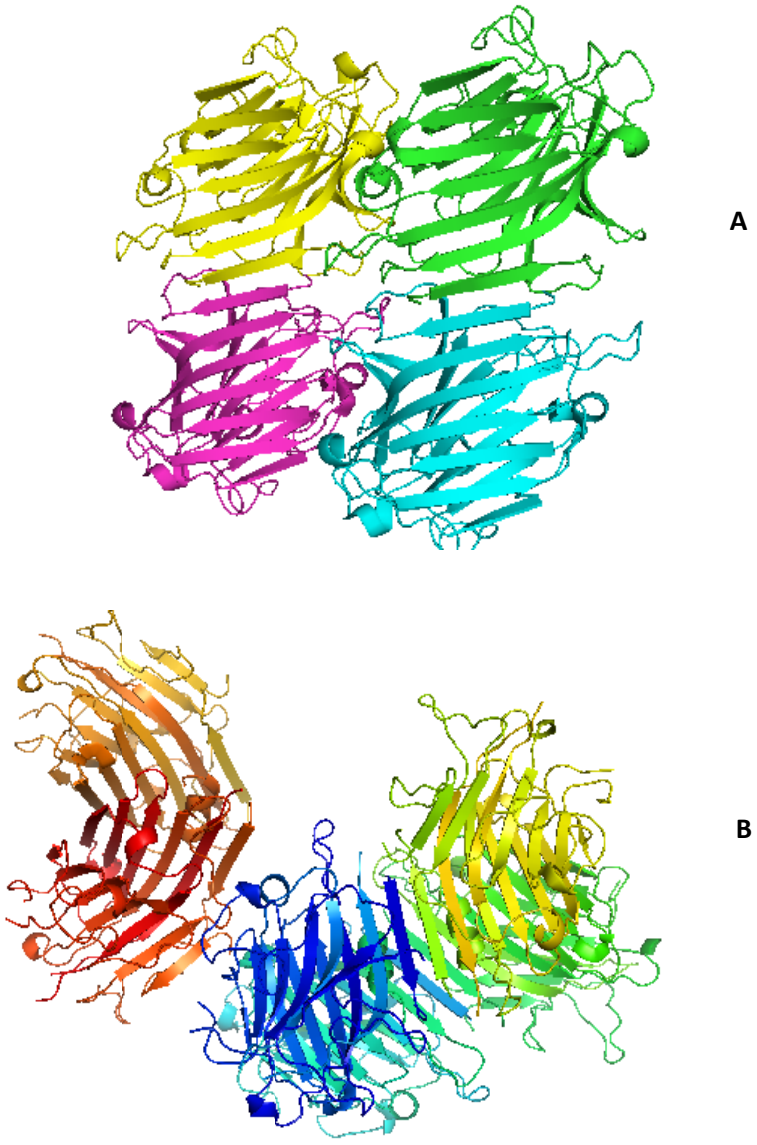


Fig 1.5 Crystal structures of representative legume lectins.

(A) Crystal structure of the *Ulex europaeus* lectin showing all the tertramers in differenent colours. The lectin is tetrameric in the asymmetric unit and displays chibose specificity (PDB code: 1JXN).

(B) The *Dolichos biflorus* lectin (DB58) crystal structure showing the arrangement of the tetramer (PDB code: 1LUL).

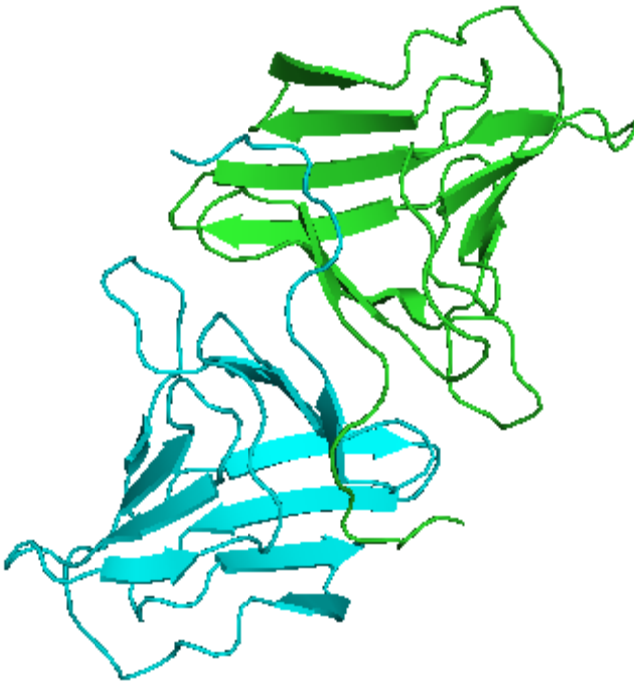


Fig 1.6 Representative structure of the Monocot-Mannose binding lectin family.

The crystal structure of mannose-binding lectin from the bulbs of *Scilla campanulatais* shown which exists as a dimer in the asymmetric unit. The structure displays a threefold symmetric β-prism made up of three antiparallel four-stranded β-sheets (PDB code: 1B2P).

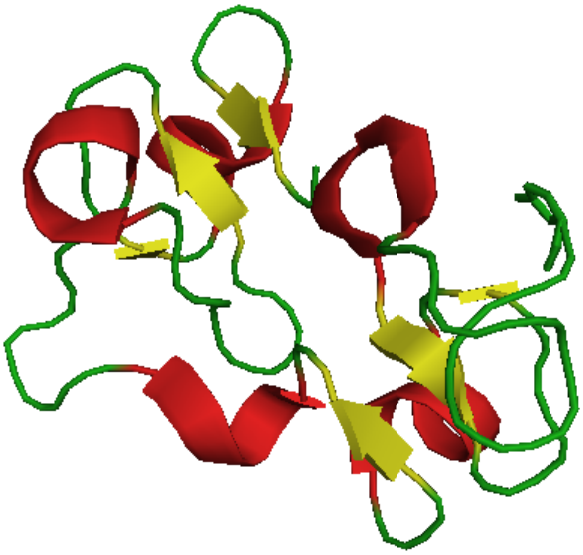


Fig 1.7 Representative structure of the Chitin binding lectins plant family.

Crystal structure of monomeric *Urtica dioica* agglutinin. The structure comprises of two hevein-like domains, each with a saccharide-binding site. (PDB code: 1EIS)

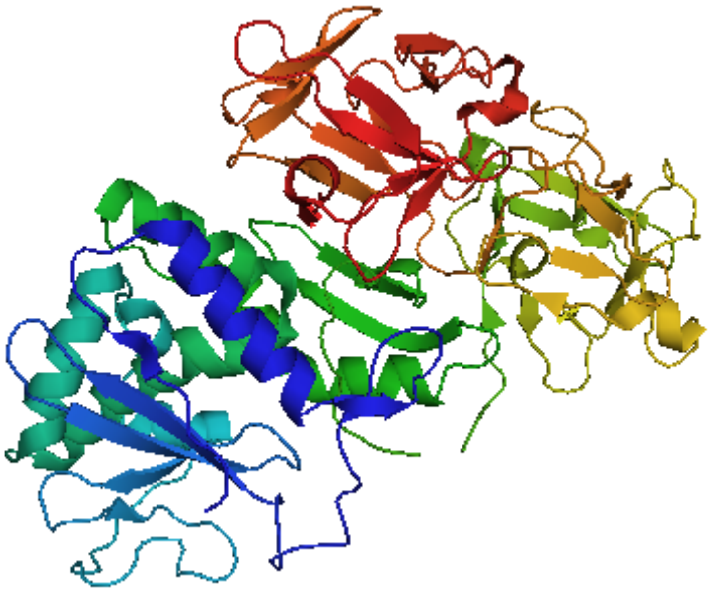


Fig 1.8 Representative structure of the Type-II RIP family

Crystal structure of HmRIP from *Viscum album* which belongs to type-II RIP family exists as a dimer. The B-chain is a galactose binding lectin and is rich in β-sheets. (PDB code: 1PC8)

Fig 1. 9 Representative structure of the Jacalin-related lectin family.

Crystal structure of *Helianthus tuberosus* lectin exhibits a threefold symmetric beta-prism fold made up of three four-stranded beta sheets. (PDB code: 1C3N).

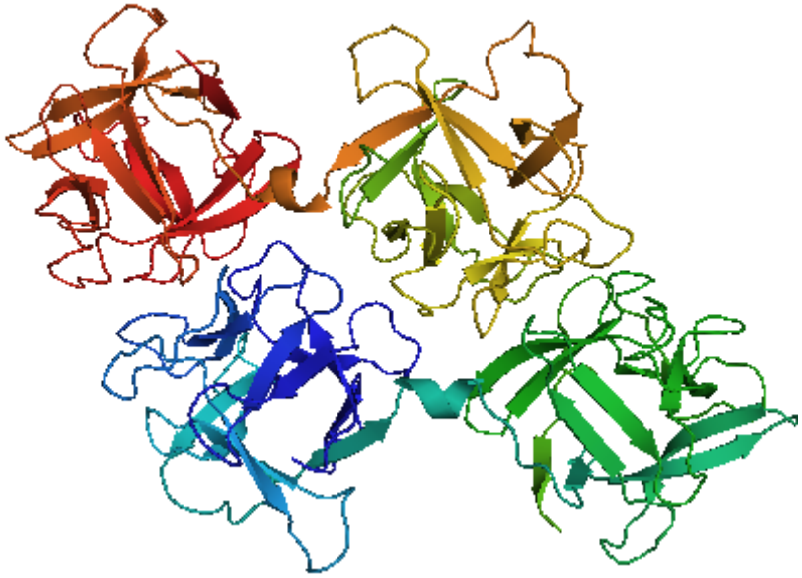


Fig 1.10 Representative structure of the Amaranthin lectin family.

Crystal structure of *Amaranthus caudatus* agglutinin showing novel arrangement of four beta-trefoil domains (PDB code: 1JLY).

Table 1.5 Overview of the structurally classified plant lectin families, lectins studied and a few examples of reported PDB structures

Lectin family and their characteristics	**Examples of lectins and few reported PDB structures**
Legume lectins This is the largest and most well studied family among the plant lectin families, and the lectins belonging to this family are evolutionarily related. Lectin is the major component of legume seeds and tubers, but their physiological function is not known yet. Well conserved lectin family, in which members share similar primary sequences, secondary structures but differ in their tertiary structures, quaternary folds and also carbohydrate specificities. Jelly roll fold/ Legume lectin fold is characteristic of this family and consists of two large β-pleated sheets sandwiching a carbohydrate site in the fold. Lectins are oligomeric (dimeric/tetrameric) and have a characteristic CBD. Generally contain transition metal ion such as Ca^{++} and Mn^{++} for interaction with	Concavalin A (ConA) (*Canavalia ensiformis*) (Edelman *et al.*, 1972; Hardman & Ainsworth, 1972).: 1I3H; Soybean (*Glycine max*)seed lectin (SBA) (Dessen *et al.*, 1995) : 1SBE ; Phytohemagglutinin (*Phaseolus vulgaris)* (Hamelryck *et al.*, 1996) : 1FAT; *Griffonia simplicifolia* lectin (GS-IB4) (Delbaere *et al.*, 1993) : 1GNZ; *Ulex Europaeus* lectin I (Dao-Thi *et al.*, 1998) : 1FX5; *E. corallodendron* lectin (EcorL) (Elgavish & Shaanan, 1998) : 1FYU, 1AXO; Peanut agglutinin (PNA) (Banerjee *et al.*, 1994) (*Arachis hypogaea)*lectin : 1BZW;

carbohydrates (Sharon and Lis, 1990). Conserved amino acids among this family are the residues involved in carbohydrate binding as well as in metal binding	
Type II Ribosome-Inactivating Proteins Type II RIPs are a group of toxic proteins from plants that catalytically inactivate eukaryotic ribosomes and thus protein synthesis. They contain an enzymatically active A chain (having RIP activity) linked to a galactose specific domain called B-chain which confers it lectin activity. Type II RIPs occur in several families, namely Euphorbiaceae, Fabaceae, Viscaceae, Passifloraceae, Ranunculaceae, Lauraceae, Sambucaceae, Cucurbitaceae and Iridaceae. They play a role in the plant's defense against plant-eating organisms and also function as plant storage proteins Some type II RIPs also show antiviral activity *in vitro* against plant viruses	*Ricinus communis* (ricin) (Rutenber *et al.*, 1991; Katzin *et al.*, 1991; Rutenber & Robertus, 1991) : 2AAI ; *Abrus precatorius* seeds (abrin) (Tahirov *et al.*, 1995) **:** 1ABR; *Trichosanthes kirilowii lectin* (Li *et al.*, 2000): 1GGP;
Monocot mannose-binding lectins These lectins are confined to a subgroup of monocots. They have been found in six different monocot families, namely, Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Liliaceae, and Orchidaceae (Van Damme *et al.*, 1998a). They have been reported to occur in various vegetative tissues, but rarely occur	*Galanthus nivalis* agglutinin (GNA) (Shibuya, 1988) : 1MSA ; Complex-sugar specific lectin from *Scilla campanulata* bulbs (SCAfet) (Wright *et al.*, 1996 Wood, 1999) : 1DLP; Dimeric garlic bulb lectin in complex with α-D-mannose (Chandra *et al.*, 1999) : 1KJ1;

in seeds. These lectins are composed of lectin subunits of approximately 12 kDa and which are mannose specific. Their structure has characteristic β-prism II fold. They show inhibitory activity on the *in vitro* replication of retroviruses as they also recognize mannose containing glycoprotein. This family of lectins has been reported from diverse plants and show marked differences in their specificity towards oligomannosides and N-glycans. The *Arisaema flavum* lectin has been reported with complex-sugar specificity. They function as storage proteins and may have role in plant defense.	
Chitin binding lectins comprising hevein domains Proteins containing at least one hevein domain. The term 'hevein' refers to a chitin-binding polypeptide, 43 amino acids long, isolated from the latex of rubber tree, *Hevea brasiliensis* (Waljuno *et al.*, 1975). "Hevein domain" refers to a structural unit of about 40 amino acid residues that exhibit sequence similarity to hevein and possess chitin-binding activity. This lectin family has high affinity for N-glycosylated animal glycoproteins. These lectins have been reported in several plant families in both monocots as well as dicots such as Gramineae, Solanaceae,	Hevein (Waljuno *et al.*, 1975). from rubber tree (*Hevea brasiliensis*) : 1Q9B Wheat germ agglutinin (WGA) complexed with sialyllactose : 1WGC *Urtica dioica* lectin(Peumans *et al.*, 1984), : 1IQB

Phytolaccaceae, Urticacea, Papavaracaceae and Viscaceae (Raikhel *et al*., 1993).	
Jacalin-related lectins This family contains all lectins that are structurally and evolutionary related to jacalin, which was the first lectin islotaed from this family (Kumar *et al*., 1982), Galactose specific lectins. These are found to occur in plants from several taxonomically unrelated families such as Moraceae (jacalin, MPA, *Artocarpus hirsuta* lectin, artocarpin), Convolvulaceae (calsepa, conarva), Asteraceae (heltuba), Gramineae (barley and wheat lectins) and Musaceae (banana lectin) (Van Damme, 1998). Three-dimensional structure has the characteristic β-prism I fold, in which three four stranded β-sheets arrange as three faces of a prism and thus form a structure resembling threefold symmetric β-prism.	Jacalin (*Artocarpus integrifolia)*(Jeyaprakash *et al.*, 2003) monomer : 1UGX ; Jacalin (*Artocarpus integrifolia)* tetramer : 1UGW; Heltuba (*Helianthus tuberosus*) lectin (Bourne *et al.*, 1999).: 1C3K; Calsepa lectin (*Calystegia sepium)* (Bourne *et al.*, 2004) resembling the canonical dimer of legume lectins : 1OUW;
Cucurbitaceae phloem lectins This is a small family of lectins present in the phloem exudates of plants belonging to the cucurbitaceae family. They are also called as also PP2 family and are chitin binding lectins. They have been characterized from the phloem exudates of some cucurbits such as *Cucumis*, *Sechium*, *Luffa, Coccinia,*	*Cucurbita maxima* lectin (Read & Northcote, 1983)

Cucurbita, and *Citrullus* species. They differ from the yype-II RIPs which are also reported in cucurbitaceae, and they do not contain the hevein domain. Structurally they consist of unglycosylated subunits of about 22-25 kDa and are dimeric in solution. The subunits may be linked by disulphide as in the case of lectin from *Cucurbita maxima* (Read & Northcote, 1983). Structure of any of the lectin from this family has not yet been solved. Their function in plants is defense.	
Amaranthins The family is named on amaranthin, lectin from *Amaranthus caudatus* seeds. All these lectins are isolated only from the seeds of the plants belonging to the family Amaranthaceae. They are specific towards GalNAc and preferentially recognize the T-antigen disaccharide Galβ (1, 3) GalNAc (Rinderle *et al*. 1990). These are homodimeric proteins composed of two identical subunits of about 30 kDa. The monomers consist of two domains (called *N*- and *C*-domains) linked by a short helix. Each domain has a β-trefoil structure like the ricin B chain.	Amaranthin (Rinderle *et al.*, 1990) from *Amaranthus caudatus* in complex with T-antigen disaccharide : 1JLX

Besides plant protection, the other role of lectin includes cellular recognition of molecules which is important for certain plant processes such as phagocytosis, cell wall elongation, cell organization in the plant tissues (Table 1.6). Being an important component of the plant storage tissue lectins serve as reserve proteins in the seeds and tubers. A few functions of plant lectins and the respective examples are tabulated below.

Table 1.6 Gives the functions of lectin activity in plants and the associated lectins

Activity	**Example of lectin**
External activities	
Protection from fungal attack	*Hevea brasiliensis* (rubber tree), *Urtica dioica* (stinging nettle), *Solanum tuberosum* (potato)
Protection from herbivorous animals	*Phaseolus vulgaris* (French bean), *Ricinus communis* (castor bean), *Galanthus nivalis* (snowdrop), *Triticum vulgare* (wheat)
Involvement in establishing symbiosis between plants and bacteria	*Pisum sativum* (common pea), *Lotononis bainesii* (miles lotononis), *Arachis hypogaea* (peanut), *Triticum vulgare* (wheat), *Oryza sativa* (rice)

Internal activities	
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 protein body membranes	*Pisum sativum* (common pea), *Lens culinaris* (lentil), *Glycine max* (soybean), *Oryza sativa* (rice)
Modulation of enzymatic activities such as phosphatase activity	*Secale cereale* (rye), *Solanum tuberosum* (potato), *Pleurotus ostreatus* (oyster mushroom), *Glycine max* (soybean), *Dolichos biflorus* (horse gram)
Participation in growth regulation	*Medicago sativa* (alfalfa), *Cicer arietinum* (chick pea)
Adjustment to altered environmental	*Triticum aestivum* (winter wheat)

conditions	

1.2.3 Plant lectins as tools in biology and medicine

Every biological species is a potential source of medicine. Any species of plants that is useful in maintenance of health, detection, prevention or cure of disease, falls in the category of medicinal plant, and can be utilized for human welfare. Many plants from the families such as Euphorbiaceae, Asteraceae, and Cucurbitaceae have been useful as a source of medicine. Hundreds of species of plants have been indicated in the indigenous system of medicine since many years. Different systems of medicine use plants for prevention and cure of disease such as Ayurveda, Unani medicine, homeopathy, naturopathy, aromatherapy, Chinese medicine etc.

Lectins with well defined specificity are excellent tools for several analytical and preparative purposes. Plant lectins have innumerable application, both direct and indirect, either in the 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 signaling events can be 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.

Versatility of plant lectins as research tools (Gabius *et al.*, 2004)

Biochemistry

- Detection of defined carbohydrate epitopes of glycoconjugates in blots or on thin-layer chromatography plates.
- Purification of lectin-reactive glycoconjugates by affinity chromatography.
- Glycan characterization by serial lectin affinity chromatography (lectin affinity capture).
- Glycome analysis (glycomics).
- Quantification of lectin-reactive glycoconjugates in enzyme-linked lectin binding assays (ELLA).
- Quantification of activities of glycosyltransferases/glycosidases by lectin based detection of products of enzymatic reaction.
- Model reagents for the assessment of the ligand functionality of carbohydrate-presenting scaffolds (e.g. glycodendrimers).
- As microarrays which can be a unique means for high throughput analysis of protein glycosylation.

Cell biology

- Characterization of intracellular assembly, routing, and cell surface presentation of glycoconjugates in normal and genetically engineered cells (glycomic profiling, spatially defined)
- Selection of cell variants (mutants, transfectants) with altered lectin binding properties as models for dissecting glycosylation machinery and glycan functionality (glycomic profiling, functionally defined).

- For profiling global changes in mammalian and bacterial cell surface glycomes.
- Fractionation of cell populations.
- Modulation of the proliferation and activation status of cells and dissection of the involved signal pathways.
- Model substratum for study of cell aggregation, adhesion, and migration.

Medicine

- Detection of disease-related alterations of glycan synthesis by lectin cyto and histochemistry.
- Histo-blood group typing and definition of secretor status.
- Quantification of aberrations of cell surface glycan presentation, e.g. in malignancy.
- Cell marker for diagnostic purposes including marking infectious agents (viruses, bacteria, fungi, parasites).
- Cell marker for functional assays to pinpoint defects in cell activities such as mediator release.

1.1. Plant lectins with focus on the lectins studied in this research

On the basis of their monosaccharide specificity plant lectins are 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.

This thesis includes work that has been carried out on lectins from five plant species, which belong to three different plant lectin families. The legume lectin from *E.indica* and the type-II RIPs family lectins from *T.dioica* and *T.anguina* possess specificitity for Galactose/*N*-acetyl Galactosamine and hence are included in the Gal/GalNAc binding lectins. Whereas, the lectins from araceaous plants *S.guttatum* and *A.tortuosum* belonging to Mannose binding monocot lectin family have sugar specificity for complex sugars as asialofetuin and are included in the Mannose/Glucose binding lectins.

Galactose/*N*-acetyl Galactosamine speccific lectins from *T.dioica* and *T.anguina*

The lectins from *Trichosanthes dioica* and *Trichosanthes anguina* belonging to the cucurbitaceae plant family are categorized as the type-II RIP lectins.

1.3.1 Type-II RIPs

On the basis of the structural analysis and molecular cloning of lectin genes, plant lectins have been divided into seven families of structurally related protein as described above. Of these, one family is the type-2 ribosome inactivating proteins or simply type-II RIP (Van Damme *et al.*, 1998). These can be classified as Galactose/*N*-acetyl Galactosamine (Gal/GalNAc) binding proteins as they consist of two chains out of which one is a galactose/*N*-acetylgalactosamine specific lectin.

Ribosome-inactivating proteins (RIPs) are plant cytotoxins having N-glycosidase activity on the large rRNA of mammalian, fungal, plant, and bacterial ribosomes that irreversibly impair protein synthesis and thus kill the cell. RIPs act on prokaryotic and eukaryotic ribosomes and are the translational inhibitors found mainly in plants. They depurinate the adenine (A4324) in a conserved (Sarcin/ricin) loop of rRNA and thus inhibit the protein synthesis at the level of

elongation, by inhibiting its interaction with the transcription elongation factors. On the basis of their primary structures they have been classified into three types. Type-I RIPs consists of a single peptide chain, while the type-II RIPs are heterodimeric proteins consisting of two chains; the A-chain has RNA N-glycosidase activity while the B-chain is a lectin. Type-III RIPs consist of a single chain with an extended carboxyl-terminal domain with no known function (Nielsen & Boston, 2001).

1.3.1.1 General Characteristics

RIPs are widely distributed in the plant kingdom; they also exist in fungi, algae, and bacteria. All the RIPs are able to cleave a specific glycosidic bond on 28S rRNA. The action results in the inactivation of the ribosome, further causing cell death (Zhang & Halaweish, 2007). In addition to the N-glycosidase activity, RIPs also possess other activities such as RNase (Wang *et al.*, 2003) and DNase (PAP and cinnamomin) (Wang & Tumer, 1999) activity, superoxide dismutase and phospholipase activities as well. Progress made in the structural characterization of these RIPs has shown that some RIPs may exhibit strong anti-HIV activity (Au *et al.*, 2000, Wang *et al.*, 2003).

Type-I RIPs contain a single polypeptide chain with a molecular weight between 26 and 32 kDa, have alkaline isoelectric points (pI> 9), and are often glycosylated. Type-II RIPs are composed of two chains (chains A and B), linked with each other by a single disulfide (S-S) bond. Chain A of type-II RIPs possesses *N*-glycosidase activity. The chain B is a galactose-specific and *N*-acetylgalactosamine-specific lectin that aids RIP to bind to galactosides on the cell surface and hence facilitates entry of A chain into the target cell.

Type-II RIPs are plant toxins such as Ricin (from *Ricinus communis*) (Rutenber *et al.*, 1991; Katzin *et al.*, 1991; Rutenber & Robertus, 1991), Abrin (*Abrus precatorious*) (Lin *et al.*, 1969; Olsnes & Pihl, 1982; Tahirov *et al.*, 1995). The anticarcinogenic activity in abrin-a is because of the inhibition of protein biosynthesis (Olsnes & Pihl, 1982) which occurs when A4324 residue is cleaved from the 28 S ribosomal RNA (Endo *et al*., 1987). Though research on RIPs has been going on since the last hundred years, their biological roles are under speculation and hence these questions as to why these proteins are synthesized and stored in plants is yet to be answered. There are some reviews giving the recent advances of research going on in RIPs and which also cover their biological activities, presence in different organisms, role in plant cytotoxicity, their structural organization, basis of ribosome inactivation and so on (Nielsen & Boston, 2001; Peumans *et al.*, 2001; Van Damme *et al.*, 2001; Girbés *et al.*, 2004; Hartley & Lord, 2004; Stirpe, 2004; Stirpe & Battelli, 2006).

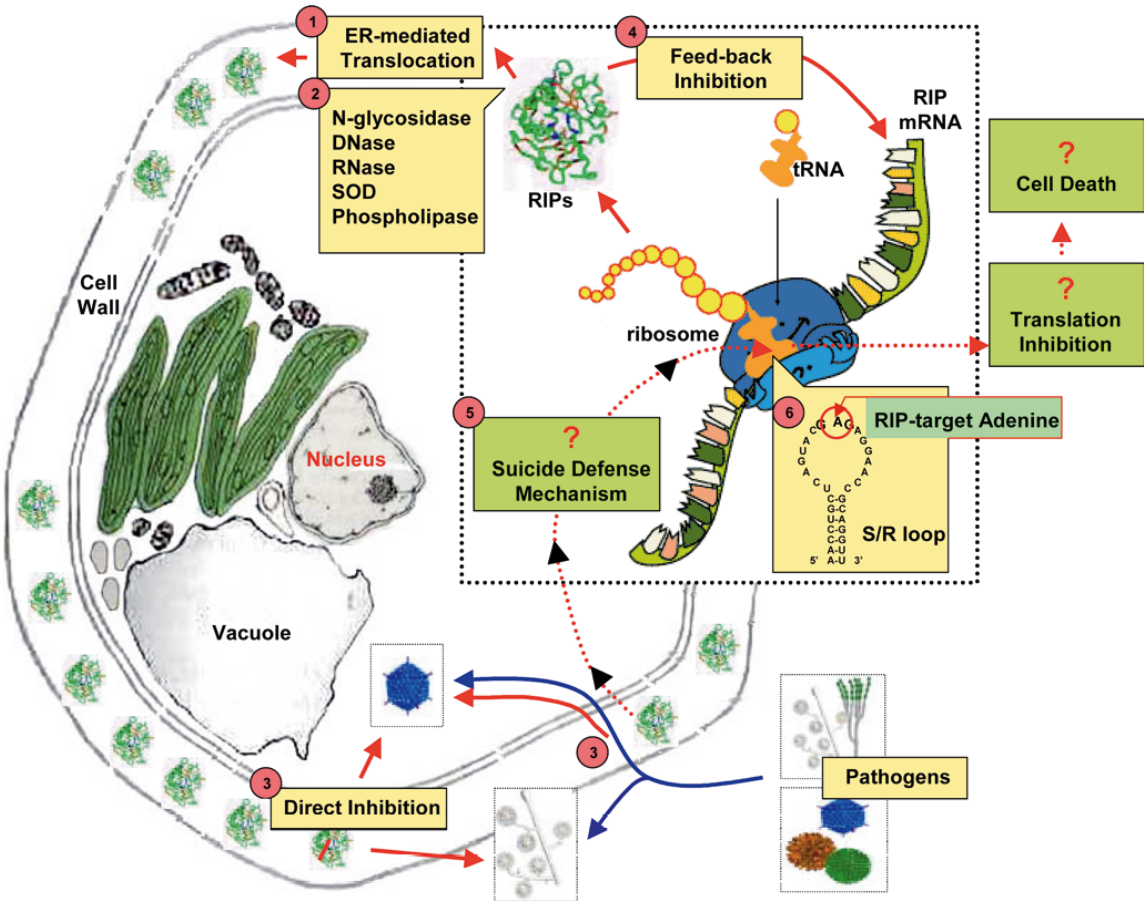


Fig.1.11 Cellular role of type-II RIPs and mode of action *in vivo*

1.3.1.2 Structural features

The type-I RIPs that have been characterized include trichosanthin and trichokirin from *Trichosanthes kirilowii* (Zhang & Wang, 1986; Casellas *et al.*, 1988), Gynostemmin (*Gynostemma pentaphyllum)*, Lagenin from the seeds of Lagenaria siceraria (Wang & Ng, 2000), Luffangulin from the seeds of ridge gourd *Luffa acutangula* seeds (Wang & Ng, 2002), Sechiumin from the seeds of *Sechium edule* (Wu *et al.*, 1998), Luffin I from the seeds of *Luffa cylindrical* (Li *et al.*, 2003a), trichoanguin (*Trichosanthes anguina*) (Chow *et al.*, 1999) and pulchellin *from Abrus pulchellus* seeds (Castilho *et al.*, 2008). The RIPs whose structures are

reported include trichosanthin (Pan *et al*., 1992; Gao *et al*., 1994; Xia *et al*., 1993) and its complex with nicotinamide adenine dinucleotide phosphate (NADPH) (Xiong *et al*., 1994).

Type-II RIPs occur in several plant families such as Euphorbiaceae, Fabaceae, Viscaceae, Passifloraceae, Ranunculaceae, Lauraceae, Sambucaceae, Cucurbitaceae and Iridaceae. Ricin, isolated from the seeds of *Ricinus communis* was the first lectin characterized from this family (Rutenber & Robertus, 1991). Some other type-II RIP have been reported from *N. tabacum* (Sharma *et.al*, 2004), *Sambucus nigra* (RIP SNAI) (Van Damme *et al.*, 1997) etc. Other type-II RIPs whose structures have been reported are listed in Table 1.7.

Abrin-a from the seeds of *Abrus precatorius* has been well characterized structurally (Tahirov *et al.*, 1995). As in all type-II RIPs the a-chian of Abrin-a is an RIP while the B-chain is a galactose binding lectin having two distinct galactose binding sites. Asn51 and Asn260 are the main residues of the two galactose binding sites in Abrin-a (Chen *et al*., 1992). The sugar chains form bridge between two Abrin-a molecules. NAG at one end of the saccharide forms hydrogen bond with Asn100 and Tyr130. Asn140 forms two hydrogen bonds with the two NAG's. Gln216 also interacts with NAG. When the oligosaccharide contains mannose additional residues such as Asp27, Asn51, Asn31 and Asp30 are also involved in interactions with sugar molecule (Tahirov *et al.*, 1995).

Mistletoe lectin I (ML-I) from *Viscum album* is also a type-II ribosome-inactivating protein. The B-chain of ML-I can bind specifically glycolipids and glycoproteins terminating in galactose on the target cell surface membranes and thus facilitates the entry of the A-chain inside the cell. Studies have shown that the B-chain appears to be formed by gene duplication of a single galactose binding domain of carbohydrate binding proteins from divergent organisms.

There are four glycosylation sites in the entire lectin that can interact with the cell surface receptors, out of which three sites have been determined. Asn96, Asn112 and Asn136 form the three sites (Eschenburg *et al.*, 1998). Aromatic residues Trp 38 and Tyr 248 provide stacking interaction similar to ricin (Krauspenhaar *et al.*, 1999).

Table 1.7 Type-II RIPs with examples and reported PDB structures

Plant Species	**Name of RIP**	**PDB Structure**
Ricinus communis (castor oil plant)	Ricin; *R. communis* agglutinin	1AAI (Rutenber & Robertus,1991) 1RTC (recombinant ricin) Mlsna *et al*., 1993; Weston *et al*., 1993); 1RZO (Gabdoulkhakov *et al.*,2003)
Abrus precatorius (jequity beans)	Abrin; *A. precatorius* agglutinin	1ABR (Tahirov *et al.,* 1995); 2Q3N (Bagaria *et al.*,2006)
Trichosanthes kirilowii	TKL-1	1GGP (Li *et al.*, 2003b)
Momordica charantia	Momordin	1MOM (Husain *et al.*,1994)
Viscum album (mistletoe)	Viscumin HmRIP	1OQL (Niwa *et al.*, 2003) 1YF8 (Mishra *et al.* ,2005)

Sambucus ebulus (dwarf elder)	Ebulin b	1HWM (Girbes *et al.*, 1993; Pascal *et al.*, 2001;)
Sambucus nigra (elderberry)	*Sambucus nigra* agglutinin II	3CAO (Maveyraud *et al.*, 2009)
Cinnamomum camphora seeds	Cinnamomin	1VLC (Xie *et.al,* 2001; Azzi *et al.*, 2009).
Shigella dysenteriae	Shiga toxin	(1R4Q) (Fraser *et al.*, 2004)
Escherichia coli (0157:H7)	Shiga-like toxin	(1RP4) (Fraser *et al.*, 2004).

1.3.1.3 Applications of RIPs

RIPs are supposed to play an important role in plant defense, which was first shown when leaves of pokeweed plant had inhibitory activity against viral infection in plants. On the basis of the study carried out on RIPs, which shows that they possess toxicity to viruses, tumor cells, insects and plant fungal pathogens, type-II RIPs are largely implicated in drug development and plant biotechnology. RIPs have also been used as a powerful probe to study the structure of ribosomes. Previous studies revealed that many RIPs are involved in defense mechanisms in plant cells (Barbieri *et al.*, 1993**;** Kumar *et al.*, 1993) as they terminate protein synthesis under appropriate physiological conditions and thus play an important role in metabolic regulation. Some RIPs accumulate in tissues such as cotyledon, bark and root (non-reproductive tissues),

where they may function as storage protein. They have good pharmacological significance as they can be implicated in the treatment of cancer (Heiny *et al*., 1998) as well as in AIDS (Elhaggar, 1993; Zhao *et al.*, 1999). However, other than defense the physiological role of RIPs in plants has yet to be deciphered.

In view of role of RIPs in the current medical scenario for humans, we have selected a galactose specific lectin from a plant *Trichosanthes dioica* belonging to Cucurbitaceae, which was studied with respect to its biophysical characteristics and whose structure was investigated by x-ray crystallography. This study would definitely shed more light on type-II RIP family of lectins.

1.3.1.4 *Trichosanthes*: About the genera & lectins studied

Lectins, type-I and II RIPs are prominent in the seeds and tubers of plants belonging to the genus Trichosanthes (cucurbitaceae) such as *T.kirilowii, T.anguina, T.dioica,* and *T.cucumerina*. From *T.kirilowii* several type I RIPs have been characterized in detail and structures for some of them have been determined. They include trichosanthin (Zhang & Wang, 1986), trichokirin (Casellas *et al.,* 1988), alpha-Kirilowin (Wong *et al.*, 1996), beta-kirilowin (Dong *et al.*, 1994), karasurin (Kondo *et al.*, 2004), trichokirin-S1 **(**Li *et al.*, 2003b) and S-Trichokirin (Tai, 2000)**,** that have been characterized in some detail. All these proteins from *T.kirilowii* exhibit strong abortifacient activity. They show similar N-terminal sequences and also similar structural domains/features. Trichosanthin is shown to possess immunomodulatory, anti-tumor and anti-human immunodeficiency virus properties (Shaw *et al.*, 1994).

The structure of β-Momorcharin (PDB: 1CF5), a single chain type I RIP, determined at 2.55 Å resolution by molecular replacement method shows similarity to those of other single-chaine RIPs and it also shows a glycosylation site (Yuan *et al.*,1999).

Two type-II RIPs have been reported from *Cucurbita foetidissima*, foetidissimin and foetidissimin II (Zhang & Halaweish, 2007) which are probably isoforms, and texanin from *Cucurbita texana,* was a type-1 RIP (Zhang & Halaweish, 2007). Structural details are not known for these cucurbitaceae RIPs yet. A novel type-1 ribosome-inactivating protein (RIP) designated cucurmosin was isolated from the sarcocarp of *Cucurbita moschata* (pumpkin) (PDB code: 1BWH) (Hou *et.al*, 2008).

Another lectin from the same family is *Trichosanthes kirilowii* lectin-1 (TKL-1) which is a type-II RIP, and whose structure has been determined at 2.7 Å resolution (Li *et al.*, 2000) (PDB code: 1GGP). The structure has been solved using the coordinates of ricin (2AAI), abrin-a (1ABR) and Trichosanthin (1TCS). The spatial arrangement of TKL-1 structure resembles that of RIPs from ricin and abrin-a. The structure of A-chain of TKL-1 is more conserved, but the Tyr74 residue essential for N-glycosidase activity, is replaced with a non-aromatic residue, hence it lacks RIP activity. The other proposed active site residues required for RIP activity are present in TKL-1. A lectin from *Sambucus sieboldiana* has been reported that has sequence homology to type-II RIPs, but has no RIP activity as in the case of TKL-1 (Rojo *et al.*, 1997).

It is thought in general that type-II RIPs are more toxic to cells than type-1 RIPs because the former are assumed to have a lectin-type chain to assist the protein to enter the cells. RIPs can be used as tools for identifying and targeting cell surface markers, and can be coupled to antibodies so that they can be delivered directly to specific cells.

The type-II RIPs have been very useful for studies of endocytosis and intracellular transport in mammalian cells as antitumor and antiviral agents in biomedical and agricultural research. Two questions still remain unanswered for type-II RIPs. The first question to be posed

is whether the difference in chain B makes some RIPs enter cells easily, and the second is whether the structural difference in chain A gives rise to a higher enzymatic activity in site-specific depurination of nucleic acids.

Evolutionary relationship of RIP and RIP like proteins that have similar structure, but lack some potential residues for biologial activity needs to be elucidated. Structure of *T.dioica* lectin may throw some light on this relationship.

1.3.2 Legume lectins

Leguminosaea is the largest and one of the most important plant families with respect to economical and agricultural aspects. Lectins belonging to this family constitute the most wide and well studied family among the plant lectins. The lectins from this family are well conserved with respect to their primary sequences. The characteristics of legume lectins have been tablulated (Table 1.2).

Plant lectins, specifically the legume lectins have been shown to be responsible for the specific association between nitrogen-fixing rhizobia and leguminous plants (Hamblin & Kent, 1973; Bohlool & Schmidt, 1974). The attachment of rhizobia to the plant roots occurs by interaction between the bacterial surface carbohydrates and lectins present in the roots of the leguminous plant and is known as the "lectin recognition hypothesis" (Sharon & Lis, 2004). Recently it has been proposed that a specific attachment of the rhizobium is achieved through the interaction between some species-specific signal molecules (lipo-chitooligosaccharide) produced by the bacteria, called as nodulation factors (Nods), and a new type of a plant root lectin found in different leguminous plants but not in plants of different families (Kalsi & Etzler, 2000). Besides nitrogen fixation, legume lectins play other important role in the plant, such as protection from attack of fungi, predators, bacteria etc.

It was first shown in peanut lectin (PNA) that lectins could be employed for fractionation of cells into biologically distinct populations. Thus, PNA could be used as a marker for thymocyte maturation (Daniels *et al.*, 2002), and it was later demonstrated that mouse bone marrow and spleen cells fractionated by PNA and soybean agglutinin (SBA) could be utilized for bone marrow transplantation across histocompatability barriers. Another recent application of plant lectin is in the form of microarrays as a unique means for high throughput analysis of protein glycosylation (Rosenfeld *et al.*, 2007) and for profiling global changes in mammalian (Tateno *et al.*, 2007) and bacterial cell surface glycomes (Sharon, 2008).

1.3.2.1 Structural features

Legume lectins are an interesting family for study as they have highly conserved primary sequences; possess similarity at the level of tertiary structures, but different quaternary structures. The primary sequence of all lectins from this family shows at least 35% sequence similarity. All legume lectins show a very similar three-dimensional structure. They are largely used as model proteins for understanding the role of primary structure for determining their mode of quaternary organization. The characteristic structural motif in this family is the Jelly-roll also called as the **Legume lectin fold**, which is formed by two large β-pleated sheets and another β-sheet which sandwiches a carbohydrate site. Since legume lectins lack α-helices they are categorized as all beta-proteins.

Besides ConA (Edelman *et al.*, 1972; Hardman & Ainsworth, 1972) another extensively studied legume lectin is the peanut (*Arachis hypogaea*) agglutinin (PNA) (Banerjee *et al.*, 1994). Till date legume lectin from 210 plants are reported whose structures have been solved in native form as well as complexed with sugars and other ligands (http://www.cermav.cnrs.fr/lectines/) by biophysical techniques, including X-ray crystallography, NMR and microcalorimetry. The x-

ray studies have shown that the folding of the polypeptide chains in the region of the carbohydrate-binding sites are also similar.

In this a curved seven-stranded β-sheet forms the front face, a flat six-stranded β-sheet forms the back face, both of which are interconnected by turns and loops. Monosaccharide binding site is conserved in all legume lectins and it is formed by four loops on the upper part of the dome shaped structure. A main hydrophobic core is located between the two β-sheets, while another hydrophobic core nests between the front β-sheet and a large loop (a V-loop). Metal binding sites in legume lectins are very well conserved; in which Ca^{++} and transition metal ion are approximately 4.5 Å apart and bridged by two aspartate residues. Both the metal ions are bridged to four residues and two water molecules.

Canonical legume lectin dimer is a structure that was first reported in ConA and then PNA, and has been observed in most legume lectin structures. The two monomers in these structures associate such that it results in a handshake mode of structural organization that is commonly called as "Canonical legume lectin dimer". The six stranded back β-sheet associate such that a 12 stranded β-sheet is formed. The side chain–side chain and side chain–main chain interactions of residues in these β-sheets are responsible for stabilization of the whole structure. Some lectins as EcorL and GS-IV do not possess this stable canonical dimer in their structures. In EcorL and other *Erythrina* lectins such as ECL, canonical dimer cannot be formed as the glycosylation site lies within the canonical dimer interface and hence hampers its formation. While in GS-IV glycosylation state of the two dimers prevents the formation of canonical dimer.

Certain legume lectins possess a site in addition to the carbohydrate binding site for binding to hydrophobic ligands such as adenine, adenine–related plant hormones as cytokines,

ANS etc. These ligands bind with a very high affinity (10^5–10^6 M) in these sites, which is many times the affinity they exhibit for mono and oligosaccharides.

Legume lectins; is a family of proteins in which even small alterations in the same tertiary structure can lead to large variations in quaternary association. Similar tetrameric association has been found in the legume lectins SBA Soybean (Glycine max) seed lectin (1SBA) (Dessen *et al.*, 1995); PHA-L from *Phaseoulus vulgaris* (1FAT) (Hamelryck *et al.*, 1996); UEA-II from *Ulex Europaeus* (1FX5) (Dao-Thi *et al.*, 1998**)**; DBL (1ILU), DB58 and DLL from *D. biflorus*, *Robinia pseudoacacia lectin* (1FNY) and *Vicia villosa* lectin (Osinaga *et al.*, 1997**) (** (1N47) and is distinctly different from the structures of either the Glc/Man-binding lectins ConA (1I3H) or peanut agglutinin (2PEL).

1.3.2.2 Carbohydrate binding site

The carbohydrate binding site in these lectins is formed from the residues contributed by four different loops (A, B, C and D loops). One of these loops contains transition metals, calcium and manganese, which keep the amino acid residues of the sugar-binding site at the required positions. It is present as a shallow depression on the surface of the lectin and located close to the concave face of the seven stranded curved β-sheet. Amino acid sequences of this loop play an important role in the carbohydrate-binding specificities of these lectins. The carbohydrate-binding sites of these lectins consist of two conserved amino acids on β-pleated sheets. The folds and the features of the carbohydrate binding site are similar in all the legume lectins (Sharma & Surolia, 1997).

Legume lectins exhibit a wide variety of carbohydrate specificities. The available structures of these family members can therefore be used for studying the structural principles

that form the basis of protein-carbohydrate recognition. Members of this family belong to the following groups in terms of specificities: (i) Glu/Man specificity (ii) Gal/GalNAc specificity (iii) Complex and other specificities such as for fucose, chitobiose etc (Loris *et al.*, 1998).

The Gal/GalNAc binding legume lectins include EcorL (*Erythrina corallodendron*) (Shaanan *et al.*, 1991), DB58 and DBL (*Dolichos biflorus* (horse gram) (Schnell & Etzler, 1987; Schnell *et al.*, 1988), CS II (*Cytisus scoparius* (scotch broom)) (Konami *et al.*, 1992) and BPL (*Bauhinia purpurea* (camel's foot tree)) (Kusui *et al.*, 1991). The difference between the binding of Gal/GalNac and Glc/Man occurs as a result of the stereochemistry of Galactose and Glucose/Mannose at C-4 which renders differential specificity for the same binding site. In case of Galactose –OH at C4 is axially placed while in Glucose and Mannose it is equatorial (Elgavish & Shaanan, 1997).

Almost the same set of residues are present in legume lectins with different sugar specificities and thus, the differences in binding a particular sugar molecule is attributed to the mode of interaction of the conserved residues with the sugar chain.

Galactose/*N*-acetyl Galactosamine speccific lectins from the genus Erythrina

1.3.2.3 *Erythrina indica* lectin

Among the legume lectins is a group of Galactose-*N*-acetyl galactosamine specific lectins that include lectins form the genus *Erythrina* (*E.corallodendron, E.cristagalli, E.indica*); the peanut agglutinin (*Arachis hypogaea*) (PNA); *Griffonia simplicifolia* lectin IV (GS-IV); Soybean agglutinin (*Glycine max*) (SBA); *Dolichos biflorus* lectin (DBL).

The two lectins from the genus that have been extensively studied structurally include those from *E.corallodendron* & from *E.cristagalli.* The structure of EcorL has been determined

in native form as well as in complex with lactose *N*-acetylgalactosamine, and *N*-acetyllactosamine at resolution 1.9 to 1.95 Å (Elgavish & Shaanan, 1998). The structure of recombinant lectin (rEcorL) has been determined at 2.55Å (Kulkarni *et al.*, 2004). EcorL shows a two fold higher affinity for GalNAc than for galactose.

E.cristalgalli lectin (ECL) has been shown to have higher affinity for fucosyllactose and fucosyllactosamine (Moreno *et al.*, 1997**;** Svensson *et al*., 2002) as compared to N-acetyllactosamine, lactose, N-acetylgalactosamine and galactose (Iglesias *et al*., 1982). The structure of ECL has been determined in native form as well as in complex with lactose and fucosyllactose *E.cristalgalli* lectin. ECL is mitogenic for human T lymphocytes (Iglesias *et al*., 1982). As legume lectins have potential role in a wide variety of aspects, we have chosen a Galactose/*N*-acetyl galactosamine specific lectin from *E.indica* seeds for structural characterization by x-ray crystallography.

1.3.3 Monocot Mannose Binding Lectins

The monocot mannose binding lectin family (MMBL) comprises of lectins from Amaryllidaceae, Lilliaceae, Alliaceae, Orchidaceae, Araceae, and Iridaceae (Van Damme *et al.*, 1995, 2000). They are shown to have sugar specificity towards α-D-mannose but show preference for α-1-3 or α -1-6-linked mannoses and they exhibit the highest affinity for oligomeric saccharides on glycoproteins (Shibuya *et al*., 1988). The members of this superfamily of monocot mannose-binding lectins possess highly conserved primary sequences and all of them contain the "β-prism II" tertiary fold, but they differ much in the pattern of organization of monomers into quaternary structures.

Monocot mannose binding lectins belonging to araceae family are a group of closely related proteins, and their sequence contains the highly conserved domains such as those present in Amaryllidaceae, Alliaceae, and Orchidaceae lectins. With respect to their specificity the Araceae lectins resemble the previously isolated Man-binding lectins from Amaryllidaceae, Alliaceae, and Orchidaceae species, except that they interact more strongly with glycoproteins such as asialofetuin.

1.3.3.1 General Characteristics

The family also goes by the name of Arum family and species in this family are often colloquially known as aroids. The family Araceae lectins have been purified and characterized in some details from *Arisaema tortuosum* Schott (Dhuna *et.al.,* 2005), *Arisaema consanguineum* Schott (ACA), *A. curvature* Kunth (ACmA), *Gonatanthus pumilus* D. Don (GPA) , *Sauromatum guttatum* Schott (SGA) (Shangary *et al.*, 1995), *Alocasia cucullata* (Kamboj *et al.* , 2005). The two lectins under investigation here are *Sauromatum guttatum* lectin (Voodoo lily) (SGA) and *Arisaema tortuosum* lectin (Himalayan cobra lily) (ATL) both being non-mannose binding monocot lectins isolated from the tubers of araceous plants *Sauromatum guttatum* and *Arisaema tortosum*.

1.3.3.2 Araceae lectins & anticancer activity

Lectins can be used in cancer research and therapy for their usefulness in detection of the alterations in the malignant cells, for reducing the cell tumorigenicity, and for prognosis of the immune status of the patient. Mitogenic lectins have implications for study of signal transduction in cells and for the *in vitro* study of events occurring during stimulation of lymphocytes. Thus investigation into this interaction between certain anticancerous and antiproliferative lectins and

their respective ligands on transformed cells will help in the elucidation of their function and mechanism of actions. The antiproliferative activity of araceous lectins opens new area of research for investigations and application of these lectins as tool in cancer research, immunomodulation, glycobiology and as a chemotaxonomic marker for diagnosis and lectin therapy in modern medicine.

Lectins specifically bind to glycoconjugates on normal and malignant cells, and can modulate many biological processes in cells, such as cell growth, adhesion, attachment, malignant transformation, and apoptosis. Lectins interact with specific carbohydrate structures on the tumor cell surface and may be used to differentiate malignant from normal cells (Gabius *et al.*, 1987). The first mitogenic agent to be described was PHA, the lectin from red kidney bean (*Phaseolus vulgaris*), by Nowell (1960). Most mitogens are lectins or plant glycoproteins such as Phytohaemagglutinin (PHA) (Hirschhorn *et al.*, 1963**)**, Concanavalin-A (Con-A) (Singh & Tevethia, 1973) and Pokeweed mitogen (PWM) (Reddy *et al.*, 1975). The mistletoe lectin has been noticed to possess cytostatic/apoptotic and immunomodulatory effects (Metzner *et al.*, 1987**)**, lectin from *Agaricus bisporus* reversibly inhibits proliferation of colonic cancer cell lines without cytotoxicity (Yu *et al.*, 1993); *Helix pomatia* agglutinin (HPA) has been useful in identifying cancers (Thies *et al.*, 2001).

It has been reported that seed lectins have been applied in the early detection of cancer as given below in Table 1.9.

Table 1.9. Plant lectins that are implicated for cancer detection

Source of plant lectins	Cancer they detect
Ricinus communis; Glycine max	Cervical cancer
Bauhinia purpurea ; Dolichos biflorus	Lung cancer
Ulex europeus	Prostrate and endometrial cancer
Arachis hypogaea	Colon cancer
Bauhinia pupurea	Thyroid cancer

1.3.3.3 Structural features

While the lectins from *Galanthus nivalis* (snowdrop) (GNA) (Van Damme *et al.*, 1987), amaryllis lectin (Chantalat *et al*., 1996), *Scilla campanulata* lectin (bluebell lectin) (Wright *et al*., 1996), all belonging to Amaryllidaceae, are tetrameric consisting of four identical subunits of 12.5 kDa, the lectins from the bulbs of garlic are either hetero-dimeric or homodimeric (Van Damme *et al*., 1992). But all the lectins subunits possess a similar structure and are found to present three potential carbohydrate binding motifs per subunit, each of which contains a consensus sequence signature Q*X*DXN*X*V*X*Y, essential for mannose binding (Ramachandraiah & Chandra, 2000). *Galanthus nivalis* lectin was the first lectin characterized from this family (Van Damme *et al.*, 1987).

Scilla campanulata agglutinin, a mannose-specific lectin from bluebell bulbs and a member of the Liliaceae family, has been crystallized in space group $P2_12_12$, and exists as a dimer in the asymmetric unit, but the lectin structure is a tetramer (Wood *et al.*, 1999). The

structure has been determined at 1.7 Å. The *Scilla campanulata* agglutinin, GNA and *Hippeastrum* (amaryllis) lectin all consists of threefold symmetric β-prism made up of three antiparallel four-stranded β-sheets. The structure of the GNA monomer can be described as consisting of a β-barrel made up of twelve-strands and it exhibits the characteristic β-prism fold. All of the three β-sheets, I, II, and III possess sugar binding sites with conserved residues. In each monomer β-sheet exist as bundles and are arranged perpendicular to the axis of the prism and three mannose binding sites are located in the cleft of these bundles. The tetramer thus posseses twelve mannose binding sites (Hester *et al.*, 1995). The structure of the GNA tetramer resembles a flattened crown with a central wide solvent channel. Some monocot lectins such as those from *Scilla campanulata* (Wood *et al.*, 1999), *Narcissus pseudonarcissus* (Sauerborn *et al.*, 1999) and garlic (Chandra *et al.*, 1999) possess three dimensional folds similar to those found in GNA.

A novel antiretroviral protein *Polygonatum cyrtonema* lectin (PCL) belonging to the monocot mannose-binding lectin superfamily has been crystallized using hanging-drop vapor-diffusion method. The crystals diffract to 2.0 Å resolution and belong to space group $P2_1$ (Ding *et al.*, 2008). Another novel mannose-binding monomeric lectin was isolated from the terminal corms of orchid *Gastrodia elata* (GAFP-1), $P2_12_12$. Four independent molecules were estimated to be present in the asymmetric unit (Liu *et al.*, 2002)

1.3.3.4 Lectins from *Sauromatum guttatum* (Voodoo lily) *Arisaema tortosum* (Himalayan cobra lily)

A few monocot mannose binding lectins having specificity for some complex glycoproteins (Shangary *et al*., 1995) have been reported. Previously lectins from *Gonatanthus pumilus* (Dhuna *et al.*, 2007)*, Alocasia indica* and *Sauromatum guttatum* have been shown to possess mitogenic potential towards human peripheral blood lymphocytes (Kamboj *et al.*, 1995).

