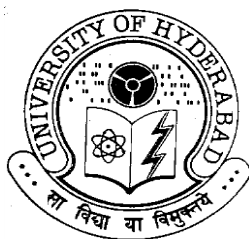


**Purification, Physico-chemical Characterization and
Primary Structure of a Lectin from Mulberry (*Morus
indica*) Latex and a Protease Inhibitor from Okra
(*Abelmoschus esculentus*) Seeds**

*A thesis
Submitted for the degree of*
DOCTOR OF PHILOSOPHY

By
DEBPARN DATTA



**School of Chemistry
University of Hyderabad
Hyderabad – 500 046
INDIA**

June 2017

To,

My loved ones....

“The highest education is that which does not merely give us information but makes our life in harmony with all existence.”

Rabindranath Tagore

Contents

Statement	i
Certificate	iii
Acknowledgments	v
Abbreviations	ix
Chapter 1: Introduction	1
Chapter 2: Biophysical characteristics and cytotoxicity of an affinity-purified α -D-galactose-specific jacalin-related lectin from the latex of mulberry (<i>Morus indica</i>)	33
Chapter 3: Elucidation of primary and N-glycan structures of a jacalin-related Lectin, MLGL by nanoESI-Q-ToF mass spectrometry	53
Chapter 4: Fluorescence and circular dichroism studies on the accessibility of tryptophan residues and unfolding of a jacalin-related α -D-galactose-specific lectin from mulberry (<i>Morus indica</i>)	79
Chapter 5: Biophysical characterization and partial amino acid sequence of a Kunitz-type protease inhibitor from okra (<i>Abelmoschus esculentus</i>) seeds	105
Chapter 6: General discussion and conclusions	127
References	135
Plagiarism Report	155
Curriculum vitae	165



STATEMENT

I hereby declare that the matter embodied in this thesis is the result of investigations carried out by me in the School of Chemistry, University of Hyderabad, Hyderabad, under the supervision of **Prof. Musti J. Swamy**.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made whenever the work described is based on the finding of other investigators. Any omission which might have occurred by oversight or error is regretted.

Hyderabad

June 2017

Debparna Datta



CERTIFICATE

This is to certify that the thesis entitled “**Purification, Physico-chemical Characterization and Primary Structure of a Lectin from Mulberry (*Morus indica*) Latex and a Protease Inhibitor from Okra (*Abelmoschus esculentus*) Seeds**” submitted by **Debparna Datta** bearing registration number **11CHPH17** in partial fulfilment of the requirements for award of Doctor of Philosophy in the School of Chemistry is a bonafide work carried out by her under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma.

Parts of this thesis have been:

A. Published in the following publications:

1. Datta D, Pohlentz G, Schulte M, Kaiser M, Goycoolea FM, Müthing M, Mormann M, Swamy MJ. *Arch Biochem Biophys*. **2016**, 609, 59-68. (**Chapter 2 and 3**)
2. Datta D, Swamy MJ. *J Photochem Photobiol B: Biol*. **2017**, 170, 108-17. (**Chapter 4**)

B. Presented in the following conferences:

1. 27th International Carbohydrate Symposium, ICS (International)
2. First Indo-Taiwan Symposium on Recent Trends in Chemical Sciences, RTCS (International)
3. 6th Asian Community of Glycoscience and Glycotechnology, ACGG (International)

Further the student has passed the following courses towards fulfilment of course work requirement for Ph.D.

Course code	Name	Credits	Pass/Fail
CY-801	Research Proposal	3	Pass
CY-802	Inorganic Chemistry Lab II	3	Pass
CY-805	Instrumental Methods A	3	Pass
CY-851	Biological Chemistry	3	Pass

Prof. Musti J. Swamy
(Thesis Supervisor)

Dean
School of Chemistry

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Prof. Musti J. Swamy, my PhD supervisor for his constant guidance, valuable suggestions and fruitful discussions throughout the course of my PhD. His continuous support and encouragement even when things did not go very well have helped me stay positive and achieve the end results of PhD.

I am particularly grateful to Dr. Michael Mormann, Institute for Hygiene, University of Muenster, Germany for allowing me to work in his lab and introducing me to mass spectrometric techniques. I would also like to thank Dr. Gottfried Pohlentz of this institute for his help with the analysis of mass spec data. The time I spent in the lab of Dr. Mormann has been a great learning experience for me.

I would like to extend my sincere thanks to Prof. Dr. Francisco Goycoolea formerly at the Institute of Plant Biotechnology and Biology, University of Muenster (currently at University of Leeds, UK) for giving me the opportunity to work in his lab. During this time I have learned about mammalian cell culture techniques and characterization of nanoparticles. I would also like to acknowledge Ms. Susana Pereira, who helped me with the cell culture work during this time.

I would like to extend my sincere thanks to Prof. Appa Rao Podile, Vice-Chancellor, University of Hyderabad for allowing me to work in his lab and learn molecular biology techniques. During this time help from his student Mr. Papa Rao is gratefully appreciated.

My sincere thanks goes out to Dr. Kay Grobe, University of Muenster and his student Dr. Shyam Bandari for help with molecular biology techniques during my stay in Germany.

I would like to thank Prof. Bruno Moerschbacher, University of Muenster and Prof. Nadimpally Siva Kumar, University of Hyderabad, coordinators for the Indo-German collaborative project IRTG-MCGS for giving me the opportunity to go on a research stay at the University of Muenster, Germany for a period of nine months.

I am highly thankful to Prof. Attipalli Reddy, University of Hyderabad for allowing me to collect latex samples from his mulberry plantation.

I am also thankful to Dr Anthony Addlagatta, Indian Institute of Chemical Technology for introducing me to protein crystallization techniques.

I am particularly grateful to Prof. Anunay Samanta of our school and his student Ms. Sneha Paul for help with the time-correlated single-photon counting instrument. I would also like to acknowledge Prof. Lalitha Guruprasad and her students Dr. Rajender Oddepally, Mr. Maaged Abdullah and Ms. Bala Divya for help with particularly protein modelling and with some other experiments as well.

I thank Prof. M. Durga Prasad and Prof. T. P. Radhakrishnan, former and present Deans of School of Chemistry for providing the infrastructure for carrying out my research work. My sincere thanks goes out to Prof. A. Samanta and Prof. S. Mahapatra, my doctoral committee members for their support.

I would like to gratefully acknowledge my senior Dr. Kishore Babu Bobbili for teaching me the techniques of biochemistry and for his timely suggestions on various things. I would also like to acknowledge Dr. Thirupathi Reddy, Dr. Pavan Kumar Nareddy, Dr. D. Sivaramakrishna, Dr. Sudheer Kumar and Dr. Bhanu Pratap Singh for their help with my research work and time to time suggestions. I am thankful to the present members of my lab, Dr. Debashree Das, Dr. Venu Gopal, Ms. Saradmoni Mondal, Mr. Sk. Alim and Mr. Ravindra Banadhua for providing a working atmosphere in the lab.

I would also like to thank all the faculty members and all the non-teaching staff of the school for their help on various occasions during my PhD.

I am particularly thankful to my friend, Ms. Arpita Ghosh who has not only supported me mentally but has also helped me with my research work in many ways. I would also like to thank all my friends who have always been there to help me and made my stay in Hyderabad a memorable one; Dr. Tulika, Dr. Subrata, Dr. Sanghamitra, Dr. Tapta, Dr. Vasudhara, Dr. Rumpa, Dr. Tanmoy, Dr. Pramiti, Dr. Manjeet kaur, Dr. Lalitha, Poorna Manasa, Suman (Sen), Rajendra Narayan, Sabari, Tasnim, Sneha, Suchana, Meena and Sugata.

I would also like to thank Beatriz, Bianca, Juan Pablo, Mona, Jule, Katja, Mathias, Christian, Lukas, Xiaofey, Stefan and Christoph for their help and support during my stay in Germany.

My sincere thanks goes out to Sudipto, Srujana, Lasya, Divya, Keshava, Raju, Navendu, Ramana, Koushik, Karunamoy da, Anitha, Suman (Ghosh), Saddam, Krishna Reddy, Subhabrata, Apurbo, Kallol, Shruthi, Swati, Amala, Leela, Anjana of this school for helping me in many ways.

This thesis would be incomplete without the help and support of my friend, Dr. Mathias Kaiser, I am forever grateful to him. I would also like to acknowledge my childhood friends Anindita, Nabanita, Taniya, Tanavi, Ria, Ganga, Pial, Pinkey and Swati for their unconditional support.

Last, but by no means least, I am eternally grateful to my parents Mr. Dilip Kumar Datta and Mrs. Tulika Datta, for being very supportive, understanding and patient with me. I would also like to acknowledge my sister, Dr. Debika Datta and my brother-in-law Mr. Anuvab Palit for their endless love and support.

Debparna Datta

ABBREVIATIONS

AA	Amino acid
ACN	Acetonitrile
ACA	<i>Allium cepa</i> agglutinin
AIDS	Acquired immune deficiency syndrome
ANS	1, 8-anilinonapthalenesulfonic acid
APA	<i>Allium porrum</i> agglutinin
ASAL	<i>Allium sativum</i> agglutinin
BBi	Bowman-Birk inhibitor
Calsepa	<i>Calystegia sepium</i> agglutinin
CD4+	Cluster of differentiation 4+
CD	Circular dichroism
CID	Collision-induced dissociation
Con A	Concanavalin A
Conrava	<i>Convolvulus arvensis</i> agglutinin
DAMP	Damage-associated molecular pattern
DMSO	Dimethyl sulfoxide
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraaceticacid
Gdn.HCl	Guanidine hydrochloride
Gdn.SCN	Guanidine thiocyanate
gJRL	Galactose-specific jacalin related lectin

GNA	<i>Galanthus nivalis</i> agglutinin
Heltuba	<i>Helianthus tuberosus</i> agglutinin
HIV-1	Human immunodeficiency virus 1
HFR-1	Hessian fly responsive protein 1
HPA	<i>Helix pomatia</i> agglutinin
IAA	Iodoacetamide
IgA1	Immunoglobulin A1
JRL	Jacalin related lectins
LOA	<i>Listera ovata</i> agglutinin
MAMP	Microbe-associated molecular pattern
MCF-7	Michigan cancer foundation-7
MDCK	Madine Darby canine kidney
MEM	Minimum essential medium
Me α Gal	Methyl- α -D-galactose
mJRL	Mannose-specific jacalin related lectin
MIL	<i>Morus indica</i> lectin
MLGL	Mulberry latex galactose-specific lectin
MPA	<i>Maclura pomifera</i> agglutinin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NanoESI-Q-ToF MS	Nano electrospray ionization quadrupole time-of-flight mass spectrometry
NICTABA	<i>Nicotiana tabacum</i> agglutinin

PAMP	Pathogen-associated molecular pattern
PI	Protease inhibitor
PNA	Peanut agglutinin
PP1	Phloem protein 1
PP2	Phloem protein 2
PRR	Pattern recognition receptor
PSA	<i>Pisum sativum</i> agglutinin
PTM	Post-translational modification
RIP	Ribosome inactivating protein
RNA	Ribonucleic acid
SBA	Soybean agglutinin
SDS-PAGE	Sodium dodecylsulphate-polyacrylamide gel electrophoresis
SKTI	Soybean Kunitz trypsin inhibitor
SNA	<i>Sambucus nigra</i> agglutinin
TF-antigen	Thomsen-Friedenreich antigen
UDA	<i>Urtica dioica</i> agglutinin
WGA	Wheat germ agglutinin

Introduction

Chapter 1

1.1 Emergence of glycomics as the new ‘omics’ science

The availability of complete sequence of the human genome revealed that there are about 20,000-25,000 protein-encoding genes, which is less than 2% of the entire genome [International Human Genome Sequencing Consortium, 2004]. The knowledge of human genome sequence data did not provide complete understanding of all the physiological processes of a cell or an organism since these are essentially governed by the proteome [Varki et. al., 2009]. The proteome is dynamic as it can vary largely within the life cycle of a cell and post-translational modifications makes it more complex. Among various post-translational modifications such as phosphorylation, ubiquitination, lipidation etc., glycosylation is the most important one [Raman et al., 2005]. Glycosylation is a process by which glycans are attached to either proteins or lipids by enzymes called glycosyltransferases [Ohtsubo and Marth, 2006]. All the glycans produced by a given cell type or an organism is termed as glycome and the study of structure and function of the glycome is referred to as glycomics [Varki et. al., 2009]. Glycans comprising linear or branched polymers of monosaccharides coat the outer layer of every single cell type and participate in many important cellular processes, for examples, molecular trafficking and clearance, cell adhesion, signal transduction, receptor activation as well as microbial pathogenesis and metastasis [Ohtsubo and Marth, 2006; Pilobello and Mahal, 2007]. This is why in post genomics era, glycomics has attracted the attention of scientists worldwide as it can yield a large amount of information on human health and diseases. However, the branched nature of the carbohydrates, the diversity of secondary modifications of sugar residues and their indirect relationship to the genome make it challenging to study the glycome in comparison to the proteome [Gupta and Surolia, 2012; Pilobello and Mahal, 2007].

Lectins, which are naturally occurring sugar binding proteins, can be used as a tool to obtain a greater knowledge of the glycome and overcome the aforementioned challenges in glycomics.

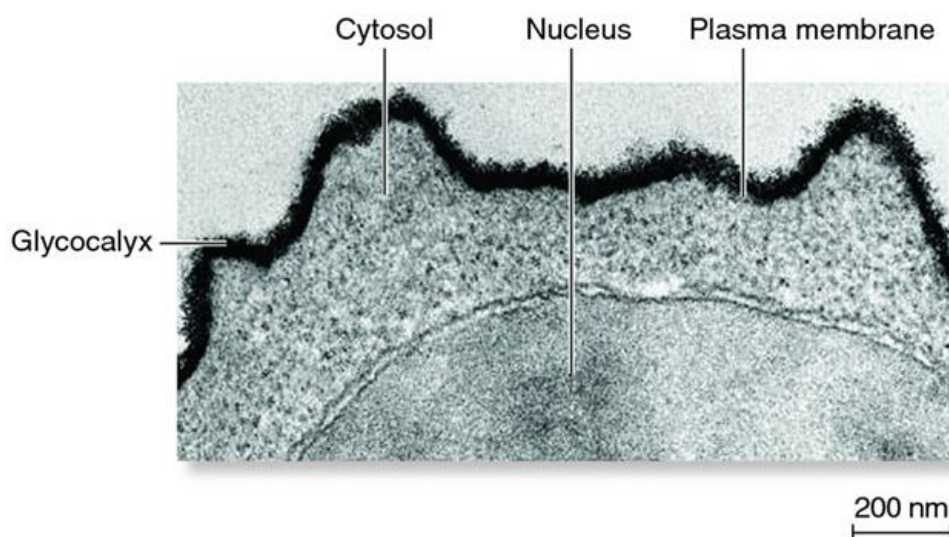


Fig. 1.1: An electron micrograph of a lymphocyte showing the carbohydrate coat around the cell known as ‘glycocalyx’ [Alberts et al., 2002]

1.2 Lectins: The decoders of cellular glycode

Lectins are a unique class of proteins which can bind to specific carbohydrate structures in a reversible manner. These proteins are different from antibodies and carbohydrate modifying enzymes since these are of non-immune origin and are devoid of any catalytic activity [Goldstein et al., 1980; Lis and Sharon, 1973]. As every single cell type is coated with a sugar layer known as the ‘glycocalyx’ (Fig 1.1), lectins are able to precipitate different types of cells and hence are also termed as cell agglutinins [Sharon and Lis, 2004]. The ubiquitous distribution of lectins in every possible living systems ranging from bacteria, viruses, fungi to plants and animals accentuates the significance of protein-carbohydrate interactions in cellular processes. In plants, these proteins are primarily found to occur in the vegetative storage organs

like seeds (Concanavalin A, ricin, jacalin), roots (*Urtica dioica* agglutinin), bulb (*Solanum tuberosum* lectin), bark (MornigaG, MornigaM) or leaves (mistletoe lectin/viscumin) with majority of the lectins being accumulated in the seeds [Rüdiger and Gabius, 2001]. However, lectins have also been found in different fluidic systems of plants such as phloem (e.g. pumpkin phloem lectin) and latex (e.g. *Morus indica* lectin) [Narahari and Swamy, 2010; Patel et al., 2012]. Inside a cell, lectins are mainly found in the protein-bodies along with other storage proteins and also in the cytoplasm, nucleus and the intercellular space [Etzler et al., 1984; Peumanns et al., 2000; Van Damme et al., 2004]. Animal lectins occur both intra- and extracellularly as well as in membrane bound form with specific endogenous roles [Drickamer and Taylor, 1993].

1.2.1 A brief history of lectins

Lectins have been known for over a hundred years now but the molecular basis of carbohydrate binding was discovered in the beginning of twentieth century and detailed investigations were taken up only from the late nineteen sixties. The first lectin that was accidentally discovered from the seeds of castor beans (*Ricinus communis*) in 1888 by Peter Hermann Stillmark is now known as the type-II ribosome-inactivating protein, ricin [Rüdiger and Gabius, 2001; Sharon, 2006]. Ricin is also a potential biological warfare agent due to its high toxicity. However, until the purification and crystallization of Concanavalin A (Con A) from the seeds of jack beans (*Canavalia ensiformis*) by the eminent biochemist James B. Sumner lectins were not known to be proteins. Sumner's work on Con A further led to the discovery that lectins are able to bind carbohydrates [Sharon and Lis, 2004]. This subsequently led to a very important finding in human biology which is the biochemical basis of ABO blood group system [Sharon and Lis,

2004]. The term lectin (Latin *legere*) which means to pick or choose, was coined by Boyd and Shapleigh in 1954 [Boyd and Shapleigh, 1954].

1.2.2 Different lectin folds

Plant lectins are divided into seven distinct families based on their amino acid sequence and structural motifs. These are legume lectins, chitin binding lectins, type 2 ribosome inactivating proteins and related lectins, jacalin related lectins, monocot mannose binding lectins, amaranthin lectins and cucurbitaceae phloem lectins [Van Damme et al., 1998]. Legume lectins, found abundantly in the leguminous plants, are either a single or double chain protomeric proteins forming either dimeric (e.g. *Pisum sativum* agglutinin) or tetrameric structures (e.g. Con A). The three dimensional structure in each case is composed primarily of β -sheets and require manganese and calcium ions for their carbohydrate binding [Agrawal and Goldstein, 1968].

Chitin binding lectins possess at least one hevein domain. Hevein is a 43 amino acid, single chain lectin isolated from rubber tree (*Hevea brasiliensis*) latex. These proteins, as the name suggests, are mainly specific for N-acetylglucosamine and its oligomers but also recognize other sugars such as sialic acid [Van Damme et al., 1998]. Notable lectins that belong to this family are wheat germ agglutinin (WGA), *Urtica dioica* agglutinin (UDA) and viscumin.

Type-II ribosome inactivating proteins (RIPs) are cytotoxic proteins comprising a catalytic chain with N-glycosidase activity and a lectin chain which specifically recognizes galactose and its derivatives. The lectin chain facilitates the protein to enter inside cell by binding to cell surface glycoconjugates whereas the catalytic chain fragments the N-glycosidic

bond of adenosine in the large subunit of ribosomal RNA, resulting in inhibited protein synthesis [Stirpe, 2004; Stirpe and Battelli, 2006]. This family includes ricin and abrin as highly toxic lectins; however, there are some non-toxic homologues as well such as ebulin and snake gourd (*Trichosanthes anguina*) seed lectin. In ebulin the non-toxic nature of the protein arises solely due to a defective carbohydrate binding site [Pascal et al., 2001], whereas in SGSL it is due to a combining effect of the catalytic chain and the lectin chain [Sharma et al., 2013].

The first lectin identified in the monocot-mannose binding family was *Galanthus nivalis* agglutinin (GNA), commonly known as snowdrop lectin [Van Damme et al., 1987]. Later several other lectins similar to GNA such as garlic lectin (*Allium sativum* agglutinin) and tulip bulb lectin have been isolated from monocots and they show exclusive preference towards mannose [Barre et al., 1996].

Jacalin related lectin family comprises lectins structurally and evolutionarily similar to the jack fruit (*Artocarpus integrifolia*) seed lectin, jacalin. As this thesis focuses on a jacalin-related lectin, a detailed discussion on this family will be given in section 1.2.7. Amaranthin from *Amaranthus caudatus* seeds is specific for N-acetylgalactosamine and closely related lectins from different *Amaranthus* family are collectively grouped into amaranthin family of lectins.

Cucurbitaceae phloem lectins are dimeric, chitin-specific proteins present in the phloem exudate of many Cucurbitaceae species including pumpkin (pumpkin phloem lectin), cucumber (cucumber phloem lectin) and ridge gourd (*Luffa acutangula* lectin). These proteins are also termed as phloem protein 2 (PP2) and are involved in wound sealing mechanism by

forming disulphide bridges with phloem protein 1 (PP1) [Narahari and Swamy, 2010; Nareddy et al., 2016].

Classification of animal lectins based on sequence similarity and structures is not very straightforward as with plant lectins due to vast structural diversity among them. Galectins are one of the largest families of animal lectins which specifically recognize endogenous β -galactosides. Galectins have been found in all organisms and even in the plant, *Arabidopsis thaliana* genome, a galectin-like sequence has been predicted. The physiological functions of galectins include regulation of immune and inflammatory responses, involvement in some types of cancer progression, animal development etc [Kilpatrick, 2002; Varki et al., 2009]. C-type lectins are another large family of lectins found in the animal kingdom that consist of the subfamilies – selectins, collectins, endocytic receptors and proteoglycans. The name C-type originated due to their dependency on calcium ion for carbohydrate binding; however, their specificity towards carbohydrates varies largely. C-type lectins are majorly involved in cellular adhesion processes [Varki et al., 2009].

1.2.3 Principles of carbohydrate binding

Hydrogen bonding and hydrophobic interactions play major roles in lectin-carbohydrate interactions. However, in many cases the reversible binding of carbohydrates to lectins also involves electrostatic interactions, metal coordination bonds and water molecules. The carbohydrate binding site is characterized by a shallow depression on the surface of the lectin [Sharon and Lis, 2003]. One of the striking features of lectin-carbohydrate interactions is that the affinity of lectins towards monosaccharides is very weak (dissociation constant in the millimolar range) even though the recognition is highly specific. However, often several fold

increase in affinity is observed for their interactions with oligosaccharides of cell surface glycoconjugates. For example, Con A shows 10^5 -fold higher affinity towards a synthetic mannose polymer than simple methyl- α -mannopyranoside [Mortell et al., 1996] and the galactose/N-acetylgalactosamine specific Ashwell lectin (the major lectin of liver parenchyma cells) shows 10^3 and 10^6 -fold higher affinity towards the bis- and tris-derivatives of the monomer, respectively [Monsigny et al., 2000]. This high affinity of lectins towards multivalent ligands is termed as 'cluster glycoside effect' [Dimick et al., 1999; Sharon and Lis, 2003]. Based on the monosaccharide specificity, lectins can be classified into five groups as well: mannose-, galactose/N-acetylgalactosamine-, sialic acid-, fucose- and N-acetylglucosamine-specific, all of which are building blocks for animal cell surface glycans.

1.2.4 Non-carbohydrate binding

Many lectins have been found to possess a non-carbohydrate ligand binding site along with the carbohydrate binding site. These non-carbohydrate ligands are generally hydrophobic in nature and they bind to a region of the protein which is distinct from the carbohydrate binding site [Komath et al., 2006]. Small hydrophobic fluorescent probes such as ANS (1, 8-anilinonaphthalenesulfonic acid) and TNS (2,6-toluidinylnaphthalenesulfonic acid), several phytohormones like abscisic acid, gibberellic acid, cytokinin, kinetin as well as the nucleic acid base adenine were found to bind several lectins with an affinity which can be compared to that of lectin-monosaccharide interactions or can be even higher sometimes. For example, the binding constants for the interaction of WGA with adenine and other phytohormones were in the range of $1.6 - 2.3 \times 10^6 \text{ M}^{-1}$ [Bogoeva et al., 2004]. In addition to this, many lectins were also shown to bind synthetic porphyrin molecules which can be used as potential photosensitizers in photo dynamic therapy. Con A and pea lectin interacted with water soluble

porphyrins with binding constants in the range of $1.2 - 6.3 \times 10^4 \text{ M}^{-1}$ [Bhanu et al., 1997]. Similarly, jacalin showed binding constants in the range of $2.4 \times 10^3 - 1.3 \times 10^5 \text{ M}^{-1}$ for its interaction with porphyrins. A porphyrin pair was found to sandwich between two symmetry related monomers of jacalin in the crystal structure of porphyrin-jacalin complex [Goel et al., 2004; Komath et al., 2000a].

1.2.5 Lectins in microorganisms

Many viruses use sialic acid specific lectins to mediate infection into the host cells, such as the influenza virus. Influenza virus hemagglutinin is a widely studied lectin which binds to cell surface sialic acids in order to be internalized by the host cells. The influenza virus also expresses the enzyme sialidase which cleaves sialic acids from glycoconjugates thereby helping the virus move through and bud from the infected host cells [Varki et al., 2009].

Bacteria also produce various toxins which are actually glycan binding proteins or lectins, such as the cholera toxin and the Shiga toxin. Shiga toxin inhibits protein synthesis in ribosomes in a way similar to the plant ribosome inactivating proteins (RIPs). The structures of both cholera and Shiga toxins are AB₅ type, where subunit A with a specific enzymatic activity is non-covalently attached to the pentameric B-subunit with ganglioside binding specificity in each monomer. The B-subunit assists in the internalization of the subunit A inside the cell by binding to membrane bound glycolipids [Girbes et al., 2004; Johannes and Roemer, 2010; Richards et al., 1979].

1.2.6 Physiological functions of plant lectins

Unlike animal lectins, plant lectins generally do not show specificity towards endogenous glycans. This lack of specificity towards plant-specific carbohydrate structures precludes the

understanding of biological functions of plant lectins. Many lectins are thought to behave as storage proteins primarily due to their occurrence in the vegetative storage organs and their subcellular localization in the vacuoles or protein bodies with other storage proteins [Van Damme et al., 2002]. However, evidences are emerging continuously on behalf of their involvement in the protection of plant against attack by fungi, insects or herbivorous animals.

Type II ribosome inactivating proteins (RIP) are the best examples in this respect. These proteins are known to be highly toxic towards higher animals including humans for a very long time. Type II RIPs exert cytotoxicity towards some insects and fungi as well. Many chitin binding lectins such as WGA, potato (*Solanum tuberosum*) lectin as well as some legume lectins like peanut agglutinin (PNA) and soybean agglutinin (SBA) have antifungal activity. Several lectins from monocot mannose-binding family (GNA, *Allium sativum* agglutinin, *Allium cepa* agglutinin etc) as well as some lectins from legumes (PSA, Con A), hevein related (WGA, UDA), jacalin related (Heltuba) and amaranthin (ACA) family were shown to possess insecticidal property against agricultural pests (see Table 1.1). Engineering of GNA in transgenic wheat plants resulted in increased resistance towards different insect species [Peumans and Van Damme, 1995; Vandenborre et al., 2011].

However, more direct evidence for the endogenous roles of plant lectins as defence proteins was obtained when some of these proteins were specifically expressed under biotic and abiotic stress conditions. As an example, the leaves of tobacco (*Nicotiana tabacum*) were found to express a lectin called Nictaba only when treated with jasmonic acid or related metabolites. The level of jasmonic acid is increased in leaf tissues after insect herbivory. Nictaba is a chitin binding lectin which shows significant homology with Cucurbitaceae phloem exudate lectins.

Table 1.1: Lectin families with established insecticidal activity [Vandenborre et al., 2011].

Lectin family	Lectins*	Pest insect
Monocot mannose binding lectins	GNA, APA, LOA, ASAL, ACA	<i>Maruca vitrata</i>
		<i>Sitobion avenae</i>
		<i>Rhopalosiphum maidis</i>
		<i>Nilaparavata lugens</i>
		<i>Nephotettix virescens</i>
		<i>Sogatella furcifera</i>
		<i>Lipaphis erysimi</i>
		<i>Eoreuma loftini</i>
		<i>Lacanobia oleracea</i>
		<i>Meligethes aeneus</i>
Legume lectins	PSA, Con A	<i>Leptinotarsa decemlineata</i>
Hevein-related lectins	WGA, OSA, UDA	<i>Tarophagous proserpina</i>
		<i>Callosobruchus maculatus</i>
		<i>Diabrotica undecimpunctata</i>
Amaranthins	ACA	<i>Ostrinia nubilalis</i>
Jacalin like lectins	Heltuba, HFR-1	<i>Aphis gossypii</i>
Nictaba related lectins	Nictaba, PP2	<i>Myzus persicae</i>
		<i>Drosophila melanogaster</i>
		<i>Spodoptera littoralis</i>
		<i>Manduca sexta</i>
		<i>Acyrtosiphon pisum</i>
		<i>Myzus persicae</i>

*Please see the list of abbreviations for full names.

It is found in the nucleus and cytoplasm of leaf parenchyma-cells and orthologues of Nictaba have been detected in the transcriptome of some other species such as *Capsicum annuum*, *Solanum tuberosum* etc [Van Damme et al., 2004; Vandenborre et al., 2011].

Similarly, a mannose-specific jacalin related lectin was detected in rice (*Oryza sativa*) when the plant was under salt or drought stress, or treated with hormones like jasmonic acid and abscisic acid. Feeding by larvae of Hessian fly also induced the expression of a mannose-specific jacalin-related lectin known as Hessian fly responsive protein 1 (HFR1) in wheat (*Triticum aestivum*) plants [Lannoo and Van Damme, 2014]. These evidences further strengthen the hypothesis that lectins are involved in the defence system of plants endogenously.

1.2.7 Jacalin-related lectins

Jacalin related lectins (JRLs) are a group of proteins that share high sequence identity with jacalin (see Table 1.2), the first lectin identified in this group from the seeds of jack fruit. This family of lectins can be further classified into two groups according to carbohydrate specificity: galactose-specific lectins (gJRLs) or mannose-specific lectins (mJRLs). gJRLs are a small subfamily of lectins synthesised as a preproprotein on the endoplasmic reticulum which subsequently undergoes co- or post-translational modifications and is directed into the vacuoles. The preproprotein first loses the signal peptide and is then cleaved into two polypeptide chains to form the protomer. The larger polypeptide chain also undergoes simultaneous glycosylation. The small polypeptide chain consisting of about 20-24 amino acids is non-covalently linked to the large polypeptide chain with 133 amino acids. These two chains together create a single protomer of a gJRL which form a homotetrameric structure by non-

covalent association. Therefore, gJRLs are tetrameric, glycosylated vacuolar proteins with double chain protomers. In contrast, mJRLs are devoid of a signal peptide and function as a single protomer. These proteins are located in the cytoplasm or nucleus and are not glycosylated. Most mJRLs are homodimers; however, Heltuba is an octamer and artocarpin forms a tetrameric structure [Raval et al., 2004; Van Damme et al., 2002].

1.2.7.1 Carbohydrate specificity of JRLs

Jacalin and other gJRLs including *Maclura pomifera* agglutinin (MPA) and MornigaG can bind to the Thomsen-Friedenreich antigen (T-antigen) with high affinity [Lee et al., 1998; Sastry et al., 1986; Singh et al., 2007]. T-antigen contains the disaccharide, Gal β 1-3GalNAc α , and is found to be expressed on the outer cell surfaces of many human carcinomas. When it comes to monosaccharide binding, gJRLs show specificity towards the α -anomeric form of galactose and its derivatives. The affinity of jacalin towards methyl- α -galactopyranoside is almost 50 times higher than simple galactose [Shankarnarayan et al., 1996]. However, a surface plasmon resonance (SPR) study carried out with jacalin also showed that jacalin can accommodate methyl- α -mannopyranoside, glucose, mannose, N-acetylneuraminic acid and N-acetylmuramic acid in its binding site as well, although with a weaker affinity [Astoul et al., 2002]. This polyspecific nature has been observed for MornigaG also, which could bind mannose and glucose in addition to galactose. The polyspecificity of MornigaG was attributed to an extended carbohydrate binding site [Rouge et al., 2003]. mJRLs show primary specificity towards mannose and its derivatives; however, minor differences in their oligosaccharide specificity can be observed [Astoul et al., 2002]. The mannose counterpart of MornigaG, MornigaM also shows polyspecificity by binding to galactose [Rouge et al., 2003].

Table 1.2: List of JRLs found in the sequence database, using α -chain of jacalin as template.*

Plant name	Accession number	No. of JRL domains	Carbohydrate specificity	Common name	Sequence length
<i>Artocarpus integrifolia</i> (Jackfruit)	P18670–1JAC Q7M1T4 1J4S	1 1 1	Galactose Mannose Mannose	Jacalin KM+ Artocarpin	133 149 149
<i>Arabidopsis thaliana</i> (thale cress)	NP_849691	3	Mannose	–	595
<i>Brassica napus</i> (Swedish turnip)	P93658	4	-	–	680
<i>Castanea crenata</i> (Japanese chestnut)	AAG40322	2	Mannose	–	309
<i>Convolvulus arvensis</i> (field bindweed)	AAG10403	1	Mannose	Conarva	152
<i>Helianthus tuberosus</i> (Jerusalem artichoke)	AAM12553	1	Mannose	Heltuba	147
<i>Ipomoea batatas</i> (sweet potato)	T10940	1	-	Ipomoelin	154
<i>Musa acuminata</i> (banana)	AAM48480	1	Mannose	BanLec	141
<i>Maclura pomifera</i> (Osage orange)	P18674–1JOT	1	Galactose	MPA	133
<i>Morus nigra</i> (black mulberry)	AAL09163 AAL10685	1 1	Galactose Mannose	MornigaG MornigaM	216 161
<i>Calystegia sepium</i> (hedge bindweed)	AAC49564	1	Mannose	Calsepa	153

*[Raval et al., 2004]

1.2.7.2 Structural features of JRLs

JRLs are characterized by a unique subunit fold known as type 1 β -prism fold. This type of fold is made up of three four-stranded β -sheets organized in a prism with a three-fold symmetry parallel to the sheets [Raval et al., 2004; Shankarnarayanan et al., 1996]. Despite their similarity in subunit fold and overall carbohydrate specificity the quaternary associations in JRLs vary significantly. The basis of carbohydrate binding and specificity in JRLs is dictated by three loop regions labelled as binding loops 1 (BL1), 2 (BL2) and 3 (BL3) which correspond to amino acid residues 47-50, 75-78 and 121-125, respectively, in the α -chain of jacalin and other JRLs. The conformation of BL3 in all the known crystal structures of JRLs (jacalin, MPA, artocarpin, Heltuba and calsepa) is found to be constant, whereas the conformation of BL1 and BL2 can vary largely. A glycine (Gly121) and an aspartic acid (Asp125) of BL3 in jacalin are found to be conserved in all JRLs and are crucial for carbohydrate binding. In fact, BL3 forms the binding pocket in all JRLs irrespective of their carbohydrate specificity. On the other hand, the amino acid residues in BL1 and BL2 are not conserved throughout but they determine carbohydrate specificity in JRLs. For example, the hydrophobic and aromatic residues of BL1 in jacalin, MornigaG and MPA make direct contacts with galactose, whereas in mJRLs (e.g. artocarpin, banana lectin, Heltuba etc.) BL1 does not participate in carbohydrate binding. BL2 is involved in making stacking interactions with the carbohydrate moiety via its Tyr78 residue (which could be replaced by phenylalanine or histidine in some JRLs) as in jacalin. In addition to this, residues corresponding to the positions 20-23 in jacalin form the influential loop (IL) which is an indirect determinant of carbohydrate specificity as well. IL is longer in gJRLs than in mJRLs. Therefore, the factors that give rise to different carbohydrate specificities in gJRLs and mJRLs can be identified as posttranslational cleavage, the aromatic stacking formed by

residues in BL2, the hydrophobic residues in BL1 and the presence of IL (see Fig. 1.2) [Raval et al., 2004].

