STRUCTURAL STUDIES ON TWO HEMAGGLUTININS FROM *CICER ARIETINUM* AND *MORINGA OLEIFERA*, AND A STUDY OF POLYMORPHISM IN THE CRYSTALS OF PLANT LECTINS

THESIS SUBMITTED TO

THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

BY

UMA V. KATRE

DIVISION OF BIOCHEMICAL SCIENCES
NATIONAL CHEMICAL LABORATORY
PUNE 411 008
INDIA

DECEMBER 2007

Dedicated to……..

Aai-Baba

&

Mama-Mami

CONTENTS

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Structural studies on two hemagglutinins from *Cicer arietinum* and *Moringa oleifera*, and a study of polymorphism in the crystals of plant lectins"** submitted by **Ms. Uma V. Katre** was carried out under my supervision. The material obtained form other sources has been duly acknowledged in the thesis.

Date

Dr. C. G. Suresh

(Research Supervisor)

Scientist

Division of Biochemical Sciences

National Chemical Laboratory

Pune 411 008

DECLARATION

I hereby declare that the thesis entitled **"Structural studies on two hemagglutinins from *Cicer arietinum* and *Moringa oleifera*, and a study of polymorphism in the crystals of plant lectins"** submitted by me to the University of Pune for the degree of Doctor of Philosophy is the record of original work carried out by me during the period from **06.08.2002** to **17.08.2007** under the supervision of **Dr. C. G. Suresh** at the Division of Biochemical Sciences, National Chemical Laboratory, Pune 411008 and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or Institution.

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

Date

Uma V. Katre

Division of Biochemical Sciences

National Chemical Laboratory

Pune 411008

ACKNOWLEDGEMENTS

During the last five years which I spent in NCL as a PhD student, many people have directly or indirectly helped me in my research work, or to make my stay in NCL a comfortable one. I will always remain grateful to them.

First and foremost, I wish to express my sincere thanks to my research supervisor, Dr. C. G. Suresh, for giving me an opportunity to work under him, for his able guidance and constant support. He always gave me freedom of everything- to think, to express, to work. More than anything, it was his trust in me which always made me work in the best possible way. I do not have enough words to express my gratitude towards him and I would always like to be indebted to him.

I express my gratitude towards Dr. Sushama Gaikwad, for her help and constant encouragement throughout my PhD work. I learnt not only the biochemical and biophysical techniques from her, but also a systematic way of conducting any experiment and to overcome the difficulties without giving up.

I wish to thank Dr. Sushama Gaikwad and Rohtas for providing me pure *Cicer arietinum* lectin whenever required.

I also wish to thank Dr. M. I. Khan for his keen interest in my research work and valuable suggestions throughout the work, also for being a member of my PhD committee. I am grateful to Dr. Vidya Gupta as well, for being in my PhD committee and for her valuable suggestions. Dr. Jayant Pal from the Department of Biotechnology, University of Pune is also acknowledged, for being in my assessment committee and for giving some useful suggestions. I am also grateful to Dr. Shama Barnabas for her concern in my research work. I thank all scientists in Division of Biochemical Sciences, for their help whenever required during my PhD course.

Thanks are also due to Dr. M. V. Badiger (Division of Polymer Sciences, NCL) for permission to use the fluorimeter and Prof. M. R. N. Murthy, Molecular Biophysics Unit, IISc for permission to use the CD facility. I thank Gayathri for

helping us to use the CD machine. Gayathri, Someshwar and other people from Prof. Murthy's group made our stay in IISc a comfortable one.

I got an opportunity to attend the CCP4 Study Weekend-2005 with the David Blow Memorial Fellowship. In this connection, I wish to thank the CCP4 team, Prof. M. Vijayan, Prof. James Naismith and Mrs. Maeri Howard. We had an opportunity to visit the leading crystallography laboratories in United Kingdom and I thank all the scientists in these labs for sharing their research experiences with us. I acknowledge Prof. Eleanor Dodson and Prof. Guy Dodson for making my stay in York a memorable one. Eleanor, with her keen interest in my research work, also helped me to analyze some of my datasets which gave me a new vision to look at the problems. Subsequently, I got an opportunity to attend a CCP4 workshop conducted in IISc, Bangalore and I thank all the tutors in the workshop for teaching us the use of computational crystallographic facilities.

I thank my seniors: Drs. Narasimha Rao, Anuradha, Manish, Priya, Suresh and Satya for their help in some way or the other. Poorva, she maintained a cheerful company in the lab as well as in the hostel, and was always with me even during my worst times. Raamesh and Nitin also helped me in many ways. Nishant has helped me during some part of my PhD work and I will be always grateful to him. Special thanks are due to Urvashi, for her help in many ways and being such a wonderful company. She is always open to ideas and ready to work on all the suggestions which I keep giving her time-to-time. I also wish to thank the short-term visitors of our lab, namely Gnanamani, Anita, Sindu, Avanti, Vidyut, Sherin, Varun, Chirag and Ashish. It was Chirag's help with which we could make a considerable progress on the analysis of polymorphism reported in this thesis.

Sharmili has been my best friend ever and it would be very formal if I thank her, though I should mention that she helped me to understand many concepts in biochemistry and helped me even otherwise, whenever needed. Also, she was always

there to listen to me whenever I felt very low. I wish to cherish our friendship for lifelong.

I wish to thank my friends in NCL for their help in some or the other ways: Anand, Sridevi, Sharath, Atul Fairenzi, Ambrish, Sachin, Ashwini Poopal, Sofia, Gayatri, Radhika, Gyanprakash, Aarohi, Anish, Ajit, Vinod, Anamika, Feroz, Sreekanth, Sajid, Ansari, Asad, Shabab, Jayprakash, Ashutosh, Shashi, Atul Thakur, Nagraj, Avinash, Sathish and Sarvesh. I also wish to mention the cheerful company of my hostel mates: Shweta, Meera, Indu, Maitri, Geetali, Suman, Shraeddha, Manaswini and Nitasha. Our tea-time conversations have been a refreshing entertainment.

I also wish to thank my BSc friends - Bhagyashree, Dhanashri, Shalaka, Kirti, Ketaki, Priya, Anjor, Shraddha and Manjiri, for being a wonderful company. Despite of our different professions and separation in terms of distance over the course of time, we still remain a group and share joys and sorrows of each other.

I acknowledge Ms. Amrita, Mrs. Indira, Mr. Mari and Ms. Satyali for their help in administrative work. Mr. Karanjkar and Mr. Trehan are acknowledged for their technical assistance.

I thank the Director, National Chemical Laboratory, and the Head, Division of Biochemical Sciences, for permitting me to carry out my PhD work in the Division of Biochemical Sciences, NCL and to submit this work in the form of the thesis. The financial assistance from the Council of Scientific and Industrial Research, India is also acknowledged.

It would be very formal to thank my family members and I have no intention to do that. Rather, I would like to always remain indebted to them.

-Uma

ABSTRACT

Many biological processes, such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilization, cancer metastasis and growth and differentiation involve interaction between carbohydrate molecules and proteins. Lectins or hemagglutinins are proteins that are involved in the carbohydrate recognition and the crystal structures of lectin-carbohydrate complexes turn out to be excellent model systems to study protein-carbohydrate interactions. About 42% of the structures reported in the 3-D lectin database (http://www.cermav.cnrs.fr/lectines/) belong to lectins from plants; the remaining ones are from animals, bacteria, fungi, algae and viruses.

Out of the several distinct families of plant lectins identified, the largest and the best-characterized family is from seeds of leguminous plants. One of the hemagglutinin studied in this thesis is from *Cicer arietinum* (chickpea), a member of the family leguminosae. This hemagglutinin shows specificity towards glycoproteins, like fetuin and desialated fetuin, but not towards any of the simple mono- or oligo-saccharides. It possesses an unusual combination of secondary structure elements for a legume lectin such as *α* helix, 34%, *β* sheet, 28% and random coil, 38%. X-ray crystallographic studies have been undertaken to determine the three-dimensional structure of this lectin.

A second hemaglutinin included in this study is from *Moringa oleifera* (drumstick) a plant cultivated in India to use its leaves, fruits, and roots as food and medicine. The hemagglutinin is isolated from the seeds of this plant. This also showed specificity towards glycoproteins like fetuin, desialated fetuin, holotransferin and thyroglobulin but not towards any of the simple mono- or oligo-saccharides. Purification and biochemical and biophysical characterization of this lectin have been carried out to reveal its structure-function relationship.