Arisaema flavum lectin (wild cobra lily) shows potent mitogenic activity towards human and BALB/c splenocytes and antiproliferative activity towards murine cancer cell lines (Singh *et al*. 2004; Kamboj *et al.*, 2004). Recently lectins from *Arisaema* species have been shown to have anti-proliferative as well as mitogenic activities. (Singh *et al*. 2004, Kaur *et al.*2005). One such lectin is from *Arisaema tortuosum* which has complex specificity towards a serum glycoprotein asialofetuin and Nacetyl-D-lactosamine (LacNAc) and inhibits in vitro proliferation of human cancer cell lines HT29, SiHa and OVCAR-5 (Dhuna *et al.*, 2005); another lectin from *Arisaema jacquemontii* Blume has anti-insect and anti-cancer properties (Kaur *et al.*, 2006).

Two lectins which have been investigated in this thesis belong to araceae and have been shown to have anti-proliferative as well as anticancerous activity. The lectins from *Sauromatum guttatum* (SGA) and *Arisaema tortuosum* (ATL) have been characterized by biophysical methods for a better understanding of their structural features.

1.4 Lectin carbohydrate interaction

The three-dimensional structures of a large number of lectins, either alone or in complex with carbohydrate molecules, have been determined using X-ray crystallography. This has provided invaluable insights into the overall structure of the lectins, as well as the topology of the binding site. Protein-carbohydrate interactions are the language that cells adopt for inter-cellular communication. Lectin-carbohydrate interaction is mediated directly or indirectly by a variety of forces such as hydrogen bonds, metal coordination bonds, van der Waals' interaction, stacking and hydrophobic interactions (Sharma & Surolia, 1997**)**. Hydrogen bonds and other interactions occur through the hydroxyl groups present on the sugars. In a protein molecule these interactions occur via the side chain of Asp or Asn and the main-chain amide hydrogens and carbonyl oxygens (Elgavish & Shaanan, 1997). The protein-carbohydrate hydrogen bonds can be

both direct and water mediated. Lectin–carbohydrate interaction in general is of great interest from the pharmaceutical point of view and it can be applied to improve the targeting and delivery of active drugs to a particular site of action.

Lectins contain a primary binding site which is also called the combining site, and provides the cavity for interaction with a specific monosaccharide through multiple interactions. The monosaccharide-binding specificity of a lectin indicates its broad functional role in the organism.

1.4.1 Galactose-specific lectins

Galactose specific lectins have been identified in three plant families, the legume lectin fanily, the type-II RIPs and the Jacalin related lectin (JRLs) family. Among other lectin families the fungal lectins from *Sclerotinia sclerotiorum, Sclerotium rolfsii, Rhizoctonia solani* and *Marasmius oreadesi* are known to contain lectins specific for galactose and its derivaties and the structure of the galactose binding site has been determined for some of them. Microbila lectins bacterium *Fusobacterium nucleatum* and *Entamoeba histolytica* and amoebic lectin Discoidin II from *Dictyostelium discoideum* have been caharcterized with respect to their galactose binding sites. From animal, galactose binding lectins have been characterized and structurally studied from roman snail (*Helix pomatia),* a tunicate *Didemnum candidumi,* earthworm (*Lumbricus terrestris*), *Caenorhabditis elegans,* horsecrab (*Tachypleus tridentatus)*, *Cucumeria echinata* (lectins CEL I-IV) and a very well characterized class is that of Galectins. The structure of the combining site, and residues involved in sugar binding have been described for these lectins.

The three types of interactions identified for a galactose molecule at the binding sites are (Lis & Sharon, 1998):

(i) Formation of hydrogen bonds with a centrally located Asp residue of the lectin.

(ii) Hydrogen bonds with other hydrogen bonding donors or acceptors from lectin positioned closer to the 4` and 3` hydroxyl groups of galactose.

(iii) The polar groups of the bound sugar such as the hydroxyls and the oxygen atoms in the hexose ring form hydrogen bonds with the backbone carbonyls and the amide groups as well as with the carboxyl, hydroxyl and amide groups of the side-chains in lectin.

Side chain of an aromatic residue in the binding site that provides further hydrophobic and stacking interaction with the sugar.

In the combining sites of Gal-specific lectins, the interaction between the aromatic residues and galactose is due to the interaction occurring between the negative charge caused by the delocalized π-electrons of the aromatics ring and the partial positive charge contributed by the aliphatic protons of the hexose ring (Weis & Drickamer, 1996).

1.4.2 *E.corallodendron* lectin

In the structures of galactose/*N*-acetylgalactosamine specific lectins from *E. corallodendron* lectin, peanut agglutinin and soybean agglutinin, it is observed that the hydrogen bonding network in the binding site is similar irrespective of whether it is a galactose-binding or mannose binding lectin. In case of the Gal/GalNAc binding lectins a conserved triad of residues Asp-Asn-Gly interacts with the O3 and O4, while in Glc/Man the interaction is with the O4 and O6 atoms, hence the position of OH at C4 is determining in both cases (Loris *et al.*, 1998).

Side chains of three invariant amino acid residues at the combining site are essential for the carbohydrate specificity of these lectins. They include an aspartic acid, an asparagine, and an

aromatic amino acid (a phenylalanine, tryptophan, or tyrosine). The fourth residue is the Glycine that hydrogen bonds with the carbohydrate via the main-chain amide hydrogen.

Residues located on four different loops are the ones involved in carbohydrate recognition (Sharma & Surolia, 1997). The invariant amino acid residues Asp from loop A, and Asn and an aromatic residue (Phe or Tyr) in loop C provides the basic architecture to recognize the common features in C4 epimers. The invariant Gly in loop B together with one or two residues in the variable region of loops D/A holds the sugar tightly at both ends. Loss of any one of these hydrogen bonds tends to weaken the binding. The difference in binding galactose or glucose can be attributed to the size as well as the location of gaps in the amino acid sequence on either side of the conserved residues in D- or A-loop. The residue variations in loop B are important for discriminating between Gal and GalNAc (Rao *et al.*, 1998**).**

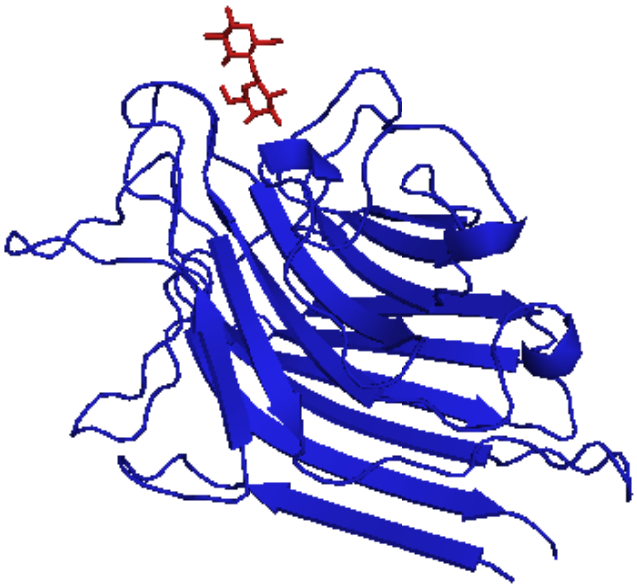


Fig. 1.12 Crystal structure of EcorL with bound lactose shown as a stick model (PDB code: 1AX1).

The residues involed for ligand binding in EcorL are Ala 88, Asp89, Gly107 and Asn133. Two key residues in carbohydrate binding are Asn and Asp. Four hydrogen bonds that stabilize the lectin-carbohydrate are formed by these two conserved residues and the main-chain NH of Gly/Arg. Hydrophobic interaction with the residues Ala88, Tyr106, Phe131, and Ala218 contribute to stabilizing the lectin sugar interaction. Besides some hydrogen bonds such as those between 3-OH and 4-OH of galactose and Asp89 (two bonds), Asn133 and Gly107 are formed. All these residues provide a basic framework for sugar binding that is conserved in legume lectin family (Shaanan *et al.*, 1991).

1.4.3 Jacalin

Seeds of jack fruit (*Artocarpus integrifolia*) contain a complex mixture of lectins. One of the lectins is jacalin, a homotetrameric protein of four identical protomers and it belongs to the Jacalin-related lectin family. The members of the jacalin-related lectin family show variation in carbohydrate specificity. They show mainly two different sugar specificities: (i) the galactose and (ii) the mannose/glucose (Peumans *et al*., 2001). Jacalin can actually be considered a polyspecific lectin as it shows specificity for galactose and its derivatives such as Galβ-1,3GalNAc, lactose, as well as glucose, mannose and oligomannosides. β-prism structure of jacalin is a sufficiently flexible structural scaffold which confers different carbohydrate-binding specificities to a single lectin. The relatively large size of the carbohydrate-binding site enables jacalin to accommodate monosaccharides with different hydroxyl conformations (Bourne *et al.*, 2002). The variability in the length and the sequence of the various loops is responsible for the varied carbohydrate specificity.

The residues in the loop region in jacalin that are involved in binding to the carbohydrate at the combining site are: 46–52; 76–82; 122–125. The surface loop Asn74-Tyr78 is involved in carbohydrate binding. The carbohydrate-binding region is formed by residues that are contributed from four different loops. Three of these loops are named as BL1, BL2 and BL3 (where BL stands for Binding loop) and the fourth is an influencing loop (IL) (Sankaranarayanan *et al.,* 1996; Jeyaprakash *et al.,* 2003). The loop BL1 in case of galactose binding lectins forms a hydrophobic lid to cover the binding pocket because of which the oligosaccharide is very well accommodated in the binding site, and this is absent in case of mannose binding lectins with similar binding site architecture. The length as well as the sequence of residues in the BL1 loop determines the carbohydrate specificity in JRLs, and the loop BL2 confers no particular specificity to the lectin. However, it contains an aromatic residue that is involved in stacking interaction with the sugar. Loop BL3 contains the sequence motif GXXXD, the invariant Asp is part of this loop. IL has a role in influencing the conformations of the residues involved in sugar binding and thus its specificity. This loop is longer in the case of galactose binding lectins such as Jacalin. Its interaction with BL1, BL2 and BL3 is essential for binding galactose and its derivatives, whereas absent in case of mannose binding lectins (Raval *et al.*, 2004**)**.

Phe47 present in the BL1 loop makes hydrophobic interactions with the galactose. Tyr78 is involved in stacking interactions with galactose. In Jacalin, the glycine residues located in the loops that connect the two β-strands are forming the sugar binding pocket. These glycines are placed facing the sugar residues, and they interact with sugar via the amide nitrogen.

The carbohydrate binding region in Jacalin is composed of three binding regions. These three binding sites in Jacalin can be categorized as: Primary site, Secondary site A and Secondary site B (Jeyaprakash *et al.,* 2003).

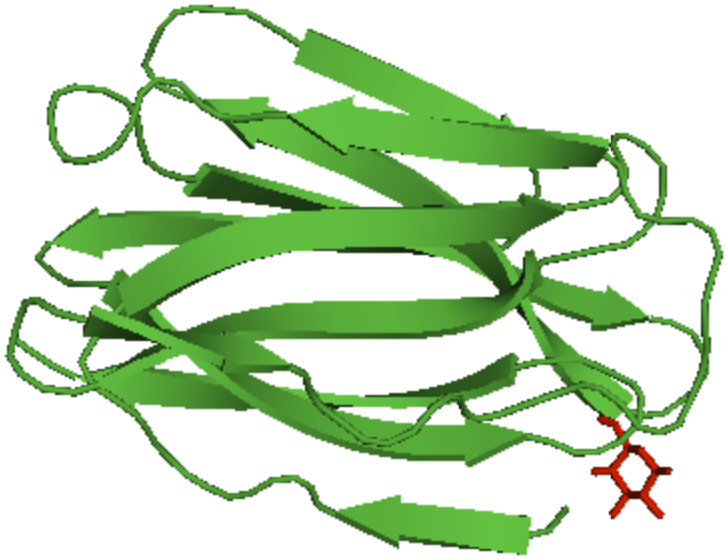


Fig.1.13 Crystal structure of Jacalin monomer with bound galactose shown as a stick model (PDB code: 1UGW).

Primary binding site: The residues in the primary binding site are the side-chains of Phe47, Tyr78 and Asp125; Tyr122 and Trp123 (main-chain nitrogen and oxygen atoms); and the amino terminus of the a-chain (α).

Secondary site A: This site involves the side-chains of the residues Tyr78, Tyr122 and Trp123. Residue Tyr78 side chain is thus involved in both the primary binding site and the secondary binding site A. these residues form a groove which accommodates the α-linked group of the sugar if present.

Secondary site B: This site is composed of Val79 (main-chain nitrogen and oxygen atoms), Ser119 (OG) and the C-terminus of the b-chain.

D-galactose, β-Met-Gal and 2-deoxy-α-D-galactose are the most potent inhibitors of the lectin. The order of oligosaccharide specifity of jacalin is reported as : *p* - nitrophenyl - α –

Galactose > α-Met – Gal> Melibiose (Gal α1, 6Glcβ)> *p* - nitrophenyl-β- galactose>GalNAc > stachyose (Gala1, 6Gala1, 6Glca1, 2Fru β) > Galactose > D-galactosamine (Bourne *et al.*, 2002).

Differences in the carbohydrate specificity here are defined by the hyper variability of the binding site loops. Each subunit contains one carbohydrate binding region. From one end of the β-prism, different loops contribute amino acids at the combining site of the lectin.

1.4.4 Ricin

The lectin from *Ricinus communis* is a type-II RIP which consists of A and B chains. The Ricin B chain (RTB) can bind oligosaccharides containing terminal galactose with much higher affinity compared to mono- or digalactosides and possesses two carbohydrate binding regions (Tahirov *et al.*, 1995). These two sites bind carbohydrate molecules forming hydrogen bonds and via hydrophobic interactions. The galactose binding pocket includes residues from two different domains, Trp37 from 1α and Tyr248 from the 2γ domain. The residues forming hydrogen bonds are Asp22 and Asp234 at the two sites, respectively. At the binding site Val23, Phe28, Trp37, and Ala45 are involved in hydrophobic interaction. For the 2nd binding site residues involved are Val235, Ala237, and Ile246 (Ganguly & Mukhopadhyay, 2007**)**. Gln47 and Gln256 strengthen the interaction between the aspartates and the galactose. Asn46 and Asn255 are also important residues in sugar binding.

RTB exhibits higher affinity for oligosaccharides (~100 times) than disaccharides (Hartley & Lord, 1994) and that can be explained by the extended stacking interaction involving residues in the proximity of the combining site such as Asp22, Asp25, Arg27, and Asp44, at the first binding site and Asp234, Arg236, Asp239, and Asp253 at the second binding site. RSA,

SNA-II, Ricin (B-chain) and earthworm lectin all contain similar domain structure (Candy *et al.*, 2001) having the β-trefoil structure.

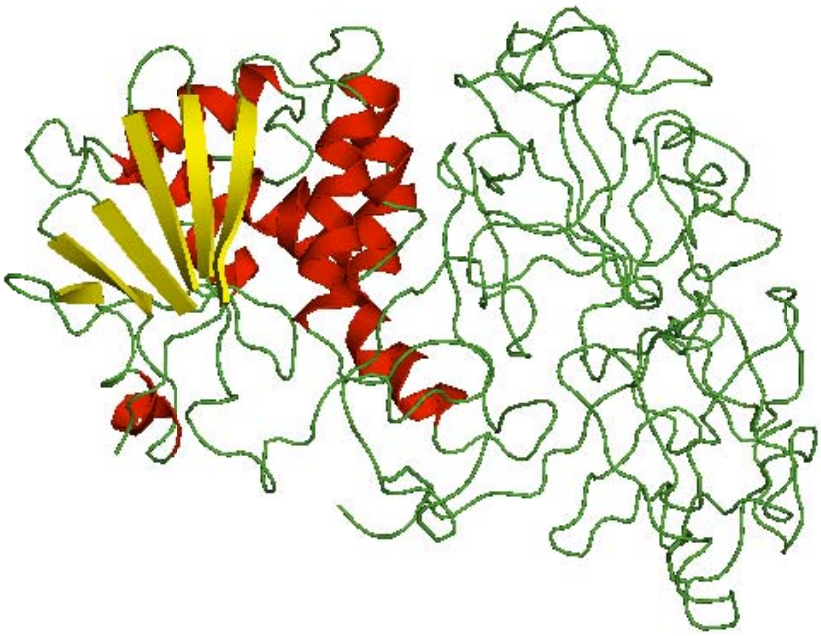


Fig. 1.14 Crystal structure of Ricin at 2.5 Å. The α-helices, β-sheets and loops are represented in red, yellow and green colours respectively. (PDB code: 2AAI).

1.4.5 *Rhizoctonia solani* agglutinin (RSA)

Lectin from the basidiomycetes *Rhizoctonia solani* (RSA) is structurally and evolutionarily related to the superfamily of proteins possessing a ricin type lectin motif. Thus on the basis of its similarity with plant lectins such as ricin in terms of sugar specificity and the structure of the combining site, it suggests an evolutionary relationship between some fungal and plant lectins. It shows structural homology to type-II RIPs and has been classified belongs to

ricin-B family on the basis of sequence similarity with the ricin-B chain. In the absence of structure it is difficult to identify the residues involved in carbohydrate binding in this lectin (Candy *et al.*, 2001**)**. The order of binding of various sugars is GalNAc >*p*-nitrophenyl-galactoside > methyl-galactoside >Gal >lactulose, raffinose, melibiose > lactose > GalNH2.

1.4.6 Galectin-9 from Mouse

Galectin-9 from *Mus musculus* is specific for galactose and derivatives as well as binds GlucNAc. It belongs to the galectins family comprising *β*-galactoside binding lectins; possessing similar carbohydrate binding domains; at least one carbohydrate recognition domain (CRD) per molecule; and has conserved sequences (Hatakeyama *et al.*, 1994). The galectin monomer shows striking structural homology with the legume lectin monomer even though there is no sequence similarity.

Galectins are widely distributed in species ranging from sponges to humans (Leonidas *et al.*, 1998). On the basis of their molecular architecture, in particular that of the CRD, galectins are classified into three groups: proto-, chimera-, and tandem repeat-type galectins. Some of the galectins that are well studied include Gal-1, Gal-2, Gal-3 Gal-7 and Gal-10 from humans (Allen *et al.*, 1987; Lobsanov *et al.*, 1993; Seetharaman *et al.*, 1998; Henrick *et al.*, 1998; Lopez-Lucendo *et al.*, 2004), toad ovary galectins (Bianchet *et al.*, 2000**)** etc**.** The amino acids located at the carbohydrate binding site in galectins and interacting directly with the bound carbohydrate are contained on the β-strands, and are highly conserved among all known galectin sequences. It is shown that the C4 hydroxyl group of galactose is hydrogen bonded to conserved residues His45, Asn47, and Arg49 (Allen *et al.*, 1987).

In this protein the carbohydrate recognition domain is formed by β-strands S4, S5, and S6, and the carbohydrate recognition mechanism is also similar to those of other galectins (Nagae *et al.*, 2006**).** His60 and Asn62 are two highly conserved residues that form hydrogen bonds with the O3 of galactose. Asn74 and His60 form hydrogen bonds, while Trp81 has stacking interaction with galactose. In case of lactose/GalNAc binding additional residues such as Arg64, Glu84, and Arg86 also form hydrogen bonds. Especially, Arg64 is shown essential for lactose binding.

In certain galectins (galectin-1, chick galectin CG16, fungal galectin CGL2) water is identified that stabilizes the amino acids interactions with sugar, just as the water molecules mimic the mode of carbohydrate binding at the combining site in the absence of the ligand (Nagae *et al.*, 2006**)**.

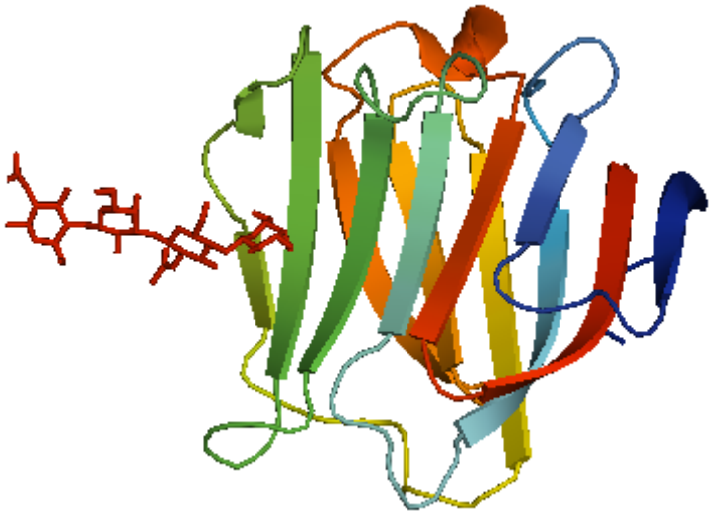


Fig. 1.15 Crystal structure of mouse galectin-9 N-terminal CRD in complex with N-acetyllactosamine dimer shown as a stick model (PDB code: 2D6O).

Human Galectin-1**:** The Gal-1 CRD consists of an antiparallel *β*-sandwich homologue of the CRD's in other galectins. The main amino acids involved in the protein-carbohydrate interaction are His44, Asn46, Arg48, His52, Asn61, Trp68, Glu71, and Arg73 (Di Lella *et al.*, 2007**).**

1.4.7 Galactosyl-binding lectins from the tunicate *Didemnum candidum*

Plasma lectins from tunicate *Didemnum candidum* DCL-I and DCL-II are shown to be specific for galactosyl residues and its galactose binding site has been characterized by extensive sugar inhibition studies (Vasta *et al.*, 1986**)**, but structural information is not available for these lectins.

They inhibit sugars in which galactose is the subterminal sugar. Galactose and oligosaccharides bearing nonreducing terminal galactose are the best inhibitors. The DCL-I combining site is comparatively smaller than the PNA combining site as it does not interact with the subterminal monosaccharide hydroxyls on C4 and C6 as PNA does. The difference in the interaction of PNA and DCL-1 can be explained as PNA interacts with hydroxyls on carbons 2, 4, and 6 of the galactosyl residues, while DCL-I interacts with 2, 3, and 4. PNA can accommodate bulkier residues at the combining site which has a large residue or group attached at the C-6, while the DCL-1 combining site does not accommodate C-6 (Vasta *et al.*, 1986**).** Asialofetuin (0.015mg/ml) and asialo-orosomucoid (0.030 mg/ml) were the best inhibitors. It inhibits sugars in which galactose is the subterminal sugar. Galactose > GalNac; β configuration of the sugars is preferred over α. methyl-α-galactoside and free galactose, or where no subterminal sugar is present, the inhibition is stronger. α-linked galactosyl terminal residues are 3-5-fold better inhibitors than the ones bearing α-linked galactosyl residues.

1.4.8 C-type lectins: CEL-I & CEL-III

CEL-I, II, III and IV constitute one of the groups of such C-type lectins that has been isolated from the sea cucumber, *Cucumaria echinata* (Hatakeyama *et al.*, 1994). C-type lectins from invertebrates play a role in host defense mechanism. As in legume lectins some divalent metal ions like Ca^{++} and Mn^{++} are involved in carbohydrate recognition near the combining site (Loris *et al.*, 1998), while in C-type lectins metal ions like Ca^{++} play a direct role in carbohydrate binding (Kolatkar & Weis, 1996)

CEL-I plays an important role in various molecular-recognition processes in organisms. It consists of two identical subunits linked by a single disulphide bond and has two carbohydrate binding sites of equal affinities. This lectin has high specificity for GalNac (Hatakeyama *et al.*, 2002). Its carbohydrate binding affinity is in the order: GalNAc > Lactose > D-Galactose > GlucNAc.

CEL-III exhibits strong hemolytic activity and cytotoxicity through pore formation in target cell membranes. Two-thirds of the N-terminal is reported to have similarity with the B-chains of ricin and abrin, which are galactose-specific plant toxic lectins. It contains the β-trefoil fold like in ricin, though it has lesser sequence similarity with ricin (Kouzuma *et al.*, 2003). One-third of the C-terminal shows no similarity with any known proteins. It consists of three domains, two N-terminal carbohydrate binding domains and one C-terminal membrane–interacting domain. N-terminal region has binding activity for specific carbohydrates. After binding to cell surface carbohydrate chains, CEL-III oligomerizes through C-terminal domains, leading to the formation of ion-permeable pores by hydrophobic interaction with the cell membrane. It is made up of 4 identical subunits each 17 kDa and linked by disulphide bonds.

Carbohydrate recognition and binding in CEL-III is similar to that observed in C-type lectins. Bound carbohydrate is stabilized by stacking interactions of aromatic residues Tyr and Trp. It differs from ricin in that the carbohydrate binding in ricin does not require Ca^{++} while CEL-III is a C-type lectins hence it requires Ca^{++} for carbohydrate recognition. CEL-III exhibits affinity for sugars in the order GalNAc>lactose> lactulose> methyl-α-galactoside (Me- α -Gal)$_2$.

Residues involved in carbohydrate binding in CEL-III are Gln101, Asp103, Glu109, Asn123, Asp124 and Trp105. At each carbohydrate-binding site, 3-OH and 4-OH of the bound carbohydrates are coordinated with a Ca^{++} ion and simultaneously form hydrogen bonds with side chains of nearby amino acid residues (Hatakeyama *et al.*, 2007).

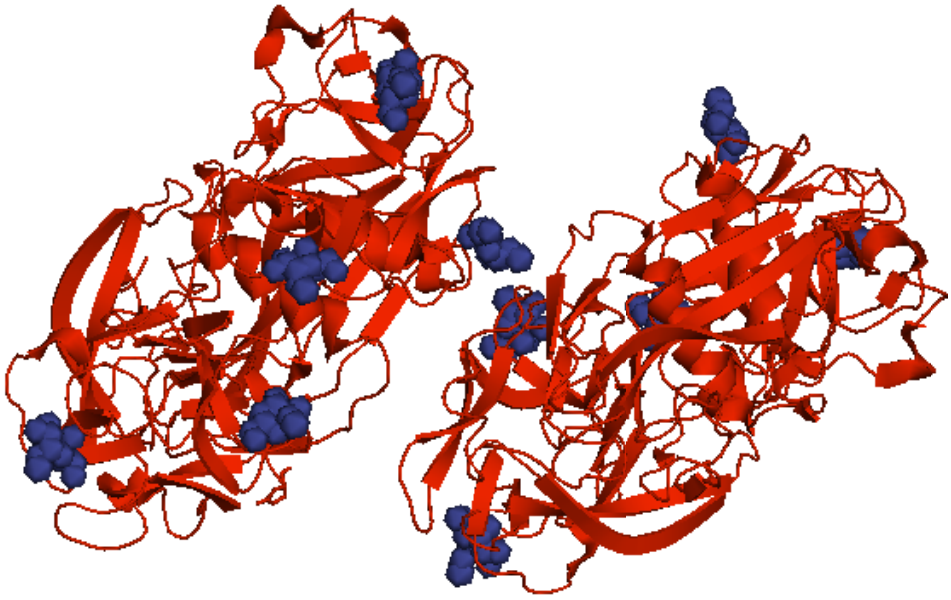


Fig 1.16 Crystal structure of CEL-III from *Cucumaria echinata* having the C-type lectin-like carbohydrate-recognition which conatin ricin-type beta-trefoil folds. The lectin is in

complex with methyl-alpha-D-galactopylanoside shown as space filling model (PDB code: 2Z49).

Table 1.9: Lectins from various organisms specific for either galactose or its derivatives are listed along with the source from which each is isolated, protein data bank identification (wherever available), lectin family it belongs to, the structural motif it possesses and the respective reference.

Lectin	Source	Sugar specificity	PDB code	Lectin family	Lectin domain/ fold	Reference
PLANT LECTINS						
Legume Lectins						
PNA	*Arachis hypogea* (peanut)	Gal/ GalNAc	1DVD, 1DVB,	Legume lectin	Legume lectin fold	Banerjee *et al.*, 1994
SBA	*Glycine max* (soybean)	Gal/ GalNAc	1SBA	Legume lectin	Legume lectin fold	Dessen *et al.*, 1995
GS IV	*Griffonia simplicifolia*	Gal/ GalNAc	1LED	Legume lectin	Legume lectin fold	Delbaere *et al.*, 1993
EcorL	*Erythrina corallodendron*	Gal/ GalNAc	1AXO, 1FYU	Legume lectin	Legume lectin fold	Shaanan & Sharon, 1991
WBA I	*Psophocarpus tetragonolobus* (winged bean)	Gal/ GalNAc	2DUO, 2DTY, 2DTW	Legume lectin	Legume lectin fold	Sharma *et al.,* 1996

DB58	*Dolichos biflorus* (horse gram)	Gal/ GalNAc	1LUL, 1G7Y	Legume lectin	Legume lectin fold	Schnell *et al.*, 1988
DBL	*Dolichos biflorus* (horse gram)	Gal/ GalNAc	1LU1, 1LU2	Legume lectin	Legume lectin fold	Schnell & Etzler, 1987
CS II	*Cytisus scoparius* (scotch broom)	Gal/ GalNAc	NA	Legume lectin	Legume lectin fold	Konami *et al.*, 1992
BPL	*Bauhinia purpurea* (camel's foot tree)	Gal/ GalNAc	NA	Legume lectin	Legume lectin fold	Kusui *et al.*, 1991
Jacalin-related lectins						
Jacalin	*Artocarpus hirsuta*	Gal, Man	2UH0, 2UH1	JRLs	β-prism-I fold	Jeyaprakash *et al.*, 2003
A.hirsuta lectin	*Artocarpus hirsuta*	GalNAc	1TOQ, 1TP8	JRLs	β-prism-I fold	Gaikwad, 2002
Type-II RIPs						
Mistletoe lectin I	*Viscum album*(Himalayan mistletoe)	Gal/GalNAc	1OQL	RIPs	β-prism-fold; RIP domain	Niwa *et al.*, 2003
Ebulin b	*Sambucus ebulus*	Gal/GalNAc	1HWM	RIPs	β-prism-fold; RIP domain	Pascal *et al.*, 2001
S.niger agglutinin I (SNA	*Sambucus nigra* (elderberry)	Gal/GalNAc	3CAO	RIPs	β-prism-fold; RIP domain	Maveyraud *et al.*,2008

II)						
Abrin-a	*Abrus precatorius*	Gal/GalNAc	1ABR	RIPs	β-prism-fold; RIP domain	Tahirov *et al.*,1995
Ricin	*Ricinus communis*	Gal/GalNAc	1AAI	RIPs	β-prism-fold; RIP domain	Rutenber & Robertus, 1991
Fungal lectins						
RSA	*Rhizoctonia solani*	GalNAc/Gal	NA	Fungal lectins	β -trefoil domain/ricin	Candy *et al.*, 2001
SSA	*Sclerotinia sclerotiorum* (phytopathogenic fungi)	GalNAc/Gal	NA	Fungal lectins	β -trefoil domain/ricin domain	Candy *et al.*, 2003
SRL	*Sclerotium rolfsii* (fungi)	GalNAc/GlcNAc	2OFD	Fungal lectins	β -trefoil domain/ricin domain	Leonidas *et al.*, 2007
Moa	*Marasmius oreades* (mushroom lectin)	Gal/GalNAc	2IHO	Fungal lectins	β -trefoil domain/ricin domain	Grahn *et al.*, 2007
Microbial lectins						
Discoidin II	*Dictyostelium discoideum* (amoeba)	Gal/GalNAc	2VMD		Discoidin (DS)domain ; β-sandwich structure	Aragão *et al.*, 2008
	Entamoeba histolytica	Gal/GalNAc	NA			Mann, 2002
	Fusobacterium	Gal	NA			Murray

	nucleatum (bacteria)					*et al.*, 1988
Animal Lectins						
DCL-1	*Didemnum candidum* (tunicate)	GalNAc	NA	C-type animal lectin	Ricin domain	Vasta *et al.*, 1986
Earthworm R-Type Lectin	*Lumbricus terrestris* (annelid)	Gal/GalNAc	2ZQO	C-type animal lectin	Ricin domain	Suzuki *et al.*, 2008
HPA	*Helix pomatia* (roman snail)	GalNAc	2CCV, 2CGZ	C-type animal lectin	β - sandwich lectin fold	Sanchez *et al.*, 2006
Tachylectin-2	*Tachypleus tridentatus*	GalNAc/GlcNAc	1TL2	C-type animal lectin	β - propeller fold	Beisel *et al.*, 1999
32-kDa GBP	*Caenorhabditis elegans*	β-galactoside-binding lecti ns	1WK1 (NMR)	C-type animal lectin	NA	Hirabayashi *et al.*, 1992
CEL-I	*Cucumeria echinata*	Gal/GalNAc	1WMY; 1WMZ	C-type animal lectin	C-type animal lectins	Hatakeyama *et al.*, 2002; Sugawara *et al.*,

						2004
CEL-II	*Cucumeria echinata*	Gal/Gal NAc	NA	C-type animal lectin	NA	Hatakeyama *et al.*, 1994
CEL-III	*Cucumeria echinata*	Gal/Gal NAc	1VLC, 2Z48, 2Z49	C-type animal lectin	β-trefoil fold, like Ricin ; but similar carbohydrate binding as in C-type lectins	Kouzuma *et al.* , 2003; Uchida *et al.*, 2004
CEL-IV	*Cucumeria echinata*	Gal/Gal NAc	NA	C-type animal lectin	C-type animal lectins	Hatakeyama *et al.*, 1995
Toad ovary galectin	*Bufo arenarum*	galactosides	1A78, 1GAN	Galectin	Legume lectin fold	Bianchet *et al.*, 2000
Mouse Galectin-9	*Mus musculus*	galactosides	2D6L	Galectin	Legume lectin fold	Nagae *et al.*, 2006
Human Galectin-1	*Homo sapiens*	galactosides	1W6N, 1W6Q	Galectin	Legume lectin fold	Lopez-Lucendo *et al.*, 2004
Human Galectin-3	*Homo sapiens*	galactosides	1A3K	Galectin	Legume lectin fold	Seetharaman *et al.*, 1998

Human Galectin-7	*Homo sapiens*	galactosides	3GAL, 2GAL, 1BKZ	Galectin	Legume lectin fold	Leonidas *et al.*, 1998

1.4.9 Mannose binding plant lectins

Glu/ Man binding lectins have been reported from a wide variety of plants and they have been well characterized structurally also. They are supposed to play an important role in plant defense by recognizing mannose terminating oligosaccharides in plants and microorganisms causing plant infection. Some legume lectins such as ConA, pea lectin, lentil lectin, *Lathyrus ochrus* lectin (LOLI) (Edelman *et al.*, 1972; Einspahr *et al.*, 1986; Loris *et al.,* 1993; Bourne *et al.*, 1990) also fall in this category of lectins that recognize mannose terminated glycans and possess a dome shaped monomeric arrangement, in which two monomers non-covalently associate forming a twelve-stranded – β-sandwich. The monosaccharide-binding site of subunit is formed by four loops that are responsible for the binding of simple sugars. Two residues are conserved in all legume lectins and essential for sugar binding are an Asp and an Asn (Asp81 and Asp125 in LoLI) (Barre *et al.,* 2001). The residues located on four loops and involved in specific recognition of mannose in LoLI are: Asp81, Gly99, Asn125, Gly208, Ala209 and Glu210. A network of seven hydrogen bonds formed by O3, O4, O5 and O6 of the sugar with these residues is responsible for stabilizing the sugar at the active site in this class of lectins. Besides this hydrophobic interaction with aromatic residue also plays an important role in stabilizing the sugar at the binding site by stacking interactions. Among the family of monocot-mannose binding lectin family the mannose binding site of GNA consists of four conserved

amino acid residues (Gln89, Asp91, Asn93 and Tyr97) which form a network of hydrogen bond with the O2, O3 and O4 of mannose. Asp91 and Asn 93 hydrogen bond with O2 of mannose and confer mannose specificity; a hydrophobic residue Val95 interacts with C3 and C4 of mannose. Lectins from the Jacalin-related lectin family such as Heltuba and some type-II RIPs such as those from iris bulbs also exhibit specificity for mannose and mannose terminated oligosaccharides (Barre *et al.,* 2001).

Chapter 2

Materials

&

Methods

2.1 Summary

This chapter presents the details of the various materials and methods used for the purification of *Trichosanthes dioica* lectin, crystallization of lectins from both *Trichosanthes dioica* and *Erythrina indica* to obtain good diffraction quality crystals, and for the biophysical experiments that have been carried out to compare the two Trichosanthes lectins: *T.dioica* and *T.anguina* lectins as well as for the two Araceae lectins: *S.guttatum* and *A.tortuosum*. This description also includes the methodology followed in X-ray diffraction, data collection, data processing, structure determination, structure refinement and analysis of the refined structure of both TDSL and EiSL.

The lectins from *Trichosanthes anguina* (SGSL)*, Erythrina indica* (EiSL)*, Sauromatum guttatum* (SGA) *and Arisaema tortuosum* (ATL) were already purified and obtained in pure form for further studies. The lectin from *Trichosanthes dioica* seeds (TDSL) was purified using various steps such as column chromatography using DEAE-Cellulose, affinity chromatography on Guar gum column, gel filtration on Sephadex G-100 and gel filtration on HPLC. Crystallization of TDSL and EiSL was carried out using hanging-drop vapour-diffusion method. The X-ray data were collected using Raxis IV^{++} detector mounted on a Rigaku rotating anode at NCL, Pune, India. The X-ray images were processed using DENZO and SCALEPACK programs (Otwinowski, 1993; Otwinowski *et al.*, 1997) in HKL suit. The crystal structures were determined using molecular replacement technique implemented in *AMoRe* (Navaza, 1994; Navaza *et al*, 1997). The REFMAC5 program (Murshudov *et al*., 1997) was used for structure refinement in cycles of REFMAC run followed by model fitting using QUANTA (Accelrys) and COOT (Emsley & Cowtan, 2004). The CCP4 suit suite (Collaborative Computational Project, Number 4, 1994) programs were used for almost all calculations.

2.2 Materials

Trichosanthes dioica seeds were purchased from the local market and were also ordered from United Chemicals and Allied products, Kolkata, West Bengal, India. Chemicals used for the purification of TDSL from crude extract of mature seeds were di-Potassium hydrogen phosphate, Potassium Hydrogen Phosphate, Ammonium sulfate (AS), Sodium acetate, Tris base, NaCl, NaOH, HCl and sodium acetate which were of analytical grade.

Sephadex G-100, Guar gum, DEAE-cellulose, phenyl-sepharose CL-4B, sodium cacodylate, polyethylene glycol 2000 MME (PEG 2KMME), polyethylene glycol 8000 (PEG 8000) etc. were purchased from Sigma, USA. Glycerol, bromo-phenol-blue (BPB), acrylamide, N,N'-methylene bisacrylamide, sodium dodecyl sulfate (SDS), acetic acid, methanol, TEMED (N,N,N',N'-Tetramethylethylenediamine), ammonium persulfate (APS), β-mercapto-ethanol, were also of analytical grade and obtained from local suppliers.

For crystallization trials the chemicals used were obtained from different sources such as : polyethylene glycols (PEG 400-20K), MPD, glycerol, ammonium sulphate, lithium sulfate (Li_2SO_4), magnesium chloride hexahydrate ($MgCl_2$.6H2O), sodiumcacodylate buffer, sodium acetate buffer, Tris HCl, isopropanol, ethelene glycol, n-octyl-β-D glucopyranoside, sucrose etc. All other chemicals used were of analytical grade and procured from local manufacturers. Multiwell trays used in crystallization experiments were purchased from Hampton research and Corning Inc. The slab gel electrophoresis unit was purchased from Tarson, India. Molecular weight marker kits for SDS-PAGE etc were purchased from Sigma chemical company, St. Louis, USA. The other specialized instruments were used in the experiments are mentioned in the appropriate places.

Silicon graphics workstations (Octane) were used for graphics display, for running the crystallographic programs and for other calculations. Programs from 'CCP4 suite' (Collaborative Computational Project, Number 4., 1994) were used extensively for calculations. In addition the program QUANTA (Accelrys) was used for graphics display, visualization of protein structures and model building.

2.3 Protein purification

The lectins from *Erythrina indica* (EiSL)*, Trichosanthes anguina* (SGSL), *Sauromatum guttatum* (SGA) *and Arisaema tortuosum* (ATL) were obtained in pure form for biophysical characterization and crystallizations respectively. *Trichosanthes dioica* lectin (TDSL) was isolated and purified from the mature seeds of the plant.

2.3.1 Extraction of lectin

The extraction of the lectin was carried out from 100 g dry mature seeds of *Trichosanthes dioica*. At each step of purification lectin activity by hemagluttination was checked to confirm the presence of lectin in the collected fractions. The seeds were finely powdered in a mixer-grinder and soaked in 500 ml 20 mM potassium phosphate buffer at pH 7.2 and 150 mM NaCl, stirred overnight at 4°C, and then filtered through four layers of cheese cloth. The filtrate was centrifuged in an SS-34 rotor at 10000 *g* for 20 min.

2.3.2 Ammonium sulfate fractionation

The supernatant containing the lectin was subjected to 0-80% ammonium sulphate precipitation (slow addition of ammonium sulphate to the suspension with continuous stirring). After overnight precipitation (8-10 h), the solution was centrifuged at 10000 *g* for 20 min in an

SS-34 rotor. The precipitate thus obtained was dissolved in 20 mM potassium phosphate buffer pH 7.2 and dialyzed against the same buffer for 20 h with three changes of buffer. The dialysate was centrifuged at 10000 *g* for 10 min in microfuge tubes and the clear supernatant having sufficient hemagluttination activity was used for further purification steps.

2.3.3 Gel filtration on Sephadex G-100

Gel filtration chromatography is a technique to determine size, homogeneity and molecular mass of globular proteins. The gel filtration matrix contains pores which permit the buffer and smaller proteins to enter into the beads of the matrix but excludes larger protein and protein complexes. Therefore, larger proteins migrate around the matrix particles and elute from the column faster than smaller proteins. This technique was performed according to the method of Whitakar (1963) and Andrews (1964). Sephadex G-100 matrix, bead size 40-120 μm (Sigma, USA) was allowed to hydrate for 48 h in 0.01 M PBS, pH 7.2 containing 0.05% sodium azide. The smaller particles, which remained suspended, were aspirated and discarded. The thick slurry was degassed using vacuum pump. The gel was packed in a chromatographic column (1.6□90.0 cm) at a flow rate of 20 ml/h, controlled by a peristaltic pump (Amersham Pharmacia). To estimate the void volume (V_o) of the column, 2 ml of blue dextran (Sigma, USA) at a concentration of 2 mg/ml dissolved in PBS containing 5% glycerol was loaded on the column. The absorbance of each fraction was read at 280 nm using UV-VIS spectrophotometer (Shimadzu, Japan) and plotted against the elution volume. The void volume (V_o) and the elution volume (V_e) were calculated by measuring the volume of eluent collected from the point of application of the blue dextran and protein respectively to the center of the eluent peak.

2.3.4 Affinity chromatography on Guar gum

Guar gum is a white to yellowish-white, nearly odourless, free-flowing powder. Guar gum was insolubilized by cross-linking with epichlorohydrine in alkaline medium. The remarkable capacity of guar gum, a galactomannan to bind α-galactosespecific lectin has been previously demonstrated by the isolation of *Ricinus communis* lectins (Appukuttan *et al.,* 1977). Guar gum contained β-(1→ 4) linked mannose units in the main chain with a single galactose moiety linked α-(1→ 6) to every alternate mannose unit (Dea *et al.*, 1975). Due to this structure it has 100-fold increased capacity to bind *Ricinus communis* lectin compared to Sepharose.

The cross-linked guar gum was packed in glass column (2.50 x12.0 cm), thoroughly equilibrated with phosphate buffer pH 7.2, and flow rate was maintained at 15-20ml/h. The supernatant obtained after 0-80% ammonium sulfate precipitation, followed by extensive dialysis in the buffer was loaded onto the column. The breakthrough was collected; it was also reloaded onto the column to ensure that lectin had completely bound to the guar gum matrix. Column was repeatedly washed with approximately 200 ml of the buffer, till the absorbance at 280nm of the elute was below 0.02. 0.2M Lactose was passed through the column maintained at 4°C and 2 ml fractions were collected. The absorbance was measured at 280 nm and the fractions having OD > 0.1 were pooled and dialyzed against phosphate buffer pH 7.2.

2.3.5 Ion Exchange chromatography on DEAE Cellulose

The fractionation of proteins by ion-exchange chromatography depends upon differences in the charge of different proteins. The charge of a protein depends upon the number and type of ionizable amino acid side chain groups. Each ionizable side chain group has a distinct pKa; that is, the pH at which half of the population of the ionizable group is dissociated. Therefore the overall number of charges on a particular protein at a particular pH will depend upon the number and type of ionizable amino acid side chain groups it contains. Since, by definition, different

proteins have different amino acid compositions, they will tend to have different charges at a given pH and so can be fractionated on this basis.

DEAE Cellulose is a type of Ion exchange chromatography using diethylaminoethyl cellulose (DEAE-cellulose) as a positively charged resin. An ion-exchange resin consists of an insoluble matrix with charge groups covalently attached to it.

The DEAE matrix was activated by washing first with 0.1M NaOH, followed by 0.1M HCl and again with 0.1M NaOH and after each step the matrix was washed thoroughly with distilled water. After activation, the matrix was packed in a glass column (3 X 20 cm), and the column was equilibrated with 20mM potassium phosphate buffer pH 7.2. The dialyzed ammonium sulphate fraction was loaded on this pre-equilibrated column washed with the same buffer and the breakthrough was collected. The fractions were collected until the absorbance at 280nm was less than 0.02. Column was then washed with increasing concentration of NaCl (100mM, 200mM, 300mM etc) starting with 100mM NaCl at pH 7.2, fractions were collected till OD at 280 nm < 0.02. Final wash was given with 1M NaCl to remove all other unbound matter to the column. The column was then washed with the starting buffer. At each step presence of lectin was monitored by hemagluttination. Fractions having activity were pooled, dialyzed against 20mM potassium phosphate buffer at pH 7.2 and concentrated.

2.4 Biochemical and biophysical techniques for protein characterization

2.4.1 Biochemical Assay

2.4.1.1 Erythrocyte preparation

Human erythrocytes of A, B and O blood groups were washed 5 to 6 times with 20 mM potassium phosphate buffer at pH 7.2 containing 150 mM NaCl. A 3% (v/v) suspension of the

erythrocytes in the above buffer was treated with different enzymes (pronase 0.05%, trypsin 0.05% and neuraminidase 0.1 U/ml) at 37 ºC for 1h, washed 3 times with the same buffer and used for further studies.

2.4.1.2 Hemagluttination assay

Hemagglutination assays were carried out using normal human A (+) or O (+) erythrocytes as well as Pronase treated erythrocytes. Two-fold serial dilution of the lectin in saline was done in a microtitre plate. 50 µl of a 3% erythrocyte suspension was added to the serially diluted lectin in saline and incubated for 30-60 min at room temperature. Hemagglutination was recorded after one hour.

2.4.1.3 Hemagluttination inhibition assay

In case of hemagglutination inhibition assays, in addition to the above described hemagluttination assay the incubation of the lectin with the sugar solution is also performed. Serially diluted sugar solutions (25 µl) were pre-incubated for 15 min at 27 °C with 25 µl of the lectin (8 U). Erythrocyte suspension (50µl) was then added, mixed and the plates read after an hour.

2.4.2 Biochemical techniques for protein characterization

The following are the biochemical techniques that have been used for estimating the proteins concentration and confirming its purity by electrophoresis.

2.4.2.1 SDS - polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis was carried out according to Laemmli (1970) using 15% gel containing 0.1% (w/v) SDS. Samples were pre-incubated with 2% (w/v) SDS, 10% (v/v) glycerol, 4.5% (v/v) β-mercaptoethanol, 0.025% (w/v) bromophenol blue and 60mM Tris buffer (pH 6.8), for about 4-10 min in a boiling water bath (95-100ºC). Gels were run at room temperature at a

constant voltage of 100 V. The gel was then fixed overnight in a mixture of methanol, glacial acetic acid and distilled water (4:1:5). It was then developed using two different procedures. When the protein concentration in individual lanes was comparatively high, Coomassie Blue R-250 was used for staining. In this procedure staining was done with 0.25% Coomassie Brilliant Blue R-250 in 40% (v/v) methanol and 10% (v/v) glacial acetic acid, and destaining was repeatedly done with a solution consisting of 40% (v/v) methanol and 10% (v/v) glacial acetic acid to visualize clear bands. In case of low protein concentration protein bands were visualized by silver staining according to Blum *et al.* (1987).

2.4.2.2 PAGE run at different pH

Polyacrylamide gel electrophoresis (PAGE) of the non-denatured lectins (native) was carried out at a slightly acidic pH (4.5) to analyze the purity of each lectin. Electrophoresis was conducted on 10 % polyacrylamide gels as per the methodology of Reisfeld *et al.* (1962). Electrophoresis was performed in 0.35 M β–alanine–0.14 M glacial acetic acid, pH 4.5 at a constant voltage of 100 V at 15ºC for 6 hours with opposite polarity of electrodes i.e. from anode to cathode. The gel was then fixed overnight in a mixture of methanol, glacial acetic acid and distilled water (40:7:53) Thereafter, the protein bands in the gels were stained in 0.25% Coomassie Brilliant Blue for 2 hours and destained by frequent changes of destainer (fixative) until clear pattern of bands was seen. SDS-PAGE was performed in 10% (w/v) polyacrylamide gel at pH 7.2 according to Weber and Osborn (1969).

2.4.3 Protein estimation

Protein concentration in samples was estimated by the method of Lowry *et al* (1951) using bovine serum albumin (BSA) as the calibration standard. In this method different dilutions of BSA solutions are prepared by using BSA stock solution (1 mg/ ml), which range from 0.05-1

mg/ ml. After adding alkaline copper sulphate reagent, it is incubated, followed by addition of Folin Ciocalteau solution (reagent solutions, again incubated and finally the OD is read at 660 nm. The absorbance is plotted against the BSA concentration to get a standard calibration curve. The concentration of our protein was determined using the standard curve plotted above.

2.5 Biophysical techniques for stability studies and characterizing tryptophan microenvironment

2.5.1 Circular Dichroism measurements

CD measurements were performed in a Jasco 810 spectropolarimeter using a cuvette of 0.1 cm path length at 30°C, unless otherwise stated. The solution was scanned in the wavelength range 200-250 nm (50 nm min^{-1}, response time 2 s) three times and the data were averaged. The protein samples (80 μg/ml) were in 20 mM phosphate buffer pH 7.2 containing 150 mM NaCl. The effect of temperature on the proteins was studied by two different methods. In one set of experiments protein samples were equilibrated at temperatures ranging from 25 to 90 °C for 10 mins and the scan was recorded separately for each sample. For thermally denatured samples, after cooling the sample to room temperature a new scan was recorded under the same conditions, for testing the reversibility of thermal transitions. For determination of T_m, the spectropolarimeter was connected to a PTC343 Peltier circulating water bath (Jasco,Tokyo,Japan). The temperatures of the protein samples were increased at the rate of 1 °C/min for the temperature ranging from 25 to 90 °C and the ellipticity was recorded at 225 nm. To study the effect of denaturants the protein samples were incubated in 0-6 M Gdn-HCl and 0-8 M urea for 4 and 16 h, respectively, and after that the far UV CD spectra were recorded. Effect of pH on the secondary structure was studied by incubating the protein samples in 50 mM

buffers in pH range 1-11, as given above, for 16 h at 30 °C. All spectra were corrected by subtracting the respective buffer baseline.

2.5.2 Steady-state fluorescence spectroscopy

The intrinsic fluorescence of the protein was analyzed at 30°C in PerkinElmer LS-50B spectrophotometer equipped with a thermostatically controlled sample holder. The effect of chemical denaturants was studied by incubating the protein samples (1.8 μM) in 0 – 6 M Gdn-HCl and Urea respectively at 30 °C and pH 7.0 for 4, 8, and 24 h. Protein samples were also incubated in 0.05 M each of glycine-HCl buffer pH 1.0, potassium phosphate buffer pH 7.0, and glycine-NaOH buffer pH 10.0 at 25 °C for 16 h. for pH denaturation studies. pH denaturation studies were carried out by incubating the protein sample in 25 mM buffers of pH 2.0, 4.0, 6.0, 7.2, 8.0, 10.0 and 12.0 for 16h after which fluorescence spectra was recorded. For temperature denaturation studies the protein samples were incubated at temperatures ranging from 25°C to 90°C for ten minutes after which fluorescence spectra was recorded. The fluorescence emission of the protein was monitored by recording the scans after every stipulated period (10 mins) in 1 cm quartz cell in 300-400 nm range after excitation at 280 nm. Excitation and emission band pass of 7 nm was used. Reaction mixtures with no protein present were used to correct the base line.

2.5.3 NBS modification

Both the lectins (300 μg) in 100 mM sodium acetate buffer pH 4.5 were titrated with 0.02 mM N-bromosuccinimide (NBS) prepared in the same buffer. The reagent was added in five installments (5 μl each) and the reaction was monitored spectrophotometrically by monitoring

the decrease in absorbance at 280 nm. The number of tryptophan residues modified was determined by assuming a molar absorption coefficient of 5500 $M^{-1}cm^{-1}$ (Spande *et al*, 1967).

2.5.4 ANS-binding assay (Hydrophobic dye binding studies)

8-anilino-1-napthalene sulfonic acid (ANS) is a charged hydrophobic dye that binds to hydrophobic clusters in proteins and hence is largely employed for characterizing and detecting partially unfolded states in proteins (Ali *et al*, 1999; Gasymov & Glasgow, 2007). ANS emission spectra were recorded in the range 400-550 nm with excitation at 375 nm using slit widths of 5 nm for emission and excitation monochromators. The change in ANS fluorescence on binding to lectin was recorded at constant concentration of protein (2 μM) and ANS (50 μM). The spectrum of ANS in buffer was subtracted from the combined protein-ANS spectrum to yield the final spectrum.