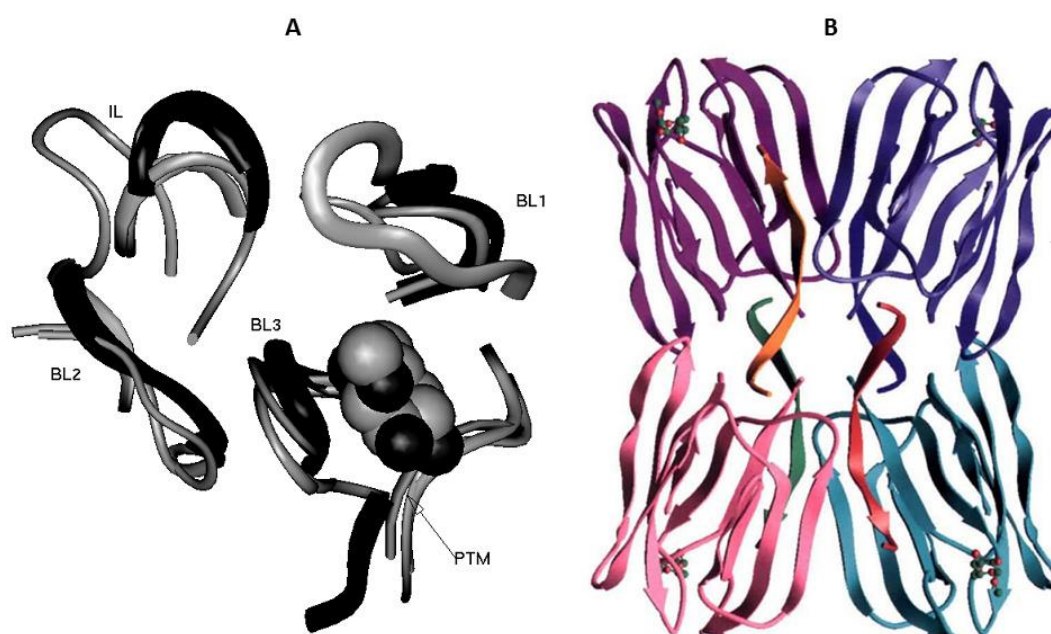


Fig. 1.2: (A) Schematic representation of the disposition of loop regions (BL1, BL2, BL3 and IL) important for carbohydrate binding in JRLs whose structures have been superposed using all C α atoms. The sugar is shown as a CPK object and PTM refers to the site of proteolytic cleavage in gJRLs. The darker shade segments correspond to jacalin and MPA. (B) Overall tetrameric structure of jacalin [Raval et al., 2004, Sankarnarayanan et al., 1996].

1.2.7.3 Applications of jacalin

Jacalin has gained a large amount of attention for its diverse applications in biomedical field. First and foremost its ability to bind T-antigenic disaccharide as mentioned earlier (Section 1.2.7.1) with very high affinity. Jacalin has been used as a marker in detecting benign and malignant lesions of the thyroid and the breast, in monitoring the progression of cervical intraepithelial neoplasia and also in distinguishing benign and malignant cells in serous

effusions [Kabir, 1998]. Jacalin can selectively bind to immunoglobulin A1 (IgA1) in human serum, hence it has been used as a tool to isolate IgA1 [Kabir, 1998]. Jacalin has also been used in AIDS research. It shows anti-HIV-1 properties by inhibiting in-vitro HIV-1 infection [Favero et al., 1993]. Jacalin can also selectively induce proliferation of human CD4⁺ T-cell lymphocytes which act as a receptor for HIV-1. Therefore, the percentage of CD4⁺ T-cell lymphocytes was used to be detected by jacalin in HIV-1 infected patients. Apart from that, jacalin can be used in identifying a number of O-linked glycoproteins including human plasminogen, bovine coagulation factor X, bovine protein Z or fetuin, all of which contain the core disaccharide Gal β 1-3GalNAc. Other than jacalin, mornigaG has been shown to be a very efficient targeting agent towards leukemic cells in photodynamic therapy when attached covalently to white light sensitized porphyrins [Evangelio et al., 2011; Poiroux et al., 2011a; Poiroux et al., 2011b].

1.2.8 Applications of lectins in biomedical and biological sciences

From their early detection time to date lectins have been used for various applications including blood typing, mitogenic stimulation of lymphocytes, bone marrow transplantation and most importantly the identification and purification of glycoconjugates and polysaccharides by affinity chromatography on immobilized lectin matrices [Sharon and Lis, 2003]. Lectins are regarded as invaluable tools for deciphering the glycode hidden in the glycome, which can give substantial amount of information on human physiological processes. Presently, lectins are used as tools in various techniques including lectin blotting, lectin microarray, lectin-based biosensors as well as in targeting tumour markers, some of which are discussed below.

1.2.8.1 As tools in targeting tumour markers

Progression of cancer is associated with aberrant changes in glycosylation pattern of cell surface glycoconjugates. Glycans participate in numerous biological processes that are involved in cancer progression including cell-cell adhesion, inter- and intra-cellular signalling and inflammation. Expression of certain glycans is associated with certain types of cancer cells and hence these glycan structures can serve as important biomarkers for diagnostic purposes [Pinho and Reis, 2015; Varki et al., 2009]. Lectins, in view of their ability to recognize carbohydrate structures with high affinity, can be used as tools to target these biomarkers for the early detection of tumour metastasis. The binding of *Helix pomatia* agglutinin (HPA), an N-acetylgalactosamine-specific lectin was shown to be associated with local axillary lymph-node metastasis of breast cancers as well as metastasis of lung and prostate cancers. *Sambucus nigra* agglutinin (SNA) and peanut agglutinin (PNA) have been used in the detection of breast cancer biomarker CA 15-3 in combination with antibodies [Dan et al., 2016].

1.2.8.2 Lectin microarray for glycan profiling

As mentioned previously (Section 1.1), cell surface glycoconjugates play vital roles in many biological communication processes and the study of the glycans associated with these glycoconjugates has given rise to a whole new field in science called glycomics. So far, high resolution mass spectrometry has been extensively used for glycomic analyses, however, the challenges encountered in mass spectrometry are with low abundant glycoproteins, the intensities of which are often overshadowed by the high abundant species precluding their detection by this method. This is why mass spectrometry often requires the liberation of glycans from the glycoproteins by enzymatic procedures [Varki et al., 2009]. Also, a complex glycoprotein sample with various glycan subtypes cannot be analysed in one mass spectrometric experiment. Therefore, to overcome these difficulties recently a new technique

called lectin microarray has emerged by which a complex set of glycans can be identified in just one experiment without the need for liberating the glycans. Lectin microarray employs lectins mostly of plant origin, which recognize complex glycans with high specificity and affinity [Gupta et al., 2010]. In this method, a panel of lectins are immobilized on a glass chip by covalent linkages. Then fluorescently labeled glycoprotein samples from cell lysate or body fluids are incubated with the lectin array. The resulting fluorescence intensity can be measured either by a confocal-type fluorescence scanner or based on evanescent-field activated fluorescence detection method on each lectin spot (see Fig. 1.3) [Hirabayashi et al., 2013; Pilobello et al., 2005]. Lectin microarray is a rapid and easy technique to identify the types of glycan epitopes present even in the crude samples containing glycoproteins; however, as in mass spectrometry, detailed structural information are not obtained by this technique. One more limitation stems out of the lectins themselves. There are only a few lectins that are commercially available while there is a growing need for careful characterization of many more lectins in terms of their oligosaccharide specificity.

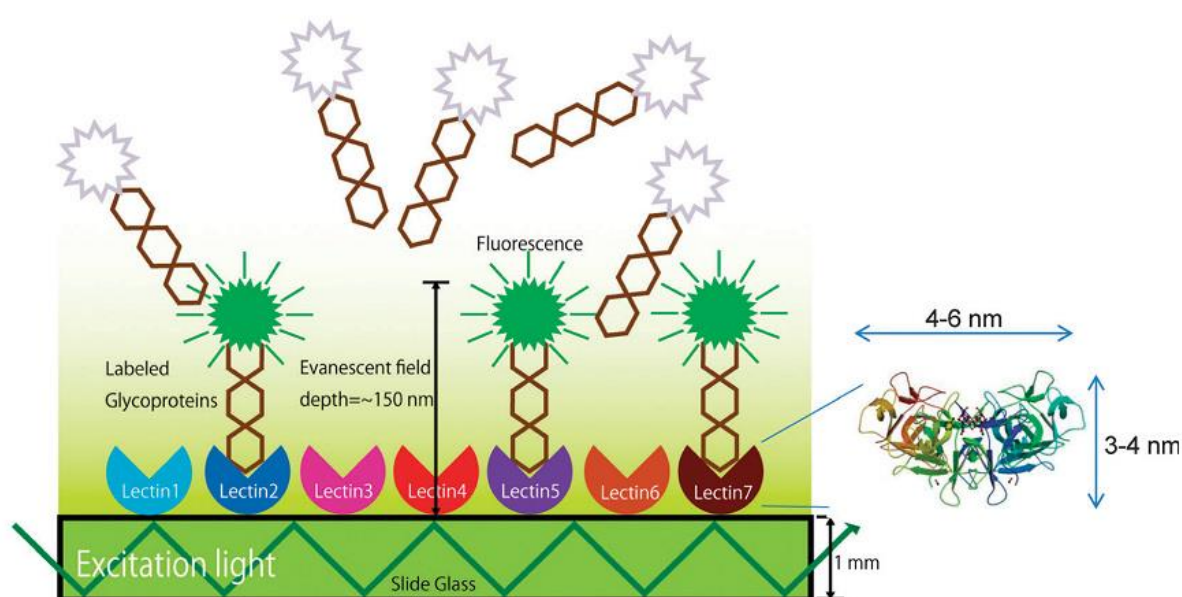


Fig. 1.3: Lectin microarray based on evanescent-field activated fluorescence detection procedure [Dan et al., 2016].

1.3 The catalytic triad formed by serine proteases

Serine proteases are a functionally diverse family of proteolytic enzymes that catalyse the hydrolysis of peptide bonds using catalytic serine residue. These enzymes are involved in many biological processes including digestion (trypsin), blood coagulation (factors VIIa, IXa, Xa, XIIa), immune responses (complement factors B, C, D), fibrinolysis (kallikrein) and reproduction (acrosin) [Polgar, 2005]. The hydrolysis of the peptide bond takes place via an addition-elimination reaction involving a tetrahedral intermediate. All the serine proteases act by following a similar mechanism where the active site consist of a nucleophile (in this case OH group of Ser), a general base and an acid. In trypsin and subtilisin families of serine proteases the active site is comprised of serine, histidine and aspartic acid residues known as the Ser-His-Asp catalytic triad (Fig. 1.4). The carbonyl carbon of the substrate peptide bond (P1 – P1') is first attacked by serine OH to form a tetrahedral oxyanion intermediate. This nucleophilic attack is facilitated by a histidine residue which deprotonates the OH group. The subsequent expulsion of the newly formed amino group is aided by a proton donation from histidine and the concerted deacylation of the acyl-enzyme complex takes place via another nucleophilic attack by a water molecule. The Ser and His residue form a strong hydrogen bond with each other to facilitate proton transfer after the substrate is bound. The aspartate residue stabilizes the charge on imidazole in the transition state before forming a tetrahedral intermediate [Perona and Craik, 1995; Polgar, 2005]. The replacement of this aspartate residue (Asp102) with neutral asparagine in trypsin was found to decrease the rate of catalysis by several fold [Craik et al., 1987].

Despite sharing a similar catalytic mechanism serine proteases show remarkable substrate specificity. Trypsin cleaves the scissile peptide bond preferentially at the C-terminal of Arg or Lys (P1) whereas chymotrypsin cleaves next to an aromatic amino acid residue like Trp, Tyr or Phe (P1) [Kraut, 1997]. This specificity is determined by the amino acid residue present in the S_1 pocket, a crevice on the enzyme surface where the substrate P1 residue fits in. In trypsin an aspartate residue (Asp189, see Fig. 1.4) in this pocket forms network of hydrogen bonds with the guanidinium group of P1-Arg [Czapinska and Otlewski, 1999].

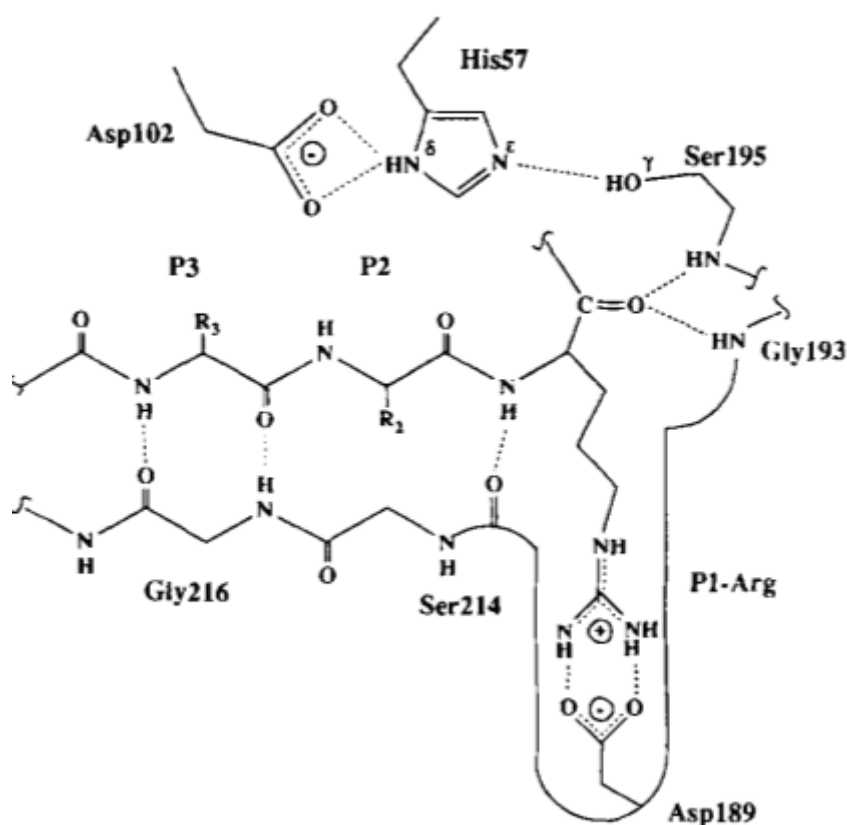


Fig. 1.4: Interaction of the catalytic groups of trypsin (Ser195, His57 and Asp102) with an oligopeptide substrate binding to P1-P4 sites. (Nomenclature for the substrate amino acid residues is $P_n, \dots, P_2, P_1, P_1', P_2', \dots, P_n'$, where P_1-P_1' denotes the scissile/reactive peptide

bond. $S_n, \dots, S_2, S_1, S_1', S_2', \dots, S_n'$ denote the corresponding enzyme binding site). [Perona and Craik, 1995]

Serine proteases are ubiquitously distributed in nature and depending on the structural fold and catalytic residues they can be classified into several subgroups again [Page and Cera, 2008]. In order to regulate the functions of these enzymes nature has devised another class of proteins known as protease inhibitors which have been discussed about in the following sections.

1.3.1 Plant protease inhibitors

Plant protease inhibitors (PIs) are a class of ubiquitous proteins that function primarily to regulate or prevent proteolysis. In plants, these proteins are also induced in response to wounding and insect herbivory. PIs are regarded as the vital defence components of plants [Laskowski and Kato, 1980; Ryan, 1990]. Every PI has at least one reactive peptide (scissile) bond on the surface designated as P1-P1' which interacts with the active site of its target protease. The hydrolysis of this peptide bond in presence of the protease is extremely slow and hence the system behaves like the free PI and the proteases are in equilibrium with the protease/PI complex [Laskowski and Kato, 1980]. Some PIs can contain more than one reactive peptide bond; in that case two protease molecules of same or different type are inhibited at the same time. Depending upon the number of reactive sites, PIs can be termed as single or double-headed inhibitors. The specificity of a PI towards its cognate enzyme is determined largely by its P1 amino acid residue. For trypsin inhibitors, P1 is Arg or Lys, for elastase inhibitors P1 is Ala or Ser and for the inhibitors of chymotrypsin P1 can be any of the following residues: Phe, Tyr, Leu, Met or Trp [Oliva et al., 2010]. Plant PIs are broadly classified into four groups

according to the mechanistic class of proteases such as serine, cysteine, aspartic acid and metallo protease inhibitors, among which serine protease inhibitors are most commonly found in plants. The serine protease inhibitors of plants can be further divided into eight subfamilies based on sequence and structural similarity as summarized in Table 1.3 [Garcia-Olmedo et al., 1987]. Soybean trypsin inhibitor (Kunitz) as well as Bowman-Birk inhibitor (BBI) families are the most studied among all families of inhibitors. A detailed discussion on these two families is presented below.

Table 1.3: Classification of serine protease inhibitors in plants [García-Olmedo et al., 1987].

Inhibitor family	Enzymes inhibited
1. Soybean trypsin inhibitor (Kunitz) family	Serine proteases and endogenous α -amylases
2. Bowman-Birk inhibitor family	Serine proteases
3. Cereal trypsin/a-amylase inhibitor family	Serine proteases and heterologous α -amylases
4. Potato inhibitor I family	Serine proteases
5. Potato inhibitor II family	Serine proteases
6. Squash inhibitor family	Serine proteases
7. Barley protein Z/ α_1 -anti trypsin family	Serine proteases
8. Ragi I-2/maize bifunctional inhibitors family	Serine proteases and heterologous α -amylases

1.3.1.1 Soybean trypsin inhibitor (Kunitz) family

The first isolation and crystallization of a proteinaceous inhibitor of trypsin was reported by Kunitz back in 1945 from the seeds of soybean (*Glycine max*) [Kunitz, 1946]. The soybean Kunitz trypsin inhibitor (SKTI) is a protein comprising 181 amino acid residues and two disulphide bonds. It shows high inhibitory activity towards trypsin ($K_i = 0.01$ nM) and

relatively weak inhibitory activity towards chymotrypsin ($K_i = 1 \mu\text{M}$). SKTI interacts with trypsin and chymotrypsin via its active site peptide bond harbouring Arg63-Ile64 at position P1-P1', respectively [Mosolov and Valueva, 2005].

Several other inhibitors belonging to this family have been isolated and characterized predominantly from leguminous plants, most of which inhibit either trypsin or chymotrypsin forming a 1:1 complex. Some inhibitors are specific only for trypsin or chymotrypsin without having any (or weak) specificity for the second enzyme while some show equal specificity towards both the enzymes. Apart from trypsin and chymotrypsin, this particular family of PI can inhibit the activity of elastase, subtilisin and α -amylase as well [Laskowski and Kato, 1980]. The first cereal inhibitor specific for subtilisin and endogenous α -amylase 2 was identified from barley (*Hordeum vulgare* L.). Later similar bifunctional inhibitors were identified from rice, wheat and rye (*Secale cereale*), all of which did not show any specificity towards the α -amylase 1 isozyme [Garcia-Olmedo et al., 1987; Mundy et al., 1983].

Similar to SKTI, majority of Kunitz protease inhibitors are single chain proteins with a M_r of ~20 kDa and two intramolecular disulphide bonds. However, two-chain inhibitors connected by a disulphide bonds have been found in the subfamily Mimosaceae of leguminous plants. In these proteins the single polypeptide chain undergoes proteolytic cleavage to form ~15 kDa and ~5 kDa subunits, where the larger subunit harbours the reactive site [Mosolov and Valueva, 2005].

The secondary structure of Kunitz protease inhibitors contains predominantly β -sheets and is devoid of α -helices. The structure of SKTI is composed of 12 antiparallel β -strands attached by long loops. Six of these strands form a short β -barrel while the other six form a lid

closing the barrel. The overall structural fold of SKTI is termed as β -trefoil fold (Fig. 1.5) which is also observed in many other Kunitz trypsin inhibitors including a trypsin-inhibitor from *Erythrina caffra* (DE-3) and a Kunitz-type chymotrypsin inhibitor from the seeds of winged bean (WCI). This structural fold is also observed in some non-homologous proteins including ricin B-chain and interleukin 1 α [Song and Suh, 1998]. The scissile peptide bond (Arg63-Ile64) in SKTI is present at one end of the structure on a bulging loop [Oliva et al., 2010; Song and Suh, 1998].

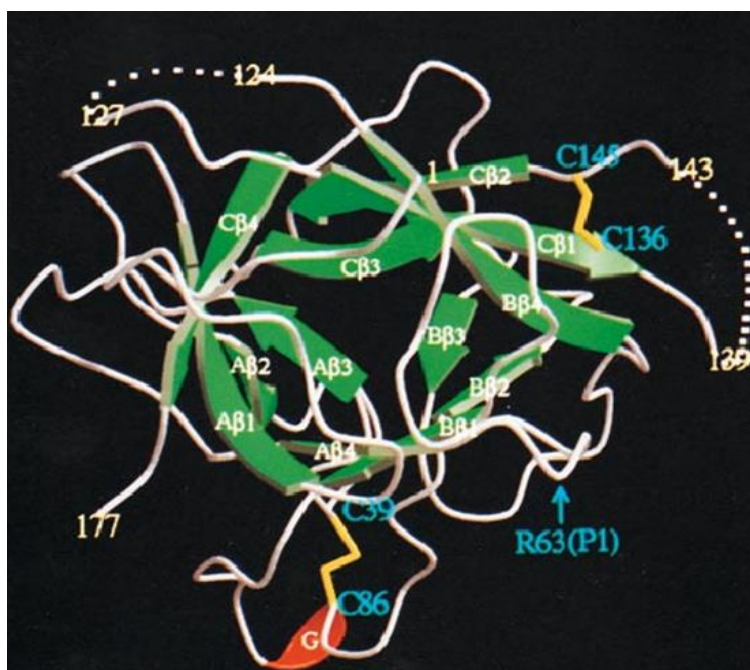


Fig. 1.5: Overall structure of SKTI shown as ribbon diagram. Green arrows, a red ribbon and white ropes correspond to β -strands, a 3_{10} -helix and the loops, respectively. The P1 residue Arg63 and two disulphide bonds (C39-C86 and C136-C145) are labeled [Song and Suh, 1998].

Protease inhibitors are products of multigene families and hence it is common to have a number of isoforms of a particular PI. There are at least 10 genes identified expressing SKTI at different parts of the plant in the genome of soybean. These isoforms can sometimes differ

in their enzyme specificity. For instance, one (DE-1) of the two isoinhibitors present in the seeds of *Erythrina latissima* L. inhibits only chymotrypsin while the other one (DE-3) is a specific trypsin inhibitor [Mosolov and Valueva, 2005; Ryan, 1990].

1.3.1.2 Bowman-Birk inhibitor (BBI) family

Bowman-Birk family of inhibitors are a group of small proteins of Mr 6-9 kDa with seven disulphide bridges and with the ability to inhibit two serine proteases, trypsin and chymotrypsin, simultaneously. The double-headed nature of these proteins is attributed to the presence of two homologous domains, a trypsin-specific N-terminal domain and a chymotrypsin-specific C-terminal domain [Garcia-Olmedo et al., 1987].

The first inhibitor in BBI family was also identified in soybean seeds by Bowman [Bowman, 1946] and later detailed characterization was done by Birk et al [1961]. The BBI in soybean is a 71 amino acid protein and one reactive site (Lys16-Ser17) at the N-terminal side is active against trypsin while another reactive site (Leu43-Ser44) at the C-terminal side is active against chymotrypsin. These two reactive sites are independent of each other and the inhibitor forms a ternary complex with both the enzymes [Birk, 1985]. The structure of soybean BBI is comprised of a similar β -hairpin motif in each of its enzyme binding domain and a short segment of antiparallel β -sheet [Werner and Wemmer, 1991]. In an experiment where the two binding domains were separated by treatment with pepsin and cyanogen bromide, the two individual fragments were still found to be active against the respective enzymes. This observation led to the idea that these double-headed inhibitors have acquired their high stability by duplicating an ancestral single-headed structure [Birk, 1985].

BBI family of inhibitors have been detected in the cereal grains as well which can be divided into two groups. Group I inhibitors are 16 kDa proteins with four consecutive enzyme binding domains whereas group II inhibitors are 8 kDa proteins and contain five disulphide bonds and a single reactive site in contrast to leguminous BBIs. Thus, group I inhibitors of BBI family in cereal grains appear as a duplicated form of group II inhibitors [Mosolov and valueva, 2005].

1.3.3 PIs are plants' natural defence components

Plants are under constant attack by various external factors among which insect herbivory is particularly challenging. Plant PIs are induced in response to insect herbivory and hence are regarded as the plants' natural defence components. The first such evidence was obtained when the leaves of tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) accumulated a large amount of PIs when attacked by Colorado potato beetles [Kessler and Baldwin, 2002; Zhu-Salzman and Zeng, 2015]. The effects of many PIs were also found to be detrimental when some insects were supplied with these PIs in their diet. For instance, Bowman-Birk soybean trypsin inhibitor inhibited the growth of meal worm larvae at 5% of the diet while SKTI was less effective when supplied at the same level [Lipke et al., 1954]. The midgut of insects, especially in the order of Lepidoptera, produces mostly serine proteases which breakdown proteins into amino acids that are crucial for the growth and development of the insect. Plant PIs can inactivate these proteases, resulting in severe amino acid deficiency which leads to subsequent retardation in the growth and reproduction [Ryan, 1990]. These proteolytic enzymes in insects can therefore be potential targets for pest management in agriculture. Several attempts have been made to incorporate the genes of plant PIs to generate pest resistant crops. When the Bowman-Birk trypsin inhibitor gene from cowpea was expressed in transgenic

tobacco plants, an increased resistance towards feeding by larvae of *Heliothis virescens* was observed [Hidler et al., 1987]. Later many transgenic crops including rice, wheat and cotton harbouring foreign PI genes were developed. However, none of these transgenic plants were made commercially available as the insects developed resistance towards these foreign genes under selection pressure [Zhu-Salzman and Zeng, 2015].

The detrimental effect of plant PIs was not only observed towards insect proteases but also towards proteases of microorganisms. The activity of extracellular proteases of the phytopathogenic fungus *Fusarium solani* was inhibited by trypsin and chymotrypsin inhibitors from the seeds of soybean, garden bean and the tubers of potato [Mosolov and Valueva, 2005]. A Kunitz-protease inhibitor (AFP-J) from the tubers of potato also suppressed the activity of yeast fungal strains including *Saccharomyces cerevisiae*, *Candida albicans* and *Trichosporon beigelii* [Park et al., 2005]. *Candida albicans* can cause vaginal, urinary, oral or oesophageal infections in humans. Therefore, apart from being plant defence components plant PIs can be exploited for their potential therapeutic use as well.

1.4 Motivational aspects and outlook of the present work

Lectins are invaluable tools in glycomic research. The unique ability of lectins to recognize simple carbohydrates or even complex glycan structures with high specificity has attracted the attention of glycobiologists for a very long time. Plant lectins are of particular interest due to their high abundance, ease of purification as well as diverse structure and carbohydrate specificity. Among different families of lectins legume lectins have been explored in great detail while other families such as jacalin-related lectins (JRLs) or phloem exudate lectins from cucurbits are known to a lesser extent. Within JRL family, the number of known gJRLs is very

small. A detailed knowledge of structure, function and carbohydrate specificity is a prerequisite in order for these lectins to be exploited in glycomic and biomedical research. Therefore, purification and biophysical characterization of new lectins from this family are of utmost importance. Mulberry (Moraceae family) is an economically important plant which is widely cultivated in India for the production of silk. Silk is produced from the cocoons of the insect, *Bombyx mori* larvae which feed on mulberry leaves. Latex is the cytoplasm of a special type of elongated secretory cells known as the laticifer, found in the leaves and stems of plants. It is generally considered to have a role in the defence against insect herbivory.

In this study, we have purified a gJRL from the latex of mulberry (*Morus indica*) and carried out biophysical investigations on its carbohydrate specificity, secondary structure and thermal stability. The primary and glycan structures of the protein were determined by nanoESI-Q-ToF mass spectrometry. The cytotoxicity of the lectin was assessed towards mammalian epithelial and cancer cell lines using MTT cell viability assays. Further, the accessibility and exposure of the Trp residues as well as chemical, acidic and thermal unfolding were investigated by fluorescence and circular dichroism spectroscopy. The results of these studies are presented in chapters 2-4.

Kunitz protease inhibitors are essential components of plant defence system. The protease inhibitors from leguminous plants have been studied extensively while there are only few reports on the protease inhibitors from mallow (Malvaceae) family. Okra belonging to this family is a common vegetable crop and has a number of medicinal properties. The fruits, seeds, pods and leaves of okra have been shown to possess antidiabetic, antihyperlipidemic, antioxidant properties and can also be used for the treatment of dysentery and diarrhoea. In

chapter 5 the purification, characterization as well as primary structure of a protease inhibitor from the seeds of okra have been reported.

Biophysical Characteristics and Cytotoxicity of an Affinity-purified α -D-Galactose-specific Jacalin-related Lectin from the Latex of Mulberry (*Morus indica*)

Chapter 2

2.1 Summary

Jacalin related lectins (JRLs) comprise a family of lectins which can bind to the tumor-associated T-antigenic disaccharide, Gal β 1-3GalNAc with high affinity. In this study, an α -D-galactose specific lectin belonging to the family of JRLs has been purified by affinity chromatography on cross-linked guar-gum. Hemagglutination-inhibition assays showed that galactose is a poor inhibitor of its lectin activity while Me α Gal and other α -anomeric derivatives of galactose can inhibit it readily. Circular dichroism spectroscopy indicated that the secondary structure of the lectin is predominantly made up of β -sheets, and differential scanning calorimetry revealed a thermal denaturation temperature of 77.6°C. MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) cell viability assays on MCF-7 and MDCK cells showed that the lectin is highly cytotoxic towards both cell lines when dosed at micromolar concentrations, suggesting that it may play a role in the defense mechanism of the plant.

2.2 Introduction

Lectins are a special class of proteins which recognize mono- or oligosaccharides in a highly specific manner and bind to them by various non-covalent interactions but are devoid of catalytic activity [Lis and Sharon, 1973]. These proteins are abundantly found in nature and show significant diversity in their structures as well as in carbohydrate specificity. Although in animals lectins are known to participate in many cellular communication processes [Kilpatrick, 2002], their functions in plants are poorly understood. Some lectins are known to be involved in the defense mechanism of plants, most likely by interacting with the cell-surface carbohydrate structures of fungi or insect herbivores [Peumans and Van Damme, 1995], whereas some others accumulate in the vegetative storage organs, e.g. seeds, bulb or bark and have been proposed to function as storage proteins [Van Damme et al., 2002]. Many plant derived lectins can interact with biologically relevant hydrophobic, non-carbohydrate ligands such as adenine, cytokinin and indole acetic acid as well as several water-soluble porphyrins, suggesting that they may also function as carriers of these hydrophobic molecules, some of which act as phytohormones [Bhanu et al., 1997; Komath et al., 2006]. Recognition of a vast array of carbohydrate structures by plant lectins, their relatively high abundance and ease of purification have made them important tools in the field of glycobiology [Ambrosi et al., 2005; Wu et al., 2009].

Jacalin-related lectins (JRLs) are a family of plant lectins which show a characteristic subunit fold, known as type-1 β -prism fold. This type of subunit fold has three four-stranded β -sheets parallel to each other arranged in a prism [Sankarnarayanan et al., 1996]. There can be two types of JRLs based on the difference in carbohydrate specificity: they are either galactose-specific (gJRLs) or mannose-specific (mJRLs). Both mJRLs and gJRLs exhibit

similar subunit folds and share high sequence identity but differ from each other in their structures of protomers and subcellular localization apart from carbohydrate specificity [Van Damme et al., 2002; Raval et al., 2004]. This difference originates due to distinct post-translational modifications. In the bark of black mulberry (*Morus nigra*) the galactose-specific JRL (MornigaG) is a two-chain protomer located in the vacuole while the mannose-specific counterpart (MornigaM) is a cytoplasmic protein containing a single-chain protomer [Van Damme et al., 2002]. gJRLs show a number of interesting properties owing to their carbohydrate specificity. Specific binding to the α -anomeric form of the tumour associated Thomsen-Friedenreich (T/TF)-antigen, Gal β 1-3GalNAc α Me is one of them [Mahanta et al., 1990; Sastry et al., 1986]. The TF-antigen is found to be expressed on the outer cell surfaces of most human carcinomas [Glinsky et al., 2001]. This explains the potential of gJRLs as targeting agents in the detection and development of therapeutics for the treatment of cancer. Apart from that, jacalin, the first gJRL to be identified and characterized extensively, exhibits other interesting properties as well. For instance, it selectively binds to human serum immunoglobulin A1, stimulates T-cell lymphocytes, inhibits HIV-1 infection *in vitro* and also shows strong binding with a variety of water-soluble porphyrins [Kabir, 1998; Komath et al., 2000]. This particular family of plant lectins is known to a lesser extent compared to other families. Over the years the number of known JRLs has increased, but very few of them were characterized properly regarding their structure, carbohydrate-binding specificity and prospective biological function. Therefore, in order to shed light on these aspects it is important to purify and characterize hitherto unknown JRLs.

The present study was aimed at the purification of a gJRL from mulberry (*Morus indica*) latex, which we shall henceforth refer to as MLGL (Mulberry Latex Galactose-specific

Lectin), by affinity chromatography and at the investigation of its secondary structure and thermal stability by circular dichroism spectroscopy and differential scanning calorimetry, respectively. The carbohydrate specificity of the lectin was assessed by performing hemagglutination-inhibition assays. Furthermore, the cytotoxic effect of the lectin was assessed towards MCF-7 and MDCK cells in order to find out its plausible function in latex.

2.3 Materials and Methods

2.3.1 Purification of MLGL

Collection of mulberry latex was performed as according to Patel et al. [2011] with a little modification. Briefly, fresh latex was directly collected by plucking young leaves into ice cold 10 mM Tris-HCl buffer containing 150 mM NaCl (TBS). The collected latex was then kept at -20°C for one day and thawed at room temperature before further processing. The crude latex was centrifuged at 10,000 rpm for 30 min at 4°C and the supernatant obtained was loaded onto a column of cross-linked guar-gum (15×3 cm) pre-equilibrated with TBS at 4°C [Appukuttan et al., 1977]. The unbound proteins were washed with Tris buffer followed by elution of the lectin with 0.2 M galactose. The eluted protein was extensively dialyzed against 20 mM sodium phosphate buffer containing 150 mM NaCl, pH 7.4 (PBS). All further experiments were performed with the protein in PBS unless specified otherwise. Purity of the protein was checked by performing SDS-PAGE according to Laemmli [1970].

