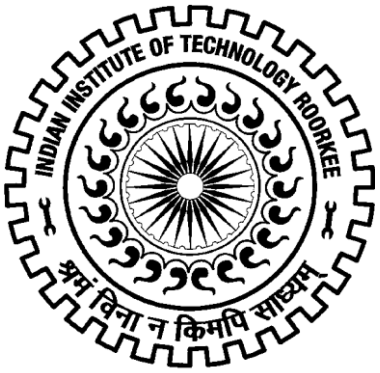
STRUCTURAL STUDIES ON 11S GLOBULIN AND 2S ALBUMIN FROM *WRIGHTIA TINCTORIA*

Ph.D. THESIS

by

PRAMOD KUMAR



DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013

STRUCTURAL STUDIES ON 11S GLOBULIN AND 2S ALBUMIN FROM *WRIGHTIA TINCTORIA*

A THESIS

Submitted in partial fulfilment of the
Requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

PRAMOD KUMAR

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013

©INDIAN INSTITUTE OF TECHNOLOGY ROORKEE, ROORKEE-2013
ALL RIGHTS RESERVED

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"STRUCTURAL STUDIES ON 11S GLOBULIN AND 2S ALBUMIN FROM *WRIGHTIA TINCTORIA*"** in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2009 to December, 2013 under the supervision of Dr. Pravindra Kumar, Associate Professor, Department of Biotechnology and Dr. B. R. Gurjar, Associate Professor, Department of Civil Engineering, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(PRAMOD KUMAR)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

(B. R. Gurjar) (Pravindra Kumar)
Supervisor Supervisor

Dated: December, 2013

The Ph.D. Viva-Voce Examination of **Mr. PRAMOD KUMAR**, Research Scholar, has been held on..... .

Signature of Supervisors Signature of Chairman, SRC Signature of External Examiner

Head of the Dept./Chairman, ODC

ABSTRACT

From the human point of view, the most important part of plant is seed that is harvested and consumed in countless forms. The seed storage protein constitutes an indispensible dietary protein that account 70% of the total intake. In addition, these proteins are sole source for the non ruminating farm animals. The storage proteins constitute a significant part of plant seeds that governs several inherent traits such as viability, nutrient value, protection from microbe and insects as well as end use properties. Although, diverse kinds of proteins are present in seed that confer metabolic and structural support, nevertheless almost every seed contains some proportion of the storage proteins. They are mostly present in high amount and facilitate store of amino acids, consumed during seed germination and seedling growth. The specific quality and texture of seed is determined by property of storage protein. In particular example of cereals, the lysine, tryptophan and threonine are limited, while in legumes methionine is the deficient amino acids. This deficiency is due to their limited proportions in seed storage proteins.

The classification of seed storage proteins (SSPs) is primarily based on the medium of solubility and extraction, the proteins are soluble in, i) dilute saline classified as globulins, ii) water classified as albumins, iii) alcohol classified as prolamin and iv) dilute acid/alkali classified as glutelins. The globulins are most profoundly distributed form of seed storage protein marks its presence across dicots, monocots as well as in lower forms of plant such as ferns. Globulins are large globular proteins further graded in two groups based on differences of sedimentation coefficients, as 11S legumin types and 7S vicilin types. The post translational processing confers major difference among these classes, which is visible in form of difference in structural organization and post translational modifications. Although both of proteins are deficient for cystine and methionine, 11S globulins are relatively higher in sulfur containing amino acid than 7S class. The 11S globulin is almost always found together with 2S albumin. The 2S albumins are smallest member of the SSP family. They are classified on the basis of sedimentation coefficient. The 2S albumins are widely known for their conserved and higher sulphur containing amino acids property. The substantial proportion of sulphur containing amino acids make them most suitable SSPs that can be exploited for improvement of seed nutrition and also for other related applications where sulphur content is required. In addition to that the high level of cystine in 2S albumins enables them to acts as a weapon in plant defence

system. A less studied regulatory role of calmodulin antagonist function has been also found from the 2S class of seed storage protein.

This thesis is focused on structural and functional studies of 11S globulin and 2S albumin from the plant *Wrightia tinctoria*. The comprehensive structure analysis and the related functions have been described in four sections of chapters. An independent work has been also incorporated under the section appendices. That contains crystallization study for structural analysis of a penicillin binding protein FmtA, responsible to confers antibiotic methicillin resistance to the gram positive bacteria *S. aureus*.

Chapter **1.** reviews the state of art in literature; describes structural fictional details of 11S globulin and 2S albumins.

Chapter **2**. describes the identification, purification and biochemical characterizations of 11S globulin from the plant *W. tinctoria*. An 11S class of seed storage protein has been purified and the relation of the protein to 11S family has been established by N-terminal as well as partial sequencing and further the gene responsible for their synthesis was cloned. The basic molecular profile has been revealed that describes the *W. tinctoria* globulin (WTG) as a hexamer made up of six monomers and further each monomers are constituted by two different alpha and beta subunits attached together by disulphide bond. It has been observed that WTG naturally exists in a very polymorphic and heterogeneous form of their constituent subunit components, as observed from internal sequencing as well SDS-PAGE analysis. The cloning of entire gene and the corresponding amino acid profile reveals that WTG is made up of 482 amino acid residues. Furthermore, the biochemical characterizations reveal that WTG encompasses a unique ability of hemagglutination. It has been observed that WTG also contains exceptional stability, concluded on the basis of circular dichroism and fluorescence studies. Only the high concentration and harsh ionic condition can sufficiently change the native state, otherwise WTG can tolerate wide range of chaotropic reagents and pH conditions.

Chapter **3**. focuses on the structural studies of WTG by crystallization and structure determination. The X-ray diffraction studies have been employed for elucidation of structural

and functional details. With a high diffraction up to 1.8 Å resolutions, the data has been solved and the final model was build. The refined model is showing all the crystallographic parameters in defined limits. The WTG encompasses conserved structural fold of jelly roll β-barrel motif that constitute a bicupin fold as a signature of cupin super family. Furthermore, the extended alpha helical domains along with cupin fold draw the structural analogy with members of 11S SSPs. The WTG encompasses six monomers, which constitute a hexameric scaffold. Further, WTG structure showed that there are three fold symmetry exists between two trimers and this view depicts a trimer of dimer model of WTG, along perpendicular to three fold a twofold symmetry has been also observed for WTG but this symmetry showed dimer of trimer arrangement of WTG. The N-terminal regions of both trimers remain funnel together at the centre of three fold axis at two opposite ends; this arrangement constitute the two poles of WTG hexamer. The initial 20 residues of N-terminal region are showing exception hydrophilic nature as well as presence of serine residue having fairly high score for possible phosphorylation. Hence it appears form the WTG model that there is a delicate series of interaction involved in hexamer formation. In addition to that dimer of trimer or trimer of dimer, both view forms three pockets related along three fold axis of symmetry. The high resolution data clearly revealed an auxin molecule uniformly present in three pockets of hexamer. Therefore, it appears that a mild change of pH or alteration of surrounding environment WTG may lead to dissociation of entire hexamer assembly.

Chapter. **4**. describes identification, purification, biochemical and biophysical characterizations of *W. tinctoria* 2S albumin (WTA). The WTA was purified to homogeneity in three steps of purification. The purified WTA was subjected to partial internal sequencing and obtained sequence were aligned to confirms the relation with albumin family of protein from the *W. tinctoria* against available EST database of NCBI. The basic secondary structure profile of WTA has been revealed by circular dichroism spectral analysis, which concludes it as a major alpha helical protein.

The highly conserved cystine residues and substantial ionic surface of WTA have been speculated for metal binding activity; hence the biochemical interaction has been studied against different metals by employing isothermal titration calorimetory. The ITC results confirm a significant binding of WTA with Copper, Cadmium and Lead. In addition to this the *in silico* modeling and structure analysis reveal that WTA contains reasonable hydrophilic surface profile that also supports the metal binding.

Appendix **A**. describes the purification and crystallization studies of FmtA from *S. aureus*. Review and related literature has been surveyed in the chapter. For the structural studies of FmtA, the protein was purified in two steps involving ion exchange and gel filtration chromatography. The purified protein was subjected to crystallization trial over 96 well sitting drop trays, and crystals were obtained. Initial diffraction data showed that the model is following the molecular weight relation to the cell volume correlation, as well the densities are coming closely similar to the related homologues and matching the sequence profile with two molecules in asymmetric unit. The more organized crystals are under optimization for the FmtA structure solution and to building of final model.

Acknowledgements

Constant inspiration and encouragement provided by a number of individuals served as a driving force in earning this day in my life. I would like to thank almighty GOD first. I raise my hands, bow down before you! "It is done"! I thank you for the struggles that made me learn how to untie them.

I would to like to acknowledge my mother Mrs. Kamla Devi and father Mr. H. C. Paswan, for raising me, for bearing with me during the challenges, and for joining me in the triumphs, for letting me be a bit too ridiculous occasionally, when my passions led me there, for never damping my sense of curiosity, for never imposing, for never having anything but confidence in me. I thank my mom for the endless home cooked food and for showing me the importance of valuing the people in my life, my dad for infecting me with a critical eye and the need to know how things and humans work. I dedicate this thesis to you both.

I would like to acknowledge my esteemed supervisor Dr. Pravindra Kumar for giving me the opportunity to work in their group and providing opportunity for developing myself in many directions during my doctoral studies that culminate from a student to Scientist. He has taught me, both consciously and artistically, how good macromolecular crystallography is done. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. I am also thankful for the excellent example he has provided as a successful crystallographer. I would also thank my co supervisor Dr. B. R. Gurjar for his stimulating suggestions and encouragement. I convey my sincere regards to Dr. Pravindra Kumar for minutely analyzing the work and investing his energy while reviewing the draft, without which it would have been impossible for me to complete this thesis and to Dr. B. R. Gurjar for his thorough checking and precious suggestions on draft.

I would like to express my deep gratitude and respect to Dr. Shailly Tomar for helping me in Biochemistry and Molecular biology and being available to guide me in my all projects.

I would like to express my sincere thanks towards Dr. A. K. Sharma and Dr. Partha Roy, Department of Biotechnology for their unflinching help and support. Further, I feel overwhelmed in thanking to Prof. R Prasad (Present), Prof. Ritu Barthwal (Former) Head, Department of Biotechnology and Dr. Ramesh Chandra Head, Institute Instrumentation Centre, for providing necessary facilities. I gratefully acknowledge the members of my Student Research Committee (SRC) Prof. G. S. Randhawa, Prof R. P. Singh, Dr. Partha Roy (Internal member) and Dr. Shishir Sinha (External member) for their scholarly suggestions.

I would also like to acknowledge Dr D. G. Kotra from York University, Canada, Dr. Girish Mishra from Department of Botany, Delhi University and Dr. Balaji Prakash, Department of Biological Science and Bio Engineering, IIT Kanpur, for their kind scientific help as well as inspiration towards my research work. I am courteous to the honourable faculty members, Professor H. S. Dhaliwal, Dr. Ranjana Pathania, Dr. Vikas Pruthi, Dr. Naveen Navani, Dr. Bijan Choudhury, Dr. Sanjoy Ghosh, Dr. Debabrata Sircar, Dr. M. S. Nair for their support and encouragement. I would like to mention all the technical and office staff of the Biotechnology Department in particular Mr. Ved Pal Singh Saini, Mr. Subhash Jain, Mrs. Shashi Prabha, Mr. Lokesh and Mr. Anil.

I am obliged to my lab mates Dr. Dipak Patil, Mr. Shivendra Pratap, Mr. Aditya Dev and Ms. Sonali Dhindwal, for friendly and helping supports. Further with thanks, I would like to appreciate, Mr. Madhusudhan, Ms. Neha, Ms. Anchal Mr. Vijay, Ms. Puja, Mr. Shahid and Mr Ankit for their sincere efforts, amiable attitude and cooperation while working in lab.

Also to be acknowledged with love and appreciation in the memorable concern, affection and care from my dear friends Dr. Santosh, Dr. Lalit, Ms. Manju, Ms. Megha, Dr.Preeti, Ms. Sweta, Mr. Naresh, Mr. Bibekanand, Mr. Paramesh, Mr. Pradeep, Dr. Satya, Dr. Rajnikant, Dr. Supriya, Mr. Prabhat, Dr. Girijesh, Dr. Vikash, Mr. Jitendra, Mr. Rajat, Mr. Rajesh, Ms. Monu, Ms. Raman and many others who made my stay in IITR memorable.

I am also grateful beyond measure to Council of Scientific and Industrial Research, Govt. of India for providing the financial support to conduct the research work.

Finally, my loving family deserve special mention for their unending love, inseparable support and blessing. With endless appreciation I would like to thank my wife Sujata for her personal support and great patience at all times. I also would like to thank to my lovely and adorable daughter Kanika. My deep appreciation goes to brother Mr. Pradeep, who is always ready with a plan, sisters Mrs. Rajni and Mrs. Pratibha, who never let me feel alone; my lovely nice Ms. Swati, my nephews Sumit, Nitin and Kukku, for their deep affections.

I never felt alone during this period due to all the special cares given by my mother-in-law Mrs. Suggi Devi, brothr-in-laws Mr. Brind Paswan, Amlesh Paswan and A. K. Saxena and cousin brother Satyendra for always giving great care. My friends, Mr. Rajnish, Mr. Ranjit, Dr. Shailesh, Mr. Jitendra, Mr. Dilip, Mr. Ashutosh, Ms. Asha and Dr. Akhilesh are special due to always provided deep affection as well as unconditional support at each turn of the road.

I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

I thank you all……

(Pramod Kumar)

LIST OF PUBLICATIONS

- **Kumar P**, Patil DN, Chaudhary A, Tomar S, Yernool D, Singh N, Dasauni P, Kundu S, Kumar P., Purification and Biophysical Characterization of an 11S Globulin from Wrightia tinctoria exhibiting Hemagglutinating Activity., *Protein Pept. Lett.* 2013, 20(5), 499-509.

- Aditya dev, Nandita Bodra, **Kumar P.,** Pravindra Kumar., Homology modeling and functional characterization of 3-D structure of DAHP Synthase from model grass Brachypodium distachyon, *Journal of Proteins and Proteomics*, 2013, 4(1), 21-25.

- **Kumar P,** Sonali Dhindwal, Girish Mishra, Ashish choudhary, B. R.Gurjar, Pravindra Kumar, Crystal structure of Wrightia tinctoria 11S globulin: homohexamer (under preparation).

- **Kumar P**, Neha Singh, B. R. Gurjar, Pravindra Kumar, The metal binding ability of 2S albumin and possible application in bioremediations (under preparation).

WORKSHOPS/CONFERENCE ATTENDED AND POSTER PRESENTATION

- **Kumar P**, Patil Dipak N, Chaudhary Anshul, Kumar Pravindra, Tomar Shailly, Purification and biophysical characterization of HAG, a 11s hexameric globulin from Kutaj (*Holarrhena antidysenterica*) seeds possessing hemagglutination and trypsin inhibitory activity (Poster presentation at Asian Biophysical Association and Indian Biophysical socity conference- 2011, New Delhi).

- **Kumar P**, Nimish Gopal, Abin Ghosh, Pravindra Kumar, Targeting the agent responsible for invasion: an approach towards cure of malaria (Poster Presentation at 2nd International conference on open source for computer-aided Translational medicine 2012, IMTECH, Chandigarh).

- **Kumar P,** Pravindra Kumar, 11S Globulins: A protein having ability to meet the need of nutrient and sequester growth hormone auxin (Oral presentation given at 42nd National Seminar on Crystallography and International Workshop on Application of X-ray Diffraction for Drug Discovery, 2013).

CONTENTS

CHAPTER - 2 **23-45**

PURIFICATION FROM NATURAL SOURCE, CLONING, BIOCHEMICAL AND BIOPHYSICAL CHARACTERISATIONS OF 11S GLOBULIN FROM *Wrightia tinctoria*

LIST OF TABLES

LIST OF FIGURES

ABBRIVIATIONS

Å	Angstrom
BME	β-mercaptoethanol
bp	Base Pairs
BSA	Bovine Serum Albumin
CAPS	(N-Cyclohexyl 1-3-aminopropanesulfonic acid)
CCP4	Collaborative Computational Project Number 4
CD	Circular Dichroism
c-DNA	Complementary DNA
CHCA	Cyano 4-Hydroxy Cinnamic Acid
cm	Centimetre
°C	Degrees Celsius
DEAE	Di-Ethyl Amino Ethane
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
ESTs	Expressed Sequence Tags
g	Gram
GdnHCl	Guanidine Hydrochloride
h	Hours
H.U.	Hemagglutination Unit
IU	International unit
Ig	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISO	International organization of standardization
ITC	Isothermal titration calorimetory
kDa	Kilo dalton
kg	Kilograms
K	Kelvin
kV	Kilovolt
L	Litres
m	Meters

mA	Milli Ampère
MALDI	Matrix-Asisted Laser Desorption/Ionization
µg	Microgram
µm	Micrometer
mL	Millilitres
mm	Millimeter
mol	Molarity
MRSA	Methicillin Resistant *S. aureus*
nm	Nanometers
NMR	Nuclear Magnetique Resonance
OD	Optical density
%	Percent
PVDF	Polyvinylidene fluoride
PEG	Polyethylene glycol
pI	Isoelectric Point
PMF	Peptide Mass Fingerprint
RMSD	Root Mean Square Deviation
rpm	Revolution per minute
SDS-PAGE	Sodium Dodecyl Sulphate-polyacrylamide gel electrophoresis
SSP	Seed Storage Protein
U.V.	Ultra Violet
v/v	Volume per Volume
w/v	Weight per Volume
WTA	*Wrightia tinctoria* Albumin

Introduction

The collections of proteins within a cell determine its function as well as constitute more than half of the cells dry weight. Proteins are the most important class of bio-molecules which confer indispensible abilities to the cells such as proper shape, structure catalysis and molecular recognition. Their orchestrated functions are indispensible for successful completion of the life span of a cell. The enormous functions of different proteins are governed by particular arrangements of twenty major types of amino acids that act as building blocks for proteins. The amino acids exhibit huge diversity due to the unique profile of their side chains. In proteins, many bonds between the contiguous amino acids constitute a long back bone arrangement. The long back bone with diverse organization of amino acids subjected to stereospecific restrains that restricts proteins to attain some limited conformations. These confirmations characterize a particular class of protein. A major function of the protein is binding ability with extreme specificity. The particular nature of proteins are governed by specific structure which is further augmented by nature of the amino acids, surrounding microenvironment and physiological conditions (Shoichet et al., 1995). To maintain the state of homeostasis proteins performs plethora of functions such as supporting the structural organisation, imparting specificity for selective enzyme substrate/inhibitor or hormonal interaction, recognition of only specific protein-protein or protein-ligand, antigen antibody interactions etc. These functions enable the protein to play pivotal role for existence of living beings.

The role of the proteins for supporting the existence in higher organisms may be considered from the beginning stage of gametogenesis, followed by fertilization and embryogenesis that leads to the formation of mature seed in case of plants. Proteins further support the progresses of seed to culminate in fully grown plant that eventually completes a life cycle. The seeds are immensely important, as a unit of propagation as well as means of dispersal of plants. Moreover, seeds store different range of compounds, which is its most important property of seeds and exploited in numerous applications ranging from food and fodder to industrial feed stocks. The most widespread compounds present in the seeds as storage material are proteins, starch and lipids. Almost all seed contains some forms of storage proteins that provide a source of nitrogen and sulphur along with the carbon skeletons. Apart

from proteins either starch or lipid also supports the storage with a wide level of variance among different species (Rochat and Boutin, 1992).

Seed proteins are most commonly classified on the basis of their solubility and the method of extraction by application of particular solvent. The earlier work especially done by T.B. Osborne (Osborne, 1924) describes seed proteins classes on the basis of extraction by using different solvents. They are broadly divided in four major groups, i) Albumins are having solubility in water and showing coagulation up on heating, present in some seeds such as cereals, legumes etc., ii) Globulins are insoluble in water but having solubility in saline solution and its presence has been observed from many seeds, iii) Prolamins like relatively strong concentration of alcohol (60-70%) for getting solubilise but not soluble in saline or water solutions and iv) Glutelins can be extracted in alkali or acidic medium rather than neutral aqueous, saline or alcohol solution. The glutelins are observed from limited sources such as wheat, rice and maize. These clear gradations does not followed by every protein on the basis of solubility, moreover strict division is also not observed due to intermittent solvent preference. Therefore this classification is no longer in regular use. A more comprehensive modern classification of these proteins is based on two major criteria, first is made on the basis of molecular/biochemical correlation and second is based on the function performed by the particular class. Therefore, it divides the seed proteins in to three major classes. (a) Storage proteins; responsible for the major function of providing store of nitrogen, carbon and sulphur required especially during the onset of germination. (b) Structural and metabolic proteins; crucial for defining the structure and growth of seeds, while the metabolic proteins regulate biosynthetic or catabolic pathway responsible for storage product biosynthesis or mobilization and (c) Defence/protective proteins; have applications for the resistance against pathogens such as microbes, invertebrate pests and sometime for desiccations tolerance also.

The seed storage proteins (SSPs) are reported from angiosperms, gymnosperms, trees, cycads, cereals and fern spores; their amount of total protein contribution to seeds vary from few percent to the major part. The globulins are large proteins having different sedimentation coefficient broadly classified in 7-8S and 11-12S as observed from cereal and leguminous family (Danielsson, 1949). Evolutionary, both the classes are related by conserved cupin folds. Furthermore, the relation of globulins with a big cupin super-family corroborates their links to other distantly related members of the family such as germins (oxalate oxidases), bacterial oxalate decarboxylases, gentistate dioxygenase and epimerases (Shutov and Kakhovskaya,

2011). On the other hand, the small proteins 2S albumins are also widespread form of SSPs in dicotyledonous plants, usually present together with 11S SSPs (Youle and Huang, 1981).

The plant *Wrightia tinctoria* belongs to the family *Apocynacea* is a small deciduous tree. The distribution of this genus has been reported throughout the world and the particular species is profoundly distributed throughout Indian continent (Varier, 1997; Deswal et al., 2013). It is exploited for the extraction of several biologically active compounds as well as for medicinal applications. The leaves acts as a source of blue dye called Indigo, the dry weight of leaves posses nitrogen 2 %, ash 9.3 % and calcium 3.8 %, moreover several amino acids has been also extracted from the leaves (Chadha, 1985). The ursolic acid and β-amyrin triterpenoids were also extracted from the leaves and pods of *Wrightia tinctoria* (Rao et al., 1966), some other triterpenoids such as olenolic acid, wrightial, cycloartenone has been reported from the pods (Kataoka et al., 2001). Moreover, many alkaloids such as indigotin, indirubin, tryptanthrin, isatin, anthranillate are reported as major constituents of this plant (Muruganandam et al., 2000). Several phytosterols have also been isolated from the seeds of *Wrightia tinctoria* like desmosterol, clerosterol, 24-methylene-25-methylcholesterol (Akihisa et al., 1988). Apart from these, many immensely important medicinal applications are known from different parts of this plant that is used as traditional medicine. The bark and seeds are used in flatulence and bilious affections, decocted leaves and bark are taken as stomachic. The aqueous and alcoholic extracts of root and leaf contains hypotensive properties (Chadha, 1976). The seeds are having special application as aphrodisiac and anthelmintic. In summary, the different parts of this plant are used for extraction of several classes of compounds like alkaloids, triterpenoids, steroids, flavanol, and phenyl propanoids as well as for many medicinal applications.

In present study, we have done extensive analysis of two proteins from the seeds of plant *Wrightia tinctoria*. **First** protein is the 11S globulin (WTG), a hexamer with molecular weight ~ 330 kDa, its biochemical and biophysical characterization and unusual stability along with hemagglutination activity has been reported by our group (Kumar et al., 2013). Further, the major findings encompass hormone binding ability of WTG, elucidated from the crystal structure that is in complex with auxin. **Second** is a small protein of 19 kDa albumin termed as *Wrightia tinctoria* albumin (WTA), concluded on the basis of observed internal sequencing results and their comparisons with the known homologues proteins of same class. The WTA has been observed with unique ability of metal binding, characterised by the Isothermal

titration calorimetory (ITC). Moreover, *in silico* studies and structure investigation reveals the major alpha helical nature of WTA.

The globulins marks widespread presence in seeds and reported in two classes as 11-12S legumin and 7-8S vicilin, divided on the basis of their sedimentation coefficients while the name legumin and vicilin is derived from the initially reported taxon *Leguminosae* and *Viciae* (Danielsson, 1949). But with the specific example of plant the globulin names are changed for both the classes (**Table. 1.1**). These two groups of globulins share many common features, one of the strong correlation found in the two classes is the basic structural features at tertiary and quaternary levels. The monomer of both globulins encompasses two structurally equivalent domains that shares a common ancestry, evolved from single domain germin-like ancestors and characterized by typical cupin fold (Shutov and Baumlein, 1999). Both the proteins are involved in molecular storage, thus act as reserve source for mainly nitrogen and carbon along with sulphur and other compounds.

The 2S albumins are small proteins, classified on the basis of their sedimentation coefficient (Youle and Huang, 1981). This class is having wide distribution among both monocots and dicots. The 2S albumins are deposited in protein bodies at the stage of seeds development and exploited by the plant as source of nutrient during germination and seedling growth. Apart from the nutritive role, these proteins are also associated with plant defence, especially against fungi (Agizzio et al., 2006).

Table 1.1. The table encompasses some of the most common examples from the 7S and 11S class of seed storage proteins.

(The 7S SSPs) **Species – Name**	**(The 11S SSPs)** **Species – Name**
Arachis hypogea -- Conarachin	*Arabidopsis thaliana* -- Cruciferin
Canavalia spp. -- Canavalin	*Arachis hypogea* -- Arachin
Glycine max -- *β*-conglycinin	*Cannabis sativa* -- Edestin
Lupinus spp. -- Conglutin β	*Cucurbita* spp. -- Cucurbitin
Medicago sativa -- Alfin	*Glycine max* -- Glycinin
Phaseolus vulgaris -- Phaseolin	*Helianthus annulus* -- Helianthinin
Pisum, Vicia spp. -- Vicilin, convicilin	*Lupinus* spp. -- Conglutin α
Theobroma cacao -- Vicilin	*Medicago sativa* -- Medicagin
	Pisum, Vicia spp. -- Legumin

The amino acid composition of this class of proteins are of special importance due to high content of sulphur-containing amino acids (Youle and Huang, 1981). The composition of cystine among 2S albumins has been reported with significant amount of approximately 13-15 %, while some species like Brazil nut contains substantial proportion up to 20% of total amino acids. This ability of 11S globulin is of immense importance for the improvement of the nutrient value of legume and other food crops. However, the engineered soybean containing Brazil nut albumin remains unsafe due to the allergenic nature of the albumin (Moreno and Clemente, 2008). The immense possibilities of seed nutrient improvisation by virtue of 2S albumin engineering is limited due to allergen nature, hence such results demand a deeper understanding for this class at molecular and structural level.

CHAPTER - 1

LITERATURE REVIEW

1.1. Introduction

The proteins cover almost all biological processes some of the major functions are catalysis, molecular transportation, storage, providing mechanical support and movement, immune protection, transmission of nerve impulse, control of growth and differentiations etc. Proteins are polymers constituted by monomers of amino acids. There are twenty different types of amino acids characterized by elemental composition of side chains while sharing common arrangement of main chain backbone. The unique sequence of amino acids and typical folding determines the structure of particular protein, which confer the function as well as governs the specific properties. The function of proteins are controlled and modulated through various strategies, such as post translational modifications, application of cofactors, protein existence in premature and get activated only when cleaved (Milne and Cook, 1979; Farley and Link, 2009). The function of proteins are always associated with specificity and the selective binding to the target, such selectivity enables each protein to perform corresponding function in the milieu of thousands of different proteins and other bio molecules. The ability of interactions is conferred by weak forces such as electrostatic, Vanderwaals forces, hydrogen bonding etc. These interactions are very crucial for specific structural organization and function of proteins. The typical location and number of residues along with weak interactions are among the major factors that governs specific activities, as concluded from several mutational studies where alteration of key interactions either lead to gain, reduction or loss of activity (Lazar et al., 1988).

The structural studies and related functional behaviour of proteins are crucial to earmark their properties. *In vivo* and *in vitro* studies are required to get the insight of functional behaviour at molecular as well cellular level. For *in vitro* studies broadly biochemical and biophysical characterizations are done by application of different techniques at different stages. The protein characterization needs the purest form that is isolated from the milieu of cellular environment. Isolation begins with lyses of cell followed by separation of debris from the supernatant. The supernatant contains mixture of several proteins from which a targeted protein needs to be separated, the selective separation is achieved by different chromatographic techniques and analysis made by different spectroscopic methods. The spectroscopic techniques such as UV-visible, fluorescence, circular dichroism, FT-IR, DLS, X-ray

crystallography, NMR are being used to get the insight of detailed structural arrangements and characterization of functional relationship. These details and extensive analysis finally becomes the basis for elucidation of proteins and enzyme nature, their mode of regulation and catalysis, to find the agonists or antagonists that finally forms the basis of principles and theories (Amsterdam and Jamieson, 1972; Kurplus and McCammon, 1983).

The present chapter describes state of art for literature under following sections, **1.2**) the globulins in plant, **1.3**) Evolution of seed storage proteins, **1.4)** the plant *Wrightia tinctoria.*

1.2. The globulins in plants

The globulins are large protein known for its globular nature and as an integral part of the plant seeds. With germination of a seeds, most of the proteins are degraded and only few proteins endure the abrupt degradation. These selected classes of proteins remains involved in nourishing the seedling and qualifies the criteria of true storage proteins. The genuine storage proteins differ by exhibiting remarkable properties of enzymatic inactivity, abundant presence, temporal regulation of synthesis and accumulation without any breakdown during seed maturation. Their degradation only takes place after a long period of dormancy that progress towards germination (Shutov et al., 2003). The globulins qualify as a true storage protein due to precisely fitting in these criteria by exhibiting structural features that govern temporal regulation and protection from proteolysis. They constitutes major storage proteins in the seeds of cereals and legumes as well as in nuts (Mandal and Mandal, 2000). As one of the main storage proteins in various food sources, the globulins are responsible and important source for the nutrition of human and farm animal. Moreover, they are considered as special source, which provide essential amino acids. Other than storage, it is also known for wide range of secondary activities such as lectin like activity, chitin binding activity, insecticidal activity, protease inhibitor activity, auxin binding activity, etc. (Moura et al., 2007; Soares et al., 2007). There are several classes of seed storage proteins as followings.

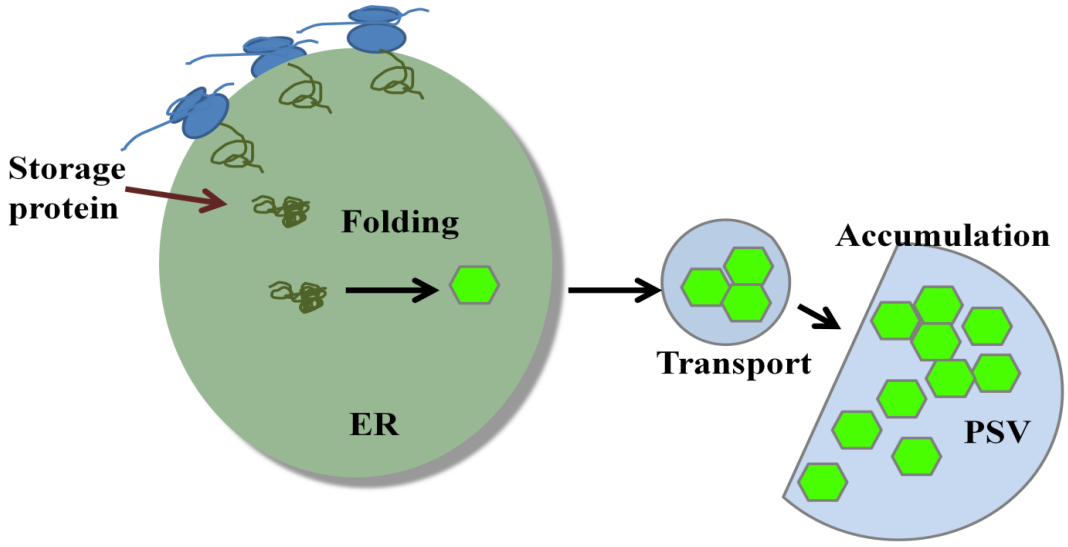


Figure 1.1. A schematic depiction of the synthesis of typical SSPs during the maturation of seeds that starts over endoplasmic reticulum followed by folding, packing, transport and finally accumulation inside the protein storage vacuoles.

1.2.1. The 11S seed storage proteins

The 11S class of SSPs are among the major form of storage proteins observed most profoundly in dicots and also found in monocots as well as primitive forms such as gymnosperm and spores of ferns (Shewry et al., 1995). There are substantial variations of total SSPs percentage with respect to the total protein content have been found among different seeds. In some species such as pumpkin and cucurbitaceous family almost entire protein proportion is shared by SSPs, while almost absence of SSPs has been reported from some cereals such as rye and barley (Danielsson, 1949). The synthesis of these globulins takes place during seed development and stopped immediately after seed maturation. The synthesis of 11S SSPs take place over endoplasmic reticulum and the proteolytic processing for the signal peptides occur inside the endoplasmic reticulum. Then they are routed for transportation to storage vacuole where they undergo processing and packed for a period of time until the germination of seedlings (**Figure 1.1**).

The 11S globulins are the largest class among SSPs, which exist in hexamer form. Each of the monomers is organized together by non-covalent interactions. The monomers are constituted by two subunits named as alpha and beta, the former is also known as acidic and

later as basic. Both the subunits are linked by disulphide bonds (**Figure 1.2. A**). A proprotein with signal sequences and intact subunits is responsible for the formation of mature alpha and beta subunits. The precursor proprotein undergoes processing of signal sequence, which is followed by folding and formation of disulphide bond between alpha and beta subunit. The formation of disulphide bond is followed by a cleavage, which is made by Asparaginyl endopeptidase at conserved Asparagine and Glycine amino acid residues (Jung et al., 1998). Some cases undergo one additional step of processing, which lead to an extra cleavage over alpha subunit, retained in mature form by non-covalent interactions (Shewry and Casey, 1999). The subunit pairs are observed with average molecular weight ~ 40 kDa for alpha subunit at N-terminal and ~ 20 kDa for beta subunit at C-terminal. The processed subunits of monomers then form a scaffold of hexameric assembly (**Figure 1.2. B and C**). Apart from the position of two subunits they also differ in extent of stability. The mature form of 11S remains stored inside the seed storage vacuole, until germination begins. Once the seed uptakes water and get imbibed, the degradation of stored globulins starts. There are many factors, which influence the regulation of storage protein degradation. But major charge for degradation of the storage protein is taken up by proteases. Furthermore, with the mobilization of storage protein the alpha subunit first undergoes degradation in just after 72 hrs, contrary to this the beta subunit which showed no obvious change even after 120 hours, as observed from soybean seed as well other members of dicot family (Kim et al., 2011). A recent finding suggests that the extent of carbonylation also governs the susceptibility of degradation, as reported from two subunits of Cruciferin (11S SSP from *Arabidopsis* seed). In addition to this, it has been concluded that alpha and beta subunits undergoes different extent of carbonylation (Job et al., 2005). These observations suggest that the mobilization of 11S subunits takes place in non uniform manner for the two subunits.