2.5.5 Fluorescence Measurement for ligand (Adenine binding studies)

Adenine is known to bind the hydrophobic amino acid residues of proteins and quench the intensity of the fluorophore. Protein (2 μM) was titrated against a 10 mM solution of adenine dissolved in 0.1 M HCl and scans were recorded. Hydrophobic ligand (5-10 μl) was added till saturation. The binding of adenine to lectins was monitored by recording changes in fluorescence after excitation at 280 nm using slit width of 7 nm at 30 °C (Watanabe *et al*, 1992). Change in fluorescence (ΔF) was noted with increasing concentrations of adenine in independent readings. The fluorescence intensity was correlated for inner filter effect due to excitation light absorbance by adenine according to the following equation:

$$\log [C]_f = -\log[K_a] + \log [(F_O - F_C)/(F_C - F_\infty)]$$

In the above equation F_0 and F_C are the fluorescence intensities of the free protein and of the protein at an adenine concentration [C]. From the ordinate intercept of the double reciprocal plot of $F_0/(F_0-F_C)$ versus 1/[C], F_∞, the fluorescence intensity upon saturation of all the adenine binding sites is obtained. The plot of $\log[(F_0-F_C)/(F_C-F_\infty)]$ versus log[C], the abscissa intercept yielded the K_d value (the dissociation constant) for these interactions, the reciprocal of which gave the K_a (the association constant). This analysis of the binding data was carried out according to Chipman *et al* (1967).

2.5.6 Tryptophan accessibility: Quenching of fluorescence intensity by small molecular quenchers

Fluorescence titrations were carried out by adding 3-5 µl of acrylamide (5 M), potassium iodide (5 M), cesium chloride (5 M) and 2.5 M of succinimide to the protein sample (2 µM) prepared in 20 mM phosphate buffer pH 7.2. Fluorescence intensity was recorded after each addition. The iodide solution contained sodium thiosulfate (200 µM) to suppress triiodate formation. The excitation wavelength was set at 295 nm; the emission spectra were recorded in the range 300 to 400 nm with both the slit widths as 7 nm at a scan speed of 200 nm/min. To eliminate contribution from background emission, the signal produced by buffer solution was subtracted.

2.5.7 Lifetime fluorescence spectroscopy

Protein samples at a concentration of 0.5 mg/ml were employed for this experiment. Lifetime measurements were carried out on Edinburg lifetime spectrofluorimeter. For samples of native and denatured (8 M urea) lectins tryptophan residues were selectively excited at wavelength 295 nm and emission recorded at 340 nm. Protein samples containing final concentrations of acrylamide were subjected to lifetime measurements. The decay curves thus

obtained were analyzed by a multiexponential iterative fitting program provided with the instrument.

2.6 Crystallographic methods

With the recent developments in the field of structural genomics, attempts at deciphering the entire complement of genes of several organisms, promises to revolutionize the fields of medicine and biology in near future. After identification of all genes from genome, the next step is the determination of molecular and cellular functions of the protein that each of those genes encodes for. The function of protein can be correlated to its three-dimensional structure. The structure-based study of proteins is therefore an important procedure to realize the function of protein. The X-ray crystallography is a major tool to explore the structure-function of macromolecule from which along with other information we can further predict the protein's molecular function (that is its biochemical and biophysical roles). Solving the structures of proteins, will help to predict their function and further they can be classified into families or groups (based on folding domains). Thus the knowledge of the three-dimensional structures of protein targets emerging from genomic data, followed by macromolecular crystallography will accelerate drug discovery to a great extent.

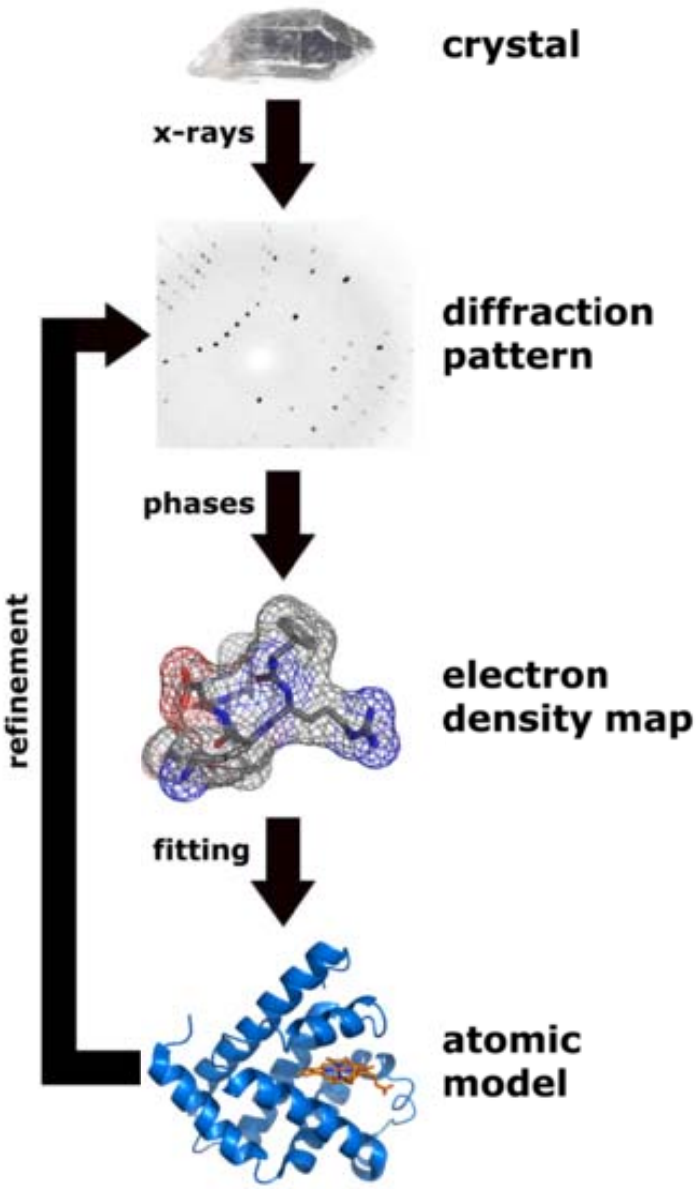


Fig 2.1 Schematic digram representing the steps showing how the final model is obtained from protein crystals (http://commons.wikimedia.org).

2.6.1 Protein crystallization

To solve the three-dimensional structure of a protein by X-ray crystallography, the first step is to obtain diffraction quality crystals of the protein. A crystal may be defined as an orderly three-dimensional array of molecules held together by non-covalent interactions. Proteins, like many organic and inorganic molecules, can be crystallized by applying appropriate conditions. The crystallization of proteins is a trial-and-error procedure in which a protein is slowly precipitated from its solution by appropriate precipitating agents. Factors that influence protein crystallization are temperature, pH, ionic strength of buffers and salts, precipitating agents, additives as well as possible unidentified factors which can affect the process and help in growing good quality protein crystals. The presence of impurities, crystallization nuclei, detergents, chaotropic agents etc. accelerates the process of crystallization of proteins from their solution (Drenth & Hass, 1998).

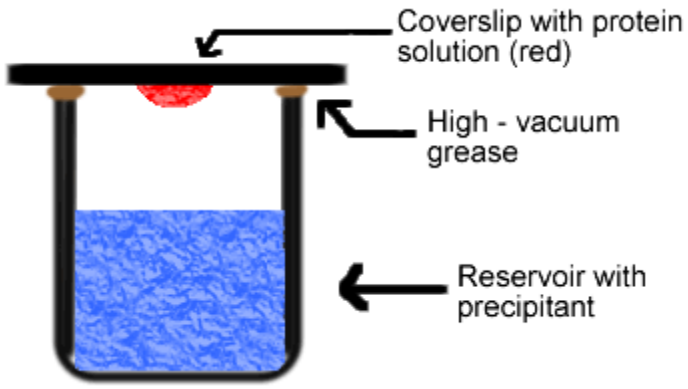


Fig. 2.2 Hanging-drop vapor-diffusion method for protein crystallization. Reservoir solution (blue) contains the buffer and precipitant. Protein solution (red) is suspended from the coverslip over the well solution containing mother liquor (www.bio.davidson.edu).

Protein crystallization occurs when the concentration of protein in solution exceeds the limit of solubility and so the protein can be considered in a superstaturated state. The purified

protein in solution undergoes slow precipitation. For crystallizing a protein, a precipitant which is an inorganic salt or an organic solvent, plays a very important role. At fixed pH and temperature, precipitant is allowed to diffuse into the protein solution, and as the protein in solution undergoes slow precipitation, crystals are obtained (Drenth & Hass, 1998). In crystals the individual protein molecules are aligned in regular repeating units called as the "Unit Cells", and thus a crystalline lattice is formed which is held together by non-covalent interactions (Rhodes, 2000).

There are several methods for crystallizing a protein such as batch crystallization, liquid-liquid diffusion, microbatch method, vapour diffusion (hanging-drop and sitting-drop), dialysis etc. among which hanging-drop vapour-diffusion method (Chayen, 1998) is the one commonly used and has been used for crystallizing the proteins described in this thesis (Fig 2.2). In this method, a solution of the protein along with the precipitant (less than 5 μl) is suspended on a coverslip over a well that contains the mother liquor with high concentration of precipitant. Water evaporates from the drop, increasing the concentration of the protein and the precipitant, and results in equilibrium between the drop and well solution. As a result both the protein and the precipitant get concentrated eventually resulting in nucleation and growth of protein crystals.

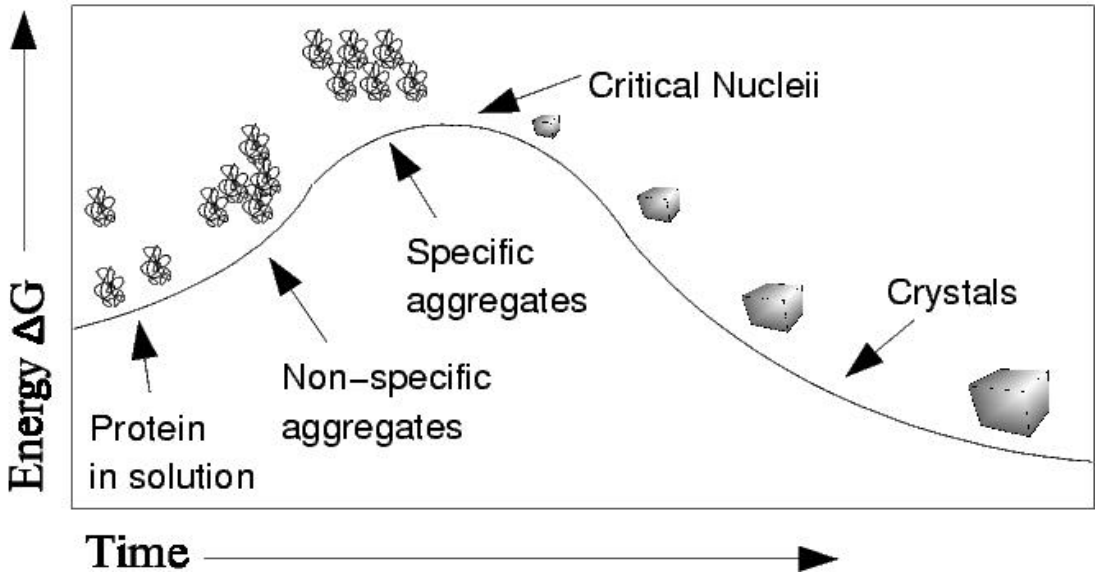


Fig 2.3 Formation of protein crystals from solution crystallization after crossing the energy barrier which must be overcome by proteins to crystallize. Saturation of protein increases (left to right) and hence the probability of nucleation is more. (http://molecularsciences.org).

The process of crystallization or crystal growth can be divided into three stages. The first stage is **nucleation**, in which a few molecules from the protein solution come together because of lowering of free energy and form a stable crystalline state, which has regular repeating units of molecules and is thermodynamically favorable. Formation of crystal form this nuclei depends on many factors such as concentration of the solute, temperature, the properties of the precipitant, pH and certain other factors. Second stage is **Crystal growth,** in which further

molecules or ions are added to the nuclei in a regular manner resulting in a good crystal. The final stage is the termination or **cessation of growth** of the crystal (Weber, 1991).

Precipitants used for crystallization are polyethylene glycol (available in a range of molecular weights 200, 400, 600 and also 3350, 6000 etc.), mono methyl ether derivative of the PEGs are also used such as (PEG 2K MME etc.), salts (like ammonium sulfate, lithium sulfate, tri-sodium citrate etc.) or organic solvent such as ethanol, 2-methyl-2, 4-pentanediol (MPD), acetone, isopropanol etc.

As described above, hanging drop vapor diffusion method was used for crystallizing the lectins TDSL and EiSL. The well solution used has 500µl or 1ml as per the availability as the screen solution. The protein concentration for both the proteins was in the range 10-15 mg/ml. 1-2 µl of the protein was mixed with either 1 or 2µl of the mother liquors on a siliconized coverslip, inverted and kept on the greased well, sealed and allowed to equilibrate. Several trials were carried out for both the lectins in which buffers at various pH and ionic strengths were varied with different precipitants also at varying concentrations. Crystallization conditions were screened using a sparse matrix screen, in which protein is subjected to widely varying pH, salts and precipitants. There are commercial screening kits available such as the screen from Hampton research and the Crystal clear screen (Molecular Dimensions Inc.) that were also used. Additives were tried for improving crystal quality, such as detergents like N-octyl-β-D-glucopyranoside.

2.6.2 Flash Cooling of Protein Crystals

After obtaining crystals, next stage is to characterize these crystals, by subjecting them to diffraction followed by data collection. During data collection, crystals are exposed to x-rays and they get damaged as the x-ray radiations are quite intense. This happens because the forces holding the crystal lattice together are weak (McPherson, 1982), and besides the disulfide bonds

holding the protein molecules are broken due to the free radicals produced in the process (Ravelli *et al.,* 2003; Ravelli & Garman, 2006). Cryocrystallography deals with collecting data at very low or cryogenic temperatures to overcome this problem. In this technique the crystal is cooled to cryogenic temperatures and cryoprotectants are added to prevent the formation of crystalline ice and thus minimize lattice damage during cryogenic cooling. Some commonly used Cryoprotectants are organic solvents such as MPD, methanol, ethanol, isopropanol, ethylene glycol, glycerol etc. (Garman & Schneider, 1997). Glycerol is one of the most common cryoprotectants, and has been employed as a cryoprotectant for approximately 50 typical protein crystallization solutions used in protein cryocrystallography (Garman & Mitchell, 1996). Other cryoprotectants used are ethylene glycol, MPD and lower molecular weight PEGs, glycerol, glucose, sucrose and xylitol etc.

2.6.3 X-Ray Diffraction

X-ray diffraction involves the set of experiments to measure intensities of Bragg reflections from the crystals. For this, in addition to crystal we require an x-ray source, an area detector (image plate or CCD), cryostream (if done at low temperature) and goniometer for oscillating the crystal. For obtaining a good quality diffraction data, we require high intensity x-ray radiation source as well as a high sensitivity area detector. The developments in the last few years that have been happening in this area have resulted in new technology for of this image plate and CCD detectors. High intensity beam from synchrotron sources, as well as crystallograophic softwares, have established x-ray crystallography as a very powerful tool for structural study of proteins (Beauchamp & Isaacs, 1999; Blundell *et al*, 2002).

The amount of details or the resolution that can be viewed is limited by the wavelength of the electro-magnetic radiation used. When electrons generated from a heated filament

accelerated by high voltage collide with atoms from a metal target such as copper, X-rays are produced. X-rays are electromagnetic radiation of wavelength about 0.1 – 10 nm (1 -100 Å) (Blundell & Johnson, 1976). For visualizing the atomic details of a protein molecule, the electro-magnetic radiation required is within wavelength of around 0.1 nm or 1 Å, which corresponds to that of X-rays, hence these X-rays are the EM radiation of choice (Blow, 2002). Since the diffraction from a single molecule is too weak to measure, a crystal is used to magnify the signal. Each molecule within the crystal contributes to the diffraction equally, and thus the diffracted X-ray beams are intense enough to be measured (Rhode, 2000). The X-rays from laboratory source used were having copper radiation of wavelength 1.5418 Å. If the crystal is well ordered, then diffraction will be measurable at high angles or high resolution and a detailed structure results. The X-rays are diffracted by the electrons in the structure and consequently the result of an X-ray diffraction experiment is a 3-dimensional map showing the distribution of electrons in the structure. Each reflection contains information on all atoms in the structure and conversely each atom contributes to the intensity of each reflection.

The data for TDSL and EiSL crystals mentioned above has been collected at Macromolecular X-ray facility at our home source which is located at NCL, Pune, India.

2.6.4 Data collection and Processing

In a macromolecular X-ray diffraction experiment the protein crystal is either soaked in a fine glass capillary (if data is to be collected at room temperature), or it is soaked in the cryoprotectant (in case of low temperature data collection) which is then lifted by a loop made up of nylon fibres. The capillary or loop is mounted on a goniometer, which allows the crystal to be positioned accurately within the X-ray beam and rotated as required. Since both the crystal and the beam are often very small, the crystal must be centered within the beam to lie within roughly

25 micrometres range, which can be done by a CCD camera focused on the crystal. The mounted crystal is then irradiated with a beam of monochromatic X-rays. When a crystal is mounted and exposed to an intense beam of X-rays, it scatters the X-rays into a pattern of spots or reflections (Bragg reflection) that can be observed on a screen behind the crystal. The intensities of these reflections may be recorded with photographic film, an area detector or with a charge-coupled device (CCD) image sensor. The peaks at small angles correspond to low-resolution data, whereas those at high angles represent high-resolution data. It is desirable to have high resolution data. The relative intensities of these spots provide the information that can be used to determine the arrangement of molecules within the crystal in atomic detail which is referred to as the space group. The diffraction pattern consists of reflections each of different intensity, and many such patterns need to be collected to cover all necessary crystal orientations. Some measure of diffraction quality can be determined at this point, such as the mosaicity of the crystal and its overall disorder, as observed in the peak widths. If there seems is a problem in the data collected from the crystal, it causes problems in solving the structure, and can be diagnosed easily during this processing. Multiple data sets may be necessary for certain phasing methods.

Knowing the wavelength and the diffraction angle of a reflection, its resolution *d* can be calculated:

$$\lambda = 2d_{hkl} \sin\theta$$

X-rays have wavelength of 1.54 Å which is in the range similar to molecular bond lengths. It is required to determine if a crystal diffracts to a good resolution, and the data obtained can be processed in the correct space group, before the final data set is collected. The initial data can often be used to determine the space group of the crystal, an important piece of information in planning data collection.

Data were collected using the rotation method employing a range of crystal rotations based on resolution, crystal orientation and mosaicity and processed using the HKL suite of programs. The oscillation images were displayed using the program XDisplayF. The Data processing steps are:

1. Indexing of the diffraction pattern and determining of the crystal orientation.

2. Refinement of the crystal and detector parameters.

3. Integration of the diffraction intensities of all the reflections.

4. Refining the relative scale factors between equivalent measurements.

5. Precise refinement of crystal parameters using all the data.

6. Merging and statistical analysis of the symmetry related reflections.

After integration, scaling of the data was done by the program SCALA from the CCP4 suite of programs (Collaborative Computational Project, Number 4, 1994).

The assessment of the high-resolution limit and quality of the diffraction pattern is by:

(i) Calculating the mean ratio of the intensity to the error; $I/\sigma(I)$.

(ii) Agreement between symmetry related reflections, *i.e.* R_{merge}.

The first parameter ($I/\sigma(I)$) provides an estimate of limit of limit of resolution.

The second parameter for accessing the X-ray data quality which is a global indicator is the merging R-factor (R_{merge}) or symmetry R-factor (R_{sym}). The merging R_{factor} is defined by the following equation (Blundell & Johnson, 1976):

$$R_{\text{merge}} = \Sigma_{hkl} \Sigma_i \mid I_i(hkl) - I(hkl) \mid / \Sigma_{hkl} \Sigma_i (I_i(hkl)$$

$\mathbf{I_i(\boldsymbol{hkl})}$ is the intensity of i^{th} observation of (h k l) reflection, and

I(*hkl*) is the mean intensity of all measured symmetry equivalents.

$\sum_{\mathbf{h}}$ is taken over all reflections.

$\sum_{\mathbf{i}}$ is taken over all observations of each reflection.

R_{merge} is commonly used to guide the decision during data reduction, such as determining upto what resolution the data are reliable. R_{merge} describes quality and reliability of data, it depends on overcoming conditions of low redundancy of data, omission of weak or partial reflections, use of sigma cut-offs in the data set all of which can lead to artificially low R_{merge} (Diederichs & Karplus, 1997). The parameter σ(I) assigned to each intensity derives its validity from the χ2's (or the goodness-of-fit). During data scaling the χ2 is brought closer to 1.0 by adjusting the parameters of the error model. R_{merge}, on the other hand, is an unweighted statistical parameter which is independent of the error model. It is sensitive to both intentional and unintentional manipulation of the data used to calculate it, and may not directly correlate with the quality of the data. The R_{merge} is less appropriate than the Chi-square, because of its dependency on the multiplicity of the data and on the symmetry of the crystal.

2.6.5 Sequence alignment

Sequence alignments of EiSL and TDSL and various other lectins form the their lectin families as well as other gala7ctose binding lectins were carried using available amino acid sequence from NCBI databse or from crystal structure determined (PDB). Sequence analysis for determining structure having highest homology for both EiSL and TDSL was performed using BLAST (Altschul *et al.,* 1997) at the NCBI site (http://www.ncbi.nlm.nih.gov/BLAST/). Protein blast was performed keeping the default values of all the parameters against as the nonredundant database, the PDB protein database and Swiss PDB database. The pair wise multiple sequence

alignment was carried out using the program ClustalW (1.83) available at European Bioinformatics Institute (http://www.ebi.ac.uk) and EsPript (Gasteiger *et al*., 2003).

2.6.6 Matthew's number

After determination of the space group and unit cell dimensions from the data, the next step is to determine the number of molecules in the crystallographic asymmetric unit and the solvent content of the protein crystals. An estimate of the number of molecules per unit cell (Z) can be made by method proposed by Matthews (1968), and it is called as the Matthews number from the equations given below.

$$V_m = (V \times z) / (MW \times n) \quad (1)$$

$$V_{solv} = 1 - (1.23 / V_m) \quad (2)$$

Where,

$\mathbf{V_m}$ is the Matthews number,

V is the unit cell volume,

MW is the molecular weight of the protein in Daltons in the asymmetric unit,

n is the number of asymmetric units per unit cell,

z is the Avogadro's number, and

$\mathbf{V_{solv}}$ is the solvent content of protein crystals

The Matthew's number and solvent content are calculated for the crystals of *T.dioica* and *E.indica* lectins. The number of molecules in the crystallographic asymmetric unit was estimated using Matthews Probability Calculator, with the resolution as an additional input (Kantardjieff & Rupp, 2003).

2.6.7 Structure determination

The data collected and processed as described above from a diffraction experiment is defined as the reciprocal space representation of the original crystal lattice. The position of each diffraction 'spot' is depended on the size and shape of the unit cell as well as the inherent crystal symmetry. The intensity of each diffraction 'spot' is recorded, and this intensity is proportional to the square of the structure factor amplitude. As we know that X-rays have wave properties, that is they have both amplitude and a phase. The structure factor is a complex number containing information relating to both the amplitude and phase of a wave. In order to obtain an interpretable electron density map, both amplitude and phase must be known. Electron density map obtained after processing is the first step towards building a starting model of the molecule. Both the amplitudes and phases of reflections need to be determined; but the information available from the diffraction pattern of the crystal contains only the intensities, from which the amplitudes can be calculated, but all phase information is lost. This is known as the Phase problem in X-ray crystallography. It is necessary to obtain phase information to calculate the electron density map.

The first step in going from experimental X-ray data to protein structure is the phase determination. Structure solution can be based on the many different techniques, depending on the data and information available. Many different programs can be used for the structure solution. The following are the different methods for structure determination:

2.6.7.1 *Ab initio* phasing (Direct method)

This method is generally used for small molecules and is hence applied to solve the phase problems in case of small proteins. If the resolution of the data is better than 1.4 Å

(140 pm), direct methods can be used to obtain phase information, by exploiting known phase relationships between certain groups of reflections.

2.6.7.2 *Ab initio* Methods (Anomalous dispersion and Isomorphous replacement)

Anomalous dispersion and Isomorphous replacement are the *ab initio* methods, which do not require any previous knowledge about the protein to phase the reflections. Depending on the number of heavy atom derivatives of the protein and/or the number of wavelengths used for obtaining the anomalous signal, these phase determination methods have been classified as follows:

(a) Isomorphous replacement methods

It includes:

(i) Multiple Isomorphous Replacement (MIR) (Blow and Crick, 1959),

(ii) Single Isomorphous Replacement and Anomalous Scattering (SIRAS), and

(iii) Multiple Isomorphous Replacement and Anomalous Scattering (MIRAS) (North, 1965).

Multiple isomorphous replacement (MIR)

Heavy atom addition can change the phases and diffraction intensities. If electron-dense metal atoms are introduced into the crystal, direct methods or Patterson-space methods can be used to determine their location and to obtain initial phases. Heavy atom derivatives are prepared from native crystals by soaking in solutions of heavy atom salts. After collecting native and derivative X-ray data sets the heavy atom positions have to be determined. This information is used to estimate phases of the native protein. When iterative refinement is done, phases are subsequently improved. MIR was first introduced in 1954 (Green *et al*., 1954) and since then it has been used widely for macromolecular structure determination.

The method of isomorphous replacement is the primary method for determining the relative phase of protein crystals by *ab initio* method. This method requires the X-ray diffraction pattern of the native protein crystal as well as that of two heavy atom derivatives. The necessary steps for isomorphous replacement are as follows:

1) Preparation of two heavy atom derivatives of the crystal.

2) Collection of x-ray intensity data for native protein crystals as well as for the derivative crystals.

3) Application of Patterson function to determine the heavy atom coordinates.

4) Refinement of the heavy atom parameters and calculation of the protein phase angles.

And lastly,

5) Calculating the electron density map of the protein.

(b) Anomalous dispersion methods

It includes two methods:

(i) Single Wavelength Anomalous Dispersion (SAD) (Hendrickson and Teeter, 1981),

(ii) Multiwavelength Anomalous Dispersion (MAD) (Hendrickson, 1991).

Structure solution by anomalous dispersion involves forming selenomethynyl derivative of proteins (Hendrickson *et al*., 1990). Recently, with the development of tunable synchrotron radiation sources (Hendrickson *et al*., 1985**;** Hendrickson, 1991) and sensitive detectors like Image plate and CCD to collect anomalous data. It has become a method of choice. Another advantage is MAD needs only one heavy atom derivative compared to MIR technique that requires more than one.

Atoms in proteins which are suitable for this purpose are sulfur or heavier atoms, for example metal ions in metalloproteins. The most commonly used atom for phase determination

via MAD is selenium, since it is usually possible to replace the natural sulfur containing amino acid methionine by selenomethionine.

At certain resonance frequencies some heavier atoms (e.g. Fe, Se, Hg, Pt, I) absorb radiation and thus cause differences in diffraction intensities. Data sets measured at different wavelengths from one crystal containing an anomalous scatterer are collected. The differences in intensities are used to calculate the positions of the anomalous atoms and finally phases can be calculated for the native protein.

By recording full sets of reflections at three different wavelengths (far below, far above and in the middle of the absorption edge) one can solve for the substructure of the anomalously diffracting atoms and hence the structure of the whole molecule. The use of the MAD technique in an experiment utilizing different wavelengths of X-rays generated at a synchrotron relieves the crystallographer from the traditional method of phase determination via Multiple isomorphous replacement (MIR), which involves the preparation of heavy atom derivatives in a trial-and-error approach.

2.6.7.3 Molecular Replacement method

It is a method of choice when structure of a homologous protein is available. If two proteins have a similar sequence (30% identity) or share a similar domain, the known structure (search model) can be used to determine the orientation and position of molecules within the unit cell of the unknown structure. The prerequisite for this method is the structure of a protein (search model) that has geometrical similarity with our protein whose structure is to be determined. The phases obtained by the method can be used to generate electron density maps. With the rapid expansion of protein data bank and with the large number of protein structures that are being deposited, the number of structures available is many; hence MR can be the

method of choice for protein structure determination in many cases. Pioneering work in this method has been carried out by Rossman and Blow (1962).

MR also makes use of the presence of non-crystallographic symmetry in the crystals to obtain phase information and to reduce uncertainties in the phase. The ideas developed by Rossmann & Blow (1962) have been extended by Bricogne (1976) to real-space symmetry averaging to improve the electron density.

The success of MR method in general relies on the following aspects (Navaza *et al*, 1997)

1. The completeness and the quality of data.
2. The extent of structural similarity between the molecular models and the actual molecules that constitute the crystal.
3. The size of each molecular model with respect to the content of the crystal cell.
4. The excellence of the criterion used as an indicator of the quality of agreement.

Several MR packages are available. *AMoRe*, X-PLOR/CNS, Merlot, PATSEE, GLRF are some of the packages. The principal difference between these packages is the way in which the calculations can be carried out, whether in real space or in reciprocal space (Turkenburg & Dodson, 1996). *AMoRe* that was designed with motto to deal with difficult problems has many novel functions incorporated in it. We have used *AMoRe* implemented in CCP4 suite (Collaborative Computational Project, Number 4., 1994) in the present study.

2.6.8 Refinement

After obtaining coordinates of the model from molecular replacement, the next step is refining the structure to get a good agreement between the observed and calculated structure factors. Refinement results in the minimization of the difference between the calculated (Fc) and observed (Fo) structure factors along with the application of certain restraints such as

minimizing the difference between the refined parameters and ideal parameters of the stereo chemistry (Drenth, 1994). Refinement techniques generally fall into two categories, depending upon whether the calculations are performed either in the real space or in the reciprocal space. Refinement methods based on reciprocal space is preferred over real space because these are computationally less expensive. To prevent the model from going into local minimum, interactive graphics was used for checking the fit of the model to the electron density. After every refinement cycle, the fitting of the model in the electron density is checked, so that the model does not go into local minima. An important crystallographic parameter is the R factor, which is monitored during every refinement cycle. The R_{factor} is defined as:

$$\mathbf{R} = \frac{\sum_{hkl} || F_o(hkl) | - | F_c(hkl) ||}{\sum_{hkl} | F_o(hkl) |}$$

Where, h, k, l, are the Miller indices of Bragg reflections, and the summation is over all the reflections. A well determined structure is expected to have an R_{factor} between 0.15 and 0.20.

2.6.9 Constraints and restraints

There are two types of crystallographic refinement: rigid body and overall B-factor. A model consists typically of five parameters for each atom: x,y,z, B, and Q. The triplet (x,y,z) specifies the position of each atom in an orthogonal coordinate system, B is the B-factor or temperature factor of each atom, and it is related to the thermal motion of the atom, and Q is the occupancy and it is the fraction of time that the atom spends at position (x,y,z).

In rigid body refinement, big sections of the protein, such as subunits, are considered separately as rigid bodies, which are refined (Head-Gordon & Brooks, 1991). In the simplest case, the entire protein is treated as one rigid body, which results in 6 degrees of freedom. Rigid body refinement procedure minimizes the R_{factor} value by refining three rotational and three translational degrees of freedom of the subunits or groups. It is possible to regard the entire molecule as a rigid entity and refine its position and orientation in the unit cell. Each molecule or subunit is treated as a continuous mass distribution located at the center of mass position.

In order to obtain a good quality model, certain constraint need to be introduced in the course of refinement, such as stereochemical constraint of amino acids bond length, bond angles, a peptide bonds torsion angles etc. (Engh & Huber, 1991). Constraints are used to hold a parameter close to the ideal value. A constraint can be supposed to be equivalent to a restraint having the value of force constant infinity. Restraints are considered when a specific parameter has restricted freedom, limited to a range of values, whereas, the parameter is constraint, when it can assume only a specific value. When applying NCS restraint, the molecules in an asymmetric unit are superposed by least squares superposition and the average coordinates of individual atoms are then computed. However, the application of NCS constraints imposes a strict restriction on the refinement of related subunits or parts, which is usually violated for macromolecular structure data with resolution better than 3.5 Å (Kleywegt, 1996). Hence in the initial stages, refinement is carried out with NCS restraint with different weights ranging from 0.3 to 0.2 and when the phases improves as indicated by drop in the R_{factor}, lower weighing term (0.1) has been applied.

2.6.10 Max-likelihood refinement

Refinement of structures was carried out using restrained maximum likelihood refinement implemented in the program REFMAC5 (Murshudov *et al.,* 1997). The program minimizes the coordinate parameters to satisfy a maximum-likelihood or least squares residuals. Refmac5 is a refinement module in which each cycle carries out two different steps:

(a) First is the estimation of the overall parameters of likelihood using the free set of reflections.

(b) And secondly it uses these parameters to build the likelihood function and refine the atomic parameters. To refine the atomic parameters only a working set of reflections are used.

Maximum likelihood method of refinement performs the calculation of the first derivative and makes an approximation of the second derivative of the likelihood function with respect to refinement parameters and then estimating the shifts to be added to the parameters. REFMAC also produces an output file with extension MTZ (named after three of its progenitors, McLaughlin, Terry and Zelinka) containing weighted coefficients for weighted mFo-DFc and 2mFo-DFc maps. About 5% of the reflections will be kept aside during refinement to calculate R_{free} for cross validation (Brunger, 1992). The values of standard deviations are calculated for the geometric restraints used during the refinement. The restraints used during refinement by REFMAC5 include distances, bond angles, torsion angles, peptide planarity, chiral volumes, Van der Waals radii and B values. These values are also the estimated standard deviations that determine the relative weights of the corresponding restraints.

The model that emerged after each cycle of model building and refinement was checked for unusual geometry at each residue, the peptide-flip values, and high temperature factors. The other parameters checked include values of torsion angles, including (Ramachandran angles φ

and ψ), side chain torsion angles such as χl, χ2 etc. which were verified between refinement cycles using the program PROCHECK.

Model building and manual fitting of the model were carried out using X-AUTOFIT module of QUANTA (Accelrys) and the addition of waters was with X-SOLVATE, again of QUANTA. Hydrogen bonds were calculated using the CONTACT program of CCP4 suite. Hydrogen bonded contacts were considered for distances between 2.5 and 3.5 Å between donor and acceptor.

2.6.11 Analysis and validation of structures

For the evaluation of the various features of refined model, it is necessary to validate the stereochemistry and geometry of the model during refinement and also finally after the model has been completely refined. The program PROCHECK (Laskowski *et al*., 1993) from the CCP4 suite of programs is routinely used for structural validation. The program compares and assesses the quality of the model with the available structures of similar or better resolution than the reference structure.

The output of the programs provides a residue by residue list of various parameters that define each residue and also the peptide bond which are evaluated after being compared to the standard values, and the deviations are also listed. These parameters include bond lengths, bond angles for individual amino acids in the structure, dihedral angles, cis-peptide bonds, the side chain and main chain angles etc.

Ramachandran plot (Ramachandran & Sasisekharan, 1968) is computed for the structure which lists out residues in the disallowed regions that can further be taken care of during fitting. Regions of the structure having higher deviation from the standard values and unusual

conformations are also listed by the program. Thus these can be corrected in the model and then further refinement can be carried out as required.

Chapter 3

Crystallization and

X-ray crystallographic studies

on a lectin from

Trichosanthes dioica

3.1 Summary

The lectin from *T.dioica* seeds has previously shown to be similar to the type-II RIPs on the basis of its biochemical characteristics. Another lectin from the cucurbitaceae family, the *T.kirilowii* lectin (TKL-1) has been structurally characterized and it has also been placed in this plant lectin family. The structure of TDSL has been determined at 2.8Å, and its structure has been solved by molecular replacement using the coordinates of TKL-1 (PDB code: 1GGP). The structure has been refined using the same model, but better values of R_{factor} and R_{free} have not been obtained in the absence of sequence information. Hence a well refined structure could not be obtained. The structure shows close similarity with type-II RIPs such as abrin-a and Ricin and it also possesses Tyr74 (of abrin-a) that is essential for adenine binding and thus the RIP activity, at the active site. Besides this other active site residues Glu164 and Arg167 are also observed in TDSL. All other active site residues and the invariant residues that have been described in all other type-II RIPs are also conserved in TDSL. On the basis of these structural features, TDSL has been categorized as type-II RIP.

3.2 Introduction

Understanding the structure function relationship of a protein is the crux of biological research especially in proteomics. Structural genomics includes screening of the complete genome of the organism, while on the other hand a structural study that aims at characterizing families of proteins from biologically important plant species also has its own significance. As already mentioned in detail in chapter 1, lectins have emerged as an important class of proteins in recent years for their importance in glycoconjugate research besides other applications such as in cell biology, cancer research, and other potential implications to clinical and biomedical applications. The legume lectins have > 200 PDB entries, while the other lectin families for

which several structures available are the β-trefoil lectins and β-prism plant lectins. For both these classes 32 PDB structures have been reported (http://www.cermav.cnrs.fr/lectines/).

Ribosome inactivating proteins (RIPs) belong to the class of β-trefoil lectins and are described as a group of naturally occurring plant proteins having the RNA-N-glycosidase activity. They depurinate the conserved sarcin/ricin loop of rRNA, thereby inhibiting protein synthesis at the stage of elongation and resulting in cell death (Stirpe *et al.*, 1992). The most well characterized activity of RIPs is the translational inhibitory activity, which is a result of a N-glycosidase on the 28 S rRNA of the eukaryotic ribosome, that in turn triggers the splitting of the A(4324) (or an equivalent base in other ribosomes), which is key residue for translation (Endo *et al*., 1987; Barbieri *et al.*, 1993). Other enzymatic activities found in RIPs are lipase, chitinase and superoxide dismutase. The proteins belonging to the RIP family share sequence and structural similarities and are shown to be phylogenetically related.

RIPs are found in different parts of plants ranging in concentration from a few micrograms to several hundred mg per 100 g of plant tissues. The highest number of RIPs has been reported in the plant families Caryophyllaceae, Sambucaceae, Cucurbitaceae, Euphorbiaceae, Phytolaccaceae and Poaceae. On the basis of their structures, RIPs have been classified into three types as already described (Fig3.1). Here we will focus more on type-II RIPs which contain an RIP domain (A-chain) and a lectin domain (B-chain). These possess either 2 or 4 polypeptide chains and have MW of approximately 60 kDa or 120 kDa. As the RIPs are cytotoxic they can be exploited for designing immunotoxins for therapeutic treatments of cancer, AIDS and other viral diseases largely affecting the human population.

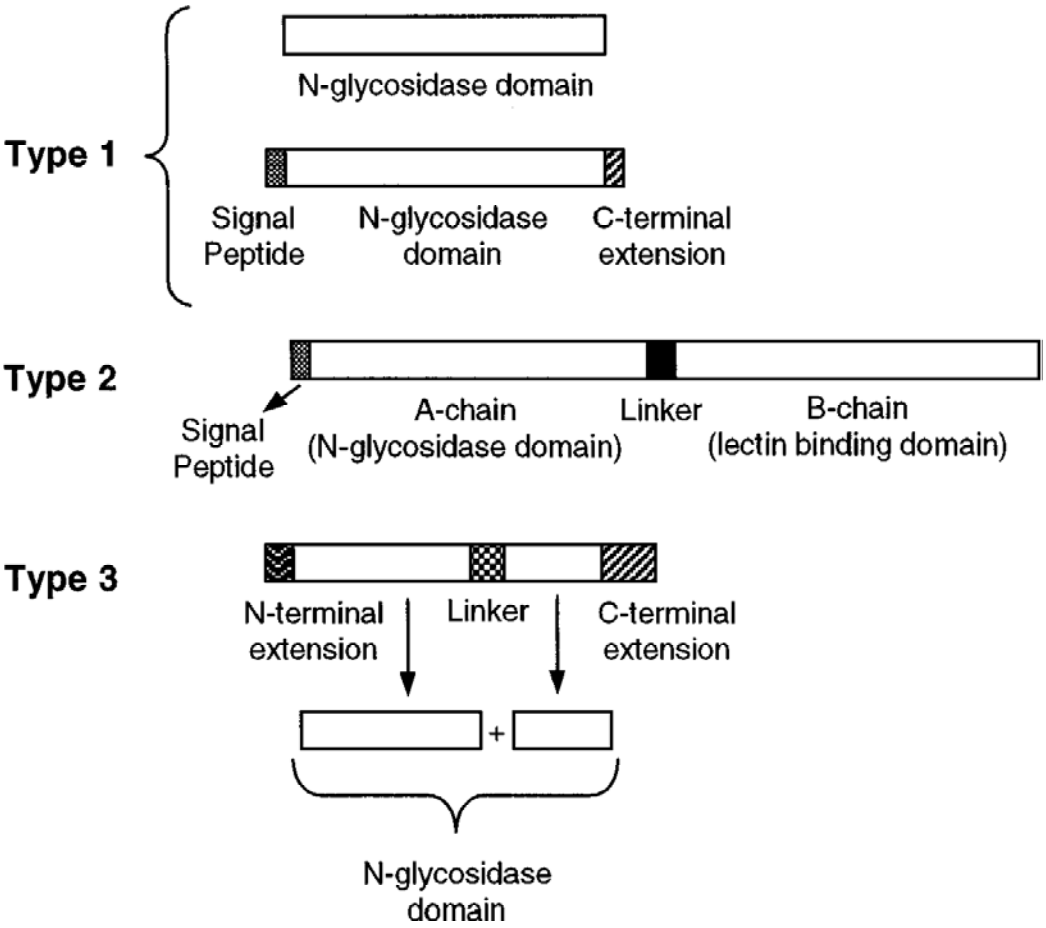


Fig. 3.1 Schematic representation of the molecular structure and the alignment of different RIPs showing a comparison of their primary structures. Filled, stippled, and hatched boxes denote regions absent in the active enzymes (shown as blank boxes) of different types of RIPs. Type 1 RIPs refer to mature, catalytically active PAP , trichosanthin etc., The type II RIPs refers to mature Ricin, Abrin-b etc., whereas that of type III RIPs refers to intact JIP60 from barley (*Hordeum vulgare*). (taken from Nielsen & Boston, 2001)

RIPs can be used as tools for identifying and targeting cell surface markers, and can be coupled to antibodies so that they can be delivered directly to specific cells. The type-II RIPs have been very useful for studies of endocytosis and intracellular transport in mammalian cells, as antitumor and antiviral agents in biomedical and agricultural research.

The cucurbitaceae family contains many type-I RIPs, but only very few type-II RIPs have been reported. They include *T.kirilowii* lectin-1(**Li *et al.*, 2000**) and *T.anguina* lectin (Manoj *et al.*, 2001) from the genus Trichosanthes, and Foetidissimin II from *Cucurbita foetidissima* (Zhang & Halaweish, 2007). These lectins have been biochemically characterized and they have been included in the type-II RIP family on the basis that they possess the β-prism-I fold, their secondary structural elements, sugar specificity (Gal/GalNAc), spatial orientation of the chains etc.

The lectin from *T.dioica* has been characterized biochemically and it has also been placed in the type-II RIP family on the basis of similar characteristics (Sultan *et al.*, 2004). In order to shed more light on structure-function relationship in this family of lectins the structural investigation of this *T.dioica* lectin was carried out.

3.3 Cucurbitaceae lectins & Type-II RIPs

Cucurbitaceae plants have been known to contain a wide variety of lectins and lectins like protein especially the type-I and type-II RIPs. Lectins from cucurbitaceae species have been purified, biochemically characterized and some of them even structurally well studied.

RIPs from other cucurbitaceae plants include beta-momorcharins (isolated from seeds of *Momordica charantia*) (Yuan *et al.*, 1999), momorchochin (isolated from tubers of *Momordica cochinchinensis*), luffaculin (isolated from seeds of *Luffa acutangula*) and luffin-a and luffin-b (isolated from seeds of *Luffa cylindrica*) (Ng *et al.*, 1992), trichoanguin from *Trichosanthes anguina* (Chow *et al.*, 1999), bryodin *(Bryonia dioica*) (Siegall *et al.*, 1994). They possess similar amino acid compositions and N-terminal amino acid sequences.

Lectins whose structures have been investigated in detail include those from various genus such as *Trichosanthes* (Falasca *et al.*, 1989; Anuradha & Bhide, 1999), *Momordica* (Ng *et al.*, 1986; Yuan iet al., 1999), *Bryonia* (Siegall *et al.*, 1994), *Coccinia* (Sanadi & Surolia, 1994), *Luffa* (Wang & Ng, 2002; Li *et al.*, 2003), *Cucurbita* (Allen, 1979; Read & Northcote, 2003; Hou *et al.,* 2008) etc. From the tuber *T.kirilowii* several type I & II RIPs have been structurally characterized and their active sites residues described in detail (Zhang & Wang, 1986; Li *et al.*, 2000; Kondo *et al.*, 2004). A few lectins have been identified from this plant and the structure of *T.kirilowii* Lectin-1 (TKL-1) has been determined at 2.7Å showing homology to type-II RIPs (Li *et al.,* 2001). The lectin from the seeds of snake gourd (*Trichosanthes anguina*) has been crystallized in the space group P622, and its structure has been determined by molecular replacement using Abrin-a as a model (1ABR). Molecular-replacement data of this lectin also shows its similarity to type II ribosome-inactivating proteins (Manoj *et al.,* 2001).

Since very few cucurbitaceae plants have been identified having type-II RIP, the identification of new lectin from these cucurbit plants will unravel the evolutionary relationship of this important family of plants. The lectins from the tubers of *T.kirilowii*, and those from the seeds of *T.anguina* and *T.dioica* have been well characterized with respect to its biochemical and biophysical properties.

3.4 Crystallization of the TDSL by Hanging-drop vapour-diffusion method

Initial crystallization trials were carried out using crystal screens obtained from Hampton Research (USA). This resulted in forked crystals growing in condition 22 of screen I, whose composition was 0.2 M sodium acetate trihydrate, 0.1 M Tris-HCl, pH 8.5 and 30% (w/v) PEG 4000. The crystal quality was improved by varying the choice and concentrations of salt and PEG. Good crystals were obtained by employing hanging-drop vapor-diffusion method in which

1 µl of protein mixed with 1 µl of well solution, comprising of 0.2 M ammonium acetate, 0.1 M Tris-HCl, pH 8.5, 25% PEG-2K-MME, was equilibrated against 1 ml of well solution. The crystals appeared elongated and forked at the edges as shown in Fig. 3.2. The well solution also had 60 µl of 0.5 % n-octyl β-D-glucopyranoside as additive. All crystallization experiments were carried out at 295 K.

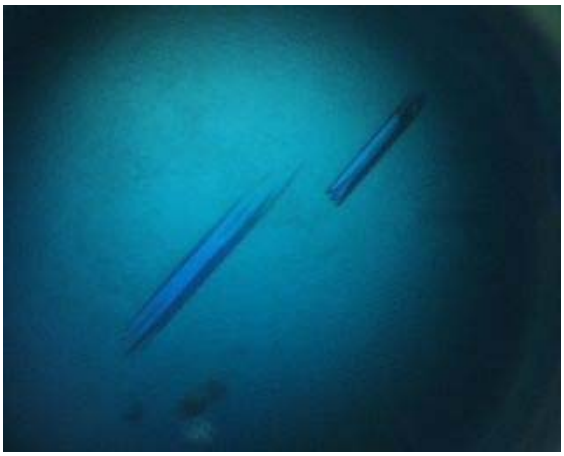


Fig.3.2: Elongated and forked crystals of *T.dioica* lectin

3.5 X-Ray diffraction, Data Collection & Processing

The data for *T.dioica* lectin crystal were collected on a Rigaku R-axis IV $^{++}$ image plate detector as described in chapter2. A single crystal was mounted in a thin walled glass capillary of 1mm diameter at room temperature (295K). X-ray diffraction data from the crystal were collected on an R-AXIS IV^{++} image plate using CuKα radiation generated by Rigaku rotating-anode X-ray generator operated at 50 KV voltages and 100 mA current. X-rays were focused using confocal mirror system. The crystal was kept frozen at 113K in liquid nitrogen cryostream, using 30% glycerol in crystallization solution as cryoprotectant, during the data collection. The crystal-to-detector distance was kept at 200 mm and an oscillation of 0.5º per frame was used.

The programs DENZO and SCALEPACK (Otwinowski, 1993; Otwinowski *et al.,* 1997) were used for processing and scaling the data. Data was processed in space group P6 using Denzo and the R-merge was 10.9 (38.4%). The unit cell dimensions were $a = b = 167.54$, $c = 77.42$ Å. The intensity data statistics are listed in Table (3.1)

Table 3.1. Summary of crystal parameters and diffraction data statistics *of T.dioica* lectin crystal

Temperature	**295 K**
X-ray source	**Rotating anode (Cu-K*α*)**
Wavelength	**1.5418 Å**
Resolution limits	**40.0-2.80 Å**
No. of observations (total no. of reflections used)	**148229**
No. of unique reflections	**30702**
Completeness of data	**99.8 (99.5)**
R_{merge}	**10.9 (38.4%)**
Average $I/\sigma(I)$	**9.9(3.7)**
Mosaicity	**0.6**
Space group	**$P6_4$**
Unit Cell Volume	**1882020**
Unit Cell Parameters	**$a = b = 167.54$, $c = 77.42$ Å.**
Matthews Coefficient	**2.85 $Å^3$ Da^{-1}**
Solvent content	**57 %**

3. 6 Alignment of the N-terminal sequence with the sequence of lectins belonging to this family

Among the cucurbitaceae family lectins those that have been structurally characterized belong to the type- I and type–II RIP family. They include several type-I RIPs such as β-Momorcharin (PDB: 1CF5) (Yuan *et al.,* 1999), trichosanthin (1TCS) (Zhang & Wang, 1986) etc., and type-II RIPs from *Trichosanthes kirilowii* lectin-1 (TKL-1) (PDB code: 1GGP) (Li *et al.,* 2000) and cucurmosin (PDB code: 1BWH) (Hou *et al.*, 2008).

From *Trichosanthes kirilowii*, three lectins have been reported out of which biochemical and structural data are available for only one lectin *i.e. Trichosanthes kirilowii* lectin-1 (TKL-1). TKL-1 has been shown to be similar to type-II RIPs as it possesses neutralizing epitopes like type-II RIPs, while it is incapable of inhibiting protein synthesis like RIPs. The lectin TDSL we have studied resembles TKL-1 in subunit composition and size of the chains, and hence it has been identified as a type-II RIP. The structure of TKL-1 has been determined at 2.7 Å resolution (Li *et al.*, 2000) by Molecular replacement method using coordinates of known structures (abrin, ricin and trichosanthin) (Li *et al.*, 2001).

The amino acid sequence of TKL-1 was queried in BLAST (Altschul *et al.*, 1997) at the NCBI site (http://www.ncbi.nlm.nih.gov/BLAST/). BlastP was performed against the PDB database and the results having significant e-values have been tabulated in Table 3.2.

Table 3.2 Top sequences from the BlastP search are tabulated. The TKL-1 sequence was searched against the PDB protein database.

Sr. No.	PDB file name	Description	Z-score (bits)	E-value	Identities (%)
1	1ABR	Crystal Structure Of Abrin-a (Tahirov *et al.*, 1995)	273	8e-74	58
2	2AAI	Crystallographic Refinement Of Ricin to 2.5 Å (Rutenber & Robertus, 1991)	231	5e-61	50
3	2Q3N	Agglutinin From *Abrus precatorius* (Apa-I) (Bagaria *et al.*, 2006)	228	4e-60	50
4	1RZO	Agglutinin From *Ricinus communis* With Galactoaza (Gabdoulkhakov *et al.*, 2003)	213	1e-55	47
5	1ONK	Mistletoe Lectin I From *Viscum album* (Krauspenhaar *et al.*, 1999)	191	74-49	42
6	2RG9	Crystal Structure Of *Viscum album* Mistletoe Lectin I In Native State At 1.95 Å Resolution (Krauspenhaar *et al.*, 1999)	179	2e-45	40
7	1PUM	Mistletoe Lectin I In Complex With Galactose (Krauspenhaar *et al.*, 1999)	179	2e-45	41
8	1HWM	Ebulin, Orthorhombic Crystal Form Model (Pascal *et al.*, 2001)	174	8e-44	35
9	3C9Z	*Sambucus nigra* Agglutinin Ii (Sna-Ii) (Maveyraud *et al.*, 2009)	161	4e-40	34

3.7 Structure Solution by Molecular Replacement

The reflections of TKL-1 have been phased using the coordinates of Ricin (2AAI), Abrin-a (1ABR) and Trichosanthin (1TCS). The spatial arrangement of TKL-1 molecule resembles that of RIPs from ricin and abrin-a. When *Trichosanthes kirilowii* lectin was used as a search model as it gave the best values for correlation and R_{factor} as compared to structurally

known type-II RIPs as models. The model chosen for molecular replacement solution of TDSL was the TKL-1 structure (1GGP) that contained one molecule in the asymmetric unit. The *AMoRe* program package (Navaza, 1994) implemented in the CCP4 suite (Collaborative Computational Project, Number 4, 1994) was used to solve the structures by MR method. The data in the resolution range 40-2.8 Å was used for determining the structure by molecular replacement.

Rotation and translation functions were calculated followed by a rigid body fit. Details of the molecular replacement solutions are given in Table 3.3. The position of the molecule in the unit cell of interest is defined in terms of three rotational and three translational parameters. *AMoRe* program from the CCP4 suit of programs has a fast rigid-body refinement procedure, which is very efficient in the evaluation of the correctness of the molecular replacement solution. It was first developed by Huber & Schneider (1985**)** and involves minimization of the misfit, where rotations and translations are calculated by interpolation and phase shifts and is not a least-squares rigid-body refinement of coordinates. The rotation function calculations (using the program 'roting' of *AMoRe*) resulted in small number of peaks with nearly similar correlation coefficient (Cc) and R-factors. The significant solutions generated by the program indicated the orientations of dimers in the asymmetric unit. Translational search calculations confirmed the dimer in the asymmetric unit. The program 'traing' of *AMoRe* was used for the translation function calculations that gave correct solutions.