2.3.2 Zeta potential measurements

The isoelectric point of the lectin was determined by measuring the zeta potential at different pH by mixed laser Doppler velocimetry and phase analysis light scattering (M3-PALS). A Malvern Zetasizer NanoZS (Malvern Instruments Ltd., UK) fitted with a red laser light ($\lambda =$

632.8 nm) and equipped with a MPT-2 autotitrator was used for this method. A protein solution of 0.1 mg/mL in PBS was titrated with 0.1 M HCl from pH 6.0 to 3.0 with pH change steps of 0.5 ± 0.2 at 25°C. The measurements were carried out in triplicates using disposable folded capillary cells (DTS1070) from Malvern.

2.3.3 Hemagglutination and hemagglutination-inhibition assays

Hemagglutination and hemagglutination-inhibition assays were performed as described previously [Sultan et al., 2009]. In short, the lectin in PBS was serially two-fold diluted in a round bottomed 96-well ELISA plate and an equal amount of a 4% erythrocyte suspension of human A, B, O (+) or rabbit blood prepared in 20 mM PBS was added. The mixture was incubated at 4°C for 1 h and the hemagglutination titre was scored visually. Hemagglutination-inhibition assays were performed in a similar way. Briefly, the sugar solution of interest was serially two fold diluted making 40 μ L as the final volume in each well. Ten μ L of lectin solution was added to each well and the mixture was incubated at 4°C for at least 10 min. Finally, 40 μ L of erythrocyte suspension was added resulting in a total volume of 100 μ L in each well. After 1 h of incubation the hemagglutination titre was scored visually.

2.3.4 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded by use of an Aviv 420 spectrometer (Lakewood, NJ, USA). For far-UV region a sample concentration of 0.1 mg/mL and for near-UV region a sample concentration of 1 mg/mL were used. All the measurements were performed using a rectangular quartz cuvette of 2 mm path length and 700 μ L capacity. Buffer scans were recorded under similar conditions and subtracted from the sample spectra before further analysis.

2.3.5 Differential scanning calorimetry

Differential scanning calorimetry (DSC) experiments were carried out as described earlier [Kavitha et al., 2010] on a differential scanning calorimeter VP-DSC from MicroCal (MicroCal LLC, Northampton, MA, USA). It is equipped with two fixed cells; a reference cell and a sample cell. The protein samples taken in PBS were loaded in the sample cell whereas PBS was loaded in the reference cell. Analyses of the thermograms were performed by a software from Origin.

2.3.6 Cell culture

All the cell cultures were maintained in a 37°C incubator (Sanyo MCO-19AIC, Panasonic Biomedical Sales Europe BV, AZ Etten Leur, Netherlands) containing 5% CO₂ and humid atmosphere. Madine-Darby canine kidney (MDCK) cells were cultured in 75 cm² culture flasks using MEM supplemented with 1% L-glutamine (200 mM), 10% fetal bovine serum and 1% penicillin-streptomycin (10000 units penicillin, 10000 units streptomycin in 0.9% NaCl). Human breast cancer cells (MCF-7) were cultured in the same way but instead of MEM, RPMI medium was used with the same supplements. Once confluence was reached the cells were detached using 0.05% trypsin in EDTA. The number of cells per 1 mL of medium was counted using a counting chamber from Improved Neubauer by diluting the cells 20-fold with trypan blue. The cells were subcultured by splitting at a ratio of 1:10 or seeded at a number of 5000 cells per well in a 96-well cell culture plate.

2.3.7 MTT cell viability assay

Cytotoxicity of MLGL towards MDCK and MCF-7 cell lines was assessed and compared with that of jacalin and wheat germ agglutinin (WGA) colorimetrically using the MTT assay as

described previously [Kaiser et al., 2015]. Briefly, 100 μ L of cell suspension was transferred to each well of a 96-well tissue culture plate (5000 cells/well) and allowed to grow overnight. Hundred μ L of each lectin, diluted in supplemented medium, was added in two desired concentrations (1.25 μ M and 0.625 μ M in tetramer) and the plates were incubated at 37°C for 72 h. After incubation the supernatant was removed from the cells and 100 μ L of fresh medium was added followed by the addition of 25 μ L of 5 mg/mL MTT dissolved in PBS. The plates were incubated for about 4 h and then after removal of the supernatants, 100 μ L of DMSO was added to each well to dissolve the dye. After 15 min of orbital shaking at 300 rpm the absorbance was measured at 570 nm using a microplate reader (Safire, Tecan AG, Salzburg, Austria). A 4% solution of Triton X-100 has used as the positive control and the respective medium has used as the negative control.

2.4 Results

2.4.1 Purification and carbohydrate-binding specificity

MLGL was purified by a simple one-step affinity chromatographic method. Crude latex was collected in ice-cold 10 mM Tris-HCl (pH – 8.0) buffer containing 0.15 M NaCl (TBS), frozen at –20°C for 24 hours and then thawed at room temperature. The sample was subjected to centrifugation and the supernatant was collected and loaded onto a guar-gum column, washed extensively with TBS and then a solution of 0.2 M galactose was used to elute the lectin. Approximately 2 mg of protein was obtained from 60 mg of crude soluble protein fraction (Fig. 2.1). Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) of MLGL revealed a single band at ~16 kDa both under reducing (β -mercaptoethanol was employed as reducing agent) and non-reducing conditions (Fig. 2.2) indicating that no intermolecular disulfide linkages are present in the protein. In gel filtration experiment the elution volume of

native MLGL was comparable to that of bovine serum albumin ($M_r \sim 66$ kDa), suggesting that the lectin is most likely a homotetramer. This is consistent with the observation that jacalin and MornigaG, two other members of the gJRL family that exhibit high sequence homology with MLGL (see Chapter 3) are also homotetramers.

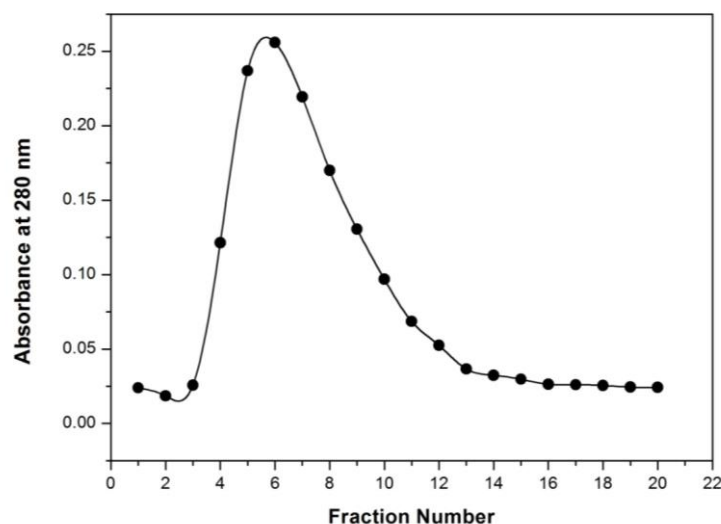


Fig. 2.1. Affinity chromatography of MLGL on cross-linked guar-gum. Fractions of 2 mL were collected.

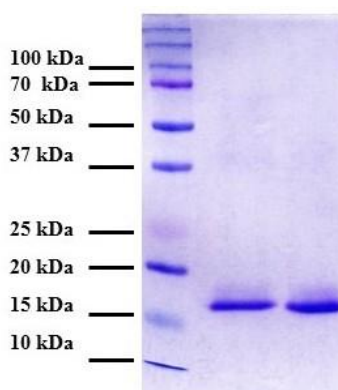


Fig. 2.2. SDS-PAGE of MLGL. Lane 1: molecular weight markers; lane 2: MLGL under reducing condition (with β -mercaptoethanol); lane 3: MLGL under non-reducing condition (without β -mercaptoethanol).

The isoelectric point (IEP) of MLGL was determined as 4.35 ± 0.07 by measurement of the electrophoretic mobility of the protein evaluating the zeta potential at different pH values (cf. Fig. 2.3).

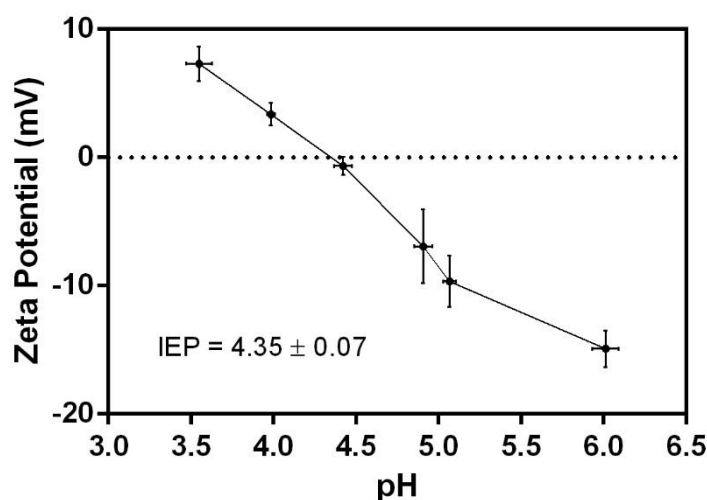


Fig. 2.3. Determination of isoelectric point of MLGL using zeta potential measurements. The protein (0.1 mg/mL in PBS) was titrated with 0.1 M HCl from pH– 6.0 to 3.0 with pH change steps of 0.5 ± 0.2 at 25°C.

MLGL could agglutinate human ABO (+) and rabbit erythrocytes with equal efficiency up to concentrations as low as 625 ng/mL. Galactose has been a poor inhibitor of the hemagglutination activity of MLGL, whereas methyl- α -D-galactopyranoside (Me α Gal, 8 mM) could readily inhibit the activity of MLGL. In contrast, methyl- β -D-galactopyranoside and lactose could not inhibit the hemagglutination activity of MLGL even at 250 mM and 125 mM, respectively, suggesting that the lectin does not recognize the β -anomer of galactose. The preferred stereospecific recognition of galactose in the α -anomeric configuration is further

corroborated by the finding that other α -D-galactose derivatives were able to inhibit the hemagglutination activity of MLGL as well (Table 2.1).

Table 2.1: Inhibition of hemagglutination activity of MLGL by various sugars. Final concentration of MLGL used was 4 μ g/mL.

Sugar ^{1,2}	Minimum Inhibitory Concentration (mM)
4-Methyl-umbelliferyl- α -galactopyranoside	0.045
4-Nitrophenyl- α -galactopyranoside	1.2
Methyl- α -galactopyranoside	8
Melibiose	16
<i>N</i> -Acetyl-galactosamine	32
Galactose	63

¹All sugars are D- sugars.

²The following sugars did not inhibit the activity of MLGL up to the resultant concentrations indicated in the parentheses: methyl- β -galactopyranoside, glucose, mannose, fructose, lactulose, fucose, 2-deoxygalactose, *N*-acetylglucosamine, methyl- α -glucopyranoside, methyl- α -mannopyranoside (all 250 mM), raffinose and lactose (both 125 mM).

2.4.2 Secondary structure

The far-UV CD spectrum of the native MLGL (Fig. 2.4, A) shows a minimum around 220 nm and a maximum around 200 nm indicating the secondary structure of the protein contains predominantly β -sheets. This interpretation was further confirmed by analyzing the CD spectral data available on the website Dichroweb (www.cryst.bbk.ac.uk/cdweb/html/) using the CDSSTR programme. The content of different secondary structural components in MLGL was yielded by this analysis as: 2% α -helix (0% regular and 2% distorted), 34% β -sheet (21%

regular and 13% distorted), 23% β -turns and 39% unordered structures. The calculated fit obtained from this analysis is also shown in Fig. 2.4 (dotted line), which is in good agreement with the experimentally obtained spectrum. The far-UV CD spectrum of MLGL in the presence of 20 mM Me α Gal showed only marginal changes, indicating that its binding to carbohydrates does not significantly alter the protein secondary structure (Fig. 2.4).

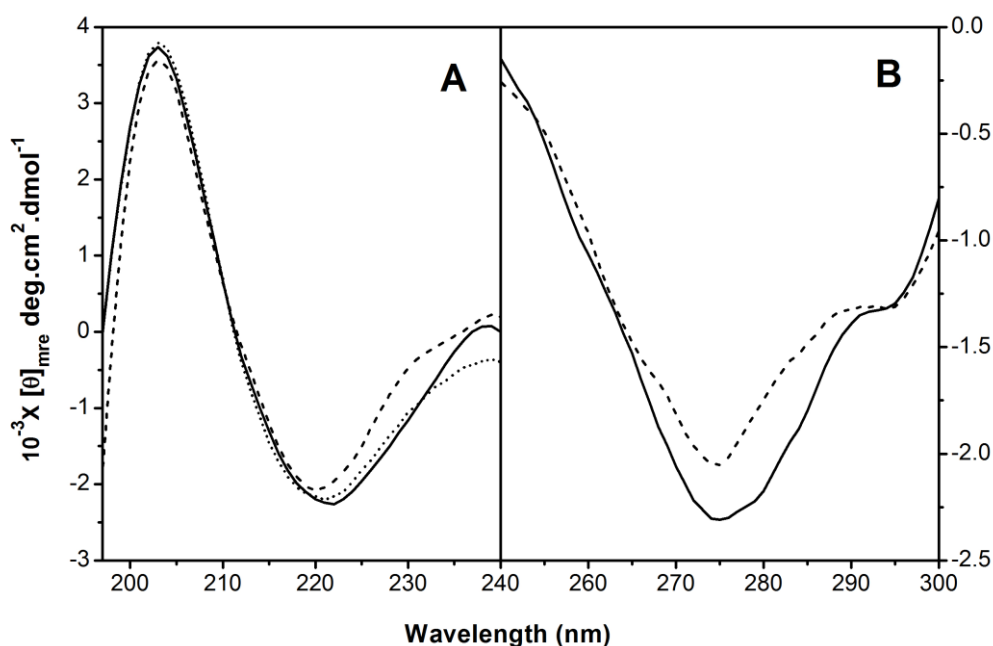


Fig. 2.4. CD spectra of MLGL and analysis by CDSSTR method in the (A) far UV and (B) near UV regions. Solid lines (-) represent experimental spectra, dotted line (....) represents the calculated fit and dashed lines (- - -) represent spectra obtained in the presence of 20 mM Me α Gal.

The CD spectrum of MLGL in the far-UV region is characterized by a minimum at ca. 275 nm and a shoulder at 294 nm (Fig. 2.4, B). In the presence of 20 mM Me α Gal, the CD signal retained its character while decreasing in intensity around 275 nm.

2.4.3 Thermal stability

Thermal stability of MLGL was investigated by differential scanning calorimetry (DSC) at a scan rate of $60^{\circ}\text{C}/\text{h}$. A relatively high scan rate of $60^{\circ}\text{C}/\text{h}$ has been used in order to avoid sample precipitation after transition temperature, which would preclude accurate determination of the transition enthalpy. At pH 7.4, the transition temperature of MLGL was observed to be 77.6°C and the thermogram obtained could be fit to a non-two state model containing a single peak as shown in Fig. 2.5 (thermogram 1). The occurrence of a single peak implies that the protein undergoes denaturation from its intact, folded form to completely unfolded form without having any intermediates. DSC of MLGL in the presence of 100 mM Me α Gal resulted in an increased thermal stability with an unfolding temperature at 79.1°C (Figure 8, thermogram 2). The calorimetric and Van't Hoff enthalpies for respective thermograms are listed in Table 2.2.

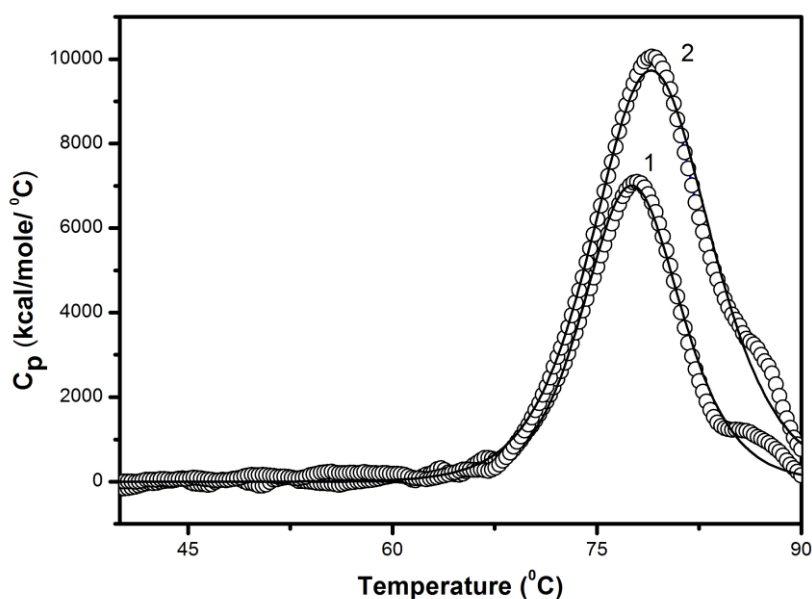


Fig. 2.5. DSC thermograms of MLGL in the absence and presence of 100 mM methyl- α -D-galactopyranoside. Thermogram 1 corresponds to MLGL alone and thermogram 2 corresponds to MLGL + Me α Gal. Data points are shown as open circles and the solid bold line represents

the best fit of DSC data to a non-two state transition model. Concentration of MLGL monomer was 31.2 μM .

Table 2.2. Thermodynamic parameters obtained from DSC studies on native MLGL and in presence of 100 mM Me α Gal.

Sample	T _m (°C)	ΔH_c (kCal/mol)	ΔH_v (kCal/mol)	$\Delta H_c/\Delta H_v$
Native MLGL	77.62	66.46	103.20	0.64
In presence of 100 mM Me α Gal	79.12	113.0	84.94	1.33

2.4.4 Cytotoxicity

Cytotoxicity of MLGL towards an epithelial cell line (MDCK) and a breast cancer cell line (MCF-7) was evaluated colorimetrically by measuring the colour intensity formed due to the reduction of MTT to purple-coloured formazan by cellular oxidoreductase enzymes. Two different concentrations (1.25 and 0.625 μM) of the lectins were used. Triton X-100 (4% w/v) has been used as the positive control. Additionally, the cytotoxicity of two well-known lectins, wheat germ agglutinin (WGA) and jacalin, was also evaluated to compare their effect with that of MLGL on the two cell lines.

WGA is an *N*-acetyl-D-glucosamine and sialic acid-specific lectin which is known for its strong binding ability to cell surface glycoproteins and hence it is used as a molecular probe in conjugation with a fluorophore in fluorescence microscopy [Kostrominova, 2011; Strathmann et al., 2002]. Jacalin, on the other hand, is a gJRL which can be expected to have similar properties as MLGL. The cytotoxic effect on both cell lines was determined after 72 h of incubation and the relative cell viabilities are shown in Fig. 2.6.

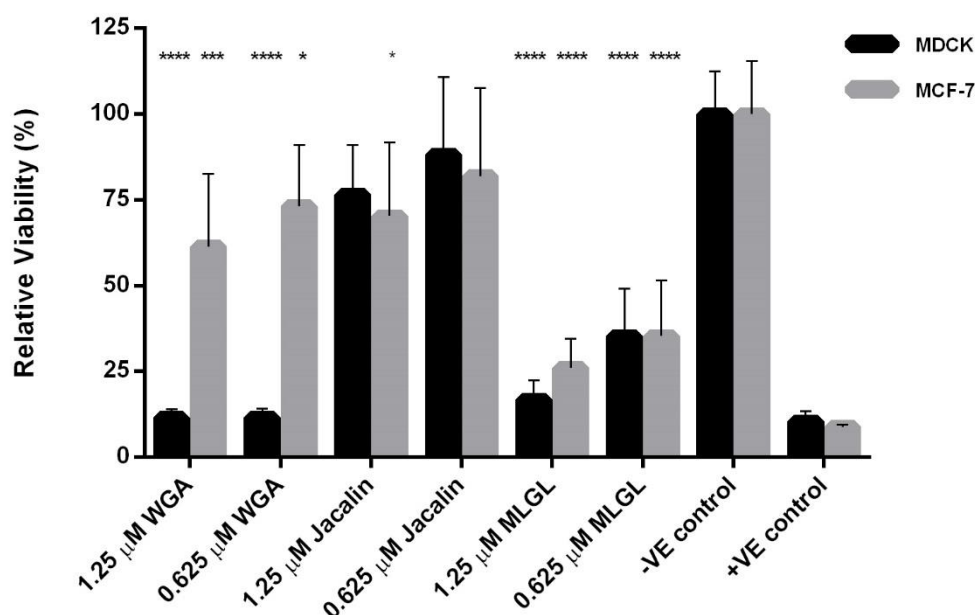


Fig. 2.6. Relative viability of MDCK and MCF-7 cells after treatment for 72 h with two different concentrations of MLGL, WGA and jacalin as determined by the MTT assay ($n = 3$). WGA and jacalin have been used as control lectins. Statistical significance: Kruskal-Wallis test (* $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$). In each case, the difference in relative viability has been compared with the untreated control (negative control). A 4% solution of Triton X-100 was used as positive control.

From the results presented in Fig. 2.6, it is clear that after treatment with the lower concentration ($0.625 \mu\text{M}$) of MLGL the viability for both cell lines reached $\sim 35\%$, whereas the greater dose reduced the viability for MDCK cells even more ($\sim 17\%$ viability) than for MCF-7 cells ($\sim 26\%$ viability). In both cases, the differences are statistically significant compared to the untreated control ($p < 0.0001$, Kruskal-Wallis test). The effect of WGA on the cell viability of MDCK cells is very pronounced ($\sim 12\%$ viability) at both concentrations ($p < 0.0001$, Kruskal-Wallis test), whereas for MCF-7 cells this effect is much less marked ($\sim 50\%$ viability)

($p < 0.001$ for $1.25 \mu\text{M}$ WGA and $p < 0.05$ for $0.625 \mu\text{M}$ WGA). In case of jacalin, by contrast, only a slight reduction in relative cell viability was observed on the two cell lines at both tested concentrations. A significant reduction ($\sim 75\%$ viability) though, was observed for MCF-7 cells at the highest tested dose of the lectin ($p < 0.05$ for $1.25 \mu\text{M}$).

2.5 Discussion

The unique ability of lectins to specifically recognize certain carbohydrates in a noncatalytic manner is consistent with their involvement in many biological phenomena that are mediated by protein-carbohydrate interactions [Lis and Sharon, 1986]. The present study reports the homogenous purification of an α -D-galactose specific lectin from mulberry (*Morus indica*) latex. It is a homotetrameric lectin with an apparent subunit weight of 16 kDa. The lectin has an isoelectric point of 4.35, implying that at physiological pH MLGL exists as a negatively charged entity. Hemagglutination-inhibition assays performed with different sugars showed that the lectin has rather weak affinity towards simple galactose, but its affinity towards Me α Gal is very high. This observation is similar to jacalin which shows almost 50 times higher binding affinity for Me α Gal than simple galactose due to formation of favourable interactions between the methyl group and a tyrosine residue in the binding pocket of the lectin [Sankaranarayanan et al., 1996].

The secondary structure of MLGL was found to consist of mainly β -sheets, with very little α -helical content. This is one of the characteristics of the type-I β -prism fold proteins. Ligand binding does not alter the secondary structure of the protein much. The CD spectrum in the near-UV region is characterized by two minima: one at ca. 275 nm and the other at ca. 294 nm which can be assigned to absorption of Tyr and Trp, respectively. Upon ligand binding,

the CD signal decreases a little without significantly affecting the spectral shape which has also been observed for jacalin and *Maclura pomifera* agglutinin [Young et al., 1989].

MLGL is stable up to 70°C and undergoes a cooperative unfolding transition centred at 77.6°C. The unfolding appears to be irreversible as a repeat scan did not show any transition up to 80°C. In the presence of Me α Gal, the lectin becomes more stable and its denaturation temperature increases by about 1.5°C. This might be due to increase of favourable interactions between the protein and the ligand.

Latex is the milky sap of plants which protects the plants by exuding under mechanical stress or while attacked by insect herbivores [Konno, 2011; Konno et al., 2006]. Previously, it was shown that latex of white mulberry (*Morus alba*) accumulates a large amount of defense proteins related to biotic stress [Kitajima et al., 2013]. Lectins, on the other hand, are proteins which might be involved in the defense system of plants. There are some lectins which act as anti-fungal or anti-bacterial ingredients or possess anti-insect activities [Peumans and Van Damme, 1995]. Some lectins can perhaps interact with the epithelial cells along the digestive tract of phytophagous insects or higher animals and exert their cytotoxicity as a part of the plant's defense mechanisms. In our study the cell viability assays carried out for MLGL exhibited that the lectin is substantially cytotoxic against the epithelial (MDCK) as well as breast cancer cells (MCF-7). Since gJRLs are able to bind to the tumor associated TF-antigen which is also expressed on the cell surface of breast cancer cells [Cazet et al., 2010], the cytotoxic effect might have been exerted due to an interaction between the TF-antigen and the lectin. Similarly, MLGL shows cytotoxic effect towards MDCK cells possibly by interacting with cell surface-exposed glycoconjugates. The cytotoxic nature of the lectin towards epithelial cells might be an indication for its involvement in the defense system by protecting the plant

from phytophagous insects or higher animals. These observations of preliminary experiments are encouraging and provide impetus for further detailed studies in this direction.

Elucidation of Primary and N-glycan Structures of a Jacalin-related Lectin, MLGL by NanoESI-Q-ToF Mass Spectrometry

Chapter 3

3.1 Summary

MLGL (Mulberry latex galactose-specific lectin) is a homotetrameric glycoprotein isolated from mulberry (*Morus indica*) latex. It belongs to the family of jacalin-related lectins (JRLs) and specifically binds to α -D-galactose and its derivatives. High resolution mass spectrometry suggested that the protein harbors two chains like all the members of galactose-specific JRLs (gJRLs). De novo sequencing of proteolytic peptides demonstrated that the heavier chain consists of 133 amino acids and the lighter chain comprises 21 to 24 amino acids. MLGL showed high sequence homology with other gJRLs, such as, MornigaG and jacalin. The heavier chain of MLGL contains one *N*-glycosylation site (Asn₄₇) occupied with either pauci-mannose type [GlcNAc₂(Fuc)Man₃(Xyl)] or complex type [GlcNAc₂(Fuc)Man₃(Xyl)GlcNAc(Fuc)Gal] *N*-glycans. The presence of core β (1,2)-linked xylose residue in the glycan structure of MLGL was further confirmed by performing dot blot assay using a xylose-specific antibody.

3.2 Introduction

Lectin domains are involved in the innate immune system of plants by being receptors to the damage, or pathogen, or microbe-associated molecular patterns (D/P/MAMPs). These molecular patterns are mostly composed of carbohydrates and could originate either from the area of damage/wound (such as cellulose or oligogalacturonides fragments) or directly from invading pathogens (such as lipopolysaccharides of Gram-negative bacteria, peptidoglycan of Gram-positive bacteria) [Lannoo and Van Damme, 2011]. Plants recognize these molecular patterns by use of pattern recognition receptor (PRR) proteins as first line of defence in a series of immune responses, termed as PAMPS-triggered immunity (PTI) [Jones and Dangle, 2006]. A lot of these PRRs have been found to carry lectin domains and exist either in soluble form or membrane-bound state [Schutter and Van Damme, 2015]. Recognition of D/P/MAMPs by PRRs activates the downstream intracellular signalling events in plants which would in turn lead to transcriptional changes, stomatal closure and cell wall reinforcement, ultimately hindering the growth of the pathogen [Schutter and Van Damme, 2015]. Apart from recognizing various carbohydrate stress agents, lectin domains are also induced directly under biotic and abiotic stresses in plants. Oryzata is a jacalin-related, mannose binding lectin which is induced either at high salt concentration or at high jasmonic acid level in the leaves of rice [Lannoo and Van Damme, 2014]. Similarly, involvement of different types of lectin domains in either PRR-based or non-PRR-based mechanisms further emphasizes the importance of lectins in plant defense.

In light of the above, it is important to recognize new lectins expressed under different stress conditions and from different sources of plants. In chapter 2, we reported the purification of a gJRL, MLGL, from mulberry (*Morus indica*) latex, which showed very high specificity

towards the derivatives of α -D-galactose. Latex is the whitish sap material of many plants exuded after mechanical damage. It contains a variety of chemicals and proteins and is believed to function primarily against insect herbivores. MLGL was found to exhibit cytotoxicity towards both normal and malignant types of mammalian cell lines *in vitro*, suggesting that it might have a role to play in the defense of the plant. Therefore, in order to correlate its function with its structure, this study was carried out to determine the complete primary as well as glycan structures of MLGL by nanoESI-Q-ToF mass spectrometry. Additionally, a dot-blot assay was performed using a xylose-specific antibody in order to confirm the presence of β (1,2) linked xylose in the glycan structure of the protein.

3.3 Materials and Methods

3.3.1 Preparation of samples for subunit mass determination

Forty μ L of a MLGL solution (\sim 500 pmol) was reduced to a volume of 10 μ L by use of a centrifugal evaporator. Samples were heated at 95°C for 5 min and 10% TFA was added to adjust the final concentration of TFA to 0.1%. Subsequently, the lectin preparation was desalted using C4 solid-phase extraction pipette tips according to the manufacturer's instructions.

3.3.2 In-solution digestion

In-solution digestions were performed as described previously [Pohlentz et al., 2016]. About 200 pmol of MLGL was rebuffered in 25 mM ammonium bicarbonate using Bio-gel P-6 spin columns. Samples were reduced by 10 mM dithiothreitol (DTT) for 1 h at 56°C. Subsequently, samples were alkylated under exclusion of light in the presence of 55 mM iodoacetamide (IAA) at ambient temperature. The reduced and alkylated samples were desalted and reconstituted in

25 mM ammonium bicarbonate buffer by use of Bio-gel P-6 spin columns. The samples were then evaporated in a centrifugal evaporator and reconstituted in 20 μ L water. This step was repeated twice. Prior to proteolysis, samples were reconstituted in 20 μ L of 25 mM ammonium bicarbonate and thermally denatured for 10 min at 95°C. Protease digestion was performed by incubation with trypsin and chymotrypsin overnight at 37°C with a substrate-to-enzyme ratio of 25:1, with thermolysin overnight at 65°C with a substrate-to-enzyme ratio of 12:1 and with endoproteinase GluC overnight at 25°C with a substrate-to-enzyme ratio of 50:1. Subsequently, the digests were dried, redissolved twice in water and dried in vacuo.

3.3.3 In-gel digestion

SDS-PAGE of MLGL was performed under reducing condition as described earlier [Sultan et al., 2004]. The gel was stained using Coomassie brilliant blue R-250 and destaining was performed using a solution containing 40% methanol and 10% acetic acid. Destained gels were thoroughly washed with water prior to performing in-gel digestion. The cut and excised bands containing MLGL were subjected to in-gel digestion by various proteases according to Shevchenko et al. [2006]. Briefly, pure acetonitrile (ACN) was added repeatedly to the gel pieces and left for 10 minutes until the gel pieces shrunk. After removal of ACN, 50 μ L of 10 mM DTT solution was added and incubated at 56°C for 30 min. After cooling down the tubes, the supernatant was discarded and 500 μ L of pure ACN was added followed by incubation for 10 min at room temperature. Iodoacetamide solution (50 μ L, 55 mM) was added and the mixture was incubated for 20 min at room temperature under exclusion of light. Subsequently, gel pieces were washed with pure ACN. Thirty μ L of 10 mM ammonium bicarbonate and 10 μ L of protease solution (trypsin and chymotrypsin: 0.1 μ g/ μ L in 1 mM HCl; elastase (0.2 μ g/ μ L in H₂O); thermolysin and endoproteinase Glu-C: 0.5 μ g/ μ L in H₂O) were added and gel

pieces were incubated for 30 min on an ice bath. If digestion buffer solution was absorbed entirely an appropriate amount of 10 mM ammonium bicarbonate and protease solution (3/1, v/v) was added to cover gel pieces completely and the reaction mixture was incubated for 90 min at ambient temperature. Twenty μL of 10 mM ammonium bicarbonate was added and digestion mixtures were incubated overnight at 37°C. If thermolysin was employed as protease, incubation was performed at 65°C. Proteolytic peptides were extracted by adding extraction buffer (5% formic acid (FA)/ACN 1:2 v/v) to each tube and incubating for 15 min at 37°C in a shaker. The supernatant was collected into a fresh tube and dried in vacuo. Finally, samples were desalted using C18 solid-phase extraction pipette tips according to the manufacturer's instructions.

3.3.4 Mass spectrometry

Nanoelectrospray ionization mass spectrometry (nanoESI Q-TOF MS) experiments were performed by use of a quadrupole time-of-flight (Q-TOF) mass spectrometer (Micromass, Manchester, UK) in the positive ion mode (ESI(+)) as described by Sharma et al. [2013]. Briefly, a Z-spray atmospheric pressure ionization (API) source was employed, with the source temperature set to 80°C and a desolvation gas (N_2) flow rate of 75 L/h. Homemade nanospray capillaries were used, with the capillary tip set to a potential of 1.1 kV and a cone voltage of 40 V. For low-energy collision-induced dissociation (CID) experiments, the peptide precursor ions were selected in the quadrupole analyzer and fragmented in the collision cell using a collision gas (Ar) pressure of 3.0×10^{-5} mbar and collision energies of 20-40 eV (E_{lab}).

Proteolytic (glyco)peptide ions were fragmented by CID and their amino acid sequences and/or glycan structures were deduced by evaluating the fragment ion series (mainly

b- and y-type ions for amide bond cleavages and B- and Y-type ions arising from glycosidic bond cleavages) observed in the corresponding CID spectra manually. Low energy CID experiments do not allow for the discrimination of the isobaric amino acids isoleucine (I) and leucine (L). Thus, they were - whenever possible - assigned as I or L by data base homology comparison. Similarly, the masses of glutamine (Q) and lysine (K) differ only by 36 mDa and are not unambiguously distinguishable with the Q-TOF instrument used. Therefore, in general, amino acids exhibiting an increment mass of 128 Da are assigned to Q by homology comparison or K, unless proven by tryptic cleavage.

High resolution mass spectra were acquired by use of a SYNAPT G2-S mass spectrometer (Waters, Manchester, UK) equipped with a Z-spray source in the positive ion sensitivity mode. Typical source parameters were: source temperature: 80°C, capillary voltage: 0.8 kV, sampling cone voltage: 20 V, and source offset voltage: 50 V.

The protein sequence data reported in this chapter will appear in the UniProt Knowledgebase under the accession number C0HK14.

3.3.5 Dot Blot assay

100 µL of MLGL of different concentrations (80 µg, 20 µg, 10 µg, 5 µg and 1 µg) were directly spotted onto a nitrocellulose membrane along with a positive control (Horseradish peroxidase) and a negative control (fetuin). The nitrocellulose membrane was incubated with blocking solution, BSA/TBS-T (3% bovine serum albumin/ 10 mM Tris, 150 mM NaCl, pH 7.5, 0.05% Tween 20) at 4°C overnight. Then it was incubated with primary antibody (the antibody was prepared as 2 µg/10 mL in BSA/TBS-T containing 0.1% BSA instead of 3%) at room temperature for three hours. The membrane was washed thrice with TBS-T solution and

incubated with secondary antibody for three hours. After washing again thrice with TBS-T and once with glycine buffer, the membrane was incubated with 0.05% of BCIP (5-bromo-4-chloro-3-indolyl phosphate) in glycine buffer for 10-15 min until it develops some colour. Then the reaction was stopped with glycine buffer.