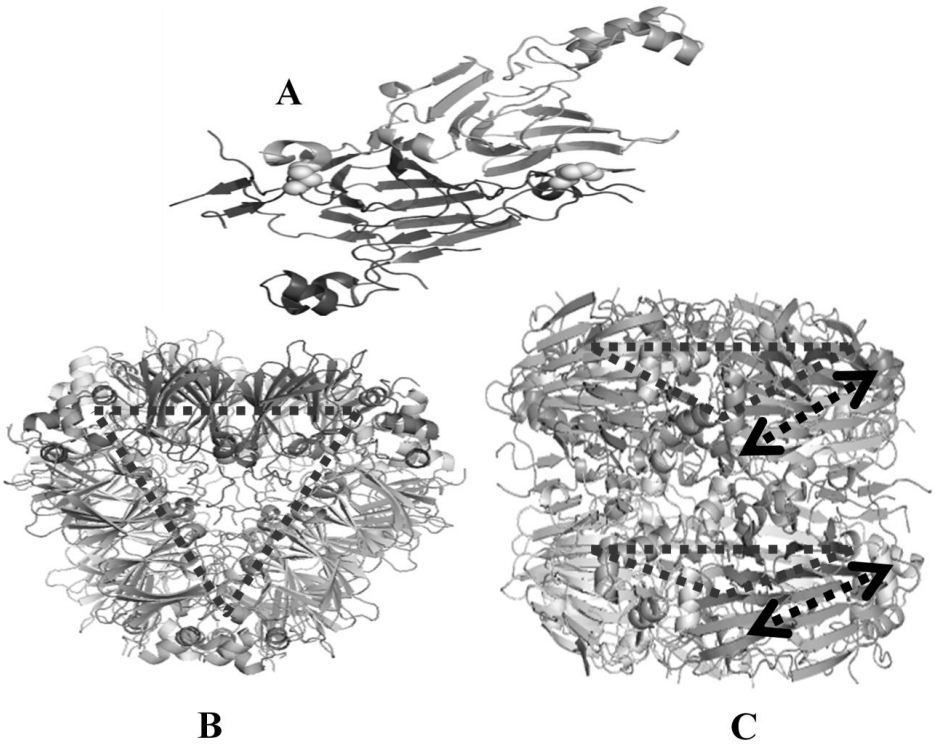


Figure 1.2. **(A)** Monomer of 11S globulin depicting alpha (dark gray lower half) and beta subunits (light gray upper half) encompasses two disulphide bonds (left side inter chain between alpha and beta subunits and right side intra chain in between alpha subunits). (**B**) Depicts a typical upper trimer of 11S globulins while lower is hidden in beneath. (**C**) The dotted triangle depicts a formation of dimer of trimer while double headed dotted arrow depicts a dimer can be repeated over three fold symmetry to form trimer of dimer of 11S hexamer.

The search for post translation modification to 11S globulins gives a conclusion that this class of storage proteins are not usually glycosylated, although as exception the lupin has been reported as glycosylated (Duranti et al., 1988). Despite of similar processing and profile of subunits of 11S globulins, there are substantial size variation has been observed in same as well different species. This size variation of 11S SSPs lead to high molecular heterogeneity, as observed from pea where 22 different alpha and 11 different beta subunits have been reported (Matta et al., 1981). Moreover, some other heterogeneous 11S SSPs are also found in glycinin from Soybean and legumin of green gram (Casey et al., 1986) **(Table 1.2)**.

Table 1.2. The list of 11S class of globulins from some common legume plants and their range of corresponding subunit size. 1-(Matta et al., 1981), 2-(Derbyshire et al., 1976), 3-(Chandna et.al, 1994), 4-(Krishna et.al, 1987), 5-(Melo et al., 1994), 6-(Matta et al., 1981), 7-(Staswick et al., 1984).

Name of Protein	**Plant**	**Subunit size (kDa)**
Legumin[1]	Broad bean	Acidic 23.0 – 58.0 Basic 21.0 – 23.0
Legumin[2]	French bean	Acidic 20.0 – 37.0 Basic 26
Legumin[3]	Grass pea	Acidic 32.0 – 89.0 Basic 20
Arachin [4]	Ground Nut	Acidic 21.4 – 47.5 Basic 21.4
α-conglutin[5]	*Lupinus albus*	Acidic 19.0 – 46.0 Basic 20-22
Legumin[6]	Pea	Acidic 37.6– 40.4 Basic 20.7 – 21.9
Glycinin[7]	Soybean	Acidic 10.0 – 42.0 Basic 20.0

1.2.2. 7S Seed storage proteins

This class of proteins are also named as Vicilin on the basis of their characterised taxonomical plants family '*Viciae*' a legume from which they were first extracted (Osborne and Campbell, 1898). They are smaller form of SSPs, marks its presence in dicotyledonous plants especially in legumes sharing 50-90% of the entire seed protein (Utsumi, 1992; Yadav et al., 2011) (**Table 1.3**). The 7S SSPs exhibit a trimeric arrangement, where each monomer interacts non-covalently to form the trimer. The typical trimer of vicilin class exhibits molecular weight range of 150-190 kDa and the monomers of 50-75 kDa and is mostly found in glycosylated form. The 7S SSPs are having highly polymorphic nature due to involvement of the multiple genes for their protein synthesis and the extensive post translational modifications. Furthermore, on the basis of post translational proteolytic processing, the 7S SSPs are divided in two types one that undergoes post translational processing such as the pea and faba bean, while the others that does not undergoes this processing like soybean and *Phaseolus* (Shewry et al., 1995).

At the gene and transcript level, the interesting correlation has been observed for the 7S SSPs. Except some convicilin like examples, most of them are expressed by genes having six exons and five introns (Utsumi, 1992). The reasonable homology has been observed among 7S SSPs genes that also display nucleotide homology at the exon/intron junctions also. A TATA box is present among all the genes, present at 30-40 base pairs upstream to transcription initiation site. Moreover an animal core enhancer sequences are also present on the genomic region responsible for the synthesis of 7S SSPs. The vicilin box is present 120-140 bp upstream to transcription initiation site in almost all the genes invariably and is believed to be involved in regulation of 7S globulin genes (Gatehouse et al., 1986).

The major difference of 7S globulin to other SSPs is absence of the cystine residues that result in absence of disulphide bonds. Furthermore, the posttranslational proteolysis and glycosylation imparts significant subunit variations to 7S SSPs in comparison to other SSPs. Despite of these differences, the 7S SSPs shares reasonably similarity with 11S globulins. These similarities encompasses, analogy between trimer of 7S globulin to initial trimeric state of 11S globulin has been observed during their processing (Gatehouse et al., 1984), sequence similarity of both the classes except four insertion region of 11S globulin, similarity for basic

chain of 11S globulin and C-terminal of 7S SSPs (Lawrence et al., 1994). So the study of 11S globulin also helps in understanding of 7S vicilin.

Table 1.3. The table is displaying a list of major studied of 7S and 11S globulins with the corresponding molecular weight range and the plant sources from which they have been extracted.

Species	**Protein**	**Molecular weight (kDa)**
	7S	
Glycine max (soybean)	β-conglycinin	150-250
Canavalia spp. (jack bean)	Canavalin	170
Pisum sativum	Convicilin	150-190
Lupines spp.	Conglutin-β	185
Phaseolus vulgaris	Phaseolin	140-163
Pisum sativum (pea)	Vicilin	150-190
Vicia faba (broad bean)		150
Vigna unguiculata (cow pea)		170
	11S	
Arachis hypogeal (peanut)	Arachin	330-350
Lupines spp.	Conglutin α	315
Brassica napus (oilseed rape)	Cruciferin	300
Cucurbita spp (melons, cucumbers)	Cucurbitin	325
Glycine max (soybean)	Glycinin	320-375
Helianthus annulus (sunflower)	Helianthinin	300-350
Pisum sativum	legumin	330-450
Vicia faba (broad bean)		320-400
Vigna unguiculata (cowpea)		300-400

1.2.3. 2S albumin

The 2S albumins was initially named as ‘albumin’ due to resemblance to hen egg albumin property of getting coagulated up on heating, but subsequently this classification has been restricted for proteins soluble in water (Osborne, 1924) and the 2S is designated due to sedimentation coefficient (Youle and Huang, 1981). The 2S albumins are one of the major classes of seed storage proteins having wide distribution among mono- and di-cotyledonous plants. The 2S albumins are related to the prolamin super-family, this super-family also encompass other allergenic proteins such as non-specific lipid transfer proteins, α-amylase/trypsin inhibitors and other SSPs from cereals. The 2S albumins share a varying percentage of total protein content as reported in 30% from Brazil nut, 20% in sunflower (Youle and Huang, 1981) and 13% in *Brassica* (Monsalve and Rodriguez, 1990). The 2S albumins are also deposited inside protein bodies in the developing seed and are subsequently utilized during the germination as a source of nitrogen and carbon.

Similar to most members of SSPs, the 2S albumins also exhibit high level of polymorphism. Several isoforms are produced due to the involvement of multi gene family for their synthesis (De Clercq et al., 1990). Furthermore, theses gene products undergo post translational modifications by proteolytic processing in most of the cases. Hence these isoforms may show varying biological properties. But despite of multiple isoforms the typical structural properties among the members of class remains same. The basic polypeptide responsible for the formation of mature protein is synthesized as single precursor of approximate molecular weight 18-21 kDa. The processing begins with the co-translational transport inside the lumen of endoplasmic reticulum, where the four intra chain disulphide bond forms by involving eight conserved cystine residues. The folded polypeptide subsequently undergoes proteolytic processing and reduced to a molecular weight range ~ 12-14 kDa, which is finally processes into larger subunit of molecular weight ~ 8-10 kDa and smaller subunit ~ 3-4 kDa (Shewry et al., 1995). Subsequently, the eight conserved cystine residues forms two inter and two intra chain disulphide bonds. The two inter chain disulphide bond connect smaller and larger subunits, while two intra chain disulphide bonds remains between larger subunit only (**Figure 1.3**).

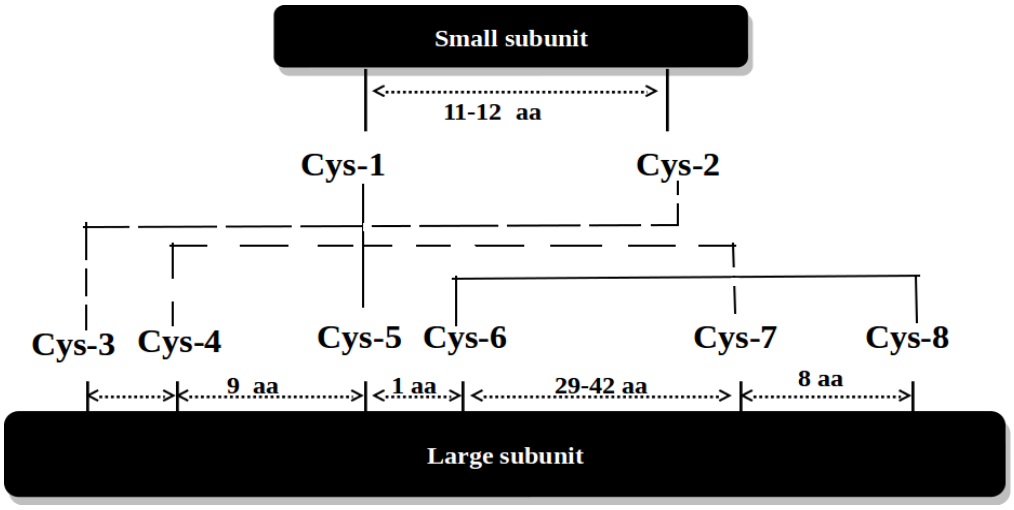


Figure 1.3. A typical arrangement of two subunits of 2S SSPs is depicting the disulphide bond arrangement for common examples.

The 2S albumins has unusually high proportion of sulphur containing residues that ranges from 6 to 13 mol %, while in few instances such as in Brazil nut they are reported to consist relatively high methionine content up to 17 mol % (**Table 1.4**). The 2S SSPs are known for the invariant presence of the skeleton of eight cystine residues, inside the α-helical arrangement of protein.

Table 1.4. Table is showing some examples of 2S albumin with the varying content of methionine. 1-(Altenbach et.al., 1990), 2-(Kortt et.al., 1990), 3-(Galau et al., 1992), 4-(Segura-Nieto et al., 1992).

Species	Protein	Methionine content (mol %)
Amaranthus[1]	MRP1	18.6
Brazil Nut[2]	2S albumin	18.89.8
Cotton[3]	Mat5	9.8
Sunflower[4]	SFA8	15.5

The 2S albumins are reported for many remarkable features. The ionic profile of overall protein has been observed as a major factor for the anti pathogenic activity, by virtue of this profile the antifungal activity has been observed from 2S SSPs and related proteins (Pelegrini et al., 2006). Due to exceptional composition of sulphur containing amino acid the 2S albumins have been considered as an excellent source, for the improvement of sulphur content in dietary proteins (Saalbach et al., 1995). Beside these, it also displays significant metal binding ability, but very limited studied has been reported from 2S albumins on this aspect (Rundqvist et al., 2012).

1.3. Evolution of seed storage proteins

The seed storage proteins belongs to a highly diverse family of proteins termed as cupin super family, the name derived from cupa a Latin synonym of barrel. The cupin super family is one of the most ancient families of proteins found across diverse forms of life, their members are present from archaea and bacteria to eukaryotes covering plants as well animals. The discovery of this class has been made on the basis of identification of sequence similarities between plant seed storage globulins, wheat germin (an exceptionally thermostable and protease resistant glycoprotein) and fungal spherulins produced in spores of slime mold *Physarum polycephalum* (Lane et al., 1991; Baumlein et al., 1995). These proteins shares a common structural arrangement of 6-stranded β-barrel. The subgroups of this family involve number of cupin domains present in proteins, hence classified as monocupins, bicupins and multicupins. The diverse members of this family constitute minimum 18 different subclasses involving various enzymatic and non enzymatic proteins. It appears that as a progenitor, the monocupins encompasses majority of cupin proteins. The members of cupin super family are monomeric, dimeric and polymeric. This diverse state attributes several kinds of functions, but most of them exhibit enzyme activity like phosphomannose isomerase and dioxygenase (Dunwell et al., 2004). Apart from this, the thermostable nature of the cupin fold is exploited by thermophilic archaea that expresses a stable cupin domain in enzymes like *Pyrococcus furiosus* phosphoglucose isomerase, reported for the maximum activity at 90 °C (Verhees et al., 2001). Their diverse forms also encompass a monocupin called auxin binding protein (ABP), which is of immense importance in plants due to auxin hormone interaction. These ABPs are found responsible for variety of plant growth regulation. They exhibit a reasonable sequence

similarity across wide range of plants. The Germins are largest member of oligomeric cupin super family. The wheat and barley germins are highly thermostable oxalate oxidases moreover the barley germin is also reported for superoxide dismutase activity (Woo et al., 2000).

The remarkable similarity of the two domain arrangement of oxalate decarboxylases (OD) to the storage globulin concludes that it may be the direct and sole progenitor of seed storage globulins (Dunwell et al., 2000). Presumably, the gene encoding cyanobacterial OD was taken up by plant symbiogenetically and transferred in to the nucleus by endosymbiosis and further evolved in to modern storage globulin of plants (Martin and Herrmann, 1998). The bicupins or 2-domain cupins are initially reported from higher plants both in dicots as well as in monocots and classified in two major classes, as 7S globulin and 11S globulin, share the common ancestry (Shutov et al., 2003) (**Figure 1.4**). The overall structure of 7S and 11S is organized in trimer and hexamer respectively. The evolution of 11S and 7S globulin shares a disparity in their common ancestry, it is interesting to observe that genes encoding 11S in monocots belong to same gene family as of dicots but the genes encodes 7S globulin in monocots has no evolutionary correlation with the 7S of dicots (Li et al., 2012). Furthermore, the amount of 11S globulin in dicots is higher in comparison to monocots. In general the higher amount in dicot is due to higher copy number of their genes, while in some particular cases higher copy number has been observed from monocot also, as observed in the examples of *Brachypodium distachyon* and *Oryza sativa* (Li et al., 2012).

At the amino acid sequence level these two classes of SSPs share less similarity with an averaged range 35-45%, but display a remarkable extent of structural similarity. The structural comparisons at primary and tertiary level conclude that 7S and 11S globulins are evolutionary related, thus shares a common evolutionary precursor (Shutov et al., 1995). Moreover a spore specific protein reported from ferns encompasses the features of 7S vicilin as well as 11S legumin, thus considered as a progenitor for both types of SSPs (Shutov et al., 1998; Shutov and Baumlein, 1999) (**Figure 1.4**).

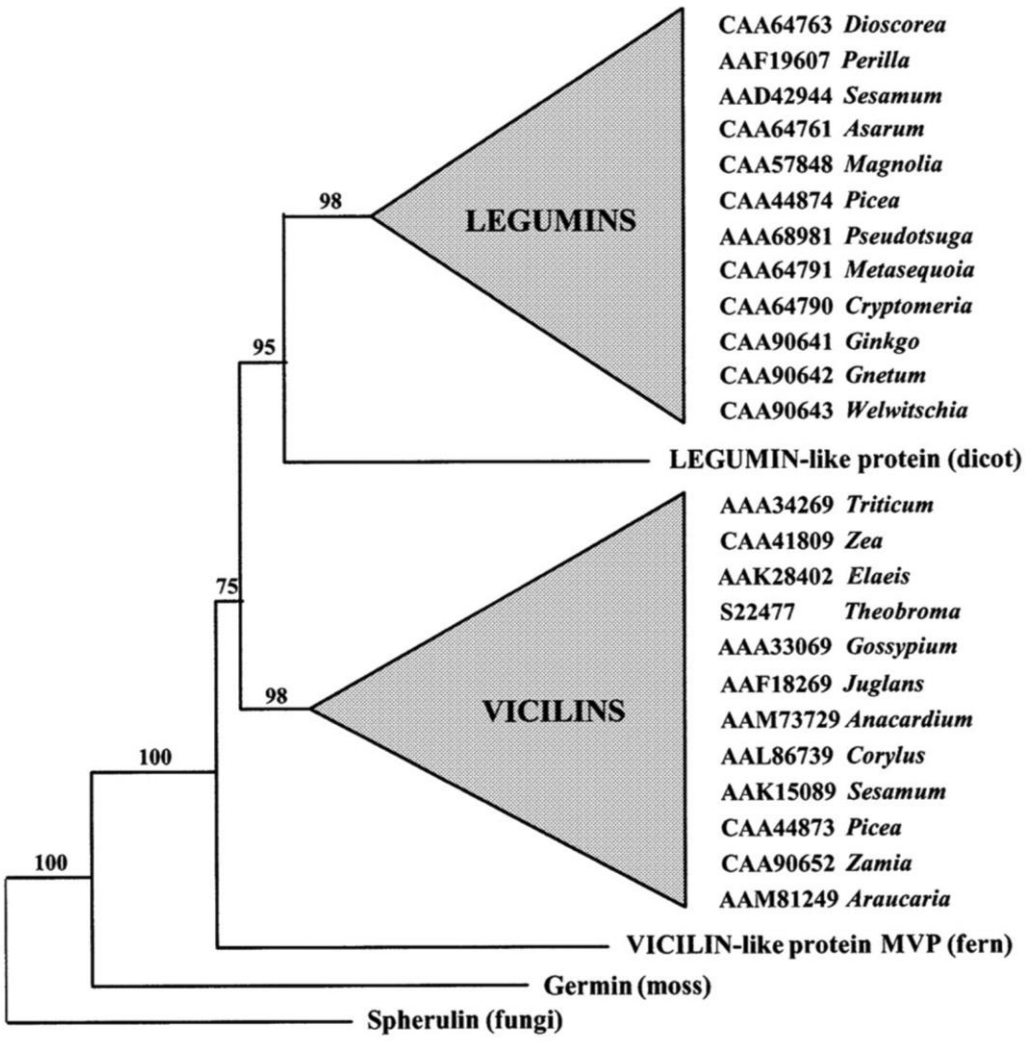


Figure 1.4. The figure depicts a common evolution of 7S vicilins and 11S globulins in monocots and dicots and their distantly related homologues (Shutov et al., 2003).

The structural homology reveals that each subunits of the storage globulin consists two similar domains, characterized by the presence of typical β-barrel made up of eight anti parallel β-strands and a group of two or three α-helices (Adachi et al., 2001). The amino acid sequence involved in constituting these domains share reasonable homology, hence it concludes that both the domains are evolved from common single domain ancestor that had undergone duplication sometime during evolution (Lawrence et al., 1994). In this respect, the comparisons of germin with 7S and 11S globulins suggest that it can be a representative of the single domain ancestor (Baumlein et al., 1995). Presumably, a small changes in functional domain had imparted

substantial diversity, as observed form several class of proteins such as transcription factors, sugar binding proteins, auxin binding protein and active enzymes like oxalate decarboxylases and ring cleaving dioxygenase (Khuri et al., 2001; Fetzner, 2012).

The seed storage protein of the 2S albumins class and related proteins displays a wide distribution across plant kingdom that covers flowering plants (Youle and Huang, 1981), gymnosperms (Flinn et al., 1993) and the spores of non vascular plant (Templeman et al., 1988). Due to evolutionary divergence now these examples only shares a limited sequence identity but the presence of conserved cystine make a strong evolutionary correlation among members of this family (Ampe et al., 1986; Dasgupta et al., 1995). The 2S albumins are encoded by multi-gene family having different numbers, as observed in *Arabidopsis thaliana* with four tandem genes (Krebbers et al., 1988), while in *Brassica Napus* there are as many as sixteen genes has been reported (Josefsson et al., 1987). Hence it has been reported previously that because of gene conversion and gene duplication this class posses a very wide diversity (Raynal et al., 1991; Dasgupta et al., 1995). But later it has been concluded that the duplication of the numbers of 2S albumin gene was occurred prior to the Sysimbrieae-Brassiceae split and after that they are homogenised and evolving in coordination (Boutilier et al., 1999).

1.4. The plant *Wrightia tinctoria*

The plant *Wrightia tinctoria* belongs to the Apocynacea family. This is a small and deciduous tree geographically distributed throughout the Indian continent. Several exceptional medicinal properties are attributed by this plant; such as antibacterial, antifungal, antipsoriasis, antiviral, antidandruff, antinociceptive, anti-inflammatory, anthelmintic, antioxidant as well as contraceptive properties (Reddy et al., 2002; Keshri et al., 2008; Selvam et al., 2009; Jain and Bari, 2010; Ponnusamy et al., 2010; Kumar, 2011; De Britto and Sebastian, 2013; Rajalakshmi and Harindran, 2013).

CHAPTER - 2

PURIFICATION FROM NATURAL SOURCE, CLONING, BIOCHEMICAL AND BIOPHYSICAL CHARACTERISATIONS OF 11S GLOBULIN FROM *Wrightia tinctoria*

2.1. Introduction

The 11S globulin has been purified from a medicinal plant *Wrightia tinctoria* and characterised for its biochemical and biophysical properties (Kumar et al., 2013). The 11S globulin extracted from *Wrightia tinctoria* seed has been named as *Wrightia tinctoria* globulin (WTG). Like other members of 11S family, the major function of WTG also entails as storage reserve. In addition to this, a unique hemagglutinating activity has also been reported for this protein (Kumar et al., 2013). The molecular weight of the two subunits of WTG are similar to most of the reported members of 11S globulin family, the alpha subunit is having molecular weight of ~ 32-34 kDa and the beta subunit is having molecular weight of ~ 22-26 kDa. The heterogeneity for both alpha and beta subunits has been observed from WTG, which is in accordance with previously studied examples of globulins from oat, soybean and pea (Utsumi et al., 1981; Walburg and Larkins, 1983; Lambert et al., 1987). The randomly united acidic and basic subunits constitute the monomers of the protein, which further assemble and augment the heterogeneity in form of hexamer. As reported from buckwheat globulin that exists in hetero oligomeric state and composed of different polypeptide with remarkable differences in relative concentration and molecular weight of subunits (Choi and Ma, 2006).

The monomers of 11S globulin are the basic unit of organization which eventually forms the hexameric structure. Each monomer has N- and C-terminal modular organization encompasses a jelly-roll β-barrel motif and helical domain. This basic structural organisation is shared by 11S as well as 7S globulin family, moreover the structural conservation of these domains put forward strong evidence that a common evolutionary lineage is shared by two classes that evolved independently after gene duplication (Adachi et al., 2001). Furthermore at sequence level the 11S showed remarkable insertions at four regions when compared to the 7S globulins (Lawrence et al., 1994), this finding again advocates that both the classes share common steps in evolution.

Despite of evolutionary correlation the overall structure of 11S storage proteins are remarkably different from 7S storage protein. The 11S always exists in hexameric state having molecular weight 300-400 kDa with each monomer 50-60 kDa while 7S exists in trimer arrangement of molecular weight 150-200 kDa and each monomer of molecular weight 40-70 kDa (Pernollet and Mosse, 1983). The difference of disulphide bond is another disparity

between the two classes. The structure of 11S globulin has been concluded to possess hexameric arrangement and the disulphides bonds confer a major support from inter chain face as well intra chain faces to rope the subunits of monomers. (Dickinson et al., 1989), contrary to this the 7S globulin is devoid of any disulphide bond, hence it appears as one of the region for inability of 7S to form hexamer (Adachi et al., 2001; Adachi et al., 2003).

The focus of present chapter is isolation and purification of 11S globulin from the seeds of *Wrightia tinctoria* as a natural source, to especially elucidate the behaviour and possible ligand interactions in native state. Further the identification and characterization has been done by extraction, purification, cloning, biochemical and biophysical studies of WTG.

2.2. Materials and Methods

2.2.1 Materials

The *Wrightia tinctoria* seeds have been obtained from the local seed distribution centre in dry form. For the isolation of c-DNA, young seed pods of *W. tinctoria* were collected from the ridge of university of Delhi, New Delhi India and immediately frozen by liquid nitrogen and stored in -80 °C for further use. The chemicals and reagents have been purchased from their authentic manufacturing companies listed as Sigma-Aldrich Corporation, St. Louis, MO USA; Himedia Laboratories India Private Limited, Mumbai, India; Bio-RAD Laboratories, Hercules, California, USA; Invitrogen, Carlsbad, CA, USA; NEB, USA; Promega, USA and Merck Limited, Worli, Mumbai, India. The columns used for the protein purification was Hi-Trap DEAE FF (1 mL), HiLoad superdex 200 16/60 columns. For the calibration of column a HMW (high molecular weight) calibration kit were purchased from GE Healthcare, AB Uppsala, Sweden. The filter and concentration has been done by using Millex syringe filters and Amicon ultra concentrators. For blotting the PVDF membranes were used from Millipore Corporation, Billerica, MA. For dialysis purpose the membranes were purchased from Pierce, Rockford, USA having cut-off range 3500 Da. The protein estimation was done by protein estimation kit purchased from Bio-RAD Laboratories, Hercules, California, USA. Molecular biology reagents, vectors and enzymes were purchased from Invitrogen, Promega and NEB, USA.

2.2.2. Methods

2.2.2.1. Purification of WTG

The purification of WTG has been done from the seeds of *Wrightia tinctoria*, which begins with the soaking of seeds for overnight at the room temperature. The soaking has been done in pH controlled environment provided by the buffer A, containing 50 mM Tris-HCl of pH 7.5, 10% glycerol, 10mM MgCl2, and 0.1 mM PMSF. The soaked seeds were minced by mortar and pestle. The minced sample was first subjected to continuous stirring for 6 hrs at 4 °C and then centrifuged at 50,000 x g at 4 °C for 45 min. The centrifugation leads to the separation of the fats as uppermost intact layer that was removed and a clear supernatant was left after three repeated rounds of centrifugations. The crude sample has been purified to homogeneity in two steps of purification; the first step involves ion exchange while second is size exclusion chromatography. The first step of purification was done by application of a 1ml Hi-Trap DEAE FF column equilibrated in buffer A. The supernatant was loaded over pre-equilibrated column in buffer A than flow through was passed out. Then 25 ml of buffer A was passed to wash the non specific impurity bound to the column. The bound sample was eluted from the column by the application of NaCl gradient for a range of 0.0 - 1.0 M by using buffer B (buffer A + 1M NaCl). The analysis of eluted fraction has been done over 15% SDS-PAGE gel. The corresponding fractions visible over SDS-PAGE were subjected to dialysis against 1 L buffer A in three repeats. The dialysed sample was concentrated to 10 mg/mL by using Amicon Ultra concentrators. The second step of purification involves HiLoad 16/60 Superdex 200 size-exclusion column, which was pre-equilibrated in buffer A. The partially purified sample has been loaded with a flow rate of 0.5 ml/minute by application of 1 ml sample loop over the ÄKTA purifier (GE Healthcare). The elution profile has been assessed by measuring the absorbance at 280 nm. The calibration of size exclusion column has been done by HMW calibration kit containing ovalbumin (44 kDa), conalbumin (75 kDa), aldolase (158 kDa), ferritin (440 kDa), and thyroglobulin, (669 kDa). These known molecular weight markers were used to estimate the unambiguous calculation and plotting of standard curve and molecular weight determination as well calculation of void volume. The fractions of peak were run over the SDS-PAGE and major peak fractions were subjected to concentration by employing Amicon ultra-15. The homogeneity of pure WTG was analysed by SDS-PAGE with the

Comassie brilliant blue staining. The concentration of protein was determined by protein assay kit provided by Bio-RAD with the application of BSA as a standard protein to draw the curve. The purified protein was concentrated and dialysed in three repeats with one litre of buffer A and stored at -20 °C.

2.2.2.2. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis

The Sodium Dodecyl Sulphate Poly-Acrylamide Gel Electrophoresis analysis of the WTG has been done at different stages of purification using method of Laemmli (Laemmli, 1970). The gel has been made by combining discontinuous gel of 5% for stacking and 15% for separating gels in the reducing as well non reducing conditions. The staining has been done by Coomassie brilliant blue, while the molecular weight calculation has been done by employing standards of different molecular weight ranges containing aprotinin (6.5 kDa), trypsin inhibitor (21 kDa), carbonic anhydrase (29 kDa), ovalbumin (35.8 kDa), bovine serum albumin (56.2 kDa), β-galactosidase (125 kDa), phosphorylase b (101 kDa) and myosin (210 kDa).

2.2.2.3. Intact Mass Determination

For the intact mass determination the purified protein was mixed in 1:1 ratio in 50% dilution with acetonitrile and 0.1% TFA with the concentration of 10 pmol/µL. The spotting of diluted protein was done by using α-cyano 4-hydroxy cinnamic acid (CHCA) over MALDI target plate. The calibration of instrument was done before data acquisition by using BSA and ProteoMass Protein MALDI/MS calibration kit (Sigma Aldrich). For intact mass determination the spectra has been acquired in the linear ion mode.

2.2.2.4. Amino Terminal and Mass Spectrometry-based Internal Sequencing

The larger-subunit of WTG has been reported as the domain present at N-terminal end of seed storage globulin family (Shewry et al., 1995), therefore it was extracted from gel and subjected to sequencing by Edman degradation over an automated protein sequencer (model 494; Applied biosystems) at Columbia university protein sequencing facility, New York, USA.

A 15% gel was used to run the SDS-PAGE in reducing condition followed by electro blotting over polyvinyl fluoride (PVDF membrane) in CAPS (N-Cyclohexyl 1-3-aminopropanesulfonic acid) buffer of concentration 10mM of pH 11 and 10% methanol (Matsudaira, 1987). The band was excised from the membrane and the initial 10 amino acids lying at N-terminal of the sample were determined. For deducing the sequence homology the NCBI database (http://www.ncbi.nlm.nih.gov/) was used.

To get the internal sequence the excised bands were de-stained in a solution of 50% acetonitrile with 50 mM ammonium bicarbonate (ABC), incubated for 1hr in three repeated change of solution (Rosenfeld et al., 1992). To reduce the protein, the supernatants were replaced with ABC containing 10mM DTT and subjected for 15 min incubation at 56 °C. The last supernatant was discarded and 20mM iodoacetamide in 50mM ABC was added and incubated for 15 min at room temperature and then dried in vacuum. The rehydration of the dried gel has been done by 3.0 µL of 12.5 ng/µL of sequencing grade modified trypsin gold (Promega) and 50 mM ABC followed by the incubation for the period of 60 min at room temperature. Later 50µL of 10% acetonitrile in 50 mM ABC was added and for digestion, the sample was subjected to incubation for 18 hrs at room temperature. After the collection of supernatant the gel piece was subjected to extraction in 0.1% TFA and 50% acetonitrile for 40 min at 37 °C in successive aliquots. Finally the concentration has been done by combining the extracts using speed Vac. The sample was resuspended in 10 µL of 50% acetonitrile and 0.1% TFA. Spotting of resuspended polypeptide has been done over 384 well MALDI plates, after mixing with α-Cyano-4-hydroxycinnamic acid matrix in 1:1 (v/v) ratio. The analysis of peptide has been made over AB Sciex 4800 Plus TOF/TOF analyzer, in reflector ion mode. For the identification of peptide, the protein pilot 2.0 was used and for inspection of peptide the sequence database of NCBI was compared in specific mode using EST blast search of translated nucleotide database against protein sequence as query. Further, the search has been limited to EST in selected species of *Wrightia tinctoria* that has given the hit with one of its EST having GenBank id HS559044.1. The biological modification along with substitution of amino acid has been incorporated in protein pilot search for the ID focus factor together with the gel based ID search. The confidence score was set to 99.9% for the protein threshold detection. Finally the homology search was made in open database using obtained peptide sequence from MS/MS analysis of trypsin digested peptide.

2.2.2.5. Chemical Modification of Trypsin Digested Protein by SPITC for De novo Sequence

The SPITC modification of peptides were made to obtain the independent de novo sequence of trypsin digested peptide of *Wrightia tinctoria* globulin, by the application of previously describes methods (Joss et al., 2006). In brief, the peptide digested by trypsin has been adsorbed over ziptip. The washing of the peptide bound over ziptip has been done by 5% methanol and water containing 0.1% TFA. Once the peptide get desalted, were subjected to incubation for 1hr in 5mg/mL SPITC (Sigma Aldrich, St.Louis,USA) in 20 mM sodium bicarbonate at 55 °C. As the reaction was completed, the ziptip was washed in three repeats in 0.1 % TFA followed by the elution of peptide in 3 µL of 50% acetonitrile and 0.1% TFA. The deposition of eluted peptide has been made in CHCA on MALDI target palate. The manual selection has been done for SPITC derivatised (+215 Da) arginine-terminated peptide.