The final solution selected was the one having high value of correlation coefficient (22.2) and low R_{factor} (53.3). The correct solution was then use as an input for in the 'fitting' program of *AMoRe* and a least-squares refinement was carried out. The parameter Cc further increased in this cycle to 38.1 and the value of low R_{factor} decreased further to 49.8 ndicating a correct

solution. Table 3.3 & 3.4 lists the MR solutions with final orthogonal rotations (rx, ry, rz) and translations (tx, ty, tz), which were applied to the MR model to generate the initial model of TDSL from the coordinates of the model.

Table 3.3 Molecular Replacement solution

Model: 1GGP.pdb (Protein Data Bank)

ROTATION FUNCTION

(Resolution Range 20 – 2.8 Å)

Solution	α	β	γ	Cc	R_{factor}
S1	44.62	87.48	146.27	10.5	0.0
S2	21.00	136.87	316.50	7.5	0.0
S3	23.05	77.64	88.79	6.0	0.0
S4	19.56	162.54	172.49	5.7	0.0
S5	50.88	74.61	172.47	5.7	0.0
S6	47.53	86.50	125.68	5.3	0.0
S7	23.67	146.62	339.00	5.3	0.0

TRANSLATION FUNCTION (for finding Space group)

(Resolution Range 40 - 2.8 Å)

Solution	Tx	Ty	Tz	Cc	R_{factor}

TF1	0.7478	0.5000	0.000	37.1	52.1
TF2	0.7165	0.7723	0.000	34.1	53.4
TF3	0.0134	0.8661	0.000	28.9	54.9
TF4	0.5469	0.0804	0.000	28.1	55.1
TF5	0.0513	0.7478	0.000	28.9	54.7
TF6	0.0067	0.9464	0.000	28.2	55.3
TF7	0.0290	0.0848	0.000	28.1	54.7

RIGID BODY FIT

(Resolution Range 20 - 2.8 Å)

Fixing

Solution	**α**	**β**	**γ**	**Tx**	**Ty**	**Tz**	**Cc**	**R_{factor}**
Sol-1	44.62	87.48	146.27	.7478	0.5000	0.000	37.1	52.1
Sol-2	21.00	136.87	316.50	0.7163	0.7710	0.000	45.1	49.7

Fitting

Solution	**α**	**β**	**γ**	**Tx**	**Ty**	**Tz**	**Cc**	**R_{factor}**
Sol-1	45.3	87.44	146.85	0.7491	0.5007	-0.0002	48.6	48.5
Sol-2	22.03	135.53	317.97	0.7153	0.7702	-0.0007	48.6	48.5

Table 3.4 Calculation of Translation function for obtaining the correct space group. The solutions with the correct space group input have comparatively higher correlation factor and lower R factor, highlighted in bold.

Input space group	**Rotation function Solution**	**Correlation factor**	**R factor (%)**
P6	1	30.0	54.7
	2	29.8	54.7

$P6_1$	1	32.9	53.5
	2	31.2	54.1
$P6_2$	1	30.0	54.5
	2	29.6	54.7
$P6_3$	1	30.3	54.4
	2	29.5	54.8
$P6_4$	**1**	**37.2**	**52.1**
	2	**34.5**	**53.2**
$P6_5$	1	30.2	54.5
	2	29.5	54.6

3.8 Problems faced due to non availability of the lectin sequence and alignment with lectins from the same genus with type II RIPs

The lectin from *T.dioica* seeds has been well characterized but its protein sequence could not be determined. In the absence of the sequence data, the molecular replacement was tried with lectin structures from related plants such as *T.kirilowii* from the same genus. The structure of a type-II RIP from *Abrus precatorius* (Abrin-a) has been determined at 2.14 Å (Tahirov *et al.*, 1995). The sequence of sequence of TDSL was aligned with that of TKL-1(Li *et al.*, 2003), abrin-a and Ricin which showed considerable sequence homology between chains in all these lectins (Fig. 3.3)

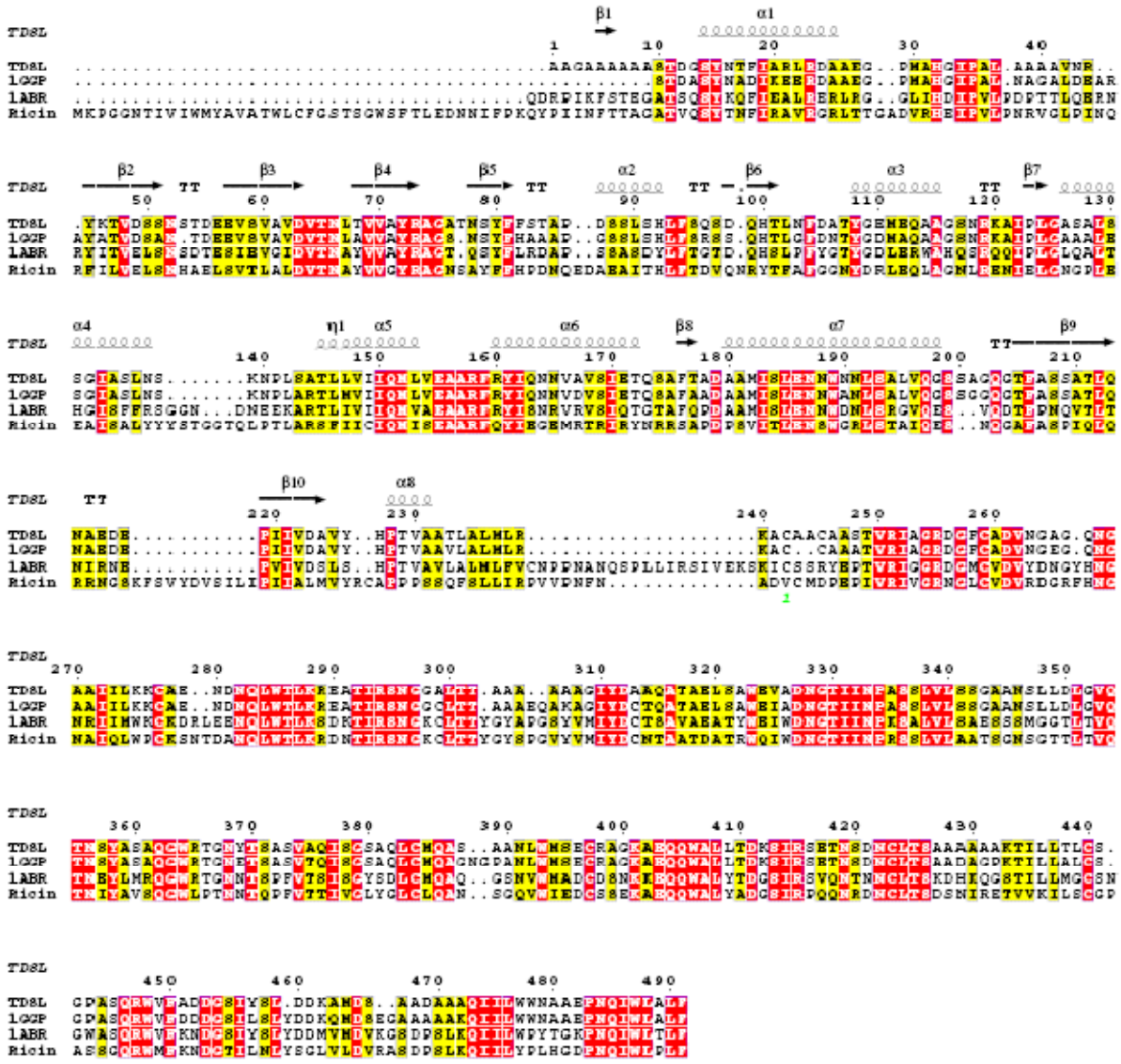


Fig. 3.3 Multiple sequence alignment of A and B chains, of TDSL with type-II RIPs from *T.kirilowii, A.precatorius* and *R.communis* using EsPript (Gasteiger *et al.*, 2003). The sequence titles are: TDSL; 1GGP (TKL-1); 1ABR (Abrin-a); Ricin

3.9 Structure Refinement

The structure of TDSL was solved using molecular replacement method as described in chapter 2 and the coordinates of TKL-1 (1GGP) was used as the search model in MR.

Refinement can be defined as an iterative process of improving agreement between the molecular model and the diffraction data. During refinement certain parameters needs to be considered in detail such as the least square adjustments of the atomic positions in the model, the occupancies of the atoms, temperature factors of individual atoms, so that there is a good agreement between the calculated and observed structure factors. Refinement is thus the minimization of these factors so that finally decrease in the values of R_{factor} and R_{free} are obtained that are indicative of a good progress of refinement. The final model should have an acceptable geometry as well. Various programs from CCP4 suite have been used for structure solution, refinement and analysis of the geometry of the final model. Refmac5 program of the CCP4 suite has been used for structural refinement, rigid body, TLS, restrained or unrestrained refinement against x-ray data, or idealization of a macromolecular structure. Refmac5 can refine an atomic model by adjusting these model parameters such as the structure coordinates, B-factors, TLS etc in order to obtain the model that best explains the experimental data (i.e. maximizes the likelihood). Certain constraints and restrains need to be introduced during structural refinement. The software Quanta (Accelrys) was used for model building and Coot (Emsley & Cowtan, 2004) was also used. In each refinement cycle of Refmac5, maps were generated. After each cycle, the model was visually inspected using the program *QUANTA* against the ($2F$o – Fc) and (Fo – Fc) difference maps.

3.9.1. Model Building and refinement

The TKL-1 is a two chain protein; and the crystal structure contains monomer in the asymmetric unit. While in TDSL structure, there are two molecules in the asymmetric unit, each having two chains. Hence the four chains corresponding to two hetero-dimers in the asymmetric unit are labeled as A, B, C and D.

As described in the above paragraphs, the initial coordinates of the molecule were obtained from 1GGP of TKL-1. Initially the values of R_{factor} and R_{free} obtained were 0.38 and 0.49. After a few refinement cycles using the non crystallographic restrains the values only slightly decreased.

It was observed that there was no electron density for residues 1-12 of the A and the C chain, and also for 1-11 of the B and the D-chains. In the remaining part of all the chains it was observed that there were many chain breaks and some amino acid residues found to be missing.

Refinement of the A and B chains was carried out using the electron density maps for modeling the protein. The sequences of TKL-1, abrin-a, ricin and other type-II RIPs were aligned using the program ClustalW(1.83). It was used for modeling our structure, especially the regions of the structure where there was ambiguity in the amino acid side chains and the loop regions where the electron densities were poorly defined.

Since the A and C chains, as well as the B and D chains are identical, here the modeling and refinement for only A and B chains is explained. Initially, when non-crystallographic restraints were used, model fitting was carried out only for the A and B chains while the C and D chains were generated from the fitted subunit by superposition (program SUPERPOSE) and by generating the coordinates using the program PDBSET, from the CCP4 suite of programs.

3.9.2 Refinement of the A chain: The RIP domain

Refinement of the A-chain was done using the (2*F*o – *F*c) and (*F*o – *F*c) difference maps. For some portions of the A-chain there was no electron density. The loop encompassing the residues corresponding to residues 41-50 of abrin-a were not located as they were not present in TKL-1 structure. But after a few refinement cycles, some deletions and repetitive cycles of

refinement and model building, electron density developed in this loop region because of improved phasing and thus amino acid residues were built. Still residue 47 was located after many more cycles when the density for it developed subsequently.

The N-terminal of the A-chain residues 1-11 was not present in the model (TKL-1); hence these residues needed to be built. A1-12 residues of the abrin-a (PDB code: 1ABR) were superposed on our model, and it was used for modeling the residues in this region along with reference electron density. With better phasing and better developed electron density in the (*2Fo – Fc*) map, 3-11 residues were successfully build initially as Ala and added side chains based on difference density and also referring to the multiple sequence alignment of the TKL-1 and other RIPs.

The residues corresponding to the loop 147-151 were also missing and were built when the electron density developed after cycles of refinement and model fitting. The other positions where break in the A chain were 81-82; 178-179; 209-210; 212-214; 238-239. The residue 213 had no density. At all these locations, the peptide bond was built when electron density developed using modules in QUANTA and geometry was corrected after every model building.

During this model building of the A-chain the sequence alignment of the TKL-1 and RIPs was considered for deciding amino-acid side-chains wherever extra electron density was observed or the side chain was outside the electron density.

3.9.3 Refinement of the B chain: Gal/GalNAc binding lectin domain

Structurally and with respect to its geometry the B-chain had more chain breaks and missing residues as compared to the RIP domain (A-chain). The N-terminal 1-11 residues of the B-chain could not be traced in our model just as in TKL-1 structure. A break in the main chain

was observed in the loop between residue 45 and 48 since no electron density (ED) was observed for these residues.

The residues corresponding to 72-77 of abrin-a forming a loop were found to be out of the electron density. After a few refinement cycles, and making zero occupancy for the residues, phases improved and these residues could be fitted in the density according to the (2*F*o – *F*c) and (*F*o – *F*c) maps. At the position of residues 160-162, that forms part of a β-turn the initial chain break was built based on the difference maps.

There was break in the chain between residue 232-233, as no density for these residues could be detected. After making zero occupancy and at the end of several refinement cycles the residue 232 could be built, but not 233. Other locations in the B-chain where the chain was broken were also built on the basis of the (2*F*o – *F*c) and (*F*o – *F*c) maps when phases improved. These portions of the B-chain were mainly located in the loop regions whereas the β-sheet regions had residues fitting well into the electron density. The positions in B-chain for which there was no electron density or there was a break in the chain are: 108-109; 131-132; 153-154; 155-156; 159-160; 163-164; 198-199; 203-204; 204-205; 209-210; 225-226; 234-235; 238-239; 245-246; 246-247.

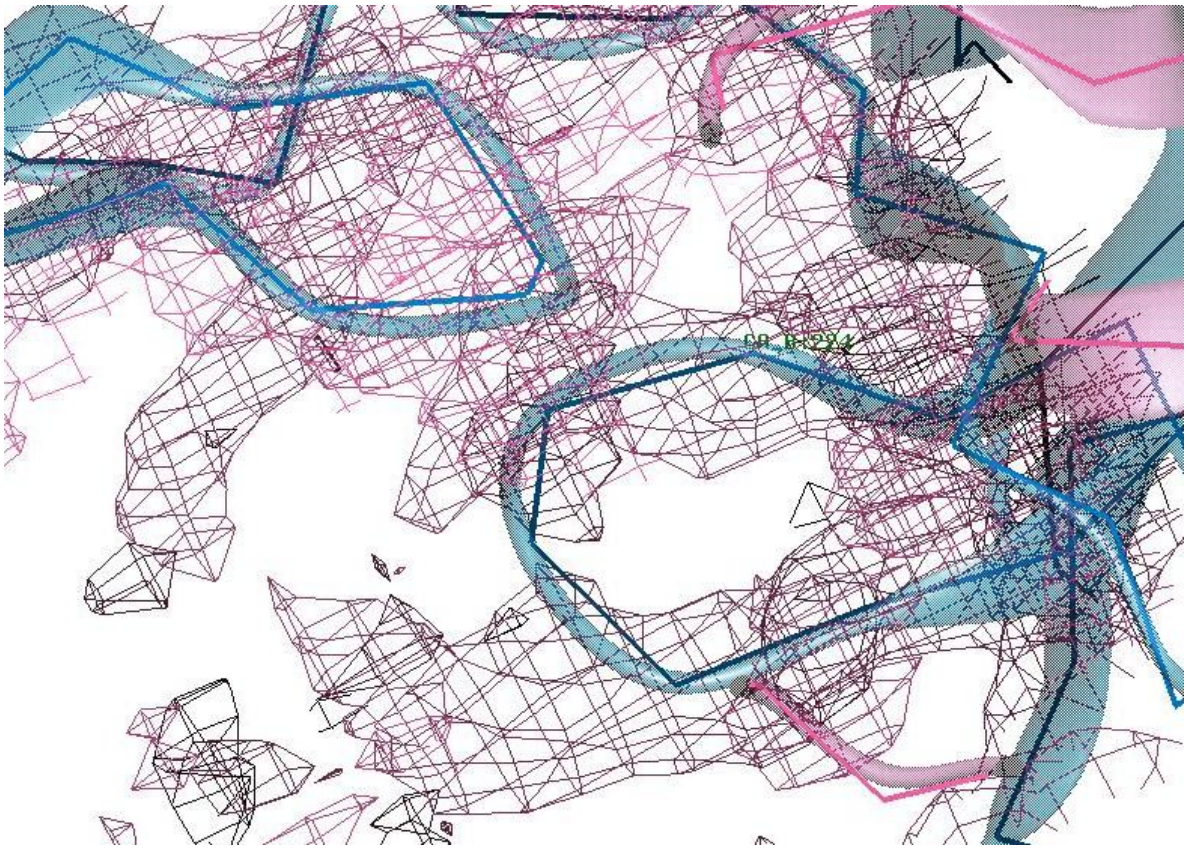


Fig. 3.4 Break in electron density is displayed in the B-chain loop 224-226. Also seen are the (*Fo-Fc)* and (*2Fo-Fc*) maps.

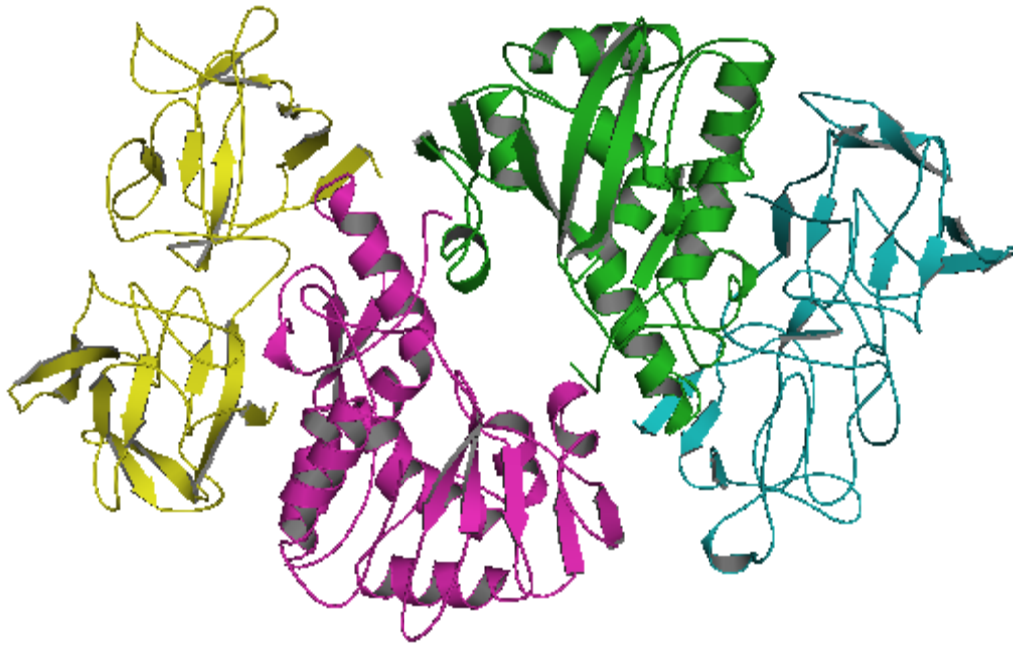


Fig.3.5 The structure of TDSL showing the four subunits A,B,C and D in Yellow, magenta, cyan and green color respectively.

3.9.4 Geometry analysis and subsequent methods to improve the geometry of the refined molecule

As the missing residues in the loops were built the geometry of other regions corresponding to β-turns and loops also improved. But even after repeated refinement cycles and model building in especially the loop region, although the value of R_{factor} decreased (0.29), the decrease in R_{free} was slow and finally remained at 0.38.

3.9.4.1 Omit Maps and assessing the correctness of the structure

In order to check whether we were proceeding in the correct direction for our model, the strategy we followed was to generate omit map for every 50 residues in both the chains A and B. Thus for generating each such map 50 residues were deleted from the model. Subsequently its refinement was carried out to see if omitting any particular segment of 50 residues lowers the

value of R_{free} for our model. These (2*F*o – *F*c) and (*F*o – *F*c) maps were then placed on our original PDB file containing all the residues (Master PDB). The decrease in values of R_{factor} and R_{free} were then recorded in all the cases and analyzed.

It was seen that in all the cases no significant drop in the values of either R_{factor} (0.31-0.33) or R_{free} (0.42-0.44) was observed. These values obtained from the omit map refinement indicate that no particular portion of our model has portions that are contributing to bad phases. Overall the structure has electron density pointing in the correct chain direction, any serious flaw in any part of the model was thus ruled out.

3.9.4.2 Poly Alanine (Polyala) Model

A polyalanine model was created in Coot. Refinement of this model was carried out using Refmac5 program. The (2*F*o – *F*c) and (*F*o – *F*c) maps that were obtained were used for model building along with the PDB (original PDB having all residues). This model was created to see if the electron density for any of the side chains is different than we have built and whether any extra density for any of the loops is in a different direction. But it was observed from these difference maps of the polyalanine model that the side chains of amino-acids as placed and modelled by us seemed correct. Any extra density for any loop of main chain was not observed. The value of R_{factor} or R_{free} after refinement of the polyalanine model were 0.31 and 0.44 respectively.

3.10 Structural details

The structure consists of two dimers in the asymmetric unit. Each subunit is hetreodimeric, made up of two chains. The entire protein is thus a homodimer of heterodimers (AB and CD). The chain A or C is linked by a single disulphide linkage to chain B or D.

3.10.1 Structure of the A and C chains: The RIP-II domain

Structurally the type-II RIPs consist of the A-chain which possesses RNA N-glycosidase activity and is responsible for cell lysis and ultimately cell death. In the structure of lectin from *T.dioica,* the A-chain can be described as composed of three folding domains as in ricin, abrin and TKL-1 structure (Rutenber & Robertus, 1991; Tahirov *et al.*, 1995; Li *et al.*, 2003). In abrin-a, the A chain contains 251 amino acids, while in TKL-1 there are only 234 residues. In TDSL, we could identify ED for 246 residues respectively in both A and C chains. Since there was no electron density developed even after several refinement cycles, the residues corresponding to A1-3 (also in C-chain) could not be built. This A-chain is made up of α-helices and β-sheets and does not contain disulphide linkages. It can be divided into three structural domains:

(i) Domain 1 consists of six-stranded β-sheet and two α helices which are interconnected by several loops.

(ii) Domain 2 predominantly consists of five α-helical regions interconnected by loops of varying lengths. This domain is conserved among various type-II RIPs

(iii) Domain 3 is made up of two helices and two β-sheets and a random coil also.

3.10.2. Structure of the B and D chains: The Lectin domain

The B-chain of type-II RIPs is a lectin and is involved in cell recognition as it binds to cell surface receptors and binding to cell surface receptors facilitate transport across the cell membrane. The B-chain has binding site for either galactose or galactose terminated oligosaccahrides while it also acts as an immunotoxin. The B- and the D-chains contain three intra-chain disulphide linkages. In abrin-a and TKL-1 there are 267 and 254 residues

respectively, while in TDSL 257 residues could be traced. B1-10 residues could not be traced in the electron density as in TKL-1 structure. The domain wise distribution of the B-chain can be described as in abrin-a and TKL-1. It is predominantly made up of β-sheets and can be divided into two homologous domains. Each domain is composed of λ, α, β and γ sub-domains. The α, β and γ sub-domains form the main body of each domain while the 2λ domains form the connection between two such domains. A hydrophobic core is found to be located in each domain and it is similar to that found in Abrin-a structure.

According to the evolutionary theory proposed by Rutenber *et al*. (1991), the B-chain in ricin is formed from α sub-domain which is the smallest structural unit possessing all the necessary structural and functional features. Gene duplication and fusion of the α sub-domain results in the $\mathbf{(\alpha\beta\gamma\lambda)_2}$ structure as observed in the B-chain of these proteins.

3.10.3. Active site characterization: Binding to Adenine

RIPs recognize a highly conserved region in the large 28S rRNA and cleaves a specific N-C glycosidic bond between an adenine and the nucleotide on the RNA and eventually the adenine residue is removed (Barbieri *et al.*, 1993). The active site of RIPs is involved in the binding of the protein to ribosomes. These proteins thus possess a cleft for adenine at the active side.

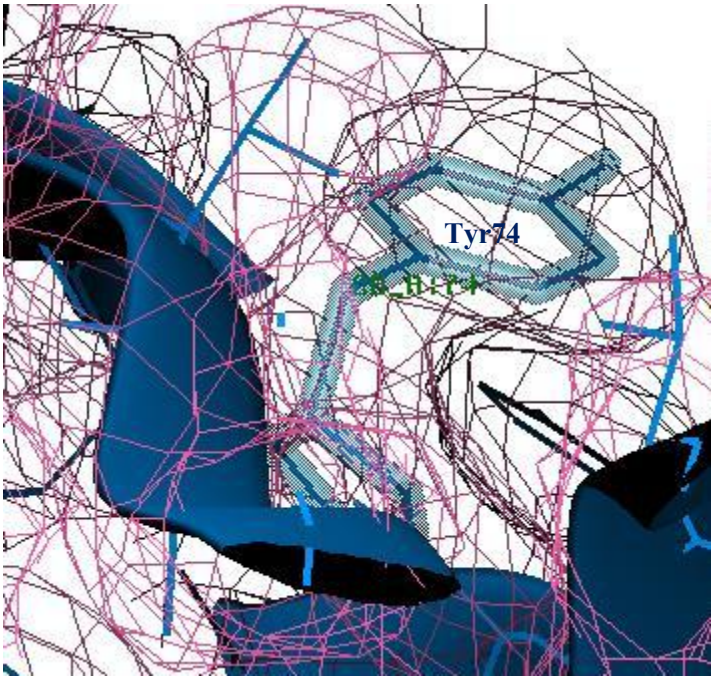


Fig 3.6 Showing the A-chain Tyr74 residue which is conserved in all RIPs and is seen in TDSL structure also.

In the RIP structures five invariant residues have been defined which are Tyr74, Tyr113, Glu164, Arg167 and Trp198. All these residues are found to be conserved in TDSL and significant one among these residues is **Tyr74** which is essential for the RIPs and is conserved among all reported RIPs. These residues are found to be invariant in many type-II RIP structures such as abrin-a and ricin (Katzin *et al.*, 1991; Funatsu *et. al*, 1991**).** Out of these conserved residues, Glu164 and Arg167 are the key residues for the catalytic processing of the N-C glycosidic bond. The residue Tyr74 is the main residue involved in binding to adenine and it is thus responsible for the ribosome inactivating activity. At the active site of RIPs, five conserved residues have been located and they are also conserved in TDSL. These conserved residues are: Asn72, Arg124, Gln160, Glu195 and Asn196. Thus TDSL resembles RIPs in having an aromatic

residue at the sugar binding cleft (Tyr74) and on this basis and also with respect to other structural similarities with type-II RIPs; it is placed in this family.

3.10.4. Disulphide bridges and stability of the lectin

A disulphide bond is found between the C-terminus of the A/C chains and the N-terminus of the B/D chains. Residue 244 in A and C is connected by a disulphide bridge to residue 11 in B and D and it is the only intra-chain disulphide bridge. A similar bond has been observed in TKL-1 structures, while in abrin-a it is found between residues A Cys248 and B Cys8. Out of four disulphide bonds that have been reported in abrin-a structure, only two of them could be defined in the electron density. These two disulphide bonds are in the B chain: B25-44 and B 189-206.

The stability of TDSL to chemical, pH and thermal denaturation can be explained on the basis of its disulphide linkages. These linkages confer high stability to the protein and maintain the protein structural integrity and active conformation under condition of high temperature (90 °C), extreme pH (pH 3.0 and pH 10) and up to 5M GdnHCl concentration.

3.11 Evolutionary relationship among the Type-II RIPs and TDSL

The proteins belonging to the RIP family share sequence and structural homologies and are shown to be phylogenetically related. The alignment of the primary sequence of TDSL with other type-II RIPs TKL-1, Abrin-a and Ricin along with the secondary structural elements of TDSL is presented in Fig 3.3. From the alignment it is quite evident that type-II RIPs are structurally well conserved and TDSL shows a good homology to these lectins representing the type-II RIP family.

Out of the various type-II RIPs compared above, a few of them do not show RIP activity while others have well defined RIP activity. The α-carbons of various RIPs are also super-imposable on each other. Hence it can be said that the tertiary structures of the different RIPs are

well conserved and these group of proteins are evolutionarily related. But these lectins differ significantly from each other with respect to their C-terminal domains, structure and composition of the various loops as well as few catalytic site residues. Thus the difference in activity of these RIPs towards various substrates as well as their difference in ligand binding can be explained on the basis of these structural differences.

3.12 Conclusion

With the available structural data on the galactose/*N*-acetyl galactosamine specific lectin from *Trichosanthes dioca* seeds, a few conclusions can be made. Firstly this structural study indicated that the lectin is well conserved with respect to the active site residues and other invariant residues with the type-II RIPs especially abrin-a and ricin, besides showing close homology with TKL-1, also from the same genus. Secondly the residue most essential for binding with the substrate adenine and responsible for the RIP activity (Tyr74 in abrin-a) has also been observed in TDSL. On this basis and other structural features TDSL can be classified as belonging to this lectin family. The lectin content of *T.dioica* fruits might not be toxic as other RIPs as it is widely consumed as a vegetable in some parts of the globe.

A therapeutic application of TDSL can be for mediating the delivery of drugs in malignant cells. Monoclonal antibodies coupled with RIP form an efficient delivery system, which form highly selective conjugate for cancerous cells. Since some RIPs have been shown to be useful for such cancer therapy, a useful implication of TDSL can be made on similar lines for effective delivery of anti-cancerous drugs for human welfare.

Chapter 4

Biophysical and stability studies on two *Trichosanthes* lectins

4.1 Summary

In this chapter a comparative account of the biophysical characterization of lectins from *Trichosanthes dioica* and *Trichosanthes anguina* with respect to their thermal, chemical and pH stability has been described. The thermal denaturation studies of these two lectins indicate that both these cucurbitaceae lectins show high thermal stability, although the disulphide linkages do not play any role in the stability of these proteins. Some RIPs have been shown to be highly thermostable. Since the *T.dioica* lectin resembles RIPs in molecular weight, subunit composition, size, and sugar specificity, it may belong to this family. As it has been pointed out that some *Trichosanthes* lectins have similarity with RIPS; here we show that like some RIPs, these lectins are resistant to wide range of chemical and pH denaturing conditions.

4.2 Introduction

Cucurbitaceae lectins have been purified and characterized in some detail from plants *Momordica charantia* (bitter gourd) (Wang & Ng, 1998), *Trichosanthes anguina* (snake gourd) (Anuradha & Bhide, 1999), *Trichosanthes kirilowii* (Wang & Ng, 2000) and *Trichosanthes cucumerina* (Kenoth & Swamy, 2003). RIPs belong to one of the subclass of plant lectins, the β-trefoil lectin (Iglesias *et al.*, 1982). Of the RIPs studied so far, some crystal structures of type II RIPs that have been determined that include Ricin (Rutenber *et al.*, 1991), abrin-a (Tahirov *et al.*, 1995), mistletoe lectin I (Krauspenhaar *et al.*, 1999**;** Meyer *et al.*, 2008) and *Trichosanthes kirilowii* lectin-1(Li *et al.*, 2000). The tuber of *Trichosanthes kirilowii* contains 3 isolectins, but only the crystal structure of lectin-1 from *T.kirilowii* has been reported.

Among the large number of lectins studied, the relationship between structure, conformation and stability has been established only for a few proteins (Kenoth & Swamy, 2003**;** Sahasrabuddhe *et al.*, 2004**;** Sultan *et al.*, 2006; Katre *et al.*, 2007; Dharker *et al.*, 2008). Lectins

from *Trichosanthes dioica* and *T.anguina* from cucurbitaceae family have been studied in detail (Komath *et al.*, 1998; Komath *et al.*, 1999; Anuradha & Bhide, 1999; Manoj *et al.*, 2001; Sultan *et al.*, 2004; Sultan & Swamy, 2005; Kavitha *et al.*, 2009). Purification of *T.anguina* lectin was carried out on Sepharose 4B (Komath *et al.*, 1998**)**. The lectin from *Trichosanthes dioica* was also purified by a similar procedure. Both these lectins are galactose specific and are heterodimeric having two subunits linked by disulphide bonds. *T.dioica* has a MW of 55,000 Da with two non-identical subunits of 24,000 and 37,000 Da as reported (Sultan *et al.*, 2004). While *T.anguina* lectin is also hetreodimeric with MW 45,000 Da, and subunits of 37,000 and 22,000 Mr. Both the lectins are galactose specific, preferentially binding the β anomer of the galactose and are heterodimers of two unidentical subunits joined by disulphide bonds. *T. anguina* lectin has a putative histidine residue (Komath *et al*, 1998) whereas *T. dioica* lectin has a tyrosine at the sugar-binding site (Sultan *et al.*, 2004). The secondary structure of the lectin estimated using CD analysis comprises of α-helix, β-sheet, β-turns and random coil.

In the present chapter, we report on the high thermostable nature of the *Trichosanthes* lectins and also the effect of chemical denaturants and pH on the relative stability of these lectins.

4.3 *Trichosanthes* lectins

Plants belonging to cucurbitaceae family are a rich source of Galactose/*N*-acetyl galactosamine specific lectins which are predominantly found in the mature seeds and sometimes in the root tubers of these plants. The lectins from some of these cucurbitaceae plants have been purified and characterized with respect to their biochemical properties and sugar specificity as well as their molecular and structural properties has been investigated. They include *Cucurbita pepo* fruit lectin (Allen, 1979**),** *Momordica charantia* (bitter gourd) seed lectin (Sultan &

Swamy, 2005; Kavitha *et al.*, 2009), *Trichosanthes kirilowii* seed and tuber lectins (Yeung *et al.*, 1986; Falasca *et al.*, 1989; Li *et al.*, 2000), *Trichosanthes dioica* seed lectin (Sultan *et al.*, 2004; Sultan & Swamy, 2005; Kavitha *et al.*, 2009), *Trichosanthes anguina* seed lectin (Komath *et al.*, 1998; Komath & Swamy, 1999; Anuradha & Bhide, 1999; Komath *et al.*, 2001; Manoj *et al.*, 2001) and *Trichosanthes cucumerina* seed lectin (Kenoth *et al.*, 2001; Kenoth & Swamy, 2003).

Besides containing lectins, these cucurbit plants have also been shown to contain various type I and type II RIPs. RIPs are proteins having dupurinating activity on the large 28S ribosomal RNA, thus inhibiting the translational activity and resulting in cell death. RIPs are toxic proteins and hence have immense potential for development of immunotoxins for cancer and HIV. The type I RIPs are small monomeric proteins of 26-30 kDa and do not possess any hemagluttination activity. While the type II RIPs are dimeric proteins in which the A chain is like the type I RIPs, while the B-chain is a galactose/*N*-acetyl galactosamine specific lectin.

Momordica charantia contains both a lectin and a type I RIP (Wang & Ng, 1998). The tubers of *Trichosanthes kirilowii* contain many type I RIPs out of which Trichosanthin has been very well studied biochemically and structurally (Zhang & Wang, 1986; Gao *et al.* 1994; Fei *et al.*, 2004), while the lectin from the same plant, *Trichosanthes kirilowii* lectin-1 has been shown to be homologous to type II RIPs (Li *et al.*, 2000).

Lectins have been characterized from the plants belonging to the genus *Trichosanthes* such as *T.kirilowii* (seeds, TKL's), (Falasca *et al.*, 1989) (root tubers, isloectins TK-I, II and III) (Yeung *et al.*, 1986), *T.anguina* (SGSL) (Komath *et al.*, 1998; Anuradha & Bhide, 1999), *T.dioica* (TDSL) (Sultan *et al.*, 2004), and *T.cucumerina* (TCSL) (Kenoth & Swamy, 2003). All

these lectins are Galactose-*N*-acetylgalactosamine specific and heterodimeric having two non-identical subunits with MW in the range of 50-56 kDa.

The information obtained from the three-dimensional structure of *T.kirilowii* lectin that has been solved at low resolution by molecular replacement using coordinates of ricin, abrin-a and trichosanthin (Li *et al.*, 2000), and those from double immune diffusion experiments has been used to categorize this lectin as a type II RIP.

4.4 *Trichosanthes dioica* and *Trichosanthes anguina* lectins

4.4.1 Plant properties & Scientific Classification

The genus *Trichosanthes* is in Greek meaning "hair flower", which describes the fragrant and delicately fringed white corolla (petals).

Trichosanthes dioica is also known as the pointed gourd, parwal (Hindi), or potol (Assamese, Oriya or Bengali) is grown as a vegetable all over India. The plant is rich in protein and vitamin A; and also contains major nutrients and trace elements (Mg, K, Cu,S and Cl). The plant parts have certain important medicinal properties. The fruits, which are commonly consumed as vegetable, are easy to digest, possess diuretic property, and have antiulcerous effects and their decoction heals skin infections. The fruit extract have property of lowering blood sugar, are used as a remedy for spermatorrhoea and these can thus be exploited for treating cancer- like conditions. The plant root finds use as a hydragogue cathartic tonic and febrifuge, while the juice of the leaf is applied to patches of alopecia areata.

Trichosanthes anguina is also known as snake gourd, serpent gourd, padwal (Hindi), Khyar, Pudal. It is widely grown as a vegetable in India and in the Orient. The fruits are generally long, narrow and cylindrical giving appearance of a snake, hence the name snake

gourd. The mature fruit has a slightly bitter taste. The fruit of this plant has substantial amount of moisture and little protein, fiber, fat and carbohydrate, and contains the minerals Ca, Fe, P, as well as carotene, thiamine, riboflavin and niacin. The medicinal properties of the leaf include the treatment of jaundice and heart disorders. A decoction of the fruits is effective in bilious fevers, as a laxative; the leaf juice is used to induce vomiting, for relieving liver congestion and in remittent fevers. Roots and seeds are used to treat diarrhoea and syphilis; leaf juice is also used in the treatment of alopecia.

The scientific classification of both the plants is as follows:

Trichosanthes dioica/anguina	
Kingdom	Plantae
Division	Magnoliophyta
Class	Magnoliopsida
Order	Cucurbitales
Family	Cucurbitaceae
Genus	*Trichosanthes*
Species	*T. dioica/ T.anguina*

(a)

(b)

(c)

Fig. 4.1 The Fruits of (a) *Trichosanthes anguina* (snake gourd) (b) *Trichosanthes dioica* (pointed gourd) (c) The whole fruit, cross sectional view of the cut fruits and seeds of snake gourd.

4.4.2 Purification & Biochemical characteristics

Trichosanthes dioica seed lectin (TDSL) was purified by affinity chromatography on cross-linked guar gum, and it eluted as a single peak corresponding to MW 55kDa on Superose-12 in the presence of 0.1 M lactose. On SDS-PAGE in presence of β-mercaptoethanol it showed two non-identical bands of MW 24 and 37 kDa, while it gave a single band in its absence (Sultan *et al.*, 2004**)**. Two non identical subunits connected by disulphide bonds have been characterized in this lectin. It has glycoprotein content of 4.7%, and is strongly inhibited by β-d-galactopyranosides, and tyrosine residues have been shown to be important for carbohydrate binding and hemagluttination activity.

Trichosanthes anguina seed lectin (SGSL) has been purified on cross-linked guar gum (Komath *et al.*, 1996**)** and eluted with lactose**.** It showed MW of 62,000 in gel filtration, while in

SDS PAGE in the absence of β-mercaptoethanol it gave one band corresponding to 53 kDa, and in presence of βME it resulted in two bands of 32 and 23 kDa. An isolectin complex specific to galactose has also been reported from *T.anguina* seeds that have been purified by affinity chromatography on Sepharose 4B followed by high affinity matrix Lactamyl Seralose (Anuradha & Bhide, 1999). These isolectins are glycosylated as well as non-glycosylated and the SDS PAGE analysis showed MW ranging from 30,000 to 50,000 +/- 1300.

The two lectins under investigation here are TDSL and SGSL.

4.4.3 Basis for their Biophysical characterizations

Various studies have been carried out on different legume lectins using different biophysical techniques which demonstrate their stability under various denaturing conditions (Sultan *et al.*, 2006**)**. Some studies on lectins from trichosanthes species have also been reported such as those from *T.cucumerina* (Kenoth *et al.*, 2001; Kenoth & Swamy, 2003). The fluorescence measurements have been carried out at steady state and also using time resolved methods and the microenvironment of tryptophan residues are investigated which shed light on their characteristics and give a better structural insight.

TDSL has been categorized as type-II RIP on the basis of its biochemical and molecular properties. It has a tyrosine residue at the active site which is involved in carbohydrate binding at the active site as well as being important for agglutinating erythrocytes. The tryptophan residues are shown to be buried in the protein interior as they were modified only after 8 M urea denaturation (Sultan *et al.*, 2004**)**.

SGSL has a histidine residue at the active site (Komath *et al.*, 1998**).** For SGSL preliminary crystallization data is available and the molecular replacement studies indicate that it

is also homologous to type-II RIPs (Manoj *et al.*, 2001). Hence characterization of microenvironment of both these lectins from cucurbitaceae will highlight their structural features as they have been placed in the type-II RIP family.

4.5 Results & Discussion

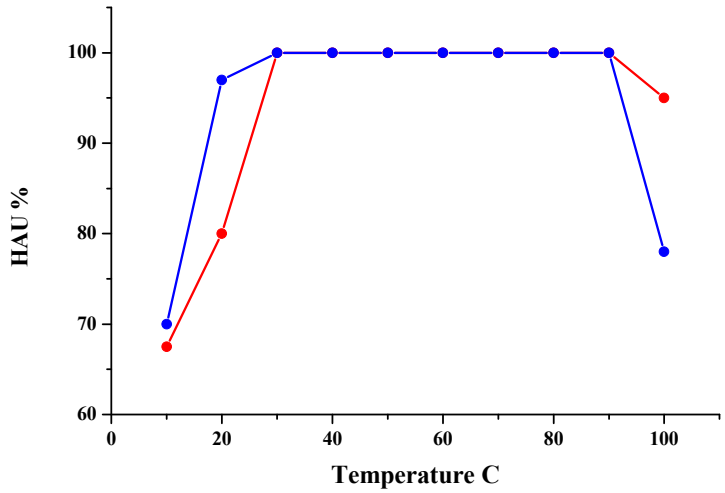
Denaturation as well as renaturation studies have been carried out on both these lectins after subjecting them to different conditions of temperature, pH and exposing to chemical denaturants such as Urea and Gdn-HCl and the spectrum has been analyzed by steady state fluorescence spectroscopy. The results are documented below.

4.5.1 Fluorimetric measurements

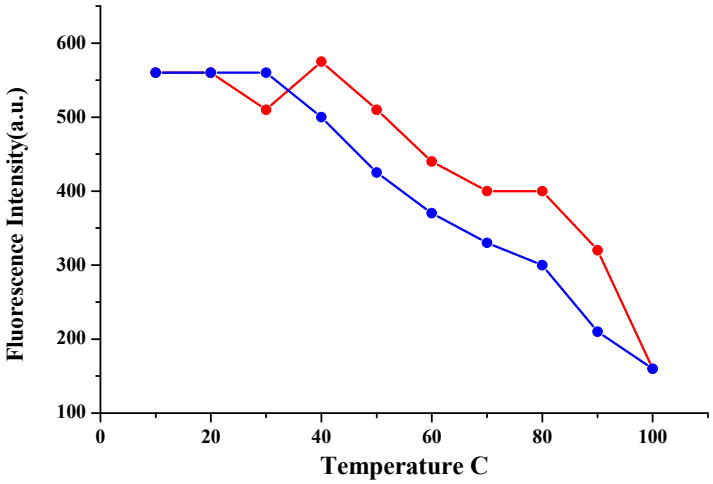
Fluorimetric measurements have been done on PerkinElmer LS-50B spectrophotometer as described under chapter 2: Materials and methods.

4.5.1.1 Spectra of native protein

The fluorescence emission spectra of the native lectins from *T. anguina* and *T. dioica* lectins showed emission maximum at 336 nm which characterizes non polar environment of Trp residues. This indicates that tryptophan residues in both these lectins are buried in the protein interior.



(A)



(B)

Figure 4.2. Effect of Temperature

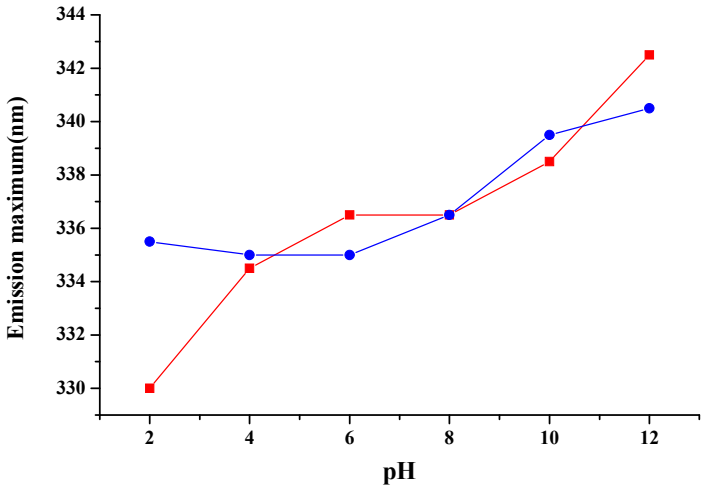
(A) Effect of thermal denaturation on the activity of both the lectins.

(B) The fluorescence intensity as a function of temperature.

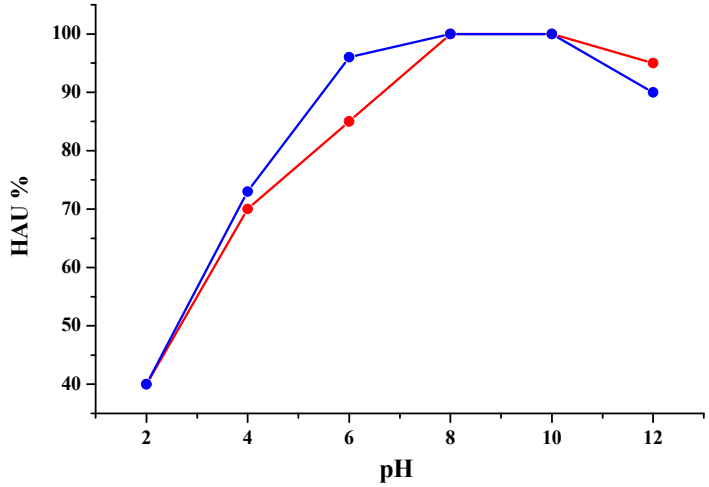
Red lines represent *T.anguina* lectin; blue represents *T.dioica* lectin.

4.5.1.2 Thermal stability

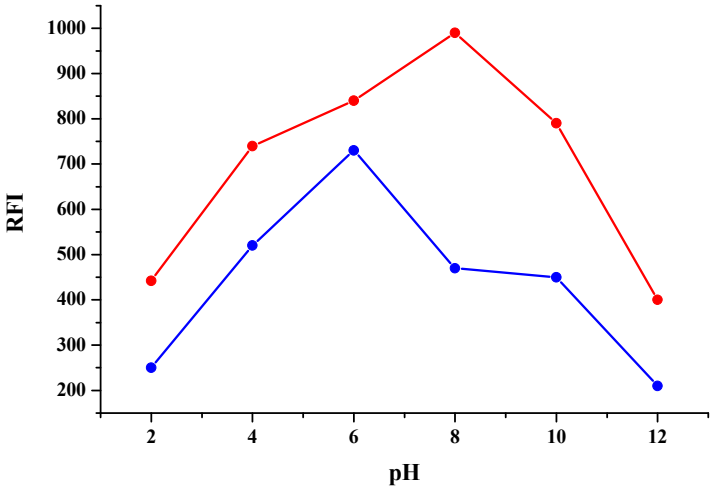
Both the lectins show hemagluttination activity upto 90°C and the fluorescence intensity gradually decreases with increase in temperature (Fig. 4.2 A and B). At 90°C the intensity is almost 25 % of the intensity at room temperature. This could be due to the deactivation of excited singlet state. ANS binding to any of the lectins was not observed at any of the temperatures, indicating no exposure of the hydrophobic patches on the surface of the proteins. On cooling the samples to 30°C the original fluorescence of the proteins is restored. No light scattering was observed in the above conditions indicating absence of aggregation of the proteins due to thermal denaturation. The hydrophobic interactions in the interior of the protein molecule holding the protein in its active structure seem to be strong.



(A)



(B)



(C)

Figure 4.3. Effect of pH

(A) The emission maximum of both the lectins under pH ranging from 2 to12.

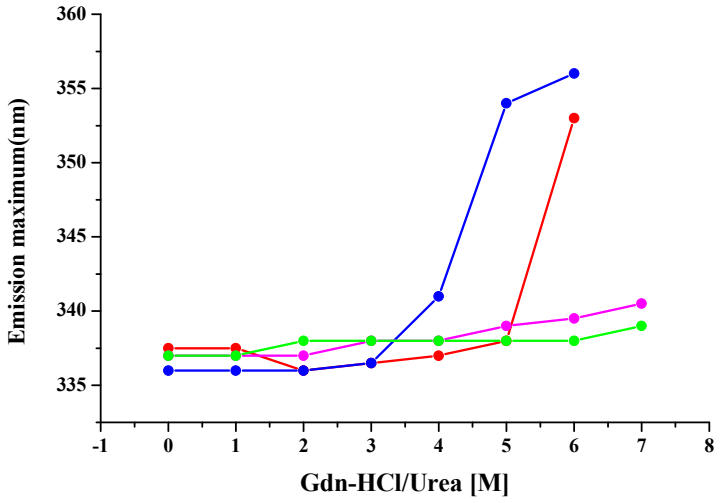
(B) Effect of denaturation at different pH on the hemagluttination activity of both the lectins.

(C) The fluorescence intensity curves at various pH.

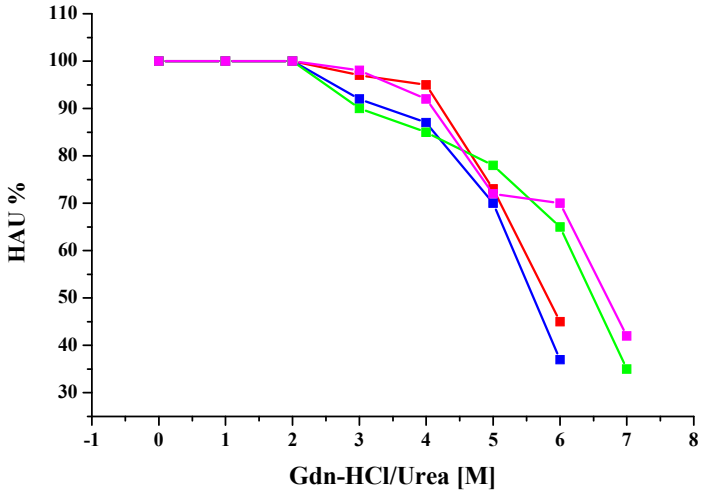
Red lines represent *T.anguina* lectin; blue represents *T.dioica* lectin.

4.5.1.3 pH Stability (Urea/ Gdn-HCl)

Both *T. anguina* and *T. dioica* lectins are stable in the pH range of 6-12. They are partially stable at pH 4, while losing 60% of the activity at pH 2 (Fig.4.3B). *T. anguina* lectin has λ_{max} 330 nm at pH 2 and it gradually shifts to 342 nm with increase in pH (Fig. 4.3A). The protein could be partially unfolded at alkaline pH, so tryptophans are exposed to the more polar environment as compared to that in the native condition of the protein. The fluorescence intensity of the lectins is much less in the highly acidic or highly alkaline pH than at closer to neutral pH range (Fig.4.3C). Thus, other than some change in the microenvironment of the Trp residue, there is no major structural transition in the protein molecule in a wide range of pH. The decrease in fluorescence intensity is due to protonated and deprotonated forms of the amino acids on the surface affecting the emission of the Trp residues.



(A)



3(B)

Figure 4.4.Urea/Gdn-HCL denaturation

(A) The emission maximum of both the lectins when subjected to 0-6M Gdn-HCl.

(B) Effect of Gdn-HCl denaturation on the hemagluttination activity of both the lectins.

Red & Magenta lines represent *T.anguina* lectin subjected to Gdn-HCl and urea denaturation respectively; blue and green color represents *T.dioica* lectin subjected to Gdn-HCl and urea denaturation respectively.

4.5.1.4 Stability against chemical denaturants (Urea/ Gdn-HCl)

Effect of Gdn-HCl:

On treatment with increasing concentrations of Gdn-HCl, both the lectins slowly start losing activity at 3M concentration of the denaturant (Fig. 4.4B). There is no shift in the wavelength corresponding to emission maximum (Fig. 4.4A). At 4M Gdn-HCl, λ_{max} of *T. dioica* lectin shifts to 340 nm and at 5M, λ_{max} of *T. anguina* lectin shifts to 342nm, indicating slight

increase in the polarity of the Trp environment. At 5M and 6M Gdn-HCl, *T. anguina* and *T. dioica* lectins, respectively, are unfolded showing red shift in the λ_{max} to 355 nm (Fig. 4.4A). At this stage, major activity loss is observed. Above 3M concentration of Gdn-HCl, even fluorescence intensity reduces substantially. Renaturation or refolding of the proteins is measured by considering the extent of reappearance of the original spectra and recovery of sugar binding activity. After dilution of the 4 and 5M Gdn-HCl treated lectins the activity was partially regained (with 10-15 % increase in the activity) and a blue shift in λ_{max} to 340- 342 nm was observed for both the lectins after 24h. Partial renaturation and reactivation after dilution of 3-5 M Gdn-HCl treated protein was observed in case of *Artocarpus hirsuta* lectin (Gaikwad *et al.*, 2002).

Effect of Urea:

As observed in case the of Gdn-HCl mediated denaturation both the lectins are stable upto 3M concentration of urea and then slowly start losing the activity at 4M. There is a major loss of activity at 7M urea, although there is no major structural change in the protein as indicated by an insignificant shift in the λ_{max} to 339 nm. The proteins are almost unfolded in 6M Gdn-HCl, while there is no major structural change in 7M urea. Still, the loss in activity is comparable to Gdn-HCl treated lectins. The reason could be that the high concentration of urea either prevented the access of RBC surface to the lectin binding site or caused a slight change in the binding site geometry of the lectin.