3.4 Results

3.4.1 Primary structure

The mass spectrum of MLGL obtained under denaturing nano electrospray ionization (nanoESI) conditions (2% formic acid and 50% acetonitrile) indicates the presence of two smaller components with molecular masses of ~ 2163 Da and 2475 Da and a larger one with ~ 16 kDa (Fig. 3.4, vide infra). This observation along with the results of de novo sequencing suggests that MLGL comprises two non-covalently linked components, *viz.* one large polypeptide chain of ~16 kDa and a smaller subunit. De novo sequencing of peptides obtained from in-solution as well as in-gel proteolysis by use of various proteolytic enzymes (trypsin, chymotrypsin, elastase, thermolysin, endoproteinase Glu-C) yielded the complete amino acid sequence of MLGL.

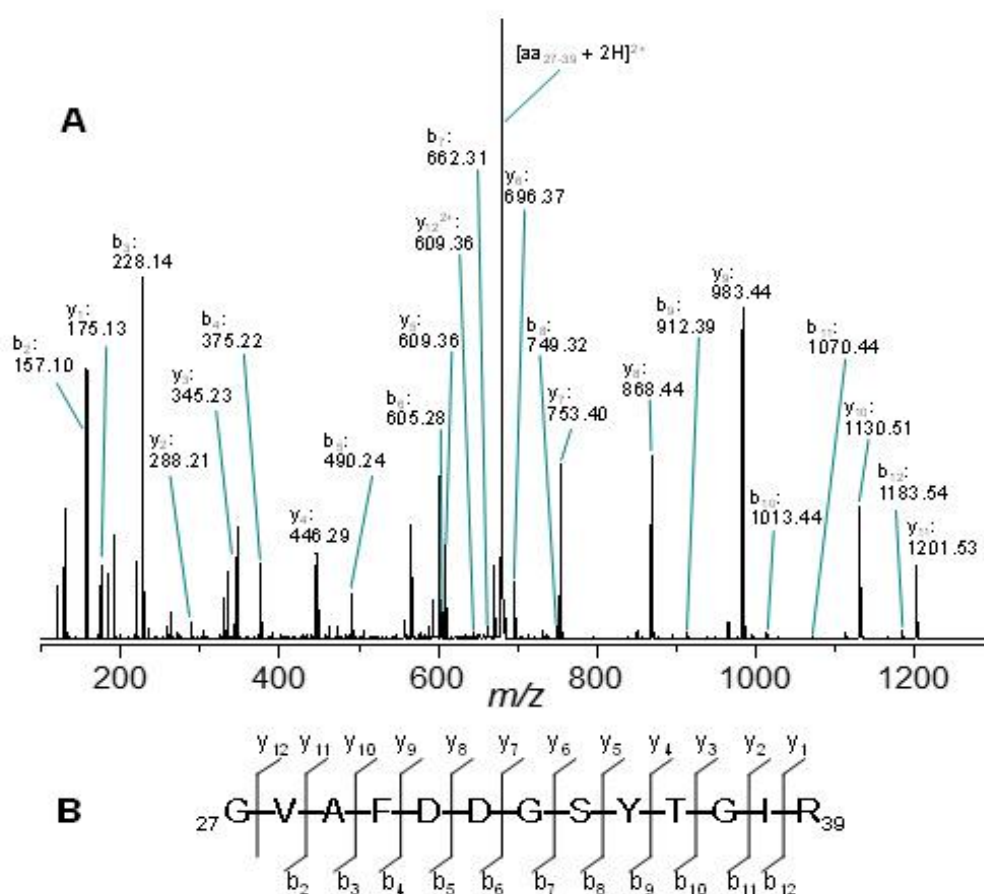


Fig. 3.1. (A) NanoESI Q-TOF fragment ion spectrum obtained from the doubly charged peptide precursor ions at m/z 679.32⁺² derived from a tryptic MLGL digest. (B) Corresponding fragmentation scheme.

Each amino acid has been determined from at least two independent proteolytic peptides. Fragmentation of the doubly charged tryptic precursor ion at m/z 679.32⁺² revealed the *N*-terminal sequence of the heavier chain of MLGL shown as a representative example in supplementary Fig. 3.1. The primary structures of the small subunit was derived by evaluation of the CID spectrum (Fig. 3.2) obtained from the doubly charged precursor ions at m/z

1082.04⁺² (calc. 1082.04⁺²) representing the most abundant ionic species (cf. Fig. 3.4, vide infra).

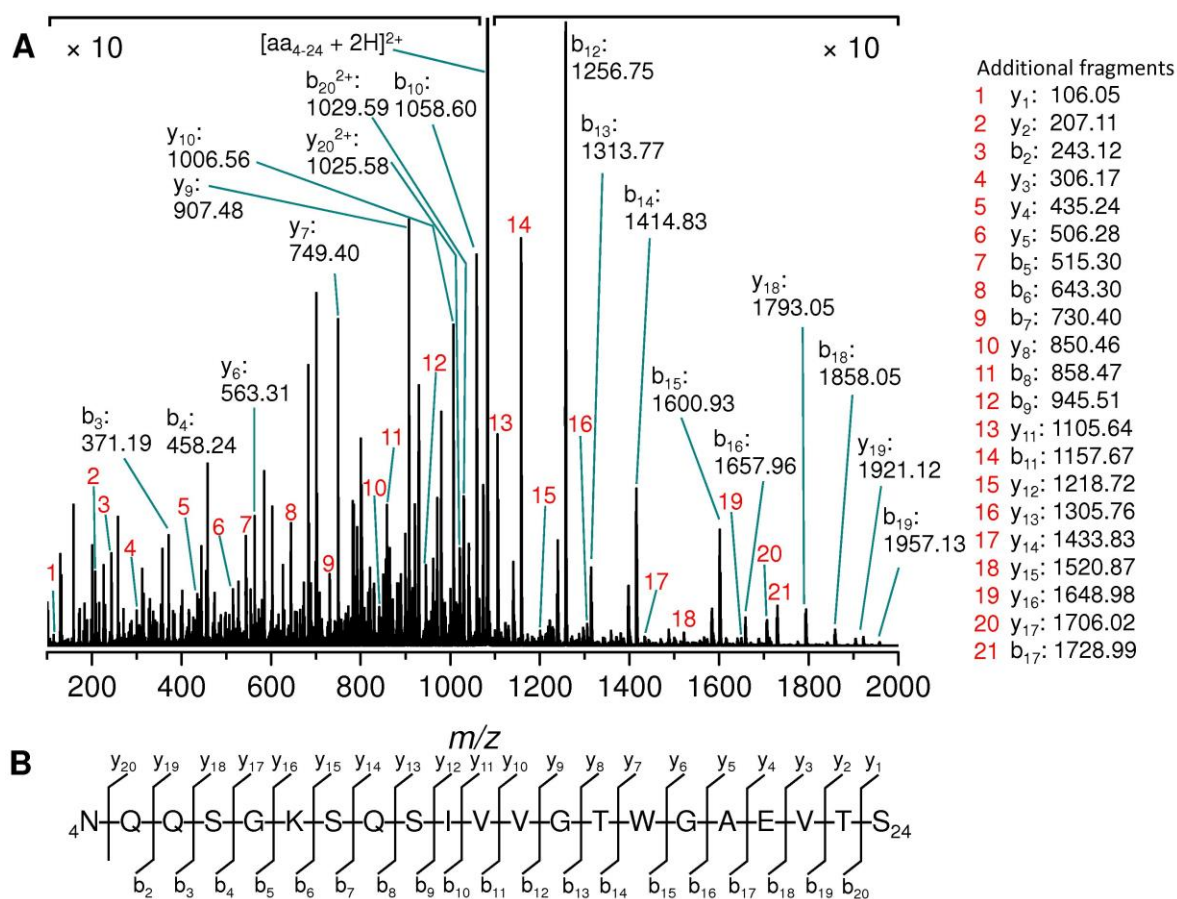


Fig. 3.2. (A) NanoESI Q-TOF fragment ion spectrum obtained from the doubly charged peptide precursor ions at m/z 1082.04⁺². The analyte comprises amino acids 4-24 of the smaller subunit of MLGL. (B) Corresponding fragmentation scheme.

Additionally, an isoform of the smaller subunit elongated by three amino acids gave rise to signals of low intensity. However, collisional activation allowed for identification of the first 3 *N*-terminal amino acids (cf. Fig. 3.3). The finally obtained sequences for both chains of MLGL and their alignment with two other JRLs *viz.* MornigaG and jacalin is shown in Fig.

Besides information on the lighter chain of MLGL, molecular weight (M_r) of the intact larger polypeptide glycoforms was obtained from the high resolution mass spectrum of purified MLGL ionized under denaturing nanoESI conditions (Fig. 3.4). M_r was determined by mass deconvolution of the centroid distribution in a nearly Gaussian shaped envelope of isotopically resolved multiprotonated species (cf. Fig. 3.4 insets B and C). The resulting values are in excellent agreement with the theoretical values calculated from the amino acid sequence displayed in Fig. 3.3 assuming post-translational modification by either a pauci-mannose ($\text{GlcNAc}_2(\text{Fuc})\text{Man}_3(\text{Xyl})$) or a complex-type ($\text{GlcNAc}_2(\text{Fuc})\text{Man}_3(\text{Xyl})\text{GlcNAc}(\text{Fuc})\text{Gal}$) glycan.

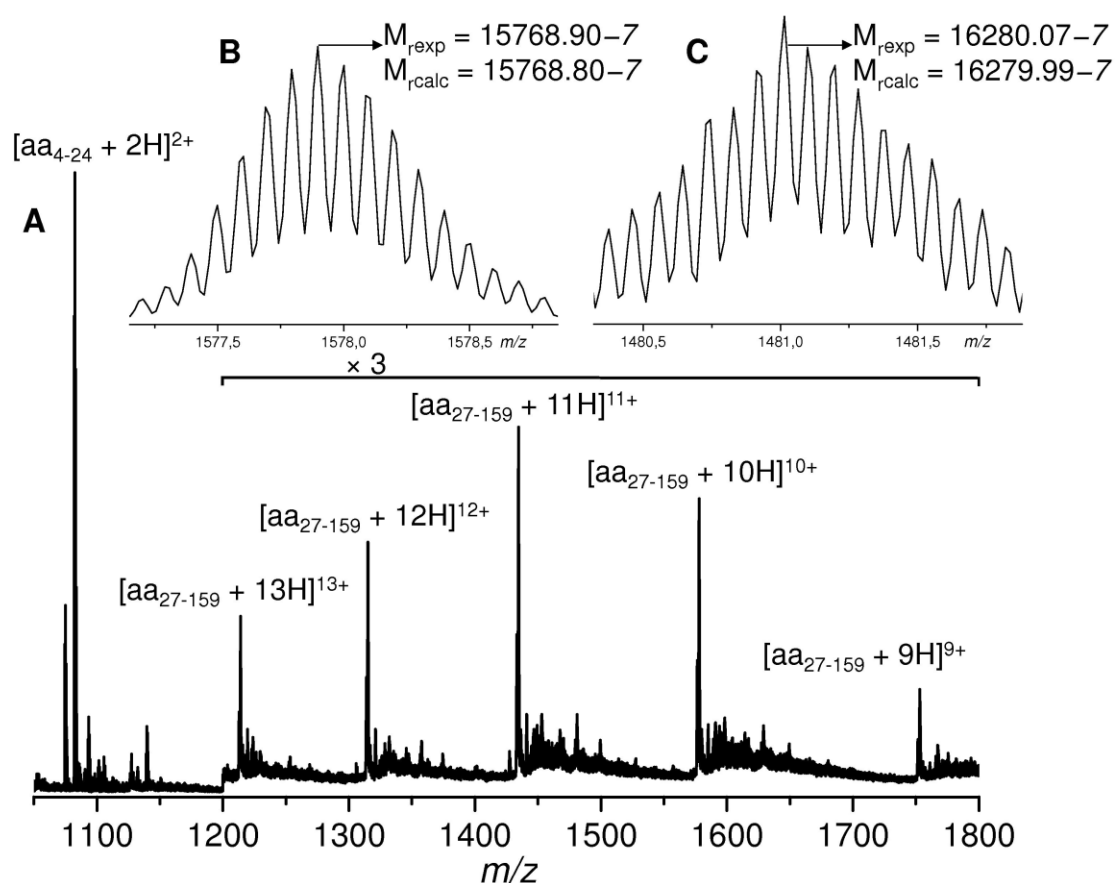


Fig. 3.4. (A) High resolution mass spectrum of purified intact MLGL dissolved in 50% acetonitrile/2% formic acid indicating the presence of different isoforms of both subunits. (B

and **C**). The insets show the expanded spectral regions for two isotopically resolved multiply charged gaseous polypeptide species harboring either a pauci-mannose type *N*-glycan (**B**, $[M + 10H]^{10+}$) or an *N*-glycan of the complex-type (**C**, $[M + 11H]^{11+}$). The exact masses were determined by comparing the experimentally obtained isotopic envelopes (ion charge states $10H^+$ and $11H^+$, respectively; the most abundant isotopologues were considered) to the theoretical isotopic distributions. The masses of the two MLGL glycoforms were determined to be $15768.90 - 7 \pm 0.03$ Da and $16280.07 - 7 \pm 0.04$ Da [theoretical values $15768.80 - 7$ Da and $16279.99 - 7$ Da calculated for the amino acid sequence depicted in Figure 4 and a modification by either $\text{GlcNAc}_2(\text{Fuc})\text{Man}_3(\text{Xyl})\text{GlcNAc}(\text{Fuc})\text{Gal-}$ or a $\text{GlcNAc}_2(\text{Fuc})\text{Man}_3(\text{Xyl})$ - glycan; the mass value is followed by the value of the mass difference (in units of 1.0034 Da) between the most abundant isotopic peak and the monoisotopic peak; thus “ $- 7$ ” indicates that the most abundant isotopologue contains 7 ^{13}C atoms].

3.4.2 Elucidation of N-glycosylation

To determine the structure of the *N*-glycan, *N*-glycopeptide ions derived from in-solution proteolytic digests were subjected to CID analysis and their glycan structure as well as their respective glycosylation site were deduced from the observed fragmentation pattern. As a representative example, Fig. 3.5, A depicts the fragment ion spectrum of CID experiments on triply charged *N*-glycopeptide precursor ions with m/z 1268.25⁺³ obtained from an in-solution chymotryptic digest. The spectrum reveals the typical fragmentation pattern of a pauci-mannose type *N*-glycan [Kumar et al., 2014]. Neutral loss of either a pentose (xylose) and/or a deoxyhexose (fucose) from the intact precursor ions as well as from fragment ions arising from subsequent cleavage of two hexoses (mannoses) were observed. Elimination of a third hexose

from the $Y_{3\alpha}/Y_{3\beta}$ fragment ions is not observed; in fact, this fragmentation process is preceded by a loss of a pentose. This observation clearly pinpoints the presence of a xylose linked to the innermost mannose. Likewise, cleavage of the proximal *N*-acetylhexosamine (*N*-acetylglucosamine, GlcNAc) giving rise to Y_0 fragment ions appears solely following the elimination of the deoxyhexose, thus substantiating the expected core fucosylation. Besides characterization of the glycan structure the detection of an almost complete b-type fragment ion series along with several y-type ions derived from the peptide backbone allowed deduction of the amino acid sequence $_{36}\text{TGIREINFEYNN}_{47}\text{ETAIGSIQVTY}_{58}$ proving Asn₄₇ to be glycosylated (Fig. 3.5, B). An additional glycopeptide species giving rise to triply charged ions at m/z 1438.61⁺³ was shown to harbor the corresponding complex-type congener attached to the same peptide (cf. Fig. 3.6). In addition to proven glycosylation at Asn₄₇, less abundant peptides harboring another potential *N*-glycosylation site at aa-position 61 have been identified. However, no evidence for modification by glycan attachment has been found. Furthermore, sequencing of most peptide ions comprising the respective peptide stretch revealed a deamidation of Asn₆₁ due to the vicinity to Gly.

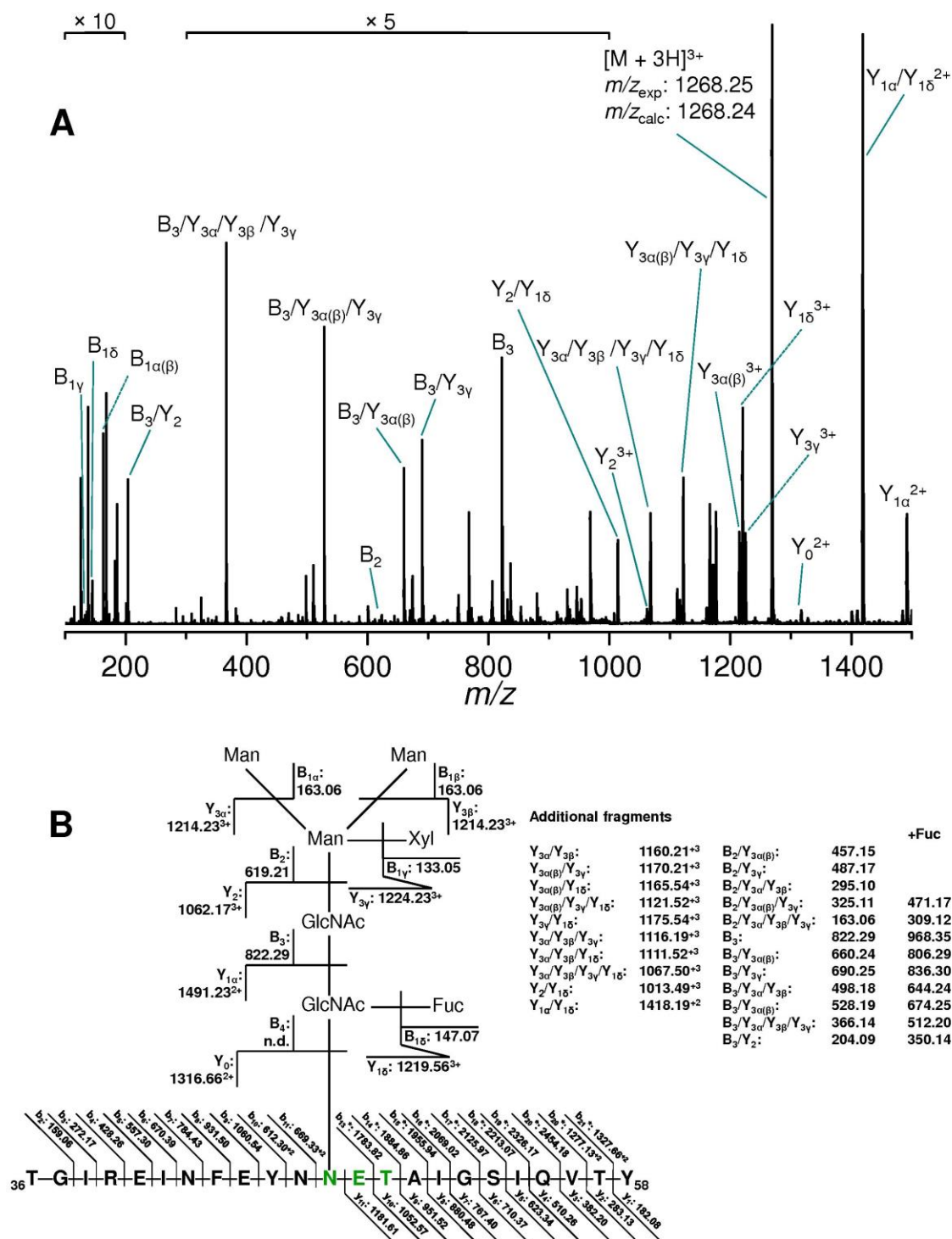


Fig. 3.5. (A) NanoESI Q-TOF fragment ion spectrum obtained from a CID experiment on the triply charged precursor glycopeptide ions with m/z 1268.25⁺³ derived from a chymotryptic in-

solution digest. **(B)** Corresponding fragmentation scheme showing pauci-mannose type N-glycan and the peptide backbone.

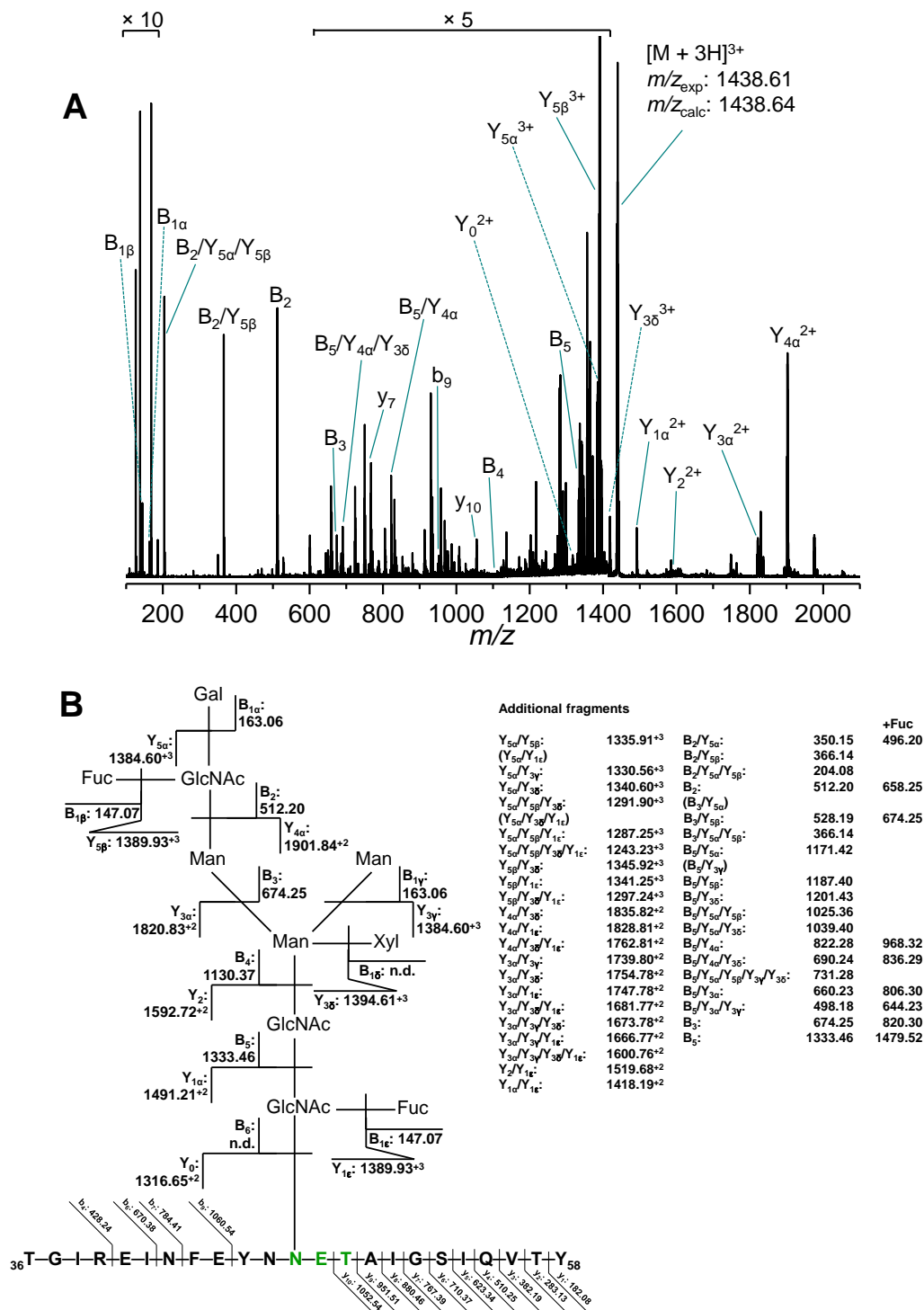


Fig. 3.6. (A) NanoESI Q-TOF fragment ion spectrum obtained from a CID experiment on the

triply charged precursor glycopeptide ions with m/z 1438.61 derived from a chymotryptic in-solution digest. (B) Corresponding fragmentation scheme showing cleavage of complex-type N-glycan and the peptide backbone.

Table 3.1: Summary of *de-novo* sequenced (glyco)peptide ions obtained from in-gel, in-solution proteolysis as well as intact species of MLGL.

(Glyco)Peptide sequence	Glycan sequence	aa	Protease	m/z_{exp}^{+z}	$m/z_{\text{calc.}}^{+z}$
GVAFDDGSYTGLR	-	1-13 b	T	679.32 ⁺²	679.32 ⁺²
HASFLTGFTPVK	-	44-55 b	T	652.83 ⁺²	652.85 ⁺²
ETYGPYGVTSPTYFK	-	91-105 b	T	835.41 ⁺²	835.39 ⁺²
GVAFDDGAYTGLR	-	1-13 b	T	671.35 ⁺²	671.32 ⁺²
AFDDGSYTGLR	-	3-13 b	T	601.29 ⁺²	601.28 ⁺²
FDDGSYTGLR	-	4-13 b	T	565.79 ⁺²	565.76 ⁺²
QHASFLTGFTPVK	-	43-55 b	T	716.90 ⁺²	716.88 ⁺²
TGFTPVK	-	49-55 b	T	749.38 ⁺¹	749.42 ⁺¹
EINFEYN NETA IG	GN ₂ (F)M ₃ (X)	14-26 b	T	1342.56 ⁺²	1342.53 ⁺²
GIREINFEYN NETA IG	GN ₂ (F)M ₃ (X)	11-26 b	T	1004.11 ⁺³	1004.09 ⁺³
GIREINFEYN NETA IG	GN ₂ (F)M ₃ (X)GN(F)G	11-26 b	T	1174.48 ⁺³	1174.49 ⁺³
VTYDVDGTPFEAKKH	-	30-44 b	T	569.63 ⁺³	569.61 ⁺³
DVDGTPFEAKKH	-	33-44 b	T	672.34 ⁺²	672.33 ⁺²
EINFEYN NETA IG	GN ₂ (F)M ₃ (X)GN(F)G	14-26 b	T	1065.78 ⁺³	1065.78 ⁺³
FEYN NETA IG	GN ₂ (F)M ₃ (X)	17-26 b	T	1164.50 ⁺²	1164.45 ⁺²
NQQSGKSQSIVVG		4-16 a	T	666.35 ⁺²	666.33 ⁺²

ETYGPYGVTSNGTHFK	-	92-105 b	T	822.34 ⁺²	822.37 ⁺²
DVDGTPF	-	33-39 b	CT	750.28 ⁺¹	750.31 ⁺¹
GVAFDDGSY	-	1-9 b	CT	930.45 ⁺¹	930.36 ⁺¹
LVEVSGY	-	65-71 b	CT	766.46 ⁺¹	766.38 ⁺¹
GVTSGETYF	-	97-104 b	CT	831.34 ⁺¹	831.37 ⁺¹
HLSL	-	130-133 b	CT	469.31 ⁺¹	469.25 ⁺¹
WLDYLGF	-	123-129 b	CT	913.42 ⁺¹	913.42 ⁺¹
VRSLTF	-	81-86 b	CT	722.42 ⁺¹	722.40 ⁺¹
GVAFDDGSYTGIREINFEYN NETA IGSIQVTY	GN ₂ (F)M ₃ (X)	1-32 b	CT	1179.29 ⁺⁴	1179.26 ⁺⁴
TGIREINFEYN NETA IGSIQVTY	GN ₂ (F)M ₃ (X)	10-32 b	CT	1268.25 ⁺³	1268.22 ⁺³
TGIREINFEYN NETA IGSIQVTY	GN ₂ (F)M ₃ (X)GN(F)G	10-32 b	CT	1438.62 ⁺³	1438.63 ⁺³
YDVNGTPFEAKKH	-	32-44 b	CT	503.02 ⁺³	502.57 ⁺³
GVAFD	-	1-5 b	CT	508.35 ⁺¹	508.22 ⁺¹
VTYDVNGTPFEAKKH	-	30-44 b	CT	569.74 ⁺³	569.27 ⁺³
SIVVGT	-	12-17 a	CT	575.46 ⁺¹	575.32 ⁺¹
AFDDGSYTGIR	-	3-13 b	CT	601.40 ⁺²	601.26 ⁺²
QSIVVG	-	11-16 a	CT	602.47 ⁺¹	602.33 ⁺¹
NQQSGKSQTIVV	-	4-15 a	CT	644.97 ⁺²	644.83 ⁺²
EINFEYN NETA IG	GN ₂ (F)M ₃ (X)	14-26 b	CT	895.34 ⁺³	895.36 ⁺³
DVDGTPFEAK	-	33-42 b	T + CT	539.81 ⁺²	539.74 ⁺²
LTGFTPVK	-	48-55 b	T + CT	431 ⁺²	431.74 ⁺²

LSLDFPSEYLVESGY	-	56-71 b	T + CT	909.52 ⁺²	909.43 ⁺²
LDFPSEY	-	58-64 b	TL	870 ⁺¹	870.37 ⁺¹
LVEVSGYTGK	-	65-74 b	TL	526.82 ⁺²	526.77 ⁺²
GVAFDDGSYTG	-	1-11 b	TL	1088 ⁺¹	1088.43 ⁺¹
LVEVSGYTGKVSQY	-	65-78 b	TL	729.90 ⁺²	729.86 ⁺²
PFEAKKH	-	38-44 b	TL	428.75 ⁺²	428.72 ⁺²
LQVTYD	-	28-33 b	TL	738 ⁺¹	738.34 ⁺¹
LSLDFPSEY	-	56-64 b	TL	1070.54 ⁺¹	1070.48 ⁺¹
LGSLQVTYD	-	25-33 b	TL	995.60 ⁺¹	995.48 ⁺¹
LDYLG	-	124-128 b	TL	580.34 ⁺¹	580.27 ⁺¹
FKLPLQDG	-	104-111 b	TL	917.56 ⁺¹	917.49 ⁺¹
FKGSVGYW	-	116-123 b	TL	943.54 ⁺¹	943.44 ⁺¹
INFEYN NETA IGSIQ	GN ₂ (F)M ₃ (X)	15-29 b	TL	1442.13 ⁺²	1442.10 ⁺²
INFEYN NETA IGS	GN ₂ (F)M ₃ (X)	15-27 b	TL	1321.55 ⁺²	1321.53 ⁺²
INFEYN NETA IGSIQVTVNGTPFEA KKH	GN ₂ (F)M ₃ (X)	15-44 b	TL	1143.51 ⁺⁴	1143.26 ⁺⁴
INFEYN NETA	GN ₂ (F)M ₃ (X)	15-24 b	TL	1192.98 ⁺²	1192.96 ⁺²
LTFKTNKETYPYG	-	84-97 b	TL	809.92 ⁺²	809.89 ⁺²
VAFDDGSYTG	-	2-11 b	TL	1031.65 ⁺¹	1031.41 ⁺¹
FEYN NETA	GN ₂ (F)M ₃ (X)	17-24 b	TL	1079.37 ⁺²	1079.40 ⁺²
GVAFDDGSYTG	-	1-11 b	TL	1088.67 ⁺¹	1088.435 ⁺¹
FDDGSYTG	-	4-11 b	TL	861.29 ⁺¹	861.30 ⁺¹

AKKHASFLT	-	41-49 ^b	GC	501.82 ⁺²	501.77 ⁺²
SIVVGTWGAE	-	12-21 ^a	GC	509.80 ⁺²	509.75 ⁺²
TAIGSIQVTYD	-	23-33 ^b	GC	584.34 ⁺²	584.28 ⁺²
SIVVGTWGAEVT	-	12-23 ^a	GC	609.88 ⁺²	609.80 ⁺²
VTYDVNGTPFE	-	30-40 ^b	GC	621.34 ⁺²	621.27 ⁺²
NQQSGKSQTIVVG	-	4-16 ^a	GC	673.41 ⁺²	673.34 ⁺²
GVAFDDGSYTGIR	-	1-13 ^b	GC	679.37 ⁺²	679.32 ⁺²
GVAFDDGSYTGIRE	-	1-15 ^b	GC	743.91 ⁺²	743.83 ⁺²
KLSLDFPSEYLVE	-	55-67 ^b	GC	770.40 ⁺²	770.38 ⁺²
TAIGSIQVTYDVNGTPFE	-	23-40 ^b	GC	956.54 ⁺²	956.45 ⁺²
NQQSGKSQSIVVGTWGAEVTS	-	4-24 ^a	Heat	1082.11 ⁺²	1082.03 ⁺²
NQQSGKSQSIVVGTWGAEVT	-	4-23 ^a	Heat	1038.60 ⁺²	1038.52 ⁺²
NQQSGKSQTIVVGTWGAQATS	-	4-24 ^a	Heat	1074.62 ⁺²	1074.54 ⁺²
GVAFDDGSYTGIREINFEYNNETAIG	GN ₂ (F)M ₃ (X)	1-26 ^b	Heat	1341.59 ⁺³	1341.56 ⁺³
GVAFDDGSYTGIREINFEYNNETAIGS IQ	GN ₂ (F)M ₃ (X)	1-29 ^b	Heat	1450.97 ⁺³	1450.95 ⁺³
SIVVGTWGAEVTS	-	12-24 ^a	Heat	653.39 ⁺²	653.32 ⁺²
GVAFDDGSYTGIREI	-	1-15 ^b	Heat	800.47 ⁺²	800.37 ⁺²
KNNQQSGKSQSIVVGTWGAEVTS	-	2-24 ^a	Heat	802.48 ⁺³	802.39 ⁺³
AKNNQQSGKSQSIVVGTWGAEVTS	-	1-24 ^a	Heat	826.15 ⁺³	826.07 ⁺³
NQQSGKSQTIVVGTWGA	-	4-20 ^a	Heat	881.02 ⁺²	880.93 ⁺²
GVAFDDGSYTGIREINFEYNNETAIGS IQ	GN ₂ (F)M ₃ (X)GN ₂ (F)G	1-29 ^b	Heat	1267.05 ⁺⁴	1267.03 ⁺⁴

DGSYTGIREINFEYNN ETA IGSIQ	GN₂(F)M₃(X)	6-29 b	Heat	1287.90 ⁺³	1287.88 ⁺³
NQQSGKSQSIVVGTWGAE –H ₂ O	-	4-21 a	Heat	929.54 ⁺²	929.45 ⁺²
NQQSGKSQSIVVGTWGAEV –H ₂ O	-	4-22 a	Heat	979.09 ⁺²	978.985 ⁺²
VTYDVNGTPFEAKKHASFLT	-	30-49 b	Heat	557.05 ⁺⁴	557.02 ⁺⁴
YDVNGTPFEAKKHASFLT	-	32-49 b	Heat	675.70 ⁺³	675.66 ⁺³
NQQSGKSQSIVVGTWGAK VT	-	4-23 a	Heat	692.40 ⁺³	692.35 ⁺³
NNQQSGKSQSIVVGTWGAEVTS	-	3-24 a	Heat	759.74 ⁺³	759.69 ⁺³

T – trypsin, CT – chymotrypsin, TL – thermolysin, GC – endoproteinase GluC, heat – heat treatment of intact MLGL, a – lighter chain, b – heavier chain.