During our studies of a jacalin type lectin from *Artocarpus hirsuta* it was observed that the lectin crystallizes in different unit cells and space groups preserving its basic tetramer association in crystal structures and solution. Several other lectins also show this tendency to crystallize in different unit cells or space group, either from the same or a slightly different crystallization condition. Polymorphism in the crystals of organic substances driven by weaker atomic

interactions or through presence of various solvent molecules is known. Similarly different parts of the protein surface can interact with neighboring molecules in different ways leading to various arrangements in a crystal lattice. An analysis of the arrangement of protein in its different crystal structures will help to identify surface groups involved in intermolecular interactions and provide information on conformational flexibility of the molecule. Plant lectins, being a structurally well studied group of proteins, and the members show tendency to crystallize in different crystal forms, they have been selected to study patterns of aggregation leading to polymorphic crystals. A study of intermolecular contacts between the symmetry related molecules in the polymorphic crystals of plant lectins was carried out to gain insight into the patterns of protein aggregation.

The tetramer association of the lectin molecules of *Artocarpus hirsuta* not only provides structural stability for its biological function but also places the four sugar binding pockets towards four corners of an approximate tetrahedron so that the binding sites are kept maximum away from each other. Such an arrangement was observed in the crystal structures of many lectins and other functional proteins. These molecular arrangements were analyzed in terms of their structural features and functional advantages. In the analysis presented here the crystal structures of various homotetrameric proteins were studied with respect to their space groups and the association of tetramers so as to identify common patterns of molecular association.

The thesis has been organized into the following chapters:

Chapter 1: General Introduction

Chapter 2: Crystallization and X-ray crystallographic studies of a hemagglutinin from *Cicer arietinum* (Chickpea).

Chapter 3: Purification and biochemical characterization of a hemagglutinin from *Moringa oleifera* (drumstick).

Chapter 4: Structural studies on hemagglutinin from *Moringa oleifera* using fluorimetric methods and circular dichroism (CD).

Chapter 5: Analysis of polymorphism in the crystal structures of plant lectins.

Chapter 6: Features of homotetrameric molecular association in the crystals of lectins and other proteins.

Chapter 1: General Introduction

This chapter presents details of known lectins with reference to their occurrence, classification and applications with emphasis on structural aspects. Special emphasis is on plant lectins with respect to their classification, structure and carbohydrate binding properties.

Chapter 2: Crystallization and X-ray crystallographic studies of a hemagglutinin from *Cicer arietinum* (Chickpea).

A hemagglutinin with complex sugar specificity was isolated and purified by two successive ion exchange chromatography steps from the mature seeds of *Cicer arietinum* (chickpea). The protein crystallized from its solution (15 mg ml^{-1}) in deionized water using hanging-drop vapor-diffusion technique where the well solution consisted of 0.1 M sodium cacodylate buffer pH 6.5, 0.2 M sodium acetate and 30 % w/v polyethylene glycol 8000. These crystals diffracted to a resolution of 2.3 Å and belonged to the space group *R*3. Unfortunately, these crystals were found to be twinned. To overcome the problem of twinning, the quality of the crystals was improved by using precipitant solution containing 0.1 M sodium cacodylate buffer at pH 6.5, 0.2 M sodium acetate and 10-15 % polyethylene glycol 8000. The crystals belonged to two different space groups, trigonal (*P*3) with unit cell dimensions *a*=*b*=80.21, *c*=69.14 Å and orthorhombic ($P2_12_12$) with unit cell dimensions *a*=70.93, *b*=73.32, *c*=86.98 Å. *N*-terminal sequence of CAL showed 90% identity with the *N*-terminal sequence of a major seed albumin (PA-2) from *Pisum sativum*. Since no three-dimensional structure of this albumin is available molecular replacement method could not be tried. Thus, it is required to solve the structure using multiple isomorphous replacement method. Crystals were grown in the presence of various heavy atom salts, namely KI, $HAuCl_4$, $Pb(NO_3)_2$, *p*-Hydroxy mercury benzoate and dichloro(ethylenediamine)platinum (II). Very often, the crystallization in presence of heavy atom or the soaking of crystals in heavy atom solutions destroyed the quality of diffraction. Only in the case of iodine the presence of heavy atom had some positive effect. Thus, only iodine derivative could be characterized for the presence of heavy atom.

Chapter 3: Purification and biochemical characterization of a hemagglutinin from *Moringa oleifera* (drumstick).

A hemagglutinin from the matured seeds of *Moringa oleifera* (MoL) was purified to homogeneity by ammonium sulfate precipitation (90% saturation) and two successive ion exchange chromatography steps using DEAE cellulose and CM Sephadex C50. The yield of the protein was about 800 mg per 100 g of the seeds with a final specific activity of 160 HU / mg. The native protein is a homotetramer with molecular mass 26 kDa as determined by gel filtration chromatography on Sephadex G-50 and subunit molecular weight 6.5 kDa determined by SDS-PAGE in the presence of 2-mercaptoethanol. MoL has a pI of 10 and is stable in the entire pH range 1-12, up to 24 hours. The protein retains its hemagglutination activity even after incubating for 30 minutes at 85^0C and is inhibited only by glycoproteins such as thyroglobulin, fetuin and holotransferin. The hemagglutination activity was lost in the presence of 0.25 M GdnHCl or urea more than 3M concentration. The *N*-terminal sequence of MoL determined is APGIMYRVQR. Similarly, the estimated amino acid composition shows that it contains number of residues per monomer as follows: Glx (15), Arg (11), Pro (6), Gly (6), Ala (5), Leu (5), Thr (4), Val (3), Ile (3), Asx (2), Ser (2), Phe (2), His (2), Tyr (1) and Met (1). Only a single tryptophan per monomer is detected in the protein, by both native and denatured conditions. No free cysteines were detected; however, three disulfide linkages detected per dimer indicated presence of 3 cysteine residues per monomer. Chemical modification studies indicated that none of the residues such as arginine, histidine, aspartic acid, glutamic acid, cysteine, tyrosine, tryptophan or serine is individually important for the hemagglutination activity of MoL.

Chapter 4: Structural studies on the hemagglutinin from *Moringa oleifera* by fluorimetric methods and circular dichroism (CD).

To study the microenvironment of the single tryptophan in MoL, solute quenching studies were carried out at pH 1.0, 7.2 and 10.0 and in the presence of 6 M urea using quenchers such as acrylamide, potassium iodide and cesium chloride. The results indicate that tryptophan is highly exposed to the solvent and present in a strongly electropositive environment. The intrinsic fluorescence of the protein was quenched upon binding sugars LacNAc (with K_a=1380M^{-1}) and

fructose (K_a=878M^{-1}), although these sugars did not inhibit the hemagglutination activity of MoL. The binding of hydrophobic dye 8-anilino-1-naphthalene sulfonic acid (ANS) using fluorescence spectroscopy shows exposed hydrophobic patches in protein that get further exposed at extreme acidic or alkaline pH but get buried in the interior in presence of 1M GdnHCl or urea. MoL also binds adenine (K_a = 7.76 x 10^3 M^{-1}). Time resolved fluorescence of the native protein shows two lifetimes indicating two different conformers of tryptophan which get merged into single one after quenching with 0.15M acrylamide.

Analysis of the far-UV CD spectrum of MoL showed secondary structure composition of *α*R (regular *α*-helix) 16%, *α*D (distorted *α*-helix) 12%, *β*R (regular *β*-sheet) 14%, *β*D (distorted *β*-sheet) 9%, turn 20% and unordered 28%. The secondary structure is not affected by extreme acidic or alkaline conditions; however, it is drastically affected by the presence of a reducing agent like dithiothreitol (DTT) (1mM) at and above pH 7.0. The far UV CD spectra of the protein incubated in the presence of different concentrations of GdnHCl shows no change in the ellipticity in the region 215 to 225 nm, although it is affected at 210 nm and less indicating an increase in the unordered structural element. The secondary structure is stable up to 3M urea concentration above which it starts changing.

Chapter 5: Analysis of polymorphism in the crystal structures of plant lectins

Many lectins, including *Cicer arietinum* hemagglutinin studied here, crystallize in different space groups or in the same space group with different unit cell dimensions. The analysis present here is carried out to understand the similarity and differences in the packing environments of molecules in the polymorphic crystals of plant lectins. A total of 236 plant lectin structures have been analyzed. Concanavalin A is the lectin that shows maximum extent of polymorphism (15 polymorphs from 47 crystal structures), whereas *Pterocarpus angolensis* lectin shows least number (2 polymorphs from 23 crystal structures).

As expected the polymorphs with any common unit cell length have similar intermolecular contacts and packing arrangement along that particular cell axis. The intermolecular contacts that occur repeatedly in the lattice of different polymorphic crystal forms of the same protein might have important role in its agglutination activity and oligomerisation in the solution.

Chapter 6: Features of homotetrameric molecular association in the crystals of lectins and other proteins.