4.5.2 Binding of ANS: a fluorimetric analysis

In none of the above conditions the hydrophobic dye was found to bind any of the proteins, indicating lack of exposure of hydrophobic patches in the protein. To study the role of disulfide linkages maintaining the unusual thermostability of these cucurbitaceae lectins, thermal and also Gdn-HCl denaturation studies were carried out in the presence of 5mM DTT as well as with 5mM β-mercaptoethanol. In both the cases, proteins were found to be stable upto 85°C ruling out any involvement of disulphide linkages in the stability of the *trichosanthes* lectins. Both the lectins contain two non-identical subunits linked by disulphide bonds (Komath *et al.*, 1998; Sultan *et al.*, 2004).

4.5. Conclusions

Thermal, chemical and pH stability of the lectins from *Trichosanthes anguina* and *Trichosanthes dioica* are studied. Hemagluttinating activity of both the lectins is stable even after incubating the lectins upto 90 °C. The tryptophans are buried in the hydrophobic environment. The relative fluorescence intensity decreases as the temperature increases due to deactivation of the excited singlet and again increases after cooling the sample to 30 °C. As the concentration of Guanidine-HCl is varied, at 4M, both the proteins start unfolding and they are totally unfolded at 6M Gdn-HCl. However, in case of urea, the lectins are only partially unfolded even at 7M concentration of the denaturant. The proteins are stable in the pH range of 6-12 and partial unfolding takes place at extreme alkaline pH. Change in the polarity of tryptophan residues for both the lectins at alkaline pH did not result in much loss of activity. Although inter-chain disulfide bonds are present in these lectins, their role in protein stability is negligible since treatment with DTT or β-mercaptoethanol reduced the thermostability only by 5°C. In none of the above conditions ANS could bind to any of the above lectins indicating unexposed

hydrophobic patches on the surface of these proteins. The unusual stability of *Trichosanthes* lectins resembles that of some of the RIPs.

Some of the RIPs are also highly thermostable (Lam & Ng, 2001). It has been pointed out that both these *Trichosanthes* lectins have similarity with RIPS (Manoj *et al.*, 2001; Sultan *et al.*, 2004**)**; here we show that like some RIPs, these lectins are resistant to a wide range of denaturing conditions.

Chapter 5

Crystallization and X-ray crystallographic studies of a Galactose/*N*-acetyl Galactosamine-specific lectin from *Erythrina indica*.

5.1 Summary

This chapter describes the structural study of the lectin from the seeds of *E.indica* (coral tree). The lectin belongs to the large family of legume lectin. The structure has been determined at 2.5 Å and solved by molecular replacement using the coordinates of EcorL (PDB code: 1FYU). EiSL is a homodimeric lectin and *N*-linked glycosylation was observed at four sites in the structure *i.e.* 17 and 113 in both the chains. A heptasachharide at B17 and a hexasaccharide at B113 were traced properly in the electron density; while A-chain at both locations showed a trisachharide comprising of Nag and Fuc residues. Structurally EiSL varies from the very closely related lectins of the same genus EcorL and ECL, with respect to a few amino acid residues. More than 90% residues have been shown to be conserved. Similar folds, conformation of loops, type of coordination bonds at the metal binding sites and residues at the sugar combining sites have been observed in EiSL as in EcorL and ECL. A galactose molecule was bound at a position that is close to the glycosylation site in a-chain (A 116 Asp) but it has been observed to be facing away from the N-linked heptasaccharide.

5.2 Introduction

Glycosylated lectins belonging to leguminoseae possess diverse types of N-linked oligosaccharide structures in spite of their homologous primary sequences, are the model systems for understanding the *N*-glycosylation patterns in proteins and the factors determining them. Structural and biochemical study of plant lectins has been going on since decades, and hence this class of carbohydrate binding proteins provides a model system for studying the molecular basis of these recognition events. Among the proteins present in the legume seeds and tubers, lectins comprise major component, posses conserved primary sequences, but they exhibit a wide range of carbohydrate specificities (Loris *et al*., 1998). The data that can be made

available from the amino acid sequences and structures of these lectins will provide the basic framework for understanding protein molecular structure and also shed light on evolutionary relationship of this class of proteins. Besides chemical and biological information, the structural data on these proteins will give us an insight into their functional role in biological systems (Sharma & Surolia, 1997; Vijayan & Chandra, 1999; Chandra *et al*., 2001).

The well conserved **Legume lectin family** comprises of lectins characterized from the seeds of these legumes and which share some unique structural characteristics. Though the function of these lectins is mainly in plant defense and as storage proteins, further functions need to be completely deciphered. Their ease of availability and detailed information of their three-dimensional structures have turned them into model systems for study of protein-carbohydrate recognition and interactions, as well as for use in immunology and glycobiology. The lectins from this family have been well studied and structurally characterized such as ConA, peanut lectin, soyabean lectin, phytohemaglutinnin (Edelman *et al.*, 1972; Einspahr *et al.*, 1986; Loris *et al.*, 1993; Dessen *et al.*, 1995; Hamelryck *et al.*, 1996). The legume lectins from especially the *Erythrina* sp. such as *E.corallodendro* and *Ecristalgalli* have been extensively studied biochemically as well as structurally and have been shown to exhibit striking similarity with each other.

The first structure of a galactose specific lectin is from the seeds of *Erythrina corallodendron* (EcorL) (Shaanan *et al.*, 1991**)**. The *E.cristagalli* lectin (ECL) from the same genus is shown to contain an N-linked heptasachharide: (αD-Man-(1→4)- αD-Man-(1→3)-[β-D-Xyl-(1→2)]- β -D-Man- β -D-GlcNAc-(1→4)-[α-L-Fuc (1→3)]- D-GlcNAc) (Ashford *et al.*, 1991).

Erythrina indica seed lectin (*EiSL*) studied here is a heterodimer of 30 and 33 kDa and shows galactose specificity (Konozy *et al.*, 2002). The *Erythrina indica* seed lectin was supplied to us in pure form for crystallographic studies at a concentration of 15 mg/ml. Here we report the structure of this lectin that was solved by x-ray crystallography.

5.3 Legume Lectins

According to the Royal Botanical Gardens the Leguminoseae is economically and agriculturally important third largest family of flowering plants comprising of 730 genera and more than 19400 species of plants. It has been commonly known as the legume family, pea family, bean family or pulse family. The members of leguminosae family in the plant kingdom are important sources of food (being a rich source of a variety of proteins) and are herbs. The legumes include plants which have wide uses, such as a large number of them have been domesticated and harvested for human consumption and animal feed, as well as to provide fiber, oil, fertilizers, dyes, fuels, biodiesels, and some varieties which have medicinal, horticultural, industrial, chemical, and aesthetic importance. Legumes have the ability to fix atmospheric nitrogen as they harbor the nitrogen fixing bacteria *Rizobium radicicola* in their roots, hence the seeds and foliage have a comparatively higher protein content making them desirable crops in agriculture. Because of these applications proteins from this plant family have gained much importance and have been well characterized by various biochemical and structural methods and molecular biological methods have dwelled on isolating and propagating the genes, for increasing their yield and increasing the useful protein content in the seeds for consumption as well as for other medicinal uses.

Lectins from this family are found as oligomers which may be dimers or tetramers (dimer of dimers) and have a characteristic carbohydrate binding domain. The *in vitro* activities such as cell agglutination, cell-receptor cross linking and stimulation of mitosis of these lectins involves binding of carbohydrate at multiple sites.

The legume lectins whose structure has been deposited in the Protein Data Bank are: concanavalin A (Edelman *et al.*, 1972) (Hardman & Ainsworth, 1972), PHA-L (Hamelryck *et al*., 1996.) , the lectins from pea (Einspahr *et al.*, 1986), lentil lectin (Loris *et al.*, 1993) (Loris *et al.*, 1994), *Lathyrus ochrus* lectin (Bourne *et al.*, 1990), peanut lectin (Banerjee *et al.*, 1994), soybean lectin (Dessen *et al.*, 1995), coral tree lectin (Shaanan *et al.*, 1991), *Griffonia simplicifolia* lectin (Delbaere *et al.*, 1993).

5.3.1 The Characteristic Legume lectin fold / Jelly roll fold

The structure of legume lectin monomer is quite conserved. A common tertiary structure that was first observed in legume lectins is the "Legume lectin fold" as it was first found in this family. The fold is structurally well conserved and consists of two large β-pleated sheets sandwiching a carbohydrate site in this fold. The jelly roll is formed by two antiparallel β-strands that arrange as two β-sheets (Srinivas *et al.,* 2001). Besides legume lectins this fold is also found in animal lectins, the Galectins, Pentraxins (Crennell *et al*., 1994) as well as some non lectins as the *Vibrio cholerae* sialidase and several other glycosidases. This loop is generally associated with carbohydrate binding activity (Loris *et al*. 1998; Vijayan & Chandra, 1999). It is also called as **Jelly roll fold** as it resembles the jelly roll topology found in structures of viral coat proteins.

The structure consists of three β-sheets: the 6-stranded back β sheet, the 7-stranded front β-sheet and a smaller 5-stranded β-sheet (S-sheet) that holds the two large sheets together in

proper conformation. One of the prominent hydrophobic cores is located between the back and the front sheet. Another hydrophobic core is formed between the front sheet and a loop.

These metal binding sites were first described in detail for concanavalin A and have been found to be extremely well conserved in all other legume lectin structures. The calcium and the transition metal ion are approximately 4.5 Å apart and bridged by two aspartate residues. Both metals have four protein ligands and two water ligands.

5.3.2 Canonical legume lectin dimer

It has been well known that legume lectins are reported to exist in oligomeric state and they generally form dimers or tetramers. Most legume lectins contain a structure called as the "Canonical legume lectin dimer". This dimeric structure has been first described in ConA (Edelman *et al.*, 1972) and later on found in many other lectins of this family such as pea lectin, lentil lectin, *Lathyrus* lectin (Einspahr *et al.*, 1986; Bourne *et al.*, 1990; Loris *et al.*, 1993) etc. It consists of two 6-stranded β-sheets that associate in a handshake mode to form a 12-stranded β-sheet. It has been observed that in lectins from *Griffonia simplicifolia* and *Erythrina corallodendron* and *E.cristagalli* (Shaanan *et al.*, 1991; Delbaere *et al.*, 1993; Svensson *et al.*, 2002) the association of the monomers is different from that observed in the canonical mode of dimerization, and it is assumed to be impaired because of the glycosylation pattern of these lectins.

5.4 *Erythrina* lectins

Lectins that have been biochemically as well as structurally characterized from the *Erythrina* genus include those from *E.corallodendron, E. cristagalli, E.indica, E.costarensis, E.variegata E.arborescens, E.lithosperma* and *E suberosa, E.speciosa, E.velutina* (Bhattacharyya *et al.*, 1981; Gilboa-Garber & Mizrahi, 1981; Bhattacharyya *et al.*, 1986). Among them, the

lectins isolated specifically from *Erythrina* sp such as *E.corallodendron* & *Ecristalgalli* have been extensively studied at biochemical and structural levels, in native sate as well as in complex with sugars, and they show striking similarity with each other (Iglesias *et al.*, 1982; Shaanan *et al.*, 1991; Adar *et al.*, 1989; Elgavish & Shaanan, 1998; Konozy *et al.*, 2002; Svensson *et al.*, 2002; Turton *et al.*, 2004; Kulkarni *et al.*, 2004). Thus this structural similarity in lectins belonging to various species from the same genus, though found in different geographical locations, indicates evolutionary relatedness between them and emphasizes the functional role of these proteins.

5.4.1. *Erythrina* lectins : *E.corallodendron, Ecristalgalli & E.indica*

Erythrina lectins, especially those from *E.corallodendron, E.cristagalli* and *E.indica* constitute a homologous group of lectins that is galactose/*N*-acetylgalactosamine specific. These are metalloproteins containing Ca^{++} and Mn^{++} ions that are strongly bound to the lectins, but are not essential for carbohydrate activity. They are dimers containing two non-covalently linked glycosylated subunits which may be of nearly identical molecular weight. All these are glycoprotein with 3-10% carbohydrate content. They exhibit similarity with respect to their N-terminal sequence (up to 14-15 amino acids) and also their primary sequence is quite identical with certain portions of the sequence being highly conserved.

They have similar amino acid composition to other legume lectins, being rich in acidic (aspartic acid, glutamic acid) and hydroxylic amino acids (serine, threonine), valine, cysteine absent and methionine being very low. The carbohydrate composition of the lectins which is also very similar shows mannose as the most abundant sugar (39.5-48% of the total sugar residues) and other sugars such as xylose, arabinose, fucose, glucose and galactose, and glucosamine as

the only amino sugar present in majority of them (Bhattacharya *et al.*, 1986). The *E.indica* lectin like the ECL & EcorL has been shown to bind specifically to terminal unsubstituted *N*-acetyllactosamine (Galβ4GlcNAc). Each subunit of the lectins contains no interacting combining sites. The *E.cristagalli* lectin has been shown to bind to thymocytes, undifferentiated carcinoma cells, and some embryonic cells, hence this lectin may be a useful marker to track changes in the surface carbohydrates during differentiation of lymphocytes and embryonic cells.

These lectins have MW in the range of 56,000-68,000 Daltons as determined by different methods. Their hemagluttination activity is also found to be similar as they are non specific towards human erythrocytes, but have a preference for blood group O. *E.cristagalli* (Iglesias *et al.*, 1982) and *E. corallodendron* (Gilboa-Garber & Mizrahi, 1981) lectins are mitogenic towards human T lymphocytes, most of this family of lectins being able to stimulate human peripheral blood lymphocytes. Our lectin closely resembles these two well studied lectins in terms of amino acid composition (Table **5.1**), sugar composition and the structure of sugar binding site.

Table 5.1. Amino acid composition of purified EcorL, ECL and EiSL in Residues/mol.

	EcorL	**ECL**	**EiSL**
Asp	60	62	63
Thr	41	43	44
Ser	47	47	51
Glu	58	55	61
Pro	39	39	34
Gly	41	39	38
Ala	40	40	40

Val	33	42	42
Met	4	6	6
Ile	29	30	29
Leu	36	37	37
Tyr	18	20	22
Phe	28	28	29
His	9	8	10
Lys	19	20	18
Arg	11	11	11

5.4.2. *Erythrina indica* seed lectin

The lectin from *Erythrina indica* seeds has been reported to be isolated by affinity chromatography on acid-treated ECD-Sepharose 6B (Bhattacharya *et al.*, 1981). In SDS-PAGE it gave two bands corresponding to 30 and 33 kDa, and the molecular weight of the entire protein as determined by sedimentation and diffusion data was 68,000 Daltons. Its amino acid composition was like other *Erythrina* sp lectins, but it did not contain any cysteine or methionine residues whereas valine was the N-terminal residue. The acidic residues and serine, threonine constituted 40% of the total amino acid residues of the lectin. Sugar content as determined was 9% neutral sugar and 1.5% glucosamine. The lectin contains two non-interacting identical binding sites for sugars. Like the other *Erythrina sp* lectins it is a metalloprotein containing Ca^{++} and Mn^{++} ion, not essential for hemagluttination.

Only a slight loss of activity occurred after exposure to 50ºC for several days indicating thermal stability at this temperature. Complete loss of activity has been observed at 80ºC.

Thermal stability was increased in the presence of lactose or galactose which is the inhibitory sugar for the lectin activity. The lectin has shown to be stable in pH range 3.0-9.0. Partial deactivation has been shown in the lectin activity below pH 3.0 and above pH 9.0, while total deactivation was observed above pH 12.0 (Ghosh & Mandal, 2006).

5.5. Dialyzing the purified lectin to remove sugars and salts

The pure protein that was obtained from our collaborators contained high amount of NaCl, and galactose in solution. It was extensively dialyzed against 5 mM sodium phosphate buffer at pH 7.2 containing 10mM NaCl, at 4°C. Overall four buffer changes of the dialysis buffer were given, one change after every four hours interval, and before the last buffer change time interval was 10h. The *E.indica* lectin thus obtained was concentrated ten folds, and used for crystallization trials.

5.6. Crystallization of the EiSL by hanging-drop vapor-diffusion method

The protein dialyzed and concentrated as described above was used for crystallization trails and crystallization screens were set up under varied conditions. Hanging-drop vapor-diffusion method was the crystallization technique employed as described in chapter 2. The protein solution was equilibrated in hanging drop against a well solution containing 0.1M Tris-HCl buffer pH 8.0 and 12% PEG 8000. 1 µl of this well solution was added to 1µl of protein solution in the drop before setting up crystallization. Crystallization was carried out at 295 K. The buffer had 0.1% sodium azide. The data collection statistics are presented in table 5.2.

Table 5.2: Summary of crystal properties and diffraction data statistics

Temperature	25 °C
X-ray source	Rotating Anode (Cu-Kα)
Wavelength	1.542 Å
Resolution limits	20 - 2.5 Å (2.66-2.5 Å)
No. of observations (total no. of reflections used)	30204
No. of unique reflections	10251
Completeness of data (%)	95.6 (91.9)
R_{merge} (%)	0.068(34.8%)
Average $I/\sigma(I)$	12.37 (2.34)
Mosaicity (°)	1.4
Space group	$P6_5$
Unit Cell Volume ($Å^3$)	1277043
Unit Cell Parameters	a=b=134.18 , c=81.66 Å
Matthews Coefficient (Vm)($Å^3$ Da^{-1})	2.13
Solvent content	42 %

5.7. X-Ray diffraction, Data Collection and Processing

The crystals under the above mentioned condition appeared in a couple of days and grew to full size in seven days. The crystals were frozen under cryostream of liquid Nitrogen (~120 K), and the cryoprotectants tested were 20-30% glycerol, out of which best quality diffraction

was obtained with 30% glycerol as cryoprotectant. Data was collected on a Rigaku R-axis IV $^{++}$ image plate mounted on a Rigaku rotating anode at 103 K. X-rays were focused using confocal mirror system. Data was 95.6% (91.9%) complete at 2.5 Å resolution. The programs DENZO & SCALEPACK were used for indexing and scaling the data as described in Chapter 2, Materials & Methods. Data was processed in space group P6 using Denzo and the estimated R-merge was 0.067(0.146). The unit cell dimensions were $a = b = 134.18$, $c = 81.66$ Å.

5.8. Alignment of the N-terminal sequence with the sequence of lectins from the same genus

The N-terminal sequence of *E.indica* lectin has already been determined as Val-Glu-Thr-IIe-Ser-Phe-Ser-Phe-Ser-Glu-Phe-Glu-Ala-Gly-Asn-Asp-X-Leu-Thr-Gln-Glu-Gly-Ala-Ala-Leu- (Konozy *et al.*, 2002). The amino acid sequence (15 residues) of 10 lectins from this genus shows an extensive homology (Adar *et al.*, 1989). In the overall sequence, at some 39 positions the residues are invariant in all these lectins. Sequence analysis of the N-terminal (25 residue sequence) of EiSL was performed using BLAST (Altschul *et al.,* 1997) at the NCBI site (http://www.ncbi.nlm.nih.gov/BLAST/). A BlastP search was performed against different databases such as the non-redundant database, the PDB protein database and Swiss PDB database. The results having significant e-values were selected. In each case the top few results obtained were of the EcorL and ECL thus indicating high sequence similarity within the same genus. Blastp against the PDB database gave various structures of lectin from *E.corallodendron and E.cristagalli* as top hits with significant e-values. The results are tabulated below (Table 5.3).

Table 5.3: Results of the BlastP search using the N-terminal sequence of *E.indica* lectin searched against the protein data bank. (The structures used in MR calculation have been shown in bold)

Sr. No.	PDB structure	Description	Z-score (bits)	Evalue	Identity (%)
1	1GZ9	Chain A, High-Resolution Crystal Structure Of *Erythrina cristagalli* Lectin In Complex With 2'-Alpha-L-Fucosyllactose	73.6	6e-15	92
2	**1FYU**	**Chain A, Crystal Structure Of *Erythrina corallodendron* Lectin In Hexagonal Crystal Form**	73.6	6e-15	92
3	1AX0	Chain A, *Erythrina corallodendron* Lectin In Complex With *N*- Actylgalactosamine	73.6	6e-15	92
4	1SFY	Chain A, Crystal Structure Of Recombinant *Erythrina corallodendron*	71	4e-14	88
5	1LTE	Chain A, Structure Of A Legume Lectin With An Ordered N-Linked Carbohydrate In Complex With Lactose	71	4e-14	88
6	**1UZY**	**Chain A, *Erythrina crystagalli* Lectin**	65.5	2e-12	83
7	1N47	Isolectin B4 From *Vicia Villosa* In Complex With The Tn Antigen	29.5	0.12	45
8	1BZW	Chain A, Peanut Lectin Complexed With C-Lactose	26.1	1.2	57
9	2PEL	Chain A, Peanut Lectin	26.1	1.2	57

5.9 Structure Solution by Molecular Replacement using *E.corL* coordinates

From the sequence alignment using blast it was clear that the N-terminal of *E.indica* lectin showed very good homology with both *E.corallodendron* lectin and *E.cristagalli* lectin. Hence the EcorL coordinates having two molecules in the asymmetric unit (1FYU) was chosen as the search model for MR. Since ECL structure (1UZY) also exhibited very good similarity, it could also be used for molecular replacement. The *AMoRe* program package (Navaza, 1994) implemented in the CCP4 suite (Collaborative Computational Project, Number 4, 1994) was used to solve the structures by MR method. The data in the resolution range 20-2.5 Å was considered for molecular replacement.

Rotation and translation functions were calculated followed by a rigid body fit. Details about the molecular replacement solutions are given in Table 5.4. As evident from the Mathews number calculations (2.13), the lectin is a dimer in the asymmetric unit and the two monomers are related by a crystallographic two-fold axis.

The position of the molecule in the unit cell of interest is defined in terms of the three rotational and three translational parameters. *AMoRe* program from the CCP4 suit of programs is a fast rigid-body refinement procedure, which is very efficient in the evaluation of the correctness of the molecular replacement solution. It was first developed by Huber & Schneider (1985) and involves minimization of the misfit, where rotations and translations are calculated by interpolation and phase shifts and is not a least-squares rigid-body refinement of coordinates. The rotation function calculations (using the program 'roting' of *AMoRe*) resulted in small number of peaks with nearly similar correlation coefficient (Cc) and R_{factor}. The significant solutions output by the program indicated the orientations of dimers in the asymmetric unit.

Translational search calculations confirmed the dimer in the asymmetric unit. The program 'traing' of *AMoRe* was used for the translation function calculations that output correct solutions.

The final solution selected was the one having high value of correlation coefficient (34.5) and with low R_{factor} (49.5). The correct solution was then used as input in the 'fitting' program of *AMoRe* and a least-squares refinement was carried out. The parameter Cc further increased in this cycle to 48.7 and the value of low R_{factor} decreased further to 45.5 that indicated a correct solution (Navaza, 1994; Navaza & Saludjian, 1997). Table 5.4 lists the MR solutions with final orthogonal rotations (rx, ry, rz) and translations (tx, ty, tz), which were applied to the MR model to generate the initial model of EiSL from the coordinates of EcorL.

Table 5.4 Molecular Replacement statistics

ROTATION FUNCTION

(Resolution Range 20 - 2.5 Å)

Solution	α	β	γ	Cc	R-free
Solution	4.40	122.83	173.61	20.9	53.8

TRANSLATION FUNCTION

(Resolution Range 20 - 2.5 Å)

Solution	Tx	Ty	Tz	Cc	R-free
Solution	-0.4478	0.4389	-0.1588	34.5	49.5

RIGID BODY FIT

(Resolution Range 20 - 2.5 Å)

Solution	α	β	γ	Tx	Ty	Tz	Cc	R-free
	4.820	122.64	173.45	-0.448	0.4389	-0.000	48.70	45.50

5.10 Structure Refinement

Refinement is the process of fitting the parameters of the model to achieve a closer agreement between the calculated and observed structure factors. The refinement of the obtained model of EiSL protein was refined to obtain a better model with good statistical values of R_{factor} and R_{free} as well as a good geometry.

5.10.1. Model Building and refinement

In the model, the two monomers in the asymmetric unit were labeled A and B respectively (as the convention in EcorL and ECL). As both these chains are symmetrically related, B could be generated from A through non-crystallographic symmetry (NCS parameter in Refmac5). The XAUTOFIT module from the Quanta was used for model building. As described above the initial coordinates of the molecule were obtained from 1FYU of EcorL. The electron density maps generated were used to model the protein. After each refinement cycle, the model was visually inspected using the program *QUANTA* using (2*F*o – *F*c) and (*F*o – *F*c) difference maps. After each such step refinement was carried out using the program Refmac5. Iterative cycles of model building and refinement was carried out till the residues in the model were fitting properly in the electron density map and all the extra density was taken care of. The

sequence alignment of EcorL and ECL was frequently referred to and accordingly mutations of amino acids were carried out wherever the side chain was either out of the electron density or electron density was present for the side chain.

For the A-chain 239 residues have been traced in the electron density while for the B-chain 238 residues could be fitted according to both (2*F*o – *F*c) and (*F*o – *F*c) difference maps. The modeling and refinement was carried out till there was a good agreement between the values of R_{factor} and R_{free}.

5.10.2 Modelling the lactose molecule in the binding cleft

During the steps of model building and refinement a difference density (*Fo-Fc*) > 3σ was observed near the residues Asp89, Asn133, Gly107 in both the chains. This is the carbohydrate binding pocket as described for both EcorL and ECL and hence a lactose molecule was modeled in this electron density in both the chains. After some refinement cycles the lactose in both the A and B chains was found to fit properly in the sugar binding pocket as it has been described for other legume lectins, and it is stabilized by forming certain hydrogen bonds with the active site residues. After the addition of the lactose a further drop in values of both the R_{factor} and R_{free} indicated correctness of model.

5.10.3 The *N*-linked sugars (the heptasachharide and other oligosaccharides)

E.indica lectin has been shown to be glycosylated, containing N-linked complex oligosaccharides containing galactosamine as the amino-sugar, fucose, xylose, glucose and galactose amongst other sugars and especially high in mannose content (Bhattacharyya *et al.*, 1986). AsN 17 and AsN 113 have been identified as the glycosylation sites in both the chains.

After many rounds of refinement, and placing the galactose in the binding site cleft further refinements were carried out. In both the A & B chains (*Fo-Fc*) > 5 σ density was observed at the all the four glycosylation sites (residues 17 and 113 in both the chains). Refinement followed every round of building the oligosaccharides. If at any N-linked sugar location after refinement, if the (*Fo-Fc*)> -5σ, it was an indication of wrongly placed sugar, and hence it was later removed. Thus only after iterative cycles of model building, refinement and visual observation of the residues, the above mentioned sugars were modeled in the respective densities. The CIF monomer library was created using the program MergeMonomer Library of CCP4 suit, and finally merged to get one LIB file.

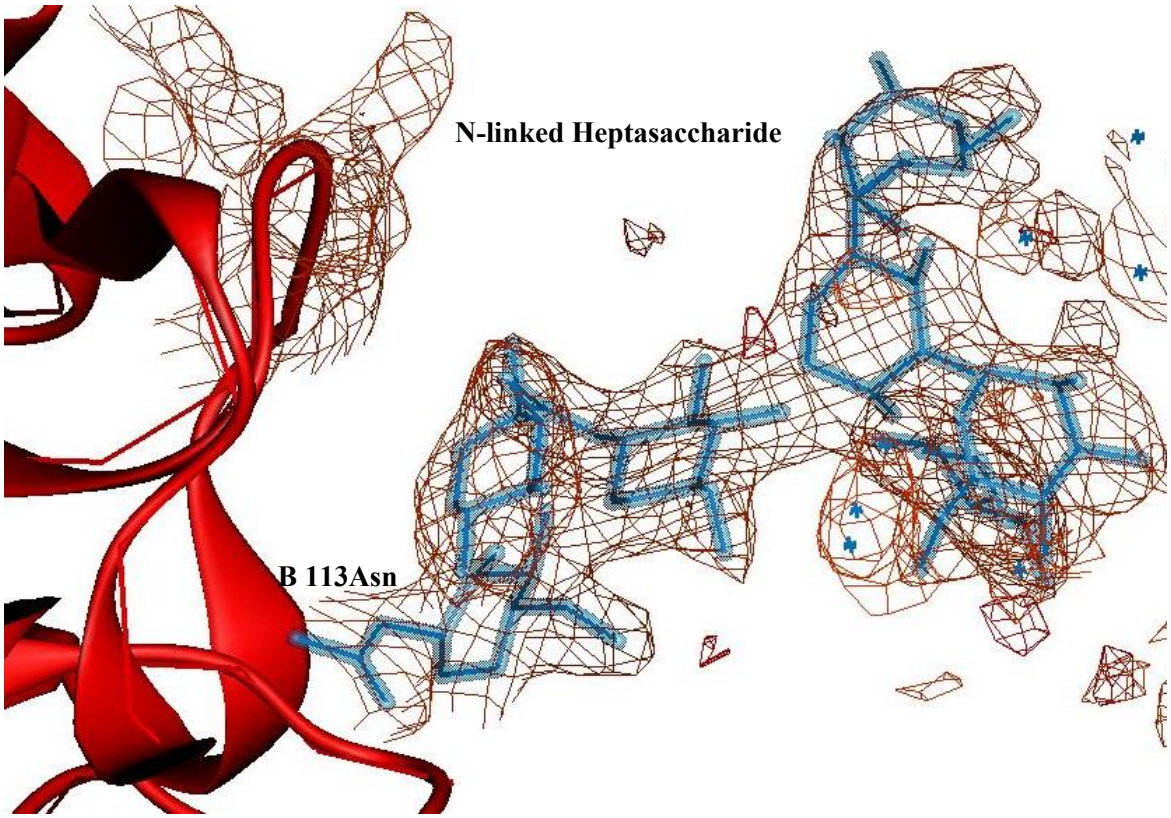


Fig 5.1 N-linked sugar at position B113 with *Fo-Fc* map countered at 1σ.

5.10.4 Placing the metal ions: Ca^{++} and Mn^{++} in the coordination sites

The electron density corresponding to (*Fo-Fc*) > 5 σ was observed in the location where Ca^{++} and Mn^{++} have been observed in both EcorL and ECL. These two metal ions have been placed in the cavity surrounded by residues Glu127; Asp129; Phe131; Asn133; Asp136; His142 such that they form coordination bonds with them. After placing the metal ions, refinement was carried out in Refmac5 program of CCP4.

5.10.5 B-Factor Refinement

After model was built and the values of R_{factor} = 0.22 and R_{free} = 0.28 were obtained. The difference Fourier map (*Fo-Fc*) seemed good at this stage with no large blobs in any part of the density. Following this refinement, individual B-factors were refined isotropically using the features from the program of CCP4 suit. Assuming here that all atomic vibrations have spherical symmetry, the B-factor was calculated. Individual B-factor refinement was carried out for the model alternatively along with positional refinement until a good convergence between the Rvalues was obtained. Now it was considered appropriate to proceed for adding solvent molecules in the final model obtained.

5.10.6 Adding solvent molecules to the refined structure

Model building and refinement was done till the major electron density blobs in the model were properly refined. When all the larger blobs of density were take care of by fitting sugar ligands and oligosaccharides, and the B-factor refinement was been carried out, water molecules were added. The XSOLVATE module of Quanta was used for adding solvent molecules to the refined model. Water molecules were added using (*Fo-Fc*) map as reference. Initially at (*Fo-Fc*) > 5σ level 80 water molecules were added, and subsequently refined. The

water molecules were then added at (*Fo-Fc*) > 3σ level. The criteria used for selecting water molecules in the structure were:

(i) When electron density is observed in both (*2Fo-Fc*) at lσ level and (*Fo-Fc*) at 5σ level initially and extended up to 3σ level in the final stages.

(ii) Only those water molecules were added satisfying distance limits of 2.4 to 3.6 Å for contacts with the protein atoms or neighboring solvent molecules.

(iii) The added water molecule forms reasonable hydrogen bonds with neighboring molecules. Geometry around water molecule for hydrogen bond formation was also considered.

(iv)The model, after adding all the water molecules was subjected to iterative cycles of B-factor refinement, positional refinement and model building using map calculated using the improved phases to obtain the final model.

A total of 192 water molecules were placed in the molecule. The model after adding all the water molecules was subjected to iterative cycles of B-factor refinement, positional refinement and model building using map from the newly improved phases to obtain the final model. R_{factor} and R_{free} for the final model were 0.1898 and 0.2485, respectively. Refinement parameters are listed in Table 5.5.

5.10.7 Ramachandran Plot and Procheck analysis of the final model

The geometry of the final model was checked using the program PROCHECK (Laskowski *et al*., 1993), part of the CCP4 package. The Ramachandran plot (Ramachandran *et al*., 1963, Ramachandran & Sasishekaran, 1968**)** shows that ~ 86.5% residues lie in the most favored region, while only 1 residue is in disallowed region (B 2Glu). 106Tyr in both the A and the B chain lies in the generously allowed region of the Ramachandran plot (Table 5.5) (Fig.5.1). All the non-glycine residues have phi and psi values in the acceptable range and fall in the

allowed region of the plot. The overall G-factor is the measure of the stereochemistry of the model, and its values as determined and listed in Table 5.5 were an indication of acceptable geometry for the model.

Table 5.5 Refinement statistics

No. of non-H protein, sugar and water atoms	**4151**
Average B-factor	**35.11**
Resolution range (Å)	**20-2.5**
R (%)	**18.98**
R_{free} (%)	**24.85**
R.m.s. deviation from ideal values	
Bond lengths (Å) Bond angles (°)	**0.212** **2.612**
Correlation coefficient	**0.8998**
Error in coordinate by Luzzati plot	**0.3107**
G-factor	
Dihedral Covalent Overall	**-0.47** **0.19** **-0.17**
Ramachandran plot statistics	

Most favoured (%)	**86.5**
Additionally allowed (%)	**12.8**
Generously allowed (%)	**0.5** (A&B 106Tyr)
Disallowed region (%)	**0.2** (B 2Glu)

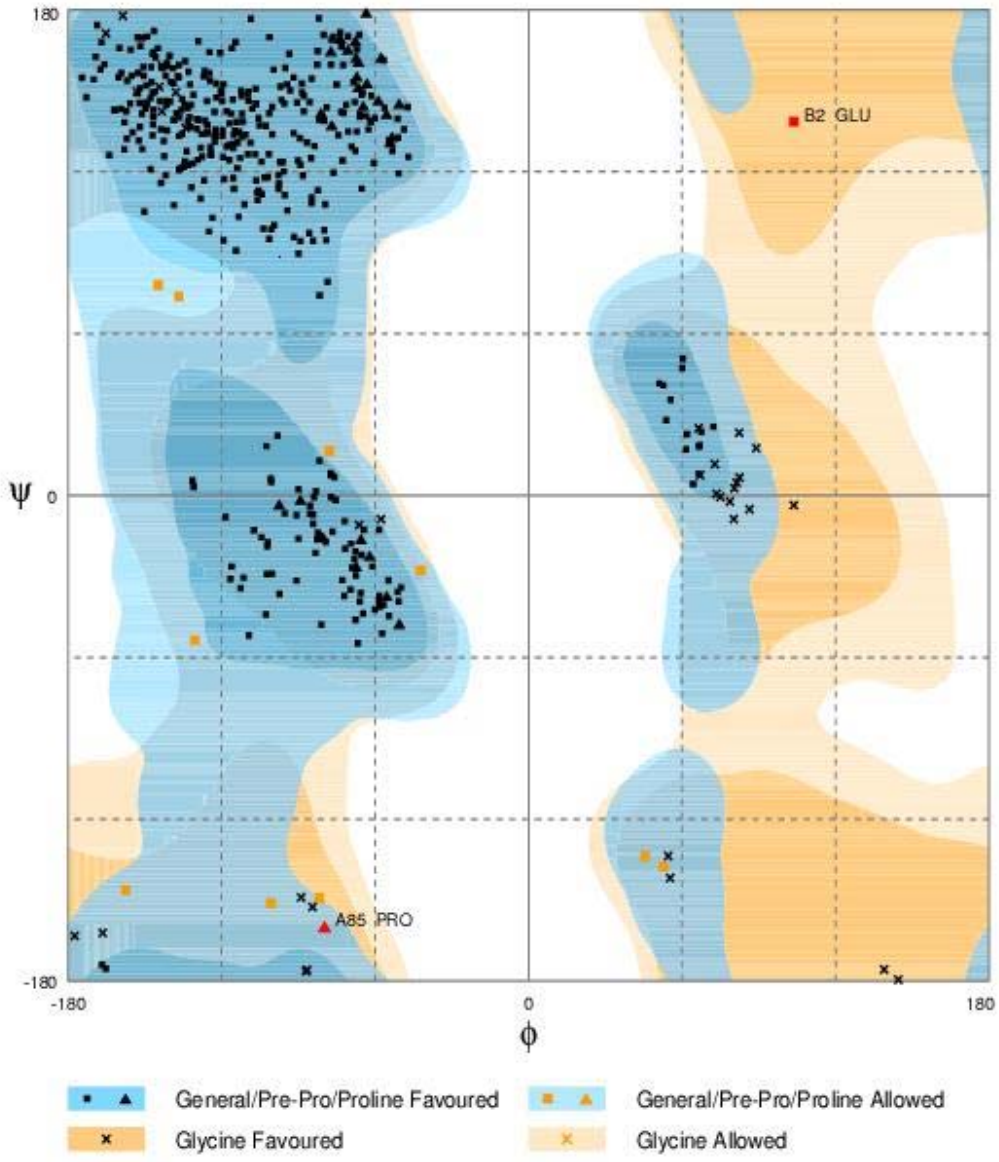


Fig.5.3 Ramachandran plot of EiSL. Triangles and squares represent glycine and nonglycine residues respectively.

5.11. Structural details

5.11.1. The subunit structure: Structure of the A & B chains

In course of these investigations it has been found that the basic structural unit of these proteins that is the monomer is conserved with respect to the three-dimensional structure. The monomer is a β-sheet containing structural unit predominantly made up of two large β-sheets.

The *Erythrina indica* seed lectin exists as a dimer in native state. It consists of two homologous domains, the A and the B chain which is predominantly made up of β-sheets. Majority of the legume lectins form a canonical dimer as described above. While forming the dimer, these chains associate in a handshake mode which is different from the canonical mode of dimerization.

Structurally each of these chains consists of three β-sheets that include each a:

(a) Front 7-stranded β-sheet,

(b) 6-stranded back β-sheet, and

(c) A small 5-stranded β-sheet (also called as S-sheet).

The first two strands of the six-stranded back β-sheets formed as the N and C-termini come closer. It is seen that both the 6- stranded and the 7-stranded sheets are placed anti-parallel to each other in the structure.

5.11.2. Active site (Architecture) characterization: *N*-linked Oligosaccharide & Binding to lactose molecule

E.indica lectin has been reported to be glycosylated, containing N-linked complex oligosaccharides containing galactosamine as the aminosugar, fucose, xylose, glucose and galactose amongst other sugars and especially high in mannose content (Bhattacharyya *et al.*, 1986). The N-linked oligosaccharide has been reported at position 17 and 113 is observed in both the monomers. One of the sites is located at Asn17 that is part of the β-strand in the B-chain. The other prominent carbohydrate binding site is Asn113 which is similar to the site in ECL, but N-linked carbohydrate was not reported at this site in the EcorL structure.

A heptasaccharide (αD-Man-(1→4)- αD-Man-(1→3)-[β-D-Xyl-(1→2)]- β -D-Man- β -D-GlcNAc-(1→4)-[α-L-Fuc (1→3)]- D-GlcNAc) could be placed at B 113 Gln, while at B17 only six out of the seven sugars of the heptasaccharide could be modeled in the electron density. In the A chain at both the locations only a trisaccharide (β-D-GlcNAc-(1→4)-[α-L-Fuc (1→3)]-D-GlcNAc) could be fitted according to the (*Fo-Fc*) > 3σ difference map.

A lactose molecule was thus modeled in this density such that the galactose moiety in the lactose interacts with the residues Asp89, (OD1 & OD2), Gly107 (NH), Asn133 (ND2, NE2) in the binding pocket. These residues along with Ala88 and Phe131 form the sugar binding cavity. The residues involved in carbohydrate binding have been found to be well conserved in all the legume lectins. Two water molecules are also involved in stabilizing the galactose moiety in the sugar binding cleft by forming hydrogen bonds. The residues Ala218 (NH) and Gln219 (NE2) are also involved in hydrogen bonding but are not a part of this basic sugar binding framework.

Fig. 5.4 Structure of EiSL showing the lectin dimers in red. Both the dimers have *N*-linked sugars at two locations which are shown in green. The bound lactose is shown as blue and a galactose reediue was traced which is shown in magenta.

Metal binding sites

The lectins from the genus Erythrina have been shown to be metalloproteins containing Ca^{++} and Mn^{++} ions (Bhattacharya *et al.*, 1986). In the structure of *E.indica* lectin each monomer contains 2 metals ions which form coordination bonds with a few residues. These residues are conserved in most of the legume lectins. Each protomer contains one Ca^{++} and one Mn^{++} ions that are located very close to the carbohydrate combining site. Each of these metal ions is stabilized by forming coordination bonds with two water molecules and also with four amino acid residues located close to the metal atom. The residues that coordinate with the Mn^{++} ion are Glu127, Asp129, Asp136 and His142, whereas the residues coordinate with Ca^{++} are Asp129,

Phe131, Asn133 and Asp136 (Fig 5.5). The tables (5.6 & 5.7) give the details of the distance between the metal atoms and the coordinating atoms in the residues specified above.

Table 5.6 : Distance between the Ca^{++} ion and coordinating atoms

Ca^{++}	Amino acid side chain	Distance (Å)
Interacting atom 1	Asp136 OD2	2.33
Interacting atom 2	Asp129 OD2	2.47
Interacting atom 3	Asp129 OD1	2.3
Interacting atom 4	Asn133 OD1	2.27
Interacting atom 5	Water (19)	2.59
Interacting atom 6	Water (51)	2.33
Interacting atom 7	Asp129 CG	2.74
Interacting atom 8	Phe131 O	2.27

Table 5.7 : Distance between the Mn^{++} ion and coordinating atoms

Mn^{++}	Amino acid side chain	Distance (Å)
Interacting atom 1	His142 NE2	2.34
Interacting atom 2	Glu127 OE2	2.02
Interacting atom 3	Asp129 OD2	2.17
Interacting atom 4	Asp136 OD1	2.06
Interacting atom 5	Water (18)	2.32
Interacting atom 6	Water (56)	2.07
Interacting atom 7	Glu127 CD	3.03
Interacting atom 8	Asp129 CG	3.23
Interacting atom 9	Asp136 CG	3.19
Interacting atom 10	His142 CE1	3.16

These metals are situated close to the carbohydrate binding site and thus maintain the correct spatial orientation of the residues in the combining site. Hence, they are required for the carbohydrate binding activity of all legume lectins and are conserved in the legume lectin structures. The distance between Ca^{++} and Mn^{++} in EiSL is 4.2 Å and 4.03 Å respectively in the A & B chains. Two Aspartic acid residues (D129 and D136) form coordination bond with both the metal ions. Hence these residues bridge the interaction between the metal ions.

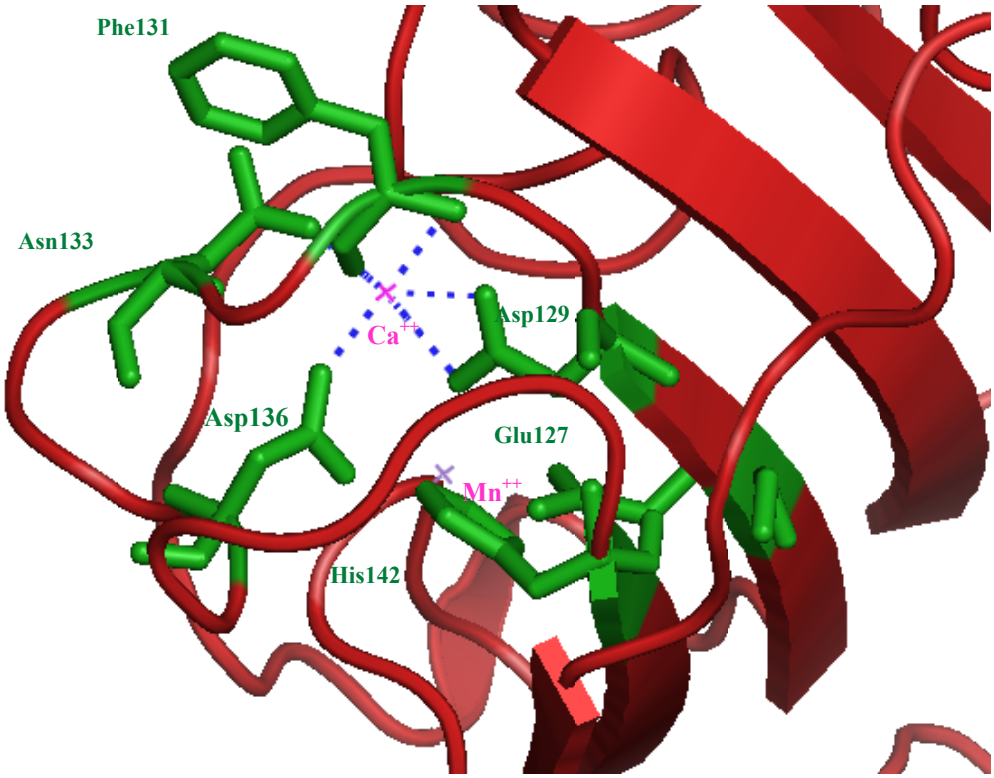


Fig. 5.5 Metal coordination bonds between the Ca^{++} and EiSL residues Asp129, Phe131, Asn133, and Asp136 are shown in the figure. The residues which are involved in coordination with the Mn^{++} are Glu127 and His 142 are also represented but the coordination bonds are not shown.

5.11.3. Thermal Stability of EiSL and correlation with structural characteristics

The legume lectins constitute a group of highly stable multimeric proteins. The presence of the well conserved **jelly roll** fold in the legume lectin structures can be correlated with the available structural information to explore reasons for their increased intersubunit and intrasubunit interactions that confers them high stability. The EiSL has shown to be structurally

stable in various ranges of temperatures, and it differs remarkably in its stability and unfolding behavior as compared to other well characterized legume lectins such as concanavalin A (ConA) (Chatterjee & Mandal, 2003), soybean agglutinin (SBA) (Chatterjee & Mandal, 2003) (Ghosh & Mandal, 2001) and *Erythrina corallodendron* lectin (EcorL). The lectin is stable under highly denaturing conditions such as in 8M Urea at 25 ºC. The hydrogen bonds between the two chains in the refined structure of *E.indica* lectin have been listed in Table 5.8. These hydrogen bonds are found between similar residues as seen in EcorL and ECL also.

Table 5.8 Hydrogen bonds between the two subunits of EiSL

Monomer A	**Monomer B**	**Distance (Å)**
R73 NH1	I191 O	3.03
R73 NH1 via water 144	I191 N	3.01 ; 3.07
R73 NH1 via two waters, 27;180	G189 O	2.95 ; 2.4 ; 2.74
R73 NH2	I191 O	3.05
K154 N via W189	D173 OD2	2.59 ; 3.08
K154 NZ via water 189	D173 OD1	2.59 ; 2.74
K154 NZ via water 189	D173 OD2	2.59 ; 3.08
Q156 NE2 via water 191	K171 NZ	3.05 ; 2.78
N167 via water 48	G189 O	2.62 ; 2.95
K171 NZ	Q156 NE2	3.24

A180 ND1	H180 ND1	3.06
A181 O	H180 NE2	3.44
T193 OG1	K171 NZ	3.01
T193 OG1	H180 NE2	3.27
T193 OG1	H180 NZ	3.01
T193 O	H180 NZ	3.24
T193 N via water 185	K171 NZ??	2.58 ; 3.23
A195 N via Water190	D173 OD2	3.18 ; 3.00
A195 N via Water190	H180 NE2	3.18 ; 3.33

5.12 Conclusion

The lectin from the seeds of *E.indica* was categorized as one of the Galactose/*N*-acetylgalactosamine specific legume. It exhibits non-canonical mode of dimerization in which the two monomers associate in handshake mode that differs from the regular dimerization pattern of majority of the legume lectins like EcorL and ECL. It is because of these *N*-linked oligosaccharides at two positions (a heptasachharide, a hexasachharide and two trisachharides) in both the chains that EiSL exhibits more structural stability as compared to other lectins from the same genus.

Chapter 6

Comparative biophysical studies on two araceae lectins from *Sauromatum guttatum* & *Arisaema tortosum*, by Fluorimetric methods and Circular dichroism (CD).

6.1 Summary

Araceae lectins are a group of closely related proteins which belong to the large superfamily of monocot mannose binding lectins whose structures contain the β-prism-II fold. They show specificity for complex sugars such as desialylated fetuin, are highly conserved with respect to their sequences possessing conserved lectin domains as in other monocot families. This chapter presents a comparative study of the structural stability and dynamics of the two araceous lectins, one from *Sauromatum guttatum* (Voodoo lily) (SGA), and another from *Arisaema tortuosum* (Himalayan cobra lily) (ATL). Among the Araceae lectins, those from *Arisaema tortuosum* schott, *Arisaema consanguineum* Schott (ACA), *A. curvature* Kunth (ACmA), *Gonatanthus pumilus* (GPA), *Sauromatum guttatum* Schott (SGA) and *Alocasia cucullata* (Shangary *et al.*, 1995; Dhuna *et al.*, 2005) have been purified and characterized. All of these lectins show mitogenic potential for human blood lymphocytes and have complex sugar specificity, and show inhibition of hemagluttination by desialylated fetuin. They are thus distinctly different from other monocot lectins in terms of carbohydrate binding specificity.

Both the lectins chosen for the study are non-mannose binding monocot lectins isolated from the tubers of the respective araceous plants. SGA and ATL consist of a mixture of isolectins differing in charge, similar to *Alocasia indica* lectin, WGA (Peumans *et al.*, 1982a, 1982b) and those from Amaryllidaceae and Alliaceae. Recently, SGA as well as some more lectins from *Arisaema* sp. have been found to show antiproliferative as well as anti-cancer activity.

The two monocot lectins have been characterized with respect to their tryptophan environment and secondary structure under chemical, pH and thermal denaturing conditions.

Transitions in the tryptophan microenvironment and secondary structure of these two lectins under different denaturing conditions were studied by steady state and time resolved fluorescence and CD spectroscopy. The lectins exist as tetramers with a single tryptophan residue estimated per monomer, present in a polar environment. Quenching with ionic quenchers showed predominantly electropositive environment for tryptophan residues. Acrylamide had maximum quenching effect. A decrease in KI quenching due to lectin denaturation indicated redistribution of charges as a result of possible conformational change. The two values for lifetimes of tryptophanyl population (1.2-1.4 and 6.3-6.4 ns) reduced substantially on quenching or denaturation. Similarly, both the lectins showed a drastic loss of secondary structure in 5M Gdn-HCl or 6M Urea or at pH 2.0 and below. For the first time araceous lectins, like legume lectins are shown to bind adenine. The presence of a compact structure at alkaline pH 10.0 - 12.0 was observed in CD spectra.

6.2 Introduction

Monocot Mannose binding lectin family is one of the seven families of plant lectins as described in the first chapter. It includes lectins from monocot families namely the Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Liliaceae, Orchidaceae and Iridaceae (Van Damme *et al.*, 1998a, 2000). Most of the plants lectins that have been characterized include lectins from dicotyledons. WGA (Wheat Germ Agglutinin from Gramineae) was the first monocot lectin to be studied (Aub *et al.*, 1965) which was later on followed by studies on other gramineae lectins from *Secale cereale, Hordeum vulgare, Oryza sativa, Brachypodium sylvaticum* and *Agropyrum repens* (Tsuda, 1979; Peumans *et al.*, 1982a; Peumans *et al.*, 1982b; Cammue *et al.*, 1985). Lectins from monocots are gaining increasing importance in recent times as their exclusive specificity towards mannose and mannose containing glycoproteins has been exploited for the

analysis and isolation of Man-containing glyconjugates (Shibuya *et al*., 1988). Some other monocot lectins that have been characterized, include lectins from the families Gramineae (Aub *et al.*, 1965), Liliaceae (Oda *et al.*, 1987), Araceae (Shangary *et al.*, 1995; Kaur *et al.*, 2005), Amarylidaceae, Orchidaceae, Alliaceae (Barre *et al.*, 1996) and Iridaceae (Van Damme *et al.*, 2000). From the studies of past few years, Araceae has emerged as a lectin-rich family with lectins constituting 70-80% of storage proteins in the tubers (Van Damme *et al.*, 1995).

Monocot mannose binding lectins have similar primary sequences but differ from each other in their quaternary organization. It would be of great interest to uncover unique structural features of this family of lectins. It is a well known fact that a protein's native conformation is stabilized by a variety of specific and non-specific interactions between the amino-acid side chains and solvent molecules, such as hydrophobic, van der Waal's, hydrogen bonds, ionic forces, being the stabilizing non-covalent forces and covalent disulphide linkages. Conformational stability of a protein is a term largely employed for denoting the three-dimensional structure that is essential for a protein's physiological and functional role. In solution, protein molecules are subjected to a wide variety of conformational fluctuations. Proteins containing tryptophan and tyrosine have an intrinsic fluorescence emission that can be measured with the excitation at 280 nm and 295 nm, respectively. Tryptophanyl residues which are generally in lesser numbers and contain the indole ring are used as molecular probes to study structural dynamics of protein in solution such as location, physical and dynamic properties of microenvironment of indole fluorophores, and the structural features and behavior of the protein molecule as a whole (Burstein *et al.*, 1973; Lakowicz & Weber, 1973; Demchenko, 1986). Small chemical moieties designated as quenchers can quench or decrease the fluorescent intensity of the tryptophans in a protein molecule and are largely employed to characterize the

microenvironment of the tryptophans. Depending on the environment of tryptophan residues in proteins, the maximum position (λ_m) of tryptophan fluorescence varies from 308 to 353 nm. Steady state fluorescence spectroscopy is a powerful tool to monitor these conformational changes occurring in native and denatured proteins.

An enhancement in the conformational flexibility causes changes in the local tryptophan population. This results in new conformational populations of fluorophores such as tryptophan which can be subsequently measured by fluorescence spectroscopy. Perturbation of the local tryptophan population by denaturants results in an enhancement of the conformational changes, which in turn alters the fluorescent properties of fluorophores such as tryptophan and promotes interconversion of substrates on a nanosecond time-scale (Wu & Brand, 1994; Ervin *et. al.*, 2002). The two decay times of tryptophan observed in water as a solvent can be interpreted as emissions from different rotamers (Eftink & Wasylewski, 1989).