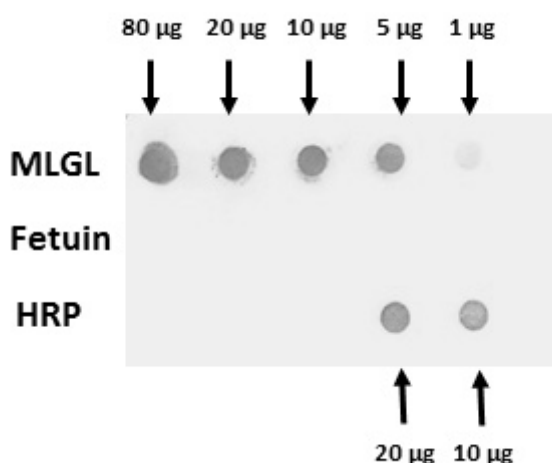


Fig. 3.7. Dot-blot assay performed with a xylose-specific antibody. Fetuin has been used as the negative control and HRP (Horse-radish peroxidase) as the positive control.

Further, the presence of β (1,2) xylose in the glycan structure of MLGL was confirmed by dot-blot assay using a xylose-specific antibody which would cross-react specifically with xylose that is attached to the core N-glycan via β (1,2) linkage. Fetuin is a glycoprotein containing fucose linked as α (1,6) without any xylose residue in its core glycan, whereas horse-

radish peroxidase (HRP) contains β (1,2)-linked xylose. Therefore, fetuin has been used as a negative control and HRP as a positive control in this assay. Different concentrations (1, 5, 10, 20 and 80 μ g) of MLGL was spotted on to a nitrocellulose membrane along with fetuin and HRP (10 and 20 μ g) and blots were developed after subsequent incubation with xylose-specific antibody as shown in Fig. 3.7. This experiment re-confirms the presence of pauci-mannose type backbone in the protein.

3.5 Discussion

Mulberry plants are widely cultivated in India to produce silk from the cocoons of *Bombyx mori* larvae. The larvae of *Bombyx mori* feed on mulberry leaves and its caterpillar is called silkworm. The silkworm makes a protective cocoon around itself from which silk filaments are obtained. While the silkworm develops a healthy relationship with mulberry leaves, the latex of mulberry accumulates a large amount of defense related proteins against herbivorous insects [Kitajima et al., 2010]. In our previous study (Chapter 2) we reported the purification and biophysical characterization of a cytotoxic lectin (MLGL) from the latex of white mulberry. In this chapter we present the complete primary structure of MLGL determined by nanoESI-Q-ToF mass spectrometry. Since MLGL is a glycoprotein its N-glycan structure and site of glycosylation were also determined by mass spectrometry.

De novo sequencing of the lectin yielded a sequence which exhibits high homology to MornigaG, a well-characterized lectin, isolated from black mulberry (*Morus nigra*) bark. In mulberry bark, the lectin (MornigaG) remains as a vacuolar storage protein which helps in the growth and development of the plant. Homology search on NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) has shown that the leaves of white mulberry (*Morus rotundiloba*) also expresses a similar type of lectin (accession number - AEE92792). MLGL

contains two polypeptide chains, the larger one consisting of 133 amino acids and the smaller one having 21 to 24 amino acids. Crystal structure analysis of two gJRLs, jacalin and *Maclura pomifera* lectin (MPA) has shown that the smaller chain of a gJRL does not contribute to the ligand binding [Sankarnarayanan et al., 1996, Lee et al., 1998], but the *N*-terminal glycine of the larger chain is important for it. Here it is pertinent to mention that initially we thought the lectin purified by us represents the *Morus indica* lectin (MIL) purified previously from *Morus indica* latex using a combination of ammonium sulphate precipitation and conventional chromatographic methods [Patel et al., 2011]. However, comparison of the partial sequence information of MIL reported by Patel et al. revealed significant differences with that of MLGL determined in this study. It was observed that in the two peptides of MIL that were sequenced covering 29 amino acid residues have 9 differences with respect to MLGL (Fig. 3.8), suggesting that MLGL is distinctly different from MIL. Further, the sequence comparison shown in Fig. 3.3 clearly indicates that MLGL resembles the bark lectin, MornigaG, in its amino acid sequence although it differs slightly from MornigaG in terms of specific carbohydrate recognition. MornigaG was shown to bind galactose and Me α Gal with similar affinity as deduced from hapten inhibition assays [Rouge et al., 2003]. Thus, MLGL shows exclusive preference for the α -anomer of galactose and hence it can be used for the selective detection of this sugar.

MLGL	KGSVGYWLDYLG FHL SL	- COOH
MornigaG	KGSVGYWLDYLG FHL SL	- COOH
MIL	LGGVGVFLENLG FHL AL	- COOH
MLGL	KETYPYGVTS GT	
MornigaG	KETYPYGVTS GT	
MIL	K-TYPYGH TS GR	

Fig. 3.8. Sequence comparison of MLGL and mornigaG (accession number - Q8LGR4) with the amino acid stretches of MIL obtained from tryptic digests [Patel et al., 2011].

MLGL is a glycoprotein which contains two types of glycan structures; pauci-mannose type and complex type. The site of glycosylation for these two types of glycans is the same residue (Asn₄₇) which indicates considerable heterogeneity in the protein in terms of its carbohydrate structures. The intact mass of the longer polypeptide chain of MLGL along with the pauci-mannose type glycan was found to be 15761.827 Da whereas along with the complex type of glycan it was 16273.045 Da. The intact mass of the most abundant species (NQQSGKSQSIVVGTWGAEVTS) of the shorter polypeptide chain was found to be 2162.08 Da. Therefore, altogether each subunit of MLGL is of ~18 kDa. Glycosylation might be an indicator for the lectin's dual role, that is, either to act as a lectin or as a receptor for terminal galactose- or mannose-binding lectins.

**Fluorescence and Circular Dichroism Studies on the Accessibility
of Tryptophan Residues and Unfolding of a Jacalin-related α -D-
Galactose-specific Lectin from Mulberry (*Morus indica*)**

Chapter 4

4.1 Summary

MLGL (Mulberry Latex Galactose-specific Lectin) is an α -D-galactose binding lectin isolated from the latex of mulberry (*Morus indica*) tree and contains two tryptophan residues in each of its subunits. The fluorescence emission maximum of native MLGL seen at 326 nm shifts to 350 nm upon incubation with 6 M guanidine thiocyanate (Gdn.SCN), suggesting that the tryptophans are located inside the hydrophobic core of the protein and become fully exposed upon denaturation. Fluorescence quenching studies revealed that the neutral acrylamide exhibits the highest quenching, with ~33% of total fluorescence in native MLGL being quenched at 0.5 M concentration, whereas iodide (~24%) and cesium (~4%) ions showed significantly lower quenching. With the denatured protein, acrylamide quenching involves both dynamic and static processes as evident from an upward curving Stern-Volmer plot. Time-resolved fluorescence studies showed two lifetime components of 3.7 ns and 1.3 ns for the native protein, while three lifetime components were observed for the denatured protein. MLGL showed high resistance to urea (up to 8 M) and guanidine hydrochloride (up to 6 M), whereas treatment with 6 M Gdn.SCN completely denatured the protein, via a broad sigmoidal transition with a transition midpoint at ~3.75 M. Circular dichroism studies and hemagglutination assays showed that the secondary and tertiary structures as well as lectin activity of MLGL were unaffected up to 70 °C. Additionally, pH dependent studies showed that the secondary structure of MLGL is unaltered in the pH range 6.2 to 8.5, but a decrease in lectin activity is observed (~50%) at pH 6.2.

4.2 Introduction

Lectins are a diverse group of proteins which can recognize specific carbohydrate structures in a non-catalytic and reversible manner [Lis and Sharon, 1973]. Their binding to specific carbohydrate molecules is mediated by various non-covalent interactions, e.g., H-bonding, hydrophobic interactions, van der Waals' forces, metal coordination bonds as well as by water molecules [Elgavish and Shaanan, 1997]. Lectin-carbohydrate interactions are highly specific but the strength of these interactions is relatively weak with a binding constant in the 10^3 - 10^4 M⁻¹ range for monosaccharides. Nevertheless, these proteins show high specificity and affinity for oligosaccharide structures of cell-surface glycoconjugates thereby agglutinating various types of cells [Ambrosi et al., 2005; Sharon and Lis, 2004]. Besides their binding to carbohydrates on cell surfaces, their interaction with different non-carbohydrate, predominantly hydrophobic molecules such as phytohormones (e.g. abscisic acid, gibberellic acid and indole acetic acid) and porphyrins may also be relevant from a physiological standpoint [Komath et al., 2006]. Lectins are ubiquitously present in all kinds of organisms and have various applications in biomedical and biological research owing to their ability to recognize specific sugars. These are, for instance, in targeting tumor markers, in affinity chromatography for the purification of glycoconjugates or oligosaccharides, in new techniques such as lectin blotting and microarray technology as well as in developing lectin-based biosensors [Dan et al., 2015].

In recent chapters 2 and 3, we reported the purification of an α -D-galactose specific lectin (MLGL) from the latex of mulberry (*Morus indica*) and characterized it in terms of its amino acid sequence, secondary structure and thermal stability [Datta et al., 2016]. MLGL belongs to the family of jacalin-related lectins (JRLs), a unique family of plant lectins which

shows high affinity towards the tumor-associated T-antigen, Gal β 1-3GalNAc [Sankarnarayanan et al., 1996; Raval et al., 2004]. MLGL is a glycoprotein with a subunit mass of 18 kDa and each of its four subunits has two polypeptide chains, a longer chain with 133 amino acids and a shorter chain with 20-24 amino acids which are non-covalently associated with each other. The longer chain contains a pauci-mannose [GlcNAc₂(Fuc)Man₃(Xyl)] or complex-type [GlcNAc₂(Fuc)Man₃(Xyl)GlcNAc(Fuc)Gal] glycan covalently attached to Asn₂₁. A 3-dimensional model of MLGL was built using the homology-modeling server MODELLER as shown in Fig. 4.1. This primarily β -sheet-rich lectin is thermally quite stable and shows a denaturation temperature of 77.6 °C. Moreover, the high cytotoxicity exhibited by MLGL towards MDCK and MCF-7 cell lines led us to speculate that this lectin might have a role to play in the protection of the plant against phytophagous insects or higher animals [Datta et al., 2016].

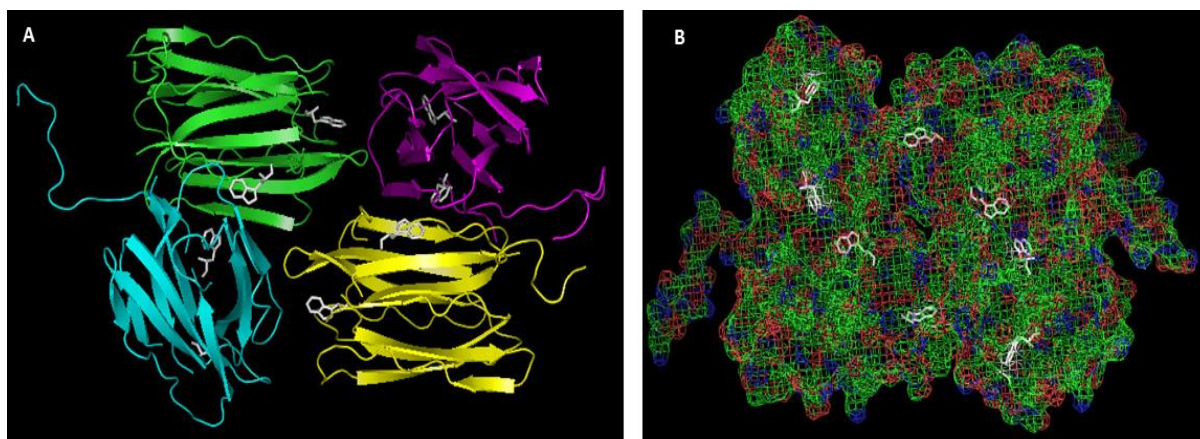


Fig. 4.1. A 3-dimensional model of MLGL (Uniprot accession number – C0HK14) generated using homology-modeling server MODELLER [Webb and Sali, 2014]. Jacalin (PDB code - 1JAC) has been used as the template. (A) A cartoon representation and (B) a mesh representation of four subunits of MLGL. Tryptophans are shown as white sticks.

MLGL has two tryptophan residues in each of its subunits, Trp₁₂₃ in the heavier chain and Trp₁₈ in the lighter chain. To investigate the exposure, accessibility and microenvironment of these tryptophans and their possible involvement in ligand binding, we have performed steady-state fluorescence quenching studies employing neutral, anionic and cationic quenchers as well as fluorescence life time measurements. Tryptophans located on the protein surface and those buried inside the hydrophobic core are not accessible to the quenchers to the same extent. Furthermore, a neutral quencher can penetrate much better into the structure of a protein than an ionic quencher due to the hydrophobic effect. Therefore, quenching of tryptophan fluorescence by neutral and ionic quenchers can yield valuable information on the structure of the protein [Lakowicz, 1999; Eftink and Ghiron, 1981]. Additionally, the unfolding behavior of MLGL was investigated by observing changes in the tryptophan emission characteristics in the presence of guanidine hydrochloride (Gdn.HCl), guanidine thiocyanate (Gdn.SCN) and urea in order to assess resistance of the protein to these chaotropic agents.

In previous work, the thermal stability of MLGL was investigated by differential scanning calorimetry [Datta et al., 2016]. In this study, circular dichroism (CD) spectroscopy was employed to observe temperature-induced changes in its secondary and tertiary structures as well as the effect of pH on the stability of MLGL. Furthermore, hemagglutinations assays were performed at different temperatures and pH to assess its lectin activity under various conditions.

4.3 Materials and Methods

4.3.1 Materials

Fresh latex was obtained from the mulberry plantation located in the campus of University of Hyderabad. Guanidine hydrochloride, urea, guanidine thiocyanate, acrylamide and potassium

iodide were obtained from Sigma (St. Louis, MO, USA). Cesium chloride was obtained from Cisco Research Laboratories (Mumbai, India). MLGL was purified using cross-linked guar-gum as described in chapter 2.

4.3.2 Steady-state fluorescence spectroscopy

All fluorescence spectra were recorded using a Fluoromax-4 fluorescence spectrometer from Horiba Jobin-Yvon. Slit widths of 3 and 5 nm were used on the excitation and emission monochromators, respectively. Absorption spectra were recorded on a Shimadzu UV3101PC double beam spectrophotometer. MLGL samples ($OD_{280} \leq 0.1$) were irradiated with 295 nm light in order to selectively excite tryptophan residues of the protein. Emission spectra were recorded above 305 nm. Quenching titrations were performed as described earlier [Padma et al., 1998]. Briefly, small aliquots were added from a 5 M stock solution of the quencher to the protein sample and fluorescence spectra were recorded after each addition. Fluorescence intensities were corrected for volume changes before further analysis of the quenching data. Fluorescence intensity values were corrected for inner filter effect by measuring the absorbance of the sample at excitation and emission wavelength. Corrections were then made using the following equation [Lakowicz, 1999]:

$$F_{\text{corr}} = F_{\text{obs}} \times \text{antilog} [(OD_{\text{ex}} + OD_{\text{em}})/2]$$

Here, F_{corr} and F_{obs} are the corrected and observed fluorescence intensities, respectively, and OD_{ex} and OD_{em} are the absorbances measured at excitation and emission wavelengths, respectively. For chemical denaturation studies, MLGL samples were incubated overnight with various concentrations of denaturants (urea, Gdn.HCl and Gdn.SCN) and the fluorescence spectra were recorded ($\lambda_{\text{ex}} = 295$ nm) thereafter.

4.3.3 Fluorescence lifetime measurements

Time-resolved fluorescence measurements were performed using a time-correlated single-photon counting (TCSPC) spectrometer from Horiba Jobin Yvon IBH (Glasgow, UK). A Delta diode laser source ($\lambda_{\text{exc}}=285$ nm) was used for excitation and a Hamamatsu photomultiplier (R3809U-50) was used to detect the fluorescence. The instrument response function, which was limited by the fwhm (full width at half maximum) of the excitation pulse, was 890 ps. Samples were excited at 285 nm and emission intensities were recorded at 325 nm (for native protein) or 350 nm (for denatured protein). A dilute solution of Ludox in water was used in place of the sample to record the lamp profile.

4.3.4 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded on an Aviv 420 spectropolarimeter (Lakewood, NJ, USA) equipped with a thermo cube and a chiller. Protein concentrations of 0.1 OD and 0.8-1.0 OD have been used for measurements in the far- and near-UV regions, respectively. A rectangular quartz cuvette of 2 mm path length was used for all the measurements. MLGL samples were dialyzed against the buffer of desired pH before spectral recording. The following buffers were used to obtain different pH conditions: 20 mM KCl-HCl (pH 2.0), 20 mM glycine-HCl (pH 3.0), 20 mM sodium acetate (pH 4.0 to 5.0), 20 mM sodium phosphate (pH 6.0 – 7.0), 20 mM Tris-HCl (pH 8.5). All buffers contained 150 mM NaCl.

4.3.5 Hemagglutination assays

Hemagglutination assays of MLGL at different pH and temperatures were carried out as described earlier [Kavitha and Swamy, 2009]. In brief, lectin samples were dialyzed against the buffer of desired pH and erythrocyte samples (human B+ cells) were also prepared in the same buffer. For thermal inactivation studies, MLGL samples were incubated at the desired

temperature for 30 min, cooled to room temperature, and any aggregates formed were separated by centrifugation. Hemagglutination assays were performed using the clear supernatants.

4.4 Results

4.4.1 Quenching of tryptophan fluorescence

When the tryptophan residues of MLGL were selectively excited by irradiating at 295 nm, the emission spectrum showed a maximum at 326 nm, indicating that the tryptophan residues are buried inside the hydrophobic core of the protein. Fluorescence spectra recorded in the presence of acrylamide, iodide and cesium ions for the native protein, and in the presence of acrylamide for the denatured protein are shown in Fig. 4.2. It can be seen from these spectra that the neutral acrylamide can penetrate better into the hydrophobic core of the protein and hence shows the highest percentage of quenching (~33%, see Table 4.1) among the three quenchers. On the other hand, iodide ion can quench about 22% of MLGL fluorescence whereas cesium ion could quench only about 4%. Binding of MLGL with its carbohydrate ligand, Me α Gal, did not significantly alter the tryptophan environment as evidenced by a lack of change in the emission maximum or intensity, indicating that the tryptophan residues are probably not directly involved in ligand binding. Denaturation of the protein with 6 M Gdn.SCN resulted in a red shift of the emission maximum to 350 nm, indicating complete exposure of the tryptophans. The percent quenching values observed with all three quenchers under different conditions are listed in Table 4.1. Denaturation of MLGL significantly enhanced the accessibility of tryptophans for acrylamide (to 84%), but only marginally for iodide ion (to 28%) and cesium ion (to 22%).

The quenching data were analyzed according to Stern-Volmer Eq. (1) as well as modified Stern-Volmer Eq. (2) [Lehrer, 1971].

$$F_o/F = 1 + K_{sv} [Q] \quad (1)$$

$$F_o/\Delta F = f_a^{-1} + (K_a f_a)^{-1} [Q]^{-1} \quad (2)$$

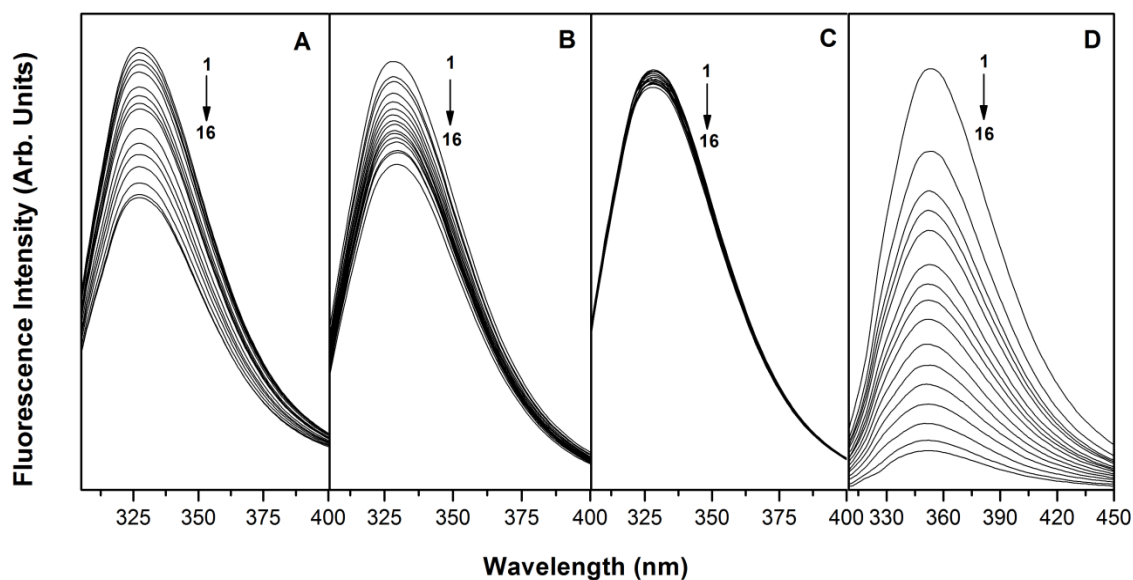


Fig. 4.2. Fluorescence emission spectra of MLGL in the presence of different quenchers. The quenchers/conditions used were: (A) acrylamide, (B) iodide ion and (C) cesium ion with native MLGL; (D) acrylamide with denatured MLGL. In each panel spectrum 1 corresponds to MLGL alone and spectra 2-16 correspond to those recorded in the presence of increasing concentration of the respective quencher.

Table 4.1. Degree of fluorescence quenching of MLGL by different quenchers under various conditions. The final concentration of quencher used was 0.5 M in each case.

Quencher	Percent Quenching (%)		
	Native	With 20 mM Me α Gal	Denatured by 6 M Gdn.SCN
Acrylamide	33.2	33.2	84.4
Iodide ion	23.8	21.4	27.9
Cesium ion	3.8	7.9	22.6

where F_0 and F are the relative fluorescence intensities in the absence and presence of quencher, respectively, $[Q]$ is the quencher concentration, K_{sv} is Stern-Volmer quenching constant of the lectin for a given quencher, $\Delta F (= F_0 - F)$ is the change in fluorescence intensity at any point in the titration, f_a is the fraction of total fluorescence intensity that is accessible to the quencher and K_a is the corresponding Stern-Volmer quenching constant for the accessible fraction of the fluorophores.

The Stern-Volmer plot for acrylamide is linear, but interestingly a biphasic pattern is seen for iodide ion (Fig. 4.3 A, B). Stern-Volmer plot for cesium ion could not be analyzed due to very low extent of quenching. However, the Stern-Volmer plots in presence of Me α Gal for both acrylamide and iodide ion are biphasic and suggest that ligand binding induces heterogeneity in the tryptophan environment. The Stern-Volmer quenching constants obtained from the linear fits shown in Fig. 4.3 A, B are listed in Table 4.2.

The Stern-Volmer plot for acrylamide quenching of denatured MLGL shows an upward curvature, indicating the involvement of dynamic as well as static quenching processes. Therefore, in this case the Stern-Volmer plot was analyzed by Eq. (3), which allows resolution of the static and dynamic components [Lakowicz, 1999]:

$$F_0/F = (1 + K_{sv}[Q])(1 + K_s[Q]) \quad (3)$$

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

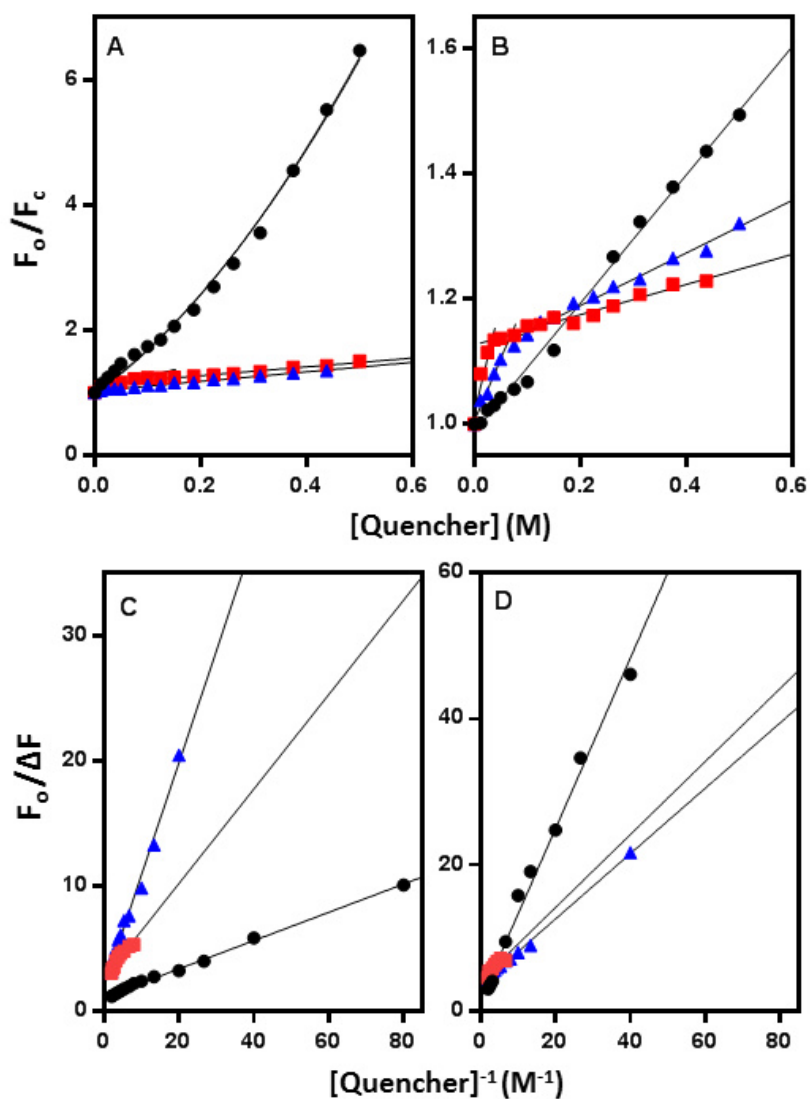


Fig. 4.3. Stern-Volmer (A, B) and modified Stern-Volmer (C, D) plots for fluorescence quenching of MLGL with acrylamide (A, C) and iodide ion (B, D). Blue triangles, native MLGL; black circles, upon denaturation with 6 M Gdn.SCN; red squares, in presence of 20 mM Me α Gal.

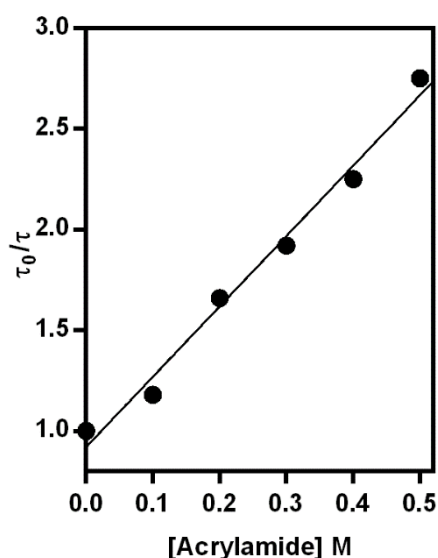


Fig. 4.4. Resolution of static and dynamic components of acrylamide quenching under denatured condition of MLGL. See text (sections 4.4.1 and 4.4.2) for details.

The Stern-Volmer constant for the dynamic component (K_{sv}) was obtained from the lifetime measurements (described in Section 4.4.2) using the following equation [Lakowicz and Weber, 1973]:

$$\tau_0/\tau = 1 + K_{sv}[Q] \quad (4)$$

where τ_0 and τ are the average lifetimes in the absence and presence of quencher at a concentration of $[Q]$. The average lifetimes were determined from the analysis of time-resolved fluorescence measurements according to equation (5) as described in section 4.4.2. A plot of τ_0/τ vs acrylamide concentration ($[Q]$) yielded a straight line with a slope of 3.5 as shown in Fig. 4.4. The value of the slope corresponding to dynamic quenching constant (K_{sv}) was then incorporated in Eq. (3) to obtain the static quenching constant (K_s) as 2.7 M^{-1} . Eq. (3) was then used to fit the Stern-Volmer plot for denatured MLGL shown as solid curved line in Fig. 4.3, A. As can be seen from the figure, the fit matches quite well with the quenching data.

Table 4.2. Different parameters of MLGL fluorescence obtained from the analysis of Stern-Volmer and modified Stern-Volmer equations.

Quencher	$K_{sv1} (M^{-1})$	$K_{sv2} (M^{-1})$	$K_s (M^{-1})$	$f_a (%)$	$K_a (M^{-1})$
Acrylamide					
1) native	0.76	-	-	53	2.11
2) Me α Gal bound	2.17	0.71	-	38	6.91
3) denatured	3.5	-	2.7	91	9.8
Iodide ion					
1) native	2.00	0.42	-	27	8.1
2) Me α Gal bound	3.48	0.24	-	24	8.3
3) denatured	1.03	-	-	63	1.34

The fraction of total fluorescence intensity (f_a) accessible to the quencher was obtained from modified Stern-Volmer plots for iodide ion and acrylamide quenching and are listed in Table 4.2. While 53% and 38% of total fluorescence is accessible to acrylamide in the native protein and in presence of ligand, respectively, denaturation increases the accessible fraction to 91%. In case of iodide ion, 27% for the native protein, 24% in the presence of ligand and 63% in the denatured state are accessible to the quencher. The corresponding Stern-Volmer constants (K_a) obtained from Eq. (2) are also listed in Table 4.2.

Table 4.3. Results of time-resolved fluorescence measurements on MLGL under different conditions.

Sample	τ_1 (ns)	α_1	τ_2 (ns)	α_2	τ_3 (ns)	α_3	τ (ns)	$\langle \tau \rangle$ (ns)	χ^2
Native MLGL	1.3	0.25	3.7	0.75	-	-	3.1	3.31	1.04
MLGL + 20 mM Me α Gal	1.98	0.29	3.6	0.71	-	-	3.13	3.3	1.03
MLGL in 6 M Gdn.SCN	1.6	0.46	2.5	0.44	0.36	0.10	1.87	2.1	1.05
MLGL + 0.5 M acrylamide	0.9	0.12	3.3	0.88	-	-	3.01	3.21	1.12
MLGL + 0.5 M KI	1.18	0.26	2.8	0.74	-	-	2.38	2.59	1.13
MLGL + 0.5 M CsCl	1.8	0.40	4.04	0.60	-	-	3.14	3.53	1.02
MLGL in 6 M Gdn.SCN + 0.5 M acrylamide	0.04	0.32	0.5	0.50	2.3	0.18	0.68	1.64	1.01
MLGL + 6 M Gdn.SCN + 0.4 M acrylamide	0.04	0.23	0.6	0.56	2.35	0.21	0.84	1.62	1.04
MLGL + 6M Gdn.SCN + 0.3 M acrylamide	0.08	0.17	0.7	0.60	2.37	0.23	0.98	1.62	1.07
MLGL + 6M Gdn.SCN + 0.2 M acrylamide	0.11	0.12	0.87	0.61	2.3	0.25	1.12	1.59	1.12
MLGL + 6M Gdn.SCN + 0.1 M acrylamide	0.2	0.03	1.4	0.87	3.6	0.09	1.58	1.82	1.09

4.4.2 Lifetimes of tryptophan fluorescence

Fluorescence lifetimes of the tryptophan residues in MLGL were determined under various conditions using an excitation laser source at 285 nm. The decay profiles of tryptophan fluorescence in the native protein, in presence of 20 mM Me α Gal and in the denatured state are shown in Fig. 4.5. The decay profiles for the native protein could be fit to a biexponential decay function ($\chi^2 \approx 1.0$) using a multi-exponential iterative program provided by the manufacturer. However, fitting the decay profiles obtained in the denatured condition of MLGL to a biexponential function gave rise to significant errors ($\chi^2 \geq 2.0$) while fitting to a triexponential function yielded better fits ($\chi^2 \approx 1.0$). The lifetimes of tryptophans in the native protein were obtained as 3.7 ns and 1.3 ns where the longer lifetime component contributes 75% to total fluorescence intensity while the shorter lifetime component corresponds to only 25%.

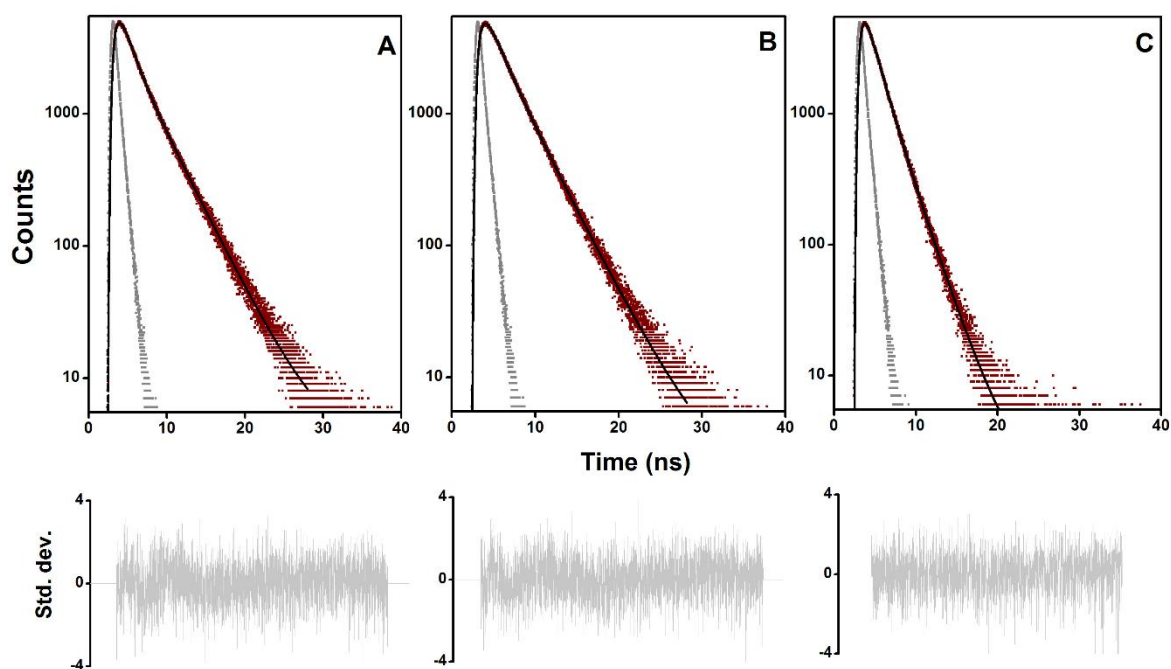


Fig. 4.5. Fluorescence decay profiles of MLGL. A) Native protein, B) in the presence of 20 mM Me α Gal and C) in the presence of 6 M Gdn.SCN.

In the ligand bound condition, the lifetimes were obtained as 3.6 ns and 1.98 ns with contributions of 71% and 29%, respectively. Table 4.3 summarizes the fluorescence lifetimes of MLGL obtained under different conditions along with the average lifetimes.

The average lifetimes of tryptophan fluorescence were calculated according to the following equations [Grinvald and Steinberg, 1974; Inokuti and Hirayama, 1965]:

$$\tau = \sum_i \alpha_i \tau_i / \sum_i \alpha_i \quad (5)$$

$$\langle \tau \rangle = \sum_i \alpha_i \tau_i^2 / \sum_i \alpha_i \tau_i \quad (6)$$

where $i = 1, 2, 3$ and α_i is the pre-exponential weighting factor of τ_i , and τ and $\langle \tau \rangle$ are amplitude and intensity average lifetimes, respectively.