2.2.2.6. Cloning of 11S globulin

The developing endosperm of the *W. tinctoria* was used for the RNA extraction. The isolation of RNA has been done by using TRIZOL and the protocols were followed as provided by manufacturers. The one µg of isolated RNA has been used for reverse transcription with the kit of SuperScript® III One-Step RT-PCR system and for c-DNA amplification Platinum® Taq High Fidelity DNA Polymerase has been used. One µL of c-DNA template has been used for 11S globulin gene amplification by employing gene specific primers designed according to available ESTs of *W. tinctoria* sequences from NCBI database. The sequences for forward and reverse primers were 5'-ATGGCCAGTACTTCTTTTCTTCAAG-3' and 5'-GCTGTTGAGAGTGCTGACGG- TGCC-3' respectively. The amplified product of the initial reverse transcription against specific primers were of the length 960 base pairs has been isolated and ligated in to pGEM-T vector (Promega) (Marchuk et al., 1991). The vectors were transformed in to *E. coli* (DH5α). The transformed colonies were picked and plasmid has been isolated and sequenced.

For the sequence determination of full length gene a rapid amplification of cDNA ends has been carried out by employing SMARTer™ RACE cDNA amplification Kit from Clontech and their provided protocol. Gene specific primer has been designed on the basis of available

partial sequencing information. For the 3' RACE the forward primer sequence was 5'GGGCAGTCGGATCAAAGAACTTATTGTGG-3' and for the 5' RACE the reverse primer sequence was 5'GGCCTCTGTTCTCAATGACATGACGAG-3'. The amplified product has been cloned and sequenced by employing the similar protocol of partial cDNA sequencing. Further the gene specific primers were designed from start to end point of translation stretch to obtain full length (1440 bp) 11s Globulin open reading frame which was isolated and cloned into pGEM-T vector.

2.2.2.7. Biochemical Characterisation of WTG

2.2.7.1. Hemagglutination Assay

The hemagglutination activity of WTG was assessed against human erythrocyte. For the detection of activity a standard serial dilution of human erythrocytes has been made in multi-well microtiter plate. The 1 ml volume of blood sample was taken in sample tube containing 6% EDTA to inhibit the coagulation in the phosphate buffer saline (PBS) of pH 7.2. Washing of mixture has been done by centrifugation at 1000 g in same buffer for 5 min at 4 °C. The upper layer of supernatant was discarded cell pellet was resuspended in 10mL of PBS buffer. The repeated dilution and centrifugation has been done to make the upper layer clear. The erythrocytes were finally resuspended in PBS buffer and final 3% (v/v) blood suspension has been made. The constant 50 µL volume of protein has been kept in each well containing a varying amount of 10 µg/mL to 100µg/ml protein was incubate with 50 µL of 3% suspension of human red blood cells and the agglutination reaction was monitored after 1 hr of incubation at 37 °C. The specific activity was defined by that minimum concentration of WTG (in µg/ml), which is showing detectable hemagglutination (Moreira and Perrone, 1977; Horta and De Azevedo, 2001) (**Table 2.1**). The BSA has been used as a negative control at each step, where WTG protein was tested for hemagglutinating activity.

Table 2.1. The hemagglutination activity for WTG at different stages of purification.

Fractions	Total Protein (mg)	H.U.	Specific Activity (H.U/mg)	Fold Purification
Crude extract	385	1540	4	1
DEAE	62.7	697	11.1	2.8
Gel filtration	40	615	15.4	3.85

2.2.2.8. Biophysical Characterisations of WTG

2.2.2.8.1. Far-UV Circular Dichroism Spectrum

The secondary structure estimation of WTG has been done by the application of circular dichroism spectral analysis. The purified protein was used for the CD analysis over the Chirascan circular dichroism spectrometer (Applied Photophysics Ltd, Surrey, United Kingdom). The range of 190 to 260 nm has been used for the collection of CD spectra, the steps was kept at 0.5 nm wavelength with the average time of 3.0 s at 25 °C using the 1mm quartz cell under the constant nitrogen purge. Protein samples has been scanned at different concentration 0.1, 0.2 and 0.3 mg/mL in 20mM potassium phosphate buffer of pH 7.5 (Matsuura and Manning, 1994; Zemser et al., 1994). The solution were every time freshly prepared and filtered through 0.45 µm Millex syringe filter moreover the degassing has also been done before data acquisition to avoid the noise of the spectrum. There are three scans have been acquired for each protein sample and averaged. The baseline obtained from buffer has been subtracted to give the final values. Mean residual activity is the method of expressing the CD data. The analysis of data has been done by web based DichroWeb software (Whitmore and Wallace, 2004) and the neural network program CDNN. The 0.2 mg/mL concentration of protein was taken to assess the effect of chemical denaturant on secondary structure of WTG, in presence of maximum 8M urea and 6M GdnHCl for 4hr at 25 °C.

2.2.2.8.2. Intrinsic fluorescence spectrum

The purified protein was subjected to emission fluorescence measurement with a concentration 50 µg/ml in 20 mM phosphate buffer of pH 7.5. The protein was exited at 280 nm and the resulted emission wavelength spectra were recorded at 290-400 nm using Varian Cary Eclipse fluorescence spectrometer (Varian, Inc., Walnut Creek, CA) at constant temperature of 25 °C. The quartz cuvette of 1 cm path length was used while the emission slits were at 5 nm. The effect of chemical denaturants has been studied by the fluorescence emission spectra against urea, GdnHCl and different range of pH. To scan the stability of WTG, the protein was incubated overnight with 1 to 8 M urea and 1 to 6 M GdnHCl, 2 hr in the different pH solutions for a range 2 to 13 at room temperature.

2.3. Results

2.3.1. Purification and sequencing of WTG

The two steps of purification lead to the successful isolation of WTG in homogeneous form. The first step involves anion-exchange followed by second step of size exclusion chromatography. The crude extract was applied over the Hi-Trap DEAE FF column. After washing the column, protein was eluted by the application of NaCl gradient. The partially purified samples were subjected to hemagglutination assay and the positive results have been observed from elution range 0.45 M to 0.6 M, the analysis of purity was done by observing corresponding protein bands over 15% SDS-PAGE (**Figure 2.1. A**). The fractions of corresponding bands were pooled together and concentrated to 10 mg/ml. The second stage of purification involves application of concentrated partially purified protein over the size-exclusion chromatography in the AKTA purifier. The target protein eluted in the major peak and impurity showed up in minor peak factions (**Figure 2.1. B**). The pure protein was pulled for subsequent use. The final yield of protein was approximately 8mg/gram of seed.

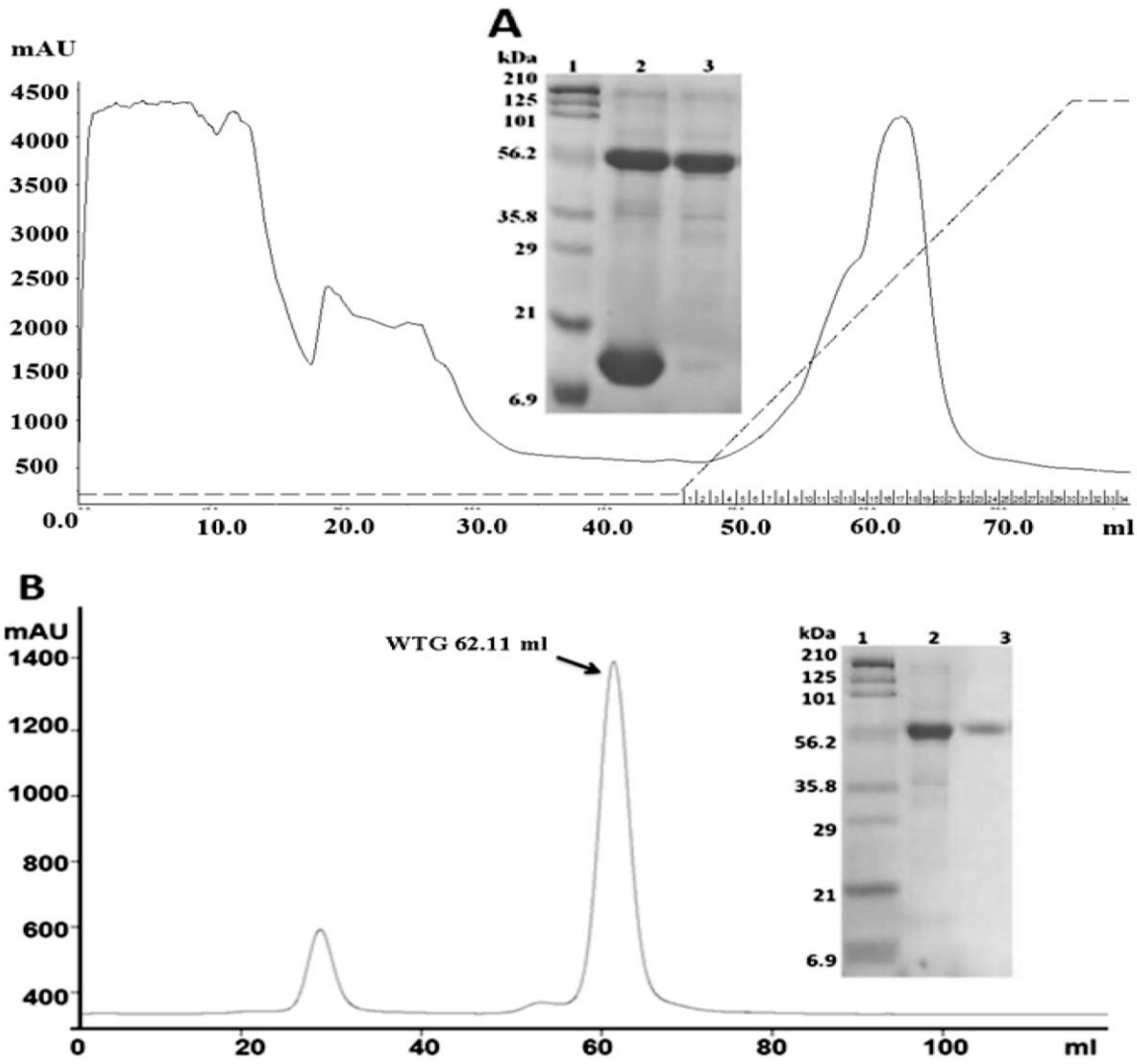


Figure 2.1. Purification profiles and SDS-PAGE of WTG: (A) DEAE chromatography profile and non-reducing SDS-PAGE in 15% gel: lane 1 - molecular weight marker, lane 2 - crude extract, lane 3 - DEAE eluted fraction. (B) Size-exclusion chromatography profile and non-reducing SDS-PAGE in 15% gel: lane 1 - molecular weight marker, lane 2 and 3, Superdex 200 peak fractions.

The isolated protein from native source was analysed on 15% SDS-PAGE in presence and absence of reducing agent (β-mercaptoethanol). In the absence of reducing agent a single band of approximate molecular weight ~ 56 kDa was observed while the presence of reducing agent splits a single band in heterogeneous populations of two major subunits. The estimated molecular weight of these peptides was ~ 32-34 kDa and ~ 22-26 kDa (**Figure 2.2**).

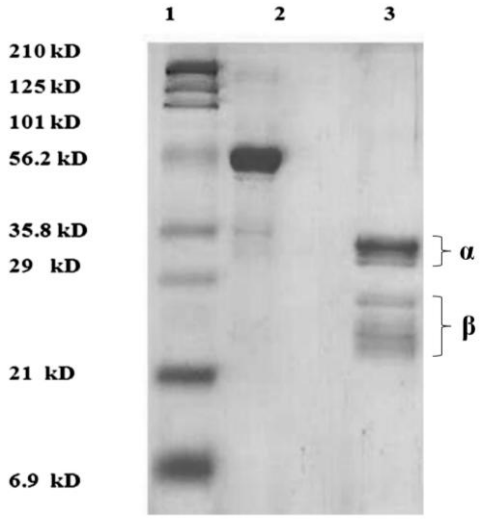


Figure 2.2. SDS-PAGE of WTG showing lane 1 - molecular weight marker, lane 2 - purified protein in the absence of reducing agent, lane 3 - purified protein in the presence of reducing agent (Subunits are designated as α and β respectively)

For the analysis of oligomerization state, the gel filtration column has been used with the known standard molecular weight proteins. The major peak of the WTG has been observed to eluted at a volume of 62.1 mL in 50 mM Tris-HCl buffer, pH 7.5 (**Figure 2.1. B**). The calculation made on the basis of standard curve suggests the approximate molecular weight of 320 kDa.

The intact mass determination of WTG has been tried by MALDI-TOF and it also concludes that a population of heterogeneous subunits are present in the sample as evident from the multiple peaks with a range of mass/charge ratio (**Figure 2.2 B**). With the assumption of highest mass/charge ratio plausibly represents the population of intact mass, the molecular weight of WTG is likely to be 49,570 Da.

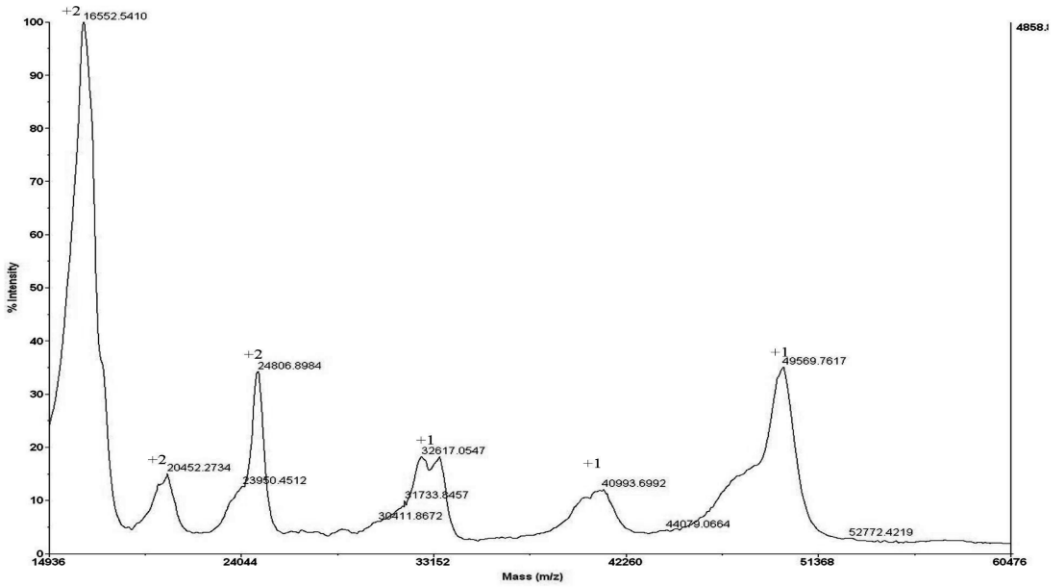


Figure 2.3. Intact mass determination by MALDI-TOF indicates the presence of multiple subunits in WTG that vary in their concentration. In the intact mass spectra of the native protein three singly charged molecular ion peaks of protein subunits at the molecular weight of approximately 33.1 kDa, 40.9 kDa and 49.5 kDa were observed. The +2 charged peaks indicated molecular weights of approximately 16.55 kDa, 20.45 kDa and 24.80 kDa, respectively.

For the determination of N-terminal sequencing the purified WTG was blotted over PVDF membrane from reducing SDS-PAGE gel and N-terminal sequence has been determined. The obtained strings of amino acids are LRTPQLNEAQ. The homology search has been done by using this sequence with the specialized blast tool in NCBI database against the expressed sequence tags for translated nucleotide database using protein query with the customize option of organism *W. tinctoria*. The 100% matched identity of ESTs of this plant with the developing embryos of following gene identities from GenBank: HS560184.1, HS557004.1, HS562364.1, HS571038.1, HS572228.1, HS563045.1, HS561676.1 and HS558020.1. Out of all these hits the ‘HS572228.1’ is showing the maximum length of 256 amino acids and has been used for the identification of the protein and it ascertains that this protein belongs to the 11S globulin family (**Figure 2.4**).

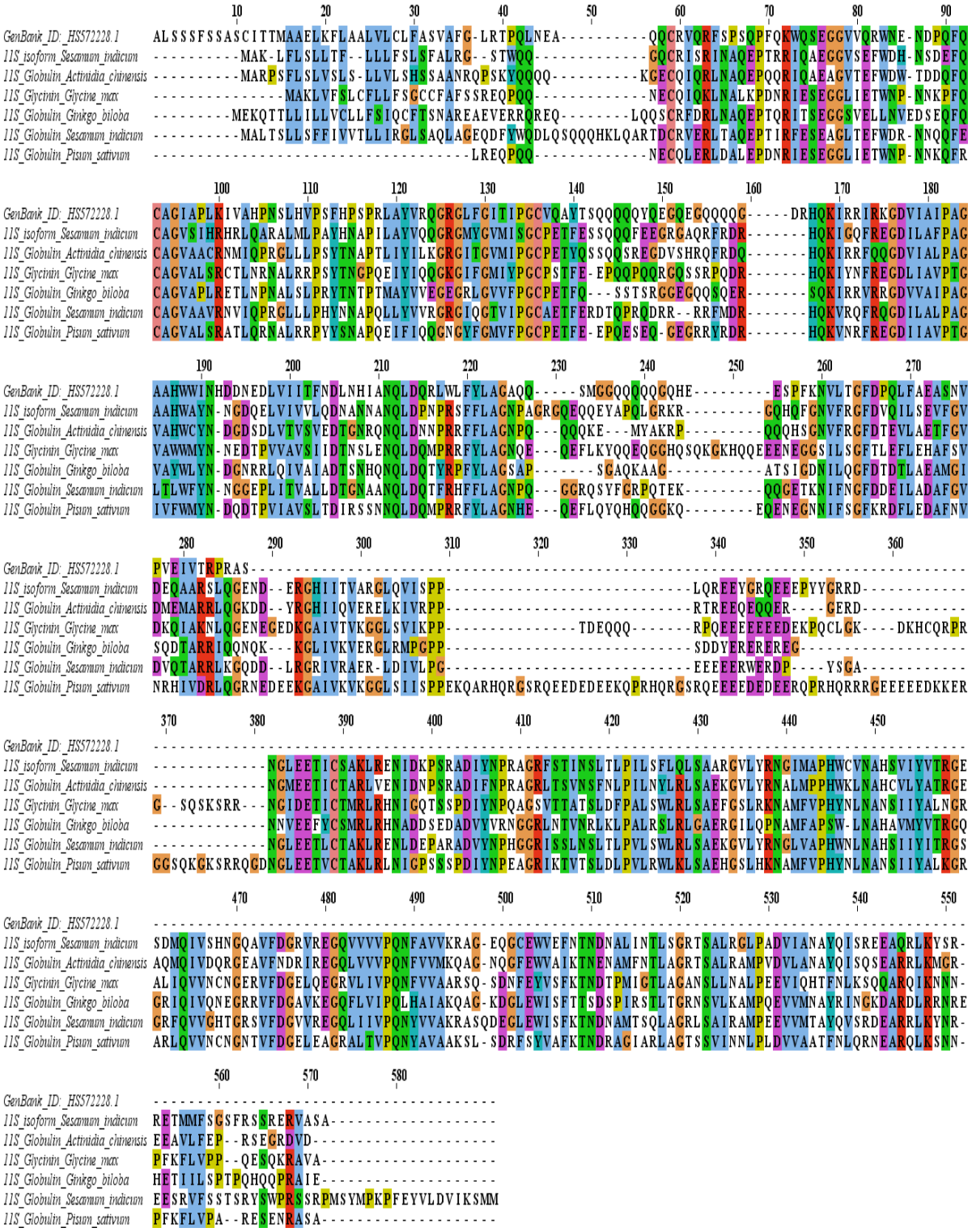
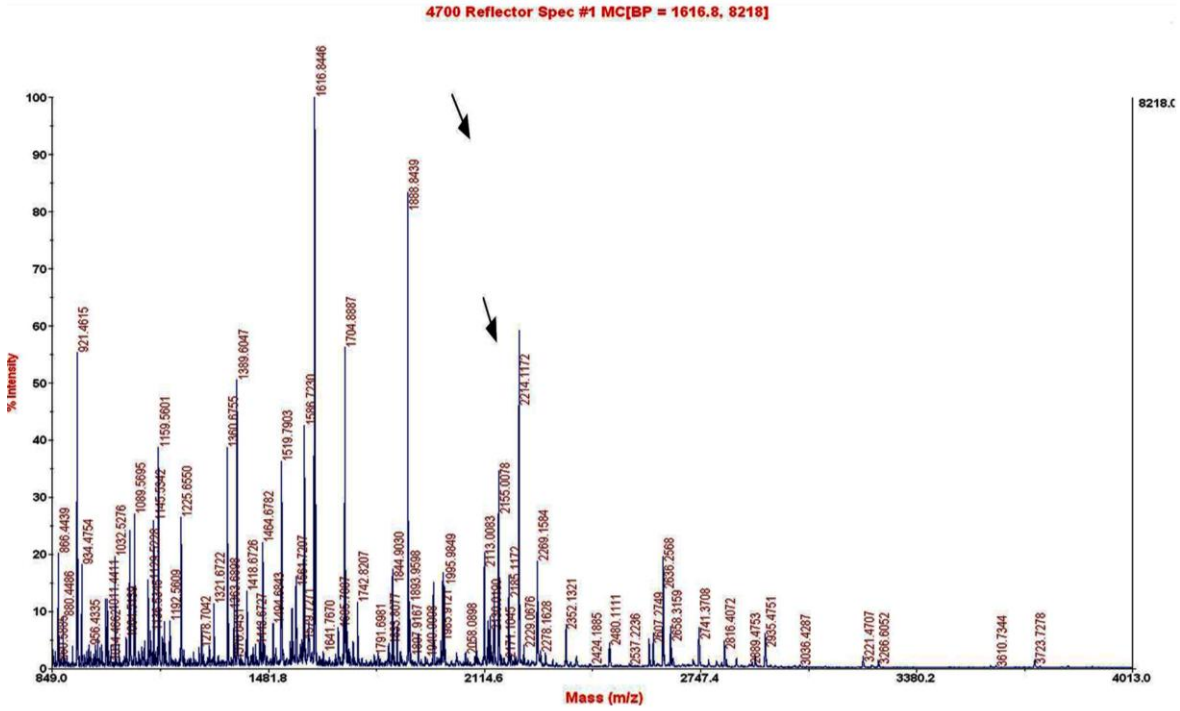


Figure 2.4. Sequence alignment profile of N-terminal amino acids of WTG based on translated EST of *W. tinctoria* with the known homologues of 11S seed storage globulins.



PMF Spectra of Trypsin digested protein

Figure 2.5. MALDI TOF/TOF analysis of WTG. Peptide mass fingerprint (PMF) of trypsin digested protein with arrows indicating the peptides used for MS/MS analysis.

To verify the EST sequence obtained on the basis of N-terminal sequencing, the partial internal sequencing has been also done by mass spectrometry. Here the MS and MS/MS spectra were acquired using trypsin digested protein (**Figure 2.5**). The sequences of some of the peptide have been generated by the MS/MS spectra. The obtained partial sequences were used for the comparisons of obtained peptide with N-terminal amino acid based translate EST, which shows 37% identity of the GenBank: HS572228.1 and a maximum identity of 69% to the GenBank: HS559044.1.

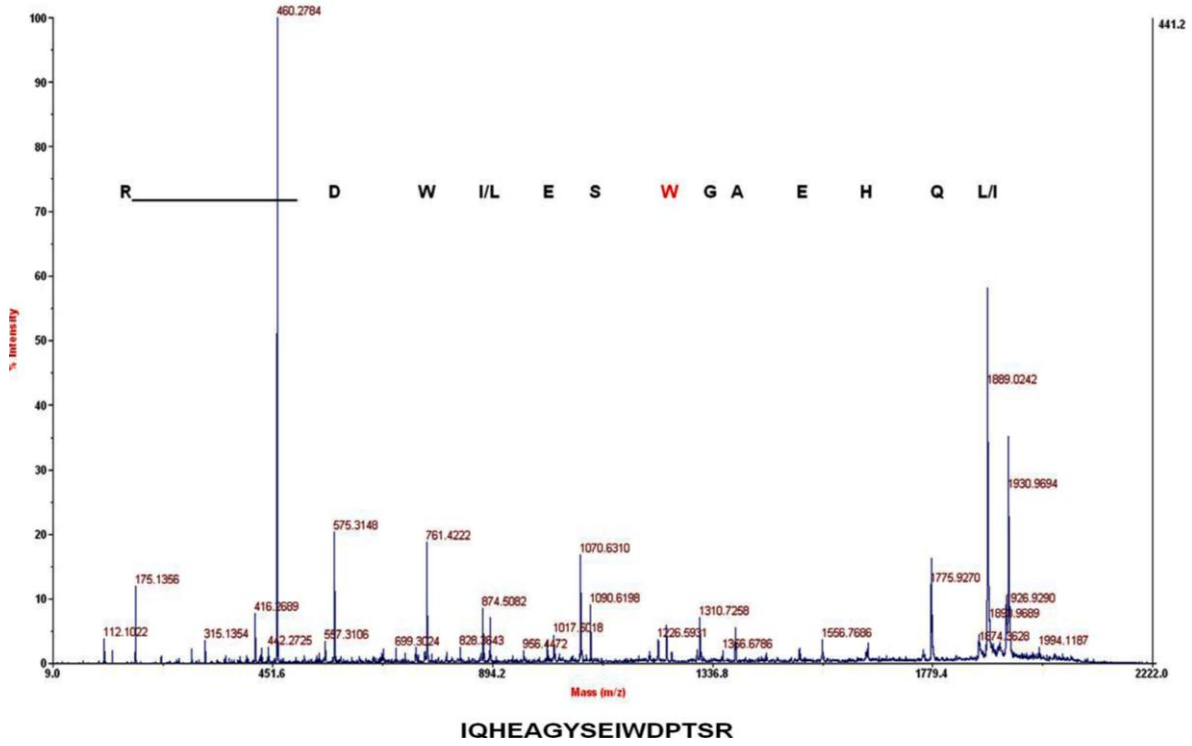


Figure 2.6. MS/MS spectra and de-novo sequence of a peptide following sulphonation by SPITC.

The *De novo* amino acid sequence of one internal peptide of WTG was also obtained (**Figure 2.6.**): IQHEAGYSEIWDPTSR after chemical modifications of peptide that helped suppress the "y" ions (Joss et al., 2006). The non redundant NCBI blasts were used against the obtained internal sequence, which confirm that the sequences are from the same 11S seed storage family.

2.3.2. Cloning of 11S globulin

The developing seeds of *W. tinctoria* have been used for m-RNA extraction. The reverse transcription has been done from the extracted m-RNA for the synthesis of cDNA. The obtained cDNA has been used as a template for amplification of gene. The specific primers were designed on the basis of available ESTs. The initial amplification has been done by using

the cDNA as template and EST based primers, which leads to the amplification product of length of 960 bp. This partially amplified product has been ligated and cloned in the vector. The sequencing results of this clone confirm that it as a part of 11S gene.

Further, the rapid amplification of cDNA end has been performed by using a specific set of primers designed on the basis of partial sequencing information. Finally the full length gene has been synthesized by using the set of primers covering start to the end of gene and amplified cDNA. Further the amplified product has been cloned and sequenced, which confirms the exact length of gene of 1440 bp that corresponds to the translation sequence of 11S globulin of WTG **(Figure 2.7**).

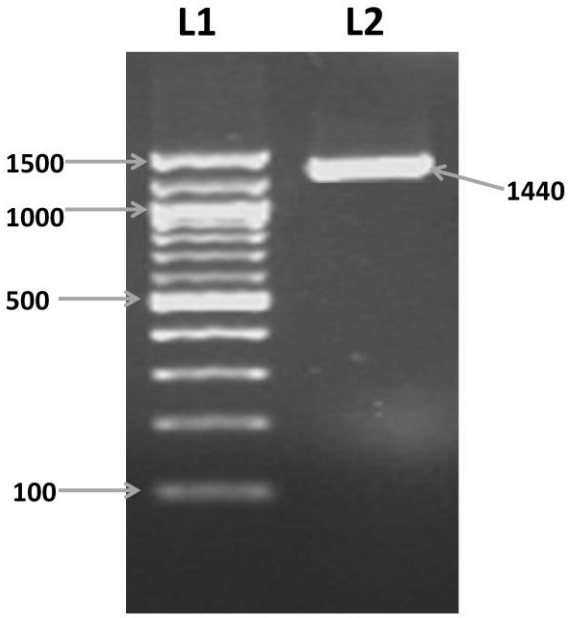


Figure 2.7. The DNA ladder depicting the range of marker from 100 base pairs to 1500 base pairs in lane 1, while lane 2 corresponds to the full length 11S globulin gene of 1440 base pairs.

2.3.3. Hemagglutination activity of WTG

The hemagglutination activity has been observed from the WTG against erythrocytes. This activity has been assessed at each stage of the purification and it has been observed that the extent of hemagglutination rises with the purity of protein (**Table 2.1**). The specificity of the hemagglutination activity conferred by WTG has been examined by the application of BSA as negative control that confirms only WTG is the responsible source for hemagglutination of erythrocytes.

2.3.4. Circular Dichroism Studies of WTG

Far-UV CD spectroscopy has been employed for the determination of secondary structure of purified WTG as well as for assessing the effects of denaturants. The maximum negative mean residual ellipticity [θ] has been showed near 210 nm for WTG. The dconvulation of the CD data by CDNN program showed that WTG secondary structure encompasses 44.2% β-sheets, 10.3% α-helix, 20.1% β-turns, and 26.6% random coil. Furthermore, the sample was incubated with different denaturants. The results of denaturation study showed that no pronounced change is visible until the globulin is subjected to higher concentration of chaotropic reagents (**Figure 2.8**).

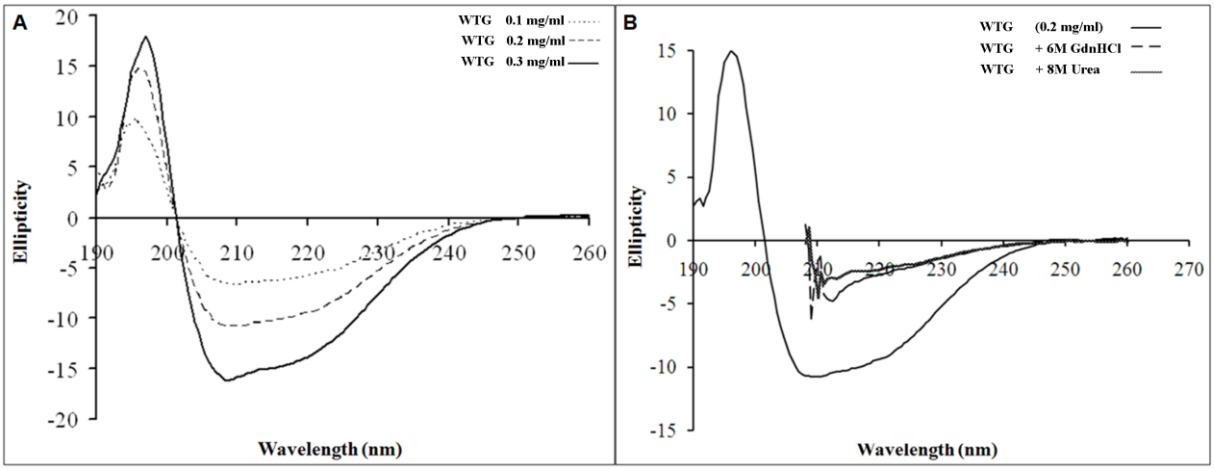


Figure 2.8. The circular dichroism spectra of WTG, **A**. Displays the molar ellipticity at different concentration of protein, **B**. Displays the disruption of secondary structure by applying the high concentration of denaturants urea and GdnHCl.

2.3.5 Fluorescence studies of WTG

The emission has been observed as a result of fluorescence which is produced due to excitation of tryptophan residues at a wavelength of 280 nm. The emission spectra of WTG have been recorded in 20 mM phosphate buffer of pH 7.5. The maximum emission λ_{max} for native protein was observed at 330 nm. To assess the effect of chaotropic reagents, the addition of 1-8 M urea has been done, which results in a redshift with maximum wavelength of λ_{max} 348 nm at the concentration of 8M urea (**Figure 2.9. A**). Furthermore, a redshift of λ_{max} 356 nm was observed for WTG against the 6M concentration of GdnHCl (**Figure 2.9. B**). The red shift values of 18 nm and 26 nm was observed for WTG in presence of 8 M urea and 6M GdnHCl respectively. The native WTG was subjected to pH range of 2 to 13, then emission fluorescence spectra has been acquired (**Figure 2.9. C**). The range of pH 5 to 10 has been observed with minimal variation, in comparison to native protein spectra, but strong shift has been observed at pH less than 5 and more than 10.

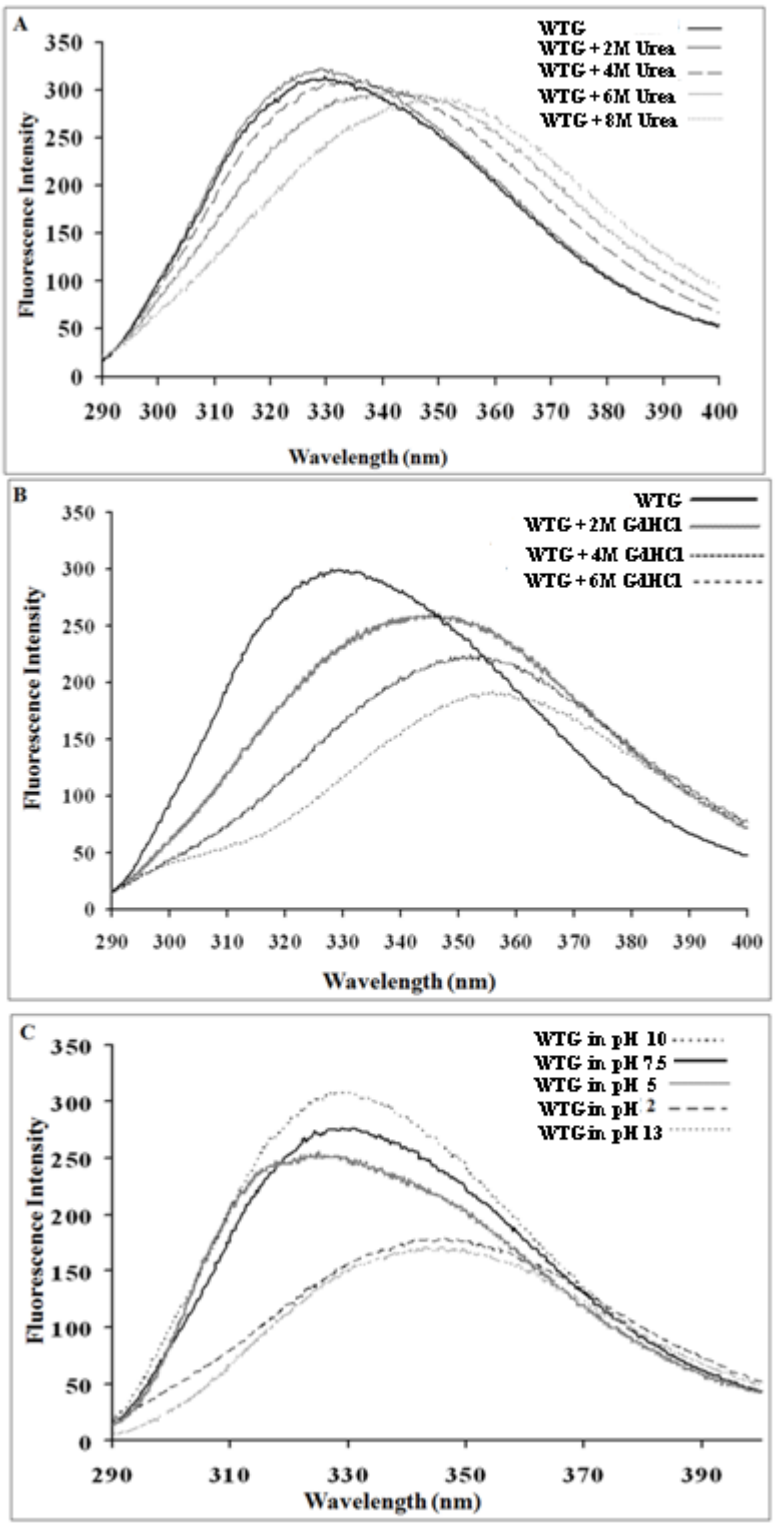


Figure 2.9. The fluorescence spectra observed with WTG against different reagents; **A**. The plot has been drawn for protein with rising concentration of urea, **B**. against GdnHCl and **C**. at different pH conditions.