The crystal structures of proteins with homotetrameric association, a common feature of many lectins, were analyzed to understand the characteristics of tetrameric association in terms of the arrangement of their subunits and its biological significance. The analysis could group the tetramer units into four categories:

1. Tetrahedral molecules, in which the four monomers form a nearly perfect tetrahedral arrangement. The angle between axes of any two monomers is ~ 109^0. Sometimes, the tetrahedral shape is distorted i. e. the tetrahedral angle deviates from ideal value, which gives the molecule a twisted shape rather than a perfect tetrahedral shape.
2. Molecules that form sandwiched dimer of dimers where the two dimers are arranged perpendicular to each other, one upon the other.
3. Planar molecules where the four monomers lie in one plane and the corresponding sides of adjacent monomers face opposite directions (that makes diagonally opposite monomers to face the same direction). This can be considered a flattened tetrahedral shape.
4. Planar closed molecules where all the four monomers lie in one plane arranged in a head-to-tail fashion in a square.

The first group is the most commonly found arrangement. The importance of each arrangement for its biological function is discussed.

List of publications

1. K. N. Rao, C. G. Suresh, U. V. Katre, S. M. Gaikwad and M. I. Khan. (2004). Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose. *Acta Crystallographica* D60, 1404-1412.
2. Uma V. Katre, S. M. Gaikwad, S. S. Bhagyawant, U. D. Deshpande, M. I. Khan and C. G. Suresh. (2005) Crystallization and preliminary X-ray characterisation of a lectin from *Cicer arietinum* (chickpea). *Acta Crystallographica* F61, 141-143.
3. Uma V. Katre, C. G. Suresh, M. I. Khan and Sushama M. Gaikwad (2007). Structure-activity relationship of a hemagglutinin from *Moringa oleifera* seeds. *International Journal of Biological Macromolecules*. (Manuscript accepted for publication).
4. Uma V. Katre, C. G. Suresh, M. I. Khan and Sushama M. Gaikwad (2007). Steady state and time-resolved fluorescence studies of a hemagglutinin from *Moringa oleifera*. *Journal of Fluorescence*. (Manuscript accepted for publication).
5. Uma V. Katre and C. G. Suresh (2007). Polymorphism in the crystals of plant lectins: Polymorphism of Concanavalin A. *Manuscript under preparation*.

List of abbreviations

AMoRe	Automated Molecular Replacement
AS	Ammonium sulfate ($(NH_4)_2SO_4$)
BME	2-mercaptoethanol (β-mercaptoethanol)
BSA	Bovine serum albumin
CAL	*Cicer arietinum* lectin
CCD	Charge-coupled device
CCP4	Colloborative Computational Project No. 4
CD	Circular Dichroism
CD	Conserved domain
COM	Center of Mass
ConA	Concanavalin A
CPB	Citrate-phosphate buffer
CRD	Carbohydrate Recognition Domain
DDW	Double distilled water
DEAE	Diethylaminoethyl
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DTNB	5,5'-Dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FFT	Fast Fourier Transform
FOM	Figure of merit
Gal	Galactose
GalNAc	N-acetyl-D-galactosamine
GdnHCl	Guanidium hydrochloride
GlcNAc	N-acetyl-D-glucosamine
GNA	*Galanthus nivalis* agglutinin
h	Hour
HPLC	High performance liquid chromatography
HX	Hemopexin
IEF	Isoelectric focusing
IP	Image plate
LacNAc	N-acetyl-D-lactosamine
LacR	Lactose operon repressor protein
MAD	Multi-wavelength anomalous dispersion
Min	Minute
MIR	Multiple isomorphous replacement
MoL	*Moringa oleifera* lectin
MPA	*Maclura pomifera* agglutinin
MPD	2-methyl-2,4-pentanediol
MR	Molecular replacement
MW	Molecular weight
NAI	*N*-acetyl imidazol
NBS	*N*-Bromosuccinimide
NCBI	National Center for Biotechnology Information
NCS	Non-Crystallographic Symmetry
NMR	Nuclear Magnetic Resonance
OD	Optical density
OD_{280}	Optical density at 280 nm

PA-2	Pea albumin 2
PAG	Polynucleotide adenosine glycosidase
PAGE	Polyacrylamide gel electrophoresis
PB	Phosphate buffer
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
PGDH	D-3-Phosphoglycerate Dehydrogenase
PMSF	Phenylmethylsulfonyl fluoride
PNA	Peanut agglutinin
RBC	Red blood cells
RIP	Ribosome inactivating protein
RNA	Ribonucleic acid
SCAfet	Fetuin-binding lectin from *Scilla campanulata* bulbs
SCAman	Mannose-binding lectin from *Scilla campanulata* bulbs
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SP	Sulfopropyl
TIMP	Tissue inhibitor of metalloproteinases
TNS	2,6-toludinylnaphthalenesulfonic acid
WGA	Wheat germ agglutinin
WRK	Woodward’s reagent K

Chapter 1

General Introduction

1.1. Genesis of the thesis

Recognition of carbohydrates by proteins is of prime importance in many biological processes, such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilization, cancer metastasis and growth and differentiation. Lectins or hemagglutinins are proteins involved in carbohydrate recognition. The crystal structures of the lectin-carbohydrate complexes turn out to be excellent model systems to study protein-carbohydrate interactions. Although lectins are known to occur in all living organisms, plant lectins were the first proteins of this class to be studied. Because of their broad distribution and ease of isolation, more lectins have been characterized from plants than from any other source (Sharon and Lis, 1989). About 42% of the structures reported in the 3-D lectin database (http://www.cermav.cnrs.fr/lectines/) belong to lectins from plants; the remaining ones are from animals, bacteria, fungi, algae and viruses. Out of the several distinct families of plant lectins identified, the largest and the best-characterized family is from seeds of leguminous plants. In this thesis, the author has presented the study on three related aspects of the structure and properties of lectins. Firstly, the X-ray crystallographic studies on a poorly soluble hemagglutinin occurring in the seeds of *Cicer arietinum* (Chickpea) which is a member of the family Leguminosae have been undertaken. Secondly, the characteristics of a highly soluble hemagglutinin from the seeds of *Moringa oleifera*, a plant cultivated in India for its use as food and medicine, have been studied. Purification and biochemical and biophysical characterization of this lectin have been carried out to reveal its structure-function relationship and ultimately crystallize this protein.

During our studies of a jacalin type lectin from *Artocarpus hirsuta* it was observed that the lectin crystallizes in different unit cells and space groups preserving its basic tetramer association either in crystal structures or in solution. Several other lectins also

show this tendency to crystallize in different unit cells or space group, either from the same or a slightly different crystallization condition. Incorporation of a macromolecule in a crystal lattice might modify its structure relative to that in the solution. The determination of a protein structure in different crystal forms can reveal conformational changes or other manifestations of dynamic behavior that may be biologically relevant. Plant lectins, being a structurally well studied group of proteins; provide a diverse example to compare the structures of closely related sets of proteins under different crystallization conditions in different crystal packing environments. A study of intermolecular contacts between the symmetry related molecules in the polymorphic crystals of plant lectins was carried out to gain insight into the patterns of protein aggregation.

In case of the *Artocarpus hirsuta* lectin it was observed that the tetramer association of the lectin molecules not only provides structural stability for its biological function but also places the four sugar binding pockets towards four corners of an approximate tetrahedron so that the binding sites are kept maximum away from each other. Such an arrangement was observed in the crystal structures of many lectins and other functional proteins. An analysis of various homotetrameric proteins was carried out with respect to the space groups and the association of tetramers so as to identify common patterns of molecular association. Thus, thirdly the polymorphism in the crystal structures of plant lectins was probed. Associated with that the quaternary associations of lectins and other proteins were also studied.

1.2. Lectins

The history of lectins dates back to 1888 when Stillmark discovered a toxic proteinaceous hemagglutinating factor in castor beans (*Ricinus communis*) which was named as “ricin” (Stillmark, 1888). In 1898, Elfstrand introduced for the first time the term

'Blutkörperchenagglutinin' (hemagglutinin) as a common name for all plant proteins that cause clumping of cells (Elfstrand, 1898). Landsteiner and Raubitschek (1907) reported for the first time the presence of nontoxic lectins in the legumes *Phaseolus vulgaris* (bean), *Pisum sativum* (pea), *Lens culinaris* (lentil), and *Vicia sativa* (vetch). Many more non-toxic plant hemagglutinins were discovered subsequently. Specificity of certain hemagglutinins toward erythrocytes of a particular human blood group within the ABO system was also established in the due course (Renkonen, 1948; Boyd and Reguera, 1949). This discovery of blood group specificity was the direct motive to the introduction of the novel term 'lectin' (from the Latin verb 'legere', which means 'to select') (Boyd and Shapleigh, 1954).