It can be seen from Table 4.3 that in its native and ligand bound states MLGL shows almost similar average lifetimes (τ) of 3.1 ns and 3.13 ns, respectively.

When denatured by 6 M Gdn.SCN, MLGL shows three lifetime components as 2.5 ns, 1.6 ns and 0.36 ns. The shortest lifetime (0.36 ns) contributes only 10% of the total intensity compared to the longer lifetimes which contribute a total of 90% (2.5 ns, 44%; 1.6 ns, 46%). The average lifetime (τ) is reduced to 1.87 ns in the denatured state as compared to 3.1 ns in the native protein. The average lifetime is decreased even further to a magnitude of 0.68 ns in presence of 0.5 M acrylamide in the denatured state of MLGL. It can also be seen from Table 4.3 that the average lifetime of native MLGL in presence of 0.5 M acrylamide is reduced to 3.01 ns whereas in presence of 0.5 M iodide ion the decrease is even more, to a magnitude of 2.38 ns. However, in presence of 0.5 M cesium ion τ remained unchanged. The average lifetimes obtained for the denatured protein in presence of different concentrations of acrylamide, listed in Table 4.3, were used in Fig. 4.4.

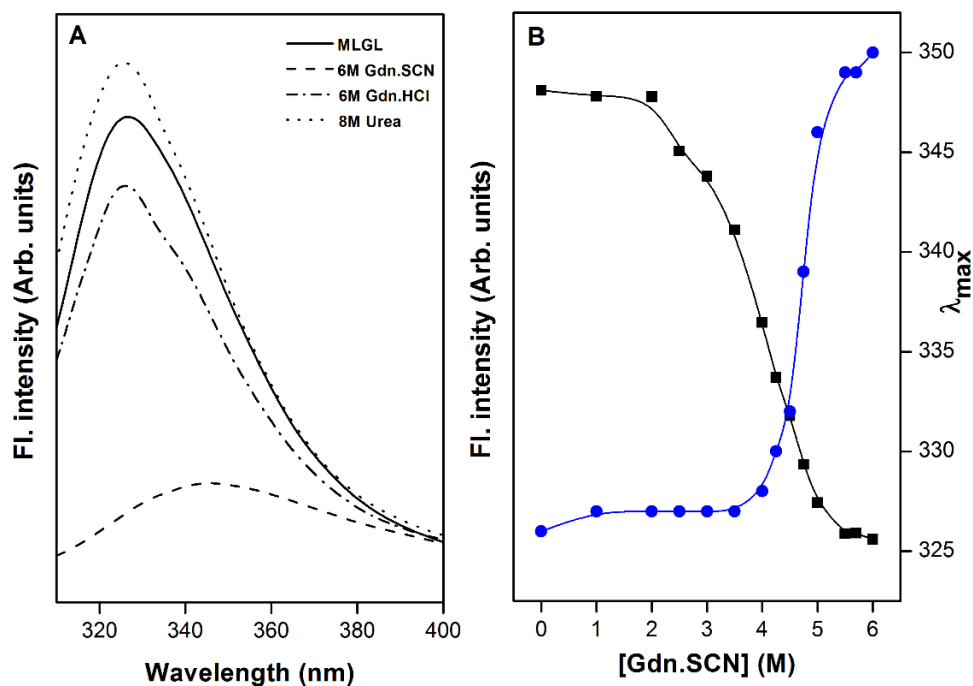


Fig. 4.6. (A) Intrinsic fluorescence spectra of MLGL in presence of different denaturants. (B) Unfolding curve of MLGL for denaturation by Gdn.SCN. Plots of emission maximum (blue circles)/emission intensity (black squares) vs concentration of Gdn.SCN are shown.

4.4.3 Chaotrope-induced unfolding of MLGL

Chemical unfolding of MLGL was investigated by monitoring changes in the protein intrinsic fluorescence in the presence of 6 M Gdn.HCl, 6 M Gdn.SCN and 8 M urea as shown in Fig. 4.6 A. While no change has been observed in the emission maximum of MLGL in the presence of 8 M urea or 6 M Gdn.HCl, a 20-nm red shift accompanied by strong quenching was observed in the presence of 6 M Gdn.SCN. Fig. 4.6 B shows the unfolding curve of MLGL in the presence of increasing concentrations of Gdn.SCN. The emission λ_{\max} of MLGL centered at 326 nm remains almost unchanged up to a concentration of 3.5 M Gdn.SCN and starts to red shift gradually above 4 M concentration of the denaturant. At 6 M Gdn.SCN, the emission maximum shifts to 350 nm, suggesting that MLGL is completely unfolded. The unfolding

curve shown in Fig. 4.6 B shows a midpoint at ~ 4.75 M and suggests that unfolding of MLGL is a two-state transition without the involvement of any intermediates. Previous DSC studies have shown that thermal denaturation of MLGL is also a two-state process [Datta et al., 2016]. Denaturation of MLGL by Gdn.SCN is a cooperative process as evidenced by the sigmoidal nature of the unfolding curve.

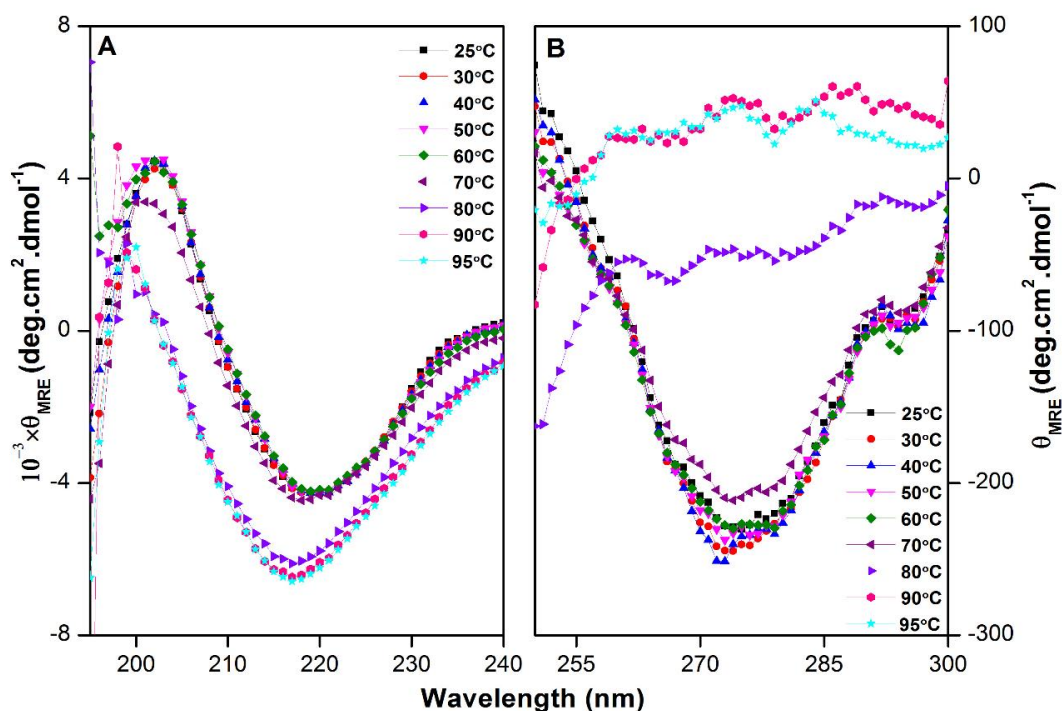


Fig. 4.7. CD spectra of MLGL recorded as a function of temperature in the far UV (A) and near-UV (B) regions. Symbols corresponding to various temperatures are indicated in the figure.

4.4.4 Effect of temperature and pH on MLGL activity and stability

Thermal stability of MLGL was investigated earlier using differential scanning calorimetry which yielded an unfolding temperature of 77.6 °C for the native protein and suggested that unfolding is a two-state process [Datta et al., 2016]. In this study, we have investigated temperature-induced changes in the structural features of MLGL by CD spectroscopy. The

spectra shown in Fig. 4.7 A, B clearly indicate that the secondary and tertiary structures of this protein are essentially unaltered up to 70 °C, whereas significant changes can be seen at temperatures ≥ 80 °C. Interestingly, in this case the (negative) signal intensity in the far-UV region increases with the minimum shifting to a lower wavelength. This might be due to stacking of β -sheets at high temperatures where the protein tends to aggregate. However, the near-UV CD spectra (Fig. 4.7 B) clearly indicate that MLGL retains its tertiary structure up to 70 °C, which is completely lost at 80 °C. These results are consistent with the previous DSC studies which showed that MLGL undergoes complete unfolding at 77.6 °C.

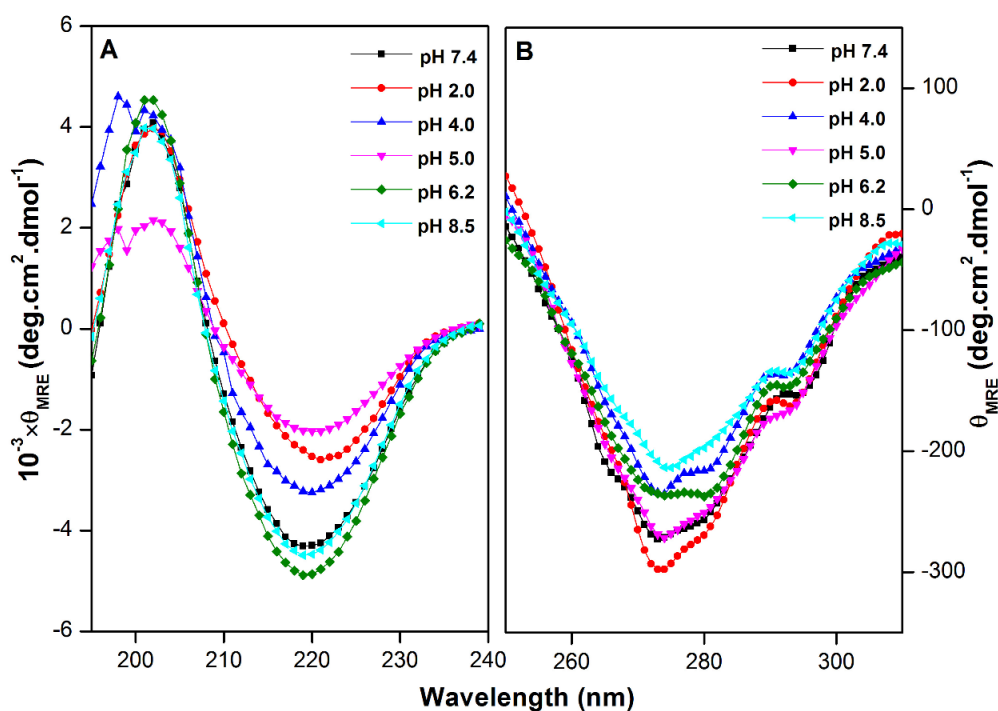


Fig. 4.8. CD spectra of MLGL recorded as a function of pH in the far-UV (A) and near-UV (B) regions. Symbols corresponding to different pHs are mentioned in the figure.

Hemagglutination assays performed with samples incubated at different temperatures showed that MLGL is fully active (100% activity) even after incubation at 70 °C for 30 min.

Incubation at 80 °C led to a near-complete loss of the activity with only about 12.5% activity being seen, whereas complete inactivation was observed upon incubation at 90 °C.

Effect of changing the pH on the structure of MLGL was also investigated by CD spectroscopy (Fig. 4.8). The far UV-CD spectra corresponding to samples of pH 6.2-8.5 show a strong negative band centered around 220 nm and a positive band at ~200 nm and remain nearly constant in intensity. However, the spectra recorded at pH 2.0, 4.0 and 5.0 show considerable loss in ellipticity without significantly altering the character of the spectra. On the other hand, the near UV CD spectra of MLGL at different pH showed similar pattern, characterized by a minimum at ~275 nm and a shoulder at ~294 nm, with only moderate differences in the intensities.

The hemagglutinating activity of MLGL was maximum (100%) and unchanged in the pH range 7.0-8.5, whereas ~50% loss in activity was observed in the pH range 5.0-6.2. Hemagglutination assays could not be performed at lower pH due to lysis of erythrocytes.

4.5 Discussion

Fluorescence quenching studies are often employed to investigate the microenvironments of tryptophans, which can give qualitative insights into the structure of a protein of interest. MLGL, a homotetrameric glycoprotein with a subunit molecular weight of 18 kDa. It contains two tryptophan residues in each of its subunits: Trp₁₂₃ in the heavier chain and Trp₁₈ in the lighter chain. A 3-D model of MLGL (Fig. 1) suggests that these tryptophans are located in loops and are buried inside the protein structure. The emission maximum of the native protein, seen at 326 nm, indicates that the Trp residues are in a hydrophobic environment, which is fully consistent with the above interpretation [Lakowicz, 1999]. Binding of the specific ligand, Me α Gal, did not bring about any change in the emission λ_{max} and upon denaturation by 6 M

Gdn.SCN the λ_{max} was shifted to 350 nm, indicating complete exposure of the tryptophans towards polar solvent molecules (in this case water) as the protein unfolds from its compact structure. Similar observations were made previously with several other lectins, e.g., snake gourd seed lectin [Komath and Swamy, 1999], *Trichosanthes dioica* seed lectin [Sultan and Swamy, 2005], pumpkin phloem lectin [Narahari and Swamy, 2005] and *Trichosanthes cucumerina* seed lectin [Kenoth and Swamy, 2003].

A linear Stern-Volmer plot (Fig. 4.3) and a decrease in fluorescence lifetimes (Table 4.3) indicate that quenching by acrylamide is essentially dynamic in nature with a Stern-Volmer quenching constant (K_{sv}) of 0.76 M^{-1} . The dynamic nature perhaps originates from the formation of a charge transfer complex at the excited state tryptophan as has been suggested for other proteins [Tallmadge et al., 1989; Eftink and Ghiron, 1976]. The Stern-Volmer plots in the presence of Me α Gal show two slopes indicating that in ligand bound state the quencher is perhaps able to differentiate the deeply buried and partially exposed Trp residues, resulting in biphasic quenching. Stern-Volmer plot for acrylamide in denatured condition shows an upward curvature indicating that quenching takes place by both dynamic and static mechanisms. In native MLGL, the tryptophans are sufficiently buried inside the protein matrix and hence are sterically shielded from acrylamide by segments of the protein [cf. Eftink and Ghiron, 1976] in order for complex formation to occur. But when MLGL is denatured, these tryptophans get exposed to the solvent and hence no such steric shielding can prevent the formation of quencher-fluorophore complex, resulting in static quenching in addition to collisional quenching. A K_s (Static quenching constant) value of 2.7 M^{-1} is indicative of the association constant for this complex formation between the tryptophan and acrylamide. This kind of upward curving profile was also observed for pumpkin phloem lectin (PPL) and

Trichosanthes cucumerina seed lectin (TCSL) [Narahari and Swamy, 2009; Kenoth and Swamy, 2003].

Since iodide ion carries a negative charge, the extent of quenching will depend not only on the exposure of the tryptophans but also on the charge of the surrounding amino acids [Möller and Denicola, 2002]. The presence of acidic amino acids near tryptophans in the protein can result in electrostatic repulsion and hence decreased quenching by iodide ion. A cartoon representation of one of the subunits of MLGL (Fig. 4.9) shows that aspartic acid and glutamic acid residues (Glu₂₂ and Asp₁₂₅ in the heavier chain, and Glu₂₁ in the lighter chain) are present in the vicinity of the tryptophans which might repel the iodide ion and prevent it from coming close to the tryptophans. The biphasic nature of Stern-Volmer plots for native and ligand bound protein in presence of iodide indicates that the extent of such charge repulsion near the two tryptophans in MLGL is somewhat different, which results in biphasic quenching profiles for the iodide ion. It may be noted that the Stern-Volmer plots obtained from iodide quenching of jacalin, which is homologous to MLGL and contains two Trp residues in each subunit [Mahanta et al., 1992], also showed a biphasic pattern in the presence as well as absence of Me α Gal [Sastry and Surolia, 1986].

Many single tryptophan containing proteins, such as adrenocorticotropin-(1-24), glucagon, human serum albumin, subtilisin etc. display two lifetime components which can be due to the presence of different conformational states of the protein [Lakowicz, 1999; Ross et al., 1981; Royer, 1993]. Native MLGL shows two lifetime components, possibly due to differences in the microenvironment of the two tryptophans; however, it is also possible that two lifetime components could simply arise from one of the tryptophan residues and the other tryptophan may not contribute to the measured lifetime at all [Lakowicz, 1999]. As the protein

denatures, a third weak component appears, this can be explained by the fact that with denaturation the conformational heterogeneity of the random polypeptide now increases even more [Lakowicz, 1999].

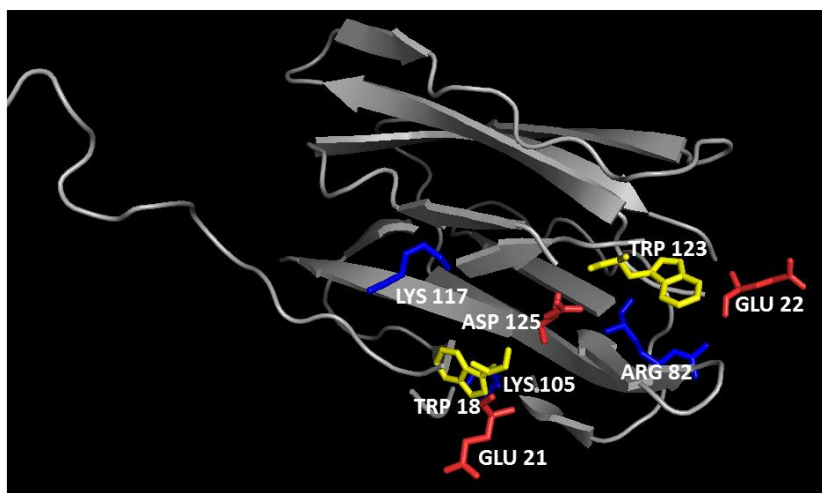


Fig. 4.9. A cartoon representation of the 3-D structure of MLGL subunit. The model was generated as described in the legend to Fig. 4.1. Tryptophans are shown as yellow sticks, positively charged amino acids are shown as blue sticks and negatively charged residues as red sticks.

The chaotropic strength of the denaturants used is in the following order: Gdn.SCN > Gdn.HCl > urea, i.e., Gdn.SCN is the strongest denaturant among the three as shown for many other globular proteins [Kumar et al., 2016; Rohamare et al., 2013]. The unfolding curve for Gdn.SCN-induced denaturation reveals that the native structure of MLGL is retained up to a concentration of 3.5 M Gdn.SCN. This observation suggests that MLGL is partially resistant even to Gdn.SCN denaturation implying high structural stability of the protein. Reports of chemical denaturation studies on other jacalin-related lectins, however, showed that jacalin, banana lectin or frutalin are easily denatured by Gdn.HCl [Khan et al., 2013; Sahasrabudhe

et al., 2004; Campana et al., 2002]. On the other hand, a serine protease isolated from *Nocardiosis* sp. NCIM 5124 (NprotI) was found to be resistant to denaturation by Gdn.HCl which was attributed to its unusual structural element, polyproline fold [Rohamare et al., 2013]. Therefore, to the best of our knowledge MLGL is the first jacalin-related lectin that is highly resistant to denaturation by Gdn.HCl and urea. This unusual stability could perhaps be attributed, at least in part, to its glycosylation as glycosylation was shown to impart stability to other lectins, viz., *Erythrina corallodendron* lectin and soybean agglutinin [Mitra et al., 2003; Sinha et al., 2005].

Although the results presented in Fig. 4.6 are consistent with a two-state transition model for the Gdn.SCN-induced unfolding of MLGL, it is also possible that any intermediate present is too metastable or not sufficiently distinct to be detected by changes in the steady-state fluorescence spectra of the protein, used in the present study. For example, chaotrope-induced unfolding of carbonic anhydrase from various sources was shown to involve a molten-globule like intermediate [Dolgikh et al., 1984; Jagannadham and Balasubramanian, 1985; Idrees et al., 2016]. A more detailed investigation can be taken up to study this aspect in the future.

The high structural stability of MLGL is also reflected in its resistance to thermal denaturation. MLGL was found to retain its secondary and tertiary structures as well as lectin activity even after incubation at 70 °C. The pH dependent CD studies showed that the tertiary structure of the lectin remains almost unchanged in all pH conditions while the secondary structural contents appear to reduce from pH 5.0. Overall, the structure of MLGL is stable in the pH range 6.2 to 8.5 although its lectin activity starts decreasing with decrease in pH from pH 6.2.

In summary, fluorescence quenching studies of a latex lectin, MLGL, showed that the tryptophan residues of this protein are relatively buried in the hydrophobic core, which get exposed upon denaturation by 6 M Gdn.SCN. The tryptophans in native MLGL show two fluorescence lifetimes of 1.3 ns and 3.7 ns whereas in the denatured protein three lifetimes of 1.6 ns, 2.5 ns and 0.36 ns are observed. The lectin is highly resistant to denaturation by Gdn.HCl or urea and the unfolding by Gdn.SCN follows a two-state, cooperative process. Furthermore, MLGL is a thermostable protein retaining its structure as well as lectin activity up to 70 °C while a reduction in secondary structure and activity is observed at lower pH values.

**Biophysical Characterization and Partial Amino Acid Sequence
of a Kunitz-type Protease Inhibitor from Okra (*Abelmoschus
esculentus*) Seeds**

Chapter 5

5.1 Summary

A Kunitz-type protease inhibitor has been purified from okra (*Abelmoschus esculentus*) seeds by a combination of ammonium sulphate precipitation, anion-exchange chromatography and reversed-phase high performance liquid chromatography (RP-HPLC). The protein shows an apparent mass of 21 kDa on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing condition. Analysis of far UV circular dichroism spectrum showed that the protein contains ~39% β -sheets but only 5% α -helices. De novo sequencing of the protein by nanoESI-Q-ToF mass spectrometry revealed that it shares 43% sequence identity with a putative 21 kDa trypsin inhibitor from *Theobroma bicolor*, considering about 83% sequence coverage has been achieved for the protein. An intramolecular disulphide linkage between Cys₁₄₈ and Cys₁₅₅ was also detected. The protein showed ~24% and ~25% sequence identity with α -amylase/subtilisin inhibitor from barley and soybean (Kunitz) trypsin inhibitor, respectively.

5.2 Introduction

Every living organism expresses a particular class of proteins which function to regulate or prevent proteolysis. These proteins are called protease inhibitors (PIs) [Laskowski and Kato, 1980]. In plants, these proteins are induced in response to insect herbivory and wounding. Plant PIs inactivate the serine proteases produced by many insects, resulting in amino acid deficiency and subsequent retardation in the growth of the insect [Zhu-Salzman and Zeng, 2015; Kessler and Baldwin, 2002]. There are five classes of proteases classified based on their active site amino acid residues. These are, serine, threonine, cysteine, aspartic acid and metallo- classes of proteases. In each case, the corresponding amino acid residue (Ser, Thr, Cys) acts as the nucleophile towards the substrate peptide bond except for aspartic acid and metallo class of proteases where an activated water molecule acts as the nucleophile [Barett et al., 1998]. Every PI has a reactive peptide bond (P1-P'1) on its surface which interacts with the active site of a protease it is specific for, which leads to the formation of a very stable enzyme/inhibitor complex [Garcia-Olmedo et al., 1987; Laskowski and Kato, 1980]. The association constant for this interaction is very high, generally being in the range of 10^7 - 10^{13} M⁻¹ [Laskowski and Kato, 1980]. Broadly, PIs can be classified into five classes according to the nature of the protease such as serine, threonine, cysteine, aspartic acid and metallo protease inhibitors. However, in plants, about eighteen nonhomologous families of inhibitors have been identified [Laskowski and Qasim, 2000]. Soybean trypsin inhibitor (Kunitz) family is one of them. The Kunitz type protease inhibitors have been studied extensively due to their high abundance in plants. These are generally a single chain or double chain polypeptide proteins with a Mr of ~ 20 kDa containing two disulphide bridges and a single reactive scissile bond (P1-P'1) [Moslov and Valueva, 2005; Oliva et al., 2010]. This particular class of protease inhibitors inhibit not

only the activity of trypsin but other enzymes, such as, chymotrypsin, elastase, subtilisin or α -amylase as well. The reactive site residue (P1) determines the specificity towards the cognate enzyme, it is an Arg or Lys residue in trypsin inhibitors, an Ala or Ser residue in elastase inhibitors whereas in chymotrypsin inhibitors P1 could be any of the following; Tyr, Phe, Trp, Leu and Met [Laskowski and Kato, 1980; Oliva et al., 2010].

Okra (*Abelmoschus esculentus*) is a widely cultivated vegetable crop in tropical and warm temperate regions around the world. It belongs to the mallow family (Malvaceae) and has a number of medicinal properties. Different parts of okra such as the fruits, seeds, pods and leaves have shown to possess antidiabetic, antihyperlipidemic and antioxidant properties and can also be used for the treatment of dysentery and diarrhoea [Roy et al., 2014]. Previously, purification and characterization of four trypsin inhibitors with molecular weight of ~ 20 kDa from okra seeds has been reported by Ogata et al. [1986], but no sequence information was available for them. Since trypsin inhibitors are products of multigene families there could be multiple isoinhibitors for one type of inhibitor. Therefore, without sequence information these protease inhibitors cannot be assigned to a particular family.

In the present study, a Kunitz-type protease inhibitor from the seeds of okra was purified by a combination of ammonium sulphate precipitation, anion-exchange chromatography and reversed-phase HPLC. The secondary structure of the protein was determined by circular dichroism (CD) spectroscopy and its inhibitory activity was checked against trypsin using N α -benzoyl-L-arginine ethyl ester (BAEE) as the substrate. Furthermore, over 80% of the primary structure was obtained by de-novo sequencing of the protein by nanoESI-Q-ToF mass spectrometry.

5.3 Materials and methods

5.3.1 Materials

Okra seeds were purchased from local vendors in Hyderabad, India. DEAE-Sephadex and PD10 columns were obtained from GE-healthcare (Little Chalfont, UK) and C18 column was obtained from Spinco Biotech Pvt Ltd (Chennai, India). Trypsin, BAEE, chymotrypsin and BTEE were obtained from Sigma-Aldrich (St. Louis, Missouri, US).

5.3.2 Purification of okra protease inhibitor

Okra seeds were ground to a fine powder and defatted using n-hexane. About 20 gm of defatted okra seed meal were suspended in 100 mM Tris-HCl (pH- 7.6) containing 150 mM NaCl and stirred overnight. Any insoluble material was separated by centrifugation at 6000 rpm for 30 minutes and the supernatant was subjected to heat treatment in the following way. The supernatant was placed in a water bath set at 60°C for 5 minutes and immediately placed in ice-cold water and kept there for at least 15 minutes. Thereafter, the heat-treated crude sample was centrifuged at 7000 rpm for 15 minutes and the supernatant was brought to 80% ammonium sulphate saturation. The material was allowed to stand in the cold (4°C) overnight and then the precipitate was collected by centrifugation at 6000 rpm for 30 minutes, dissolved in minimum volume of water and dialyzed extensively against distilled water containing 0.02% sodium azide. A DEAE-Sephadex column was pre-equilibrated with 20 mM sodium phosphate buffer (pH 7.4) without NaCl. The clear supernatant was loaded onto the ion-exchange column after dialysis and the mixture of okra protease inhibitors were eluted with 20 mM phosphate buffer containing 0.3 M NaCl. Fractions showing high absorbance at 280 nm were pooled, concentrated and reconstituted in water using a PD10 column before further purification by HPLC. Reversed-phase HPLC (RP-HPLC) was performed on a Shimadzu UFLC system (Kyoto, Japan) equipped with a diode array detector. A semi-preparative C18 column (250×10

mm, 10 μ m particle size) was used with 0.1% trifluoroacetic acid (TFA) as Solution A and 66% acetonitrile (ACN) in 0.1% TFA as Solution B. The trypsin inhibitor mixture was separated using a 70 min programme with a constant flow rate of 5 ml/min. The peaks were collected manually and purity of the proteins was assessed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli [1970].

5.3.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded on an Aviv 420 spectropolarimeter (Lakewood, NJ, USA) equipped with a thermo cube and a chiller. Protein concentrations of 0.1 OD and 0.8-1.0 OD were used for measurements in the far and near-UV regions, respectively. A rectangular cuvette of 2 mm path length was used for all the measurements.

5.3.4 Trypsin inhibitory assays

Trypsin inhibitory activity has been assayed as described by Yanes et al. [2007]. Briefly, 1.8 μ g/mL trypsin was incubated with or without protease inhibitor at room temperature for 5 minutes in activity buffer (0.1 M Tris, 20 mM CaCl_2 , pH 8.0). Subsequently, 0.38 mM BAEE was added to the solution, mixed gently and the absorbance at 253 nm was measured over a period of 5 minutes.

5.3.5 Preparation of samples for mass spectrometry

The samples of okra protease inhibitor mixture were both reduced and alkylated or to characterize disulfide linkages solely alkylated prior to acid hydrolysis or in-solution digestion by various proteases as described previously [Pohlentz et al., 2016]. Reduction of disulfide bridges was achieved by reducing the protein in a solution containing 6 M guanidium hydrochloride, 250 mM trizma base/HCl and 65 mM dithiothreitol (DTT) adjusted to pH 8.6.

The resulting mixture was incubated at 56°C for one hour. Thereafter, reduced thiol groups were alkylated by adding 2.5 mg iodoacetamide to the final mixture and incubating it at room temperature for 45 min under exclusion of light. In a separate set of experiments, DTT was omitted and samples were alkylated by use of iodoacetamide as described above. Desalting was achieved by size-exclusion chromatography using Biogel P-6 spin columns (Bio-Rad) according to the manufacturer's instructions. Subsequent in-solution digestions were performed according to Datta et al. [2016] and acid hydrolysis was performed according to Kumar et al. [2014]. Trypsin, chymotrypsin, endoproteinase GluC and thermolysin were employed as proteases. For acid hydrolysis, 25 µL of 20% acetic acid was added to the treated or untreated samples of the trypsin inhibitor mixture reconstituted in deionized water and the resulting mixtures were incubated at 95°C for two hours. Subsequently, samples were dried in vacuo and stored at -20°C prior to mass spectrometric experiments.

5.3.6 In-gel digestion

SDS-PAGE of okra protease inhibitor mixture was carried out under reducing condition as described previously [Sultan et al., 2004]. The gel was stained using Coomassie brilliant blue R-250 and excised bands were subjected to in-gel digestion by various proteases according to Shevchenko et al. [2006] with some slight modification as described recently by Pohlentz et al. [2016]. Finally, samples were desalted using C18 solid-phase extraction pipette tips according to the manufacturer's instructions.

Mass spectrometric experiments on an nanoESI-Q-ToF MS instrument have been essentially performed as described in chapter 3.

5.4 Results

5.4.1 Purification of okra protease inhibitor

Okra protease inhibitor was purified from the defatted okra seed meal. About 600 mg of soluble protein was loaded on to the DEAE-Sephadex column and a batch separation with increasing concentrations of sodium chloride was carried out. The fraction collected with 0.3 M sodium chloride in 20 mM phosphate buffer showed two protein bands on SDS-PAGE between Mr 25 and 20 kDa as shown in Fig. 5.2 (lane 3) and can be referred to as protease inhibitor mixture. About 35 mg of protease inhibitor mixture was obtained from anion-exchange chromatography which was subsequently separated by RP-HPLC. The separation was achieved by using the following programme:

<u>Time (min)</u>	<u>Solution B (%)</u>
0	5
5	20
10	40
15	50
55	70
60	100
60.1	5
70	5
70.1	Stop

The protease inhibitor mixture gave rise to three distinct peaks at retention times (RT) of 37.2, 42.6 and 48.1 minutes on HPLC chromatogram as shown in Fig. 5.1. The highest intensity peak (peak 3, RT 48.1 min) named as okra protease inhibitor (OPI), was concentrated and characterized further. On SDS-PAGE, under reducing conditions, OPI showed a band corresponding to a molecular weight of ~ 21 kDa (Fig. 5.2, lane 1), which matches the lower band of the protease inhibitor mixture. The mobility of peak 1 and peak 2 on SDS-PAGE matched the higher molecular weight band of the trypsin inhibitor mixture (data not shown).

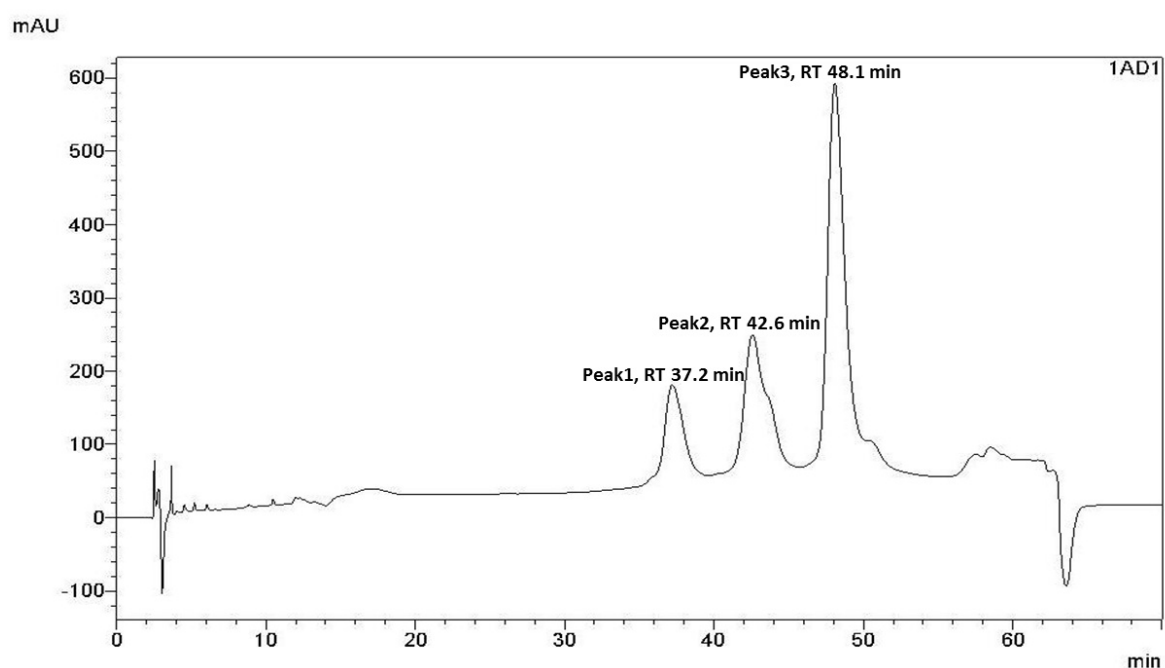


Fig. 5.1. HPLC-chromatogram for the purification of okra protease inhibitor (OPI). The column was eluted with an increasing gradient of 66% ACN in 0.1% TFA. Peak 3 was designated as okra protease inhibitor.