2.4. Discussion

The two steps of purification were performed to get the pure and homogenous form of WTG. The basic unit of WTG has been observed with a molecular weight of ~ 56 kDa that splits in the presence of reducing agent. It is interesting to observe that the intact monomer of WTG remains almost homogeneous without a profound heterogeneity, but the divided subunits showed significant heterogeneity with a molecular weight range ~ 32-34 kDa for alpha subunit and ~ 22-26 kDa for beta subunit (**Figure 2.2**). The higher molecular weight polypeptide are referred as alpha or acidic subunit and smaller as beta or basis subunit, as reported form oat (Peterson, 1978; Brinegar and Peterson, 1982), pea (R. Casey, 1979) and broad bean (Wright and Boulter, 1974). Presumably, the natural form of WTG is generated by the combination of different size of subunits, but eventually they combine in precise combination to form homogeneous monomer by some unknown mechanism, as appeared from SDS-PAGE analysis. The clear reason for homogeneous monomer of WTG has not been found, as best of our knowledge. Although, the acidic and basic subunits has been reported to impart the heterogeneity has been well studied (Langston-Unkefer and Gade, 1984), furthermore in case of buck wheat a large heterogeneity has been reported due to hetero-oligomeric nature of globulin (Choi and Ma, 2006). Beside this, the WTG exist in oligomeric form, made up of six monomers as observed from gel filtration analysis (**Figure 2.1. B**).

MALDI-TOF based intact mass determinations, further concludes the presence of heterogeneous populations, as observed from the multiple peaks of mass/charge ratio, where and the highest mass/charge ratio likely to be molecular weight of WTG (**Figure 2.3**). Beside this the N-terminal and internal sequencing results confirm that WTG belongs to 11S class of SSPs (**Figure 2.4)**. Further the clear identity and of WTG has been established by cloning and sequencing, which establishes its relation to the 11S class of seed storage proteins (**Figure 2.7**).

The WTG has been observed as a unique example of 11S globulin due to hemagglutinating activity. This activity has been found against different types of erythrocytes. The hemagglutinating activity of WTG is confirmed by formation of lattice against erythrocytes, this may be further correlated with the lectin like activity of WTG. To the best of our knowledge, the 11S globulin showing hemagglutinating activity is first of this kind in case of WTG (Kumar et al., 2013). Apart from this, in some other plants the lectins have been

reported for the fungicidal and insecticidal activity (Chen and Barkley, 1998; Banerjee et al., 2011; Lam and Ng, 2011; Pohleven et al., 2011).

The majority of secondary structure of WTG is constituted by β-sheet and a very few part is made up of α-helices and loops, as revealed by CD analysis. This facts were confirmed by a maximum negative mean residual ellipticity [θ] near 210 nm, which makes a characteristic feature of WTG that is invariantly observed from the family of 11S globulins (Zirwer et al., 1985; Clara Sze et al., 2007). The CD spectra acquired with the high concentration of denaturants (8M urea and 6 M GdnHCl) leads to loss of ellipticity at 210 nm and also impart substantial change in the shape of CD spectra (**Figure 2.8**).

The extent of exposure to the solvent gives a correlation of natively folded protein in particular environment, which can be assessed by measuring a fluorescence spectrum of tryptophan. The sensitive nature of tryptophan is exploited in the fluorescence studies as natural chromophore that undergoes differential solvent exposure (Chen et al., 2009). The native protein was showing emission λ_{max} value of 330 nm (**Figure 2.9**). This value of λ_{max} is a standard point where it has been reported that tryptophan is spanning inside the hydrophobic pocket of natively folded proteins (Dufour et al., 1994). The solvent accessibility was assessed in the presence of denaturants. For WTG only at high concentration of denaturants the significant access of solvent has been observed, further an additional randomization of protein has been observed in presence of 6M GdnHCl compared to 8M urea. The Brazil nut 11S globulin has also showed similar results in presence of urea (Sharma et al., 2010), moreover strong influence of GdnHCl has been speculated due to stronger denaturant ability in comparison to urea (**Figure 2.9. A and B**).

The influence of different pH environment has been observed over WTG, which showed that WTG is structurally quite stable for a wide range of pH i.e. from 5 to 10. However the pH greater than 10 and less than 5 showed a distinct decrease of λ_{max} for emission spectra with a substantial redshift of 18 nm at pH 2 and 13 nm at pH 13. Here it is interesting to find that a shift of 18 nm at pH 2 was similar to the effect imparted by 8M of urea (**Figure 2.9. A and C**).

CHAPTER - 3

CRYSTALLIZATION AND STRUCTURE DETERMINATION OF 11S GLOBULIN FROM THE PLANT *Wrightia tinctoria*

3.1. Introduction

The 11S globulin is one of the significant proteins present in seed and majorly known for providing the nutrient reservoir during seed germination and seedling growth. But they are also responsible for the determination of the nutrient value, quality of seed and some time elicitation of allergy also. The structure of 11S globulin has been known from many sources, but only limited details are available for many key aspects such as, fate of the protein, structural basis for allergenic nature (Palomares et al., 2007), and hormone binding ability, as reported from homologous proteins (Woo et al., 2002). Therefore, it becomes a necessity to reveal the precise structural details that will help in comprehensive understanding of related functions conferred by 11S globulins. To reveal the three dimensional structural details the application of X-ray crystallography is an indispensable biophysical technique, which can elucidate structural details at atomic level (Jeffery, 1971).

The X-ray diffraction based structure elucidation has been employed to gain an insight of basic organization and related function of 11S globulin from *Wrightia tinctoria*. This plant is a member of *Apocynacea* family and is known for several significant medicinal applications. For the structural studies, 11S WTG has been purified from the plant *Wrightia tinctoria* and extraction has been made from the seeds, as a natural source of protein. The WTG has been reported as a member of cupin super family (Kumar et al., 2013).

For structural studies of 11S globulin from *Wrightia tinctoria*, the crystals were raised from the purified protein by vapour diffusion method. The condition for WTG crystallization has been screened by the application of commercial crystal screens, which were further optimized by the combination different pH and concentration of precipitants. Once the good quality diffracting crystals were raised, the diffraction data have been collected at home source. The molecular replacement method has been employed for the phase determination that forms the basis for structure solution. The final model has been validated for basic stereochemistry and the backbone arrangement using the Ramachandran plot (Davis et al., 2007). The refined model has been used for the analysis of structural details as well as for the comparison with the known members of 11S SSP family and other related homologous proteins.

3.2. Materials and Methods

3.2.1. Materials

Sodium citrate tribasic dihydrate, Tris Hydrochloride, PEG 400 were purchased from Sigma-Aldrich Pvt. Ltd. Cryoloops, crystallization screens, crystallization plates were purchased from Hampton Research, USA. The purification columns and gel filtration standard kits were purchased from GE Healthcare, USA. General chemicals were bought from Himedia Chemicals India.

3.2.2. Methods

3.2.2.1. Crystallization of WTG

Crystallization trials were done in the 96 well crystal trays by vapour diffusion method at the temperature of 293 K. A drop of 2 µL were kept in each well with 1:1 ratio of reservoir to protein solution, the protein is having final concentration of 20 mg/ml. Each well of the reservoir contained 100 µL solutions. The initial crystallization conditions were obtained from the Hampton commercial crystal screens (Wooh et al., 2003). Initially the small crystals were obtained from the Crystal Screen 1, in the well having 0.2 M sodium citrate tribasic dihydrate, 100 mM Tris Hydrochloride of pH 8.5 and 30% (v/v) PEG 400. The good quality diffracting crystals were then raised by the optimization of initial hits with varying pH and precipitant concentration.

3.2.2.2. Data Collection and Analysis

Before data collection, the protein crystals were cryoprotected by direct transfer to the mother liquor drop containing 7% glycerol as a cryoprotectant. Diffraction data were collected under cryogenic conditions (100 K). The data were collected over the MAR 345 dtb imaging-plate system using CuKα radiation generated by a Bruker Microstar H rotating-anode

generator, operated at 45 kV and 60 mA equipped with Helios optics. Data were collected with 200 images at a distance of 200 nm with the oscillation of 0.5° per image and time of exposure was kept 5 min. The crystal was diffracted to 1.8 Å resolutions. The initial indexing, integration and scaling of the diffracted images were processed by using program HKL2000 (Otwinowski and Minor, 1997).

3.2.2.3. Structure determination and refinement

The structure solution was done by molecular replacement method by employing program MOLREP (CCP4, 1994; Vagin and Teplyakov, 1997), using poly-Ala model of available 11S globulin structure. Initially the residues were added to the model as guided by the electron density and with the help of multiple sequence alignment of known globulin sequences. Later, the model was refined with the protein sequence according to the obtained cDNA sequence. The best solution was obtained by using Pru du amandin as a search model which is an allergen from *Prunus dulcis* (PDB ID: 3EHK) as the search model. The structure solution was achieved through initial rigid body refinement followed by iterative cycles of restrained atomic parameters refinement using REFMAC and manual electron density fitting using a molecular graphic program COOT (Murshudov et al., 1997; Emsley and Cowtan, 2004). The refined structure has been verified by assessing the stereochemistry using the MolProbity server (Lovell et al., 2003). The solved structure lies in the region of good crystal packing with no clashes of symmetry related molecule, the data reduction and refinement statistics has been summarised in (**Table 3.1)**.

Table 3.1. The data collection and refinement statistics for final refined model of WTG.

Crystallographic data	
Space group	$P2_12_12_1$
Resolution (Å)	1.80
Cell dimensions (Å)	*a* =111.22 , *b* = 114.25, *c* = 202.47
Unique reflections	240037
Completeness (%) (Last shell)	95.3 (71.1)
R_{sym}(%) (Last Shell)	3.7 (31.2)
I/σ (Last shell)	18.0 (2.0)
Multiplicity (Last shell)	2.3 (1.8)
No. of Residues	2337
Water Molecules	788
Resolution range (Å)	28.06-1.80
R_{cryst} (%)	20.5
R_{free} (%)	24.8
Average B-factor (Å^2)	A 20.23
	B 23.382
	C 23.24
	D 22.15
	E 22.02
	F 22.21
Water molecule	32.92
All molecule	22.60
R.M.S. deviations on bond length (Å)	0.0047
R.m.s. deviations bond angles (Å)	1.147
Ramachandran plot (%)	
Favored	99.9
Outlier	0.1

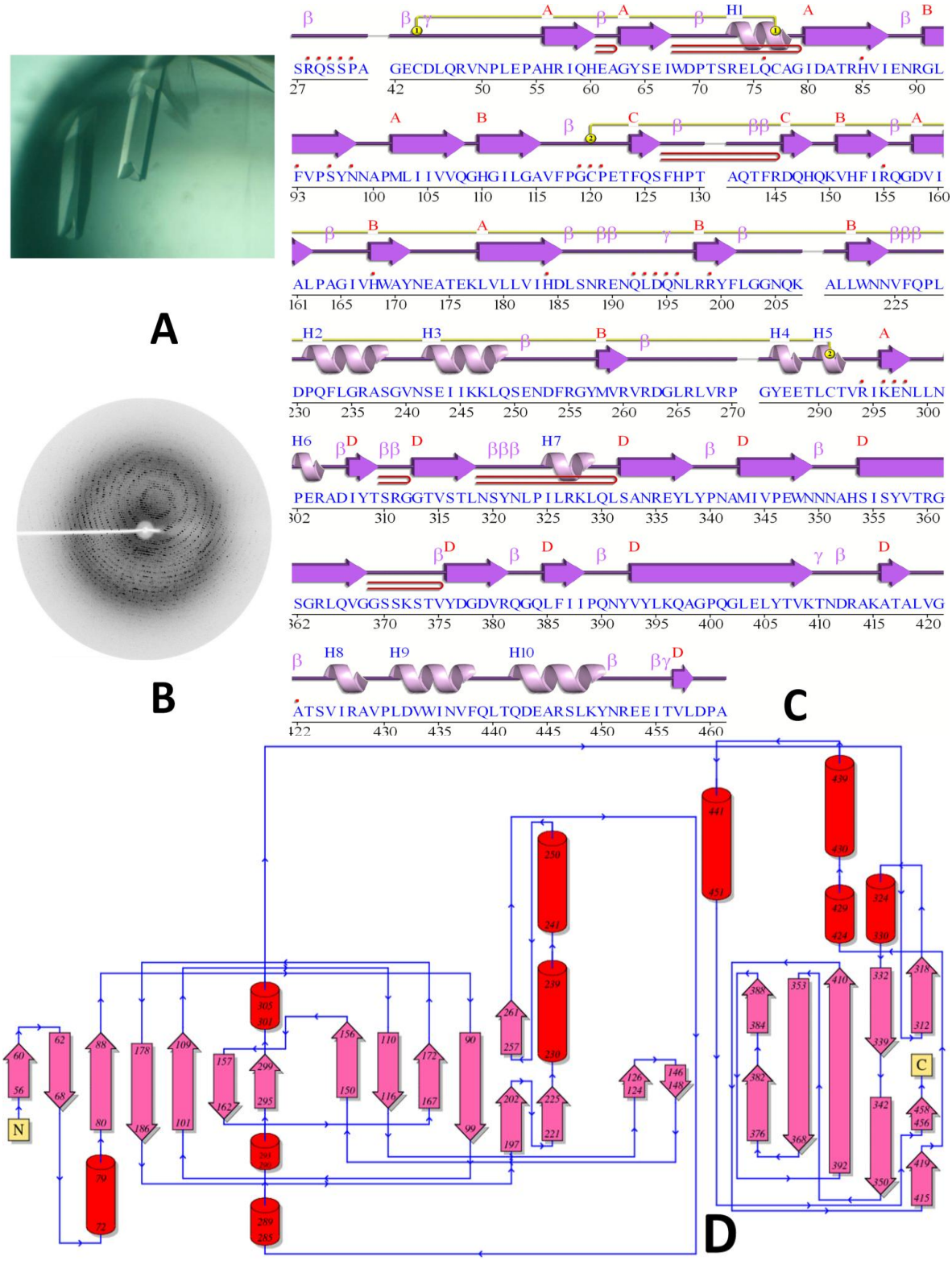


Figure 3.1. The figure shows; **A.** crystals of WTG with a dimension range of 120 - 160 µm, **B.** diffraction pattern observed at 1.8 Å resolution, **C.** secondary structure assignment for the monomer. **D.** the topology diagram depicting chain A of WTG.

3.3. Results

3.3.1. Three-dimensional structure of WTG

3.3.1.1 Quality of WTG Structure

The X-ray diffraction pattern of WTG crystals has been obtained from native crystal at maximum resolutions of 1.8 Å (**Figure 3.1**). The final structure was refined with the values of R_{cryst} (%) 20.5 and R_{free} (%) 24.8. The information for the data collection and refinement has been summarized in **Table 3.1**. The crystal of WTG has six molecules in an asymmetric unit with the space group $P2_12_12_1$. There are four disordered regions present in each monomer i.e. residues 33-42, 130-141, 207-219 and 270-285. The residues lying in the disordered regions were not resolved in the electron density map, probably due to their position at the surface. The Ramachandran plot depicts that 99.9% residues are lying in favoured region and 0.1% in outlier (**Table 3.1**)

3.3.1.2. Overall structure of WTG

The overall structure of WTG is organized in two stacked layer of trimers (**Figure 3.2**). The trimers are facing opposite to each other and the monomer of each trimer is organized in reverse orientation from N-terminal to C-terminal (**Figure 3.2. B**). Overall superposition of two trimers ABC over DEF shows a RMSD value of 0.12 Å with 1014 equivalent C_α-atoms, while average RMSD for each monomer with respect to chain A is 0.12 Å. These observations advocate that all the monomers are essentially identical in the WTG hexamer. Therefore, the chain A has been considered as a monomeric unit for structure analysis and representation unless otherwise mentioned.

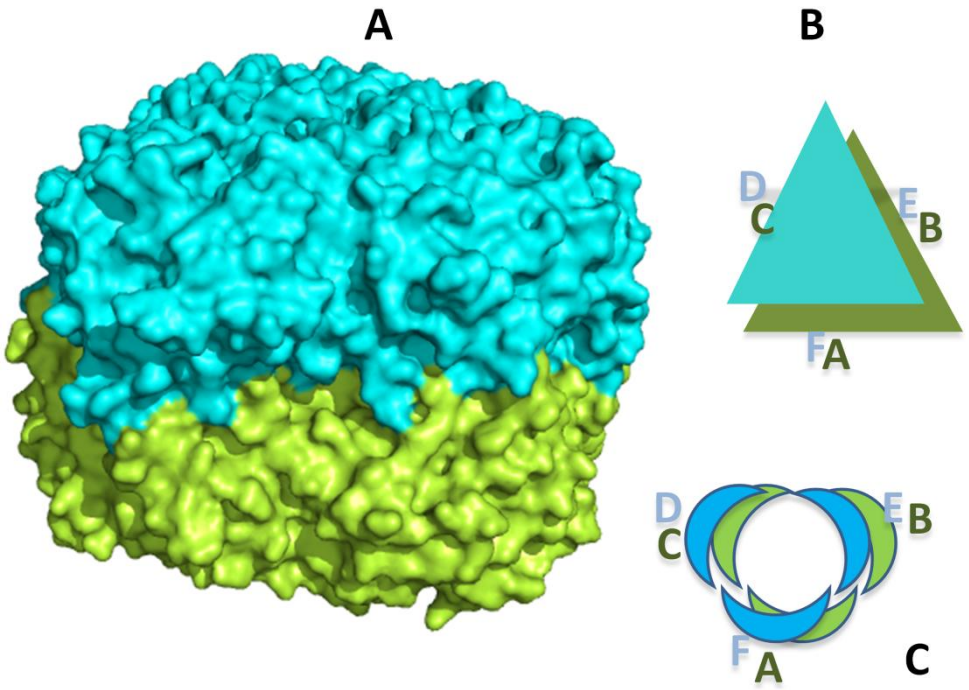


Figure 3.2. The hexameric assembly of 11S globulin; **A.** Surface view of two stacked layer of trimers, **B.** A depiction of dimer of trimers, **C.** A depiction of trimer of dimer.

Each monomer of WTG encompasses ten helices, out of them four helices constitutes the characteristic extended helical domains; particularly the N-terminal domain contains 2nd and 3rd helices and C-terminal domain contains 9th and 10th helices (**Figure 3.3**). The extended helical domains are involved in a head to tail interaction, which assists in the formation of trimeric assembly. The N- and C-terminal helical domains exist from the residue Pro-231 to Gln-249 and Leu-431 to Tyr-450 respectively. There are 26 beta strands present in each monomer and most of them are involved in formation of highly conserved double sheeted cupin barrel, a characteristic feature of cupin super family. In each monomer, there are two cupin folds independently spanning over N- and C-terminals (**Figure 3.3**).

The alpha subunit encompasses 284 residues while beta subunit remains short in length with 198 residues. To assess the overall ionic profile of subunits at neutral pH, the isoelectric point (PI) has been deduced by using bioinformatics tool Protparam accessed via Expasy bioinformatics resource portal (Gasteiger et al., 2005).

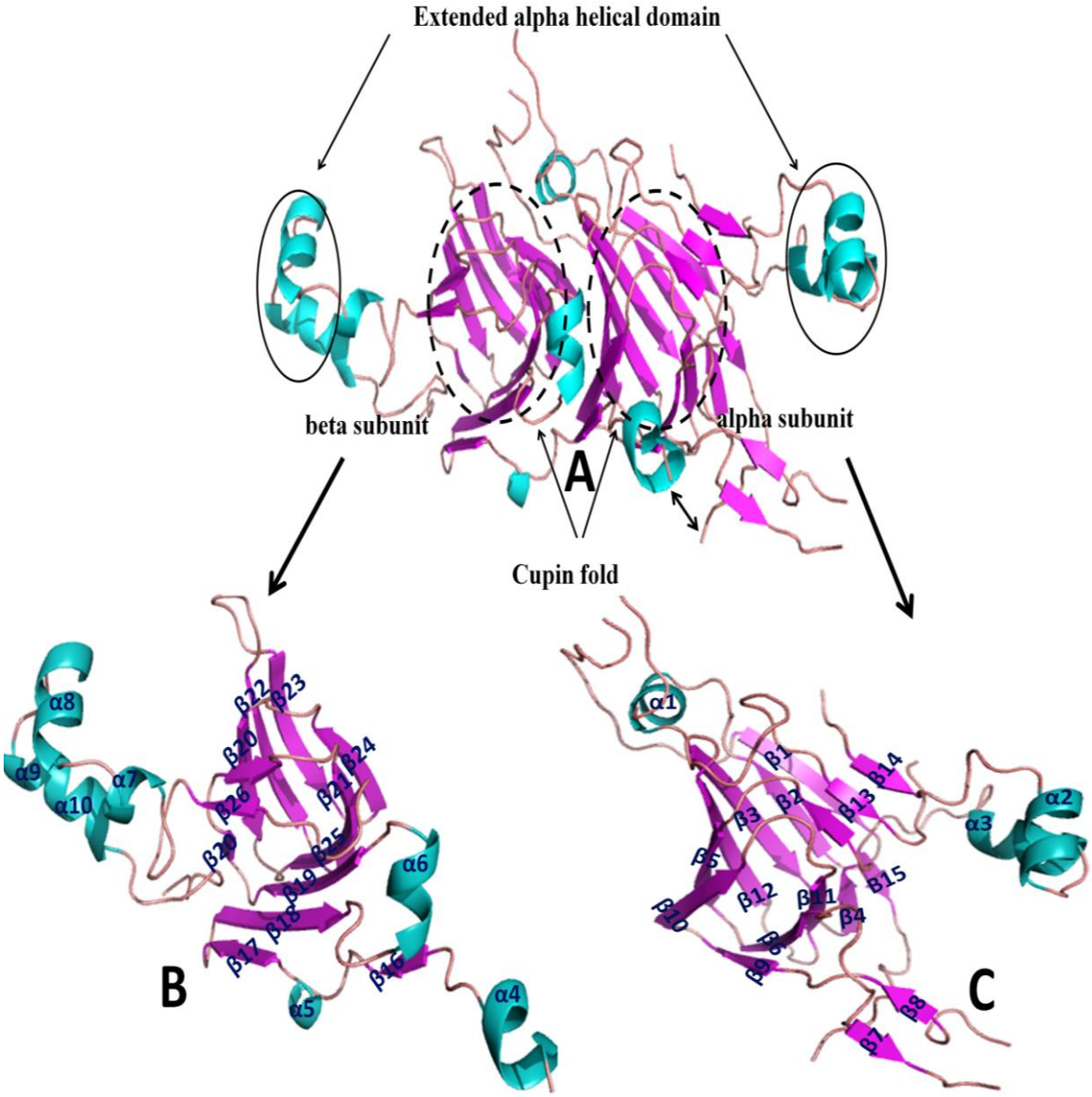


Figure 3.3. **A.** The monomer of WTG is displaying extended alpha helical domains (solid rings) and cupin folds (dotted circles), with the cleavage region of two subunits (double headed arrow) **B.** A smaller beta subunit of monomer is having one cupin fold and one extended alpha helical domain **C.** The larger alpha subunit is having extra long N-terminal stretch along with alpha helical domain and cupin fold.

The value of PI for alpha and beta subunit were 6.16 and 8.59 respectively, obtained by providing a customize sequence information for individual subunits.

The alpha and beta subunits of a monomer are connected by disulphide bond involving Cys-120 and Cys-291. Moreover, an additional disulphide bond has been observed in alpha subunit involving the Cys-44 and Cys-77, which connects the flexible loop of N-terminal to the 1st alpha helix lying in the vicinity (**Figure 3.4. A**). Both the disulphide bonds can be symmetrically positioned along a pseudo-dyad axis (**Figure 3.4. B**). The inter subunit disulphide bond is present at a key position that exhibit interactions with nearby chains, therefore this face of trimer has been termed as inter chain face (IE). While the intra subunit disulphide bond does not face any chain other than itself, has been termed as intra chain face (IA) of the monomer (**Figure 3.4. C**). The name of disulphide bonds spanning at different faces has been adopted form Glycinin (Adachi et al., 2001).

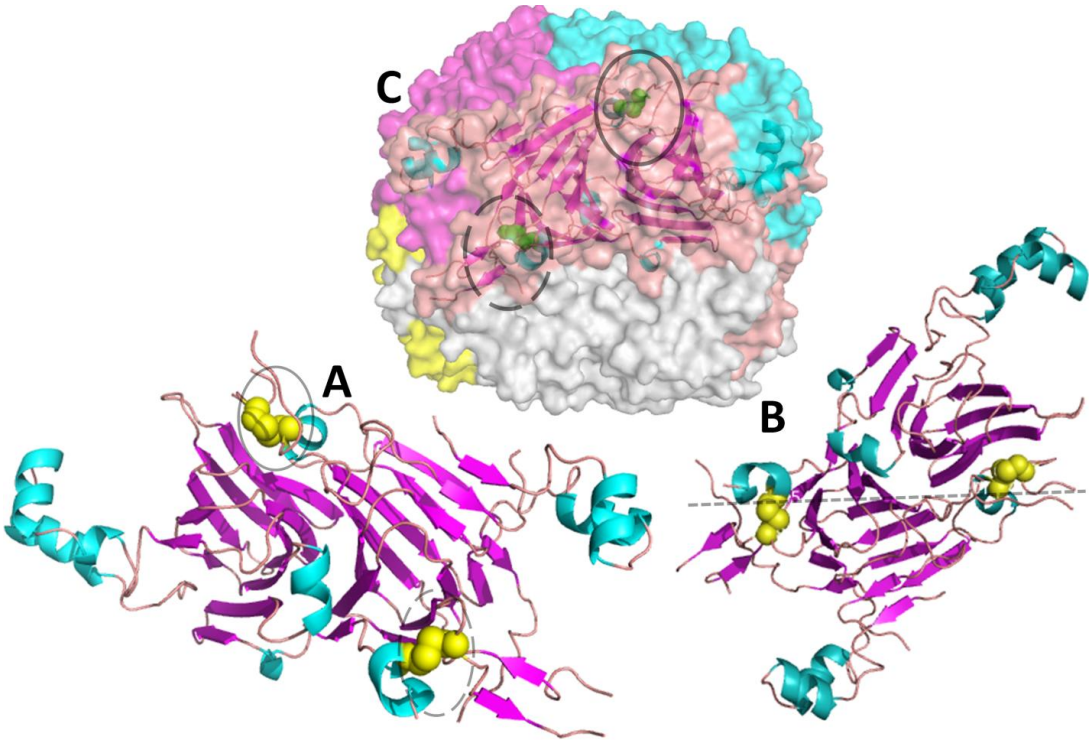


Figure 3.4. **A.** The two disulphide bonds present in monomer of WTG has been shown in yellow coloured spheres. **B.** An imaginary pseudo dyad axis has been shown along the disulphide bond **C.** A hexamer view of WTG surface displaying cartoon of chain A with inter chain interface spanning disulphide bond (dotted circle) and intra chain interface disulphide bond (solid circle).

The monomers of WTG interact together by involving different kinds of interactions; out of them the Salt bridge formation between the monomers imparts a significant interaction as observed from trimeric assembly. Such interactions have been uniformly observed between the monomers related by three fold axis of symmetry. An example of this interaction has been represented between Asp-432 of chain-A, spans over the extended alpha helical domain and Arg-266 of chain B, present on flexible loop at N-terminal end of alpha subunit (**Figure 3.5**).

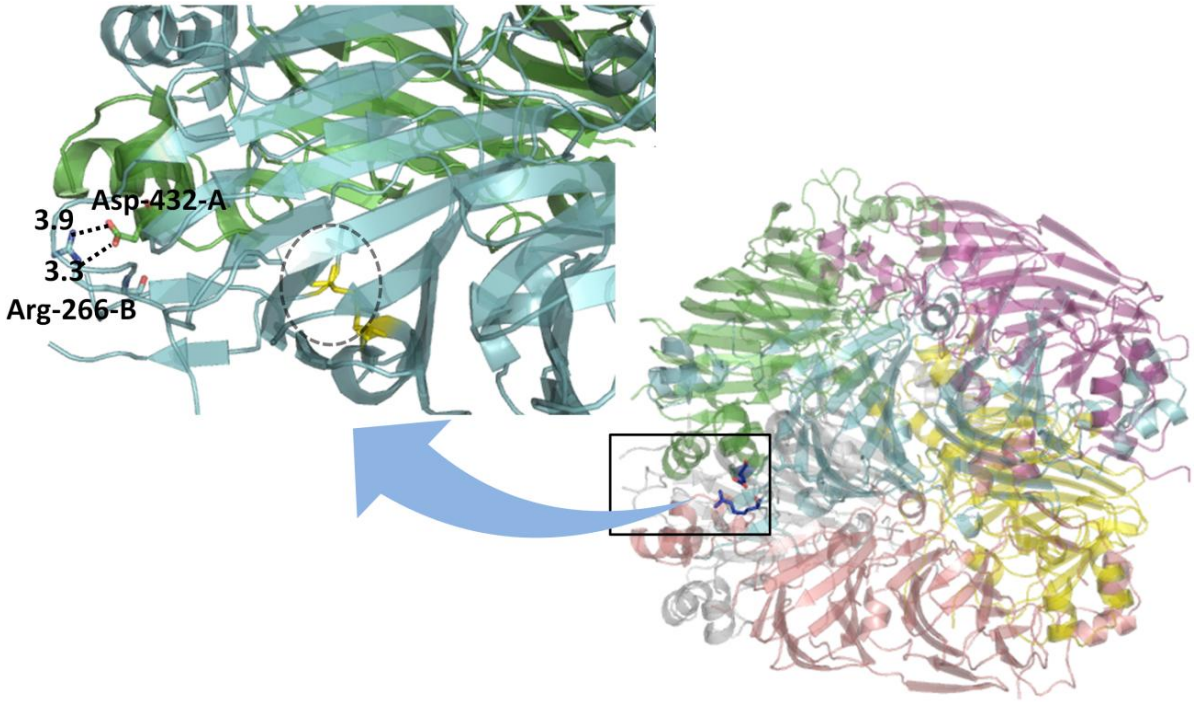


Figure 3.5. The hexamer of WTG showing a salt bridge formation between extended helical domain of chain A (green) and a loop region of chain B (cyan) as well as a disulphide bond in vicinity (dotted circle).

The N-terminal end of each chain in the trimers involves a floppy region between 34-41residue, along the three fold axis of symmetry. These N-terminal ends of first layer of trimer and second layer of trimer are lying at two opposite poles of hexamer and share together a threefold axis of symmetry (**Figure 3.6**).

The structure superposition has been employed to extrapolate the relative profile of WTG (**Table 3.2**). The comparison between WTG and other members of SSPs reveals that it encompasses a longer floppy region at N-terminal. Furthermore, the initial 60 residues have been observed as a highly dynamic stretch, due to their unique position and amino acid profile.

Table 3.2. Structural comparison between WTG and related member of seed protein family.

Protein	**RMSD**	**Number of atoms**
11S SSP		
Procruciferin PDB ID: 3KGL	0.89	328
Glycinin PDB ID: 1OD5	0.80	325
7S SSP		
Canavalin PDB ID: 1CAU	2.5	230
Phaseolin PDB ID: 2PHL	4.5	202

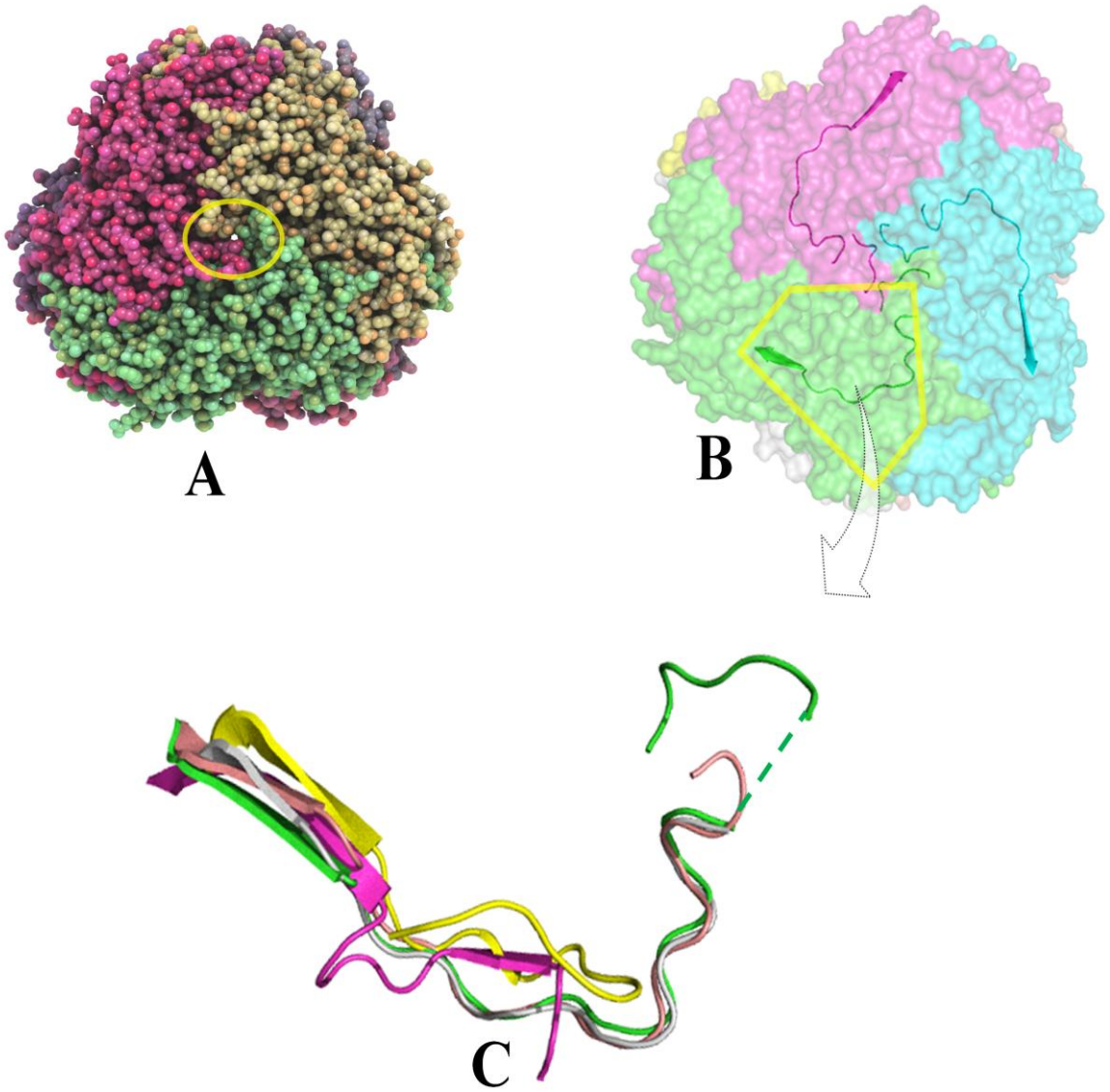


Figure 3.6. **A.** Space fill model depicting centre pole (yellow circle). **B.** The N-terminal floppy and highly hydrophilic region showing 60 residues of the trimer from N-terminal (yellow marked area). **C.** The N-terminal residues displaying dynamic region of WTG (in green) with related members of SSPs (Phaseolin PDB ID: 2PHL in pink, Canavalin PDB ID: 1CAU in yellow, Procruciferin PDB ID: 3KGL in orange, Glycinin PDB ID: 1OD5 in gray colour), green dotted line represents a mobile disordered region between 34th to 41st amino acid a in electron density.