Several definitions have been coined for the term "lectin" in the course of time. According to the first proper definition, which was based primarily on the sugar specificity and inhibition of the agglutination reaction, lectins were defined as "carbohydrate-binding proteins (or glycoproteins) of non-immune origin that agglutinate cells and/or precipitate glycoconjugates" (Goldstein *et al.*, 1980). However, this definition was confined to multivalent carbohydrate-binding proteins. To overcome the shortcomings of this definition, Kocourek and Horejsi (1983) proposed a modified version of the definition as "Lectins are proteins of nonimmunoglobulin nature capable of specific recognition and reversible binding to carbohydrate moieties of complex carbohydrate without altering the covalent structure of any of the recognized glycosyl ligands". However, when it was observed that some lectins contain a second type of binding site that interacts with noncarbohydrate ligands, the definition proposed by Kocourek and Horejsi was also found to be too restrictive. Hence, lectins were redefined as "carbohydrate-binding proteins other than antibodies or enzymes" (Barondes, 1988). Later the discovery of some plant enzymes (like the type 2 ribosome-inactivating proteins [RIP] and the class I chitinases) which are fusion proteins built up of a carbohydrate-binding domain tandemly

arrayed with a catalytic domain implied that the definition of lectins cannot exclude all enzymes. Thus, the lectins are presently defined as “all proteins possessing at least one non-catalytic domain, which binds reversibly to a specific mono- or oligosaccharide” (Peumans and Van Damme, 1995).

Based on the overall structure of the mature lectins, they are subdivided into the following categories (Peumans and Van Damme, 1995):

1. Merolectins

They consist of a single carbohydrate-binding domain. By definition they are monovalent and hence cannot precipitate glycoconjugates or agglutinate cells. Hevein, the small chitin-binding protein from the latex of the rubber tree (*Hevea brasiliensis*) (Van Parijs *et al.*, 1991) is a typical merolectin.

2. Hololectins

Although they also are exclusively built up of carbohydrate-binding domains, they contain at least two such domains which are either identical or very homologous and bind either the same or structurally similar sugar(s). Hololectins are di- or multivalent and hence agglutinate cells and/or precipitate glycoconjugates. Most plant lectins belong to the subgroup of hololectins (Van Damme *et al.*, 1998a).

3. Chimerolectins

These are fusion proteins consisting of one or more carbohydrate-binding domain(s) tandemly arrayed to an unrelated domain. The unrelated domain can have well defined biological activity and functions independently of the carbohydrate binding domain. Chimerolectins can behave as merolectins (for example class 1 plant

chitinases) or hololectins (for example type 2 RIP) depending upon the number of carbohydrate binding domains.

4. Superlectins

They consist exclusively of at least two carbohydrate-binding domains, which recognize structurally unrelated sugars. Hence superlectins can also be considered a special group of chimerolectins composed of two tandemly arrayed structurally and functionally different carbohydrate-binding domains.

Yet another classification of lectins, based on their molecular structure, was introduced by Sharon and Lis in 1998. In this classification, lectins are grouped into three classes: simple lectins, mosaic (or multidomain) lectins and macromolecular assemblies.

A. Simple lectins

They consist of a small number of subunits, of molecular weight usually below 40 kDa, which may not be necessarily identical and may contain an additional domain besides their carbohydrate binding site(s). Practically all known plant lectins as well as the galectins (formerly S-lectins), a family of galactose-specific animal lectins, come under this class.

B. Mosaic (or multidomain) lectins

These are all composite molecules with a wide range of molecular weights, consisting of several kinds of protein modules or domains, only one of which possesses a carbohydrate binding site. This group includes diverse proteins from different sources, like viral hemagglutinins and animal lectins of the C-, P-, and I-

type. Although many of these lectins are monovalent, since they are embedded in membranes, they act in a multivalent fashion.

C. Macromolecular assemblies

These lectins are commonly seen in bacteria in the form of fimbriae or pili. These are filamentous, heteropolymeric organelles which consist of helically arranged subunits of several different types, assembled in a well defined manner (Ofek and Sharon, 1990; Ofek and Doyle, 1994; Gaastra and Svennerholm, 1996). The bulk of the fimbrial filament (shaft) is made up of polymers of the major subunit, which thus plays a structural role. Only one of the subunits, usually a minor component of the fimbriae, possesses a carbohydrate combining site and is responsible for the binding activity and sugar specificity of the fimbriae.

Although lectins were first isolated from plants, soon the presence of these proteins outside plants was demonstrated. They are ubiquitous in occurrence, being found in all types of living organisms as well as viruses. A brief account of the lectins occurring in different types of organisms and their roles is given.

1.2.1. Viral lectins

The interaction of influenza viruses with their target cells provides the oldest and perhaps best characterized example of lectin-carbohydrate recognition system (Paulson, 1987). Human influenza virus binds to erythrocytes and other cells by recognizing *N*-acetyl neuraminic acid present on the cell surface and this binding is a prerequisite for initiation of infection. Influenza A and B viruses and paramyxoviruses bind gangliosides containing *N*-acetyl neuraminic acid, whereas influenza C and bovine corona viruses bind receptors containing *N*-acetyl-9-*O*-acetyl neuraminic acid, the 9-*O*-acetyl group of

which is critical for mediating cellular attachment (Schultze *et al*., 1993).

The detailed knowledge of the sialic acid-hemagglutinin interaction provides a possible basis for the design of antiviral drugs that would block viral attachment to cells. An inhibitor targeted to the conserved amino acids of the combining site or to the part of the cellular receptor essential for interaction with the hemagglutinin might be effective against influenza viruses of all subtypes. It would be independent of the antigenic changes that accompany the recurrent epidemics for which these viruses are renowned (Sharon and Lis, 1989).

1.2.2. Bacterial lectins

Many bacterial species, especially enterobacteria and *Salmonella* species are capable of producing surface lectins, in the form of submicroscopic hair-like appendages known as fimbriae (pili), present on the surface of the cells (Sharon, 1987). Fimbriae are usually 5 to 7 nm in diameter and 100 to 200 nm in length. Several types of fimbriae have been characterized. Type 1 fimbriae from *E. coli* are mannose specific and preferentially bind oligomannose and hybrid oligosaccharides of animal cell surface glycoproteins, while P fimbriae from *E. coli* interact specifically with glycolipids containing Galα1,4Gal. S fimbriae of *E. coli* are specific for NeuAcα2,3Gal, and type 2 fimbriae of oral actinomycetes, specific for Galβ1,3GalNAc (Sharon and Lis, 1989).

Bacterial surface lectins play a key role in the initiation of infection by mediating bacterial adherence to epithelial cells of the host, for example, in the urinary and gastrointestinal tracts. The fimbriated strains of *E. coli* and *Klebsiella pneumoniae* are more infective than their non-fimbriated counterparts (Sharon, 1987).

1.2.3. Lectins occurring in protozoa

Among the numerous protozoa that infect humans and animals, the occurrence of lectins has been well-studied in the pathogenic amoeba *Entamoeba histolytica*, which causes dysentery in humans by disruption and invasion of the colonic mucosa. Inhibition of amoebic infection by different saccharides suggests the role of lectin-sugar interactions in amoebic adherence. Among the two lectins isolated and characterized from *E. histolytica*, one is specific for β-1,4 linked oligomers of GlcNAc (Kobiler and Mirelman, 1981) and the other for Gal and GalNAc (Petri *et al*., 1989).

1.2.4. Lectins occurring in slime molds

Aggregation of slime molds is a key event in the differentiation of these organisms from their single-cell, vegetative form to an aggregated form, and is an example of involvement of lectins in cell-cell recognition. *Dictyostelium discoideum* produces a developmentally regulated galactose-specific lectin, discoidin I. The lectin is present on the surface of aggregating cells, and in its isolated form it agglutinates aggregating slime mold cells however, not vegetative cells (Sharon and Lis, 1989).

1.2.5. Fungal lectins

In fungi, lectins are known to occur in mycelium (Candy *et al.*, 2003), conidia (Tronchin *et al.*, 2002), sporomes (Raszeja, 1958), basidiomes (Konska, 1985) and fruiting bodies (Wang *et al.*, 2003). Fungal lectin are reported to participate in the formation of primordia, creation of mycelium structures to facilitate, penetration of parasitic fungi into the host organism as well as mycorrhization (Konska, 2006, Guillot and Konska, 1997).

1.2.6. Animal lectins

A variety of lectins are produced in animals, both intracellular and extracellular. The intracellular lectins are classified into four families, namely calnexin family, M-type, L-type and P-type. They are located in luminal compartments of the secretory pathway and function in the trafficking, sorting and targeting of maturing glycoproteins. The extracellular lectins are classified as C-type, R-type, siglecs and galectins. These are either secreted into the extracellular matrix or body fluids, or localized to the plasma membrane, and mediate a range of functions including cell adhesion, cell signalling, glycoprotein clearance and pathogen recognition (http://www.imperial.ac.uk/research/animallectins/default.html).

Lectins found in invertebrate animals are mainly present in the hemolymph and sexual organs. These animals solely rely on innate immunity for defense against microbial infections in which lectins are known to play an important role (Vasta *et al.*, 2004).