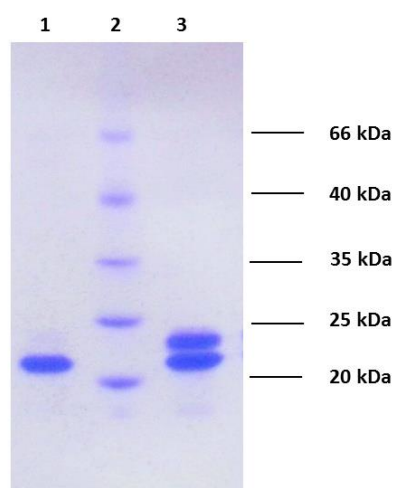


Fig. 5.2. SDS-PAGE of okra protease inhibitor mixture and purified okra protease inhibitor (OPI). Lane 1: OPI; lane 2: MW marker; lane 3: okra trypsin inhibitor mixture.

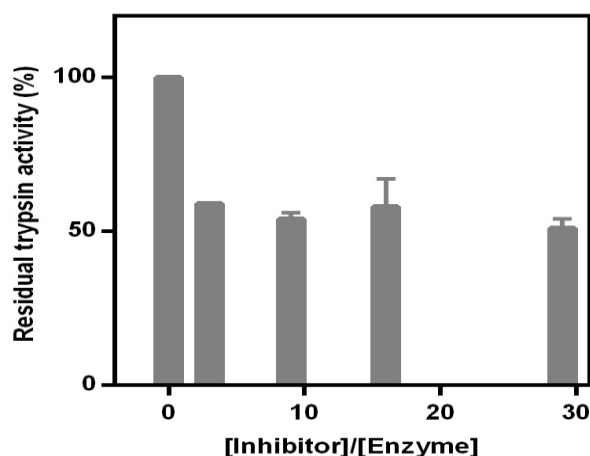


Fig. 5.3. Residual trypsin activity against BAEE measured after 60 sec in presence of various concentration of OPI.

The trypsin inhibitory activity of OPI was studied by monitoring the rate of hydrolysis of BAEE by trypsin at 253 nm colorimetrically and adding increasing concentration of OPI to the reaction mixture. The percent residual trypsin activity was calculated thereafter and plotted against the molar ratio of OPI and trypsin as shown in Fig. 5.3. It can be seen from the plot that in the molar ratio range 3 to 29, the residual trypsin activity remains ~50% suggesting that OPI inhibits the activity of trypsin rather weakly.

5.4.2 Secondary structure

The secondary structure of OPI was determined by circular dichroism (CD) spectroscopy. The far-UV CD spectrum of the protein was characterized by a negative band around 195 nm and a slightly positive shoulder around 230 nm (Fig. 5.4, A). Analysis of the CD spectral data employing the CDSSTR programme [Sreerama and Woody, 2000] available online (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) showed that the secondary structure of

this protein contains about 5% α -helix, 39% β -sheets, 23% turns and 34% unordered structures (Fig. 5.4, B).

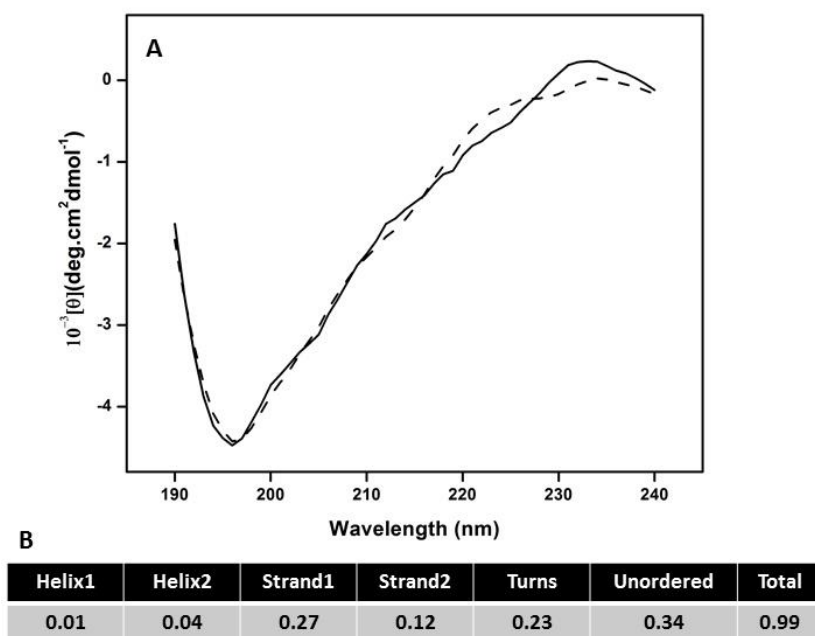


Fig. 5.4. Far-UV CD spectrum of native OPI (A) and its secondary structure content deduced from analysis by CDSSTR method (B).

5.4.3 De-novo sequencing and primary structure

The amino acid sequence of OPI was obtained by *de-novo* sequencing of peptides derived from proteolytic cleavages or chemical fragmentation, i.e. acid hydrolysis. Various proteolytic enzymes (trypsin, chymotrypsin, thermolysin, endoproteinase Glu-C and mixtures of trypsin and chymotrypsin) were employed in this study. Fragmentation of doubly charged precursor peptide ions with m/z 1116.04⁺² derived from thermolysin digest of reduced and alkylated OPI demonstrated the peptide sequence AAHDAVLVDVDGDELLVGVPYY (aa₁₋₂₁) (cf. Fig. 5.5) corresponding to the N-terminal sequence.

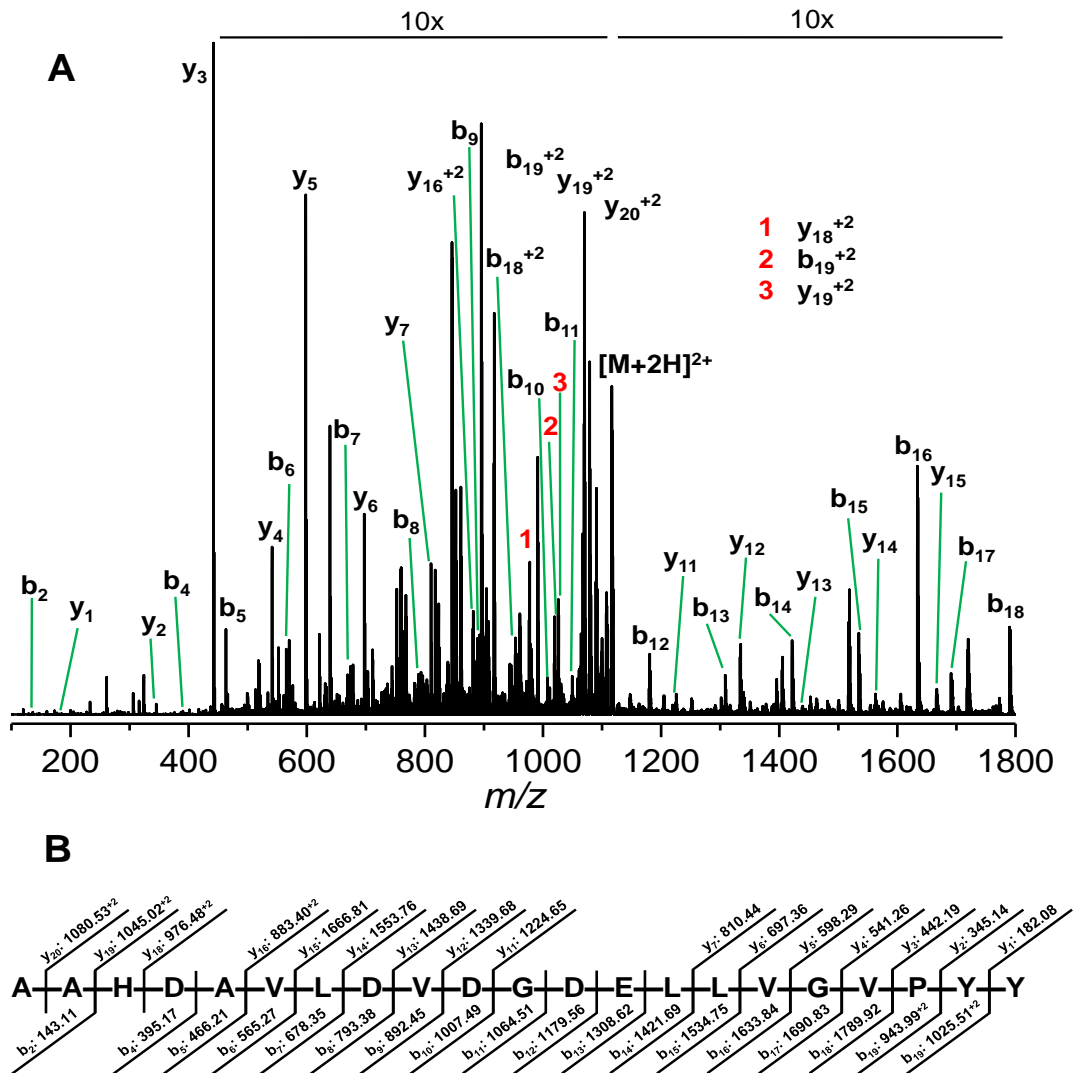


Fig. 5.5. A) NanoESI fragment ion spectrum obtained from a low-energy CID experiments on the doubly charged precursor peptide ions with m/z 1116.04 derived from an in-solution proteolytic thermolysin digestion of reduced and alkylated OPI comprising the *N*-terminus (aa: 1-21). B) Corresponding fragmentation scheme.

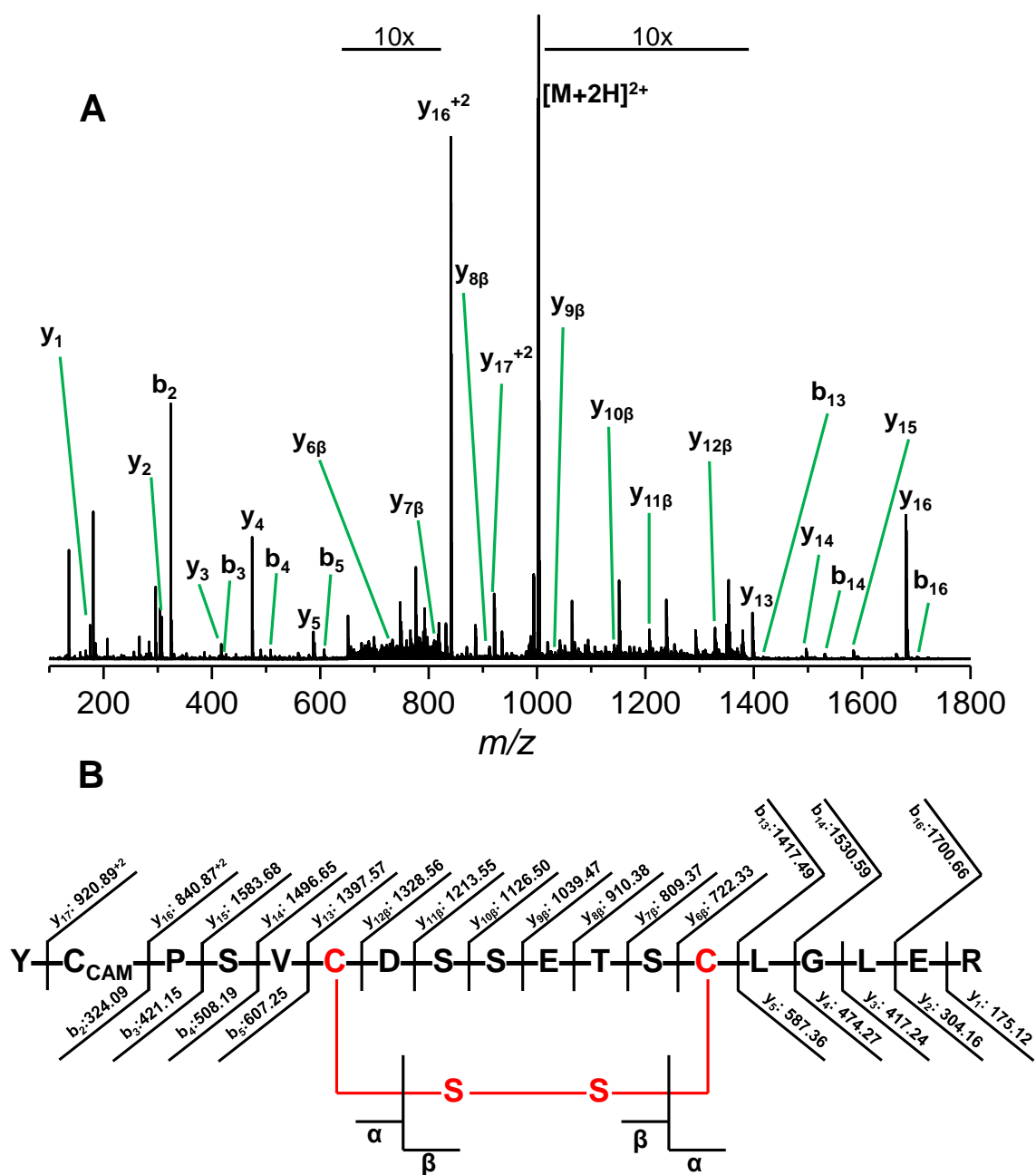


Fig. 5.6. A) NanoESI fragment ion spectrum obtained from a low-energy CID experiments on the doubly charged precursor peptide ions with m/z 1002.41 derived from an in-solution proteolytic digestion of alkylated OPI using trypsin as protease. B) Corresponding fragmentation scheme. The intrapeptide disulfide bridge is highlighted in red (C_{CAM}: carboxymethylated cysteine).

Three peptides were identified harboring an intramolecular disulfide bridge that joins C₁₄₈ and C₁₅₅. The S—S bridge was determined by low energy CID experiments as shown previously [Mormann et al., 2008]. Fragmentation of the doubly charged precursor peptide ion at m/z 1002.41⁺² derived from a tryptic digest of alkylated OPI yielded product ions that allowed determination of the disulfide linkage (cf. Fig. 5.6). Besides formation of b- and y-type ions formed by cleavage of the amide bonds outside the disulfide loop, diagnostic fragment ions giving rise to information on the amino acid sequence inside the disulfide loop are observed. Asymmetric cleavage of the disulfide bond is followed by the formation of a dehydroalanine residue on the C—S cleavage site and a modified cysteine comprising a sulfhydryl moiety associated with mass shifts of −34 u for b-type ions and +32 u for y-type ions, respectively (cf. Fig. 5.6). The observed fragmentation pattern unequivocally indicates a disulfide linkage between C₁₄₈ and C₁₅₅.

Evaluation of the fragment ion spectra of almost 50 different peptide species (cf. Table 5.1) and subsequent sequence alignment to a putative 21 kDa trypsin inhibitor from *Theobroma bicolor* (Q3BD85) yielded over 80% of the amino acid sequence. The finally obtained sequence of OPI and its alignment with 4 other proteins with known trypsin inhibitor or α -amylase/subtilisin inhibitor activities are depicted in Fig. 5.7.

OPI	-----AAHDAVLVDVDGDELLVGVPYYVVSVLW-----	32
Q3BD85	MKTATAVVLLLLAFTSKSYFFG---VANAANSPVLDTDGDELRTGVRYVVSISWGAGGG	57
A0A061G2K6	MKTATAVVLLLLAFTSKSYFFG---VANAANSPVLDTDGDELQTGVQYYVLSSISGAGGG	57
1604472A	MGSRR-----AGSSSSPLFWPAPPSRAADPPPVHDTDGHELADANYVLSANR-AHGG	53
1BA7_B	-----DFVLDNEGNPLENGGTYIILSDITA--FG	27
OPI	-----FCPELVVQNADSLDKANPVVFSNADSS--DGVVRVSSDVRLEFVGPR--	87
Q3BD85	GLALGRATGQSCPEIIVVQRRSDLNNGTPVIFSNADSK--DGVVRLSTDINIEFVPIRD-	113
A0A061G2K6	GLALGRATGQSCPEIIVVQRRSDLNNGTPVIFSNADSK--DDVVRVSTDVNIEFVPIRD-	113
1604472A	GLTMAPGHGRHCPLFVSQDPNGQHDGFPVRITPYGVAPSDKIIRLSTDVRISFRAYT--	110
1BA7_B	GIRAAPTGNERCPLTVVQSRNELDKIGITIISSPYRIR---FIAEGHPLSLKFDSFAVI	83
OPI	-----QVQGPGESSGKRWVELGGSEGEPCD TVKSWFKLESANM--PGSYKFKYC	144
Q3BD85	RLC-STSTVWKLDNYDNSAGKWWVTDDGVRGQPGPNTLT SWFKIERAGV--LG-YKFRFC	169
A0A061G2K6	RLC-STSTVWRLDNYDNSAGKWWVTDDGKGEPCGPNLTCSWFKIEKAGV--LG-YKFRFC	169
1604472A	-TC-LQSTEWHDSEL-AAGRRHVIT-GPVKDPSPSGRENAFRIEKYSGAEVHEYKLMSC	166
1BA7_B	MLCVGIPTEWSVVEDLPE-GP-----AVKIGENKDAMDGWFRLERVSDDEFNNYKLVFC	136
OPI	PSVCDSSSETSCGLGLER--DEDLLVLSDTVDFPWP--WVFLKAYGKD-----	188
Q3BD85	PSVCDSCKTLCSDIGRHSDDDGQIRLALSDNGWP--WMFKKASKTIKQVVNAKH	221
A0A061G2K6	PSVCDSCCTLCSDIGRHSDDDGQIRLALSDNEWA--WMFKKASKTIKQVVNAKH	221
1604472A	GD-----WCQDLGVFRDLKGGAWFLGATEPYHV-VVFKKAPPA-----	203
1BA7_B	PQQAED--DKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESL-----	181

Fig. 5.7. Sequence alignment of okra protease inhibitor (OPI) with a putative trypsin inhibitor from *Theobroma bicolor* (Q3BD85), trypsin inhibitor from *Theobroma cacao* (A0A061G2K6), barley α -amylase/subtilisin inhibitor (1604472A) and soybean (kunitz) trypsin inhibitor (1BA7_B). Sequence similarities are highlighted in yellow. Gaps (-) were introduced to maximize homology and in case of OPI (-) represents undetermined amino acid residue.

Table 5.1. Summary of *de-novo* sequenced peptide ions obtained from in-gel digests and products of acid hydrolysis of okra protease inhibitor.

Peptide sequence	aa ^a	Protease ^b	m/z _{exp} ^{+z}	m/z _{calc.} ^{+z}
Y(CAM)PSVCDSSSETSCGLGLER ^{c,d}	143-160	T	1002.41 ⁺²	1002.41 ⁺²
Y(CAM)PSV(CAM)DSSETS(CAM)LGLER ^c	143-160	T	1060.45 ⁺²	1060.43 ⁺²
F(CAM)PELVVQNADSLDKANPVVFSNADSSDGVVR ^c	43-74	T	1150.56 ⁺³	1150.56 ⁺³

LEFVGPR	81-87	T	409.24 ⁺²	409.23 ⁺²
(MAM)(CAM)PELVVQNADSLDK ^{c,e}	43-57	T	592.61 ⁺³	592.62
LESANM _{ox} PGSYK ^f	131-141	T	606.78 ⁺²	606.77 ⁺²
NADSLDK	51-57	T	762.37 ⁺¹	762.36 ⁺¹
FKY(CAM)PSVCDSSETSCGLER ^{c,d}	141-160	T	760.33 ⁺³	760.00 ⁺³
PG(CAM)DTVK ^c	120-126	T	776.36 ⁺¹	776.35 ⁺¹
PVVFSN ADSSDGVVR	60-74	T	774.89 ⁺²	774.88 ⁺²
FKY(CAM)PSV(CAM)DSSETS(CAM)LGLER ^c	141-160	T	799.02 ⁺³	799.01 ⁺³
ANPVVFSNADSSDGVVR	58-74	T	867.43 ⁺²	867.43 ⁺²
WVELGGSEGEPPG(CAM)DTVK ^c	110-126	T	910.41 ⁺²	910.40 ⁺²
AHDAVLDDVDGDELLVGVPY	2-20	T	999.00 ⁺²	999.00 ⁺²
VELGGSEGEPPG(CAM)DTVKSWF ^c	111-129	CT	1027.47 ⁺²	1027.46 ⁺²
QVQGPGESSGKRW	98-110	CT	472.59 ⁺³	472.57 ⁺³
AHDAVLDDVDGDELLVGVPY ^g	2-20	CT	678.99 ⁺³	679.02 ⁺³
NADSLDKANPVVF	51-63	CT	695.36 ⁺²	695.35 ⁺²
SNADSSDGVVRVSSDVRLEF	64-83	CT	713.69 ⁺³	713.68 ⁺³
VLSDTVYVPWPW	168-179	CT	731.36 ⁺²	731.37 ⁺²
YVVSVLW	21-27	CT	865.49 ⁺¹	865.47 ⁺¹
RLVLSDTVYVPWPW	166-179	CT	865.96 ⁺²	865.96 ⁺²
AHDAVLDDVDGDELLVGVPYY	2-21	TL	1080.53 ⁺²	1080.52 ⁺²
AAHDAVLDDVDGDELLVGVPYY	1-21	TL	1116.05 ⁺²	1116.04 ⁺²
LESANMPGSYK	131-141	TL	598.78 ⁺²	598.78 ⁺²
FSNADSSDGVVR	63-74	TL	627.29 ⁺²	627.29 ⁺²
VQGPGESSGKRW(-OMe) ^h	99--110	TL	651.32 ⁺²	651.33 ⁺²
VVFSNADSSDGVVR	61-74	TL	726.36 ⁺²	726.36 ⁺²
LGGSE(-OMe)GEPG(CAM)DTVKSW ^{c,h}	113-128	TL	846.87 ⁺²	846.88 ⁺²
LVVQNADSLDKANPVVFSNADSSDGVVRVS	47-76	GluC	1034.89 ⁺³	1034.86 ⁺³
LGGSEGEPPG(CAM)DTVKS ^c	113-127	GluC	746.84 ⁺²	746.83 ⁺²

PVVFSNADSSDGVVVRVSSDVRLE	60-82	GluC	812.09 ⁺³	812.08 ⁺³
SANM _{ox} PGSYKFKYCPSVCDSE ^{c,f}	133-152	GluC	810.68 ⁺³	810.67 ⁺³
Y(CAM)PSVCDSSSETSCGLER ^{c,d}	143-160	T+CT ⁱ	1002.43 ⁺²	1002.41 ⁺²
M _{ox} (CAM)PELVVQNADSLDKANPVVF ^{c,f}	43-63	T+CT	1181.63 ⁺²	1181.57 ⁺²
VSSDVRLEF	75-83	T+CT	526.30 ⁺²	526.27 ⁺²
SNADSSDGVVVR	64-74	T+CT	553.78 ⁺²	553.76 ⁺²
KLESANM _{ox} PGSY ^f	130-140	T+CT	606.81 ⁺²	606.77 ⁺²
SSDVRLEFVGPR	76-87	T+CT	681.39 ⁺²	681.36 ⁺²
KY(CAM)PSV(CAM)DSSETS(CAM)LGLER ^c	142-160	T+CT	750.02 ⁺³	749.99 ⁺³
M(CAM)PELVVQNADSLDK ^c	43-57	T+CT	859.94 ⁺²	859.90 ⁺²
WVELGGSEGEPPG(CAM)DTVK ^c	110-126	T+CT	910.45 ⁺²	910.40 ⁺²
AHNAVLDVDGDELTLGAPY	2-20	T+CT	985.51 ⁺²	985.48 ⁺²
AHDAVLDVDGDELLVGVPY	2-20	T+CT	999.02 ⁺²	999.00 ⁺²
FPWPWVFLKAYGKD	175-188	AH	585.34 ⁺³	585.31 ⁺³
TVDFPWPWVFLKAYGKD	172-188	AH	690.38 ⁺³	690.35 ⁺³
DEDLLVLSDTVDFPWPWVFLKAYGKD	163-188	AH	767.93 ⁺⁴	767.89 ⁺⁴
SLDKANPVVFSNADSSD	54-70	AH	883.45 ⁺²	883.41 ⁺²

^aaa – amino acid; ^bT – trypsin, CT – chymotrypsin, TL – thermolysin, GC – endoproteinase
GluC, AH – acid hydrolysis; ^cCAM – carbamidomethyl cysteine; ^dcysteines forming disulfide
bridge; ^eMAM – carbamidomethyl methionine; ^fM_{ox} – methionine in oxidised form, ^gions
detected as [M+2H+K]³⁺; ^hOMe – *O*-methylated form; 1:1 (vol:vol) mixture of trypsin and
chymotrypsin solution.

5.5 Discussion

Kunitz protease inhibitors are a distinct family of PIs which are active against mostly serine
proteases. Kunitz PIs are abundant in leguminous plants and have been studied extensively.
However, PIs from other families of plants are comparatively less explored. The present study

was carried out to isolate and characterize a Kunitz protease inhibitor from the seeds of okra, a common edible vegetable in India and around the world which belongs to the mallow family.

Other than okra, the family of Malvaceae (mallow) includes two very well-known plants: cotton and cacao. Reports on the nucleic acid sequence of a 21 kDa, Kunitz-type protease inhibitor from cacao (*Theobroma cacao*) showed that the protein shares significant sequence homology with barley and wheat α -amylase/subtilisin inhibitors as well as with soybean (Kunitz) trypsin inhibitor [Spencer and Hodge, 1991; Tai et al., 1991]. Although the exact protease specificity of this 21 kDa protease inhibitor was not known at that time, a later report on the major ~20 kDa albumin protein from *Theobroma cacao* which shared high sequence homology (95%) with the former was shown to possess potent trypsin inhibitory activity with a dissociation constant (K_i) of 9.5×10^{-8} M [Kochhar et al., 2000]. In all cases, the mature polypeptide was shown to undergo post-translational modification by trimming off the signal peptide. On the other hand, previously reported isolation and subsequent characterization of four inhibitors of trypsin from okra (*Abelmoschus esculentus* L.) showed strong trypsin inhibitory activity forming a 1:1 complex and Arg residues were found to be important for their activity [Ogata et al., 1986]. However, due to lack of even partial amino acid sequence information for any of those trypsin inhibitors it is difficult to know whether the protease inhibitor purified by us is similar to any previously purified okra trypsin inhibitor.

The trypsin inhibitory activity of OPI assayed by monitoring activity of trypsin against hydrolysis of BAEE suggested that OPI is a rather weak inhibitor of trypsin. A thorough investigation on the inhibitory activity of OPI against other serine proteases such as chymotrypsin and subtilisin or α -amylase is required before its protease specificity can be clearly established.

The structure of soybean trypsin inhibitor (Kunitz) and some other Kunitz trypsin inhibitors is composed of intricate folds of β -sheets which is known as β -trefoil fold [Song and Suh, 1998]. In case of OPI, the majority of secondary structural elements were found to consist of β -sheets with very little α -helical content. The far-UV CD spectrum of OPI is characterised by a sharp minimum at the lower wavelength side (195 nm) of the spectrum. Similar observations were reported for a serine protease inhibitor from *Derris trifoliata* L. seeds with potent trypsin and chymotrypsin inhibitory activity [Bhattacharyya and Babu, 2009].

Mass spectrometric study on OPI has allowed for the determination of a total 162 amino acid residues in the primary structure of the protein. Online homology search on the amino acid sequence of OPI led to the identification of a 21 kDa putative trypsin inhibitor from *Theobroma bicolor* (Q3BD85) as its closest match with 43% sequence identity, considering the mature polypeptide of Q3BD85 is devoid of the 25-amino acid long signal peptide. This also indicates that about 83% of OPI has been sequenced. Similarly, OPI shared ~43% sequence identity with a trypsin inhibitor from *Theobroma cacao* (A0A061G2K6) and ~24% sequence identity with the barley α -amylase/subtilisin inhibitor (1604472A). Its homology with the soybean (Kunitz) trypsin inhibitor (~25% identity), its apparent molecular weight and secondary structure further indicate that OPI belongs to the family of Kunitz protease inhibitors. Kunitz protease inhibitors generally contain two intramolecular disulphide linkages [Mosolov and Valueva, 2005]. However, only one disulphide linkage between C₁₄₈ and C₁₅₅ in OPI was determined. In soybean trypsin inhibitor (1BA7_B) two disulphide bridges between C₃₉-C₈₆ and C₁₃₆-C₁₄₅ are formed [Song and Suh, 1998]. Fig. 5.7 indicates that C₃₉ is conserved in all the protease inhibitors and so is C₈₆, however, that position remains undetermined in OPI.

It should be noted that we have referred to the ion-exchange eluted fraction containing two protein bands (Fig. 5.2, Lane 3) as the okra protease inhibitor mixture throughout the text due to the fact that partial amino acid sequence obtained for the other two peaks observed in HPLC by nanoESI-Q-ToF mass spectrometry revealed that their N-terminal sequences are quite similar to that of OPI (See Fig. 5.8).

OPI	AAHDAVLVDVGDELLVGVPYYVVSVLW
Peak1/2	AAGNAVLDRDGDEVLTGVRYFVVSALW
Peak1/2	AHDAVMDADGDELLVGAPYY

Fig. 5.8. Comparison of N-terminal sequences of OPI with two other protein peaks (peak 1 and peak 2) seen in HPLC-chromatogram.

Based on sequence homology with other known protease inhibitors, the physiological function of OPI can be assigned as the inhibition of foreign protease activity or regulation of endogenous proteolytic activities. Its occurrence in the seeds further confirms its role as a storage protein. Therefore, OPI might remain as a storage protein in the seeds of okra and also act as a PI to protect the seeds from foreign proteases whenever necessary.

General Discussion and Conclusions

Chapter 6

6.1 General discussion and conclusions

This thesis presents work done on the purification, macromolecular characterization and primary structure determination of two proteins namely, mulberry (*Morus indica*) latex galactose-specific lectin (MLGL) and okra (*Abelmoschus esculentus*) protease inhibitor (OPI).

Plant latex is a milky sap exuded under mechanical stress or insect herbivory. It is stored in the laticifer tissue of plants and contains many proteins including enzymes and chemicals such as terpenoids and alkaloids. The function of plant latex has been the subject of much speculation. Excretion of waste metabolites, protection of damaged tissue and defense against insect herbivores and pathogens are few of the proposed hypotheses for its role in plants. However, most evidence support its involvement in the defense against insect herbivores [Konno, 2011]. For example, the latex of mulberry (*Morus* spp.) contains mulatexin (MLX56), a 56 kDa protein with an inactive chitinase-like domain and two hevein-related chitin binding domains. MLX56 is found to be highly toxic towards many insects in the order of Lepidopteran [Wasano et al., 2009]. An interesting fact is that MLX56 is not toxic to *Bombyx mori*, which feeds on mulberry leaves to subsequently produce silk filaments and perhaps has grown adapted to the mulberry defense system. In another study, the latex of white mulberry (*Morus alba*) was shown to accumulate defense related proteins under biotic stress [Kitajima et al., 2013]. Lectins have been characterized extensively from the vegetative organs of plants, such as, the seeds, leaves, bark or tubers. However, other than storage proteins no definitive function could be attributed to most of these lectins. Our laboratory has been working on the Cucurbitacea phloem proteins for almost a decade now [Narahari and Swamy, 2010; Narahari et al., 2011; Nareddy et al., 2017], which appear to play a major role in wound sealing. In view

of this, we chose to purify lectins from a relatively less explored source of plants, such as the latex.

Chapter 2 deals with the purification, macromolecular characterization, carbohydrate specificity and cytotoxicity of MLGL, a jacalin-like lectin from mulberry (*Morus indica*) latex. Previously, purification of MIL, another jacalin-related lectin from mulberry latex using ion-exchange and hydrophobic interaction chromatography was reported by Patel et al. [2011]. Prior to purification, we presumed that the protein to be eluted through the guar-gum column is MIL. However, due to lack of detailed carbohydrate specificity of MIL and later comparison of its available amino acid sequence stretches with that of MLGL (see chapter 3), we inferred that MIL is different from MLGL. The results presented in this chapter demonstrates that MLGL shows anomeric preference towards α -form of galactose and weakly binds to simple galactose in which the β -isomer prevails. The surface of MLGL is negatively charged at neutral pH and its secondary structure is predominantly made up of β -sheets. MLGL is thermally quite stable, showing a denaturation temperature of 77.6°C as determined by differential scanning calorimetry. MLGL belongs to the galactose-specific jacalin related lectins (gJRL) family, which recognizes the Thomsen-Friedenreich antigen (Gal β 1-3GalNAc) with high affinity. TF/T-antigen is one of the extensively studied tumor markers. It is found on the surface of breast cancer cells at a higher level than normal cells. Therefore, in order to check if MLGL can selectively exert cytotoxic effect towards breast cancer cells than other normal cells we have performed MTT cell viability assays with MDCK and MCF-7 cell lines after incubating them with MLGL for 72 h. Interestingly, we observed that MLGL is cytotoxic towards both the cell lines even at micromolar concentrations. However, the mechanism of MLGL cytotoxicity cannot be explained by MTT assays alone but it can be assumed that interactions

with cell surface glycans play a major role in exhibiting such behaviour. Nevertheless, the cytotoxic nature of MLGL can perhaps be related to defense of mulberry latex against biotic stress.

The complete primary as well as the glycan structures of MLGL determined by nanoESI-Q-ToF mass spectrometry have been presented in **Chapter 3**. MLGL is a homotetrameric lectin with a subunit mass of 18 kDa. Each of its subunits has two amino acid chains, a longer one with 133 amino acid residues and a shorter one with 21-24 residues. The longer chain is glycosylated on Asn₂₁ with either a pauci-mannose type or a complex type glycan. Pauci-mannose is a plant specific glycan structure and has been found in many other plant proteins such as jacalin, jack bean α -mannosidase [Kumar et al., 2014] and *Trichosanthes dioica* seed lectin (unpublished data). However, in case of MLGL, two types of glycan structures were detected at the same glycosylation site (Asn₂₁) which indicates the presence of heterogeneity in MLGL even in terms of its glycan structures. In addition to mass spectrometric analysis of the glycan structure, the presence of core β (1, 2) linked xylose was assessed by performing dot-blot assay with an anti-xylose antibody. We have been able to determine the complete primary structure of MLGL, which resembles that of MornigaG with a few minor differences in the amino acid residues. MornigaG is a gJRL obtained from the bark of black mulberry (*Morus nigra*). However, MornigaG differs from MLGL in that it recognizes both simple galactose and Me α Gal with similar affinity whereas MLGL binds to simple galactose rather weakly.

Chapter 4 describes studies on the exposure and accessibility of tryptophan residues in MLGL and its chemical, thermal as well as pH-induced modes of denaturation assessed by fluorescence and circular dichroism spectroscopy. The tryptophan residues in proteins could

reside either on the surface or inside the hydrophobic core of the protein giving rise to different microenvironments which can be detected by monitoring tryptophan fluorescence characteristics and also by employing different quenchers. The two tryptophan residues of MLGL were probed by neutral (acrylamide) and ionic (iodide and cesium ions) quenchers in the native and denatured conditions as well as in the presence of Me α Gal. The results of such experiments suggested that tryptophan residues are buried inside the protein and upon denaturation become fully exposed. Also, the tryptophan residues are surrounded by charged amino acids which hinder the ionic quenchers to access these residues. In all cases, quenching takes place by collisions; however, in the denatured protein acrylamide can complex with the fluorophores, resulting in static quenching alongside dynamic quenching. Even though MLGL is devoid of any disulphide linkages, the protein was found to be highly stable. Chemical denaturants such as urea and guanidine hydrochloride were unable to denature the protein even at their saturated concentrations, whereas guanidine thiocyanate, a relatively stronger denaturant than the other two could denature the protein above a concentration of 4 M. This high stability of MLGL could be attributed to its glycosylation, which has been shown to impart structural stability in many proteins including some lectins [Mitra et al., 2003; Sinha et al., 2005]. Changes in the secondary and tertiary structures of MLGL with increasing temperature and at different pH conditions were monitored by circular dichroism spectroscopy. Both secondary and tertiary structures of MLGL were stable up to 70°C and were disrupted above this temperature corroborating with the observed denaturation temperature of 77.6°C seen in DSC. The structure of MLGL was found to be stable in the pH range 6.2 to 8.5 whereas below this pH the secondary structure content started reducing while the tertiary structure was found

to remain intact. MLGL retained its carbohydrate binding ability even when heated to 70°C as evident from its hemagglutination activity at this temperature.