3.3.1.3. WTG binding to auxin

The three dimensional structure of WTG obtained from native protein source has shown that auxin forms a complex with 11S of globulin obtained from *W. tinctoria*. The three molecules of auxin were found in the crystal structure of native WTG protein. The pockets of WTG are related by threefold axis of symmetry and each pocket uniformly encompasses a molecule of auxin (**Figure 3.7**). These pockets are constituted by four nearby chains. The non-covalent interactions hold auxin inside the pocket. If we consider a pocket constituted by the chain B, C, D and E, then the carboxyl end of auxin is stabilized by electrostatic interactions involving Arg-427 (chain D) and Glu-122 (chain E). The aromatic ring of auxin has been stabilized by stacking interactions between Pro-121 (chain C) and Pro-121 (chain E), which intercalates the auxin. Such interactions are symmetrically observed between all chains of WTG clasping the auxin invariantly (**Figure 3.8**).

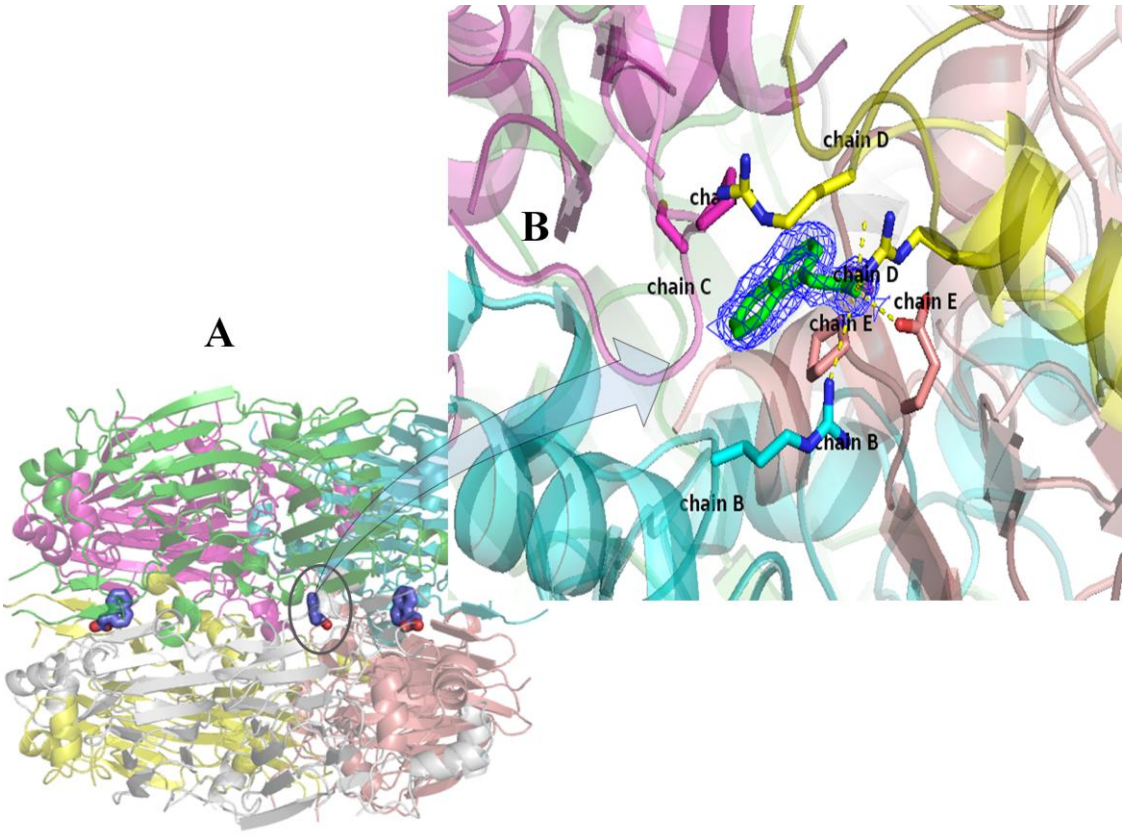


Figure 3.7. The structure is representing WTG hexamer in complex with auxin. **A**. The three symmetrical position of the auxin present inside the three pockets. **B**. the auxin molecule has been shown at the junction of four chains with electron density map contoured at 1 σ.

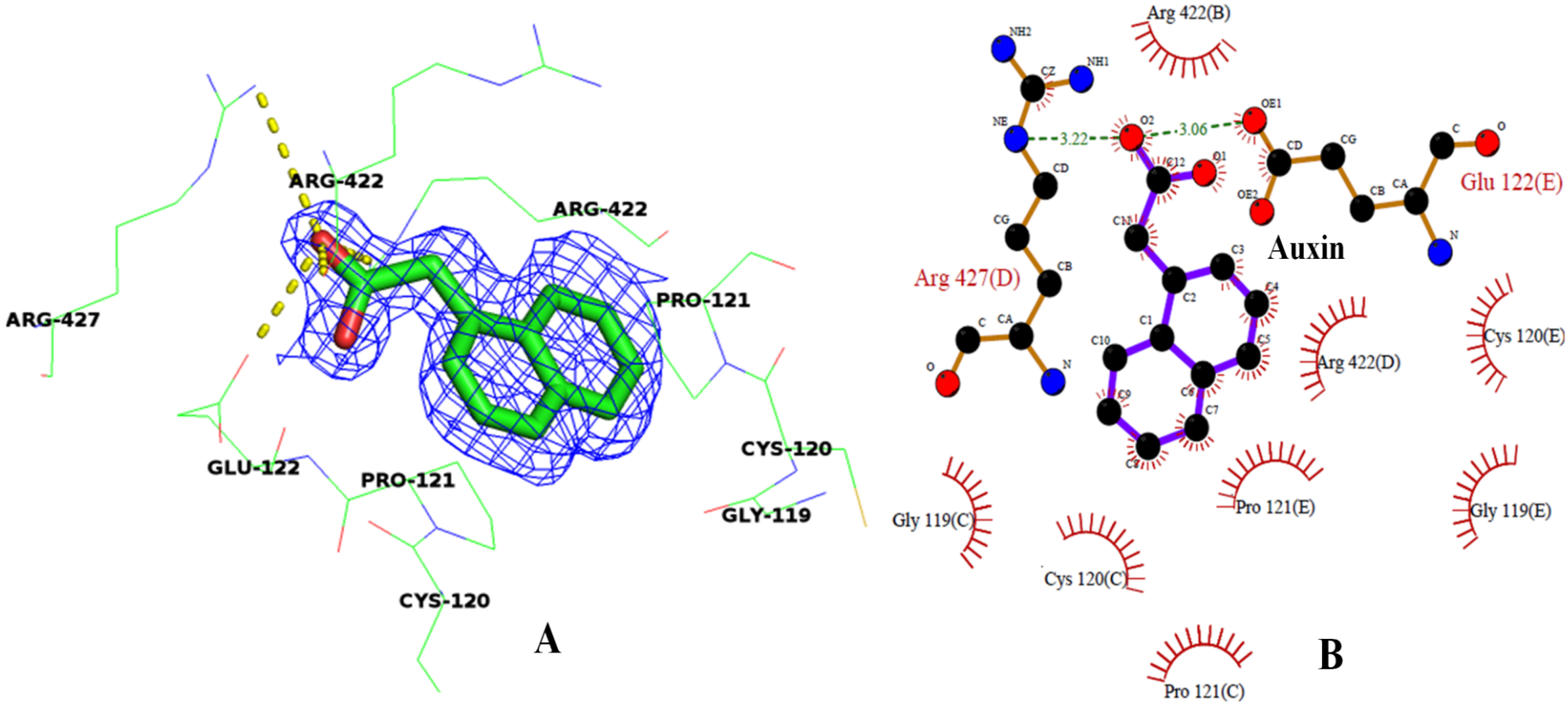


Figure 3.8. The interactions of auxin; **A**. The amino acid residues involved in auxin interaction has been depicted from a pocket of three dimensional structure of WTG (electron density may around auxin contour at 1 σ). **B**. The four chains and corresponding residues clasping the auxin in pocket.

3.4. Discussion

3.4.1. Overall structure of WTG

With the aim to get insight of 11S globulin structure and possible ligand in natural state, the WTG has been purified from the seed of *Wrightia tinctoria* as a natural source. The crystal structure reveals a basic arrangement WTG, which involves two stacking layer of trimers and the dimer of trimer like organization (**Figure 3.1**). The monomer interacts to each other in a triangular manner, with head to tail association along three fold axis of symmetry. The other view, which is a dimer of trimer, displays twofold symmetry along stacking layer. The monomers of WTG encompass conserved cupin folds and extended alpha helical domains; these extended helical domains and cupin folds together to constitute the signature scaffold of 11S globulin. These signature structures reveal that WTG is a type of seed storage protein, which is a member of cupin super family. (**Figure 3.3**). Further, in accordance with earlier reports, the monomers are predominantly composed of beta strands followed by alpha helices (Tandang-Silvas et al., 2010).

As a general method of processing, each monomer of WTG is synthesized in a precursor polypeptide that undergoes post translational processing of signal sequence inside the endoplasmic reticulum. The cleavage of N-terminal signal peptide is an intermediate stage of processing that leads to the formation of pro-protein. Subsequently, the pro-protein undergoes post translational cleavage by asparaginyl endopeptidase, between the conserved residues Asn-284 and Gly-285. This processing leads to formation of alpha subunit at N-terminal and beta subunit at C-terminal region connected by disulphide bond. When the theoretical isoelectric point of alpha and beta subunit has been predicted individually, the values of alpha and beta subunits were found as 6.16 and the 8.59 respectively. These unique values of PI become a basis of characterization for two subunits as acidic and basic domain uniformly present in each monomer. These observations are supported by the earlier findings regarding the processing of 11S globulin and their isoelectric points (Jung et al., 1998; Adachi et al., 2001).

3.4.2. Hexameric structure of WTG

The hexamer is made up of two stacking layers of trimers facing opposite to each other and also lying in reverse orientation. Such arrangements allow co-operative interaction, which accommodate the protruded as well as depressed surface regions over each other (**Figure 3.2. A**). Both the faces of trimers can be designated on the basis of position of disulphide bonds spanning over a monomer, as inter chain disulphide containing face (IE) and intra chain disulphide containing face (IA). The IE face and IA face are lying perpendicular to the three fold axis of symmetry. The monomers of WTG hexamer are invariantly similar among themselves. A monomer of WTG encompasses four disordered regions between 33-42, 130-141, 207-219 and 270-285 residues. The last mobile disorder stretch (270-285) has been observed as a crucial region because, i) it encompasses conserved amino acid Asn-284 and Gly-285 that act as recognition site for the proteolytic enzyme 'asparaginyl endopeptidase'(Jung et al., 1998) , ii) it facilitates efficient packing by providing a mobile region of 14 amino acid residues between Pro-270 and Gly-285 at the IE face and iii) it imparts a mobility of extended helical domain at alpha subunit that assists in formation of heat to tail association between the trimers, which eventually assemble to WTG hexamer (**Figure 3.9**).

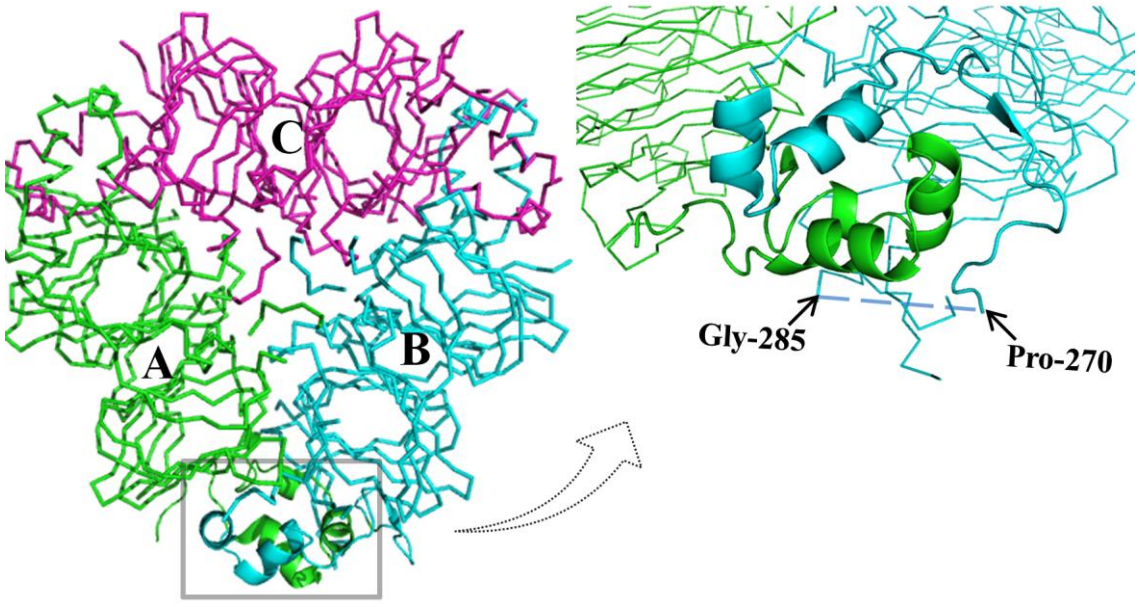


Figure 3.9. The trimer made by chain ABC has been shown, the inset emphasize an extended alpha helical region between chain A and B exhibiting mobile random region between Gly-285 and Pro-270.

Along three fold axis of WTG, the N-terminal end of three chains are uniformly funnelled together to constitute a limited ordered region of seven residues followed by mobile disorder region of eight residues that subsequently merge in cupin fold **(Figure 3.6**). It appears that if the N-terminal loops are absent, the centre pore like structure remains open as observed from Canavalin (Ko et al., 1993) and Phaseolin (Lawrence et al., 1994) like structures. Contrary to this, WTG has been observed with longer N-terminal floppy loop that is followed by mobile disordered region. It has been observed that the N-terminal loop appeared in electron density are started just after the signal peptide cleavage site, as assessed by *in silico* predictions and analysis (Petersen et al., 2011). These observations ascertain that possibly there is no disorder region present before the N-terminal loop appeared in the model. Therefore, just after density of seven N-terminal residues, there is only one disordered region of eight residues. It is interesting to find that together these fifteen residues constitute a significantly hydrophilic stretch. This stretch further becomes dynamic due to the presence of two Serines (Ser-31 and 34). Interestingly only density for Ser-31 was appeared in the loop, while Ser-34 remains in disordered region. When the *in silico* prediction for phosphorylation of these two residues has been done, a significant score has been obtained (**Figure 3.10. A**) (Blom et al., 1999). Additionally, a reasonably high possibility of disorder has been noticed from N-terminal stretch (**Figure 3.10. B**), which also concludes the dynamic nature of the region. This property has been invariantly observed from N-terminal loops of all the six chains of WTG. Since the hexamer encompasses two stacking layer of trimers, therefore N-terminal dynamic regions at two ends constitute an opposite poles like centre. Moreover, the centre of the poles is present exactly at the three fold axis of symmetry. Since symmetrical and dynamic assembly at poles are possibly forming a set of interactions susceptible for phosphorylation and rearrangement of interactions, therefore an influence of slight change of pH (as reported by Adachi et al., 2003) or ion balance during germination may led to alteration of opening of pore and allow the disintegration of hexamer (**Figure 3.6**) (Adachi et al., 2003).

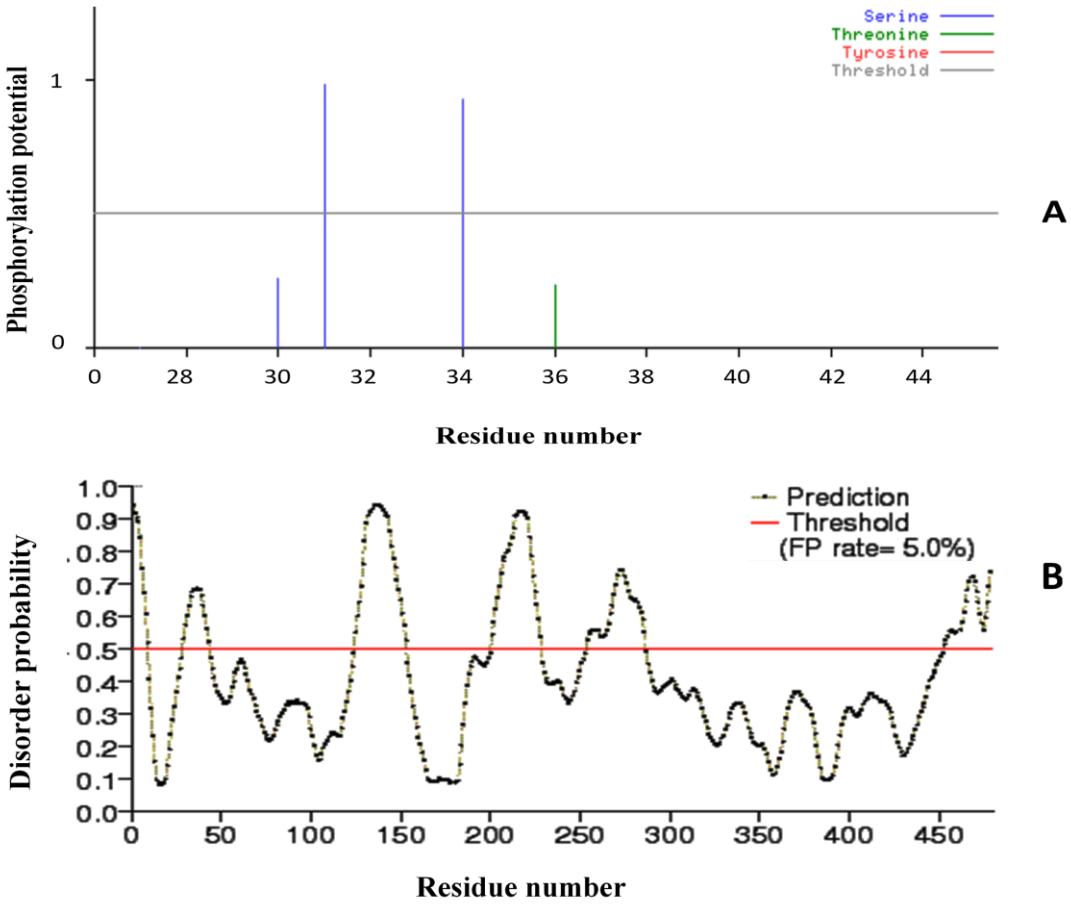


Figure 3.10. **A.** *In silico* prediction of phosphorylation potential of initial N-terminal residues involved in formation of floppy region (X- axis corresponds to amino acid position and Y- axis represents phosphorylation potential. **B.** The disorder probability prediction with respect to the position of amino acid residues (X-axis corresponds to residue number and Y-axis displays probability score for disorder.

3.4.3. Stacking of two trimers

The WTG hexamer can be divided in three equal parts along the three fold axis of symmetry and each part encompasses a dimer. The pair of chains involved in formation of three equal parts is constituted by A-F, B-E and C-D. An additional two fold symmetry further separates the hexamer in two stacking layers of trimers, made by the chains ABC and DEF. The numbers of residues involved in formation of contacts at the interface of dimers are generally 45 except the chain A and C that interacts by only 43 residues. The total area of interface

provided by each monomer showed very limited variation range with minimum 2366 $Å^2$ for chain C and maximum 2433 $Å^2$ for chain E (**Table 3.3**).

Table 3.3. The interaction profile between three dimers has been shown that lead to formation of hexamer.

Dimer formed between	**Number of residues involved**	**Number of hydrogen bonds**	**Number of non bonded contacts**	**Area of interface ($Å^2$)**
A-F	43 and 45	28	320	2410 and 2400
B-E	45 and 45	28	314	2414 and 2433
C-D	43 and 45	25	320	2366 and 2375

The WTG interactions can also be divided in two pairs of trimer lying at the two fold axis of symmetry, formed between the groups of chains ABC and DEF. The extended helical region is known as a characteristic feature of globulin/vicilin class of SSPs and invariantly marks its presence in all the members, this helical region also acts as a major part of WTG structure by having substantial extent of hydrogen bonds and non bonded contacts. In each pair of trimers, the head to tail interaction region encompasses significant interactions confer the stability to the trimer (**Table 3.4**). Beside this, The WTG consist salt bridge interactions, where each monomer linked by head to tail association (**Table 3.5**). The salt bridge interactions confer structural integrity, which exists in between the trimer.

Table 3.4. The list of interacting residue of each trimer and the area of interface.

Trimer pair between chain	Number of residues at interface	Number of hydrogen bonds	Number of non bonded contacts	Interface area ($Å^2$)
A-B	57 and 51	34	371	2723 and 2824
B-C	50 and 48	33	355	2558 and 2622
A-C	52 and 48	31	345	2638 and 2600
D-E	50 and 49	31	336	2539 and 2621
E-F	47 and 48	32	348	255 and 2591
D-F	47 and 50	29	354	2643 and 2558

Table 3.5. The list of residues involved in salt bridge formation.

Trimer	Residues involved
ABC	Asp-432 (Chain A) - Arg-266 (Chain B) Asp-432 (Chain B) - Arg-266(Chain C) Asp-432 (Chain C) - Arg-266 (Chain A)
DEF	Asp-432 (Chain D) - Arg-266 (Chain E) Asp-432 (Chain E) - Arg-266(Chain F) Asp-432 (Chain F) - Arg-266 (Chain D)

3.4.4 General comparisons of WTG with homologous proteins

The WTG shares structure and sequence identity with the several classes of proteins mainly by virtue of beta barrel domain that comes under cupin super family. The structural similarity encompasses prokaryotic as well as eukaryotic proteins involved majorly in sugar and other compound binding (Dunwell, 1998). The seed storage proteins are among one of the largest example of this super family. WTG also shares great similarity with allergens that characterizes WTG also as a possible allergen like protein. This conclusion has been drawn on the basis of resemblance to major studied allergens such as Pis v 2.0201 from *Pistacia vera*, Pru du amandin from *Prunus dulcis*, Ara h3 from *Arachis hypogea*, Ana o 2 from *Anacardium occidentale* and allergenic protein from *Fagopyrum tataricum* (**Table 3.6**). WTG has been observed with similar amino acid residues present in known allergens, which is reported as potent antigen due to binding of immunoglobulin-E (IgE) (Asero et al., 2007). The comparison with these allergens suggests that both linear and conformational antigenic determinants are present over WTG, as reported from other 11S allergen globulins (Robotham et al., 2009). The peanut allergen Ara h3 has been mapped for linear epitope using 15 mer overlapping peptides against the different patients (Rabjohn et al., 1999). The sequence alignment of WTG to the Ara h3 suggests the remarkable existence of these epitope over WTG also. These findings suggest that few specific stretches of WTG can also elicit the allergic response due to IgE binding (**Figure 3.11**).

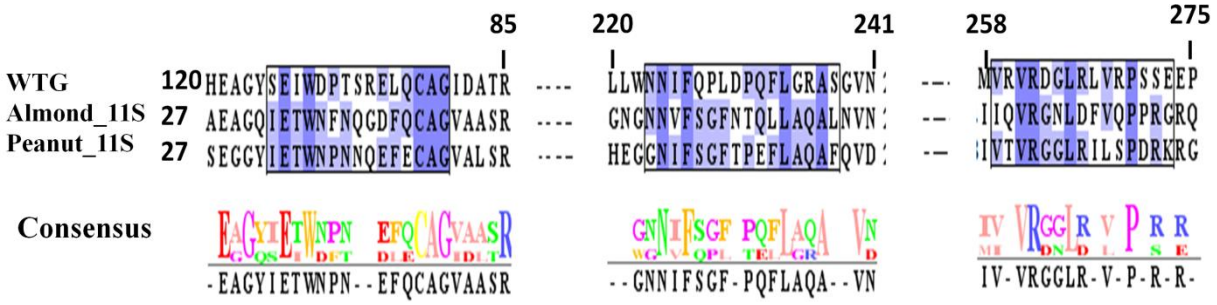


Figure 3.11. The comparison of WTG with the almond and peanut allergen depicts a consensus for the epitopic determinants.

Table 3.6. A list of selected allergens is showing the structural similarity (a 3D structure comparison server by Holm et. al, 1995).

Plant source	Name of allergen	Similarity
Pistacia vera	Pis v 2.0201	46%,
Arachis hypogea	Ara h3	41%
Prunus dulci	Pru du amandin	33%
Anacardium occidentale	Ana o 2	40%
Fagopyrum tataricum	Globulin	39%

3.4.5. Structure specific comparisons of WTG with related proteins

The WTG most closely resemble with the allergens, Ara h3 from *Arachis hypogea* and prunin from *Prunus dulcis*. The prunin and Ara h3 shares the similar loop structure at the N-terminal flexible region, but prunin remains shorter in length at the N-terminal flexible region. The Ara h3 and WTG remains almost same for the N-terminal flexible region. The significant allergenic profile of WTG has been observed on the basis of structure and sequence similarity with the Ara h3 binding. There are four IgE binding epitopes has been characterized over Ara h3 (Rabjohn et al., 1999). The Ara h3 has been observed to share significant similarity over all the four epitope determinants with WTG. Due to the limitation of available region of model in electron density map, out of the four regions only 1st and 3rd has been appeared. Surprisingly both the model structure displays random disordered region for second and third determinants (**Figure 3.12**).

The first epitopic region between strand B and C of Ara h3 is characterized as epitope 1 for IgE binding. The WTG shares significant structural and sequence similarity at this region (**Figure 3.12. A**). The second epitopic determinant has neither appeared in Ara h3 available structure (PDB ID: 3C3V) nor in WTG. The third region has been reported as most potent site for IgE binding (Rabjohn et al., 1999). The WTG displays maximum structural and sequence similarity with third region of epitopic determinant. The forth and last region remains obscured over WTG as well as at Ara h3. These observations ascertain that WTG can also display the allergenic profile similar to Ara h3 like allergen from peanut.

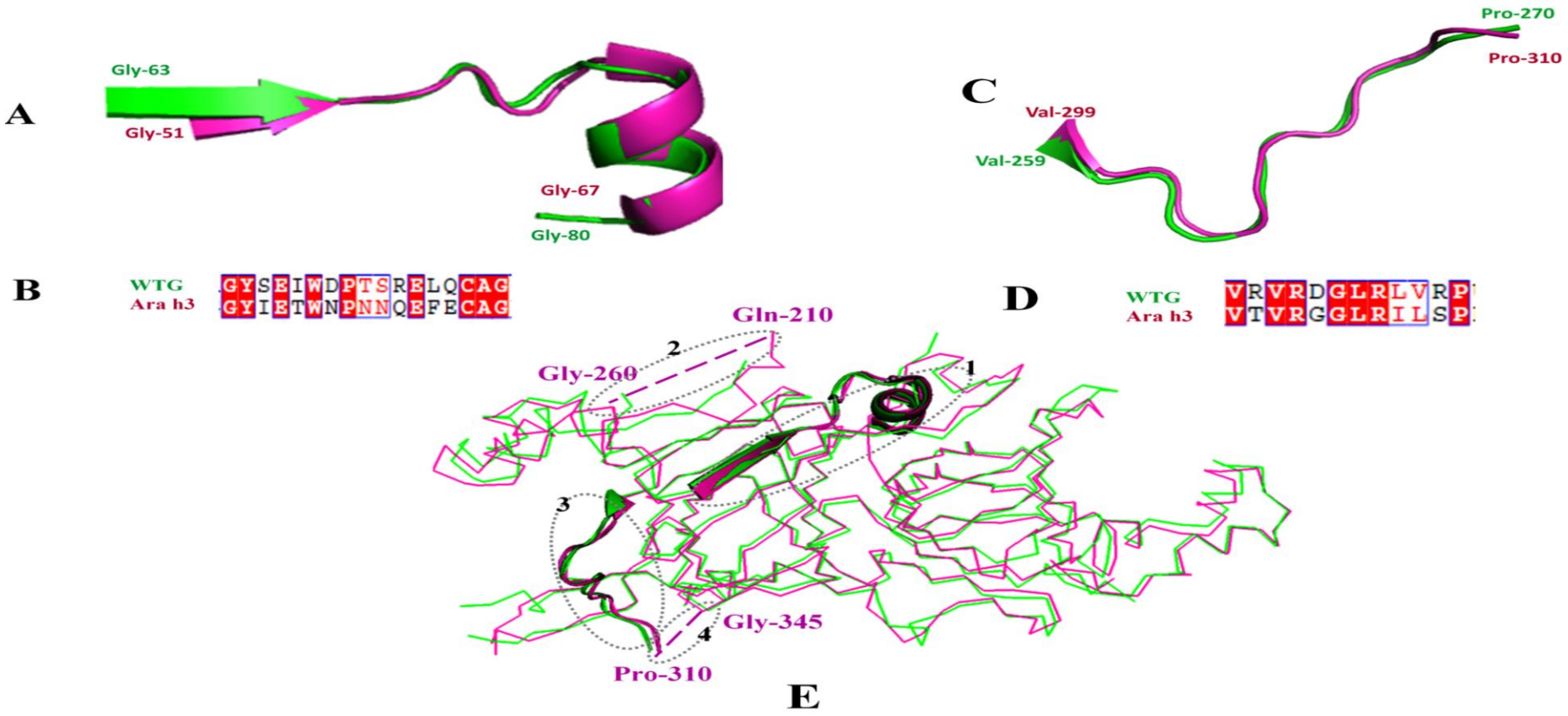


Figure 3.12. Structure and sequence similarity profile between WTG (in green) and allergen Ara h3 (in pink), **A.** The structural similarity of WTG and Ara h3 has been shown by superposition of 1st epitope region of Ara h3 and with β1-strand and first helix of WTG. **B.** Depicts a sequence similarity of the superposed region. **C.** WTG superposition with third epitope region of Ara h3. **D.** The amino alignment is shown between WTG and Ara h3 residues spanning over 3rd epitope region. **E.** The overall superposition between WTG and Ara h3 is depicting the four epitopic determinants (circled area 1, 2, 3 and 4).

The WTG is share remarkable similarity with other proteins, such as proglobulin from *Amaranthus hypochondriacus*, prolegumin from *Pisum sativum*, glycinine from *Glycine max*, procruciferin from *Brassica napus* and pumpkin 11S globulin. Despite the similarities, the WTG is unique for its initial N-terminal region, rarely observed from other known members of the family. The germin and auxin binding proteins (ABP1) are the exception, which encompasses WTG like similar N-terminal floppy region. The structure of WTG has been characterized as a storage protein but, some distantly related functional protein also share key features. Germin is one of the most important functional members of this family. Unlike other member of this family the germins contains single cupin fold. The superposition with WTG reveals that cupin domain present in beta subunit is closer to germin than alpha subunit. The superposition with beta subunit of WTG has been considered for structural comparison with germin, which shares 19% sequence identity as well as a reasonable structural alignment root mean square deviation of 2.1 Å. The N-terminal region of both the proteins is floppy but the germin possess small organised helical region of five residues from Pro 28 to Ala 32 whereas WTG is having entire loop region. As a closet functional relative of WTG, the germin also encompasses similar arrangement of two beta sheets that forms a typical cupin fold. The key residues responsible for enzymatic functions as well as for evolutionary conservation of germin and germin like proteins present in two conserved signature domain, first contains 21 residue stretch as G(X)5HXH(X)11G, while second contains 16 residue stretch as G(X)5P(X)4H(X)3N (where X is any amino acid residue) (Dunwell et al., 2000). Similar to germin, the WTG consist first stretch between strands β-20 and β-22 and second stretch between β-25 and β-26 (**Figure 3.13**). It appears that during the course of evolution, several residues are altered from these signature stretches. Among different mutated residues, Histidine appears as of special important due to active involvement in reaction and ligand interaction. Since the Histidine amino acid residues has been noticed for the metal binding as well as substrate interaction from functional member of cupin family, therefore its mutation appears as one of the major cause for non functioning of SSPs (**Figure 3.13. B**). Despite of the difference at residue level, the overall structure remains reasonably preserved (**Figure 3.13. C**).

The comparison of WTG with one of the significant member of cupin family Auxin binding protein (ABP1) reveals that overall sequence identity is 17% while the RMSD is 2.9 Å. The ABP1 is also a functional member of cupin super family. The Histidine cluster enables this protein to form a Zn co-ordination. Contrary to this, WTG is devoid of such cluster. These

finding establishes the basis of functional difference between WTG and other members of cupin super family.