1.2.7. Plant lectins

Lectins occurring in plants differ in their molecular structure, biochemical properties, and carbohydrate-binding specificity, hence are considered a complex and heterogeneous group of proteins (Van Damme *et al.*, 1998a). Apart from seeds, plant lectins occur in virtually all types of vegetative tissues, such as leaves, bark, stems, rhizomes, bulbs, and tubers (Etzler, 1985 and 1992; Peumans and Van Damme, 1998 and 1999). In seeds, lectins constitute about 1-10%, sometimes even higher (up to 50%) part of the total seed proteins. Vegetative tissues (for example, bark, bulbs, tubers, rhizomes and corms) also contain 1-20% of total proteins as lectins (Peumans and Van Damme, 1998). The lectin content of both seeds and vegetative tissues is often

developmentally regulated and exhibit seasonal variations in their concentrations (Peumans and Van Damme, 1995).

Plant lectins can be classified based on different characteristics. According to one scheme, they are classified based on their carbohydrate-binding specificity, into several specificity groups. Using this criterion, plant lectins have been distinguished as mannose-, mannose/glucose-, mannose/maltose-, Gal/GalNAc-, GlcNAc/$(GlcNAc)_n$-, fucose-, and sialic acid binding lectins (Goldstein and Poretz, 1986; Van Damme *et al.*, 1998b). Although this subdivision is helpful for the use of lectins as tools, it is artificial and does not consider the possible evolutionary relationships between lectins (Van Damme *et al.*, 1998a). Hence another subdivision was proposed based on the analysis of the amino acid sequence data. According to this classification, most of the currently known plant lectins can be divided into seven families of evolutionary and structurally related proteins (Goldstein and Poretz, 1986; Van Damme *et al.*, 1998a). These families include: the legume lectins, the monocot mannose-binding lectins, the chitin-binding lectins composed of hevein domains, the type 2 RIP, the jacalin related lectins, the amaranthin lectin family, and the Cucurbitaceae phloem lectins. A brief review of each of these families with an emphasis towards the salient features of their three-dimensional structures is presented below.

I. Legume lectins

This is the largest and best characterized family of plant lectins. Historically, most of the pioneering work in the field of biochemistry, physiology, and molecular biology of plant lectins has been achieved with legume lectins. Concanavalin A (ConA) was the first plant lectin to be purified and crystallized (Sumner and Howell, 1936) and was also the first lectin whose primary structure and three-dimensional structure were resolved (Edelman *et al.*, 1972; Hardman and Ainsworth, 1972). The first plant lectin gene to be

sequenced was that encoding for the soybean seed lectin. (Vodkin *et al.*, 1983).

Legume lectins occur exclusively in plants from the family Leguminosae (Fabaceae), although not all lectins found in plants from this family belong to the legume lectins (Peumans and Van Damme, 1998; Van Damme *et al.*, 1998a). Most legume lectins have been isolated from seeds, found to be sequestered in storage protein vacuoles (Etzler, 1986). Several legumes contain two or more different seed lectins (for example, *Phaseolus vulgaris*). Some of the legume lectins have also been found in vegetative tissues like leaves, stems, bark, roots and root nodules (for example, bark lectins from *Robinia pseudoacacia*, leaf lectins from *Griffonia simplicifolia*) (Van Damme *et al.*, 1998a).

Legume lectins are dimeric or tetrameric proteins, each subunit (25-30 kDa) consisting of one carbohydrate binding site. Most of them are “single chain proteins” (for example, ConA); while in some, this single chain is broken into two smaller peptides (for example, lectins from Vicieae tribe). Tightly bound Ca^{2+} and Mn^{2+} (or another transition metal) ions are necessary for the interaction of lectin with carbohydrates (Sharon and Lis, 1990). Amino acid residues involved in metal ion binding are found to be highly conserved in all legume lectins.

A wide range of carbohydrate-binding specificities is observed in legume lectins. This is in contrast to their sequence similarity and close evolutionary relationship. However, it has been demonstrated that substitutions of amino acids involved in carbohydrate-binding as well as variations in the length of a particular loop change the structure of the binding site without affecting the overall three-dimensional structure of the monomer (Young and Oomen, 1992; Sharma and Surolia, 1997).

All legume lectins show a very similar three-dimensional structure. The front face of the monomer is built from a curved seven-stranded β-sheet, whereas the back face is formed from a flat six-stranded β-sheet (Fig. 1.1. (A)). They are interconnected by turns

and loops and form flattened dome shaped structure, which is known as the “jelly-roll motif”, commonly found in viral coat proteins. Four loops located at the upper part of the dome form the monosaccharide binding site. Legume lectins contain almost no α-helix; hence belong to the class of all beta-proteins.

Dimeric lectins are found mainly in Vicieae tribe and are bivalent, whereas most other legume lectins form tetramers having four carbohydrate-binding sites. Two monomers associate non-covalently to form a “canonical dimer” (Fig. 1.1. (B)); two such dimers associate through their back walls in such a way that two monosaccharide-binding sites occur at both sides of the tetramer (Fig. 1.1. (C)). This gives rise to a 222 (D2) symmetry in the tetrameric molecule and in some cases is responsible for a perfect or nearly-perfect tetrahedral arrangement of subunits, carbohydrate-binding sites occurring at the four corners of the tetrahedron. An exception to this is the quaternary structure of peanut agglutinin (PNA; Fig. 1.1. (D)) which has neither a 222 (D2) nor a four-fold (C4) symmetry (Banerjee *et al.*, 1994).

Apart from binding to the carbohydrate ligands, legume lectins are also known to bind hydrophobic molecules like 1,8-anilinonaphthalenesulfonic acid (ANS), 2,6-toludinylnaphthalenesulfonic acid (TNS), adenine and phytohormones like cytokinin (Roberts and Goldstein, 1983a, b). These hydrophobic binding sites are distinctly different from sugar-binding sites. In addition, some lectins, for example ConA and PNA are found to bind porphyrins, three-dimensional structures of which are also available (Goel *et al.*, 2001 and 2005) as well as bind certain designed peptides which act as carbohydrate mimetics (Kaur *et al.*, 1997; Jain *et al.*, 2000).

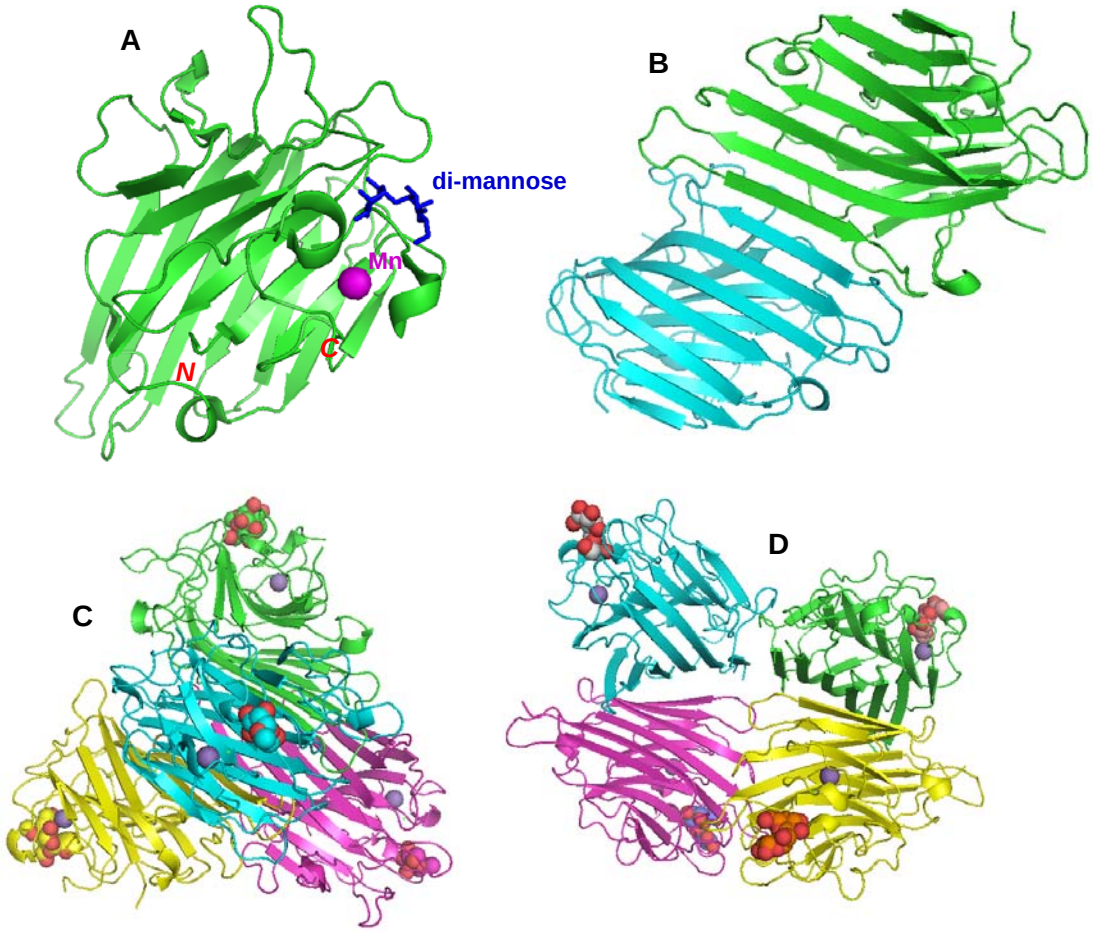


Fig. 1.1. (A) The crystal structure of ConA (PDB code 1I3H) displaying a monomer with jelly-roll motif. Two mannose molecules are bound to the carbohydrate-binding site defined by four loops. A manganese atom bound to the molecule is also shown. ***N*** and ***C*** denote the *N*- and *C*- terminals respectively.