Unlike lectins, the role of protease inhibitors as defense proteins is already well established. Protease inhibitors accumulate in high amount in specific tissues after insect herbivory or wounding [Zhu-Salzman and Zeng, 2014]. **Chapter 5** deals with the purification and structural characterization of a Kunitz-type protease inhibitor from okra seeds. Okra is a common vegetable eaten around the world and has been reported to exhibit medicinal properties. There has been a report on the purification and trypsin inhibitory activity of four protease inhibitors from okra seeds; however, no sequence information was available for them. Since protease inhibitors are products of multigene families there could be a number of isoinhibitors for a particular inhibitor family [Ryan, 1990]. In addition to this, there could be protease inhibitors of different families as well. This can make the number of protease inhibitors present in a particular organ very high. Therefore, without sequence information these proteins cannot be assigned to a particular family. We have also followed a different purification method than what was reported for those four trypsin inhibitors. We have used a method combining ammonium sulphate precipitation, anion-exchange chromatography and RP-HPLC to purify the okra protease inhibitor (OPI). The secondary structure of OPI was found to consist of mainly β -sheets and its partial amino acid sequence determined by nanoESI-Q-ToF mass spectrometry suggested that it belongs to the Kunitz protease inhibitor family. Approximately 83% of the protein was sequenced and an intramolecular disulfide linkage between C₁₄₈ and C₁₅₅ could be detected as well. Although its inhibitory activity towards trypsin was found to be weak, its sequence homology with some known protease inhibitors such as soybean trypsin (Kunitz) inhibitor and barley α -amylase/subtilisin inhibitor and a

putative trypsin inhibitor from *Theobroma cacao* suggested that it might function as an inhibitor of serine proteases *in vivo*.

6.2 Future prospects

Although a thorough investigation on the carbohydrate-specificity of MLGL was performed by hemagglutination-inhibition assays using various mono- and di-saccharides, important thermodynamic parameters for binding such as the binding constant, stoichiometry, enthalpy and entropy of binding could not be determined due to sample limitation. Therefore, it would be interesting to determine these parameters employing isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) and also using other carbohydrates such as the T-antigenic disaccharide. With the complete amino acid sequence available now, MLGL can be cloned and expressed in high yield using bacterial expression system. Furthermore, from mutational analysis of recombinant MLGL more information can be obtained on the amino acid residues important for its carbohydrate binding. Crystal structure of MLGL can also be solved by X-Ray crystallography as structures of very few gJRLs are known so far.

The inhibitory potency of OPI against various serine proteases such as chymotrypsin, subtilisin etc. or other enzymes such as α -amylase can be investigated by colorimetric methods. Also, it would be interesting to isolate and characterize the other two protease inhibitors designated as peak1 and peak2 on HPLC-chromatogram.

References

- Alberts B, Johnson A, Lewis J, Walter P, Raff M, Roberts K. (2002) Molecular Biology of the Cell. 4th edition. International student edition. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26878/>.
- Ambrosi M, Cameron NR, Davis BG. (2005) Lectins: tools for the molecular understanding of the glycode. *Org Biomol Chem.* **3**: 1593-1608.
- Appukuttan PS, Surolia A, Bachhawat BK. (1977) Isolation of two galactose-binding proteins from *Ricinus communis* by affinity chromatography. *Ind. J. Biochem. Biophys.* **14**: 382-384.
- Astoul CH, Peumans WJ, Van Damme EJM, Barre A, Bourne Y, Rougé P. (2002) The size, shape and specificity of the sugar-binding site of the jacalin-related lectins is profoundly affected by the proteolytic cleavage of the subunits. *Biochem J.* **367**: 817-824.
- Barre A, Van Damme EJM, Peumans WJ, Rouge P. (1996) Structure-function relationship of monocot mannose-binding lectins. *Plant physiol.* **112**: 1531-1540.
- Barrett AJ, Rawlings ND, Woessner JF. Handbook of Proteolytic Enzymes. Academic, New York, (1998).
- Bhanu K, Komath SS, Maiya BG, Swamy MJ. (1997) Interaction of porphyrins with concanavalin A and pea lectin. *Curr Sci.* **73**: 598-602.
- Bhattacharyya A, Babu CR. (2009) Purification and biochemical characterization of a serine proteinase inhibitor from *Derris trifoliata* Lour. seeds: insight into structural and antimalarial features. *Phytochemistry*, **70**: 703-712.
- Birk Y. (1985) The Bowman-Birk inhibitor. Trypsin-and chymotrypsin-inhibitor from soybeans. *Int J Pept Protein Res.* **25**: 113-131.

References

- Bogoeva, VP, Radeva MA, Atanasova LY, Stoitsova SR, Boteva RN. (2004) Fluorescence analysis of hormone binding activities of wheat germ agglutinin. *Biochim Biophys Acta Proteins Proteom.* **1698**: 213-218.
- Boyd WC, Shapleigh E. (1954) Specific precipitating activity of plant agglutinins (lectins), *Science*, **119**: 419.
- Campana PT, Moraes DI, Monteiro-Moreira AC, Beltramini LM. (2002) Unfolding and refolding studies of frutalin, a tetrameric d-galactose binding lectin. *Eur. J. Biochem.* **269**: 753-758.
- Cazet A, Julien S, Bobowski M, Burchell J, Delannoy P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res.* **12**: 204.
- Craik CS, Rocznik S, Largman C, Rutter WJ. (1987) The catalytic role of the active site aspartic acid in serine proteases. *Science*, **237**: 909–913
- Czapinska H, Otlewski J. (1999) Structural and energetic determinants of the S1-site specificity in serine proteases. *Eur J Biochem.* **260**: 571-595.
- Dan X, Liu W, Ng TB. (2016) Development and applications of lectins as biological tools in biomedical research. *Med Res Rev.* **36**: 221-247.
- Datta D, Pohlentz G, Schulte M, Kaiser M, Goycoolea FM, Müthing J, Mormann M, Swamy MJ. (2016) Physico-chemical characteristics and primary structure of an affinity-purified α -D-galactose-specific, jacalin-related lectin from the latex of mulberry (*Morus indica*). *Arch Biochem Biophys.* **609**: 59-68.

- Dimick SM, Powell SC, McMahon SA, Moothoo DN, Naismith JH, Toone, EJ. (1999) On the meaning of affinity: cluster glycoside effects and concanavalin A. *J Am Chem Soc.* **121**: 10286-10296.
- Dolgikh DA, Kolomiets AP, Bolotina IA, Ptitsyn OB. (1984) 'Molten-globule' state accumulates in carbonic anhydrase folding. *FEBS Lett.* **165**: 88-92.
- Drickamer K, Taylor ME. (1993) Biology of animal lectins. *Annu Rev cell biology*, **9**: 237-264.
- Elgavish S, Shaanan B. (1997) Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.* **22**:462-467.
- Eftink MR, Ghiron CA. (1976) Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry* **15**: 672-680.
- Eftink MR, Ghiron CA. (1981) Fluorescence quenching studies with proteins. *Anal. Biochem.* **114**:199–227.
- Evangelio E, Poiroux G, Culerrier R, Pratviel G, Van Damme EJM, Peumans WJ, Barre A, Rougé P, Benoist H, Pitié M. (2011) Comparative study of the phototoxicity of long-wavelength photosensitizers targeted by the mornigag lectin. *Bioconjugate Chem.* **22**: 1337-1344.
- Favero J, Corbeau P, Nicolas M, Benkirane M, Travé G, Dixon JF, Aucouturier P, Rasheed S, Parker JW, Liautard JP, Devaux C. (1993) Inhibition of human immunodeficiency virus infection by the lectin jacalin and by a derived peptide showing a sequence similarity with GP120. *Eur J Immunol.* **23**: 179-185.
- García-Olmedo F, Salcedo G, Sánchez-Monge R, Gómez L, Royo J, Carbonero P. (1987) Plant proteinaceous inhibitors of proteinases and alpha-amylases. *Oxf Surv Plant Mol Cell Biol.* **4**: 275-334.

- Girbés 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.
- Glinsky VV, Glinsky GV, Rittenhouse-Olson K, Huflejt ME, Glinskii OV, Deutscher SL, Quinn TP. (2001) The role of Thomsen-Friedenreich antigen in adhesion of human breast and prostate cancer cells to the endothelium. *Cancer Res.* **61**: 4851-4857.
- Goel M, Anuradha P, Kaur KJ, Maiya BG, Swamy MJ, Salunke DM. (2004) Porphyrin binding to jacalin is facilitated by the inherent plasticity of the carbohydrate-binding site: novel mode of lectin–ligand interaction. *Acta Crystallogr Sect D: Biol Crystallogr.* **60**: 281-288
- Grinvald A, Steinberg IZ. (1974) On the analysis of fluorescence decay kinetics by the method of least squares. *Anal. Biochem.* **59**: 583–598.
- Gupta G, Surolia A. (2012) Glycomics: An Overview of the Complex Glycocode. In: Sudhakaran P., Surolia A. (eds) *Biochemical Roles of Eukaryotic Cell Surface Macromolecules*. Advances in Experimental Medicine and Biology, vol 749. Springer, New York, NY.
- Gupta G, Surolia A, Sampathkumar SG. (2010) Lectin microarrays for glycomic analysis. *OMICS.* **14**: 419-36.
- Hilder VA, Gatehouse AM, Sheerman SE, Barker RF, Boulter D. (1987) A novel mechanism of insect resistance engineered into tobacco. *Nature*, **330**: 160-163.
- Hirabayashi J, Yamada M, Kuno A, Tateno H. (2013) Lectin microarrays: concept, principle and applications. *Chem Soc Rev.* **42**: 4443-4458.

- Idrees D, Prakash A, Md. Haque A, Islam A, Md. Hassan I, Ahmad F. (2016) GdnHCl-induced unfolding intermediate in the mitochondrial carbonic anhydrase VA. *Int. J. Biol. Macromol.* **91**: 1151-1160.
- Inokuti M, Hirayama F. (1965) Influence of energy transfer by the exchange mechanism on donor luminescence. *J. Chem. Phys.* **43**: 1978–1989.
- International Human Genome Sequencing Consortium. (2004) "Finishing the euchromatic sequence of the human genome." *Nature*, **431**: 931-945.
- Jagannadham MV, Balasubramanian D. (1985) The molten globular intermediate form in the folding pathway of human carbonic anhydrase B. *FEBS Lett.* **188**: 326-30.
- Jones JD, Dangl JL. (2006) The plant immune system. *Nature*, **444**: 323-329.
- Johannes L, Römer W. (2010) Shiga toxins-from cell biology to biomedical applications. *Nature Rev Microbiol.* **8**: 105-116.
- Kabir S. (1998) Jacalin: a jackfruit (*Artocarpus heterophyllus*) seed-derived lectin of versatile applications in immunobiological research. *J Immunol Meth.* **212**: 193-211.
- Kaiser M, Pereira S, Pohl L, Ketelhut S, Kemper B, Gorzelanny C, Galla HJ, Moerschbacher BM, Goycoolea FM. (2015) Chitosan encapsulation modulates the effect of capsaicin on the tight junctions of MDCK cells. *Sci Rep.* **5**: 10048.
- 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.* **26**: 1075-1084.
- Kavitha M, Bobbili KB, Swamy MJ. (2010) Differential scanning calorimetric and spectroscopic studies on the unfolding of *Momordica charantia* lectin. Similar modes of thermal and chemical denaturation. *Biochimie.* **92**: 58-64.

- Kenoth R, Swamy MJ. (2003) Steady-state and time-resolved fluorescence studies on *Trichosanthes cucumerina* seed lectin. *J. Photochem. Photobiol. B: Biol.* **69**: 193-201.
- Kessler A, Baldwin IT. (2002) Plant responses to insect herbivory: the emerging molecular analysis. *Annu Rev Plant Biol.* **53**: 299-328.
- Khan JM, Qadeer A, Ahmad E, Ashraf R, Bhushan B, Chaturvedi SK, Rabbani G, Khan RH. (2013) Monomeric banana lectin at acidic pH overrules conformational stability of its native dimeric form. *PLoS One* **8**: e62428.
- Kilpatrick DC. (2002) Animal lectins: a historical introduction and overview. *Biochim Biophys Acta.* **1572**: 187-197.
- Kitajima S, Kamei K, Taketani S, Yamaguchi M, Kawai F, Komatsu A, Inukai Y. (2010) Two chitinase-like proteins abundantly accumulated in latex of mulberry show insecticidal activity. *BMC Biochemi.* **11**: 6.
- Kitajima S, Yamamoto Y, Hirooka K, Taki C, Hibino S. (2013) Laticifers in mulberry exclusively accumulate defense proteins related to biotic stresses. *Plant Biotechnol.* **30**: 399-402.
- Kochhar S, Gartenmann K, Juillerat MA. (2000) Primary structure of the abundant seed albumin of *Theobroma cacao* by mass spectrometry. *J Agric Food Chem.* **48**: 5593-5599.
- 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: Biol.* **50**: 108-118.
- Komath SS, Kavitha M, Swamy MJ. (2006) Beyond carbohydrate binding: new directions in plant lectin research. *Org Biomol Chem.* **4**: 973-988.

- Komath SS, Bhanu K, Maiya BG, Swamy MJ. (2000) Binding of porphyrins by the tumor-specific lectin, jacalin [Jack fruit (*Artocarpus integrifolia*) agglutinin]. *Biosci Rep.* **20**: 265-276.
- Konno K. (2011) Plant latex and other exudates as plant defense systems: roles of various defense chemicals and proteins contained therein. *Phytochemistry.* **72**: 1510-1530.
- Konno K, Ono H, Nakamura M, Tateishi K, Hirayama C, Tamura Y, Hattori M, Koyama A, Kohno K. (2006) Mulberry latex rich in antidiabetic sugar-mimic alkaloids forces dieting on caterpillars. *Proc Natl Acad Sci USA.* **103**: 1337-1341.
- Kostrominova TY. (2011) Application of WGA lectin staining for visualization of the connective tissue in skeletal muscle, bone and ligament/tendon tissue studies. *Microsc Res Techniq.* **74**: 18-22.
- Kraut J. (1977) Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem.* **46**: 331-358.
- Kumar BSG, Pohlentz G, Schulte M, Mormann M, Kumar NS. (2014) Jack bean α -mannosidase: Amino acid sequencing and N-glycosylation analysis of a valuable glycomics tool. *Glycobiology.* **24**: 252-261.
- Kumar CS, Sivaramakrishna D, Ravi SK, Swamy MJ. (2016) Fluorescence investigations on choline phospholipid binding and chemical unfolding of HSP-1/2, a major protein of horse seminal plasma. *J. Photochem. Photobiol. B: Biol.* **158**: 89-98.
- Kunitz M. (1946). Crystalline soybean trypsin inhibitor. *J Gen Physiol.* **29**: 149.
- Laemmli UK. (1970) Most commonly used discontinuous buffer system for SDS electrophoresis. *Nature.* **227**: 680-685.
- Lakowicz JR. (1999) Principles of Fluorescence Spectroscopy, 2nd Edition, Kluwer Academic Publishers, New York, 1-698 pp.

- 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.
- Lannoo N, Van Damme EJM. (2014) Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* **5**: 1-16.
- Laskowski Jr M, Kato I. (1980) Protein inhibitors of proteinases. *Ann. Rev Biochem.* **49**: 593-626.
- Laskowski M, Qasim MA. (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1477**: 324-337.
- Lee X, Thompson A, Zhang Z, Ton-that H, Biesterfeldt J, Ogata C, Xu L, Johnston RA, Young NM. (1998) Structure of the Complex of Maclura pomifera Agglutinin and the T-antigen Disaccharide, Gal β 1, 3GalNAc. *J Biol Chem.* **273**: 6312-6318.
- Lehrer SS. (1971) Solute perturbation of protein fluorescence. The quenching of tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry*, **10**: 3254–3263.
- Lipke H, Fraenkel GS, Liener IE. (1954) Growth inhibitors. Effect of soybean inhibitors on growth of *Tribolium confusum*. *J Agric Food Chem.* **2**: 410-414.
- Lis H. and Sharon N. (1973) The biochemistry of plant lectins (phytohemagglutinins). *Ann Rev Biochem.* **42**: 541-574.
- Lis H, Sharon N. (1986) Lectins as molecules and as tools. *Ann Rev Biochem.* **55**: 35-67.
- Mahanta SK, Sanker S, Rao NV, Swamy MJ, Surolia A. (1992) Primary structure of a Thomsen-Friedenreich-antigen-specific lectin, jacalin [*Artocarpus integrifolia* (jack fruit) agglutinin]. Evidence for the presence of an internal repeat. *Biochem. J.* **284**: 95-101.

- Mahanta SK, Sastry MK, Surolia, A. (1990) Topography of the combining region of a Thomsen-Friedenreich-antigen-specific lectin jacalin (*Artocarpus integrifolia* agglutinin). A thermodynamic and circular-dichroism spectroscopic study. *Biochem J.* **265**: 831-840.
- Mitra N, Sharon N, Surolia A. (2003) Role of N-linked glycan in the unfolding pathway of *Erythrina corallodendron* lectin. *Biochemistry* **42**: 12208–12216.
- Möller M, Denicola A. (2002) Protein tryptophan accessibility studied by fluorescence quenching. *Biochem. Mol. Biol. Educ.* **30**: 175-178.
- Monsigny M, Mayer R, Roche AC. (2000) Sugar-lectin interactions: sugar clusters, lectin multivalency and avidity. *Carbohydr Lett.* **4**: 35-52.
- Mormann M, Eble J, Schwöppe C, Mesters RM, Berdel WE, Peter-Katalinić J, Pohlentz G. (2008) Fragmentation of intra-peptide and inter-peptide disulfide bonds of proteolytic peptides by nanoESI collision-induced dissociation. *Anal Bioanal Chem.* **392**: 831-838.
- Mortel KH, Weatherman RV, Kiessling LL. (1996) Recognition specificity of neoglycopolymers prepared by ring-opening metathesis polymerization. *J Am Chem Soc.* **118**: 2297-2298.
- Mosolov VV, Valueva TA. (2005) Proteinase inhibitors and their function in plants: a review. *Appl Biochem Microbiol.* **41**: 227-246.
- Mundy J, Svendsen IB, Hejgaard J. (1983) Barley α -amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Res Commun.* **48**: 81-90.
- Narahari A, Swamy MJ. (2010) Rapid affinity-purification and physicochemical characterization of pumpkin (*Cucurbita maxima*) phloem exudate lectin. *Biosci Rep.* **30**: 341-349.

- Narahari A, Swamy MJ. (2009) Tryptophan exposure and accessibility in the chitooligosaccharide-specific phloem exudate lectin from pumpkin (*Cucurbita maxima*). A fluorescence study. *J. Photochem. Photobiol. B: Biol.* **97**: 40-47.
- Narahari A, Nareddy PK, Swamy MJ. (2011) A new chitooligosaccharide specific lectin from snake gourd (*Trichosanthes anguina*) phloem exudate. Purification, physico-chemical characterization and thermodynamics of saccharide binding. *Biochimie*, **93**: 1676-1684.
- Nareddy PK, Bobbili KB, Swamy MJ. (2017) Purification, physico-chemical characterization and thermodynamics of chitooligosaccharide binding to cucumber (*Cucumis sativus*) phloem lectin. *Int J Biol Macromol.* **95**: 910-919.
- Ogata F, Imamura H, Hirayama K, Makisumi S. (1986) Purification and characterization of four trypsin inhibitors from seeds of okra, *Abelmoschus esculentus* L. *Agric Biol Chem.* **50**: 2325-2333.
- Ohtsubo K, Marth JD. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell*, **126**: 855-867.
- Oliva MLV, Silva MC, Sallai RC, Brito MV, Sampaio MU. (2010) A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie.* **92**: 1667-1673.
- Padma P, Komath SS, Swamy MJ. (1998) Fluorescence quenching and time-resolved fluorescence studies on *Momordica charantia* (bitter gourd) seed lectin. *IUBMB Life*, **45**: 911-922.
- Page MJ, Di Cera E. (2008) Serine peptidases: classification, structure and function. *Cell Mol Life Sci.* **65**: 1220-1236.
- Park Y, Choi BH, Kwak JS, Kang CW, Lim HT, Cheong HS, Hahm KS. (2005) Kunitz-type serine protease inhibitor from potato (*Solanum tuberosum* L. cv. Jopung). *J Agric Food Chem* **53**: 6491-6496.

- Pascal JM, Day PJ, Monzingo AF, Ernst SR, Robertus JD, Iglesias R, Pérez Y, Ferreras JM, Citores L, Girbés T. (2001) 2.8-Å crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin I. *Proteins: Struct, Funct, Bioinf.* **43**: 319-326.
- Patel AK, Singh VK, Bergmann U, Jagannadham MV, Kursula P. (2011) Purification, crystallization and preliminary X-ray crystallographic analysis of MIL, a glycosylated jacalin-related lectin from mulberry (*Morus indica*) latex. *Acta Crysta Sec F: Struct Bio Cryst Comm.* **67**: 608-612.
- Perona JJ, Craik CS. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**: 337-360.
- Peumans WJ, Van Damme EJM. (1995) Lectins as plant defense proteins. *Plant Physiol.* **109**: 347-352.
- Pilobello KT, Mahal LK. (2007) Deciphering the glycode: the complexity and analytical challenge of glycomics. *Curr Opin Chem Biol.* **11**: 300-305.
- Pilobello KT, Krishnamoorthy L, Slawek D, Mahal LK. (2005) Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem*, **6**: 985-989.
- Pinho SS, Reis CA. (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nature Rev Cancer*, **15**: 540-555.
- Pohlentz G, Marx K, Mormann M. (2016) Characterization of Protein N-Glycosylation by Analysis of ZIC-HILIC-Enriched Intact Proteolytic Glycopeptides. *Proteom Syst Biol: Meth Protoc.* **1394**: 163-179.
- Poiroux G, Pitié M, Culerrier R, Lafont E, Ségui B, Van Damme EJM, Peumans WJ, Bernadou J, Levade T, Rougé P, Barre A. (2011)a. Targeting of T/Tn antigens with a plant lectin to kill human leukemia cells by photochemotherapy. *PloS one*, **6**: 23315.

- Poiroux G, Pitié M, Culerrier R, Ségui B, Van Damme EJM, Peumans WJ, Bernadou J, Levade T, Rougé P, Barre A, Benoist H. (2011)b. Morniga G: A Plant Lectin as an Endocytic Ligand for Photosensitizer Molecule Targeting Toward Tumor-Associated T/Tn Antigens. *Photochem Photobiol.* **87**: 370-377.
- Polgár L. (2005) The catalytic triad of serine peptidases. *Cell Mol Life Sci.* **62**: 2161-2172.
- Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nature Methods*, **2**: 817-824.
- Raval S, Gowda SB, Singh DD, Chandra NR. (2004) A database analysis of jacalin-like lectins: sequence–structure–function relationships. *Glycobiology*, **14**: 1247-1263.
- Richards RL, Moss J, Alving CR, Fishman PH, Brady RO. (1979) Cholera toxin: a bacterial lectin. *Proc Natl Acad Sci.* **76**: 1673-1676.
- Rohamare SB, Dixit V, Nareddy PK, Sivaramakrishna D, Swamy MJ, Gaikwad SM. (2013) Polyproline fold–In imparting kinetic stability to an alkaline serine endopeptidase. *Biochim. Biophys. Acta.* **1834**: 708-716.
- Ross JA, Rousslang KW, Brand L. (1981) Time-resolved fluorescence and anisotropy decay of the tryptophan in adrenocorticotropin-(1-24). *Biochemistry*, **20**: 4361-4369.
- Rougé P, Peumans WJ, Barre A, Van Damme EJM. (2003) A structural basis for the difference in specificity between the two jacalin-related lectins from mulberry (*Morus nigra*) bark. *Biochem Biophys Res Commun.* **304**: 91-97.
- Roy A, Shrivastava SL, Mandal SM. (2014) Functional properties of Okra *Abelmoschus esculentus* L.(Moench): traditional claims and scientific evidences. *Plant Sci Today*, **1**: 121-130.
- Royer CA. (1993) Understanding fluorescence decay in proteins. *Biophys. J.* **65**: 9-10.

- Rüdiger H, Gabius HJ. (2001) Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj J*. **18**: 589-613.
- Ryan CA. (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu Rev Phytopathol*. **28**: 425-449.
- Sahasrabuddhe AA, Gaikwad SM, Krishnasastri MV, Khan MI. (2004) Studies on recombinant single chain Jacalin lectin reveal reduced affinity for saccharides despite normal folding like native Jacalin. *Prot. Sci*. **13**: 3264-3273.
- 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 β -prism fold. *Nature Struct Mol Biol*. **3**: 596-603.
- Sastry MV, Surolia A. (1986) Intrinsic fluorescence studies on saccharide binding to *Artocarpus integrifolia* lectin. *Biosci. Rep*. **10**: 853-860.
- Sastry MV, Banarjee P, Patanjali SR, Swamy MJ, Swarnalatha GV, Surolia A. (1986) Analysis of saccharide binding to *Artocarpus integrifolia* lectin reveals specific recognition of T-antigen (beta-D-Gal (1----3) D-GalNAc). *J Biol Chem*. **261**: 11726-11733.
- Sharma A, Pohlentz G, Bobbili KB, Jeyaprakash AA, Chandran T, Mormann M, Swamy MJ, Vijayan M. (2013) The sequence and structure of snake gourd (*Trichosanthes anguina*) seed lectin, a three-chain nontoxic homologue of type II RIPs. *Acta Crystallogr Sect D: Biol Crystallogr*. **69**: 1493-1503.
- Sharon N, Lis H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, **14**: .53R-62R.
- Sharon N, Lis H. (2003) *Lectins*. Springer Science & Business Media.
- Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. **1**: 2856-2860.

- Singh T, Wu JH, Peumans WJ, Rougé P, Van Damme EJM, Wu AM. (2007) Recognition profile of *Morus nigra* agglutinin (Morniga G) expressed by monomeric ligands, simple clusters and mammalian polyvalent glycotopes. *Mol Immunol.* **44**: 451-462.
- Sinha S, Mitra N, Kumar G, Bajaj K, Surolia A. (2005) Unfolding studies on soybean agglutinin and concanavalin A tetramers: a comparative account. *Biophys. J.* **88**: 1300–1310.
- Song HK, Suh SW. (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J Mol Biol.* **275**: 347-363.
- Spencer ME, Hodge R. (1991) Cloning and sequencing of the cDNA encoding the major albumin of *Theobroma cacao*. *Planta*, **183**: 528-535.
- Sreerama N, Woody RW. (2000) Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal Biochem.* **287**: 252-260.
- Stirpe F. (2004) Ribosome-inactivating proteins. *Toxicon*, **44**: .371-383.
- Stirpe F, Battelli MG. (2006) Ribosome-inactivating proteins: progress and problems. *Cell Mol Life Sci.* **63**: 1850-1866.
- Strathmann M, Wingender J, Flemming H. (2002) Application of fluorescently labelled lectins for the visualization and biochemical characterization of polysaccharides in biofilms of *Pseudomonas aeruginosa*. *J Microbiol Methods.* **50**: 237-248.
- Sultan NM, Swamy MJ. (2005) Energetics of carbohydrate binding to *Momordica charantia* (bitter gourd) lectin: An isothermal titration calorimetric study. *Arch. Biochem. Biophys.* **437**: 115-125.

- Sultan NM, Swamy MJ. (2005) Fluorescence quenching and time-resolved fluorescence studies on *Trichosanthes dioica* seed lectin. *J. Photochem. Photobiol. B: Biol.* **80**: 93-100.
- Sultan NM, 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.
- Tai H, McHenry L, Fritz PJ, Furtek DB. (1991) Nucleic acid sequence of a 21 kDa cocoa seed protein with homology to the soybean trypsin inhibitor (Kunitz) family of protease inhibitors. *Plant Mol Biol.* **16**: 913-915.
- Tallmadge DH, HuEbnert JS, Borkman RF. (1989) Acrylamide quenching of tryptophan photochemistry and photophysics. *Photochem. Photobiol.* **49**: 381-386.
- Thyrock A, Ossendorf E, Stehling M, Kail M, Kurtz T, Pohlentz G, Waschbüsch D, Eggert S, Formstecher E, Müthing J, Dreisewerd K. (2013) A new Mint1 isoform, but not the conventional Mint1, interacts with the small GTPase Rab6. *PloS one*, **8**: e64149.
- 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 EJM, Allen AK, Peumans WJ. (1987) Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS letters*, **215**: 140-144.
- Van Damme EJM, Barre A, Rouge P, Peumans WJ. (2004) Cytoplasmic/nuclear plant lectins: a new story. *Trends Plant Sci.* **9**: 484-489.

References

- Van Damme EJM, Hause B, Hu J, Barre A, Rougé P, Proost P, Peumans WJ. (2002) Two distinct jacalin-related lectins with a different specificity and subcellular location are major vegetative storage proteins in the bark of the black mulberry tree. *Plant Physiol.* **130**: 757-769.
- De Schutter K, Van Damme EJM. (2015) Protein-carbohydrate interactions as part of plant defense and animal immunity. *Molecules*, **20**: 9029-9053.
- Vandenborre G, Smaghe G, Van Damme EJM. (2011) Plant lectins as defense proteins against phytophagous insects. *Phytochemistry*, **72**: 1538-1550.
- Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, editors. (2009) Essentials of Glycobiology. 2nd edition. Cold Spring Harbor Laboratory Press; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1908/>
- Wasano N, Konno K, Nakamura M, Hirayama C, Hattori M, Tateishi K. (2009) A unique latex protein, MLX56, defends mulberry trees from insects. *Phytochemistry*, **70**: 880-888.
- Webb B, Sali A. (2014) Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 5.6.1-5.6.32.
- Werner MH, Wemmer DE. (1992) Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution. *Biochemistry*, **31**: 999-1010.
- Wu AM, Lisowska E, Duk M, Yang Z. (2009) Lectins as tools in glycoconjugate research. *Glycoconjugate J.* **26**: 899-913.
- Yanes O, Villanueva J, Querol E, Aviles FX. (2007) Detection of non-covalent protein interactions by 'intensity fading' MALDI-TOF mass spectrometry: applications to proteases and protease inhibitors. *Nat Protoc.* **2**: 119-130.

- Young NM, Johnston RA, Szabo AG, Watson DC. (1989) Homology of the D-galactose-specific lectins from *Artocarpus integrifolia* and *Maclura pomifera* and the role of an unusual small polypeptide subunit. *Arch Biochem Biophys.* **270**: 596-603.
- Zhu-Salzman K, Zeng R. (2015) Insect response to plant defensive protease inhibitors. *Annu Rev Entomol.* **60**: 233-252.

Curriculum Vitae

***Debparna Datta****School of Chemistry**University of Hyderabad**Hyderabad-500046**Email: debparnadatta@gmail.com*

Debparna Datta was born in Suri, West Bengal on 24th September, 1987. She finished her schooling from Patha-Bhavana and Uttar Siksha Sadan, Visva-Bharati in 2006. That same year she joined Department of Chemistry, Visva-Bharati and finished her M.Sc degree in Chemistry with a specialization in Organic Chemistry in 2011. She qualified CSIR-JRF in August, 2011 and subsequently joined the lab of Prof. Musti J. Swamy, School of Chemistry, University of Hyderabad in March, 2012. Her major research focus was on the purification and biophysical characterization of plant lectins, however, she has also worked on plant protease inhibitors. During her PhD she was selected for IRTG-MCGS student exchange program, through an MoU between University of Hyderabad and University of Muenster, Germany. She has spent nine months at the University of Muenster where she got acquainted with mass spectrometric techniques.

List of Publications

1. **Debparna Datta**, Gottfried Pohlentz, Mona Schulte, Mathias Kaiser, Francisco M Goycoolea, Johanness Müthing, Michael Mormann, Musti J Swamy. (2016) Physico-chemical characteristics and primary structure of an affinity-purified α -D-galactose-specific, jacalin-related lectin from the latex of mulberry (*Morus indica*). *Arch Biochem Biophys.* **609**: 59-68.
2. **Debparna Datta**, Musti J Swamy. (2017) Fluorescence and circular dichroism studies on the accessibility of tryptophan residues and unfolding of a jacalin-related α -d-galactose-specific lectin from mulberry (*Morus indica*). *J Photochem Photobiol B: Biology*, **170**: 108-17.
3. **Debparna Datta**, Gottfried Pohlentz, Saradmoni Mondal, Michael Mormann, Musti J Swamy. (2017) Biophysical characterization and partial amino acid sequence of a Kunitz-type protease inhibitor from okra (*Abelmoschus esculentus*) seeds. (Manuscript under preparation).
4. S N Raju Kutcherlapati, Rambabu Koyilapu, Uma Maheswara Rao Boddu, **Debparna Datta**, Ramu Sridhar Perali*, Musti J Swamy*, Tushar Jana*. (2017) Glycopolymer Grafted Nanoparticles: Synthesis using RAFT Polymerization and Binding Study with Lectin. (Submitted to *Macromolecules*).
5. Musti J Swamy, Akkaladevi Narahari, Kishore Babu Bobbili, **Debparna Datta**. (2017) Review on Cucurbitaceae phloem exudate lectins. (To be communicated).

Symposia and Conferences Attended

1. Affinity purification and characterization of *Morus indica* latex lectin, a jacalin-related galactose-specific lectin.
Poster presentation 27th International Carbohydrate Symposium (ICS), held at Indian Institute of Sciences, Bangalore; January 2014.
2. Affinity purification and characterization of *Morus indica* latex lectin, a jacalin-related galactose-specific lectin.
Poster presentation First Indo-Taiwan Symposium on Recent Trends in Chemical Sciences (RTCS), held at School of Chemistry, University of Hyderabad, November 2014.
3. Affinity purification and characterization of *Morus indica* latex lectin, a jacalin-related galactose-specific lectin.
Poster presentation 6th Asian Community of Glycoscience and Glycotechnology (ACGG), held at School of Life Sciences, University of Hyderabad, December 2014.
4. Affinity purification and characterization of *Morus indica* latex lectin, a jacalin-related galactose-specific lectin.
Poster presentation DRILS Science Café-2015, held at Dr. Reddy's Institute of Life Sciences, University of Hyderabad, August 2015.
5. Affinity purification, physico-chemical characterization and primary structure of an α -D-galactose specific jacalin-related lectin from the latex of mulberry
Poster presentation 13th Annual In-House Symposium Chemfest-2016, held at School of Chemistry, University of Hyderabad, March 2016.
6. Biochemical and biophysical studies on an α -D-galactose-specific lectin from mulberry (*Morus indica*) latex.
Oral presentation 14th Annual In-House Symposium Chemfest-2017, held at School of Chemistry, University of Hyderabad, March 2017.