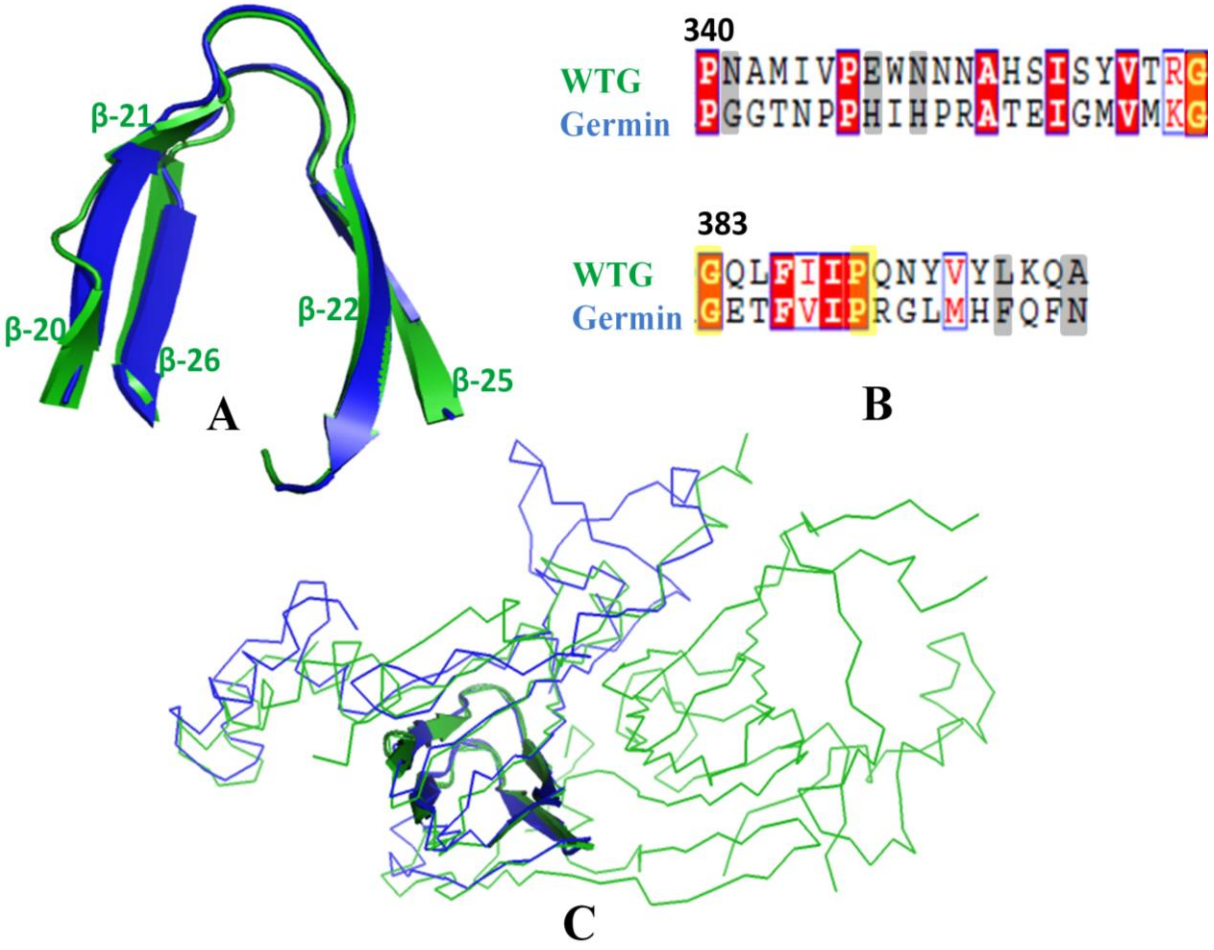


Figure 3.13. **A.** The superposition of signature stretch of Germin (blue) is made over WTG (green) to depict the structural similarity. **B.** The comparative residue profile depicting signature amino acids (grey highlights the substituted residues, yellow highlight the identical residue). **C.** The overall superposition of germin over the beta subunit of WTG monomer.

3.4.3. The auxin interaction with of WTG

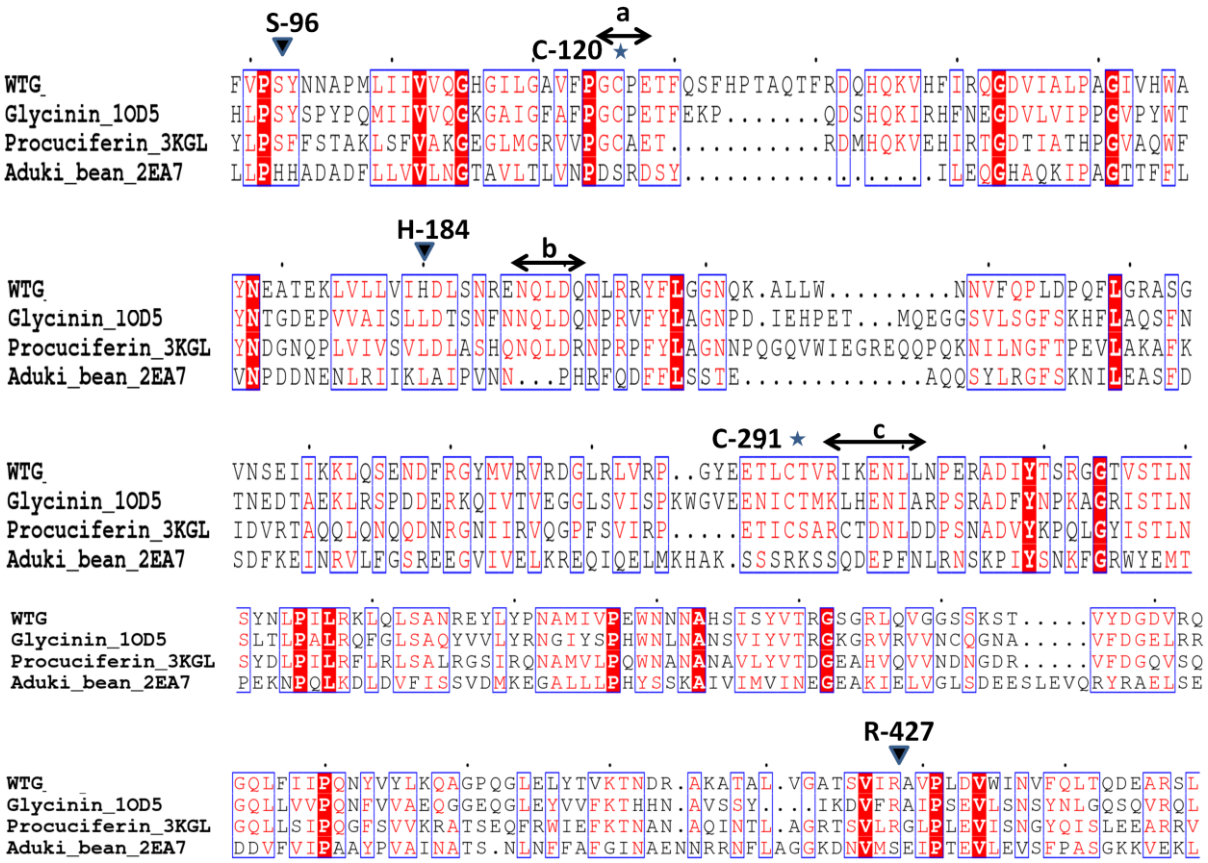


Figure 3.14. The sequence obtained from Protein Data Bank has been used for the structure based sequence comparison among the members of 11S (Glycinin PDB ID: 1OD5, Procruciferin PDB ID: 3KGL) and 7S (Aduki bean PDB ID: 2EA7) SSPs featuring the residues involved in interaction with auxin (the representation made in triangle for single residues, star for Cystine residue and the double headed arrow for small stretches involved in interaction with auxin).

The presence of auxin has been invariably observed from all the three position of the three pockets formed by the two stacking layer of trimers related by three fold axis of symmetry. The auxin is occupying the cavity constituted by all the four chains at three pockets, which is made by chains ABEF, BCDE and CADF (**Figure 3.7**). Moreover, all the pockets keep a similar orientation and position of auxin. In particular pocket BCDE the aromatic ring of auxin is stabilized by two chains C and E by involving ring of Pro 121 from both the chains

(**Figure 3.8 A**). This, the carboxyl end of the auxin is stabilized inside the pocket by the Arg-427 of chain D and Glu-122 of chain E. The other interactions involve chains B (Arg-422), chain C (Cys-120, Gly-119), chain D (Arg-422) and chain E (Gly-119, Cys-120). Therefore, auxin nailed together by four chains, invariably observed across all the three pockets from WTG native crystal structure. Furthermore, the IE face disulphide bonds and the cleaved end of acidic and basic subunits are also lying in close proximity to the pocket (**Figure 3.5**). In addition to this the extended alpha helical domain of chain D and B contributes same Arg-422 from two ends, which further stabilizes the complex. Beside this when the residues involved in interaction has been compared to other known crystal structure of the member of 11S as well as 7S super family, the results conclude a substantial correlation for the conservation of residues (**Figure 3.10**). Hence, the observed model depicts a precisely balanced interaction of auxin by virtue of carboxyl terminal as well as stacking of auxin ring.

CHAPTER – 4

PURIFICATION AND CHARACTERISATION OF

Wrightia tinctoria ALBUMIN (WTA)

4.1. Introduction

The seed storage proteins have always been known for their significant role in nutrition of humans and livestock. Recent studies unfold many interesting applications of this class of proteins. Due to their high level of expression, they have been exploited as an experimental system to explore gene regulation. The strict regulation in temporal and tissue specific manner, and also responsiveness to the mild change of environment or biochemical conditions, makes it further an interesting target for several applications (Morton et al., 1995).

The classification of seed storage proteins were originally based on solubility in defined solvents as well as method of extraction (Osborne, 1924). By definition, albumins are characterized for the solubility in water, prolamins in alcohol/water, globulins in saline and glutelins in alkali. The significant variation of size and shape also makes another basis to differentiate the storage proteins. Therefore, characteristic value of sedimentation coefficient has been used for classification. The major classes of seed storage proteins are globulins and albumins. Globulins are further divided in 11S globulins and 7S vicilins, while the albumins remain a class with typical sedimentation coefficient of 2S.

The evolutionary correlation of SSPs to other proteins converge most likely at the fern storage protein due to the significant homology observed at amino acids level among spore storage proteins and 2S albumins (Templeman et al., 1988; Rodin and Rask, 1990). They employ common mode of synthesis for their proteins, which takes place in large quantities, either by involving few genes with high level of transcription, or by lower level of expression from multiple genes (Bewley, 1997).

Among the SSPs, 2S albumins are smallest storage proteins having a molecular weight range of 9 to 20 kDa and known for most abundant form of seed proteins (Youle and Huang, 1981). The 2S albumins make a super family due to; i) amino acid sequence and conserved eight cystine residues (**Figure 4.1**), ii) basic structure made up of small and large subunits connected by disulphide bonds, and iii) strict seed specific expression. Furthermore, the amino acid sequence comparisons of 2S albumin reveal their relation to prolamin family. There are several low molecular weight proteins, which are also member of this family. A few water

soluble proteins are also member of this family such as trypsin inhibitor and α-amylase inhibitor (Dunwell et al., 2000).

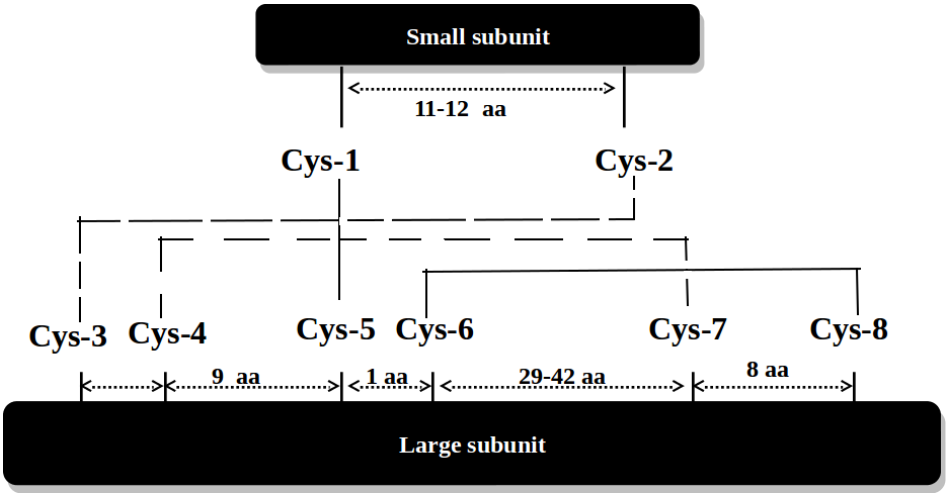


Figure 4.1. A typical arrangement of two subunits of 2S albumin depicting the disulphide bonds.

As a typical feature of seed storage protein, the 2S albumins are synthesized over rough endoplasmic reticulum and transported inside the lumen followed by processing of N- terminal signal sequence. The signatures of conserved eight cystine residues of 2S albumin get folded inside the endoplasmic reticulum that is followed by the formation of disulphide bonds (**Figure 4.1**). The protein is then transported through Golgi body system to vacuole and eventually stored in protein bodies (Reyes et al., 2011). They remain stored inside protein bodies during the maturation of seed and act as nutrient reservoir especially as a source of amino acid and carbon skeleton during the germination of seed and post germination seedling growth.

The glycosylation of 2S albumins has not been reported but post translational processing do occurs in most of the cases through proteases, which lead to the formation of a larger and smaller subunit (E. N. Clare Mills and Shewry, 2008). Moreover, the two subunits remain attached together by disulphide bonds. The 2S albumin from Sunflower has been observed as an exception to the common mode of processing that does not undergo cleavage (Gueguen et al., 1996). Besides this, 2S albumins show substantial polymorphism that leads to

the mixture of structurally related proteins synthesized by small gene families, as demonstrated in case of *Arabidopsis* (Krebbers et al., 1988).

Apart from seed storage, the 2S albumins have been reported for secondary activities from different sources. The anti-pathogenic activity has been reported from radish and *B. napus*, from which 2S albumin has been reported for antifungal activity by virtue of its ability to make the membrane of fungal hyphae extremely permeable (Terras et al., 1993). The wheat and barley thionins further showed augmented anti-pathogenic activity against growth of gram positive bacteria, when applied together with 2S albumin of radish and *B. napus* (Terras et al., 1993). Moreover, A function of trypsin and alpha amylase inhibitor has also been observed from 2S albumin due to significant amino acid homology with trypsin and alpha amylase inhibitor, (Kreis et al., 1985).

The 2S albumins are also reported for calmodulin inhibitory activity. This antagonistic behaviour has been inversely correlated with the presence of calmodulin, as it has been observed that level of calmodulin significantly rises and calmodulin inhibitors steeply decreases during the germination (Cocucci and Negrini, 1988). Moreover, The phosphorylation has also been observed over many serine residues of 2S albumin, occurred by calcium-dependent protein kinase (CDPK) at peculiar serine residues present in specific signature stretch that contains nearby basic amino acid residues (Arg, His and Lys). The phosphorylation of 2S albumin is believed as a signal for the processing of precursor, protein folding and proteolysis during germination (Polya et al., 1993). The phosphorylation is also correlated with anti-fungal activity of 2S albumin (Neumann et al., 1996).

Furthermore, the elicitation of allergy by 2S albumins is one of the major studied behaviour. It has been reported that this class of SSPs are having ability to sensitize the gastro intestinal tract leading to allergenic profile (**Table 4.1**) (Moreno and Clemente, 2008). The conserved eight cystine residue forms robust four disulphide bonds, which confer exceptional stability to the 2S albumins. The stable nature of 2S albumins helps to skip the harsh acidic and salt condition in stomach, which leads to allergy. Furthermore, the immunoglobulin IgE binds to many members of the 2S albumin family by recognizing the conserved amino acid stretch 'RGEE' (stands of Arginine-Glycine-Glutamate-Glutamate amino acids) (Moreno and Clemente, 2008). The IgE recognition of conserved immunodominant region has been observed from several plant species such as cashew nut, Brazil nut, pecan nut or castor bean (Barre et al.,

2005). These observations affirm that several members of 2S SSPs possess the allergy eliciting property, which demands a more comprehensive understanding about this class.

Table 4.1. Most common classes of some well studied allergens and their corresponding source.

Class of protein	**Plant source**
2S albumins	Brazil nut - Ber e1
	English walnut - Jug r1
	Yellow mustard - Sin a1
	Oriental mustard - Bra j1
Vicilins	Peanut - Ara h1
	English walnut - Jugr2
Inhibitors	Soybean – Kunitz trypsin inhibitor family
	Cereals – trypsin/α-amylase inhibitors
Peroxidases	Wheat
	Barley
Profilins	Peanut Ara h5, soybean, hazelnut, apple, pumpkin seeds.

The exceptional compositions of cystine residues in 2S albumins also confer unique ability for charge based interaction with metals. The 2S albumin has been observed from Brazil nut allergen Ber e 1, and reported first time for significant metal affinity (Rundqvist et al., 2012). Furthermore, it has been shown that the metal is having specific interaction due to binding inside the deep hydrophobic pocket without altering the overall structure. Apart from this, a metal binding property of 2S albumins by virtue of cystine residues may be colligated from unrelated metallothioneins class of protein, where conserved cystine residues in signature

stretch has been found responsible for metal binding. The metallothioneins are special class of small protein having exceptional ability for metal binding and also detoxification. On the basis of cystine signature repeat they are divided in three classes as; Cys-Cys, Cys-X-Cys and Cys-X-X-Cys (x-any amino acid) (Rundqvist et al., 2012). The 2S albumins have been observed to contain as many as two similar signature sequences i.e. Cys-Cys and Cys-X-X-Cys. These correlations may pave a way for possible application of 2S albumin in metal detoxification.

The aim of present chapter is characterization of *Wrightia tinctoria* 2S albumin (WTA). The studies of this chapter focus on purification, partial sequencing, biochemical and biophysical analysis of WTA. It has been observed that 2S albumins have substantial ability to bind the metals like copper in case of Brazil nut (Rundqvist et al., 2012). We have exploited the WTA as a member of this family to assess binding ability against toxic metals with the purpose of bioremediation. The binding properties have been studied against copper, cadmium and lead, which are widely known for their high environmental pollution potential and toxicity. The findings of this study can be used to propose a model that can be applied in developing synthetic or extracted proteins/peptides, which can be expressed in microbes or plants for the bio remedial applications against toxic metals.

4.2. Materials and methods

4.2.1. Materials

The WTA seeds were obtained from an authorised distributor of seeds. For purification of protein the MacroPrep DEAE matrix, standards for the molecular weight and protein estimation kit were obtained from Bio-RAD (Bio-Rad Laboratories, Hercules, CA, USA). The gel filtration has been done by using a superdex 75 column purchased from GE healthcare. The protein samples were concentrated by concentrators and centipreps purchased from Amicon, Millipore (Beverly, MA). Membranes employed for dialysis has been purchased from pierce. Reagent and buffers like tris-chloride, potassium phosphate and sodium phosphate, Ammonium sulphate, $CuCl_2$, $PbCl_2$, and CdCl2 were purchased from Himedia chemicals, India.

4.2.2. Methods

4.2.2.1. Extraction and purification of WTA

The seeds of *Wrightia tinctoria* were soaked overnight in 50 mM Tris-HCl, pH 7.5, 10% glycerol, 10 mM $MgCl_2$, 0.1 mM PMSF (Buffer A). After removal of seed-coat the kernels were minced by mortar and pestle in buffer A. The crushed samples were subjected to stirring for 6 hr followed by centrifugation at 50,000 x g at 4 °C for 45 min. A clear layer of fat was get separated and surfaced out from the supernatant. After removal of uppermost layer, sample was centrifuged in two repeat to get a clear supernatant. The supernatant was then subjected to column chromatography through MacroPrep DEAE matrix, which had been pre equilibrated in buffer A. The major fraction came out as albumin in flow-through while most of the impurities remained bound to the column. The partially purified fractions were subjected to ammonium sulphate precipitation with the percentage range of 0-20, 20-40, 40-60. The targeted protein majorly elutes in the 20-40 percentage range of ammonium sulphate. The purity of fractions were assessed over SDS-PAGE and kept for dialysis in one litre volume of buffer A in three repeats. Further, the dialysed samples were concentrated to final volume of 1ml and loaded over the superdex 75 column for gel filtration chromatography. The molecular weight has been estimated on the basis of standard marker run over SDS- PAGE.

4.2.2.2. SDS-PAGE analysis of WTA

A 15% sodium dodecyl sulphate polyacrylamide gel was run as described by Laemmli (Laemmli, 1970). The relative molecular weight was determined by the application of standard molecular weight marker. Molecular weight determination has also been done over calibrated gel filtration column in Tris-HCl buffer. The molecular weight standards used were in increasing order i.e. aprotinin (7.1 kDa), lysozyme (20.6 kDa), trypsin inhibitor (28 kDa), carbonic anhydrase (34.0 kDa), ovalbumin (49 kDa), bovine serum albumin (80 kDa), β-galactosidase (124 kDa), and myosin (209 kDa). The proteins and standards were detected by SDS-PAGE followed by staining with 0.1% Coomassie brilliant blue R-250. The molecular weight determination has been carried out by comparing the relative distance travelled by protein to the molecular weight standards.

4.2.2.3. Protein estimation

The estimation of protein has been done by employing Bio-RAD (Bradford). The standards have been drawn against the known concentration of Bovine Serum Albumen (BSA). The WTA sample volume was reduced before the estimation then the concentration was checked by using Bio-Rad (Bradford) assay.

4.2.2.4. Partial internal sequencing and analysis

An SDS-PAGE was run with purified WTA protein sample. The protein bands were excised and partially digested by Trypsin. The partially digested peptide were subjected to two dimensional liquid chromatography ESI-MS followed by reverse phase separation. The ionization of peptides was done by liquid phase electro-spray ionizer followed by their entry into ion trap that lead to fragmentation (MS/MS) and give results. The sequence data were obtained and the partial internal sequencing was analysed by the application of MASCOT search engine at the protein sequencing facility National Botanical Research Institute, Lucknow, India. The obtained sequencing results were used for comparisons against the available ESTs data base of *Wrightia tinctoria* accessed through NCBI BLAST server (Johnson et al., 2008). Furthermore, the correlation with the related members and sequences were fetched by using same server. Moreover, the super family characterisation has been achieved on the basis of alignment by using super family 1.75 HMM library and customized genome assignment server (Gough et al., 2001). The alignment profile and depiction have been generated by Jalview version 2 (Waterhouse et al., 2009), Clustal (Chenna et al., 2003) and ESPript version 3 webserver (Gouet et al., 2003).

4.2.2.5. Far-UV Circular Dichroism spectrum

Circular dichroism studies were performed for the WTA at the Chirascan circular dichroism spectrophotometer (Applied Photophysics, UK). A region of far-UV CD spectra (180-260) were recorded using 1mm quartz cell with the bandwidth of 1nm and the average time of 3 sec at constant temperature of 25 °C. The CD spectra were acquired by using constant protein concentration of 0.2 mg/ml at the pH 7.5. Three repeats were performed for each scan, which were averaged out for final signal acquisition. The base line values of buffer

were subtracted to get the final values. The web based software Dichroweb and neural network program CDNN were used for analysis of data (Whitmore and Wallace, 2004).

4.2.2.6. Isothermal titration calorimetry

To measure the affinity of WTA to metal, the isothermal titration calorimetry has been employed by using the Microcal ITC 200 from GE healthcare. The data were acquired at constant temperature 25 °C with the varying concentration of ligand while protein remains constant. All the solutions were degassed prior to the calorimetric analysis. The constant protein concentration of 0.3 mM has been used to fill the sample cell and used for titration against different ligands. The ligand and protein were kept in 20 mM potassium phosphate of buffer of pH 7.5. The titrations were done in 10-20 injections containing 2 µL of ligand for each injection with 100 sec of interval. The data correction has been done for each of the injections related heat of dilution produced as a result of continuous injections. The processing of data and model fitting has been done by ORIGIN software provided with ITC software from GE healthcare.

4.2.2.7. *In silico* Structural modelling of WTA

The WTA model has been built *in silico* by using a template structure (PDB ID: 1PSY), which is a three dimensional structure of 2S albumin 'RicC3' from Ricinus communis. The MODELLER was used for the comparative modeling (Sali et al., 1995).The generation of models of WTA was followed by minimization using Swiss Pdb Viewer 4.01 (Guex and Peitsch, 1997). The best model was taken on the basis of stereochemical properties and used for further refinement in MODELLER's loop refinement tool (Fiser et al., 2000). The model was validated by Ramachandran plot, ERRAT PLOT and ProSA analysis (Luthy et al., 1992; Laskowski et al., 1993; Sippl, 1993). The generated model was visualized, inspected and analyzed by PyMOL (DeLano, 2002). The three dimensional generated structure was subjected to energy minimization and finally validated by Ramachandran plot analysis by PROCHECK server.

4.3. Results and discussions

4.3.1. Purification of WTA

The *W. tinctoria* 2S albumin has been purified from the seeds as a natural source of protein. The pure and homogeneous form of WTA has been obtained by employing three steps of purifications. The first step of purification involves ion exchange chromatography. The partially purified protein was obtained after ion exchange chromatography, in the form of flow through while the impurities remains bound to the column. The second step involves ammonium sulphate precipitation, which reduces the volume and also produces more pure protein. The pure and homogeneous protein was obtained in third step of purification, by employing the gel filtration chromatography. The purity of protein samples during the process of purification has been monitored by SDS-PAGE. The analysis of possible disulphide bond and related subunit profile has been assessed by running the SDS-PAGE in reducing as well as non reducing condition. The results demonstrate that WTA is a heterodimeric protein and the dimers are attached through disulphide bond. It has been observed that WTA remains united in non reducing condition with molecular weight ~ 19 kDa that splits in presence of reducing agent to a larger subunit of molecular weight ~ 12 kDa and a smaller subunit of molecular weight ~ 7 kDa (**Figure 4.2**).

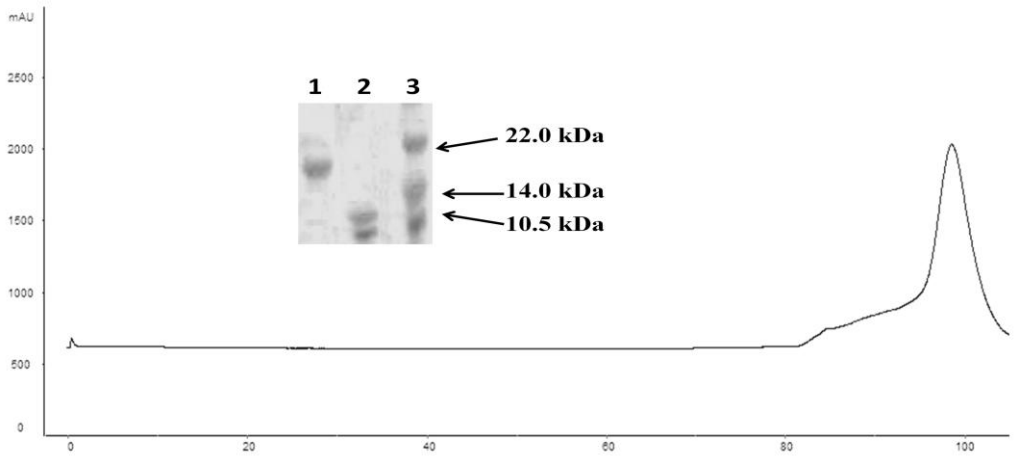


Figure 4.2. The WTA purification to homogeneity by gel filtration chromatography; the inset figure displays the protein profile in absence (lane 1) and presence (lane 2) of reducing agent and molecular weight standard (lane 3).

4.3.2. Partial internal sequencing and analysis

The partial internal sequencing results have given three peptide stretches; the identity of protein has been established by sequence alignment based on these stretches. The three best stretches obtained from partial sequencing are Q/*K*-C-C-Q/K-P-L/*I*-R, I/*L*-F-Q/*K*-E-Q/*K*-V-R and N-L/*I*-P-Q/*K*-E-C-N-I/*L*-G-T-E-C-Q/*K*-I/*L*-R (the options in italic letters are substituted) (**Figure 4.3 A**). The identified sequence was used for the comparisons with homologous proteins. The relation with homologous members has been established by assessing the alignment profile through NCBI-BLAST tool that has given best alignment with 2S albumins (**Figure 4.3 B**). Furthermore, the careful sequence analysis for the related member of 2S albumin establishes their link to a big super family designated as 'Bi-functional inhibitor/lipid-transfer protein/seed storage 2S albumin super family'. The cystine has been marked as a major residue exhibits invariable presence. In addition to this, the buried existence of CC signature residues inside the pocket has been invariantly observed from entire super family (**Figure 4.4 A and B**). Beside this, by virtue of similar hydrophilic and hydrophobic profile and conserved aligned residues, WTA may also share somewhat similar features of related members. These features encompass surface interaction, nature of pocket as well as ligand interactions and role in enzyme inhibition etc. (**Figure 4.4. A**).

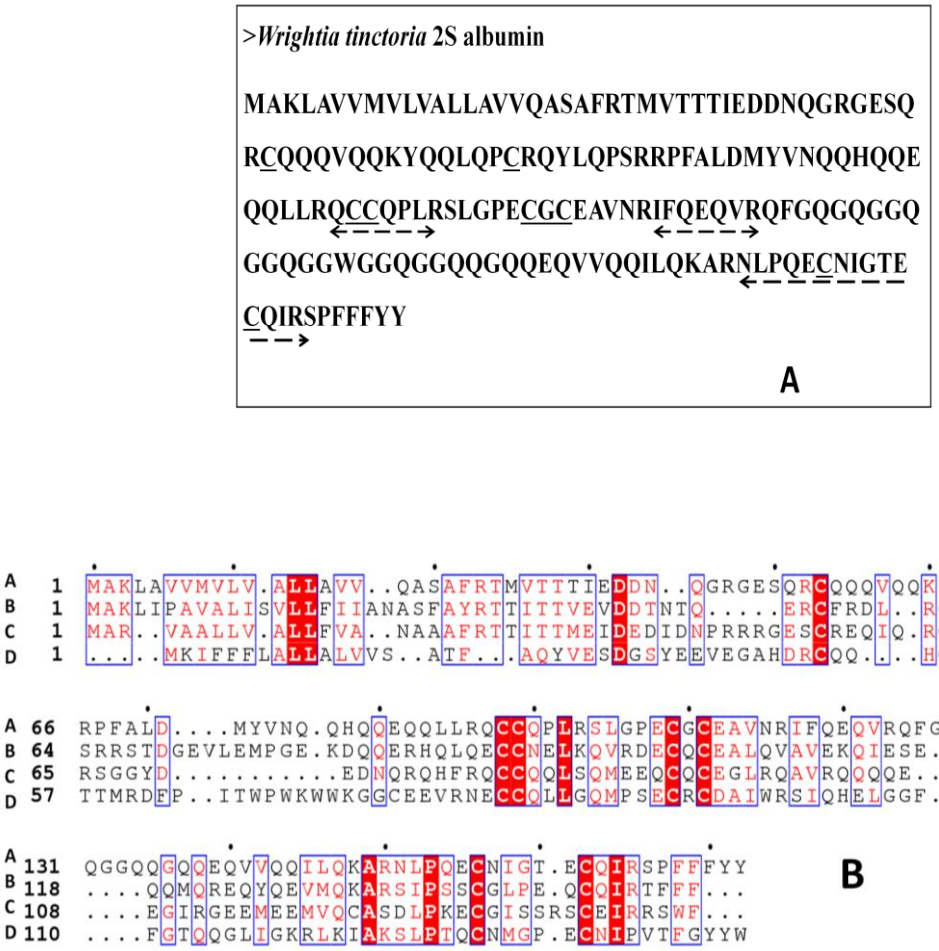


Figure 4.3. WTA sequence identity established by; **A**. the peptides stretches obtained from internal sequencing of WTA as represented by dotted arrow, signature cystine residues has been represented in underlined residues. **B.** the sequence comparison with related member of 2S albumin family depicts conserved residues (A-WTA, B-2S albumin from *Ricinus. Communis*, C-Allergen I1 from *Carya illinoinensis.* D-Tryptophanin from *Avena sativa*).

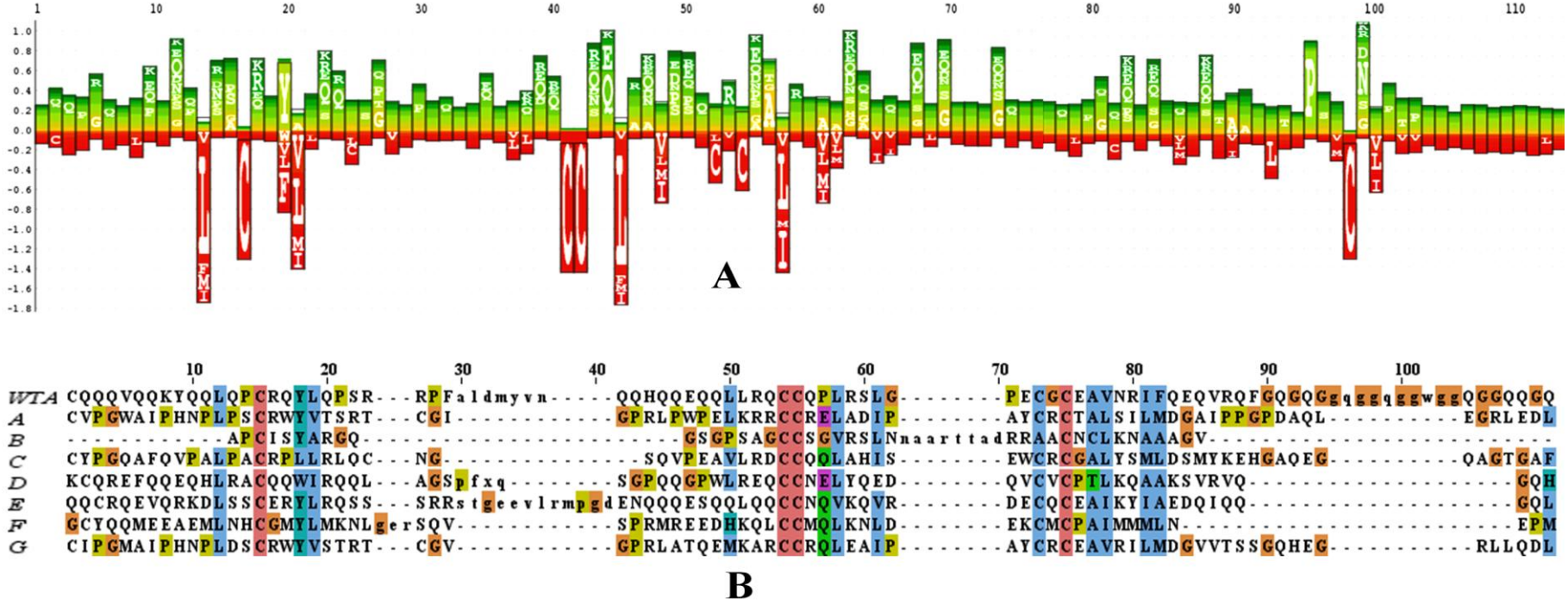


Figure 4.4. The plot **A.** showing match probabilities in the 2S albumin family along with amino acids in each column ordered in the form of most hydrophilic (top) to most hydrophobic (bottom) (here size of each column is proportional to the difference of conservation). **B**. A window of most conserved amino acids displaying the 2S albumin superfamily (The first row represent the sequence of WTA followed by A-Hageman factor inhibitor from *Zea mays*, B-phospholipid transfer protein 1 from *Zea mays*, C-Alpha-amylase inhibitor *Aegilops tauschii*, D-Napin from *Brassica napus*, E-Ricc3 from *Ricinus communis*, F-*Helianthus annuus* Methionine-rich 2S protein, G-Bifunctional proteinase inhibitor trypsin and alpha-amylase from *Eleusine Coracana*).