(B) A canonical dimer of ConA (PDB code 1APN). Two monomers are associated non-covalently by formation of an antiparallel β-sheet through intermolecular strands.

(C) The near-perfect tetrahedral arrangement of subunits as seen in the quaternary structure of ConA (PDB code 1BXH).

(D) The unusual quaternary structure of peanut lectin (PDB code 1BZW).

These and all subsequent figures from PDB files were prepared using PyMOL (DeLano, 2002).

Although legume lectins are considered to be mediators of symbiosis in plants and microorganisms (Diaz *et al.*, 1989; Brewin and Kardailsky, 1997), until now there is no conclusive evidence that legume lectins play a deterministic role in this process. Due to their abundance in seeds and other storage tissues, legume lectins are considered to be storage proteins. Because of their preferential specificity toward typical animal glycans, it has been also proposed that legume lectins play a role in the plant's defense against insects and/or predating animals (Chrispeels and Raikhel, 1991; Peumans and Van Damme, 1995). Although they are known to bind hydrophobic ligands, direct evidence for the binding of auxins, gibberlins or other plant growth factors to lectins *in vivo*, which may be correlated to their role in plant development, is still not available (Komath *et al.*, 2006).

II. Monocot mannose-binding lectins

These lectins are a superfamily of mannose-specific lectins which are mostly confined to a subgroup of monocots. So far, these lectins have been found in six different monocot families, namely, Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Liliaceae, and Orchidaceae (Van Damme *et al.*, 1998a). They are reported to occur in various vegetative tissues such as leaves, flowers, ovaries, bulbs, tubers, rhizomes, roots (Van Damme *et al.*, 1995), and even in nectar (Peumans *et al.*, 1997a), however, rarely in seeds.

Galanthus nivalis agglutinin (GNA) was the first lectin of this family to be crystallized (Wright *et al.*, 1990) and analyzed by X-ray diffraction (Hester *et al.*, 1995). The three-dimensional structure of this lectin shows a β-prism II fold, a characteristic of this family. The monomer consists of three tandemly arrayed subdomains (I, II, and III) each of which consists of a four-stranded β-sheet. Each of the subdomain has a carbohydrate recognition domain (CRD). The three subdomains have a local three-fold

symmetry, and form three faces of a prism. They are connected by loops and form a 12-stranded β-barrel which exhibits three mannose-binding sites located in the clefts formed by the three bundles of β-sheet. Two such monomers form tight dimers through hydrogen-bond contacts stabilized by *C*-terminal strand exchange. These dimers further associate into tetramers mainly through hydrophobic interactions. Thus, the tetrameric GNA molecule has 12 mannose-binding sites (Fig. 1.2. (A)). Other tetrameric mannose-binding lectins, namely those from daffodil (*Narcissus pseudonarcissus*) (Sauerborn *et al.*, 1999) and bluebell (*Scilla campanulata*; SCAman) (Wood *et al.*, 1999) show a similar three-dimensional structure. Although these lectins are specific towards mannose, they also recognize mannose containing glycoproteins due to which they are able to show inhibitory activity on the *in vitro* replication of retroviruses (Balzarini *et al.*, 1991 and 1992). The specificities for the complex glycans originate from the oligomerisation of the molecule. While the tetrameric snowdrop, daffodil and bluebell lectins bind a surface glycoprotein on HIV, gp120, with a high degree of affinity, the dimeric garlic lectin (Fig. 1.2. (B)) cannot (Vijayan and Chandra, 1999).

Apart from mannose-specific lectins, certain lectins with complex-sugar specificity have also been reported to occur in some monocots, for example *N*-acetyl-D-lactosamine (LacNAc) specific lectins from *Arisaema flavum* (Singh *et al.*, 2004); *Alocasia cucullata* (Kaur *et al.*, 2005a); *Arisaema tortuosum* (Dhuna *et al.*, 2005); *Arundo donax* (Kaur *et al.*, 2005b) etc. The crystal structure of one such complex-sugar specific lectin from *Scilla campanulata* bulbs (SCAfet) has been reported (PDB code 1DLP; Wright *et al.*, 2000). Unlike most monocot mannose-binding lectins (for example, GNA and SCAman) which fold into a single domain (each domain consisting of three subdomains), SCAfet contains two domains (each domain consisting of three subdomains) with approximately 55% sequence identity, joined by a linker peptide. Each domain is made up of a 12-stranded β-prism II fold, with three putative carbohydrate-

binding sites, one on each subdomain (Fig. 1.2. (C)). The lack of interaction of SCAfet with simple sugars could be due to the replacement of key amino acid residues (Asn, Asp, Gln, Tyr) within the monosaccharide-binding pocket by hydrophobic residues.

Because of their abundance in the storage tissues, monocot mannose- binding lectins are considered to be the storage proteins. They might also be involved in the plant's defense against sucking insects (Rahbé *et al.*, 1995) and invertebrates (Hilder *et al.*, 1995).

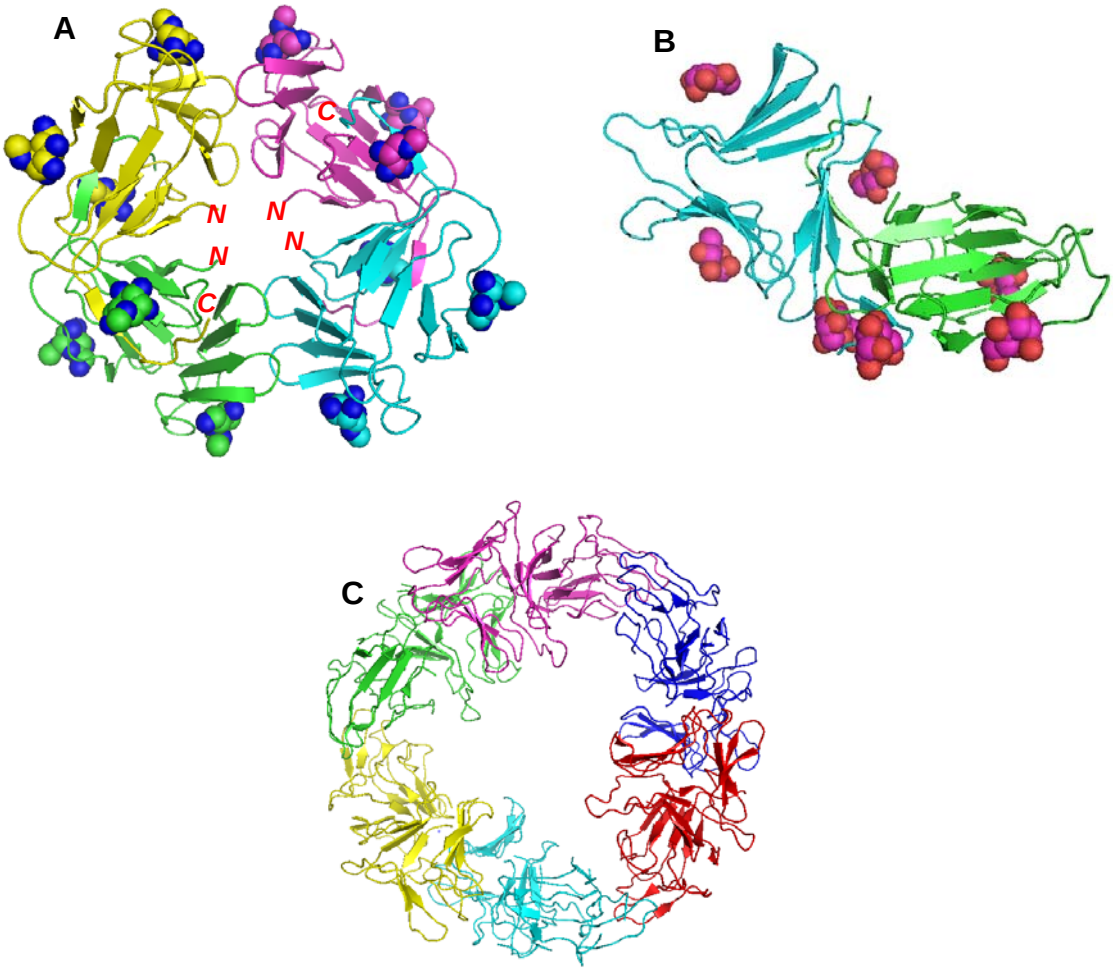


Fig. 1.2. (A) The crystal structure of GNA (PDB code: 1MSA) in complex with O1-methyl-mannose, shown as space filling models.
(B) The dimeric garlic bulb lectin (PDB code 1KJ1) in complex with α-D-mannose shown as space filling models.
(C) The crystal structure of SCAfet (PDB code: 1DLP).