In this chapter we report studies by steady-state fluorescent spectroscopy, lifetime fluorescent spectroscopy and CD which have been used to characterize on a comparative basis the structural properties of the two araceous lectins with respect to their chemical, pH, thermal denaturation, tryptophan environment and secondary structure.

6.3 Araceae Lectins: *Sauromatum guttatum* and *Arisaema tortosum* lectins

Some lectins from Araceae species that have been characterized are *Arum maculatum*, *Colocasia esculenta*, *Xanthosoma agittifolium*, *Dieffenbackia sequina* (Van Damme *et al.*, 1995), *Arisaema consanguineum* Schott (ACA), *A. curvature* Kunth (ACmA), *Gonatanthus pumilus* D. Don (GPA), *Sauromatum guttatum* Schott (SGA) (Shangary *et al.*, 1995), *Arisaema tortosum* schott (Dhuna *et al*., 2005), *Alocasia cucullata* (Kaur *et al.*, 2005).

The two lectins under investigation here are *Sauromatum guttatum* lectin (SGA) and *Arisaema tortosum* lectin (ATL) both being non-mannose binding monocot lectins isolated from the tubers of araceous plants *Sauromatum guttatum* (Voodoo lily) and *Arisaema tortosum* (Himalayan cobra lily). Recently lectins from *Arisaema* sp. has been shown to have antiproliferative as well as mitogenic activity (Singh & Kamboj, 2004; Kaur *et. al*., 2005)

6.4 *Sauromatum guttatum* and *Arisaema tortosum* lectins

The general characteristics of both SGA and ATL, and the anticancerous properities of lectins in general with special mention of the araceae lectins are documented below.

6.4.1 Purification & Biochemical characteristics

The lectins have been extracted from the tubers of *Sauromatum guttatum* Schott (SGA) and *Arisaema tortuosum* and purified by affinity chromatography as described by Shanghary *et.al* (1995) on a column of asialofetuin-linked amino activated silica column (0.8 cm x 3 cm). The unbound protein was removed by washing with 0.01 M PBS (pH 7.2), followed by elution with 0.1 M glycine-HC1 buffer (pH 2.5) with a flow rate of 40 ml hr^{-1} that eluted the lectin.

Both the lectins are homotetramers, with the tetramer MW 49 KDa (SGA) and 54 KDa (ATL) as determined by gel filtration chromatography (Shangary *et al.*, 1995; Dhuna *et al.*, 2005) and it consists of four homotetrameric subunits not linked by disulphide linkages. Pure proteins gave a single band on SDS-PAGE at pH 4.5 corresponding to MW 13 KDa and 13.5 KDa, respectively, while double bands were obtained on non-denatured PAGE pH 8.3.

SGA & ATL agglutinate sheep, rabbit and rat RBC's but are inactive against human ABO erythrocytes (Shanghary *et al.*, 1995; Dhuna *et al.*, 2005). Both these lectins did not require metal ions for hemagluttination activity.

6.4.2 Sugar Specificity & Thermal stability of these lectins

Both these lectins, like other araceae lectins, differ from other monocot lectins in carbohydrate specificity, in that they do not bind to and are thus not inhibited by monosachharides, disachharides, oligosachharides and other sugars as chitin, fetuin and porcine mucin. Their hemagluttination activity was inhibited by desialylated fetuin (125 □g $m1^{-1}$ inhibitory concentration), but by none of the simple sugars (Dhuna *et al.*, 2005). Both these proteins are glycosylated having 0.47% carbohydrate (Singh & Kamboj, 2004; Kaur *et al*., 2005)

The thermal stability of SGA shows that it is stable at 55 °C for 15 min without any loss of hemagluttination activity. When incubated in a boiling water bath for 15 mins, it still retains 25% residual hemagluttination activity (Shangary *et al*., 1995).

6.4.3 Anticancer properties

Some plant lectins possess mitogenicity i.e. they have the unique ability to induce quiescent lymphocytes to divide (Kilpatrick, 1997). Besides, some selected plant lectins have shown to have anti-proliferative (anti-tumor) activity, *i.e.* an inhibitory effect on tumor growth. Some others have show to possess anticarcinogenic activity *i.e.* they possess an inhibitory effect on the induction of cancer by carcinogens. Preliminary investigations on some of them suggest that these lectins are able to detect alterations of malignant cells and they can also decrease the tumorigenicity of the cancerous cell, hence they can be of use in detection of cancer (Gabius *et al*., 1987). Some of the mitogenic lectins are being investigated for their use in cancer research and therapy.

Glycoconjugates on tumor cell surface are functionally important for the interaction of the tumor cell with its environment. Several studies have demonstrated that particular

carbohydrate residues on primary cancers are associated with metastasis. Identification of such residues is possible using lectins as valuable prognostic markers in different types of cancers (Mitchell *et al*., 1998).

Because of their recognition of specific carbohydrates, lectins may also be used as carriers for targeted drug delivery, depending on the glycosylation pattern of the cells and the lectin specificity (Hussain *et al*., 1997; Nantwi *et al*., 1997). Apparently, keeping in view of the diversity of glyco profile expression of various tumors, the detection of lectins with novel sugar specificities assumes great significance.

Plant lectins can distinguish between malignant and normal cells, a property that is attributed to their cell surface carbohydrate molecules (Aub *et al*., 1963).

ATL has been shown to possess *in vitro* anti-cancer potential towards established cancer cell lines. It has anticancer activity against three cell lines; OVCAR-5 (Ovary), SiHa (Cervix) and HT-29 (Colon) (Dhuna *et al.*, 2005). This lectin has shown specificity towards LacNAc, which is one of the most important cancer markers studied so far (Ito *et al*., 1996).

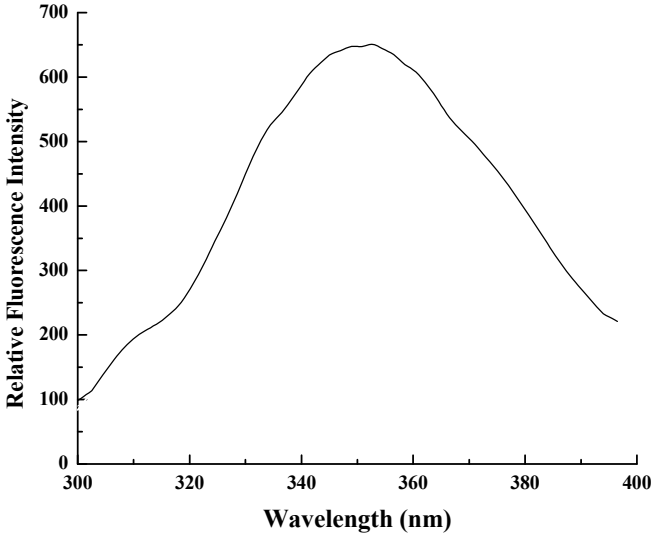
6.4.4 Basis for their Biophysical characterizations

Many lectins and especially the legume lectins have been well characterized with respect to their intrinsic fluorescence and the microenvironment of tryptophan residues under native and denaturing conditions. This has been done at steady state and also by life time measurements of these lectins. These studies give information about the protein structure and conformational changes induced by change in the environment induced by denaturation of the lectin. In these proteins some tryptophans are already exposed to environment while some residues are buried in the protein matrix. Here we have carried out stability studies on these two lectins from

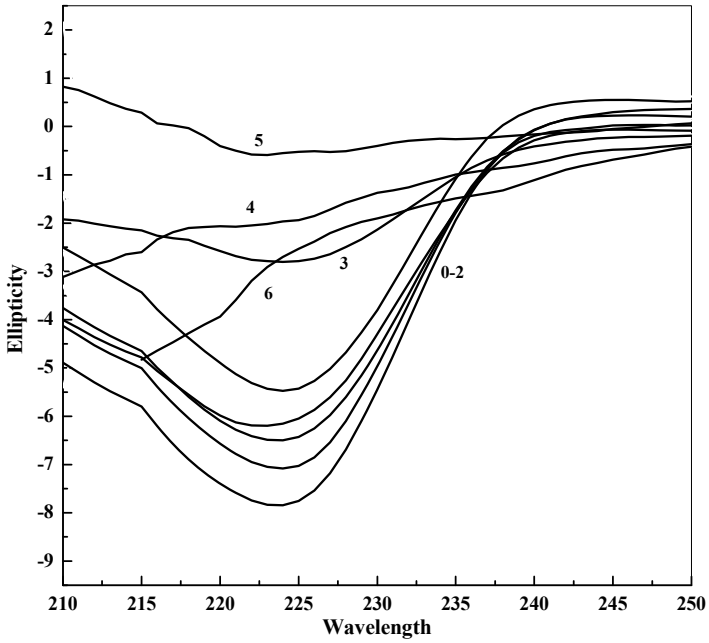
Sauromatum guttatum & *Arisaema tortosum* and we have also probed the microenvironment of tryptophan of these lectins by fluorescence spectroscopy and CD. This is the first report of any such characterizations in Araceae lectin family.

6.5 Results & Discussion

6.5.1 Fluorimetric measurements & Analysis of secondary structures by Circular Dichroism



(A)



(B)

Figure 6.1.

(A) Fluorescence spectra of Native SGA (1.8 μM) at pH 7.2

(B) Far-UV CD scans of SGA, incubated in 0-6 M Gdn-HCl, pH 7.2. The scans were averaged after three accumulations.

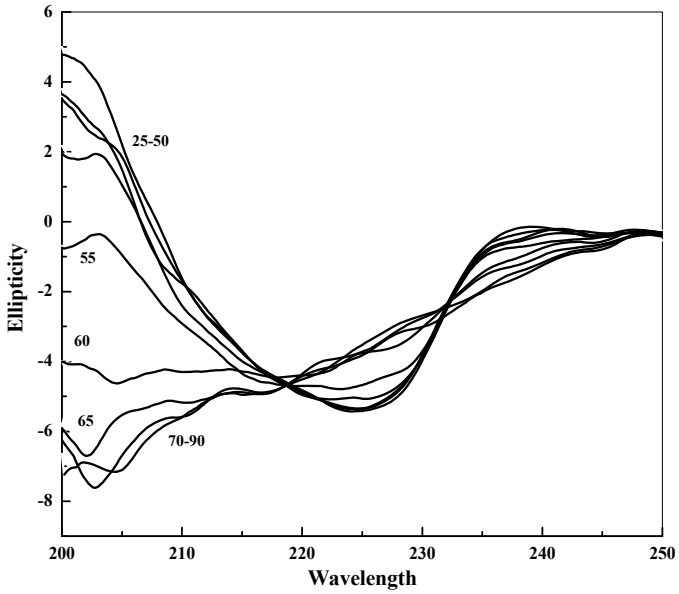
6.5.1.1 Spectra of native protein

The fluorescence emission maxima of SGA and ATL at pH 7.0 were at 350 nm (Fig. 6.1 A), indicative of exposed tryptophan residues existing in a relatively polar environment. The NBS modification of tryptophan in both the native and urea denatured states of the two lectins indicated presence of single tryptophan per monomer and the lectins were tetramers in the native state. These results further confirmed the surface exposed state of tryptophans in these araceous lectins.

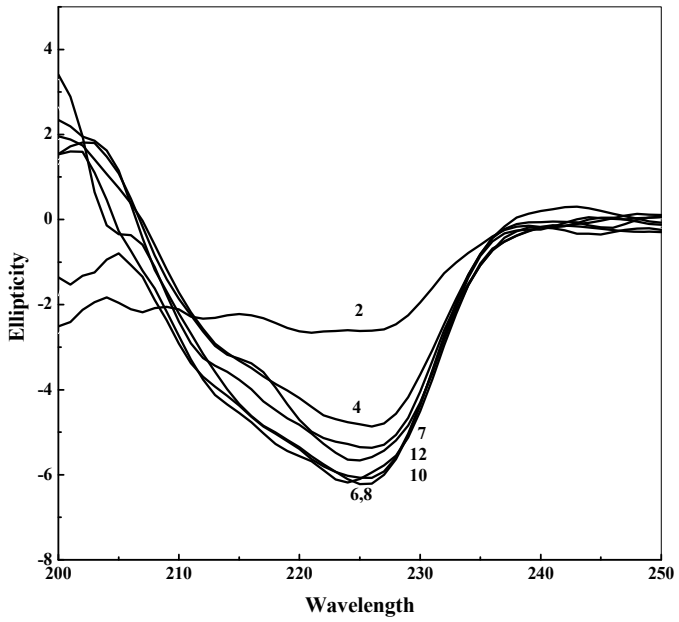
The far-UV CD spectroscopy for secondary structure prediction showed maxima of positive ellipticity at 201 nm for both the lectins, while the minima were 226 nm for SGA, and 225 nm for ATL (Fig 6.1B). The percentage of secondary structural elements in SGA estimated from the far-UV CD spectra showed α-helix: 2.8%, β-sheet: 49% turns: 22.9% and random coil: 25.2%. Similar results were obtained for ATL. The far-UV CD spectra of SGA and ATL resemble the spectra of a synthetic tetrapeptide *tuftsin*, which is shown to be an important immunomodulator and stimulates phagocytosis and is shown to elicit *in vivo* as well as *in vitro* antitumor effects. A negative trough at 225 nm and maxima at 200 nm are found to be associated with type III β turns as in the case of tuftsin **(**Siddiqui *et al.*, 1996).

6.5.1.2 Thermal stability

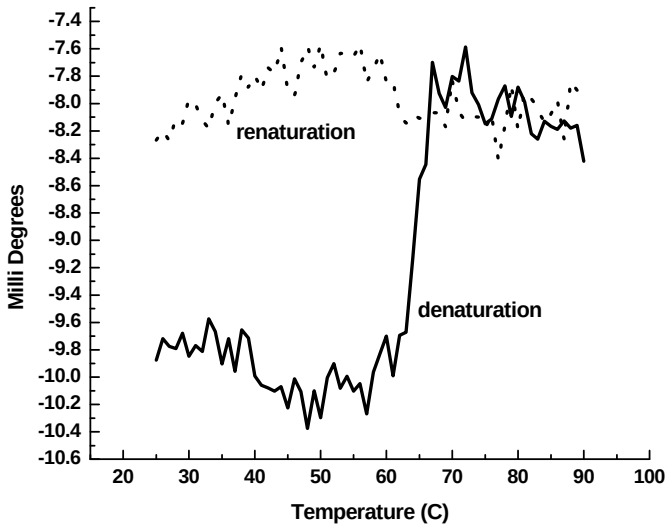
The fluorescence intensity was found to decrease with increasing temperature without any shift in λ_{max} for both these lectins. The decrease in the fluorescence intensity could be due to deactivation of the excited singlet state. Since, the increasing temperature has no effect on the emission maximum of the protein, the polarity of tryptophan environment can be assumed to remain same.


Ellipticity
Wavelength
25-50
55
60
65
70-90

(A)


Ellipticity
Wavelength
2
4
7
12
10
6,8

(B)



(C)

Figure 6.2.

(A) Far-UV CD spectra of SGA (1.8µM) incubated in buffers of pH from 2-12 for 16 h at 30 °C.

(B) Effect of temperature on the secondary structure of SGA (1.8µM): Far-UV CD scans of SGA at temperatures ranging from 25-90 °C.

(C) Denaturation & Renaturation of ATL, ellipticity at 225 nm: Temperature of the protein was increased at the rate of 1°C/min from 25-90 °C.

In CD spectra the negative ellipticity of the lectins at 226 nm remains same while positive ellipticity at 200-205 nm decreases gradually with increase in the temperature till 50 °C indicating increase in the random coil element. At 55 °C and above the negative ellipticity at 226 nm also decreases indicating the subsequent unfolding of the protein (Fig 6.2A). From the denaturation and renaturation experiments carried out, the Tm for unfolding of SGA and ATL was determined as 55°C and 60°C. Hemagluttination of SGA and ATL at 55 °C did not show any change (Shangary *et al*, 1995, Dhuna *et al.,* 2005). The partial loss of secondary structure at

these temperatures is not destroying activity fully. However, at temperatures above 55 ºC far-UV CD shows distortion in secondary structure as well as loss of hemagluttination activity.

6.5.1.3 pH Stability (Urea/ Gdn-HCl)

A 2-3 nm blue shift in λ_{max} to 347 nm at pH 3 is suggestive of decrease in polarity of the Trp residues due to protonation. At extreme alkaline pH (10-12) a 2-3 nm red shift in λ_{max} to 353 nm was observed indicating increased surface-exposed Trps.

Both these lectins retain hemagluttination activity at both acidic (pH 2-4) and alkaline pH (10-12) (Shangary *et al*, 1995; Dhuna *et al*., 2005**).** Thus, it may be reasonable to assume that the apparent structural loss at acidic pH observed in the CD spectra can be attributed to some disruption away from the sugar binding site not affecting its integrity.

The far-UV CD scans of SGA at various pH is shown in Fig 6.2B. These far-UV CD studies of both the lectins suggest that these lectins exhibit a compact structure at alkaline pH (10-12). The secondary structures of ATL as well as SGA were drastically affected at acidic pH as evident from the change in ellipticity at 210 nm in far-UV CD spectra.

6.5.1.4 Stability against denaturants (Urea/ Gdn-HCl)

Characterization of the structural and conformational changes occurring in both these araceous lectins in the presence of Gdn-HCl was carried out to probe the changes occurring in the tryptophan environment. A red shift in λ_{max} of fluorescence of the protein from 350 nm to 356 nm with increasing concentration of Gdn-HCl was observed in the spectrum indicating increased polarity of tryptophan environment. A 50% loss of hemagluttination activity of ATL has been reported in presence of 3M Gdn-HCl concentration (Dhuna *et al,* 2005). A simultaneous drop in

the fluorescence intensity observed could be due to structural changes in protein with increasing concentration of Gdn-HCl (1-6M) leading to solvent shielding of Trps.

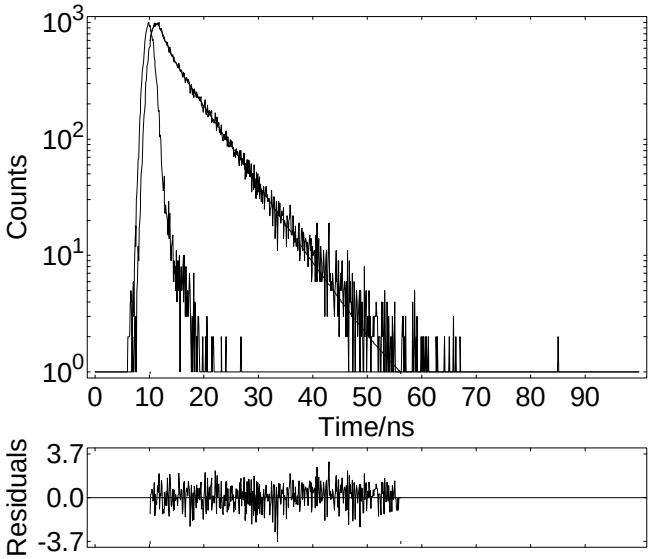
The change in negative ellipticity caused by Gdn-HCl in the case of SGA is depicted in Fig 6.1B. At 6 M concentration of the denaturant major distortion in the secondary structure along with a corresponding drop in the ellipticity at 225 nm was observed. In the case of ATL also Gdn-HCl mediated denaturation profiles are similar. 3M Gdn-HCl caused only partial unfolding of both the lectins, concomitant with loss of hemagluttination activity indicating disruption of structural and functional integrity.

6.5.2 Lifetime measurements of fluorescent decay

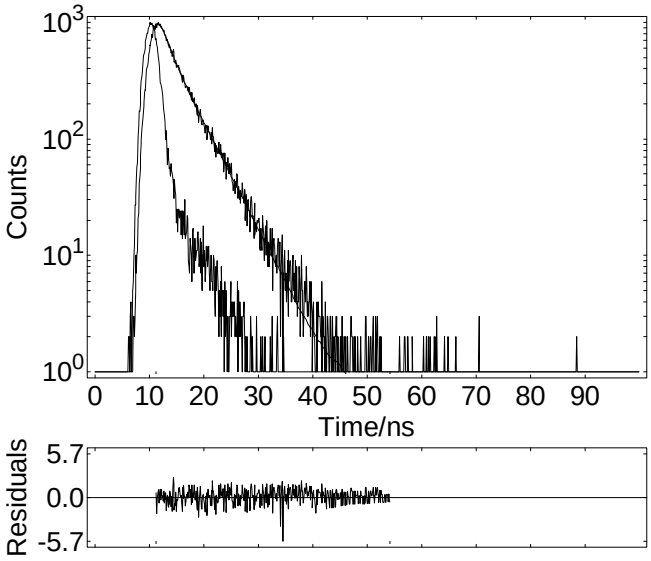
The fluorescent decay of the tryptophan residues on a nanosecond time scale for both the lectins in native, denatured and quenched states, obtained from time resolved measurements are presented in Fig. 6.3. A biexponential curve could be fitted to the time resolved fluorescence profiles of both the lectins ($\chi 2 = 1.02$) giving values of shorter ($\tau_{1)}$ and longer (τ_2) lifetimes for SGA (1.25 and 6.48 ns) which are similar for ATL (Table 6.1). The relative contributions of these components to the overall fluorescence were ~20% for shorter and ~80% for longer component. Thus majority of the Trp residues take longer time to decay. The values of individual tryptophan lifetimes and their averages for SGA are tabulated in Table 6.1. The corresponding values for urea denatured SGA were 1.26 and 4.23 ns. The values for ATL under the same conditions were similar. On denaturation the population components shifted to ~30% for shorter and 70% for the longer lifetime. The observed change in values of relative amplitudes (α_1 and α_2) could be due to change in the protein conformation after subjecting to urea denaturation.

The decay times τ_1 and τ_2 for Gdn-HCl-denatured SGA were 1.52 and 3.50 ns with populations 66 % and 34 %. Denatured ATL also showed two lifetimes, which changed like SGA. For both these lectins a marginal increase in the shorter lifetime (1.25 to 1.52 ns for SGA) and a pronounced decrease in the longer lifetime (6.48 to 3.50 ns for SGA) were observed. The remarkable change in the percentage contributions on denaturation is suggestive of conformational changes affecting the local Trp environment.

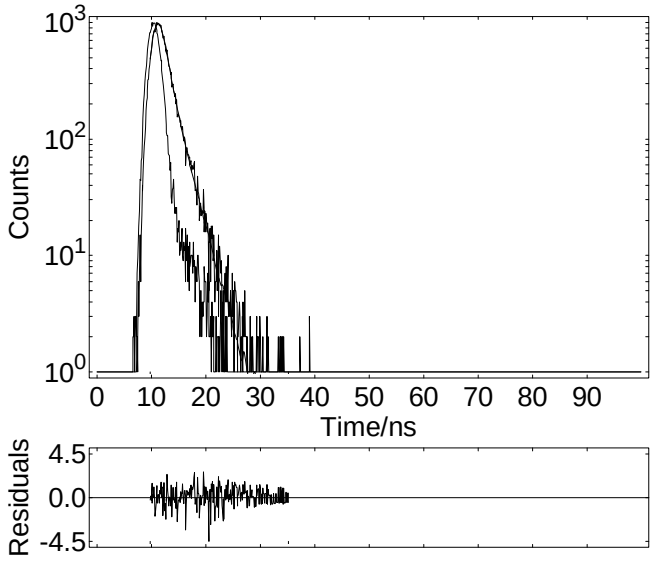
Analysis of the decay profiles and determination of corresponding fluorescent lifetimes obtained after quenching with the neutral quenchers acrylamide and succinimide as well as the ionic quenchers I^- and Cs^+ was also carried out (Fig.6.3C). In case of acrylamide quenching for SGA it is found that the values of τ_1 and τ_2 were drastically reduced to 0.71 and 2.03 ns and more or less same changes have been observed for ATL also. The life of short time decay decreased from 1.25 to 0.71 ns (that is by 17.6%) whereas longer lifetime decreased from 6.48 to 2.03 ns (that is by 31.3 %). Quenching with the cation Cs^+ resulted in decrease in longer lifetime from 6.48 to 4.99 ns and increases in the shorter lifetime 1.25 to 1.58 ns when compared to native SGA. Thus, the effect was not as pronounced as observed with neutral and anionic quenchers.


Counts
10^3
10^2
10^1
10^0
0 10 20 30 40 50 60 70 80 90
Time/ns
Residuals
3.7
0.0
-3.7

(A)


Counts
10^3
10^2
10^1
10^0
0 10 20 30 40 50 60 70 80 90
Time/ns
Residuals
5.7
0.0
-5.7

(B)



(C)

Figure 6.3. Time resolved fluorescence decay profile. The solid lines correspond to the nonlinear least square fit of the exponential data. The lower panel represents the residual.

(A) Native SGA.

(B) SGA in the presence of 6 M Gdn-HCl.

(C) SGA quenched with 0.5 M Acrylamide.

Table 6.1: Lifetimes of fluorescent decay of SGA and ATL under different conditions and the corresponding pre-exponential factors along with the calculated average lifetimes.

Sample Description	α_1	τ_1	α_2	τ_2	τ	$<\tau>$	χ^2
Native SGA	0.031	6.48 (± 0.086)	0.050	1.25 (± 0.055)	3.25	5.23	1.003
Native ATL	0.039	6.36 (± 0.19)	0.046	1.42 (± 0.095)	3.69	5.32	1.003
Denatured SGA (8M Urea)	0.039	1.26 (± 0.15)	0.035	4.26 (± 0.17)	2.68	3.65	1.003
Denatured ATL (8M Urea)	0.041	1.85 (± 0.17)	0.037	4.79 (± 0.14)	3.24	3.90	1.001
Denatured SGA (6M Gdn-HCl)	0.045	1.52 (± 0.15)	0.038	3.50 (± 0.065)	2.42	2.82	1.001
Denatured ATL (6M Gdn-HCl)	0.046	1.27 (± 0.17)	0.043	3.38 (± 0.11)	2.23	2.77	1.004
SGA + 0.5M Acrylamide	0.013	2.03 (± 0.057)	0.132	0.71 (± 0.093)	0.83	0.99	1.002
ATL + 0.5M Acrylamide	0.134	0.82 (± 0.055)	0.013	2.15 (± 0.25)	0.93	1.09	1.001

SGA +0.5M Succinimide	0.133	0.666 (± 0.042)	0.019	2.61 (± 0.144)	0.90	1.36	1.003
SGA + 0.5M KI	0.142	0.757 (± 0.043)	0.005	2.70 (± 0.36)	0.82	0.97	1.005
SGA + 0.5M CsCl	0.049	1.57 (± 0.11)	0.027	4.99 (± 0.1)	2.78	3.74	1.005

For both these lectins the average lifetimes were calculated using two different approaches (τ and ‹τ›) from the data obtained from the fluorescent decay under various conditions and presented in Table 6.1., using the following equations (3, 4):

$$\boldsymbol{\tau = \sum_i \alpha_i \tau_i / \sum_i \alpha_i} \qquad \textbf{(3)}$$

$$\boldsymbol{\langle\tau\rangle = \sum_i \alpha_i \tau^2_i / \sum_i \alpha_i \tau_i} \qquad \textbf{(4)}$$

Where i= 1, 2 …

From Table 6.1 it can be seen that for native SGA the average lifetimes τ and ‹τ› are 3.25 and 5.23 ns, respectively. The average lifetimes τ and ‹τ› of these proteins are affected after denaturation as well as on treatment with acrylamide, succinimide and KI, while they are not much affected by Cs^+.

6.5.3 Binding of ANS & Adenine to these lectins: A fluorimetric analysis

Hydrophobic dye binding

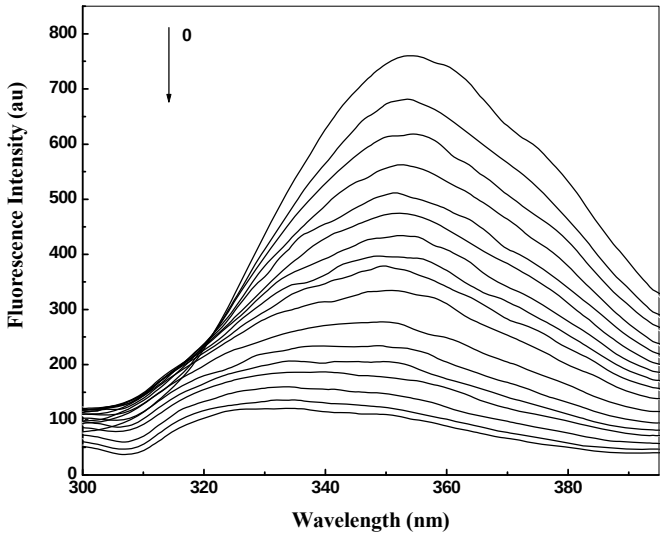
8-anilino-1-napthalene sulfonic acid (ANS) is a charged hydrophobic dye that binds to hydrophobic clusters in proteins and hence is largely employed for characterizing and detecting partially unfolded states in proteins **(**Ali *et al.*, 1999; Gasymov & Glasgow, 2007). Enhancement of the intrinsic fluorescence of ANS upon binding to proteins at acidic and alkaline pH was studied. Although the secondary structure at pH 2.0 gets distorted as evident from the far-UV CD spectrum, this unfolded structure does not result in the exposure of any hydrophobic patches which could be accessed by the hydrophobic dye. Even the compact structure at alkaline pH did not exhibit fluorescence enhancement in presence of the hydrophobic dye. Hence this compact structure at extreme alkaline pH cannot be categorized as a molten globule like intermediate.

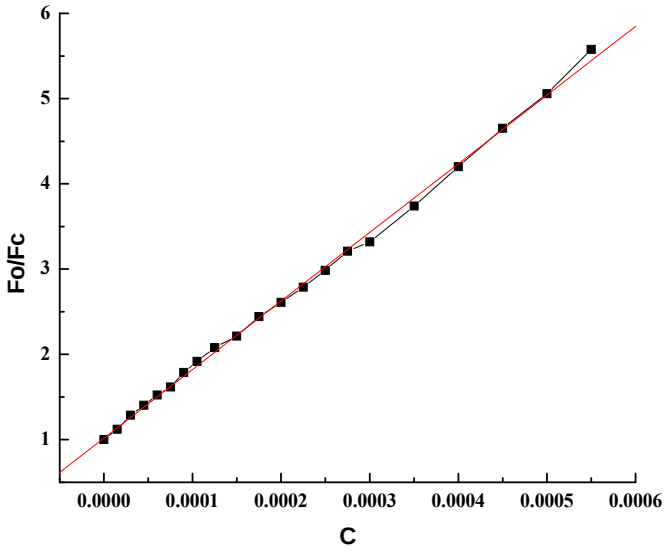
Adenine binding

Titrations of SGA and ATL were carried out with Adenine and the decrease in the intensity was measured (Fig. 6.5(A)). Adenine-binding resulted in considerable quenching of the fluorescence intensity (80-90%) in both the lectins. The estimated binding constant K_a was 6.61 x 10^3 M^{-1} for native SGA and 1.0 x 10^5 M^{-1} for ATL.

Adenine and adenine-derived plant growth regulators are hydrophobic ligands known to bind lectin in sites different from the carbohydrate binding site and this binding could be significant for the storage of plant growth regulators (Gegg *et al.*, 1992). The binding affinity is reported to be of the order of 10^5-10^6 and hence may be of physiological importance in these plants. Both Lima bean lectin from *Phaseolus lunatus* (Roberts & Goldstein, 1983) (K_a= 8.3 x 10^4 M^{-1}) and *Dolichos biflorus* lectin (Gegg *et al.*, 1992) (K_a= 7.31 x 10^5 M^{-1}) have high affinity

for adenine. Winged bean agglutinin also exhibits adenine-quenching without any shift in its λ_{max} at 330nm, K_a =1.5 x 10^4 M^{-1} (Puri & Surolia, 1994**).** Thus, SGA and ATL behave like legume lectins in binding adenine. As far as we know, this is the first report of any araceae lectin binding adenine.





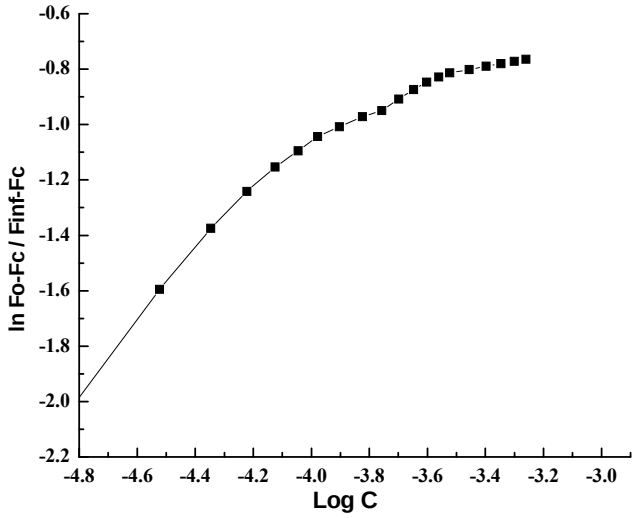


Figure 6.5 Determination of association constant for the binding of adenine (10 mM stock in methanol) to SGA

A. Fluorescence quenching of SGA on addition of aliquots of adenine to the lectin solution.

B. The plot of ($F_0/\Delta F$) vs $[C]^{-1}$ for adenine.

C. The plot of log ($\Delta F/F_C$-F_∞) vs log(C) for adenine.

6.5.4 Solute quenching studies: Native & Denatured lectins

Quenching experiments on native and denatured SGA and ATL were carried out using two neutral quenchers of different sizes (acrylamide and succinimide) and two ionic quenchers (iodide and cesium ion). Acrylamide being a small molecule can penetrate into protein interior and quench the fluorescence of even buried Trps, while succinimide affects intensities of mainly partially exposed tryptophans (Eftink & Ghiron, 1977).The ionic quenchers such as iodide and cesium are also larger in size and cannot penetrate protein matrix and thus can quench fluorescence of only surface exposed tryptophan residues located in the vicinity of positively and negatively charged residues, respectively.

The intrinsic florescence of SGA and ATL lectins were quenched by all quenchers, without any change in the λ_{max} of 350nm. The percentage quenching in case of native SGA was 80%, 65.2%, 72.5% and 29.5% for acrylamide, succinimide, iodide, and cesium, respectively. For ATL the percentage of quenching values were approximately the same except in case of succinimide and cesium ion, the values were 72.0 % and 15.4%. A slightly higher percentage of quenching obtained with succinimide is indicative of more easily accessible and non-polar tryptophans.

The change in the response to denatured lectins was negligible with most of the quenchers, presumably due to already exposed Trp residues. However, iodide quenching showed a decrease (from 72 to 65%) at 6M Gdn-HCl whereas cesium quenching showed an increase (from 29 to 34 %) at 8 M urea accompanied with a spectral shift of 4-5 nm. Already the CD analysis has shown that the lectins get denatured at these concentrations of the denaturant. A possible explanation for the above results can be given as redistribution of charges in the vicinity of Trp residues due to the change in conformation.

Analysis of the Steady state quenching data

Quenching data thus obtained was analyzed by the Stern-Volmer equation (1) (Lis *et al.*, 1966**)** as well as modified Stern-Volmer Equation (2) (Lehrer & Leavis, 1978; Eftink & Ghiron, 1981; Townsend *et al.*, 1986) to quantify the extent of quenching.

$$\mathbf{Fo/Fc = 1 + Ksv\ [Q]} \qquad (1)$$

$$\mathbf{Fo/\ (Fo\text{-}Fc) = } f\mathbf{a}^{-1} \mathbf{+ (Ka}f\mathbf{a)}^{-1}\ \mathbf{[Q]}^{-1} \qquad (2)$$

Where F_o and Fc are the respective fluorescence intensities, corrected for dilution, in the absence and presence of quencher, respectively, [Q] is the resultant quencher concentration, *K*sv is the Stern-Volmer quenching constant of the lectin for given quencher, *f*a refers to the fraction of the total fluorescence that is accessible to the quencher and Ka is the corresponding quenching constant. For both the lectins, Stern-Volmer plots for quenching with the various quenchers are shown in Fig 6.4. Slopes of Stern-Volmer plots yield *K*sv values (Eq.1), whereas the slopes of modified Stern-Volmer plots give $(Ka f a)^{-1}$ and their ordinate give values of $1/f a$ (Eq.2).

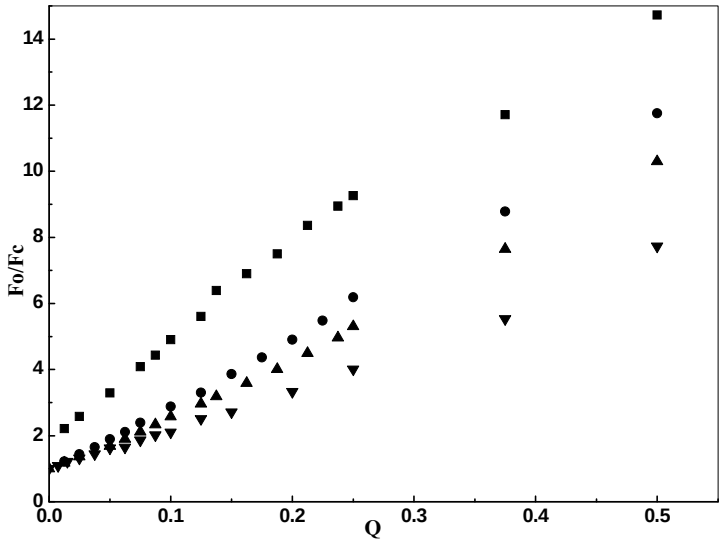
Quenching with Acrylamide

Quenching with acrylamide gave a linear stern-volmer plot in case of SGA, indicative of dynamic/collisional quenching of a homogenous population of tryptophans. The values of the bimolecular quenching constant (K_q) were calculated by dividing K_{sv} with τ_0 which is the average fluorescence lifetime in the absence of the quencher. The values were calculated from time resolved fluorescence data described below. The value of this constant (K_q) for native SGA and ATL is 5.36 x 10^9 M^{-1} s^{-1} and 5.51 x 10^9 M^{-1} s^{-1} respectively, has not changed much for denatured lectin (4.64 x 10^9 M^{-1} s^{-1} and 3.51 x 10^9 M^{-1} s^{-1}) (Table 6.3). Similar values were obtained for ATL also.

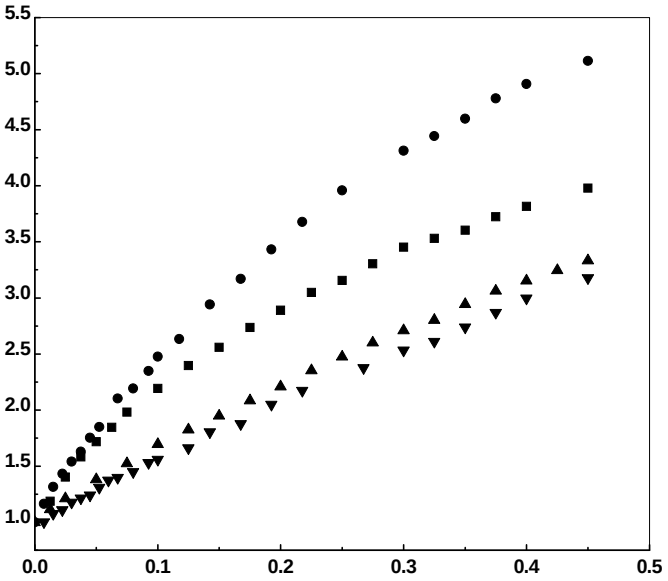
Quenching with Succinimide

The Stern-Volmer plot for succinimide quenching of SGA gave a downward curve, whereas it was straight-line for ATL. The straight-line is indication of dynamic quenching. The values of K_{sv1} and K_{sv2} calculated for SGA were 6.87 and 2.78 M^{-1}. These values did not change for denatured SGA. The high value of K_{sv1} compared to K_{sv2} suggests a differentially exposed Trp population that is easily quenched. The values of bimolecular rate constants, K_{q1} and K_{q2} for

native SGA 2.0 x10^9 M^{-1} $s^{-1}$and 0.7 x10^9 M^{-1} s^{-1}, respectively, decreased slightly after denaturation. For ATL the values of K_{sv} and K_q were 9.87 M^{-1} and 2.68 x 10^9 M^{-1} s^{-1}, respectively. In the case of native SGA and ATL the quenching constant K_a calculated from modified SV plots was lesser for succinimide compared to acrylamide. The reason may be lesser quenching efficiency and lesser penetrating capacity of succinimide compared to acrylamide. As expected, the K_a values were lesser for denatured lectins.



(A)



(B)

<u>Figure 6.4.</u> Stern volmer plot for the quenching of native and 8 M urea denatured SGA and ATL (Native SGL --■--, Native ATL -●-, 8 M Urea denatured SGL --▲—ATL --▼--)

(A) Acrylamide

(B) KI.

Quenching with Iodide

For Iodide quenching, the Stern-Volmer plot curves downwards for both native and denatured SGA and native ATL, whereas it is a straight line for urea denatured ATL. The values of K_{sv1} and K_{sv2} are 11.08 M^{-1} and 4.96 M^{-1} for native SGA & 15.44 M^{-1} and 10.47 M^{-1} for native ATL. The higher values of K_{sv1} suggests an easily accessible and rapidly quenching population of tryptophans while a lower value is indicative of a late and slower quenching Trp population. For both lectins in denatured state the Stern-Volmer plot is a straight line and the value of K_{sv1} is 5.14 M^{-1} and 5.29M^{-1} respectively for SGA and ATL. The estimated K_a values were 21.75 and 2.0 for SGA and 28.64 and 1.48 for ATL, when quenched by KI and CsCl, respectively.

Table 6.2. The percentage quenching obtained with different quenchers in the native and 8 M urea denatured SGA and ATL. (Calculated from raw data)

Quencher (Final concentration)	SGA Quenching % (Native)	SGA Quenching% (Denatured)	ATL Quenching % (Native)	ATL Quenching % (Denatured)
Acrylamide (0.25M)	80 (± 1.5)	80.21 (± 1.3)	83.03 (± 1.9)	80.39 (± 1.8)
Succinimide (0.3M)	65.2 (± 1.2)	61.64 (± 1.6)	72.05 (± 2.0)	NA
KI (0.4M)	72.49 (± 2.0)	65.18 (± 1.7)	74.74 (± 1.2)	59.66 (± 1.5)
CsCl (0.33M)	29.5 (± 1.1)	34.34 (± 1.1)	15.38 (± 0.9)	15.7 (± 1.0)

Table 6.3: Summary of comparative parameters for SGA and ATL obtained from Stern-Volmer and modified Stern-Volmer analysis of the intrinsic fluorescence quenching with different quenchers.

Sample Description	K_{sv1} (M^{-1})	k_{q1} (x 10^9 $M^{-1}s^{-1}$)	K_{sv2} (M^{-1})	k_{q2} (x 10^9 $M^{-1}s^{-1}$)	f_a	K_a (M^{-1})
Acrylamide						
Native SGA	17.42 (± 0.12)	5.36 (± 0.08)	-	-	0.98	19.08 (± 0.51)
Native ATL	20.32 (± 0.31)	5.51 (± 0.21)	-	-	1	16.55 (± 0.74)
Denatured SGA	15.07 (± 0.14)	4.64 (± 0.10)	-	-	1	13.19 (± 0.17)
Denatured ATL	12.94 (± 0.27)	3.51 (± 0.21)	-	-	0.98	11.36 (± 0.29)
Succinimide						
Native SGA	6.5 (± 0.15)	2 (± 0.12)	2.25 (± 0.17)	0.7 (± 0.05)	0.77	12.16 (± 0.44)
Native ATL	9.87 (± 0.07)	2.68 (± 0.05)	-	-	0.79	15.83 (± 0.44)
Denatured SGA	6.27 (± 0.20)	1.93 (± 0.15)	2.41 (± 0.16)	0.65 (± 0.05)	0.81	9.7 (± 0.31)
KI						
Native SGA	11.08	3.4	4.96	1.53	0.81	21.75

	(± 0.61)	(± 0.57)	(± 0.19)			(± 0.4)
Native ATL	15.44	4.19	10.47	2.84	0 .79	28.64
	(± 0.51)	(± 0.41)	(± 0.18)			(± 0.54)
Denatured SGA	5.14	1.58	-	-	0.85	12.05
	(± 0.01)	(± 0.01)				(± 0.27)
Denatured ATL	5.29	1.59	-	-	0.83	4.28
	(± 0.09)	(± 0.05)				(± 0.47)
CsCl						
Native SGA	1.44	0.44	-	-	0.83	2
	(± 0.03)	(± 0.02)				(± 0.12)
Native ATL	0.99	0.27	-	-	0.72	1.48
	(± 0.02)	(± 0.01)				(± 0.16)
Denatured SGA	1.59	0.9	0.77	0.24	0.65	3.07
	(± 0.04)	(± 0.02)	(± 0.05)			(± 0.17)

Bimolecular quenching constants (K_{q1} and K_{q2})

The bimolecular quenching constants K_{q1} and K_{q2} for both the lectins quenched by various quenchers are presented in Table 6.3. For both the lectins the value of K_{q1} for KI quenching is found to decrease on denaturing the protein. In case of ATL, quenching with acrylamide showed a decreasing trend in the values of these constants. The values of K_{q1} and K_{q2} remained unaltered for other quenchers. Since K_q values can be associated with collisional frequency, a decrease in collisional frequency with denaturation in some cases is only non-specific.

6.6 Conclusions

In conclusion, this is the first report of structural investigations on these araceaous lectins to probe the tryptophan environment and secondary structure of the members of this monocot family. Both SGA and ATL are tetrameric lectins, containing four trp residues that show fluorescence maxima at 350 nm corresponding to tryptophan residues exposed in polar environment. For both these lectins the far-UV CD spectra indicate a predominant β-sheet structure and type-III β-turns. Investigations into the secondary structure after various modes of denaturation reveal that these lectins possess structure with no hydrophobic core, labile at high acidic pH and stable at extreme alkaline pH. Adenine-binding, a feature of legume lectins has been observed in members of this family for the first time. This property is important for binding plant growth regulators. Complete accessibility of the tryptophans to the neutral quencher acrylamide confirmed the surface exposure of the residues. The fluorescent lifetimes were significantly altered after quenching with acrylamide. The predominantly positive environment of the tryptophans was confirmed by quenching with iodide ion. Average lifetimes were not significantly affected with Cs^+ quenching.

Chapter 7

Comparisons of plant lectin characteristics studied and Conclusions

7.1 Summary

This chapter outlines the comparative analysis of the structures, the sugar-binding of lectins, as well as the biophysical characteristics of lectins that have been studied in this thesis, belonging to three structurally different plant lectin families.

Comparison of the sugar binding sites of Galactose/*N*-acetyl Galactosamine specific lectin TDSL and EiSL was done with lectins having similar sugar specificity. The structural features of the active site of *T. dioica* lectin was compared with binding sites of galactose specific lectins, a type-II RIP Abrin-a and a C-type animal lectin CEL-III from *Cucumaria echinata*. The active site geometry and residues involves in sugar binding for EiSL have been compared with that of Jacalin, mouse galectin-9, a tunicate lectin DCL-I from *Didemnum candidum* and C-type lectins.

The three-dimensional structure of *E. indica* lectin was similar to that of lectins from *E. cristagalli* and *E. corallodendron* in terms of the N-linked heptasacchride, conserved water molecules and residues in the sugar binding site. Minor structural differences have been highlighted and compared to show the higher stability of this lectin as compared to EcorL.

We have for the first time carried out structural investigations on the two araceaous lectins from *S. guttatum* and *A. tortuosum,* to probe the tryptophan environment and secondary structure of the members of this monocot family. For both these lectins the far-UV CD spectra indicate a predominantly β sheet structure and type-III β-turns. Adenine-binding, a feature of legume lectins has been observed in members of this family for the first time. The two lectins belonging to araceae and differing from other mannose specific lectins show similarity with

respect to their tryptophan environments as well as structural stability with respect to temperature, pH as well as presence of chemical denaturants.

7.2 Structural comparison of EiSL with the lectins from the *Erythrina* genus

We have determined the structure of *E.indica* seed lectin (EiSL), a member of the well conserved legume family and also belonging to the group of *Erythrina* lectins. Two other lectins that have been well studied structurally include the lectins from *E.corallodendron* (EcorL) and another from *E.cristalgalli* (ECL) and these lectins have been crystallized in the native state and also in complex with various sugars and, their recombinant variety has also been well characterized structurally. It is seen that the tertiary and the quaternary structures of the native lectin are rarely affected by glycosylation. Buried surface area, shape complementary, and interaction energy are considered to be key parameters in the evaluation of protein–protein interactions. Both *E.corallodendron* lectin and *E.cristagalli* lectin have been crystallized in the native state, in complex with various monosaccharide and oligosaccharides, as well as structures of the recombinant firms reported (Shaanan *et al.*, 1991; Elgavish & Shaanan, 1998; Svensson *et al.*, 2002; Turton *et al.*, 2004). A comparative study of the structures of the *E.indica* lectin with the lectins from *E.corallodendron* and *E.cristagalli* is presented below.

7.2.1 Primary structure comparison

The structure of EiSL reported here is compared with the already reported structures of *E.corallodendron* lectin (PDB code: 1FYU) and *E.cristagalli* lectin (PDB code: 1UZY). Sequences of both the A and B chains of all these lectins have been aligned using the program ClustalW (1.82). The residues in EiSL are based on the electron density. The closeness of the sequences is reflected in the structures as well. The lectin from *E.indica* differs from EcorL and ECL in a few aspects. These have been mentioned below.

```
EiSL            VETISFSFSEFEAGNDNLTLEGAALITQSGVLQLTKINANGMPAWDSTGRTLYTKPVHIW 60
EcorL           VETISFSFSEFEPGNDNLTLQGAALITQSGVLQLTKINQNGMPAWDSTGRTLYAKPVHIW 60
ECL             VETISFSFSEFEPGNNDLTLQGAAIITQSGVLQLTKINQNGMPAWDSTGRTLYTKPVHIW 60
                ************.**::***:***:************* **************:******

EiSL            DSTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQGYGYLGVFNNSDNDNS 120
EcorL           DMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQGYGYLGIFNNSKQDNS 120
ECL             DMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQGYGYLGVFNNSKQDNS 120
                * ***********************************************:****.:***

EiSL            YQTLAVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQVANVVIKYDASSKILH 180
EcorL           YQTLGVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQVANVVIKYDASSKILH 180
ECL             YQTLAVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQVANVVIKYDASSKILL 180
                ****.**************************************************

EiSL            AVLVYPSDGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAETHDVYSWSFQASLP-- 238
EcorL           AVLVYPSSGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAETHDVYSWSFQASLPET 240
ECL             AVLVYPSSGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAETHDVYSWSFHASLPET 240
                *******.*******************************************:****


EiSL            -------------EVETISFSFSEFEAGNDNLTLQGAALITQSGVLQLTKINANGMPAW 284
EcorL           NDAVIPTSNHNTFAIVETISFSFSEFEPGNDNLTLQGAALITQSGVLQLTKINQNGMPAW 300
ECL             ND------------VETISFSFSEFEPGNNDLTLQGAAIITQSGVLQLTKINQNGMPAW 287
                              ************.**::*******:************* ******

EiSL            DSTGRTLYTKPVHIWDSTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQG 344
EcorL           DSTGRTLYAKPVHIWDMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQG 360
ECL             DSTGRTLYTKPVHIWDMTTGTVASFETRFSFSIEQPYTRPLPADGLVFFMGPTKSKPAQG 347
                ********:******* ****************************************

EiSL            YGYLGVFNNSDNDNSYQTLAVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQV 404
EcorL           YGYLGIFNNSKQDNSYQTLGVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQV 420
ECL             YGYLGVFNNSKQDNSYQTLAVEFDTFSNPWDPPQVPHIGIDVNSIRSIKTQPFQLDNGQV 407
                *****:****.:*******.****************************************

EiSL            ANVVIKYDASSKILHAVLVYPSDGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAET 464
EcorL           ANVVIKYDASSKILHAVLVYPSSGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAET 480
ECL             ANVVIKYDASSKILLAVLVYPSSGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAET 467
                ************** *******.************************************

EiSL            HDVYSWSFQASLP----------------- 477
EcorL           HDVYSWSFQASLPETNDAVIPTSNHNTFAI 510
ECL             HDVYSWSFHASLPETND------------- 484


                ********:****
```

Fig.7.1 Alignment of the amino acid sequences of EcorL (PDB Code: 1FYU) and ECL (PDB Code: 1UZY) with the sequence of EiSL as obtained from the refined structure. This has been done using the program CLUSTALW-1.83 (**Thompson *et al.*, 1994**). **"*"** denotes that the residues in that column are identical, **":"** conserved substitutions are labeled in Red, **"."** semi-conserved substitutions are in Blue.

7.2.1.1 Non-canonical mode of dimerization

The mode of formation of dimers in EcorL is reported to be different from most legume lectins. This same pattern of non-canonical dimerization has been observed in ECL from this genus, and now in EiSL also. The structure of *E.indica* lectin resembles the EcorL as well as ECL structures and thus differs from other legume lectin structures in not forming the canonical lectin dimer, because the heptasachharide bound in the sugar binding site (Asn113) prevents the two monomers from associating in a handshake mode.

7.2.1.2 Glycosylation site

In both EcorL and ECL, four glycosylation sites have been predicted at locations A13, A117, B13 and B117. Similarly in the structure of EiSL also we found N- linked glycosylation at all the four sites. There are differences in the N-linked glycosylated carbohydrate traced at four locations in our structure which differs from those reported for EcorL and ECL structures.

As in EcorL and ECL, EiSL also shows glycosylation at two sites in both the chains (17 and 113) and the characteristic glycosylation site motif is conserved (Asx-Ser/Thr) (Shaanan *et al.*, 1991; Turton *et al.*, 2004). In EcorL structure a heptasaccharide has been traced only at the glycosylation site B17, while in ECL only a hexasacchride could be fitted in the electron density. In the present structure we could model oligosaccharides at all the four sites. As described earlier, a trisachharide (NAG-Fuc-NAG) was located at A17 and A113; a hexasachharide at B17 and a heptasachhride as in EcorL was traced at B113.

Inter subunit interactions have been assumed to play a major role in the structural stability of lectins. The inter chain hydrogen bonds have a role in maintaining protein stability. The structural stability of EiSL has been studied by denaturation under equilibrium and kinetic

conditions suggesting that the lectin has a high kinetic stability as compared to other similar legume lectins such as EcorL, ConA, and SBA.