4.3.3. Far UV circular dichroism analysis

The basic structural components of WTA have been evaluated by circular dichroism studies. The deconvolution results of the CD data for the range of 180 to 260 nm showed that WTA secondary structure proportion predominantly encompasses alpha helix with 98.10%, while beta turns constitutes 4.7%, random coil 1% and the parallel beta sheet is limited to 0.30%. No anti-parallel beta sheets have been observed. The results of secondary structure profile display the typical nature of WTA as observed from other members of the 2S albumin family (**Figure 4.5**). In addition to this, the modeling studies of WTA further correlate to similar secondary structural profile.

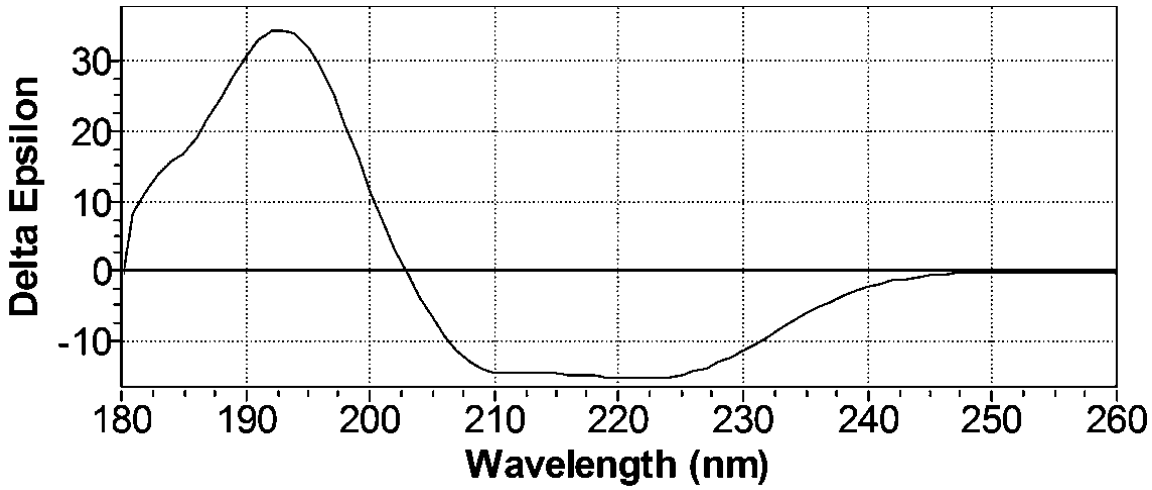


Figure 4.5. The Circular dichroism spectrum depicting secondary structure profile of WTA.

4.3.4. Isothermal titration calorimetry

The isothermal titration calorimetry has been used to assess a metal binding activity of WTA. The binding affinity of 2S albumin for copper has been reported from an allergen Ber e1of the plant Brazil nut (Rundqvist et al., 2012). We have examined the copper binding affinity for 2S albumin of WTA by employing ITC. Further, we have exploited this feature of 2S albumin to bind other heavy metals known for toxicity and environmental pollution. The significant binding of metals has been observed. The binding affinity (M^{-1}) has been observed

by metals in decreasing order from copper ($K > 10^5$) followed by cadmium ($K > 10^4$) than lead ($K > 10^3$) (**Figure 4.6**). These findings present the WTA as a protein having reasonable metal binding property (especially for the studied heavy metals) and thus have a potential to use it for the bioremediation related applications. It appears that cystine residues of WTA play a key role for metal interaction, as observed from other metal binding proteins (Mejare and Leif, 2001). Further, the invariant presence of cystine among 2S albumin concludes that this class is having significant role for metal interaction (**Figure 4.3**).

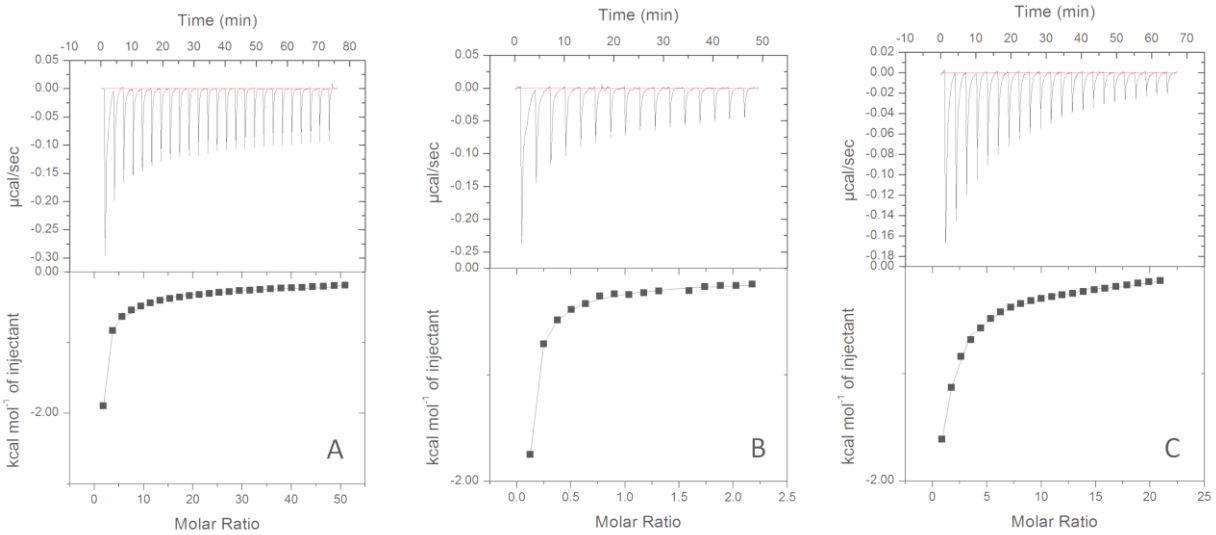


Figure 4.6. The isothermal titration calorimetry plots displaying affinity of WTA for different metals (A-Copper, B-Cadmium and C-lead).

4.3.5. The *in-silico* modeling and structural characterization of WTA

The model of WTA was built by employing comparative modeling. The three dimensional structure of model was generated by providing WTA sequence and closest structural homologue having PDB ID: 1PSY (NMR structure of Ric C3, a member of 2S SSP family from *Ricinus communis*) from RCSB database. The modelled structure was validated by stereo chemical evaluation, which has been done by PROCHECK that shows 99.2% residues are lying in most favoured and allowed regions in Ramachandran plot and 0.8 % in disallowed regions. Furthermore, the energy function of the model has been assessed by ProSA server, which showed the overall Z-score -7.8 i.e. a negative Z-score of energy that confirmed good

overall quality of the model. The model of WTA exhibits similar feature of presence of cystine residue as reported from the members of the 2S SSP family. The cystine residues held responsible for exceptional stability of 2S albumins (**Figure 4.7)**.

The structural profile of the model was found in accordance with the bio-physical characterization details of WTA, as observed from circular dichroism studies. The WTA belongs to alpha class of structure. Furthermore, the conserved positions of cystine earmark the relation with 2S albumin and also confer a compact three dimensional structure. The WTA scaffold comprises bundle of five α-helices (**Figure 4.7. B**). Apart from structural strength, the disulphide bonds confer the stability against food processing that becomes a basis of allergen nature, known as a characteristic feature of this class (Pantoja-Uceda et al., 2004). The structural superposition WTA shares significant resemblance to the allergens as observed with the allergens Ara h2 from *Arachis hypogea* (PDB ID: 3OB4) and Ber e1 from *Bertholletia excelsa* (PDB ID: 2LVF) (**Figure 4.7. C**).

The 2S albumin has also been concluded as an extended structure encompasses ample amount hydrophilic residues at surface, which is stabilized by electrostatic interactions, while the lack of significant hydrophobic environment has been observed. This electrostatic profile is in accordance with a reported member of 2S albumin 'napin' from *Brassica juncea* (Jyothi et al., 2007). Such profile has also been profoundly observed across 2S albumin family. Hence it appears that hydrophilic nature and supported ionic interaction may also participate in metal ion interaction as observed from the WTA biochemical studies by isothermal calorimetory analysis.

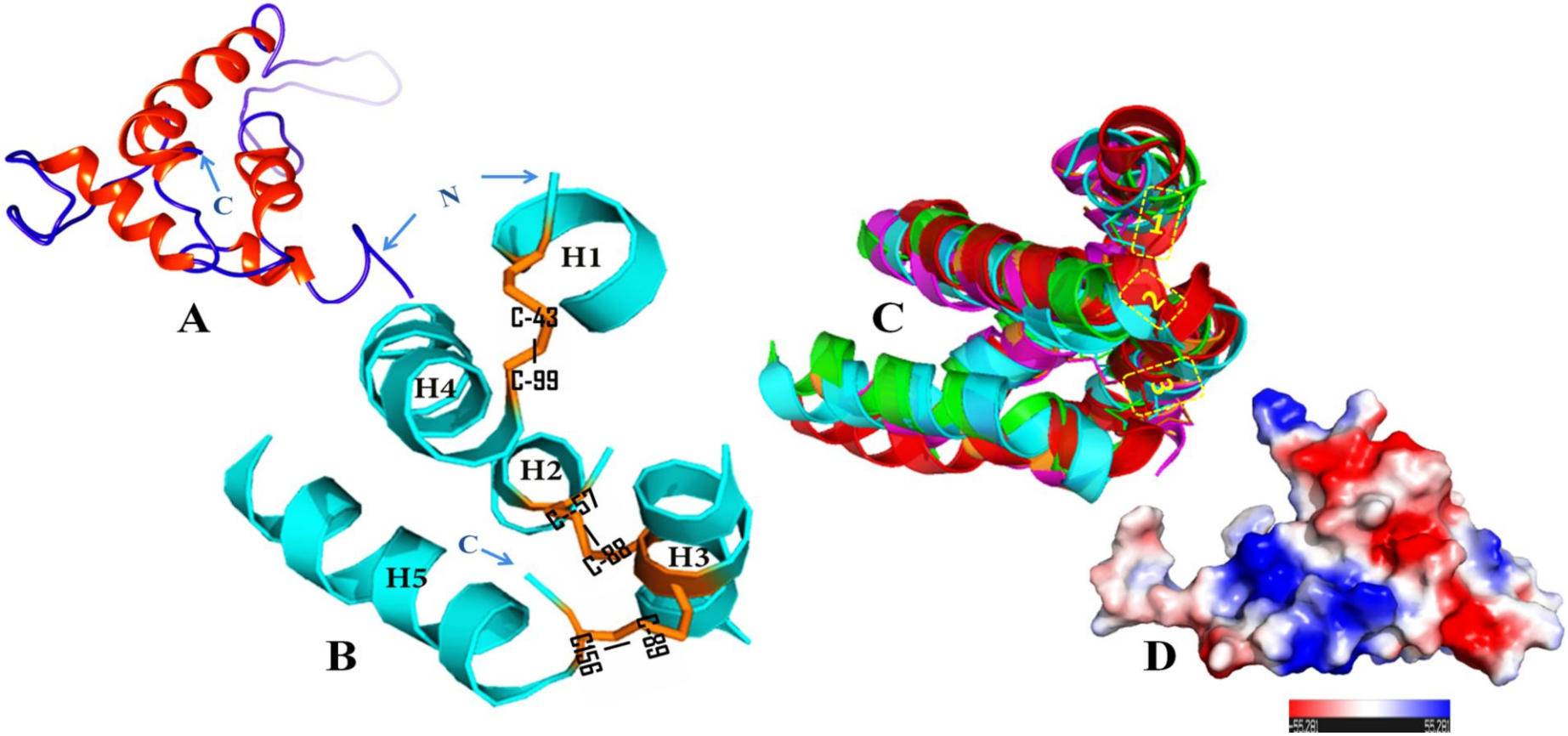


Figure 4.7. **A**. Overall model of WTA, **B**. A typical helical arrangement of protein with cystine residues, **C**. The structural comparisons displaying most invariant three disulphide bonds of WTA (in orange colour) with RicC3 from *Ricinus communis* (PDB ID: 1PSY in magenta), Mabinlin II from *Capparis masaikai* (PD ID: 2DS2 in cyan), allergen Ara h 2 from *Arachis hypogea* (PDB ID: 3OB4 in green), Ber e 1 from *Bertholletia excels* (PDB ID: 2LVF in red), **D**. The overall electrostatic surface profile of WTA (colour coded electrostatic surface unit -55.2 in red to 55.2 in blue).

CONCLUSIONS

PURIFICATION FROM NATURAL SOURCE, CLONING, BIOCHEMICAL AND BIOPHYSICAL CHARACTERISATIONS OF 11S GLOBULIN FROM *Wrightia tinctoria*

- The globulin was purified from the plant *Wrightia tinctoria* and identified as a member of 11S family. The WTG shares substantial sequence homology with the member of this family and encompasses conserved bicupin fold.

- The WTG exists in hexameric form having overall molecular weight of ~ 320 kDa, where each monomer has a molecular weight ~ 56 kDa. The WTG encompasses two subunits; a larger alpha subunit and a smaller beta subunit with molecular weight ~ 32 and ~ 24 kDa respectively.

- The monomer of WTG encompasses two heterogeneous subunits connected together by disulphide bond. The monomer remains in homogenous form while the subunits display significant heterogeneity.

- The WTG exhibits a significant hemagglutination behaviour, which may confer it an ability to act as lectin like protein; hence it can be considered for future studies to find a possible role in plant defence system.

- The three dimensional structure of WTG reveals that it shares maximum similarity with allergens.

- The WTG encompasses epitopic determinants both as linear epitope over continuous amino acid residues as well as structural epitope by virtue of allergy eliciting fold and loops. The alpha subunit of WTG contains most of the allergenic determinants.

- The allergenic profile analysis on the basis of structural and corresponding amino acid residues conclude that WTG may be exploited as a possible model for allergen and related studies.

- Structural stability of mature WTG depends over the post translational processing by asparaginyl endopeptidase. Apparently this processing governs, i) precise packing of subunits and monomer for stable hexamer structure, ii) release of steric constraints that confers mobility to the extended alpha helical region involved in head to tail association, iii) formation of a cavity that holds the auxin.

- Unlike other known SSPs, the N-terminal of WTG contains a surprisingly dynamic region. They are having significant hydrophilic nature as well as such amino acid profile that may positively undergo phosphorylation. They are having high potential for structure mobility also. Since this dynamic region is lying at the poles of hexamer, therefore this delicate position may transfer any small change over the entire hexamer. In view of these observations, it is may be concluded that the dynamic region from N-

terminal of WTG can leverage possible splitting of entire scaffold, in influence of mild change of pH or salt like balance of nearby environment. This splitting may also be associated with release of auxin.

- The details of breakdown of WTG and there correlation to N-terminal flexible dynamic residues can be studied, which will elucidate the basis of dissociation of hexameric scaffold and release of auxin. Therefore, a better understanding of the structure/function relations of WTG remains of significant importance as well as target for future research.

PURIFICATION AND CHARACTERISATION OF *WRIGHTIA TINCTORIA* ALBUMIN (WTA)

- The *Wrightia tinctoria* albumin was purified from the plant *Wrightia tinctoria* and identified as a member of 2S seed storage protein family.
- The WTA is a small protein that belongs to bi-functional inhibitor/lipid-transfer protein/seed storage 2S albumin super family.
- The basic organization of WTA involves N-terminal smaller and C-terminal larger subunits, together encompasses four signature disulphide bonds.
- The identity of protein has been concluded on the basis of obtained sequence, which reveals the presence of signature cystine residues C-Xn-C, C-C, C-X-C and C-Xn-C (X can be any amino acid residue/s with n length) shared by entire super family.

- The WTA is an alpha helical protein, which displays substantial metal binding ability. The scope of metal binding with WTA has been concluded for Copper, Cadmium and Lead with significant affinity.

- These findings pave the way for possible application of WTA in bioremediation of toxic compounds containing heavy metals like copper, cadmium and lead.

- The major advantage of small protein and further smaller subunit size make WTA a suitable protein that can be engineered and expressed either in bacteria or plant to possibly circumvent the heavy metal pollution. Further studies and validation may make the WTA a suitable agent to get exploited for bioremediation related applications and to avoid, reduce and mitigate the environmental pollution.

REFERENCES

References

Adachi M, Kanamori J, Masuda T, Yagasaki K, Kitamura K, Mikami B, Utsumi S. (2003), Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. *Proceedings of the National Academy of Sciences*, (100), 7395-7400.

Adachi M, Okuda E, Kaneda Y, Hashimoto A, Shutov AD, Becker C, Müntz K, Utsumi S. (2003), Crystal structures and structural stabilities of the disulfide bond-deficient soybean proglycinin mutants C12G and C88S. *Journal of Agricultural and Food Chemistry*, (51), 4633-4639.

Adachi M, Takenaka Y, Gidamis AB, Mikami B, Utsumi S. (2001), Crystal structure of soybean proglycinin A1aB1b homotrimer. *Journal of molecular biology*, (305), 291-305.

Agizzio AP, Da Cunha M, Carvalho OA, Oliveira MA, Ribeiro SFF, Gomes VM. (2006), The antifungal properties of a 2S albumin-homologous protein from passion fruit seeds involve plasma membrane permeabilization and ultrastructural alterations in yeast cells. *Plant Science*, (171), 515-522.

Akihisa T, Ahmad I, Singh S, Tamura T, Matsumoto T. (1988), 14α-Methylzymosterol and other sterols from *Wrightia tinctoria* seeds. *Phytochemistry*, (27), 3231-3234.

Ampe C, Damme J, Castro LAB, Sampaio MJAM, Montagu M, Vandekerckhove J. (1986), The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* HBK). *European Journal of Biochemistry*, (159), 597-601.

Amsterdam A, Jamieson JD. (1972), Structural and functional characterization of isolated pancreatic exocrine cells. *Proceedings of the National Academy of Sciences*, (69), 3028-3032.

Asero R, Ballmer Weber BK, Beyer K, Conti A, Dubakiene R, Fernandez-Rivas M, Hoffmann-Sommergruber K, Lidholm J, Mustakov T, Oude Elberink JNG. (2007), IgE-Mediated

food allergy diagnosis: Current status and new perspectives. *Molecular nutrition & food research*, (51), 135-147.

Banerjee N, Sengupta S, Roy A, Ghosh P, Das K, Das S. (2011), Functional alteration of a dimeric insecticidal lectin to a monomeric antifungal protein correlated to its oligomeric status. *PloS one*, (6), e18593.

Barre A, Borges J-P, Culerrier R, Rouga P. (2005), Homology modelling of the major peanut allergen Ara h 2 and surface mapping of IgE-binding epitopes. *Immunology letters*, (100), 153-158.

Baumlein H, Braun H, Kakhovskaya IA, Shutov AD. (1995), Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. *Journal of Molecular Evolution*, (41), 1070-1075.

Berger-Bachi B, Rohrer S. (2002), Factors influencing methicillin resistance in staphylococci. *Archives of microbiology*, (178), 165-171.

Berger-Bachi B, Strassle A, Gustafson JE, Kayser FH. (1992), Mapping and characterization of multiple chromosomal factors involved in methicillin resistance in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, (36), 1367-1373.

Berger-Bachi B, Tschierske M. (1998), Role of Fem factors in methicillin resistance. *Drug Resistance Updates*, (1), 325-335.

Bewley JD. (1997), Seed germination and dormancy. *The plant cell*, (9), 1055-1066.

Blom N, Gammeltoft S, Brunak S. (1999), Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*, (294), 1351-1362.

Boutilier K, Hattori J, Baum BR, Miki BL. (1999), Evolution of 2S albumin seed storage protein genes in the Brassicaceae. *Biochemical systematics and ecology*, (27), 223-234.

Brinegar AC, Peterson DM. (1982), Separation and characterization of oat globulin polypeptides. *Archives of Biochemisty and Biophysics.*, (219), 71-79.

Casey R, Domoney C, Ellis N. (1986), Legume storage proteins and their genes. *Oxford surveys of plant molecular and cell biology. Volume 3.*, 1-95.

CCP4. (1994), Collaborative Computational Project Number 4, The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallographica Section D*, (50), 760–763.

Chadha YR. (1976), Wealth of India, Raw Material. *Y.R. Chadha (Ed) CSIR, New*, (10), 588.

Chadha YR (1985) The Wealth of India. Raw materials. Vol. I: A. Publications & Information Directorate, CSIR

Chen J, Liu B, Ji N, Zhou J, Bian H, Li C, Chen F, Bao J. (2009), A novel sialic acid-specific lectin from *Phaseolus coccineus* seeds with potent antineoplastic and antifungal activities. *Phytomedicine*, (16), 352-360.

Chen Y, Barkley MD. (1998), Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, (37), 9976-9982.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. (2003), Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research*, (31), 3497-3500.

Choi SM, Ma CY. (2006), Extraction, purification and characterization of globulin from common buckwheat (*Fagopyrum esculentum* Moench) seeds. *Food research international*, (39), 974-981.

Clara Sze KW, Kshirsagar HH, Venkatachalam M, Sathe SK. (2007), A circular dichroism and fluorescence spectrometric assessment of effects of selected chemical denaturants on soybean (*Glycine max* L.) storage proteins glycinin (11S) and -conglycinin (7S). *Journal of Agricultural and Food Chemistry*, (55), 8745-8753.

Cocucci M, Negrini N. (1988), Changes in the levels of calmodulin and of a calmodulin inhibitor in the early phases of radish (*Raphanus sativus L.*) seed germination effects of ABA and fusicoccin. *Plant physiology*, (88), 910-914.

Danielsson CE. (1949), Seed globulins of the *Graminae* and *Leguminosae*. *Biochemical Journal*, (44), 387-400.

Dasgupta J, Dasgupta S, Ghosh S, Roy B, Mandal RK. (1995), Deduced amino acid sequence of 2S storage protein from *Brassica species* and their conserved structural features. *Indian journal of biochemistry & biophysics*, (32), 378.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS. (2007), MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research*, (35), W375-W383.

De Britto AJ, Sebastian SR. (2013), Phytochemical evaluation and in vitro antibacterial screening of *Wrightia tinctoria* (roxb.) r. br. against enteric pathogens. *International journal of pharmaceutical, chemical and biological sciences*, (3), 210-214.

De Clercq A, Vandewiele M, De Rycke R, Van Damme J, Van Montagu M, Krebbers E, Vandekerckhove Jl. (1990), Expression and processing of an *Arabidopsis* 2S albumin in transgenic tobacco. *Plant physiology*, (92), 899-907.

DeLano WL. (2002), The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.

Deswal R, Gupta R, Dogra V, Singh R, Abat JK, Sarkar A, Mishra Y, Rai V, Sreenivasulu Y, Amalraj RS. (2013), Plant proteomics in India and Nepal: current status and challenges ahead. *Physiology and Molecular Biology of Plants*, (19), 1-17.

Dickinson CD, Hussein EH, Nielsen NC. (1989), Role of posttranslational cleavage in glycinin assembly. *The Plant Cell Online*, (1), 459-469.

Dufour E, Hoa GHB, Haertle T. (1994), High-pressure effects on [beta]-lactoglobulin interactions with ligands studied by fluorescence. *Biochim. Biophys. Acta*, (1206), 166-172.

Dunwell J. (1998), Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnology & Genetic Engineering Reviews*, (15), 1-32.

Dunwell JM, Khuri S, Gane PJ. (2000), Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiology and molecular biology reviews*, (64), 153-179.

Dunwell JM, Purvis A, Khuri S. (2004), Cupins: the most functionally diverse protein superfamily? *Phytochemistry*, (65), 7-17.

Duranti M, Guerrieri N, Takahashi T, Cerletti P. (1988), The legumin-like storage protein of *Lupinus albus* seeds. *Phytochemistry*, (27), 15-23.

E. N. Clare Mills, Shewry PR. (2008), Plant Food Allergens. *Wiley-Blackwell*, 248.

Emsley P, Cowtan K. (2004), Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, (60), 2126-2132.

Fan X, Liu Y, Smith D, Konermann L, Siu KWM, Golemi-Kotra D. (2007), Diversity of Penicillin-binding Proteins resistance factor FmtA of staphylococcus aureus. *Journal of Biological Chemistry*, (282), 35143-35152.

Farley AR, Link AJ. (2009), Identification and quantification of protein posttranslational modifications. *Methods in enzymology*, (463), 725-763.

Fetzner S. (2012), Ring-cleaving dioxygenases with a cupin fold. *Applied and Environmental Microbiology*, (2012),

Fiser A, Do RKG, Sali A. (2000), Modeling of loops in protein structures. *Protein science*, (9), 1753-1773.

Flinn BS, Roberts DR, Newton CH, Cyr DR, Webster FB, Taylor IEP. (1993), Storage protein gene expression in zygotic and somatic embryos of interior spruce. *Physiologia Plantarum*, (89), 719-730.

Fuda CCS, Fisher JF, Mobashery S. (2005), β-Lactam resistance in Staphylococcus aureus: the adaptive resistance of a plastic genome. *Cellular and molecular life sciences*, (62), 2617-2633.

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. *In* The proteomics protocols handbook. Springer, pp 571-607.

Gatehouse JA, Croy RRD, Boulter D, Shewry PR. (1984), The synthesis and structure of pea storage proteins. *Critical Reviews in Plant Sciences*, (1), 287-314.

Gatehouse JA, Evans IM, Croy RRD, Boulter D. (1986), Differential expression of genes during legume seed development. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, (314), 367-384.

Gouet P, Robert X, Courcelle E. (2003), ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic acids research*, (31), 3320-3323.

Gough J, Karplus K, Hughey R, Chothia C. (2001), Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of molecular biology*, (313), 903-919.

Gueguen J, Popineau Y, Anisimova IN, Fido RJ, Shewry PR, Tatham AS. (1996), Functionality of the 2S albumin seed storage proteins from sunflower (*Helianthus annuus* L.). *Journal of agricultural and food chemistry*, (44), 1184-1189.

Guex N, Peitsch MC. (1997), SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, (18), 2714-2723.

Horta ANACG, De Azevedo MR. (2001), Isolation and partial characterization of a lectin from *Bauhinia pentandra* (Bong) vog. Ex. Steua. *Revista Brasileira de Fisiologia Vegetal*, (13), 262.

Jain PS, Bari SB. (2010), Anti-inflammatory effects of wood stem extracts of *Wrightia tinctoria*. *Asian Journal of Traditional Medicines*, (5), 132-137.

Jeffery JW (1971) Methods in X-ray Crystallography. ACS Publications

Job C, Rajjou L, Lovigny Y, Belghazi M, Job D. (2005), Patterns of protein oxidation in *Arabidopsis* seeds and during germination. *Plant physiology*, (138), 790-802.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. (2008), NCBI BLAST: a better web interface. *Nucleic acids research*, (36), W5-W9.

Josefsson L-G, Lenman M, Ericson ML, Rask L. (1987), Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*. *Journal of Biological Chemistry*, (262), 12196-12201.

Joss JL, Molloy MP, Hinds LA, Deane EM. (2006), Evaluation of chemical derivatisation methods for protein identification using MALDI MS/MS. *International Journal of Peptide Research and Therapeutics*, (12), 225-235.

Jung R, Scott MP, Nam YW, Beaman TW, Bassuner R, Saalbach I, Muntz K, Nielsen NC. (1998), The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell*, (10), 343-358.

Jyothi TC, Singh SA, Appu Rao AG. (2007), Conformation of napin (*Brassica juncea*) in salts and monohydric alcohols: Contribution of electrostatic and hydrophobic interactions. *Journal of agricultural and food chemistry*, (55), 4229-4236.

Kataoka M, Hirata K, Kunikata T, Ushio S, Iwaki K, Ohashi K, Ikeda M, Kurimoto M. (2001), Antibacterial action of tryptanthrin and kaempferol, isolated from the indigo plant (*Polygonum tinctorium Lour.*), against *Helicobacter pylori*-infected *Mongolian gerbils*. *Journal of gastroenterology*, (36), 5-9.

Keshri G, Kumar S, Kulshreshtha DK, Rajendran SM, Singh MM. (2008), Postcoital interceptive activity of *Wrightia tinctoria* in Sprague Dawley rats: a preliminary study. *Contraception*, (78), 266-270.

Khuri S, Bakker FT, Dunwell JM. (2001), Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. *Molecular biology and evolution*, (18), 593-605.

Kim HT, Choi U-K, Ryu HS, Lee SJ, Kwon O-S. (2011), Mobilization of storage proteins in soybean seed (*Glycine max* L.) during germination and seedling growth. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, (1814), 1178-1187.

Ko TP, Ng JD, McPherson A. (1993), The three-dimensional structure of canavalin from jack bean (Canavalia ensiformis). *Plant physiology*, (101), 729-744.

Komatsuzawa H, Ohta K, Labischinski H, Sugai M, Suginaka H. (1999), Characterization of fmtA, a gene that modulates the expression of methicillin resistance in Staphylococcus aureus. *Antimicrobial agents and chemotherapy*, (43), 2121-2125.

Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vandekerckhove J. (1988), Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. *Plant physiology*, (87), 859-866.

Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR. (1985), Molecular evolution of the seed storage proteins of barley, rye and wheat. *Journal of molecular biology*, (183), 499-502.

Kumar L. (2011), Anti oxidation activity of *Wrightia tinctoria* Roxb bark and *Schrebera swietenoides* Roxb bark extract. *Journal of Pharmacy Research*, (4), 2.

Kumar P, N Patil D, Chaudhary A, Tomar S, Yernool D, Singh N, Dasauni P, Kundu S, Kumar P. (2013), Purification and Biophysical Characterization of an 11S Globulin from Wrightia tinctoria Exhibiting Hemagglutinating Activity. *Protein and Peptide Letters*, (20), 499-509.

Kurplus M, McCammon JA. (1983), Dynamics of proteins: elements and function. *Annual review of biochemistry*, (52), 263-300.

Laemmli UK. (1970), Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, (227), 680-685.

Lam SK, Ng TB. (2011), Lectins: production and practical applications. *Applied Microbiology and Biotechnology*, (89), 45-55.

Lambert N, Plumb GW, Wright DJ. (1987), Application of high-performance liquid chromatography to the assessment of subunit heterogeneity in plant 11S storage globulins. *Journal of Chromatography A*, (402), 159-172.

Lane BG, Bernier F, Dratewka-Kos E, Shafai R, Kennedy TD, Pyne C, Munro JR, Vaughan T, Walters D, Altomare F. (1991), Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain *Physarum spherulins*. *Journal of Biological Chemistry*, (266), 10461-10469.

Langston-Unkefer PJ, Gade W. (1984), A seed storage protein with possible self-affinity through lectin-like binding. *Plant physiology*, (74), 675-680.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993), PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, (26), 283-291.

Lawrence MC, Izard T, Beuchat M, Blagrove RJ, Colman PM. (1994), Structure of Phaseolin at 2. 2 Å. resolution: implications for a common Vicilin/Legumin structure and the genetic engineering of seed storage proteins. *Journal of molecular biology*, (238), 748-776.

Lazar E, Watanabe S, Dalton S, Sporn MB. (1988), Transforming growth factor alpha: mutation of aspartic acid 47 and leucine 48 results in different biological activities. *Molecular and Cellular Biology*, (8), 1247-1252.

Li C, Li M, Dunwell JM, Zhang Y-M. (2012), Gene duplication and an accelerated evolutionary rate in 11S globulin genes are associated with higher protein synthesis in dicots as compared to monocots. *BMC evolutionary biology*, (12), 15.

Lovell SC, Davis IW, Arendall WB, De Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. (2003), Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, (50), 437-450.

Lowy FD. (1998), Staphylococcus aureus infections. *New England Journal of Medicine*, (339), 520-532.

Lowy FD. (2003), Antimicrobial resistance: the example of Staphylococcus aureus. *Journal of Clinical Investigation*, (111), 1265-1273.

Luthy R, Bowie JU, Eisenberg D. (1992), Assessment of protein models with three-dimensional profiles. *Nature*, (356), 83-85.

Mandal S, Mandal RK. (2000), Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. *Current sciences-Bangalore*, (79), 576-589.

Marchuk D, Drumm M, Saulino A, Collins FS. (1991), Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic acids research*, (19), 1154-1154.

Martin W, Herrmann RG. (1998), Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant physiology*, (118), 9-17.

Matsudaira P. (1987), Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *The Journal of Biological Chemistry*, (262), 10035-10038.

Matsuura JE, Manning MC. (1994), Heat-induced gel formation of beta.-lactoglobulin: A study on the secondary and tertiary structure as followed by circular dichroism Spectroscopy. *Journal of Agricultural and Food Chemistry*, (42), 1650-1656.

Matta NK, Gatehouse JA, Boulter D. (1981), Molecular and subunit heterogeneity of legumin of *Pisum sativum* L.(Garden pea), a multi-dimensional gel electrophoretic study. *Journal of experimental botany*, (32), 1295-1307.

McCallum N, Berger-Bachi B, Senn MM. (2010), Regulation of antibiotic resistance in *Staphylococcus aureus*. *International Journal of Medical Microbiology*, (300), 118-129.

Mejare M, Leif B. (2001), Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. *Trends in Biotechnology*, (19), 67-73.

Milne JA, Cook RA. (1979), Role of metal cofactors in enzyme regulation. Differences in the regulatory properties of the *Escherichia coli* nicotinamide adenine dinucleotide specific malic enzyme depending on whether magnesium (2+) or manganese (2+) serves as divalent cation. *Biochemistry*, (18), 3604-3610.

Monsalve RI, Rodriguez R. (1990), Purification and characterization of proteins from the 2S fraction from seeds of the Brassicaceae family. *Journal of experimental botany*, (41), 89-94.

Moreira RDA, Perrone JC. (1977), Purification and partial characterization of a lectin from Phaseolus vulgaris. *Plant Physiology*, (59), 783-787.

Moreno FJ, Clemente A. (2008), 2S albumin storage proteins: what makes them food allergens? *The open biochemistry journal*, (2), 16–28.

Moreno FJ, Clemente A. (2008), 2S albumin storage proteins: what makes them food allergens? *The open biochemistry journal*, (2), 16-28.

Morton RL, Quiggin D, Higgins TJV. (1995), Regulation of seed storage protein gene expression. *Seed Development and Germination*, 103-138.

Moura FT, Oliveira AS, Macedo LLP, Vianna ALBR, Andrade LBS, Martins-Miranda AS, Oliveira JTA, Santos EA, De Sales MP. (2007), Effects of a chitin-binding vicilin from *Enterolobium contortisiliquum* seeds on bean bruchid pests (*Callosobruchus maculatus* and *Zabrotes subfasciatus*) and phytopathogenic fungi (*Fusarium solani* and *Colletrichum lindemuntianum*). *Journal of agricultural and food chemistry*, (55), 260-266.