III. Chitin-binding lectins composed of hevein domains

This family consists of all proteins containing at least one hevein domain (Raikhel *et al*., 1993). Hevein is a small 43 amino acid protein from the latex of the rubber tree (*Hevea brasiliensis*) (Waljuno *et al.*, 1975). "Hevein domain" refers to a structural unit of about 40 amino acid residues that exhibit sequence similarity to hevein and possess chitin-binding activity. There are other chitin binding lectins without hevein domain, for example chitin binding legume lectins and Cucurbitaceae phloem lectins, which have no sequence similarity to the hevein domain and hence are not considered in this family (Van Damme *et al.*, 1998a).

Chitin-binding lectins are ubiquitously found in plants. Their occurrence has been reported in several plant families, namely Gramineae, Solanaceae, Phytolaccaceae, Urticacea, Papavaracaceae and Viscaceae (Peumans *et al*., 1996; Raikhel *et al*., 1993). These lectins occur in seeds as well as vegetative tissues. They belong to different classes, namely merolectins, hololectins as well as chimerolectins.

Hevein is a typical merolectin; in addition, chitin-binding antimicrobial peptides have also been described (Broekaert *et al.*, 1992). The three-dimensional structure of hevein contains two short α-helices and a stretch of amino acid residues located at the *N*-terminal end of the polypeptide chain that forms two strands of antiparallel β-sheet followed by one more α-helix (Fig. 1.3. (A)). Four disulfide bonds (colored yellow in Fig. 1.3. (A)) stabilize the overall structure of hevein.

Hololectins consisting of polypeptides built up of two, three, four and seven tandemly arrayed hevein repeats occur in various plant species, for example, the monomeric lectin with two hevein domains from *Urtica dioica* (Peumans *et al.*, 1984), the dimeric Wheat Germ Agglutinin (WGA) and other Graminae lectins containing four hevein domains (Reikhel *et al.*, 1993). WGA exists in three isoforms, namely WGA-1, WGA-2 and WGA-3, which differ by 5-8 amino acid residues. Overall three-dimensional

structure of these isoforms is same. The monomer is made up of four hevein domains each folded into a compact globule through four disulfide linkages and five or six β-turns (Fig. 1.3. (B)). The four domains are organized into a helical assembly. Monomers associate in a head-to-tail fashion to form dimers such that the pairs of domains in contact are quasi two-fold related. Each of the four domains in a monomer possesses a carbohydrate-binding site (Wright, 1990).

Chimeric lectins occurring in the family of chitin binding lectins are of two types. Class I chitinases have a single *N*-terminal chitin binding domain linked to the catalytically active chitinase domain (Collinge *et al.*, 1993; Beintema, 1994). The second type consists of dimeric Solanaceae lectins, for example, *Lycopersicon esculentum* agglutinin and *Solanum tuberosum* agglutinin. These proteins are built up of chimeric polypeptides consisting of an *N*-terminal chitin binding domain with three hevein repeats linked to a highly *O*-glycosylated serine-hydroxyproline-rich domain (Kieliszewski *et al.*, 1994; Allen *et al.*, 1996). Three-dimensional structure of none of these has been reported so far.

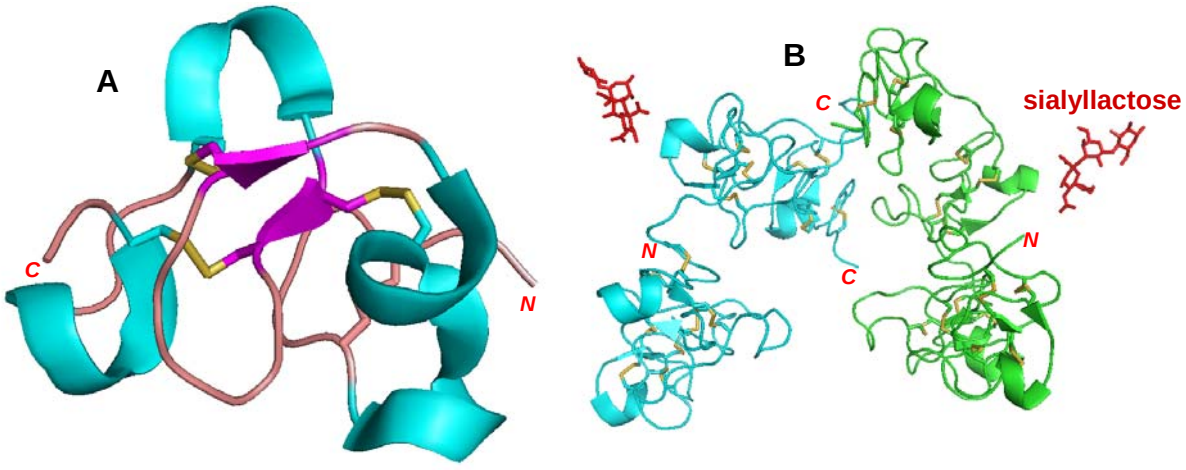


Fig. 1.3. (A) Three-dimensional structure of hevein (PDB code 1Q9B). α-helices are colored in cyan and β-sheets are colored in magenta. The disulfide bonds are shown in yellow. **(B)** WGA complexed with sialyllactose (PDB code 1WGC).

Class I chitinases are involved in the plant's defense against fungi (Collinge *et al.*, 1993). Other non-enzymatic chitin binding lectins are also supposed to be a part of a more complex defense system in which they act synergistically with other antifungal proteins (Van Damme *et al.*, 1998a). Some chitin binding lectins may also be involved in the plant's defense against bacteria, for example the seed lectin from *Datura stramonium* (Broekaert and Peumans, 1986) and insects (Murdock *et al.*, 1990; Huesing *et al.*, 1991).

IV. Type 2 RIP and related lectins

Ribosome-inactivating proteins (RIP) catalytically inactivate eukaryotic ribosomes (Barbieri *et al*., 1993). Type 1 RIPs consist of a single polypeptide of about 30 kDa with polynucleotide adenosine glycosidase (PAG) activity, whereas Type 2 RIPs contain an enzymatically active A chain (Endo *et al*., 1987) linked to a galactose specific domain called B-chain (Lord *et al*., 1994) and hence possess lectin activity. The A and B chains are linked by disulfide bonds. Type 2 RIPs occur in several families, namely Euphorbiaceae, Fabaceae, Viscaceae, Passifloraceae, Ranunculaceae, Lauraceae, Sambucaceae, Cucurbitaceae and Iridaceae. The first lectin studied of this type was "ricin", isolated from the seeds of *Ricinus communis*. The three-dimensional structure of ricin (Rutenber *et al.*, 1991; Katzin *et al.*, 1991; Rutenber and Robertus, 1991) consists of two subunits (A and B) of molecular weight ~32 kDa held by a disulfide bridge (Fig. 1.4) and has specificity for Gal/GalNAc (Lord *et al*., 1994). The ricin A chain contains regular secondary structure and is responsible for the catalytic activity, whereas the B chain contains no regular secondary structures such as α-helices or β-sheets and consists mainly of coil structures linked by turns and loops and is involved in the carbohydrate binding. The two domains in the B chain show a tertiary folding characteristic of the β-trefoil family (Murzin *et al.*, 1992).

Three-dimensional structure of type 2 RIP from *Abrus precatorius* seeds (abrin) is similar to that of ricin (Tahirov *et al.*, 1995). Lectins from seeds of certain cucurbitaceae family plants, for example *Trichosanthes anguina* (Komath *et al.*, 1996), *T. cucumerina* (Padma *et al.*, 1999) and *T. dioica* (Sultan *et al.*, 2004) are also considered to be type 2 RIPs. Some of these lectins are resistent to a wide range of denaturing conditions (Dharkar *et al.*, 2006).

Because of the extreme cytotoxicity of certain type 2 RIPs like ricin and abrin, these are considered to play a role in the plant's defense against plant-eating organisms. Some type 2 RIPs also show antiviral activity *in vitro* against plant viruses, hence might be involved in the plant's defense against these viruses (Barbieri *et al.*, 1993; Kumar *et al.*, 1993). Because of their abundance in seeds and vegetative storage tissues, they may act as storage proteins as well.