7.2.1.3 Conserved waters and Metal coordination bonds

The water molecules which form hydrogen bonds with the bound sugar (lactose), those situated at dimer interface and which also play role in stabilizing the dimeric interface, and coordinating with the Ca^{++} and Mn^{++} ions are conserved in EiSL as in both EcorL and ECL lectin structures. Both the Ca^{++} and Mn^{++} coordinate with the residues in the vicinity of these ions. In EiSL the distance between the two metal ions is 4.2 Å (A-chain) and 4.03 Å (B-chain) which is closer to that in EcorL (4.5Å) and ECL (4.22 Å) (Turton *et al.*, 2004). The hydrogen bonds listed in Table 5.7 of chapter 5, between the two subunits seen to be conserved in other two *Erythrina* lectins.

7.2.1.4 Amino acid substitutions

More than 90% of the residues in all the three lectins are invariant. Some residues are found to be conserved in two, but different in other.

Residues in EiSL that vary from those of EcorL and ECL:

Out of these substitutions the significant one appears to be 13 Alanine which is Proline in both EcorL and ECL and is part of a large loop connecting the two β-sheets. A residue 116 Lysine which is a basic residue in EiSL is replaced by Aspartate, an acidic one in EcorL and ECL structures and is adjacent to the glycosylation site 117 on both the chains. Residue 180 Histidine is a Leucine in ECL structure, while this residue adopts an alternate conformation in EiSL. Glutamine 234 in both EcorL and EiSL is a Histidine in ECL. The residues at all these

three positions are all substitutions of the hydrophobic residues by the similar type of residue. Other amino acid substitutions seem not important as they involve acidic residue (Asp or Glu) being replaced by amide residues (Asn or Gln).

7.3 Lectin- carbohydrate interaction

The criterion on the basis of which a protein is classified as a lectin is its carbohydrate (especially oligosaccharide) binding property (Lis & Sharon, 1998). Because of their carbohydrate binding specificity lectins mediate an array of biological processes and are involved in many cell recognition events. It is also known that lectins possess varying carbohydrate specificities, where on one hand some lectins bind to simple sugars (galactose, glucose, mannose etc.) while others bind to complex carbohydrates such as asialofetuin. As described earlier in chapter 1, lectins can be grouped into five families on the basis of their carbohydrate specificity. Galactose/*N*-acetylgalactosamine lectins form one important class and include lectins from diverse organisms. Lectins with varying structures and different sugar specificities are spread over in both plants and animals. Plant lectins have been more extensively studied than animal lectins because of their comparatively easy availability.

7.4 Galactose-specific plant lectins

The galactose/*N*-acetylgalactosamine specific lectins are reported from three of the seven plant families described previously. They include the legume lectins, the type-II RIPs and the Jacalin-related lectins (JRLs). Apart from plants some lectins from microorganisms and fungi and the animal lectins such as galectin also show specificity towards galactose and its derivatives.

7.4.1 Comparison of the binding sites of various Gal/GalNAc lectins

Thus the following comparison is based on the structures of various lectins belonging to diverse families that show specificity for either galactose or any of its derivatives and the basis of their specificity for galactose is analysed based on the available carbohydrate bound structures. This analysis will also bring out the information that could be important for designing or engineering proteins for desired carbohydrate specificity.

7.4.2. *Erythrina indica* seed lectin

The binding site in EiSL is similar to that found in other members of the genus such as the EcorL and ECL. The monomer in this lectin interacts in a handshake-mode, and the glycosylation site is at the monomer-monomer interface which in other lectins like the ConA is involved in forming canonical dimers (Shaanan *et al.*, 1991). In EiSL which belongs to the legume lectins, the carbohydrate recognition domain (CRD) is approximately 120-125 residues long and is part of a single domain. This type of CRD has also been observed in jacalin, galectins and the C-type lectins (Vijayan & Chandra, 1999). While in the RIP type-II family to which TDSL belongs, as well as monocot mannose binding lectins and cereal lectins, CRD is shorter (40-50 residues) and is present on more than one domains or subunits.

(a) **Binding sites and residues of EiSL and Jacalin**

In both EiSL and Jacalin, residues located only on four loops are involved in carbohydrate recognition (Fig 7.2 & 7.3). While in EiSL (legume lectins) these loops are named as the A, B, C and D; in Jacalin the nomenclature for these loops is BL1, BL2, BL3 and IL (Jeyaprakash *et al.*, 2003).

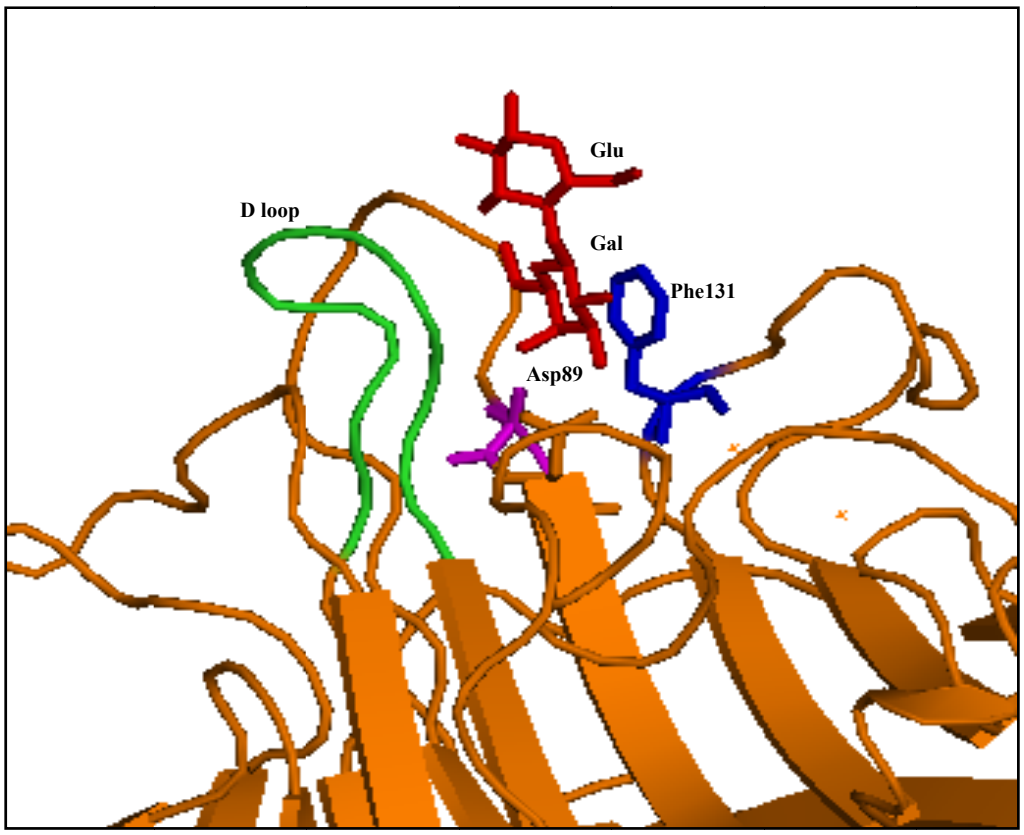


Fig 7.2 Residues involved in binding to Galactose at the active site in EiSL. The sugar and all residues are shown as stick models. The galactose and glucose of the lactose moiety are shown in red color. The residues Phe 131 and Asp89 are represented in blue and magenta colors respectively. The D loop of legume lectins is shown in green color.

- The A loop of EiSL corresponds to BL3 of Jacalin as both these loops contain the conserved Asp residue (Asp89 in EiSL and Asp125 in Jacalin) which forms hydrogen bonds with the sugar.
- The loop B is same as loop BL1 consisting of hydrophobic residues, and they form a hydrophobic pocket for accommodating the non-polar face of the sugar residue.

- The loop C, similar to the loop BL2 of Jacalin, has an aromatic residue (Phe131 in EiSL and Tyr78 in Jacalin) which stabilizes the sugar through stacking interactions. Both these loops are not involved in any manner which provides specificity for the lectin-carbohydrate interaction.
- Loop D of EiSL (Thr216–Glu224) and the loop IL in Jacalin both possess characteristic structure which is responsible for the specificity of the respective lectin for binding to the sugar residue which is galactose terminated sugar in EiSL or galactose or mannose terminating oligosaccharide in case of Jacalin. The length and residue composition of this loop decides its carbohydrate specificity (Raval *et al.*, 2004**)**.

The D loop as in legume lectins is located between Thr216–Glu224 in the lectins of genus *Erythrina* and has been called as the monosaccharide specificity loop (Sharma & Surolia, 1997). The residues in this loop form hydrogen bonds with the main-chain NH and CO and the carbohydrates.

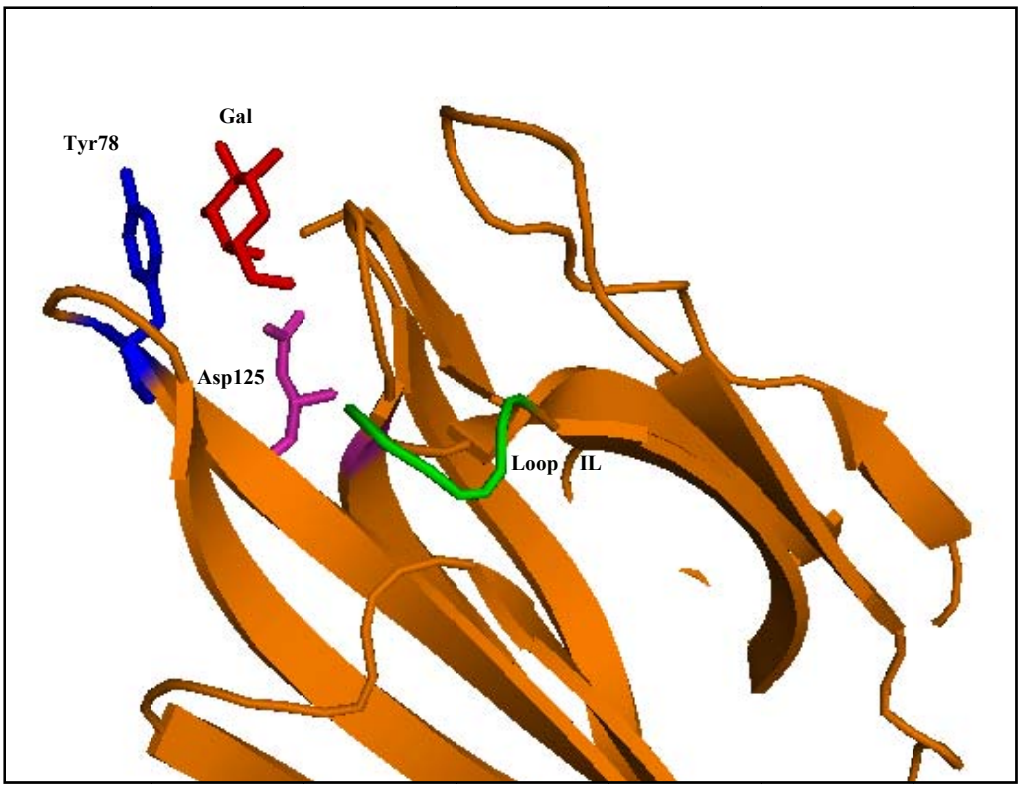


Fig 7.3 Residues involved in binding to Galactose at the active site in monomer of Jacalin (PDB code: 2UGW). The sugar and all residues are shown as stick models. The residues Tyr78 and Asp125 are similar to those found in two loops of EiSL represented in blue and magenta colors respectively. Loop IL is homologous to the D loop of legume lectins and is shown in green color.

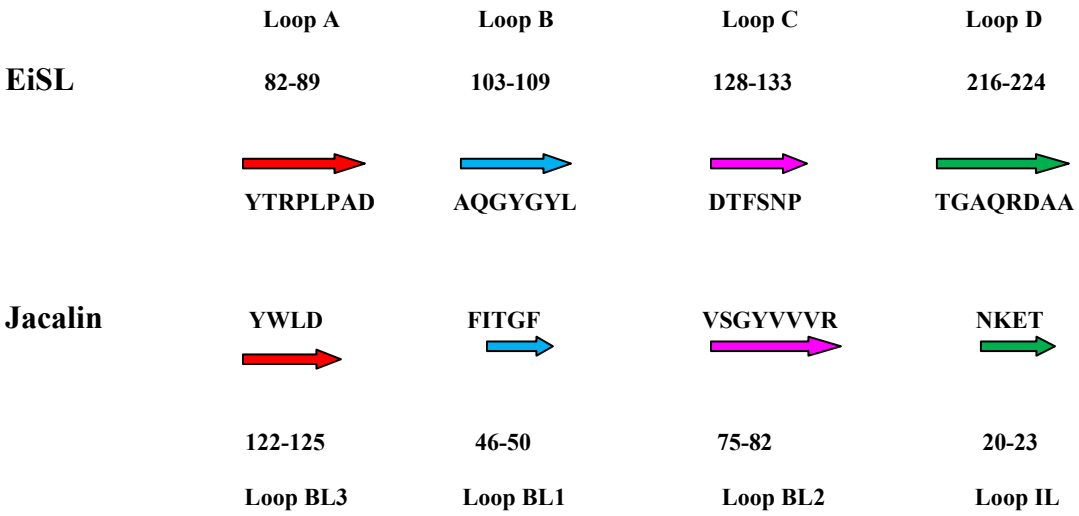


Fig 7.4 The various loops of EiSL which are involved in carbohydrate recognition, and the corresponding loops of Jacalin are shown. Highlighted residues in the loops represent the conserve amino-acid in both the families having specific function in carbohydrate binding.

(a) Binding sites and residues of EiSL and Galectins

Legume lectins and galectins share similar structural domains that are the legume lectin fold made up of three β-sheets. One similarity in the mode of carbohydrate binding between legume lectins and galectins is the water mediated hydrogen bonding. Water molecules located close to the combining site also play important role in meditating hydrogen bonds between the mono or oligosaccharides and the amino acid side chains in both these lectin families.

In mouse galectin-9 the carbohydrate binding site is formed by residues from three β-strands S4, S5, and S6 (Nagae *et al.*, 2006). But in EiSL as in other members of the legume lectin family four conserved loops play significant role in sugar binding at the active site. In mouse galectin-9, a histidine residue is common at the active site, His60, which with another residue Asn62 is involved in hydrogen bonding interactions (Fig 7.5). In EiSL, only Asp, Asn and Gly

residues are seen to be forming polar interactions, and histidine residue or any aromatic ring containing residue is not part of hydrogen binding, but provides stacking interactions.

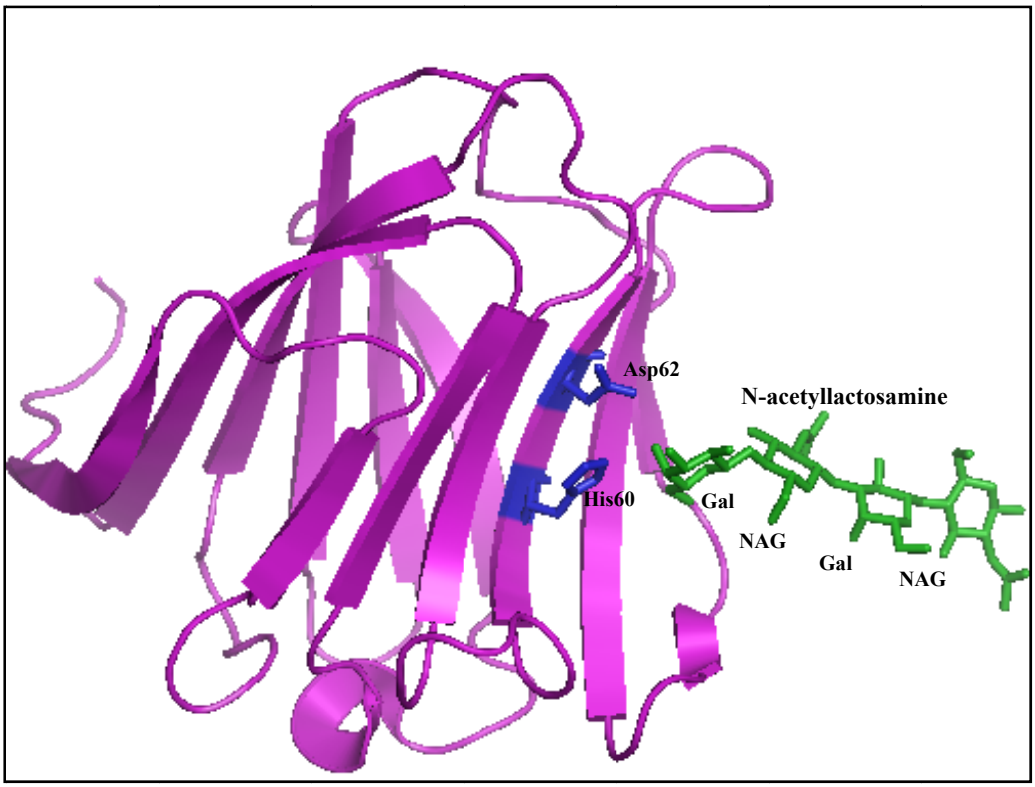


Fig 7.5 The residues at the active site of Mouse Galectin -9 (PDB code: 2D6O) and involved in binding to galactose terminated oligosaccahrides (*N*-acetyllactosamine) represented in green color. His 60 and Asp62 are shown as stick models in blue.

(a) EiSL and tunicate lectin

An animal lectin, DCL-I from the tunicate *Didemnum candidum* (Vasta *et al.,* 1986) has a combining site like PNA and other legume lectins. But DCL-I combining site is smaller as it cannot accommodate galactose derivatives with bulky groups at C-6 which is possible in case of EiSL, EcorL, PNA and other legume lectins.

(b) EiSL and C-type lectins

The legume lectins from Erythrina and C-type lectins both contain Ca^{++} and Mn^{++} in their structures. In EiSL and other legume lectins these metal ions are involved indirectly in forming hydrogen bonds near the combining site, while in the C-type lectins these ions are interacting directly with the sugar residue.

7.4.3 TDSL (a type-II RIPs) and C-type lectins

The lectins such as abrin-a, ricin and the lectin from *T.dioica* seeds (TDSL) all have been characterized as type-II RIPs. The C-type lectins are animal lectins require Ca^{++} for binding to carbohydrates. From *Cucumaria echinata*, CEL I-IV have been characterized (Hatakeyama *et al.*, 1994) and have been shown to possess structural similarity with the B-chains of ricin and abrin-a (Rutenber & Robertus, 1991; Tahirov *et al.*, 1995) the type-II RIPs, in spite of very less sequence similarity. Their mode of interaction with the galactosides is similar to that in ricin, but they differ in the requirement of divalent ion for interacting with sugars which is totally absent in TDSL or other type-II RIPs.

7.4.3.1 Residues involved in binding of sugar residues in TDSL and CEL-III

Gly26 is conserved in CEL-III and TDSL whereas Asp23 in the former is substituted by Arg in TDSL. The main chain carbonyl oxygen and amide nitrogen of these residues are involved in formation of hydrogen bonds with the 4-OH of the sugar residues. The two residues Asp39 and Asp43 important for sugar binding are absent in TDSL and other type-II RIPs such as Ricin. The residue Asp39 forms hydrogen bond with the 3-OH of the bound sugar while both these residues are involved in binding with the metal ion and forming coordination bonds mediated by water molecules that stabilize the carbohydrate binding(Uchida *et al.*, 2004). It may be noted that type-II RIPs lack metal ions also in their structures. The aromatic residue involved

in stacking with the hexose ring is Trp269 in CEL-III which corresponds to Trp236 in TDSL. Fig 7.6 shows the interaction of α-methyl galactose with Ca^{++} and the active site residues of CEL-III at three different locations.

In CEL-III binding of the sugar is achieved by hydrogen binding as well as coordination bonds with the metal ions (Kouzuma *et al.,* 2003), while in TDSL only hydrogen bonding fixes the galactose in the binding grove. But despite this difference in binding pattern, the residues involved in stabilizing the carbohydrate at the active site are similar in both these lectins from different families.

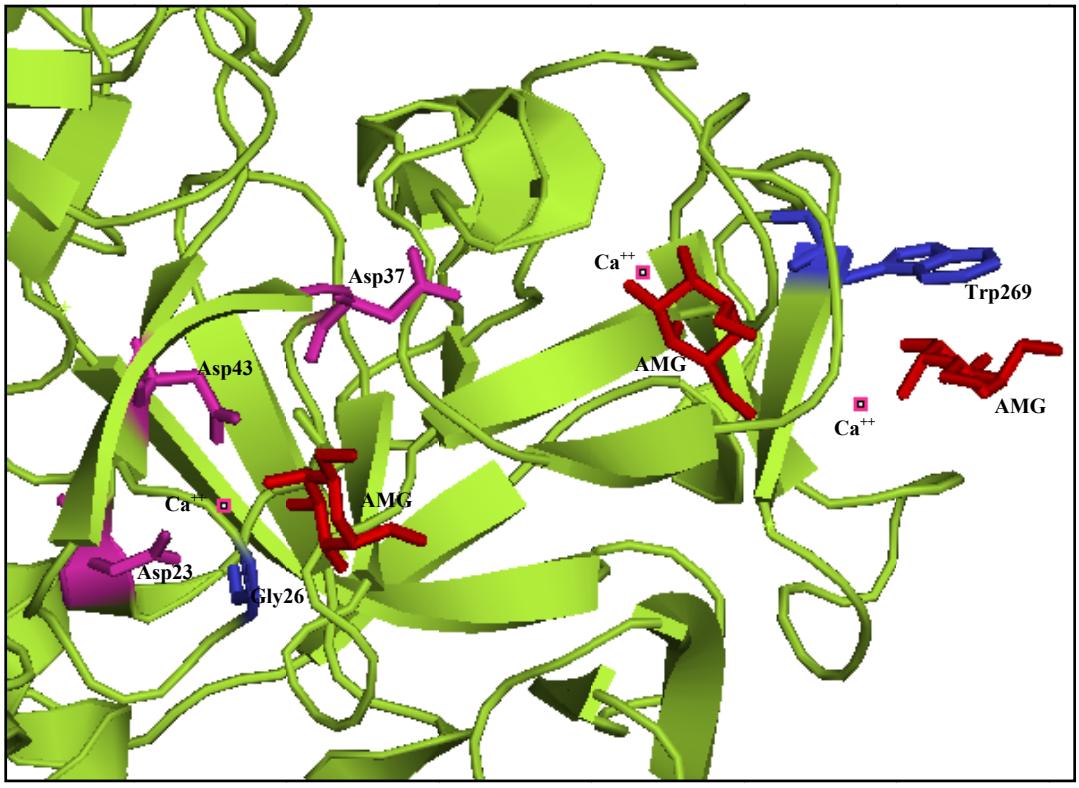


Fig. 7.6 Showing the Galactose binding site residues in CEL-III (PDB code: 2Z49). All the residues involved in sugar binding are shown as stick models. The sugar α-methyl galactose (AMG) is bound at three locations and is shown in red. The coordinating Ca^{++} ions are

shown in magenta at all three locations near the sugar binding site and participate in stabilizing the sugar at the active site. Asp23, Asp37 and Asp43 which are different from those of TDSL are represented in magenta. While residues Gly26 and Trp 269 that are are similar location in TDSL are colored in blue.

Table 7.1 lists the PDB structures of lectins from different families that have been taken for structural analysis for the comparison of the galactose binding site.

PDB ID	Description	Resolution
1AXO	*Erythrina corallodendron* lectin in complex with *N*-Acetylgalactosamine	1.9
1AXZ	*Erythrina corallodendron* lectin in complex with D-galactose	1.95
1FYU	Crystal structure of *Erythrina corallodendron* lectin in hexagonal crystal form	2.6
1SFY	Crystal structure of recombinant *Erythrina corallodandron* Lectin	2.55
1UGW	Crystal structure of Jacalin- Gal complex	1.7
1UHO	Crystal structure of jacalin- Me-alpha-GalNAc complex	2.8
1TOQ	Crystal structure of a galactose specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose	2.5
1ABR	Crystal structure of Abrin-a	2.14
2AAI	Crystallographic refinement of Ricin to 2.5 Å	2.50
1OQL	Mistletoe Lectin I from *Viscum album* complexed with galactose	3.0
1SZ6	Mistletoe lectin I from *Viscum album*. Crystal structure at 2.05 Å resolution	2.05
1PC8	Crystal Structure of a novel form of mistletoe lectin from Himalayan *Viscum album* L. at 3.8 Å resolution	3.8
3CAI	*Sambucus nigra* agglutinin II (SNA-I I)- tetragonal crystal form- complexed to galactose	1.55
3CA3	Crystal structure of *Sambucus Nigra* Agglutinin II (SNA-II)-tetragonal crystal form- complexed to N-Acetylgalactosamine	1.55
3CA5	Crystal structure of *Sambucus nigra* agglutinin II (SNA-	1.55

	II)-tetragonal crystal form- complexed to alpha1 methylgalactose	
3C9Z	*Sambucus nigra* agglutinin II (SNA-II), tetragonal crystal form	1.35
1HWN	Ebulin complexed with Galactose, trigonal crystal form	2.8
1HWM	Ebulin,orthorhombic crystal form model	2.8
2OFD	The Crystal Structure of *Sclerotium rolfsii* lectin in complex with *N*-acetyl-D-galactosamine	
1IHO	Crystal structure of MOA, a lectin from the mushroom *Marasmius oreades* in complex with the trisaccharide Gal(1,3)Gal(1,4)GlcNAc	2.41
2VMD	Structure of the complex of Discoidin II from *Dictyostelium discoideum* with Beta-methyl-Galactose	1.9
2VMC	Structure of the complex of Discoidin II from *Dictyostelium discoideum* with *N*-acetyl-galactosamine	1.9
2VME	Structure of the wild-type Discoidin II from *Dictyostelium discoideum*	2.45
2ZQO	Crystal structure of the earthworm R-type lectin C-half in complex with GalNAc	1.8
2CCV	Structure of *Helix pomatia* agglutinin with zinc and *N*-acetyl-alpha-D-Galactosamine (galnac)	
2CE6	Structure Of *Helix pomatia* Agglutinin with no ligands	2.4
1TL2	Tachylectin-2 from *Tachypleus tridentatus* (japanese horseshoe crab)	2.0
1WK1	Solution structure of Lectin C-type domain derived from a hypothetical protein from *C. elegans*	NMR
1W6N	X-ray crystal structure of C2S Human Galectin-1	1.65
1W6Q	X-ray crystal structure of R111H Human Galectin-1	2.10
1BKZ	Crystal Structure of Human Galectin-7	1.9
2GAL	Crystal structure of Human Galectin-7 in complex with Galactose	2.00
1A3K	X-ray crystal structure of the Human Galectin-3 Carbohydrate Recognition Domain (CRD) at 2.1 angstrom resolution	2.10
2D6L	Crystal structure of mouse galectin-9 N-terminal CRD (crystal form 2)	2.5

1A78	Complex of toad ovary Galectin with Thio-digalactose	2.00
1GAN	Complex of toad ovary Galectin with *N*-Acetylgalactose	2.23
1WMY	Crystal Structure of C-type Lectin CEL-I from *Cucumaria echinata*	2.00
1WMZ	Crystal Structure of C-type Lectin CEL-I complexed with *N*-acetyl-D-galactosamine	1.7
1VLC	Crystal Structure of Hemolytic Lectin CEL-III	1.7
2Z48	Crystal Structure of Hemolytic Lectin CEL-III Complexed with GalNac	1.7
2Z49	Crystal Structure of Hemolytic Lectin CEL-III Complexed with methyl-alpha-D-galactopylanoside	1.95

7.5 Monocot Mannose binding lectins with complex sugar specificity: Araceae lectins from *S.guttatum* & *A.tortuosum*

Determination of the forces that stabilize the native conformation of a protein can provide a good understanding of its conformational stability. The stability of a protein can be generally determined by the analysis of denaturant induced or thermally induced unfolding transitions occurring in the protein that are measured either spectroscopically or calorimetrically. The changes occurring in the three secondary structures as a result of these externally induced destabilizing forces can be measured by CD spectroscopy. Information from all these methods can be well correlated and comparison with the data on already available proteins provides a better understanding of such transitions occurring in one particular protein family.

The araceae lectins SGA and ATL have been biophysically characterized by fluorescence and CD spectroscopy. For both these lectins the fluorescence emission maximum at pH 7.0 is 351nm which indicates that the tryptophan residues are exposed to the polar solvent. On denaturation with 8M Gdn-HCl or 6M Urea the lectins are unfolded and the tryptophan is further exposed, as indicated by the red shift in the fluorescence spectrum.

Table 7.2: Comparative account of λ_{max} (native & denatured protein) and extent of quenching of fluorescence by different quenchers in native and 8M urea denatured conditions for Jacalin, SGSL, TDSL, TCSL, DBL AHL, MOL and MCL.

Lectin	λ_{max} (Native)	λ_{max} (Denatured)	Acrylamide Quenching (%)		Iodide Quenching (%)		Cesium Quenching (%)	
			Native	De-natured	Native	De-natured	Native	De-natured
Jacalin	334	356	100	ND	0	ND	5.0	ND
SGSL	331	361.8	45.0	86.3	10.0	58.3	12.5	31.8
TDSL	328	343	27.2	72.7	4.4	49.7	9.5	20.9
TCSL	331	347	57.5	93.0	11.5	57.4	15.7	43.3
DBL	328	352	48.9	74.7	10.6	57.5	1.9	21.0
AHL	335	356	100	-	-	-	-	-
MOL	351	355	82.32	88.04	85.85	86.32	No quenching	ND
MCL	335	365	46	ND	17	ND	12	ND

Some lectins that have been characterized with respect to their tryptophan environments are Jacalin (Sahasrabuddhe *et al.*, 2004), *T. anguina* lectin (SGSL) (Komath & Swamy, 1999), *T. dioica* lectin(TDSL) (Swamy & Swamy, 2005), *T. cucumerina* lectin(TCSL) (Kenoth & Swamy, 2003), *Doilichos* Lectin(DBL) (Sultan *et al.*, 2006), *Artocarpus hirsuta* lectin (AHL) (Gaikwad *et al.*, 2003), *Moringa oliefera* lectin (MOL) (Katre *et al.,*2007), *Momordica charantia* lectin (MCL) (Padma, 1998)

Out of the lectins presented in the table 7.2, it is seen that most of these lectins have tryptophans buried in the hydrophobic core as seen by their λ_{max} in the range of 328-335 nm. Upon denaturation the shift in the λ_{max} to 347-356 has been observed for Jacalin, SGSL, TDSL, TCSL, DBL, AHL and MCL. While in case of MOL as in araceae lectins SGA and ATL the tryptophan environment is polar and they are exposed to the solvent as evident by their λ_{max} in the range of 350-351 nm, and after exposure to denaturants a further unfolding results in the copmplete exposure of the tryptophan (λ_{max} ~ 356nm).

When the accessibility of tryptophan residues is calculated in presence of the neutral quencher acrylamide, the values for Jacalin, AHL and MOL correspond to > 80% quenching which is similar to that observed in araceae lectins. This indicates that SGA and ATL resemble these lectins in that the Trp environment is non-polar and they are solvent exposed.

The anionic quencher I^- results in very less quenching in the members of the cucurbitaceae like TDSL,SGSL,TCSL,MCL and the legume lectin DBL (4.4-11.5 %) , that is indicative of the Trp environment being negatively charged. For SGA, ATL like MOL

quenching with the I^- is in the range 72-80% indicating that these araceae lectins are similar to MOL in having Trp in electropositive environment.

A higher percentage of fluorescence quenching was observed with the iodide ion in case of both the aracea lectins (72.49% for SGA, 74.74% for ATL) as in the case of *M. oleifera* lectin, 85%, (Katre *et al.*, 2008), as compared to the available values for other lectins such as *Doilichos lablab* lectin,10.6%, (Sultan *et al.*, 2006), *T. cucumerina* lectin,11.5%, (Kenoth & Swamy, 2003), *T.anguina* lectin,10.0%, (Komath & Swamy, 1999) and *T.dioica* lectin, 4.4%, (Sultan *et al.*, 2005). Higher quenching by iodide as compared to Cs^+ ions is an indication of predominantly electropositive environment of Trp residues.

References

- Abdu-Allah HH, Tamanaka T, Yu J, Zhuoyuan L, Sadagopan M, Adachi T, Tsubata T, Kelm S, Ishida H, Kiso M (2008). Design, synthesis, and structure-affinity relationships of novel series of sialosides as CD22-specific inhibitors. *J Med Chem* **51**: 6665-6681.
- Adar R, Richardson M, Lis H, Sharon N (1989). The amino acid sequence of *Erythrina corallodendron* lectin and its homology with other legume lectins. *FEBS Lett*. **257**:81-85.
- Agrawal BBI and Goldstein IJ (1965). Specific binding of concanvalin A to cross-linked dextran gels. *Biochem J* **96**:23C–15C.
- Ali V, Prakash K, Kulkarni S, Ahmad A, Madhusudan KP, Bhakuni V (1999). 8-anilino-1-naphthalene sulfonic acid (ANS) induces folding of acid unfolded cytochrome c to molten globule state as a result of electrostatic interactions. *Biochemistry* **38**: 13635-13642.
- Allen AK (1979) A lectin from the exudate of the fruit of the vegetable marrow (*Cucurbita pepo*) that has a specificity for beta-1,4-linked N-acetylglucosamine oligosaccharides. *Biochem J.* **183**: 133-7
- Allen HJ, Cywinski M, Palmberg R, DiCioccio RA (1987). Comparative analysis of galactoside-binding lectins isolated from mammalian spleens. *Arch Biochem Biophys* **256**: 523-533.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**:3389-3402**.**
- Andrews P (1964). Estimation of the molecular weights of proteins by Sephadex gel-filtration. *Biochem J* **91**: 222-233.
- Anuradha P, and Bhide SV (1999). An isolation complex from (*Trichosanthes anguina*) seeds. *Phytochemistry.* **52**: 751-758
- Appukuttan PS, Surolia A, Bachawat BK (1977). Isolation of two galactose-binding proteins from *Ricinus communis* by affinity chromatography. *Indian J Biochem Biophys* **14**: 382-384.

- Aragão KS, Satre M, Imberty A, Varrot A (2008). Structure determination of Discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain, *Proteins*.**73**: 43-52
- Ashford DA, Dwek RA, Rademacher TW, Lis H, Sharon N. (1991) The glycosylation of glycoprotein lectins. Intra- and inter-genus variation in N-linked oligosaccharide expression. *Carbohydr Res*.**213**:215-227
- Attrill H, Takazawa H, Witt S, Kelm S, Isecke R, Brossmer R, Ando T, Ishida H, Kiso M, Crocker PR, van Aalten DM (2006). The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. *Biochem J.* **397**:271-8.
- Au TK, Collins RA, Lam TL, Ng TB, Fong WP, Wan DCC (2000). The plant ribosome inactivating proteins luffin and saporin are potent inhibitors of HIV-1 integrase. *FEBS Lett* **471**:169–172.
- Aub J C, Sanford BH, and Cote MN (1965). Studies on reactivity of tumor and normal cells to a wheat germ agglutinin. *Proc Natl Acad Sci USA* **54**:396-399.
- Aub JC, Tieslau C, Lankester A (1963). Reactions of Normal and Tumor Cell Surfaces to Enzymes. I. Wheat-Germ Lipase and Associated Mucopolysaccharides. *Proc Natl Acad Sci U S A* **50**: 613-619.
- Azzi A, Wang T, Zhu DW, Zou YS, Liu WY, Lin SX (2009). Crystal structure of native cinnamomin isoform III and its comparison with other ribosome inactivating proteins. *Proteins* **74**: 250-255.
- Bagaria A, Surendranath, K, Ramagopal UA, Ramakumar S, Karande AA (2006). Structure-function analysis and insights into the reduced toxicity of *Abrus precatorius* agglutinin I in relation to abrin. *J Biol Chem* **281**: 34465-34474.
- Banerjee R, Das K, Ravishankar R, Suguna K, Surolia A, Vijayan M (1996) Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. *J Mol Biol*. **259**:281-296.
- Banerjee R, Mande SC, Ganesh V, Das K, Dhanaraj V, Mahanta SK, Suguna K, Surolia A, Vijayan M (1994). Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc Natl Acad Sci USA* **91**:227–231.
- Barbieri L, Battelli MG, Stirpe F (1993). Ribosome inactivating proteins from plants. *Biochim Biophys Acta* **1154**: 237-282.

- Barre A, Bourne Y, Van Damme EJ, Peumans WJ, Rougé P (2001). Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. *Biochimie.* **83**:6 45-51
- Barre A, Van Damme EJ, Peumans WJ, Rougé P (1996) Structure-function relationship of monocot mannose-binding lectins. *Plant Physiol*.**112**:1531-40
- Beauchamp JC, Isaacs NW (1999). Methods for X-ray diffraction analysis of macromolecular structures. *Curr Opin Chem Biol* **3**: 525-529.
- Beisel HG, Kawabata S, Iwanaga S, Huber R, Bode W (1999). Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J.* **18**: 2313-22
- Bernard C, Geraldes A, Boiron M (1964). Action of Phytohemagglutinin "in Vitro" on the Lymphocytes of Chronic Lymphoid Leukemia. *Nouv Rev Fr Hematol* **17**: 69-76.
- Bhattacharyya L, Das PK, Sen A (1981) Purification and properties of D-galactose-binding lectins from some Erythrina species: comparison of properties of lectins from *E. indica, E. arborescens, E. suberosa,* and *E. lithosperma*. *Arch Biochem Biophys*. **211**:459-470.
- Bhattacharyya L, Ghosh A, Sen A, (1986) A comparative study on lectins from four Erythrina species. *Phytochemistry* **25**: 2117-2122.
- Bianchet MA, Ahmed H, Vasta GR, Amzel LM (2000). Soluble beta-galactosyl-binding lectin (galectin) from toad ovary: Crystallographic studies of two protein-sugar complexes. *Proteins* **40**: 378-388.
- Blow D (2002). *Outline of Crystallography for Biologist*, 1 edn. Oxford University Press: New York.
- Blow DM, Crick FHC (1959). The treatment of errors in the isomorphous replacement method. *Acta Crystallographica* **12**: 794-802
- Blow DM, Rossmann MG (1961). The single isomorphous replacement method. *Acta Crystallographica* **14**: 1195-1202.
- Blum H, Beier H, Gross HJ (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**, 93-99.

- Blundell TL, Jhoti H, Abell C (2002). High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **1**: 45-54.
- Blundell TL, Johnson LN (1976). Protein Crystallography**,** Vol. 3, Crystallization of Proteins. pp. 59-82. Academic Press
- Bohlool BB and Schmidt EL (1974). Lectins: a possible basis for specificity in the Rhizobium-legume root module symbiosis. *Science* **188**: 269–271.
- Bonavida B (1977). Concanavalin A-mediated activation of antigen-primed lymphocytes into secondary cytotoxic lymphocytes. *J Exp Med* **145**: 293-301.
- Bourne Y, Abergel C, Cambillau C, Frey M, Rougé P, Fontecilla-Camps JC (1990) X-ray crystal structure determination and refinement at 1.9 A resolution of isolectin I from the seeds of *Lathyrus ochrus*. *J Mol Biol.* **214**:571-584.
- Bourne Y, Astoul CH, Zamboni V, Peumans WJ, Menu-Bouaouiche L, Van Damme EJ, Barre A, Rougé P (2002). Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. *Biochem J* **364**: 173-180.
- Bourne Y, Roig-Zamboni V, Barre A, Peumans WJ, Astoul CH, Van Damme EJ, Rouge P (2004). The crystal structure of the *Calystegia sepium* agglutinin reveals a novel quaternary arrangement of lectin subunits with a β-prism fold. *J Biol Chem* **279**:527-533.
- Bourne Y, Zamboni V, Barre A, Peumans WJ, Van Damme EJM, Rouge P (1999). *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure* **7**:1473-1482.
- Boyd WC and Shapleigh E (1954). Separation of individuals of any blood group into secretors and non-secretors by use of a plant agglutinin (lectin). *J Immunol* **73**: 226-231.
- Boyd WC and Shapleigh E (1954). Specific precipitating activity of plant agglutinins (lectins). *Science* **119**: 419.
- Bricogne G (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallographica Section A* **32**: 832-847.
- Brunger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475.
- Bullough PA, Hughson FM, Skehel JJ, Wiley DC (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**: 37-43.

- Burger MM, and Martin GS (1972). Agglutination of cells transformed by Rous sarcoma virus by wheat germ agglutinin and concanavalin A. *Nat New Biol* **237**:9-12.
- Burstein EA, Vedenkina NS, Ivkova MN (1973). Fluorescence and the location of tryptophan residues in protein molecules. *Photochem Photobiol* **18**: 263-279.
- Cammue B, Stinissen HM, Peumans WJ (1985). A new type of cereal lectin from leaves of couch grass (*Agropyrum repens*). *Eur J Biochem* **148**: 315-322.
- Candy L, Peumans WJ, Menu-Bouaouiche L, Astoul CH, Van Damme J, Van Damme EJ, Erard M, Rougé P (2001). The Gal/GalNAc-specific lectin from the plant pathogenic basidiomycete *Rhizoctonia solani* is a member of the ricin-B family. *Biochem Biophys Res Commun* **282**: 655-661.
- Cao Z, Qiu J, Domeniconi M, Hou J, Bryson JB, Mellado W, Filbin MT (2007). The inhibition site on myelin-associated glycoprotein is within Ig-domain 5 and is distinct from the sialic acid binding site. J Neurosci. **27**: 9146-54
- Casellas P, Dussossoy D, Falasca AI, Barbieri L, Guillemot JC, Ferrara P, Bolognesi A, Cenini P, Stirpe F. (1988). Trichokirin, a ribosome-inactivating protein from the seeds of *Trichosanthes kirilowii Maximowicz*. Purification, partial characterization and use for preparation of immunotoxins. *Eur J Biochem* **176**: 581-588.
- Castilho PV, Goto LS, Roberts LM, Araujo AP (2008). Isolation and characterization of four type 2 ribosome inactivating pulchellin isoforms from *Abrus pulchellus* seeds. *FEBS J* **275**: 948-959.
- Chandra NR, Prabu MM, Suguna K, Vijayan M (2001) Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng*.**14**:857-866.
- Chandra NR, Ramachandraiah G, Bachhawat K, Dam TK, Surolia A, Vijayan M (1999). Crystal structure of a dimeric mannose-specific agglutinin from garlic: quaternary association and carbohydrate specificity. *J Mol Biol* **285**: 1157-1168.
- Chantalat L, Wood SD, Rizkallah P, Reynolds CD (1996). X-ray structure solution of amaryllis lectin by molecular replacement with only 4% of the total diffracting matter. *Acta Crystallogr D Biol Crystallogr* **52**: 1146-1152.

- Chatterjee M, Mandal DK (2003) Kinetic analysis of subunit oligomerization of the legume lectin soybean agglutinin. *Biochemistry* **42**:12217-12222.
- Chayen NE (1998). Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr D Biol Crystallogr* **54**: 8-15.
- Chen YL, Chow LP, Tsugita A, Lin JY (1992). The complete primary structure of abrin-a B chain. *FEBS Letters* **309:** 115–118.
- Chipman DM, Grisaro V, Sharon N (1967). The binding of oligosaccharides containing N-acetylglucosamine and N-acetylmuramic acid to lysozyme. The specificity of binding subsites. *J Biol Chem* **242**: 4388-4394.
- Chow LP, Chou MH, Ho CY, Chuang CC, Pan FM, Wu SH, Lin JY (1999). Purification, characterization and molecular cloning of trichoanguin, a novel type I ribosome-inactivating protein from the seeds of *Trichosanthes anguina*. *Biochem J* **338**: 211-219.
- Collaborative (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **50**: 760-763.
- Collaborative Computational Project, Number 4 (1994). The *CCP*4 suite: programs for protein crystallography. *Acta Cryst* D **50**:760-763.
- Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, Lloyd SJ, Shin DS, Getzoff ED, Yeager M, Forest KT, Tainer JA (2003). Type IV Pilin Structure and Assembly: X-Ray and EM Analyses of *Vibrio cholerae* Toxin-Coregulated Pilus and *Pseudomonas aeruginosa* PAK Pilin. *Mol Cell* **11**: 1139-1150.
- Crennell S, Garman E, Laver G, Vimr E, Taylor G (1994) Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure*. **2**:535-544.
- Daniels MA, Hogquist KA and Jameson CS, 2002. Sweet 'n' sour: the impact of differential glycosylation on T cell differentiation. *Nat. Immunol*. **3**: 903–910
- Dao-Thi MH, Rizkallah P, Wyns L, Poortmans F, Loris R (1998). Quaternary structure of UEA-II, the chitobiose specific lectin from gorse. *Acta Crystallogr D Biol Crystallogr* **54**: 844-847.
- Dea ICM, Morrison A, Tipson RS, Derek H (1975). Chemistry and Interactions of Seed Galactomannans. *Advances in Carbohydrate Chemistry and Biochemistry*. Academic Press.

Vol Volume 31, pp 241-312.

- Dec M, Wernicki A (2006). Conglutinin CL-43 and CL-46--three bovine collectins. *Pol J Vet Sci* **9**: 265-275.
- Delbaere LT, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z (1993) Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 A resolution. *J Mol Biol*.**230**:950-965.
- Demchenko AP (1986). Fluorescence analysis of protein dynamics. *Essays Biochem* **22**: 120-157.
- Dessen A, Gupta D, Sabesan S, Brewer CF, Sacchettini JC (1995) X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry*. **34**:4933-4942.
- Dharker PN, Gaikwad SM, Suresh CG, Dhuna V, Khan MI, Singh J, Kamboj SS (2008). Comparative Studies of Two Araceous Lectins by Steady State and Time-Resolved Fluorescence and CD Spectroscopy. *J Fluoresc*. **19**: 239-248
- Dhuna V, Bains JS, Kamboj SS, Singh J, Kamboj S, Saxena AK (2005). Purification and characterization of a lectin from *Arisaema tortuosum* Schott having in-vitro anticancer activity against human cancer cell lines. *J Biochem Mol Biol* **38**: 526-532.
- Dhuna V, Kamboj SS, Kaur A, Saxena AK, Bhide SV, Shanmugavel, Singh J (2007). Characterization of a lectin from *Gonatanthus pumilus D*. Don having anti-proliferative effect against human cancer cell lines. *Protein Pept Lett* **14**: 71-78.
- Di Lella S, Martí MA, Alvarez RM, Estrin DA, Ricci JC (2007). Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy. *J Phys Chem B* **111**: 7360-7366.
- Diederichs K, Karplus PA (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol*. **4**: 269-275.
- Ding JJ, Bao JK, Zhu DY, Zhang Y, Wang DC (2008). Crystallization and preliminary x-ray diffraction analysis of a novel mannose-binding lectin with antiretroviral properties from *Polygonatum cyrtonema hua*. *Protein Pept Lett* **15**: 411-414.
- Dong TX, Ng TB, Yeung HW, Wong RN (1994). Isolation and characterization of a novel

ribosome-inactivating protein, beta-kirilowin, from the seeds of *Trichosanthes kirilowii*. *Biochem Biophys Res Commun* **199**: 387-393.

- Drenth J (1994). Principles of protein crystallography. Springer-Verlag, New York, pp. XIII + 305.
- Drenth J, Haas C (1998). Nucleation in protein crystallization. *Acta Crystallogr D Biol Crystallogr* **54**: 867-872.
- Duncan MJ, Mann EL, Cohen MS, Ofek I, Sharon N, Abraham SN (2005). The distinct binding specificities exhibited by enterobacterial type 1 fimbriae are determined by their fimbrial shafts. *J Biol Chem* **280**: 37707-37716.
- Edelman GM, Cunningham BA, Reeke GN Jr, Becker JW, Waxdal MJ, Wang JL. (1972) The covalent and three-dimensional structure of concanavalin A. *Proc Natl Acad Sci USA*. **69**: 2580-2584.
- Edelman GM, Wang JL (1978). Binding and functional properties of concanavalin A and its derivatives. III. Interactions with indoleactetic acid and other hydrophobic ligands. *J Biol Chem* **253**:3016–3022.
- Eftink MR, Ghiron CA (1977). Exposure of tryptophanyl residues and protein dynamics. *Biochemistry* **16**: 5546-51.
- Eftink MR, Ghiron CA (1981). Fluorescence quenching studies with proteins. *Anal Biochem* **114**: 199-227.
- Eftink MR, Wasylewski Z (1989). Fluorescence lifetime and solute quenching studies with the single tryptophan containing protein parvalbumin from codfish. *Biochemistry* **28**: 382-391.
- Ehrlich P (1891). Experimentelle Untersuchungen ¨uber Immunit¨at. II.Ueber Abrin. *Dtsch Med Wschr* **17**: 1218–9.
- Einspahr H, Parks EH, Suguna K, Subramanian E, Suddath FL (1986) The crystal structure of pea lectin at 3.0-A resolution. *J Biol Chem*.**261**:16518-16527.
- Elgavish S, Shaanan B (1997). Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem Sc* **22**: 462-467.

- Elgavish S, Shaanan B (1998) Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides. *J Mol Biol*. **277**:917-932.
- Elhaggar S (1993). Treatment / control of HIV infection. *Arch. AIDS Res*. **7**: 120-121.
- Emsley P, Cowtan K (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.
- Endo Y, Mitsui K, Motizuki M, Tsurugi K (1987). The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem* **262**: 5908-5912.
- Endo Y, Tsurugi K (1986). The mechanism of action of Ricin and related toxic lectins on eukaryotic ribosome. *Nucleic Acids Symp Ser*. **17**:187-90.
- Engh RA, Huber R (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A* **47**: 392-400.
- Ervin J, Larios E, Osvath S, Schulten K, Gruebele M (2002). What causes hyperfluorescence: folding intermediates or conformationally flexible native states? *Biophys J* **83**: 473-483.
- Eschenburg S, Krauspenhaar R, Mikhailov A, Stoeva S, Betzel C, Voelter W (1998). Primary structure and molecular modeling of mistletoe lectin I from *Viscum album*. *Biochem Biophys Res Commun* **247**: 367-372.
- Falasca AI, Abbondanza A, Barbieri L, Bolognesi A, Rossi CA, Stirpe F (1989**).** Purification and partial characterization of a lectin from the seeds of *Trichosanthes kirilowii* Maximowicz. *FEBS Lett.* **246**: 159-62
- Fei X, Hill M, Ma X, Teng L, Li Q, Stöckigt J (2004). Isolation of a putative ribosome inactivating protein from dried roots of *Trichosanthes kirilowii* used in Traditional Chinese Medicine. *Planta Med*.**70**: 364-5.
- Fraser ME, Fujinaga M., Cherney M.M., Melton-Celsa AR, Twiddy EM., O`Brien AD, James MNG (2004). Structure of Shiga Toxin Type 2 (Stx2) from *Escherichia coli* O157:H7. *J.Biol.Chem*. **279**: 27511-27517.
- Funatsu G, Islam MR, Minami Y, Sung-Sil K. & Kimura M **(1991)**. Conserved amino acid residues in ribosome inactivating proteins from plants. *Biochimie.* **73**: 1157–1161.

- Gabdoulkhakov AG, Savochkina Y, Konareva N, Krauspenhaar R, Stoeva S, Nikonov SV, Voelter W, Betzel C, Mikhailov AM. Structure-function investigation complex of agglutinin from *Ricinus communis* with galactoaza. *TO BE PUBLISHED*
- Gabius HJ, Gabius S, Vehmeyer K, Schauer A, Nagel GA (1987). Endogenous tumor lectins: new tumor markers and targets in cancer therapy? *Onkologie* **10**: 184-185.
- Gabius HJ, Siebert HC, André S, Jiménez-Barbero J, Rüdiger H (2004).Chemical biology of the sugar code. *Chembiochem*. **5**:740-764.
- Gaikwad SM, Gurjar MM, Khan MI (2002). *Artocarpus hirsuta* lectin. Differential modes of chemical and thermal denaturation. *Eur J Biochem.* **269**: 1413-7.
- Ganguly D, Mukhopadhyay C (2007). Extended binding site of ricin B lectin for oligosaccharide recognition. *Biopolymers* **86**: 311-320.
- Gao B, Ma XQ, Wang YP, Chen SZ, Wu S, Dong YC (1994). Refined structure of trichosanthin at 1.73 A resolution. *Sci China B*. **37**:59-73
- Garman EF, Mitchell EP (1996). Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *Journal of Applied Crystallography* **29**: 584-587.
- Garman EF, Schneider TR (1997). Macromolecular Cryocrystallography. *Journal of Applied Crystallography* **30**: 211-237.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**: 3784-3788.
- Gasymov OK, Glasgow BJ (2007). ANS fluorescence: potential to augment the identification of the external binding sites of proteins. *Biochim Biophys Acta* **1774**: 403-411.
- Gegg CV, Roberts DD, Segel IH, Etzler ME (1992). Characterization of the adenine binding sites of two *Dolichos biflorus* lectins. *Biochemistry* **31**: 6938-6942.
- Ghosh M, Mandal DK (2001) Analysis of equilibrium dissociation and unfolding in denaturants of soybean agglutinin and two of its derivatives. *Int J Biol Macromol*. **29**:273-280.
- Ghosh S, Mandal DK (2006) Kinetic stability plays a dominant role in the denaturant-induced unfolding of *Erythrina indica* lectin. *Biochim Biophys Acta*. **1764**:1021-1028.

- Gilboa-Garber N, Mizrahi L (1981) A new mitogenic D-galactosephilic lectin isolated from seeds of the coral-tree *Erythrina corallodendron*. Comparison with *Glycine max* (soybean) and *Pseudomonas aeruginosa* lectins. *Can J Biochem* **59**:315-320.

- Girbés T, Citores L, Iglesias R, Ferreras JM, Muñoz R, Rojo MA, Arias FJ, García JR, Méndez E, Calonge M (1993). Ebulin 1, a nontoxic novel type 2 ribosome-inactivating protein from *Sambucus ebulus* L. leaves. *J Biol Chem*. **268:**18195-9
- Girbes T, Ferreras JM, Arias FJ, Stirpe F (2004). Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. *Mini Rev Med Chem* **4**: 461-476.
- Gottschalk A (1965). The significance of neuraminic acid, serine and threonine for a glycoprotein inhibiting influenza virus haemagglutination. Aust J Exp Biol Med Sci.**43**: 391-8.