Murshudov GN, Vagin AA, Dodson EJ. (1997), Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography*, (53), 240-255.

Muruganandam AV, Bhattacharya SK, Ghosal S. (2000), Indole and flavanoid constituents of *Wrightia tinctoria, W. tomentosa* and *W. coccinea*. *Indian Journal of Chemistry. Section B, Organic including Medicinal*, (39), 125-131.

Neumann GM, Condron R, Thomas I, Polya GM. (1996), Purification and sequencing of multiple forms of *Brassica napus* seed napin small chains that are calmodulin antagonists and substrates for plant calcium-dependent protein kinase. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, (1295), 23-33.

Osborne TB. (1924), The Vegetable Proteins. *London: Longmans, Green,*

Osborne TB, Campbell GF. (1898), Proteids of the pea. *Journal of the American Chemical Society*, (20), 348-362.

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

Palomares O, Vereda A, Cuesta-Herranz J, Villalba M, Rodriguez R. (2007), Cloning, sequencing, and recombinant production of Sin a 2, an allergenic 11S globulin from yellow mustard seeds. *Journal of allergy and clinical immunology*, (119), 1189-1196.

Pantoja-Uceda D, Palomares O, Bruix M, Villalba M, Rodriguez R, Rico M, Santoro J. (2004), Solution structure and stability against digestion of rproBnIb, a recombinant 2S albumin

from rapeseed: relationship to its allergenic properties. *Biochemistry*, (43), 16036-16045.

Pelegrini PB, Noronha EF, Muniz MAR, Vasconcelos IM, Chiarello MD, Oliveira JTA, Franco OL. (2006), An antifungal peptide from passion fruit (*Passiflora edulis*) seeds with similarities to 2S albumin proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, (1764), 1141-1146.

Pernollet JC, Mosse J. (1983), Structure and location of legume and cereal seed storage proteins. *Seed proteins/edited by J. Daussant, J. Mosse, J. Vaughan*, 155-119.

Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011), SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, (8), 785-786.

Peterson DM. (1978), Subunit structure and composition of oat seed globulin. *Plant physiology*, (62), 506-509.

Pinho MG, Filipe SR, de Lencastre H, Tomasz A. (2001), Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (PBP2) by the drug resistance protein PBP2A in Staphylococcus aureus. *Journal of bacteriology*, (183), 6525-6531.

Pohleven J, Brzin J, Vrabec L, Leonardi A, okl A, Strukelj B, Kos J, Saboti J. (2011), Basidiomycete Clitocybe nebularis is rich in lectins with insecticidal activities. *Applied Microbiology and Biotechnology*, (91), 1141-1148.

Polya GM, Chandra S, Condron R. (1993), Purification and sequencing of radish seed calmodulin antagonists phosphorylated by calcium-dependent protein kinase. *Plant physiology*, (101), 545-551.

Ponnusamy K, Petchiammal C, Mohankumar R, Hopper W. (2010), In vitro antifungal activity of indirubin isolated from a South Indian ethnomedicinal plant *Wrightia tinctoria* R. Br. *Journal of ethnopharmacology*, (132), 349-354.

R. Casey. (1979), Genetic variability in the structure of the -subunits of legumin from *Pisum*-A two-dimensional gel electrophoresis study. *Heredity*, (43), 265-272.

Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, Bannon GA. (1999), Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *Journal of Clinical Investigation*, (103), 535-542.

Rajalakshmi GR, Harindran J. (2013), In vitro Anthelmintic Activity Of *Wrightia tinctoria*. *In vitro*, (5), 308-310.

Rao MN, Rao EV, Rao VS (1966) Triterpenoid components of the leaves and pods of *Wrightia tinctoria*, Ed 20 Vol 35

Raynal M, Depigny D, Grellet F, Delseny M. (1991), Characterization and evolution of napin-encoding genes in radish and related crucifers. *Gene*, (99), 77-86.

Reddy YSR, Venkatesh S, Ravichandran T, Murugan V, Suresh B. (2002), Antinociceptive activity of *Wrightia tinctoria* bark. *Fitoterapia*, (73), 421-423.

Reyes FC, Chung T, Holding D, Jung R, Vierstra R, Otegui MS. (2011), Delivery of prolamins to the protein storage vacuole in maize aleurone cells. *The Plant Cell Online*, (23), 769-784.

Robotham JM, Hoffman GG, Teuber SS, Beyer K, Sampson HA, Sathe SK, Roux KH. (2009), Linear IgE-epitope mapping and comparative structural homology modeling of hazelnut and English walnut 11S globulins. *Molecular immunology*, (46), 2975-2984.

Rochat C, Boutin JP. (1992), Temporary storage compounds and sucrose starch metabolism in seed coats during pea seed development (*Pisum sativum*). *Physiologia Plantarum*, (85), 567-572.

Rodin J, Rask L. (1990), Characterization of matteuccin, the 2.2 S storage protein of the ostrich fern. *European Journal of Biochemistry*, (192), 101-107.

Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. (1992), In-gel digestion of proteins for internal sequence analysis after one-or two-dimensional gel electrophoresis. *Analytical biochemistry*, (203), 173-179.

Rundqvist L, Tengel T, Zdunek J, Bjorn E, Schleucher J, Alcocer MJC, Larsson G. (2012), Solution structure, copper binding and backbone dynamics of recombinant Ber e 1-the major allergen from Brazil Nut. *PloS one*, (7), e46435.

Saalbach I, Waddell D, Pickardt T, Schieder O, Muntz K. (1995), stable expression of the sulphur-rich 2S albumin gene in transgenic *Vicia narbonensis* increases the methionine content of Seeds. *Journal of plant physiology*, (145), 674-681.

Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. (1995), Evaluation of comparative protein modeling by MODELLER. *Proteins: Structure, Function, and Bioinformatics*, (23), 318-326.

Selvam P, Murugesh N, Witvrouw M, Keyaerts E, Neyts J. (2009), Studies of antiviral activity and cytotoxicity of *Wrightia tinctoria* and *Morinda citrifolia*. *Indian journal of pharmaceutical sciences*, (71), 670–672.

Sharma GM, Mundoma C, Seavy M, Roux KH, Sathe SK. (2010), Purification and Biochemical Characterization of Brazil Nut (*Bertholletia excelsa* L.) Seed Storage Proteins. *Journal of Agricultural and Food Chemistry*, (58), 5714-5723.

Shewry PR, Casey R (1999) Seed proteins. Springer

Shewry PR, Napier JA, Tatham AS. (1995), Seed storage proteins: structures and biosynthesis. *The plant cell*, (7), 945.

Shoichet BK, Baase WA, Kuroki R, Matthews BW. (1995), A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences*, (92), 452-456.

Shorr AF. (2007), Epidemiology of staphylococcal resistance. *Clinical Infectious Diseases*, (45), S171-S176.

Shutov AD, Baumlein H (1999) Origin and evolution of seed storage globulins. *In* Seed Proteins. Springer, pp 543-561.

Shutov AD, Baumlein H, Blattner FR, Muntz K. (2003), Storage and mobilization as antagonistic functional constraints on seed storage globulin evolution. *Journal of experimental botany*, (54), 1645-1654.

Shutov AD, Braun H, Chesnokov YV, Baumlein H. (1998), A gene encoding a vicilin-like protein is specifically expressed in fern spores. *European Journal of Biochemistry*, (252), 79-89.

Shutov AD, Kakhovskaya IA. (2011), Evolution of seed storage globulins and cupin superfamily. *Molecular Biology*, (45), 529-535.

Shutov AD, Kakhovskaya IA, Braun H, Baumlein H, Muntz K. (1995), Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. *Journal of Molecular Evolution*, (41), 1057-1069.

Sippl MJ. (1993), Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, (17), 355-362.

Soares EL, Freitas CDT, Oliveira JS, Sousa PAS, Sales MP, Barreto-Filho JDM, Bandeira GP, Ramos MV. (2007), Characterization and insecticidal properties of globulins and albumins from *Luetzelburgia auriculata*(Allemao) Ducke seeds towards *Callosobruchus maculatus* (F.)(Coleoptera: Bruchidae). *Journal of Stored Products Research*, (43), 459-467.

Strynadka NCJ, Adachi H, Jensen SE, Johns K, Sielecki A, Betzel C, Sutoh K, James MNG. (1992), Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution. *Nature*, (359), 186-190.

Swain M, Ross NW. (1995), A silver stain protocol for proteins yielding high resolution and transparent background in sodium dodecyl sulfate polyacrylamide gels. *Electrophoresis*, (16), 948-951.

Tandang-Silvas MRG, Fukuda T, Fukuda C, Prak K, Cabanos C, Kimura A, Itoh T, Mikami B, Utsumi S, Maruyama N. (2010), Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, (1804), 1432-1442.

Templeman TS, Stein DB, DeMaggio AE. (1988), A fern spore storage protein is genetically similar to the 1.7 S seed storage protein of *Brassica napus*. *Biochemical genetics*, (26), 595-603.

Terras FRG, Schoofs HME, Thevissen K, Osborn RW, Vanderleyden J, Cammue BPA, Broekaert WF. (1993), Synergistic enhancement of the antifungal activity of wheat and barley thionins by radish and oilseed rape 2S albumins and by barley trypsin inhibitors. *Plant physiology*, (103), 1311-1319.

Utaida S, Dunman PM, Macapagal D, Murphy E, Projan SJ, Singh VK, Jayaswal RK, Wilkinson BJ. (2003), Genome-wide transcriptional profiling of the response of Staphylococcus aureus to cell-wall-active antibiotics reveals a cell-wall-stress stimulon. *Microbiology*, (149), 2719-2732.

Utsumi S. (1992), Plant food protein engineering. *Advances in food and nutrition research*, (36), 89-208.

Utsumi S, Inaba H, Mori T. (1981), Heterogeneity of soybean glycinin. *Phytochemistry*, (20), 585-589.

Vagin A, Teplyakov A. (1997), MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography*, (30), 1022-1025.

Varier PS. (1997), Indian Medicinal Plants - a compendium of 500 species. *Coll. No., AVS 1647:*, (5), 417- 418.

Verhees CH, Huynen MA, Ward DE, Schiltz E, de Vos WM, van der Oost J. (2001), The phosphoglucose isomerase from the hyperthermophilic archaeon *Pyrococcus furiosus* is a unique glycolytic enzyme that belongs to the cupin super family. *Journal of Biological Chemistry*, (276), 40926-40932.

Walburg G, Larkins BA. (1983), Oat seed globulin subunit characterization and demonstration of its synthesis as a precursor. *Plant physiology*, (72), 161-165.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. (2009), Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, (25), 1189-1191.

Whitmore L, Wallace BA. (2004), DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research*, (32), W668-W673.

Woo EJ, Dunwell JM, Goodenough PW, Marvier AC, Pickersgill RW. (2000), Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nature Structural & Molecular Biology*, (7), 1036-1040.

Woo EJ, Marshall J, Bauly J, Chen JG, Venis M, Napier RM, Pickersgill RW. (2002), Crystal structure of auxin-binding protein 1 in complex with auxin. *The EMBO journal*, (21), 2877-2885.

Wooh JW, Kidd RD, Martin JL, Kobe B. (2003), Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallographica Section D: Biological Crystallography*, (59), 769-772.

Wright DJ, Boulter D. (1974), Purification and subunit structure of legumin of *Vicia faba* L.(broad bean). *Biochemical Journal*, (141), 413-418.

Yadav S, Tomar AK, Jithesh O, Khan MA, Yadav RN, Srinivasan A, Singh TP, Yadav S. (2011), Purification and partial characterization of low molecular weight Vicilin-like glycoprotein from the seeds of *Citrullus lanatus*. *The protein journal*, (30), 575-580.

Youle RJ, Huang AHC. (1981), Occurrence of low molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species. *American Journal of Botany*, (68), 44-48.

Zemser M, Friedman M, Katzhendler J, Greene LL, Minsky A, Gorinstein S. (1994), Relationship between functional properties and structure of ovalbumin. *Journal of protein chemistry*, (13), 261-274.

Zirwer D, Gast K, Welfle H, Schlesier B, Dieter Schwenke K. (1985), Secondary structure of globulins from plant seeds: A re-evaluation from circular dichroism measurements. *International Journal of Biological Macromolecules*, (7), 105-108.

APPENDIX

CHAPTER – A

PURIFICATION AND CRYSTALLIZATION OF FmtA FROM *Staphylococcus aureus*

A.1. Introduction

The hospital and community acquired pathogenic infections is one of the major medical problems for decades and the *Staphylococcus aureus* has been identified as a most common pathogen responsible for such infections. It is a gram-positive facultative aerobic bacteria, which can cause diverse range of diseases such as sepsis, pneumonia, endocarditis etc. (Lowy, 1998) (**Figure A 1**). The pathology of *S. aureus* becomes further grimmer due to the exceptional ability of antibiotic resistance, displayed by it. The resistance of *S. aureus* has been reported against several antibiotics. Therefore to cure its infections an antibiotic methicillin has been developed. This antibiotic is especially applied against those strains of *S. aureus*, which has been already diagnosed with the resistance to other β-lactam based antibiotics. But soon after the introduction of methicillin, this microbe again developed resistance. Furthermore, the methicillin enduring strains of *S. aureus* are also reported for their tolerance to almost every kind of beta lactams based antibiotics used in clinical practice (Lowy, 2003; Fuda et al., 2005). It has been reported that more than 63 % of this microbe isolated from critically ill patients can behave as methicillin resistant *S. aureus* (MRSA) (Shorr, 2007). Therefore the methicillin antibiotic resistance is one of the most challenging cases.

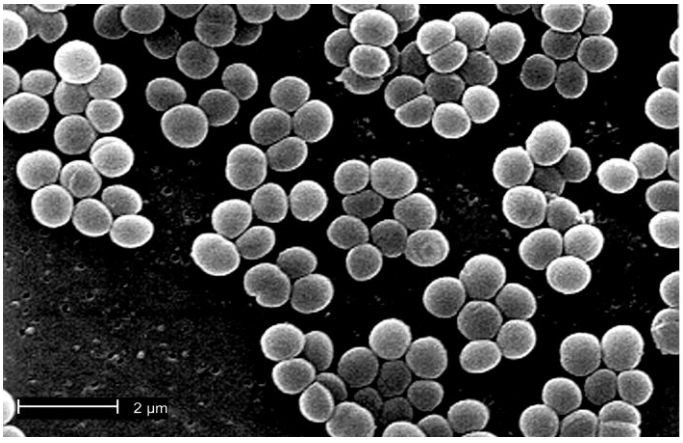


Figure A.1. The typical cocci of *S. aureus*, depicting their morphology.

The penicillin binding protein PBP2a is a major factor, that facilitate methicillin resistance to *S. aureus* (Pinho et al., 2001). The last step of peptidoglycan biosynthesis is carried out by PBP2s, hence considered as a good target for antibiotics having β-lactam structure. Moreover, it has been reported that in case of antibiotic treatment this family of proteins take over the charge of peptidoglycan biosynthesis (Pinho et al., 2001). Apart from this, many other factors also involved in the *S. aureus* antibiotics resistance (Berger-Bachi et al., 1992; Berger-Bachi and Rohrer, 2002). The *fem* genes have been reported for production of some major factors responsible for the augmentation of methicillin resistance (Berger-Bachi and Tschierske, 1998; Berger-Bachi and Rohrer, 2002). Further, the *fmt* (methicillin resistance factor) group of genes have been particularly observed as an integral part of cell wall stimulon that gets activated in the presence of β-lactams and other cell wall targeting antibiotics (Utaida et al., 2003; McCallum et al., 2010). Furthermore, the *fmtA* gene has been particularly identified, for direct involvement in the cell wall synthesis during antibiotic stress. The role of this factor has been proved by their disruption, which leads to the altered structure of peptidoglycans (Komatsuzawa et al., 1999). The FmtA is a key factor having ability to directly alter the penicillin activity by forming covalent complexes. The basic mode of interaction involves the formation of acyl-enzyme complex between lactam ring and Serine residue present at the active site. This Serine is strictly conserved among penicillin binding proteins and β-lactamases with the signature residues, Ser-X-X-Lys (Strynadka et al., 1992; Fan et al., 2007).

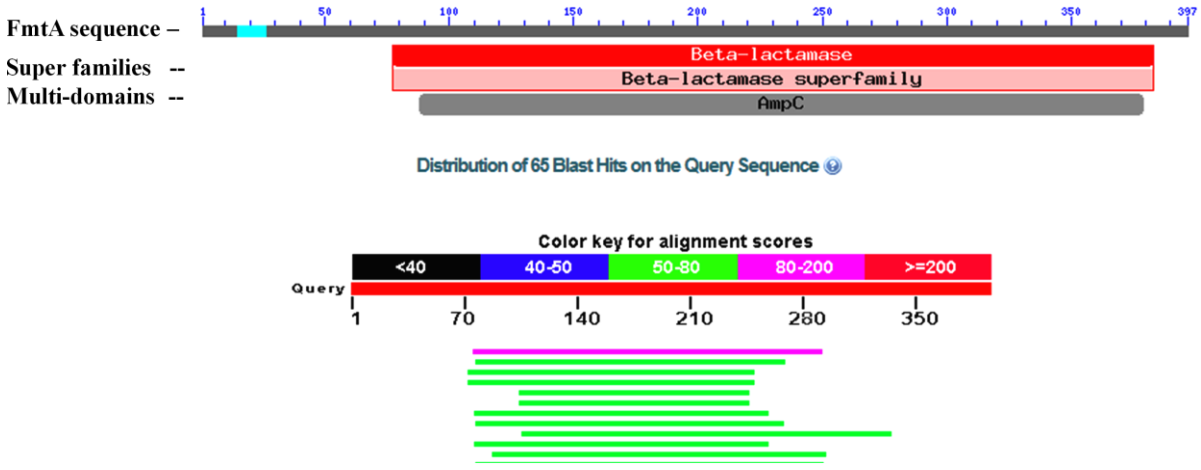


Figure A.2. The FmtA sequence alignment against Protein Data Bank entries, displaying list of available homologues and query coverage.

Despite of the major involvement in antibiotic resistance the structure of the FmtA has not been determined so far (**Figure. A2**). The present study has been focused on the purification and crystallization of fmtA protein from *S. aureus*.

A.2. Material and methods

A.2.1. Materials

The chemical and reagents has been purchased from the authentic distributors of manufacturing companies, Sigma- Aldrich Corporations, St. Louis, MO USA: Himedia Laboratories India Private Limited, Mumbai, India; Bio-RAD Laboratories, Hercules, California, USA and Merck Limited, Worli, Mumbai, India. The columns employed for the protein purification were HP-SP (5 mL) and Hiload superdex 75 16/60, the calibration was done by a HMW (high molecular weight) kit purchased from GE Healthcare, AB Uppsala, Sweden. The millex syringe filters, amicon ultra concentrators and PVDF membranes were purchased from Millipore Corporation, Billerica, MA. For dialysis, the membranes were purchased from Pierce, Rockford, USA with the cut-off range of 3500 Da. The protein estimation was done by Bio-Rad protein assay kit purchased from Bio-RAD Laboratories, Hercules, California, USA. The crystallization trays, crystal screens and cryoloops were purchased from Hampton Research, CA, USA. The kanamycin and IPTG were purchased from Sigma Aldrich.

A.2.2. Methods

A.2.2.1. Purification of FmtA

For the purification of the FmtA protein, the plasmid containing *fmtA* gene in pET24a vector has been provided by Dr. Dasantila Golemi-Kotra, York University, Canada, in which the signal peptide sequence as initial 27 amino acid residues, has been omitted during the cloning of *fmtA* gene. The cloned plasmid was transformed in *E. Coli*-BL21 expression system.

The transformed colony were picked from the Petri plates and inoculated to the 10 ml seed culture that contains antibiotic kanamycin (30 μg/mL). The seed culture has been raised overnight and next day used for the inoculation on large scale. Large culture was raised using the seed culture with 30 μg/mL concentration of kanamycin, 0.4 M D-sorbitol and 2.5 mM betaine. The cells were allowed to grow till the absorbance of media reaches the OD_{600nm} of value ~ 0.6 to 0.7 at a temperature of 37 °C with gentle shaking at 120 rpm. Once the culture attains OD, the temperature was reduced to 25 °C and induced by 1mM concentration of Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16hr at 120rpm. Finally, the cells were harvested by centrifugation at 8000 rpm for 10 min at 4 °C. The pellets were stored at -80 °C.

The cell pellet was washed in buffer A (50 mM sodium phosphate pH7.2), then dissolved in lysis buffer containing buffer A and 20% of buffer B (50 mM sodium phosphate, 1 M NaCl pH7.2). Suspended cells were disrupted using high pressure cell disruption by applying the mechanical pressure of 20,000 PSI using the cell disruptor from Constant system ltd., Northants, United Kingdom. The disrupted cells were centrifuged at 50,000 g. The supernatant was loaded over HP-SP column pre-equilibrated with buffer A (50 mM sodium phosphate pH7.2) containing 20% of buffer B. The elution of protein has been made by applying the gradient of range 25-80% by buffer B against buffer A. The eluted fractions were run on SDS-PAGE. The partially purified protein fractions were pooled together and kept for overnight dialysis in 2 litres of buffer A. The dialysed samples were concentrated to the final volume of 1ml and loaded over Superdex 75 16/60 Hi load column pre-equilibrated with buffer A. The fractions of FmtA were analysed by SDS-PAGE for the purity of sample.

A.2.2.2. Crystallization of FmtA

The pure protein was used for crystallization trials in a 96 well crystal trays at a temperature of 20 °C. The crystal screens from Hampton were used for initial crystallization trials. A drop size of 2 μL containing 1:1 ratio of reservoir to protein was kept. The initial hits obtained from screens were optimized subsequently to get good diffraction quality crystals.

The raised crystals were crushed and run over SDS-PAGE followed by visualization using silver staining (Swain and Ross, 1995). The list of reagents used for the developing and staining of gel are categorised as i) fixer; containing 50% methanol, 12% acetic acid, 35%

formaldehyde, ii) washing solution; contains 35% ethanol, iii) sensitizer; 0.02 % sodium-thio-sulphate in water, silver nitrate solution; contains 0.2% silver nitrate, 35% formaldehyde, iv) developer solution; sodium bicarbonate 6%, 35% formaldehyde and sodium thiosulphate and v) blocking solution containing 50% methanol, 12% acetic acid, 35% formaldehyde. For the silver staining, sample present in the gel was fixed by applying the fixing solution, followed by washing and sensitization. Finally the gel was stained with silver nitrate, followed by processing in developer and blocking solutions. The silver stained gel was used for extra sensitive sample analysis.

A.2.2.3. Data Collection of FmtA

The crystal was soaked in cryoprotectant prior to the freezing with liquid nitrogen cooled stream. The data was collected over MAR 345 dtb imaging plate system with the X-ray generator from Bruker Microstar H with Cu*K*α rotating anode generator operated at 45 kV with the current of 60 mA that is equipped with Helious optics. Data were collected for 200 images with a distance of 250 nm with the oscillation of 1° per images and the time of exposure was kept 10 min for each image. The crystal was diffracted to maximum resolution of 3.0 Å. The initial indexing, integration and scaling was done by using the software HKL-2000 (Otwinowski and Minor, 1997). The initial structure solution was tried by molecular replacement program MOLREP available through program suite CCP4 (CCP4, 1994; Vagin and Teplyakov, 1997), with the related homologous proteins from Protein Data bank having PDB ID: 4GDN, 2QMI. The preliminary statistics has been shown in **Table A.1**.

A.3. Results and discussions

A.3.1. FmtA purification

The two step purification has been used to get the pure and homogenous form of FmtA. The first step involves ion-exchange chromatography, which gives a partially purified protein and in second step the pure protein was obtained through gel filtration chromatography. At the first stage of purification, a partially purified protein was obtained with the gradient of 25% to

80% of buffer B. The analysis of purification was done by SDS-PAGE (**Figure A.3. A and B**), Further, the protein was purified to homogeneity by gel filtration chromatography (**Figure A.3. C**).

Once the purity of protein is confirmed over SDS-PAGE, the concentration of protein was estimated and volume was reduced to get the optimal protein concentration of approximately 30 mg/ml. The concentrated protein was then subjected to crystallization trails.

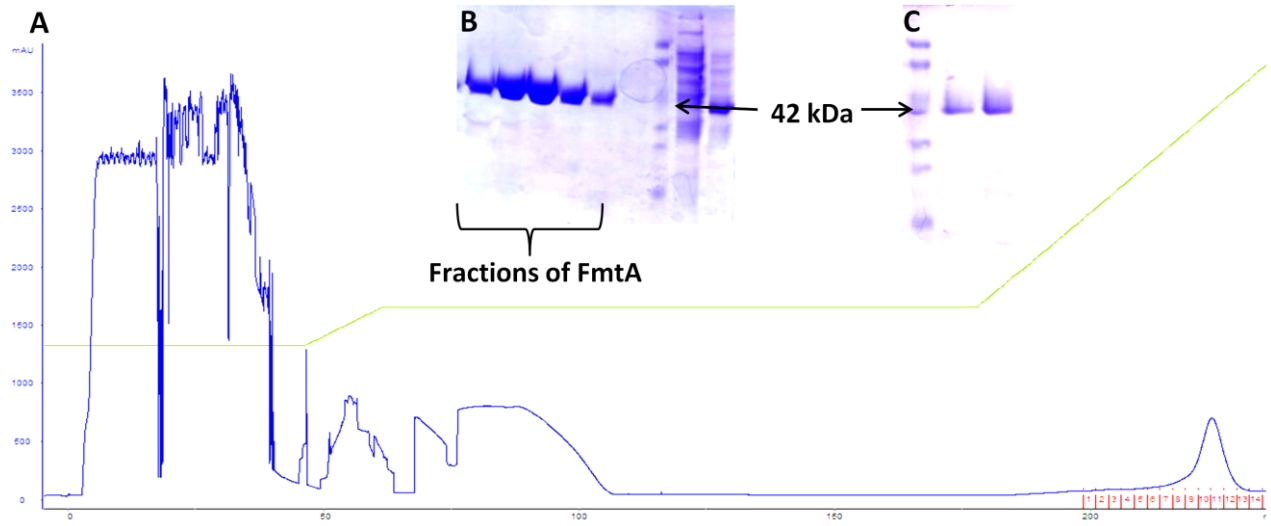


Figure A.3. The purification profile of FmtA has been shown; A. The peak obtained under the gradient of NaCl corresponds to the FmtA, the inset B is SDS-PAGE eluted ion exchange fractions, inset C corresponds to SDS-PAGE filtration of pure protein obtained after gel filtration.

A.3.2. Crystallization of FmtA

The initial crystallization trials of FmtA were done with the commercial screen sets. The crystals start appearing only after a period of three months from setting the trays. Crystals were obtained in the condition containing 0.2 M NaCl, 0.1 M Tris pH 8.5, and 25% PEG 3350. Further the crystallization conditions were optimized from the initial crystal hits. The PEG ion gradient with different concentration range of protein along with pH gradient was used for the improvisation of crystals of FmtA (**Figure A.4**).

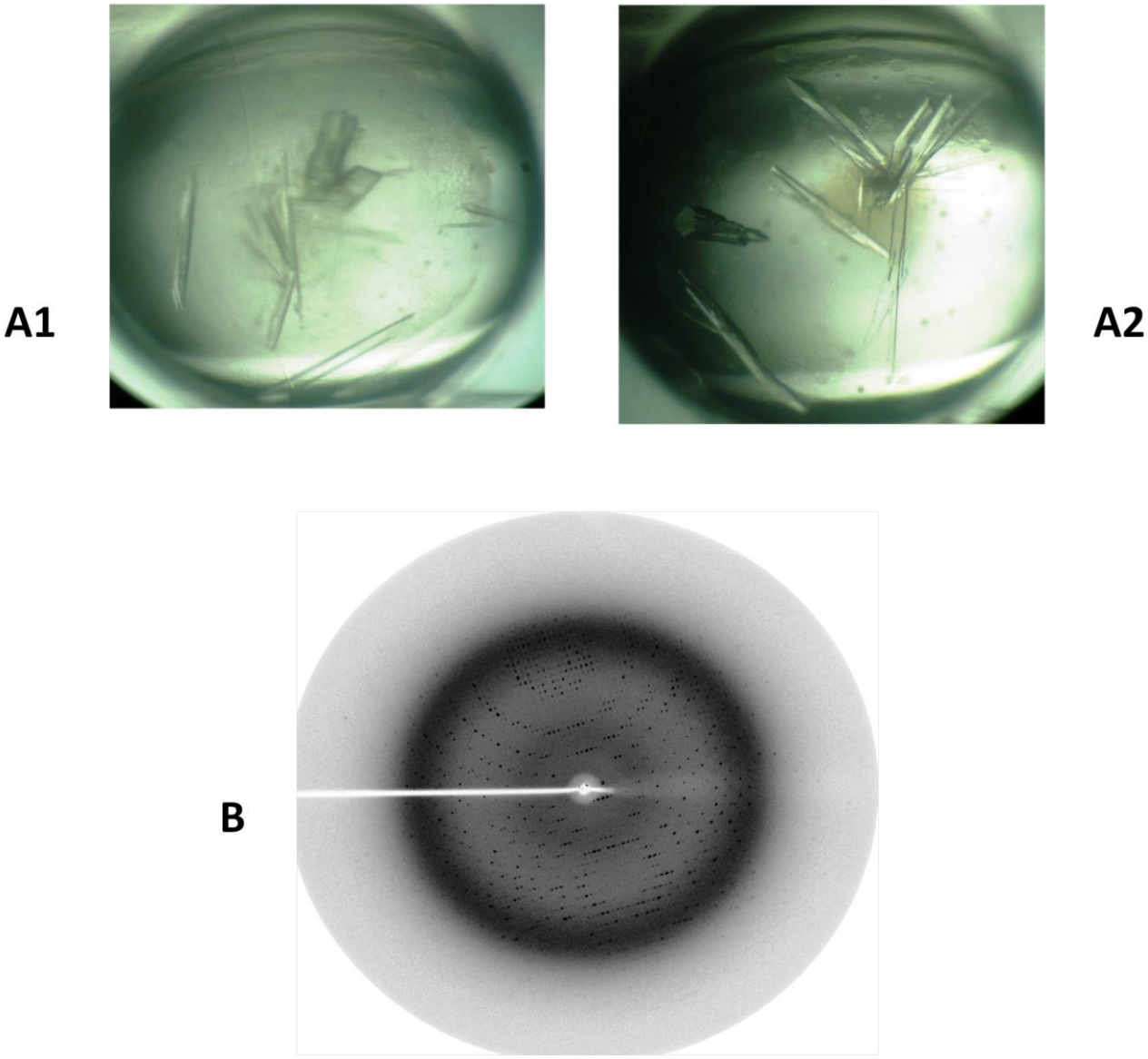


Figure A.4. The figure A1 is showing initial crystals, while the A2 is displaying the optimized crystals of FmtA; B is a diffraction pattern image of FmtA crystals.

Table A.1. Initial data collection statistics of FmtA have been summarized in table.

Space group	$C222_1$
Unit-cell parameters (Å)	a =105.4, b = 128.3 c = 107.6
Resolution range (Å)	50-2.9 (3.00-2.97)
Completeness (%)	97.6 (96.3)
$R_{sym}^{\dagger}$ (%)	8.9 (30.5)
Mean $I/\sigma(I)$	13.8 (2.9)
No. of observed reflections	538439
No. of unique reflections	122995 (6204)
Molecules per ASU	2
Mathews coefficient ($Å^3$ Da^{-1})	2.11
Solvent content (%)	41.87
Redundancy	4.3 (2.4)
Mosaicity (º)	1.2

(† $R_{sym} = \Sigma_{\mathbf{h}} \Sigma_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}} \rangle| / \Sigma_{\mathbf{h}} \Sigma_l \langle I_{\mathbf{h}} \rangle$, where I_l is the lth observation of reflection **h** and $\langle I_{\mathbf{h}} \rangle$ is the weighted average intensity for all observations l of reflection **h.** Values in parentheses are for the highest resolution shell (3.00-2.97)).

A.3.3. The preliminary crystallization profile of FmtA

The single crystal of FmtA was diffracted at home source with the maximum resolution of 3.0 Å. The initial data processing reveals that the crystal parameter follows orthorhombic crystal system having $C222_1$ space group. The cell parameter were observed as a =105.4 Å, b=128. Å, c=107.6 Å, with two molecules in an asymmetric unit, this corresponds to the crystal volume of 3.53 $Å^3$ Da^{-1} with the given molecular weight of 43 kDa of protein. The structure solution was attempted by molecular replacement. The statistics for data collection has been described in **Table A.1**.

Some other crystals were picked from same well and washed. The crystals were crushed and ran over SDS-PAGE. The gel was then stained by silver staining method to get the extra sensitive visualization of protein bands. The standard molecular weight marker has been employed for the assessment of corresponding molecular weight of FmtA. The results confirm that molecular weight of 43 kDa from crushed crystal sample and ascertained the homogeneity of FmtA (**Figure A.5**).

The preliminary structure has obtained from molecular replacement method using models from Protein Data Bank having identity number PDB ID: 4GDN as well as 2QMI. The results show that the FmtA exists in a dimeric form.

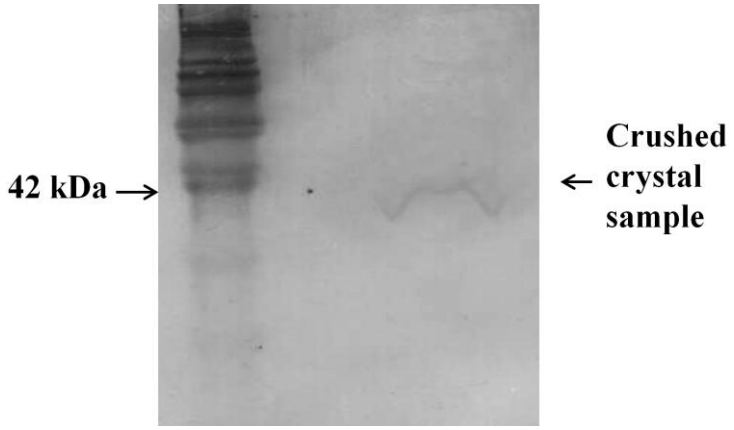


Figure A.5. The SDS-PAGE gel displaying the 43 kDa protein standard and corresponding sample of FmtA obtained by crushing the crystals used for diffraction analysis.

A.4. Conclusion

The FmtA protein, known for penicillin resistance has been targeted from *S. aureus*. There are no structures available for FmtA or related protein up to sufficient homology. (**Figure A.2**). Therefore, to get the first structure of methicillin resistance related protein, the crystallization has been achieved. The preliminary characterization of symmetry and systematic absences reveals that FmtA crystal belonged to $C222_1$ space group having two molecules in an asymmetric unit. The structure solution is in progress as search model is having low sequence identity. The heavy atom derivatives are also being made for isomorphous replacement method.