- Grahn E, Askarieh G, Holmner A, Tateno H, Winter HC, Goldstein IJ, Krengel U, Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope*, J Mol Biol*. **369**: 710-21.
- Green DW, Ingram VM, Perutz MF (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. Proc. Roy. Soc. Lond., A. **225**, 287-307.
- Guillot J and Konska G (1997). Lectins in higher fungi. *Biochem. Syst. Ecol*. **25**: 203-230.
- Hamblin J, and Kent S.P (1973). Possible role of phytohaemagglutinin in *Phaseolus vulgaris Nat. New Biol*. **245**: 28–30.
- Hamelryck TW, Dao-Thi MH, Poortmans F, Chrispeels MJ, Wyns L, Loris R (1996) The crystallographic structure of phytohemagglutinin-L. *J Biol Chem*.**271**:20479-20485.
- Hamelryck TW, Loris R, Bouckaert J, Dao-Thi MH, Strecker G, Imberty A*, et al* (1999). Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J Mol Biol* **286**: 1161-1177.

- Hardman KD, Ainsworth CF. (1972) Structure of concanavalin A at 2.4-A resolution. *Biochemistry*. **11**:4910-9.
- Hartley MR, Lord JM (2004). Cytotoxic ribosome-inactivating lectins from plants. *Biochim*

Biophys Acta **1701**: 1-14.

- Hatakeyama T, Kohzaki H, Nagatomo H, Yamasaki N (1994). Purification and characterization of four Ca(2+)-dependent lectins from the marine invertebrate, Cucumaria echinata. *J Biochem* **116**: 209-214.
- Hatakeyama T, Matsuo N, Shiba K, Nishinohara S, Yamasaki N, Sugawara H, Aoyagi H (2002). Amino acid sequence and carbohydrate-binding analysis of the N-acetyl-D-galactosamine-specific C-type lectin, CEL-I, from the Holothuroidea, *Cucumaria echinata*. *Biosci Biotechnol Biochem* **66**: 157-163.
- Hatakeyama T, Ohuchi K, Kuroki M, Yamasaki N (1995). Amino acid sequence of a C-type lectin CEL-IV from the marine invertebrate *Cucumaria echinata*. *Biosci Biotechnol Biochem* **59**: 1314-1317.
- Hazes B, Boodhoo A, Cockle S.A, Read R.J (1996). Crystal structure of the pertussis toxin-ATP complex: a molecular sensor. *J.Mol.Biol*. **258**: 661-671.
- Head-Gordon T, Brooks CL (1991). Virtual rigid body dynamics. *Biopolymers* **31**: 77-100.
- Heiny B.M, Albrecht V. and Beuth J (1998). Correlation of immune cell activties and beta-endorphin release in breast carcinoma patients treated with galactose-specific lectin standardized mistletoe extract. *Anticancer Res*. **18**: 583-586.
- Hellin H (1891). Der Eiwesskorpe Abrin; Sieve Wirkung auf Blut. Dissertation, Dorpat.
- Hendrickson WA (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51-58.
- Hendrickson WA, Horton JR, LeMaster DM (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three dimensional structure. *EMBO J.*
- Hendrickson WA, Smith JL, Sheriff S, Harold W. Wyckoff CHWH, Serge NT (1985). Direct phase determination based on anomalous scattering. *Methods in Enzymology*. Academic Press. Vol 115, pp 41-55.
- Hendrickson WA, Teeter MM (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**: 107-113.
- Henrick K, Bawumia S, Barboni EA, Mehul B, Hughes RC (1998). Evidence for subsites in the galectins involved in sugar binding at the nonreducing end of the central galactose of

oligosaccharide ligands: sequence analysis, homology modeling and mutagenesis studies of hamster galectin-3. *Glycobiology* **8**: 45-57.

- Hester G, Kaku H, Goldstein IJ, Wright CS (1995). Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat Struct Biol* **2**: 472-479.
- Hirabayashi J, Satoh M, Kasai K (1992). Evidence that *Caenorhabditis elegans* 32-kDa beta-galactoside-binding protein is homologous to vertebrate beta-galactoside-binding lectins. cDNA cloning and deduced amino acid sequence, *J Biol Chem* **267**: 15485-15490
- Hirschhorn K, Kolodny RL, Hashem N, Bach F (1963). Mitogenic action of phytohaemagglutinin. *Lancet* **2**: 305-306.
- Hohenester E, Hutchinson WL, Pepys MB, Wood SP (1997). Crystal structure of a decameric complex of human serum amyloid P component with bound dAMP. *J Mol Biol* **269**: 570-578.
- Hou X, Meehan EJ, Xie J, Huang M, Chen M, Chen L (2008). Atomic resolution structure of cucurmosin, a novel type 1 ribosome-inactivating protein from the sarcocarp of *Cucurbita moschata*. *J Struct Biol* **164**: 81-87.
- http://www.ebi.ac.uk
- http://commons.wikimedia.org
- http://molecularsciences.org
- http://www.cermav.cnrs.fr/lectines/
- http://www.imperial.ac.uk/research/animallectins/default.html
- http://www.ncbi.nlm.nih.gov/BLAST/.
- Huber, R., and Schneider M. (1985) A group refinement procedure in protein crystallography using Fourier transforms. *J. Appl. Cryst.* **18**: 165-169.
- Hudgin R.L, Pricer W.E. Jr, Ashwell G, Stockert R.J, and Morell A.G (1974). The isolation and properties of a rabbit liver binding protein specific for asialoglycoproteins. *J. Biol. Chem*. **249**: 5536–5543.
- Hume DA, Weidemann MJ (1980). *Mitogenic lymphocyte transformation* Amsterdam: Elsevier/North-Holland Biomedical Press

- Husain J, Tickle I. J, Wood S. P (1994). Crystal structure of momordin, a type I ribosome inactivating protein from the seeds of *Momordica charantia*. *FEBS Lett*. **342**: 154-158.
- Hussain N, Jani PU, Florence AT (1997). Enhanced oral uptake of tomato lectin-conjugated nanoparticles in the rat. *Pharm Res* **14**: 613-618.
- Iglesias JL, Lis H, Sharon N (1982) Purification and properties of a D-galactose/N-acetyl-D-galactosamine-specific lectin from *Erythrina cristagalli*. *Eur J Biochem*. **123**:247-252.
- Ippoliti G, Marini G, Casirola G, Ascari E, Invernizzi R (1975). Lymphocyte reactivity in chronic lymphocytic leukaemia: effect of repeated and prolonged stimulation with PHA and PWM. *Haematologica* **60**: 31-37.
- Ito N, Imai S, Haga S, Nagaike C, Morimura Y, Hatake K (1996). Localization of binding sites of *Ulex europaeus* I, *Helix pomatia* and *Griffonia simplicifolia* I-B4 lectins and analysis of their backbone structures by several glycosidases and poly-N-acetyllactosamine-specific lectins in human breast carcinomas. *Histochem Cell Biol* **106**: 331-339.
- Jeyaprakash AA, Katiyar S, Swaminathan CP, Sekar K, Surolia A, Vijayan M (2003). Structural basis of the carbohydrate specificities of jacalin: an X-ray and modeling study. *J Mol Biol* **332**: 217-228.
- Kalsi G and Etzler M.E (2000). Localization of a Nod factor-binding protein in legume roots and factors influencing its distribution and expression. *Plant Physiol*. **124**: 1039–1048.
- Kantardjieff KA, Rupp B (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**: 1865-1871.
- Katre UV, Suresh CG, Khan MI, Gaikwad SM (2008). Steady state and time-resolved fluorescence studies of a hemagglutinin from *Moringa oleifera*. *J Fluoresc* **18**(2): 479-485.
- Katzin B.J, Collins E.J, Robertus J.D (1991). Structure of Ricin A-chain at 2.5 Å. *Proteins*. **10**: 251–259.
- Kaur A, Kamboj S.S, Singh J, Saxena A.K, Dhuna V (2005a). Isolation of a novel N-acetyl-D-lactosamine specific lectin from *Alocasia cucullata* (Schott.) *Biotechnol. Lett*. **27**:1815-1820.

- Kaur M, Singh K, Rup P.J, Kamboj S.S, Saxena A.K, Sharma M, Bhagat M, Sood S.K, Singh J (2006b). A tuber lectin from *Arisaema jacquemontii* blume with anti-insect and anti-proliferative properties. *J. Biochem. Mol. Biol.* **39**:432-440.
- Kaur M, Singh K, Rup PJ, Saxena AK, Khan RH, Ashraf MT, Kamboj SS, Singh J (2006). A tuber lectin from *Arisaema helleborifolium* Schott with anti-insect activity against melon fruit fly, *Bactrocera cucurbitae* (Coquillett) and anti-cancer effect on human cancer cell lines. *Arch Biochem Biophys* **445**:156-65

- Kavitha M, Sultan NA, Swamy MJ (2009). Fluorescence studies on the interaction of hydrophobic ligands with *Momordica charantia* (bitter gourd) seed lectin. *J Photochem Photobiol* B. **94**: 59-64

- Kavitha M, Swamy MJ (2009). Spectroscopic and differential scanning calorimetric studies on the unfolding of *Trichosanthes dioica* seed lectin. Similar modes of thermal and chemical denaturation. *Glycoconj J.*

- 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.
- Kawagishi H, Nomura A, Mizuno T, Kimura A and Chiba S (1990). Isolation and characterization of a lectin from *Grifola frondosa* fruiting bodies. *Biochim. Biophys. Acta*. **1034**: 247-252.
- Kenoth R, Komath SS, Swamy MJ (2003). Physicochemical and saccharide-binding studies on the galactose-specific seed lectin from *Trichosanthes cucumerina*. *Arch Biochem Biophys* **413**: 131-138.
- Kenoth R, Raghunath Reddy D, Maiya BG, Swamy MJ (2001). Thermodynamic and kinetic analysis of porphyrin binding to *Trichosanthes cucumerina* seed lectin. *Eur J Biochem.* **268**: 5541-9.

- Kenoth R, Swamy MJ (2003). Steady-state and time-resolved fluorescence studies on Trichosanthes cucumerina seed lectin. *J Photochem Photobiol B* **69**: 193-201.

- Kilpatrick DC (1997). Use of Lectins as Mitogens for Lymphocytes. Book Title: Lectin

Methods and Protocols. Series: Methods in Molecular Medicine. **9:** 385-392

- Kleywegt GJ (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr* **52**: 842-857.
- Kobert R (1906). Lehrbuch der Intoxikationen, Enke, Stuttgart.
- Kolatkar AR, Weis WI (1996). Structural basis of galactose recognition by C-type animal lectins. *J Biol Chem* **271**: 6679-6685.
- Komath S.S, Nadimpalli S.K, Swamy M.J (1996). Purification in high yield and characterisation of the galactose-specific lectin from the seeds of snake gourd (*Trichosanthes anguina*). *Biochem. Mol. Biol. Int* .**39**:243-252.
- Komath SS, Kenoth R, Swamy MJ (2001). Thermodynamic analysis of saccharide binding to snake gourd (*Trichosanthes anguina*) seed lectin. Fluorescence and absorption spectroscopic studies. *Eur J Biochem*. **268**:111-9.
- Komath SS, Nadimpalli SK, Swamy MJ (1998). Identification of histidine residues in the sugar-binding site of snake gourd (*Trichosanthes anguina*) seed lectin. *Biochem. Mol. Biol.Int* **44** :107-116.
- Komath SS, Swamy MJ (1999). Fluorescence quenching, time-resolved fluorescence and chemical modification studies on the tryptophan residues of snake gourd (*Trichosanthes anguina*) seed lectin. *J Photochem Photobiol B* **50**: 108-118.
- Konami Y, Yamamoto K, Osawa T, Irimura T (1992). The primary structure of the *Cytisus scoparius* seed lectin and a carbohydrate-binding peptide. *J Biochem.* **112**: 366-75.
- Kondo T, Kurihara S, Yoshikawa T, Mizukami H (2004). Effect of N- and C-terminal deletions on the RNA N-glycosidase activity and the antigenicity of karasurin-A, a ribosome-inactivating protein from *Trichosanthes kirilowii var. japonica*. *Biotechnol Lett* **26**: 1873-1878.
- Konozy EH, Mulay R, Faca V, Ward RJ, Greene LJ, Roque-Barriera MC, Sabharwal S, Bhide SV. (2002) Purification, some properties of a D-galactose-binding leaf lectin from *Erythrina indica* and further characterization of seed lectin. *Biochimie*.**84**:1035-1043.

- Kouzuma Y, Suzuki Y, Nakano M, Matsuyama K, Tojo S, Kimura M, Yamasaki T, Aoyagi H, Hatakeyama T (2003). Characterization of functional domains of the hemolytic lectin CEL-III from the marine invertebrate *Cucumaria echinata*. *J Biochem* **134**: 395-402.
- Krauspenhaar R, Eschenburg S, Perbandt M, Kornilov V, Konareva N, Mikailova I, Stoeva S, Wacker R, Maier T, Singh T, Mikhailov A, Voelter W, Betzel C (1999). Crystal structure of mistletoe lectin I from *Viscum album*. *Biochem Biophys Res Commun* **257**: 418-424.
- Kulkarni KA, Srivastava A, Mitra N, Sharon N, Surolia A, Vijayan M, Suguna K (2004) Effect of glycosylation on the structure of *Erythrina corallodendron* lectin, *Proteins*. **56**: 821-827.
- Kumar A.M, Timm D.E, Neet K.E, Owen W.G, Peumans W.J, Rao A.J (1993). Characterization of the lectin from the bulbs of *Eranthis hyemalis* (winter aconite) as an inhibitor of protein synthesis. *J Biol Chem* **268**: 25176-25183.
- Kumar S.G, Appukuttan P.S, Basu D (1982). α-D-galactose specific lectin from jack fruit (*Artocarpus integrifolia*) seed. *J. Biosci* **4**: 257-261.
- Kusui K, Yamamoto K, Konami Y, Osawa T (1991). cDNA cloning and expression of *Bauhinia purpurea* lectin. *J Biochem*.**109** : 899-903.
- Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Lakowicz JR, Weber G (1973). Quenching of protein fluorescence by oxygen. Detection of structural fluctuations in proteins on the nanosecond time scale. *Biochemistry* **12**: 4171-4179.
- Lam SK, Ng TB (2001). Hypsin, a novel thermostable ribosome-inactivating protein with antifungal and antiproliferative activities from fruiting bodies of the edible mushroom *Hypsizigus marmoreus*. *Biochem Biophys Res Commun*. **285**:1071-5
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**: 283-291.
- Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 282-291.

- Lehrer SS, Leavis PC (1978). Solute quenching of protein fluorescence. *Methods Enzymol* **49**: 222-236.
- Leonidas DD, Elbert BL, Zhou Z, Leffler H, Ackerman SJ, Acharya KR (1995). Crystal structure of human Charcot-Leyden crystal protein, an eosinophil lysophospholipase, identifies it as a new member of the carbohydrate-binding family of galectins. *Structure* **3**: 1379-1393.
- Leonidas DD, Vatzaki EH, Vorum H, Celis JE, Madsen P, Acharya KR (1998). Structural basis for the recognition of carbohydrates by human galectin-7. *Biochemistry* **37**: 13930-13940.
- Levi G, Teichberg VI (1981). Isolation and physicochemical characterization of electrolectin, a beta-D-galactoside binding lectin from the electric organ of Electrophorus electricus. *J Biol Chem* **256**: 5735-5740.
- Li F, Yang XX, Hu WG, Xia HC, Li Z, Zhang ZC (2003). Purification and characterization of trichokirin-S1, a novel ribosome-inactivating peptide from seeds of *Trichosanthes kirilowii*. *Acta biochimica et biophysica Sinica* **35**: 841-846.
- Li F, Yang XX, Xia HC, Zeng R, Hu WG, Li Z*,* Zhang ZC (2003). Purification and characterization of Luffin P1, a ribosome-inactivating peptide from the seeds of *Luffa cylindrica*. *Peptides* **24**: 799-805.
- Li M, Chai JJ, Wang YP,Wang KY, Bi RC (2001). Crystal structure of *Trichosanthes kirilowii* Lectin-1 and its relation to the type 2 ribosome inactivating proteins. *Protein and Peptide Letters*. **8:** 81-87
- Li M, Wang YP, Chai JJ, Wang KY, Bi RC (2000). Molecular-replacement studies of *Trichosanthes kirilowii* lectin 1: a structure belonging to the family of type 2 ribosome-inactivating proteins. *Acta Crystallogr D Biol Crystallogr* **56**: 1073-1075.
- Li YF, Poole S, Rasulova F, McVeigh AL, Savarino SJ, Xia D (2009). Crystallization and preliminary X-ray diffraction analyses of several forms of the CfaB major subunit of enterotoxigenic Escherichia coli CFA/I fimbriae, Acta Crystallogr Sect F. *Struct Biol Cryst Commun.***65**:242-247.
- Lin JY, Ju ST, Wu HL, Tung TC (1973). The binding of abrin and ricin by Ehrlich ascites tumor cells. *Cancer Res* **33**: 2688-2691.

- Lin JY, Lei LL, Tung TC (1969) Purification of abrin from *Abrus precatorius* L. Leguminosae. *Taiwan Yi Xue Hui Za Zhi*. **68**: 518-521.
- Lis H, Sharon N (1977). *The Antigens* Acad. Press: New York. Vol 4.
- Lis H, Sharon N (1998). Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**(2):637-674.
- Lis H, Sharon N, Katchalski E (1966). Soybean hemagglutinin, a plant glycoprotein. I. Isolation of a glycopeptide. *J Biol Chem* **241**: 684-689.
- Liu W, Hu YL, Wang M, Xiang Y, Hu Z, Wang DC (2002). Purification, crystallization and preliminary X-ray diffraction analysis of a novel mannose-binding lectin from *Gastrodia elata* with antifungal properties. *Acta Crystallogr D Biol Crystallogr* **58**: 1833-1835.
- Lobsanov YD, Gitt MA, Leffler H, Barondes S, Rini JM (1993). Crystallization and preliminary X-ray diffraction analysis of the human dimeric S-Lac lectin (L-14-II). *J Mol Biol* **233**: 553-555.
- Lopez-Lucendo MIF, Solis D, Andre S, Hirabayashi J, Kasai K, Kaltner H, Gabius HJ, Romero A (2004). Growth-Regulatory Human Galectin-1: Crystallographic Characterisation of the Structural Changes Induced by Single-Site Mutations and Their Impact on the Thermodynamics of Ligand Binding. *J Mol Biol* **343**: 957.
- Lord JM, Roberts LM, Robertus JD (1994). Ricin: structure, mode of action, and some current applications. *Faseb J* **8**: 201-208.
- Loris R, Hamelryck T, Bouckaert J, Wyns L (1998). Legume lectin structure. *Biochim Biophys Acta* **1383**: 9-36.
- Loris R, Steyaert J, Maes D, Lisgarten J, Pickersgill R, Wyns L (1993) Crystal structure determination and refinement at 2.3-A resolution of the lentil lectin. *Biochemistry* **32**:14229.
- Loris R, Van Overberge D, Dao-Thi MH, Poortmans F, Maene N, Wyns L (1994) Structural analysis of two crystal forms of lentil lectin at 1.8 A resolution. *Proteins.* **20**:330-346.
- Lowe M, Nilsson B (1983). The structure of the complex type oligosaccharide from rabbit hepatic binding protein. A re-examination. *J Biol Chem* **258**: 1885-1887.
- Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951). Protein measurement with the

Folin phenol reagent. *J Biol Chem* **193**: 265-275.

- Mann BJ (2002). Structure and function of the *Entamoeba histolytica* Gal/GalNAc lectin, *Int Rev Cytol.* **216**: 59-80

- Manoj N, Jeyaprakash AA, Pratap JV, Komath SS, Kenoth R, Swamy MJ, Vijayan M (2001) Crystallization and preliminary X-ray studies of snake gourd lectin: homology with type II ribosome-inactivating proteins. *Acta Crystallogr D Biol Crystallogr*. **57**: 912-4.

- Matthews BW (1968). Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.
- Maveyraud L, Niwa H, Guillet V, Svergun DI, Konarev PV, Palmer RA, Peumans WJ, Rougé P, Van Damme EJ, Reynolds CD, Mourey L (2009). Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*. *Proteins* **75**(1):89-103.

- McPherson A (1982). Preparation and analysis of protein crystals. John Wiley & Sons.
- Metzner G, Franz H, Kindt A, Schumann I, Fahlbusch B (1987). Effects of lectin I from mistletoe (ML I) and its isolated A and B chains on human mononuclear cells: mitogenic activity and lymphokine release. *Pharmazie* **42**: 337-340.
- Meyer A, Rypniewski W, Szymański M, Voelter W, Barciszewski J, Betzel C (2008). Structure of mistletoe lectin I from *Viscum album* in complex with the phytohormone zeatin. *Biochim Biophys Acta*. **1784**: 1590-5.

- Miller K (1983). The stimulation of human B and T lymphocytes by various lectins. *Immunobiology* **165**: 132-146.

- Mishra V, Bilgrami S, Sharma RS, Kaur P, Yadav S, Krauspenhaar R, Betzel C, Voelter W, Babu CR, Singh TP (2005). Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site. *J.Biol.Chem.* **280:** 20712-20721

- Mitchell BS, Brooks SA, Leathem AJ, Schumacher U (1998). Do HPA and PHA-L have the same binding pattern in metastasizing human breast and colon cancers? *Cancer Lett* **123**: 113-119.

- Mitchell SW (1860). Researches about the venom of the rattlesnake. *Smithsonian Contributions to Knowledge* **XII:** 89–90.

- Mlsna D, Monzingo AF, Katzin B, Ernst SR, Robertus JD (1993). Structure of recombinant Ricin A-chain at 2.3A° . *Protein Sci.* **2:** 429–435.

- Mo H, Van Damme EJ, Peumans WJ, Goldstein IJ (1993). Purification and characterization of a mannose-specific lectin from Shallot (*Allium ascalonicum*) bulbs. *Arch Biochem Biophys* **306**: 431-438.

- Montfort W, Villafranca JE, Monzingo AF, Ernst SR, Katzin B, Rutenber E, Xuong NH, Hamlin R, Robertus JD (1987). The three-dimensional structure of ricin at 2.8A° . *J. Biol. Chem.* **262:** 5398–5403.

- Monzingo AF, Robertus JD (1992). X-ray analysis of substrate analogs in the ricin A-chain active site. *J.Mol. Biol.* **227:** 1136–1145.

- Moreno E, Teneberg S, Adar R, Sharon N, Karlsson KA, Angström J (1997) Redefinition of the Carbohydrate Specificity of *Erythrina corallodendron* Lectin Based on Solid-Phase Binding Assays and Molecular Modeling of Native and Recombinant Forms Obtained by Site-Directed Mutagenesis. *Biochemistry* **36**: 4429-4437.
- Moulaei T, Botos I, Ziółkowska NE, Bokesch HR, Krumpe LR, McKee TC, O'Keefe BR, Dauter Z, Wlodawer A (2007). Atomic-resolution crystal structure of the antiviral lectin scytovirin. *Protein Sci.* **16**: 2756–2760

- Murray PA, Kern DG, Winkler JR (1988). Identification of a galactose-binding lectin on *Fusobacterium nucleatum* FN-2. *Infect Immun*.56: 1314-9
- Murshudov GN, Vagin AA, Dodson EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.
- Nagae M, Nishi N, Murata T, Usui T, Nakamura T, Wakatsuki S, Kato R (2006). Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from *Mus musculus* reveals the basic mechanism of carbohydrate recognition. *J Biol Chem* **281**: 35884-35893.

- Naidoo KJ, Denysyk D, Brady JW (1997) Molecular dynamics simulations of the N-linked oligosaccharide of the lectin from *Erythrina corallodendron*. *Protein Eng.* **10**:1249-1261.
- Nantwi PK, Cook DJ, Rogers DJ, Smart JD (1997). Lectins for drug delivery within the oral cavity--investigation of lectin binding to oral mucosa. *J Drug Target* **5**: 45-55.
- Navaza J (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallographica Section A* **50**: 157-163.
- Navaza J, Saludjian P, Charles W. Carter, Jr. (1997). AMoRe: An automated molecular replacement program package. *Methods in Enzymology*. Academic Press. Vol Volume 276, pp 581-594.
- Navaza, J., and Saludjian, P. (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol*. **276**, 581–594.
- Ng TB, Chan WY, Yeung HW (1992). Proteins with abortifacient, ribosome inactivating, immunomodulatory, antitumor and anti-AIDS activities from Cucurbitaceae plants. *Gen Pharmacol.* **23**:579-90
- Ng TB, Wong CM, Li WW, Yeung HW (1986). Isolation and characterization of a galactose binding lectin with insulinomimetic activities. From the seeds of the bitter gourd *Momordica charantia* (Family Cucurbitaceae). *Int J Pept Protein Res.* **28**: 163-72
- Nicolson GL, Blaustein J (1972). The interaction of *Ricinus communis* agglutinin with normal and tumor cell surfaces. *Biochim Biophys Acta* **266**: 543-547.
- Nielsen K, Boston RS (2001). Ribosome-inactivating proteins: a plant perspective. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **52:** 785-816.
- Nimrichter L, Burdick MM, Aoki K, Laroy W, Fierro MA, Hudson SA, Von Seggern CE, Cotter RJ, Bochner BS, Tiemeyer M, Konstantopoulos K, Schnaar RL (2008). E-selectin receptors on human leukocytes. *Blood* **112**: 3744-3752.
- Niwa H, Tonevitsky AG, Agapov, II, Saward S, Pfuller U, Palmer RA (2003). Crystal structure at 3 A of mistletoe lectin I, a dimeric type-II ribosome-inactivating protein, complexed with galactose. *Eur J Biochem* **270**: 2739-2749.
- North A (1965). The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. *Acta Crystallographica* **18**:

212-216.

- Nowell PC (1960). Phytohaemagglutinin: an initiator of mitosis in cultures of normal human leucocytes.*Cancer Res.* **20**: 462-466.

- Oda Y, Minami K, Ichida S, Aonuma S (1987). A new agglutinin from the *Tulipa gesneriana* bulbs. *Eur J Biochem* **165**: 297-302.

- O'Hare M, Roberts LM, Lord JM (1992). Biological activity of recombinant *Ricinus communis* agglutinin A chain produced in *Escherichia coli*. *FEBS Lett* **299**: 209-212.
- Olsnes S, Pihl A (1982). Toxic lectins and related proteins. In Molecular Action of Toxins and Viruses (Cohen, P. & Van Heyningen, S., eds), pp. 51–105, Elsevier Scientific Publishing Co., New York,USA.
- Olson LJ, Zhang J, Lee YC, Dahms NM, Kim JJ (1999). Structural basis for recognition of phosphorylated high mannose oligosaccharides by the cation-dependent mannose 6-phosphate receptor. *J Biol Chem* **274**: 29889-29896.
- Osinaga E, Tello D, Batthyany C, Bianchet M, Tavares G, Duran R, Cervenansky C, Camoin L, Roseto A, Alzari PM (1997). Amino acid sequence and three-dimensional structure of the Tn-specific isolectin B4 from *Vicia villosa*. *FEBS Letters* 412: 190-196.
- Otwinowski Z (1993). *Proceedings of the CCP4 study weekend*. 56-62.
- Otwinowski Z, Minor W, Charles W. Carter, Jr. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*. Academic Press. Vol Volume 276, pp 307-326.
- Padma P, Komath SS, Swamy MJ (1998). Fluorescence quenching and time-resolved fluorescence studies on *Momordica charantia* (bitter gourd) seed lectin. *Biochem Mol Biol Int.***45**: 911-22

- Pan KZ, Lin YJ, Zhou KJ, Fu ZJ, Chen MH, Huang DH, Huang DR (1992). Crystal and molecular structure of trichosanthin at 2.6A° resolution. *Sci. China Ser. B.* **36:** 1071–1081.

- Pascal JM, Day PJ, Monzingo AF, Ernst SR, Robertus JD, Iglesias R, Perez Y, Ferreras JM, Citores L, Girbes T (2001). 2.8-A crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin l. *Proteins* **43:** 319-326.

- Peumans WJ, De Ley M, Broekaert WF (1984). An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. *FEBS Lett* **177**:99-103.

- Peumans WJ, Hao Q, Van Damme EJM (2001). Ribosome-inactivating proteins from plants: more than RNA N-glycosidases? FASEB J. ***15***: 1493-1506 .

- Peumans WJ, Spaepen C, Stinissen HM, Carlier AR (1982a). Isolation and partial characterization of a lectin from a false brome grass (*Brachypodium sylvaticum*). *Biochem J* **205**: 635-638.

- Peumans WJ, Stinissen HM, Carlier AR (1982b). Isolation and partial characterization of wheat-germ-agglutinin-like lectins from rye (*Secale cereale*) and barley (*Hordeum vulgare*) embryos. *Biochem J* **203**: 239-243.

- Peumans WJ, Van Damme EJ, Barre A, Rougé P (2001). Classification of plant lectins in families of structurally and evolutionary related proteins. *Adv Exp Med Biol* **491**: 27-54.

- Puri KD, Surolia A (1994). Amino acid sequence of the winged bean (*Psophocarpus tetragonolobus*) basic lectin. Adenine binding and identification of the active-site tryptophan residue. *J Biol Chem* **269**: 30917-30926.

- Raikhel NV, Lee HI, Borekaert WF (1993). Structure and function of chitin binding proteins. *Ann Rev Plant Phys Plant Mol Bio* **44**: 591-615.

- Ramachandraiah G, Chandra NR (2000). Sequence and structural determinants of mannose recognition. *Proteins* **39**: 358-364.
- Ramachandran GN, Ramakrishnan C**,** Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**:95-99.
- Ramachandran GN, Sasisekharan V (1968). Conformation of polypeptides and proteins. *Adv Protein Chem* **23**: 283-438.
- Rao VS, Lam K, Qasba PK (1998). Architecture of the sugar binding sites in carbohydrate binding proteins--a computer modeling study. *Int J Biol Macromol* **23**: 295-307.

- Raval S, Gowda SB, Singh DD, Chandra NR (2004). A database analysis of jacalin-like lectins: sequence-structure-function relationships. *Glycobiology* **14**: 1247-1263.
- Ravelli RB, Garman EF (2006). Radiation damage in macromolecular cryocrystallography. *Curr Opin Struct Biol* **16**: 624-629.
- Ravelli RB, Leiros HK, Pan B, Caffrey M, McSweeney S (2003). Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **11**: 217-224.
- Read SM, Northcote DH (1983). Subunit structure and interactions of the phloem proteins of *Cucurbita maxima* (pumpkin). *Eur J Biochem* **134**: 561-569.
- Reddy MM, Goh KO, Poulter C (1975). Mitogenic stimulation of lymphocytes in cancer patients. *Oncology* **32**: 47-51.
- Reisfeld RA, Lewis UJ, Williams DE (1962). Disk electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature* **195**: 281-283.
- Reisner Y, Itzicovitch L, Meshorer A, Sharon N (1978). Hemopoietic stem cell transplantation using mouse bone marrow and spleen cells fractionated by lectins. *Proc Natl Acad Sci USA* **75**: 2933–2936.
- Reisner Y, Linker-Israeli M, Sharon N (1976). Separation of mouse thymocytes into two subpopulations by the use of peanut agglutinin. *Cell Immunol* **25:** 129–134.
- Renkonen KO (1948). Studies on hemagglutinins present in seeds of some representatives of leguminoseae. *Ann Med Exp Fenn (Helsinke)* **26***: 66-72.*
- Rhodes G (2000). *Crystallography Made Crystal Clear* Academic Press: London.
- Rinderle SJ, Goldstein IJ, Remsen EE (1990). Physicochemical properties of amaranthin, the lectin from Amaranthus caudatus seeds. *Biochemistry* **29**: 10555-10561.
- Rini JM. (1995) Lectin structure. *Annu Rev Biophys Biomol Struct*.**24**:551-577.
- Roberts DD, Goldstein IJ (1983). Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J Biol Chem* **258**: 13820-13824.

- Roberts DL, Weix DJ, Dahms NM, Kim JJ (1998). Molecular basis of lysosomal enzyme recognition: three-dimensional structure of the cation-dependent mannose 6-phosphate receptor. *Cell* **93**: 639-648.
- Rojo MA, Yato M, Ishii-Minami N, Minami E, Kaku H, Citores L, Girbés T, Shibuya N (1997). Isolation, cDNA cloning, biological properties, and carbohydrate binding specificity of sieboldin-b, a type II ribosome-inactivating protein from the bark of Japanese elderberry (*Sambucus sieboldiana*). *Arch Biochem Biophys* **340**: 185-194.
- Rosenfeld R, Bangio H, Gerwig GJ, Rosenberg R, Aloni R, Cohen Y, Amor Y, Plaschkes I, Kamerling JP, Maya RB (2007). A lectin array-based methodology for the analysis of protein glycosylation. *J Biochem Biophys Methods* **70**: 415–426.
- Rossmann MG, Blow DM (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* **15**: 24-31.
- Rutenber E, Katzin BJ, Collins EJ, Mlsna D, Ready MP, Robertus JD (1991). Crystallographic refinement of Ricin to 2.5 Å. *Proteins* **10**: 240–250.
- Rutenber E, Robertus JD (1991). Structure of Ricin B-chain at 2.5 Å resolution. *Proteins* **10**: 260–269.
- Sahasrabuddhe AA, Gaikwad SM, Krishnasastry MV, Khan MI.(2004) Studies on Recombinant single chain Jacalin lectin reveal reduced affinity for saccharides despite normal folding like native Jacalin. *Protein Science* **13**: 3264-3273.
- Sanadi AR, Surolia A (1994). Studies on a chitooligosaccharide-specific lectin from *Coccinia indica*. Thermodynamics and kinetics of umbelliferyl glycoside binding. *J Biol Chem.* **269**: 5072-7.
- Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M (1996). A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold. *Nat Struct Biol* **3**: 596-603.
- Sato T, Ozaki H, Hata Y, Kitagawa Y, Katsube Y, Shimonishi Y (1994). Structural characteristics for biological activity of heat-stable enterotoxin produced by enterotoxigenic *Escherichia coli*: X-ray crystallography of weakly toxic and nontoxic analogs. *Biochemistry*

33*:* 8641-8650.

- Satoh T, Cowieson NP, Hakamata W, Ideo H, Fukushima K, Kurihara M, Kato R, Yamashita K, Wakatsuki S (2007). Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36. *J Biol Chem* **282**: 28246-28255.

- Sauerborn MK, Wright LM, Reynolds CD, Grossmann JG, Rizkallah PJ (1999). Insights into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: the crystal structure at 2 A resolution in complex with alpha1-3 mannobiose. *J Mol Biol* **290**: 185-199.
- Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, Park SJ, Skehel JJ, Wiley DC (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.
- Schnell DJ, Etzler ME (1987). Primary structure of the *Dolichos biflorus* seed lectin. *J Biol Chem.* **262**: 7220-5
- Schnell DJ, Etzler ME (1988) cDNA cloning, primary structure, and in vitro biosynthesis of the DB58 lectin from *Dolichos biflorus*. *J Biol Chem.* **263**:14648-14653
- Schrag JD, Bergeron JJ, Li Y, Borisova S, Hahn M, Thomas DY, Cygler M (2001). The Structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol Cell* **8**: 633-644.

- Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM (1998). X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. *J Biol Chem* **273**: 13047-13052.
- Shaanan B, Lis H, Sharon N (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*. **254**:862-866.
- Shangary S, Singh J, Kamboj SS, Kamboj KK, Sandhu RS (1995). Purification and properties of four monocot lectins from the family Araceae. *Phytochemistry* **40**: 449-455.

- Sharma M, Bhasin D, Vohra H (2008). Differential induction of immunoregulatory circuits of phagocytic cells by Gal/Gal NAc lectin from pathogenic and nonpathogenic Entamoeba. *J Clin Immunol* **28**: 542-557.

- Sharma N, Park SW, Vepachedu R, Barbieri L, Ciani M, Stirpe F, Savary BJ, Vivanco JM (2004). Isolation and characterization of an RIP (ribosome-inactivating protein)-like protein from tobacco with dual enzymatic activity. *Plant Physiol* **134**: 171-181.

- Sharma V, Srinivas VR, Surolia A (1996). Cloning and sequencing of winged bean (*Psophocarpus tetragonolobus*) basic agglutinin (WBA I): presence of second glycosylation site and its implications in quaternary structure. *FEBS Lett* **389**: 289-92.
- Sharma V, Surolia A. (1997) Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J Mol Biol.* **267**:433-445.
- Sharon N (1987). Bacterial lectins, cell-cell recognition and infectious disease. *FEBS Lett.* **217**:145-57.
- Sharon N (2008). Lectins: past, present and future. *Biochem Soc Trans* **36:** 1457–1460.

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

- Sharon N, Lis H (2004). History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **14:** 53–62.

- Sharon N, Lis H, Lotan R (1974). On the structural diversity of lectins. *Coll Int CNRS* **221:** 693–709.

- Shaw PC, Chan WL, Yeung HW, Ng TB (1994). Minireview: trichosanthin- a protein with multiple pharmacological properties. *Life Sci* **55**: 253-262.
- Shibuya N, Goldstein IJ, Van Damme EJ, Peumans WJ (1988). Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J Biol Chem* **263**: 728-734.

- Shibuya N, Goldstein IJ, Van Damme EJ, Peumans WJ (1988). Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J Biol Chem* **263**: 728-734.

- Siddiqui MZ, Sharma AK, Kumar S (1996). Solution conformation of tuftsin. *Int J Biol Macromol* **19**: 99-102.

- Siegall CB, Gawlak SL, Chace D, Wolff EA, Mixan B, Marquardt H (1994). Characterization of ribosome-inactivating proteins isolated from *Bryonia dioica* and their utility as carcinoma-reactive immunoconjugates. *Bioconjug Chem*. **5**: 423-9
- Singh Bains J, Singh J, Kamboj SS, Nijjar KK, Agrewala JN, Kumar V, Kumar A, Saxena AK (2005). Mitogenic and anti-proliferative activity of a lectin from the tubers of Voodoo lily (*Sauromatum venosum*). *Biochim Biophys Acta* **1723**: 163-174.

- Singh J, Kamboj SS (2004). A novel mitogenic and antiproliferative lectin from a wild cobra lily, *Arisaema flavum*. *Biochem Biophys Res Commun* **318**: 1057-1065.

- Singh J, Singh J, Kamboj SS (2004). A novel mitogenic and antiproliferative lectin from a wild cobra lily. *Arisaema flavum Biochem Biophys Res Commun.* **318**: 1057–1065.

- Singh SB, Tevethia SS, Cytotoxicity of concanavalin A - activated hamster lymphocytes. *Infect Immun* **7**: 46-52.
- Spande TF, Witkop B, Hirs CHW (1967). Determination of the tryptophan content of proteins with N-bromosuccinimide. *Methods in Enzymology*. Academic Press. Vol Volume 11, pp 498-506.
- Srinivas, V. R., Reddy, G. B., Ahmad, N., Swaminathan, C. P., Mitra, N. and Surolia, A. (2001) Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta* **1527**:102–111.
- Stehle T, Harrison SC (1997). High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *Embo J* **16**: 5139-5148.
- Stillmark H (1888). Uber Rizin eun giftigens Ferment aus den Samen von Ricin communis L und einige anderen Euphorbiaceen, Inang. Diss., Dorpat.

- Stirpe F (2004). Ribosome-inactivating proteins. *Toxicon* **44**: 371-383.
- Stirpe F, Barbieri L, BattelliMG, Soria M, Lappi DA. (1992). Ribosome-inactivating proteins from plants: present status and future prospects. *Biotechnology.* **10**: 405– 12
- Stirpe F, Battelli MG (2006). Ribosome-inactivating proteins: progress and problems. *Cell*

Mol Life Sci **63**: 1850-1866.

- Stockert RJ, Morell AG, Scheinberg IH (1974). Mammalian hepatic lectin. *Science* **186**: 365–366.

- Sueyoshi S, Tsuji T, Osawa T (1985). Purification and characterization of four isolectins of mushroom (*Agaricus bisporus*). *Biol Chem Hoppe-Seyler* **366**: 213-221.

- Sugawara H, Kusunoki M, Kurisu G, Fujimoto T, Aoyagi H, Hatakeyama T (2004). Characteristic Recognition of N-Acetylgalactosamine by an Invertebrate C-type Lectin, CEL-I, Revealed by X-ray Crystallographic Analysis. *J Biol Chem* **279**: 45219-45225.
- Sultan NA, Rao RN, Nadimpalli SK, Swamy MJ (2006). Tryptophan environment, secondary structure and thermal unfolding of the galactose-specific seed lectin from *Dolichos lablab*: fluorescence and circular dichroism spectroscopic studies. *Biochim Biophys Acta* **1760**: 1001-1008.

- Sultan NA, Swamy MJ (2005). Energetics of carbohydrate binding to *Momordica charantia* (bitter gourd) lectin: an isothermal titration calorimetric study. *Arch Biochem Biophys*. **437**: 115-25.

- Sultan NA, Swamy MJ (2005). Fluorescence quenching and time-resolved fluorescence studies on *Trichosanthes dioica* seed lectin. *J Photochem Photobiol B* **80**: 93-100.

- Sultan NAM, Kenoth R, Swamy MJ (2004) Purification, Physicochemical characterization, saccharide specificity, and chemical modification of a Gal/GalNAc specific lectin from the seeds of *Trichosanthes dioica. Arch. Biochem. Biophys.* **432**: 212-221.
- Sumner J.B (1918). The globulins of the jack bean, *Canavalia ensiformis*. *J. Biol. Chem*. **37**: 137-142.
- Sumner JB, Howell SF (1936). The identification of the hemagglutinin of the jack bean with concanavalin A. *J Bacteriol* **32**: 227-237.

- Surolia A, Sharon N, Schwarz FP (1996) Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J Biol Chem*. **271**:17697-17703.

- Suzuki R, Kuno A, Hasegawa T, Hirabayashi J, Kasai K, Momma M, Fujimoto Z (2009). Sugar-complex structures of the C-half domain of the galactose-binding lectin EW29 from the earthworm *Lumbricus terrestris. Acta Crystallogr Sect.D Biol Crystallogr* **65**: 49-57.
- Svensson C, Teneberg S, Nilsson CL, Kjellberg A, Schwarz FP, Sharon N, Krengel U (2002) High-resolution crystal structures of *Erythrina cristagalli* lectin in complex with lactose and 2'-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J Mol Biol*. **321**:69-83.
- Tabary F, Balandreau J, Bourrillon R (1984). Purification of the rice embryo lectin and its binding to nitrogen-fixing bacteria from the rhizosphere of rice. *Biochem Biophys Res Commun* **119**: 549-555.
- Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY (1995). Crystal structure of Abrin-a at 2.14 A. *J Mol Biol **250**: 354-367.*

- Tai NW, Li F, Li Z, Zhuang DH, Zhang ZC (2000). Purification and Partial Characterization of S-trichokirin, A New Small Ribosome-inactivating Protein, from Seeds of *Trichosanthes kirilowii..Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* . **32**: 495-498
- Tateno H, Uchiyama N, Kuno A, Togayachi A, Sato T, Narimatsu H, Hirabayashi J (2007). A novel strategy for mammalian cell surfaces glycome profiling using lectin microarray. *Glycobiology* **17:** 1138–1146.

- Thies A, Moll I, Berger J, Schumacher U (2001). Lectin binding to cutaneous malignant melanoma: HPA is associated with metastasis formation. *Br J Cancer* 84: 819-823.
- Thompson JD**,** Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res*. **22**:4673-4680.
- Townsend RR, Hardy MR, Wong TC, Lee YC (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 a triantennary structure. *Biochemistry* **25**: 5716-5725.

- Tsuda M (1979). Purification and characterization of a lectin from rice bran. *J Biochem* **86**: 1451-1461.

- Turkenburg JP, Dodson EJ (1996). Modern developments in molecular replacement. *Current Opinion in Structural Biology* **6**: 604-610.
- Turton K, Natesh R, Thiyagarajan N, Chaddock JA, Acharya KR (2004) Crystal structures of *Erythrina cristagalli* lectin with bound N-linked oligosaccharide and lactose. *Glycobiology*. **14**:923-929.
- Uchida T, Yamasaki T, Eto S, Sugawara H, Kurisu G, Nakagawa A, Kusunoki M, Hatakeyama T (2004) Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate *Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism. *J Biol Chem*. **279**: 37133-41
- Van Damme E J M, Hao Q, Barre A, Vandenbussche F, Desmyter S, Rougé P Peumans WJ (2001). Ribosome inactivating proteins: a family of plant proteins that do more than inactivate ribosomes. *Crit. Rev. Plant Sci.* **20**: 395–465.
- Van Damme EJ, Astoul CH, Barre A, Rouge P, Peumans WJ (2000). Cloning and characterization of a monocot mannose-binding lectin from *Crocus vernus* (family Iridaceae). *Eur J Biochem* **267**: 5067-5077.
- Van Damme EJ, Goossens K, Smeets K, Van Leuven F, Verhaert P, Peumans WJ (1995). The major tuber storage protein of araceae species is a lectin. Characterization and molecular cloning of the lectin from *Arum maculatum* L. *Plant Physiol* **107**:1147-58
- Van Damme EJ, Roy S, Barre A, Citores L, Mostafapous K, Rougé P, Van Leuven F, Girbés T, Goldstein IJ, Peumans WJ (1997). Elderberry (*Sambucus nigra*) bark contains two structurally different Neu5Ac(alpha2,6)Gal/GalNAc-binding type 2 ribosome-inactivating proteins. *Eur J Biochem* **245**: 648-655.
- Van Damme EJM, Peumans WJ, Barre A, Rouge P (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci*. **17**: 575–692.
- Van Damme EJN, Peumans WJ, Bardocz S, Pusztai A (1998). Handbook of plant lectins: properties and biomedical applications , pg 23-23
- Vasta GR, Ahmed H, Odom EW (2004). Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Curr Opin Struc*

Biol **14**: 617-630.

- Vasta GR, Hunt JC, Marchalonis JJ, Fish WW (1986). Galactosyl-binding lectins from the tunicate *Didemnum candidum* purification and physicochemical characterization. *J Biol Chem* **261**: 9174-9181.
- Verrey F, Drickamer K (1993). Determinants of oligomeric structure in the chicken liver glycoprotein receptor. *Biochem J* **292** : 149-155.
- Vijayan M, Chandra N (1999) Lectins. *Curr Opin Struct Biol*. **9**:707-714.
- Waljuno K, Scholma RA, Beintema J, Mariono A, Hahn AM (1975). Amino acid sequence of hevein. *Proc Int Rubber Conf*, (Kuala Lumpur) **2**: 518–531.

- Wang H, Ng TB (1998). Ribosome inactivating protein and lectin from bitter melon (*Momordica charantia*) seeds: sequence comparison with related proteins. *Biochem Biophys Res Commun.* **253**: 143-6

- Wang H, Ng TB (2002). Luffangulin, a novel ribosome inactivating peptide from ridge gourd (*Luffa acutangula*) seeds. *Life Sci* **70**: 899-906.
- Wang H, Ng TB, Liu Q (2003). A novel lectin from the wild mushroom *Polyporus adusta*. *Biochem Biophys Res Commun* **307**: 535-539.

- Wang H, Ng TB, Ooi VE, Liu WK (2000). Effects of lectins with different carbohydrate-binding specificities on hepatoma, choriocarcinoma, melanoma and osteosarcoma cell lines. *Int J Biochem Cell Biol* **32**: 365-372.

- Wang HX, Ng TB (2000). Heterogeneity of the ribosome-inactivating protein trichosanthin in *Trichosanthes kirilowii* tubers. *Comp Biochem Physiol B Biochem Mol Biol*. **127**: 519-24.

- Wang HX, Ng TB (2000). Lagenin, a novel ribosome-inactivating protein with ribonucleolytic activity from bottle gourd (*Lagenaria siceraria*) seeds. *Life Sci* **67**: 2631-2638.
- Wang HX, Ng TB, Cheng CH, Fong WP (2003). Contamination of ribosome inactivating proteins with ribonucleases, separated by affinity chromatography on red sepharose. *Prep Biochem Biotechnol* **33**: 101-111.

- Wang P, Tumer NE (1999). Pokeweed antiviral protein cleaves double-stranded supercoiled DNA using the same active site required to depurinate rRNA. *Nucleic Acids Res* **27**: 1900-1905.
- Watanabe K, Honjo E, Tsukamoto T, Funatsu G (1992). Fluorescence studies on the interaction of adenine with ricin A-chain. *FEBS Lett* **304**: 249-251.
- Watkins WM, Morgan W (1952). Neutralization of the anti-H agglutinin in eel serum by simple sugars. *Nature* **169**: 825-826.

- Weber K, Osborn M (1969). The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem* **244**: 4406-4412.
- Weber PC (1991). Physical principles of protein crystallization. *Adv Protein Chem* **41**: 1-36.
- Weimar T, Bukowski R, Young NM (2000). The conformation of the T-antigen disaccharide bound to *Maclura pomifera* agglutinin in aqueous solution. *J Biol Chem* **275**: 37006-37010.
- Weis WI, Drickamer K (1996). Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **65**: 441-473.

- Weston SA, Tucker AD, Thatcher DR, Pauptit RA (1993). Ricin A refined at 1.7A°. In *Collected Abstracts*, pp. 89, XVI Congress and General Assembly IUCR, Beijing, China.

- Whitaker JR (1963). Determination of Molecular Weights of Proteins by Gel Filtration of Sephadex. *Analytical Chemistry* **35**: 1950-1953.
- Wilchek M, Spande TF, Witkop B, Milne GW (1967). Chemical conversion of tyrosine to 6-hydroxyindoles. *J Am Chem Soc* **89**: 3349-3350.
- Wimmerova M, Mitchell E, Sanchez JF, Gautier C, Imberty A (2003) Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**: 27059-27067.
- Wong RN, Dong TX, Ng TB, Choi WT, Yeung HW (1996). alpha-Kirilowin, a novel ribosome-inactivating protein from seeds of *Trichosanthes kirilowii* (family Cucurbitaceae): a comparison with beta-kirilowin and other related proteins. *Int J Pept Protein Res* **47**: 103-109.
- Wood SD, Wright LM, Reynolds CD, Rizkallah PJ, Allen AK, Peumans WJ, Van Damme

EJ. (1999). Structure of the native (unligated) mannose-specific bulb lectin from Scilla campanulata (bluebell) at 1.7 A resolution. *Acta Crystallogr D Biol Crystallogr* **55**: 1264-1272.

- Wright CS, Schroeder MR, Raikhel NV (1993). Crystallization and preliminary X-ray diffraction studies of recombinant barley lectin and pro-barley lectin. *J Mol Biol* **233**: 322-324.
- Wright CS.(1977) The crystal structure of wheat germ agglutinin at 2-2 A resolution. *J Mol Biol*. **111**:439-457.
- Wright LM, Wood SD, Reynolds CD, Rizkallah PJ, Peumans WJ, Van Damme EJ, Allen AK. (1996). Purification, crystallization and preliminary X-ray analysis of a mannose-binding lectin from bluebell (*Scilla campanulata*) bulbs. *Acta Crystallogr D Biol Crystallogr* **52**: 1021-1023.
- Wu P, Brand L (1994). Conformational flexibility in a staphylococcal nuclease mutant K45C from time-resolved resonance energy transfer measurements. *Biochemistry* **33**: 10457-10462.
- Wu TH, Chow LP, Lin JY (1998). Sechiumin, a ribosome-inactivating protein from the edible gourd, Sechium edule Swartz--purification, characterization, molecular cloning and expression. *Eur J Biochem* **255**: 400-408.
- www.bio.davidson.edu
- Xia ZX, Zhang L, Zhang ZM, Wu S, Dong YC (1993). The three-dimensional structure of trichosanthin refined at 2.7A° resolution. *Chinese J. Chem.* **11**: 280–288.
- Xie L, Wang BZ, Hu RG, Ji HB, Zhang L, Liu WY (2001). Structural and functional studies of cinnamomin, a new type II ribosome-inactivating protein isolated from the seeds of the camphor tree. *Eur J Biochem.* **268**: 5723-33
- Xiong JP, Xia ZX, Wang Y (1994). Crystal structure of trichosanthin-NADPH complex at 1.7 Å resolution reveals active-site architecture. *Nat Struct Biol.* **1**: 695-700
- Yeung HW, Ng TB, Wong DM, Wong CM, Li WW (1986). Chemical and biological characterization of the galactose binding lectins from *Trichosanthes kirilowii* root tubers. *Int J Pept Protein Res*. **27**: 208-20.

- Young NM, Oomen RP (1992) Analysis of sequence variation among legume lectins. A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J Mol Biol*. **228**: 924-934
- Young NM, Watson DC, Yaguchi M, Adar R, Arango R, Rodriguez-Arango E, Sharon N, Blay PK, Thibault P. (1995) C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in Escherichia coli. Characterization of "ragged ends" by mass spectrometry. *J Biol Chem*. **270**:2563-2570.
- Yu L, Fernig DG, Smith JA, Milton JD, Rhodes JM (1993). Reversible inhibition of proliferation of epithelial cell lines by *Agaricus bisporus* (edible mushroom) lectin. *Cancer Res* **53**: 4627-4632.
- Yuan YR, He YN, Xiong JP, Xia ZX (1999). Three-dimensional structure of beta-momorcharin at 2.55 A resolution. *Acta Crystallogr D Biol Crystallogr* **55**: 1144-1151.
- Zhang D, Halaweish FT (2007). Isolation and characterization of ribosome-inactivating proteins from Cucurbitaceae. *Chem Biodivers*. **4**: 431-42
- Zhang RG, Scott DL, Westbrook ML, Nance S, Spangler BD, Shipley GG, Westbrook EM (1995). The three-dimensional crystal structure of cholera toxin. *J Mol Biol.* **251**:563-73
- Zhang XJ, Wang JH. (1986). Homology of trichosanthin and ricin A chain. Nature. **321**: 477-8
- Zhao J, Ben LH.,Wu YL, Hu W, Ling K, Xin SM., Nie H.L, Ma L and Pei GJ (1999). Anti-HIV agent trichosanthin enhances the capabilities of chemokines to stimulate chemotaxis and G protein activation, and this is mediated through interaction of trichosanthin and chemokine receptor. *J Exp Med*. **190**: 101-111.