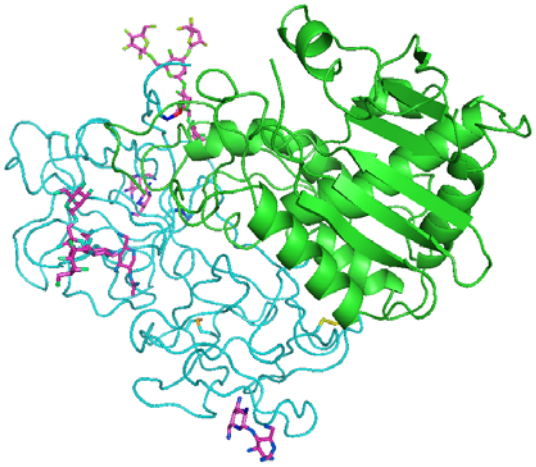


Fig. 1.4. Three-dimensional structure of ricin (PDB code 2AAI). Chain A, shown in green, has a regular secondary structure, whereas chain B has no secondary structure, however binds carbohydrates.

V. Jacalin related lectins

Jacalin is one of the two lectins which occur in mature seeds of *Artocarpus integrifolia* (jackfruit). The term "jacalin related lectins" is used for all lectins that are structurally and evolutionary related to jacalin. Based on their specificities, the lectins in this family are divided into two subgroups. The galactose-specific subgroup comprises jacalin and a few other Moraceae lectins, which exhibit specificity towards galactose and are built up of subunits consisting of a short β chain of about 20 residues and a long α chain of 133 residues. Lectins belonging to the mannose-binding subgroup occur in different plant families, exhibit an exclusive specificity towards mannose and are built up of subunits consisting of a single polypeptide chain (Bourne *et al*., 1999). Galactose specific lectins are cytoplasmic proteins, whereas mannose specific ones are located in storage vacuoles (Peumans *et al*., 2000).

Jacalin-related lectins are reported to occur in plants from several taxonomically unrelated families, namely Moraceae (jacalin, MPA, *Artocarpus hirsuta* lectin, artocarpin), Convolvulaceae (calsepa, conarva), Asteraceae (heltuba), Gramineae (barley and wheat lectins) and Musaceae (banana lectin) (Peumans and Van Damme, 1998). Jacalin was the first lectin in this family to be isolated (Kumar *et al*., 1982), sequenced, crystallized and three-dimensional structure determined in complex with methyl-α-D-galactose (Sankaranarayanan *et al*., 1996).

The three-dimensional structures of all jacalin-related lectins show the characteristic β-prism I fold, which is defined by a 3-fold symmetric β-prism made of three four stranded β-sheets, which are arranged like the three faces of a prism. The strands are nearly parallel to the 3-fold axis (Fig. 1.5. (A)). Each monomer has a single carbohydrate binding site. Most of the lectins of this family form tetramers with 222 symmetry as their functional biological units (Fig. 1.5. (B)) and hence are tetravalent. In contrast, the lectin from *Helianthus tuberosus* (Heltuba) is octameric and eight

monomers are assembled as a donut-shaped octamer (Fig. 1.5. (C)) with eight solvent-exposed carbohydrate-binding sites (Bourn *et al.*, 1999). The lectin from *Calystegia sepium* (Calsepa) exhibits a novel dimeric assembly that mimics the canonical 12-stranded β-sandwich dimer (Fig. 1.5. (D)) typically found in legume lectins (Bourn *et al.*, 2004).

Although jacalin-related lectins occur in many tissues, Moraceae lectins are particularly abundant in the seeds, whereas Convolvulaceae lectins are abundant in rhizomes. In vitro studies indicate that Moraceae lectins interact with human and animal cells and are capable of inducing specific processes (Van Damme *et al.*, 1998a). Feeding trials with artificial diets have demonstrated that the Moraceae lectins have anti-insect properties. For example, MPA was found to inhibit growth of larvae of the cowpea weevil (*Callosobruchus maculatus*) (Murdock *et al.*, 1990), jacalin and MPA inhibited larval growth of the Southern corn rootworm (*Diabrotica undecimpunctata*) (Czapla and Lang, 1990), and *Artocarpus hirsuta* lectin had insecticidal activity against the larvae of red flour beetle (*Tribolium castaneum*) (Gurjar *et al*., 2000). Hence it can be concluded that Moraceae seed lectins are storage proteins with an additional defensive function against potential seed predating animals and/or insects. Similarly, Convolvulaceae lectins have mitogenic activity (Peumans *et al.*, 1997b). Therefore they might be (rhizome-specific) storage proteins with a possible protective activity against potential predating soil-borne vertebrates or invertebrates.

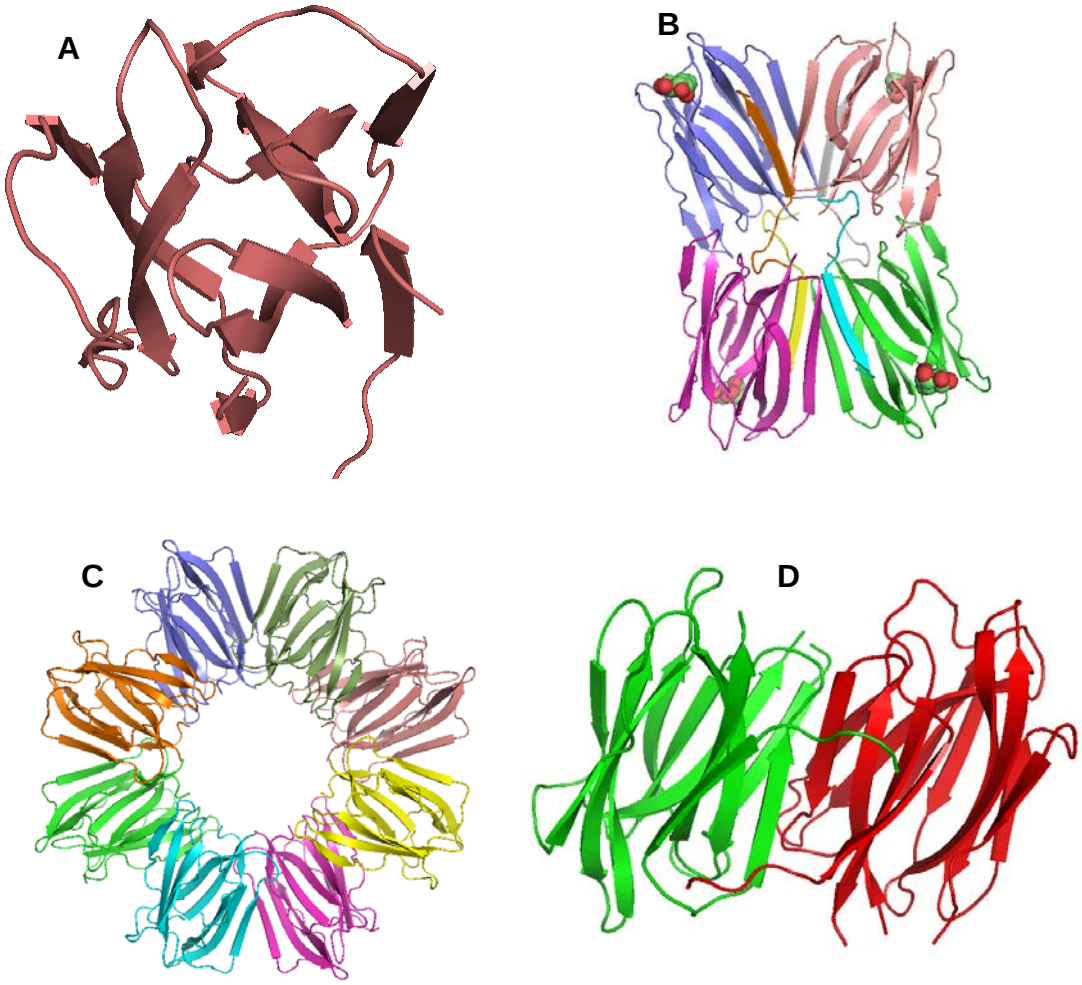


Fig. 1.6. (A) 3-fold symmetric β-prism I fold as observed in the monomer of jacalin (PDB code: 1UGX).

(B) Three-dimensional structure of the functional tetramer of jacalin (PDB code: 1UGW).

(C) Octameric structure of Heltuba lectin (PDB code: 1C3K). The octamer was constructed from the monomeric asymmetric unit using symmetry operations in QUANTA.

(D) Dimeric assembly of Calsepa lectin showing resemblance to canonical dimer of legume lectins (PDB code: 1OUW).

VI. Amaranthin lectins

The amaranthin lectin family is named after amaranthin, which is the lectin from *Amaranthus caudatus* seeds. These lectins exclusively occur in seeds of several plants from the family Amaranthaceae. These are homodimeric proteins composed of two identical subunits of about 30 kDa and exhibit specificity towards GalNAc (Rinderle *et al*., 1989, 1990). The three-dimensional structure of amaranthin at 2.2 Å resolution is available (Transue *et al*., 1997). The monomers consist of two domains (called *N*- and *C*-domains) linked by a short helix. Each of these domains has a β-trefoil structure similar to that of the two domains of the ricin B chain. However, unlike the ricin B chain (which has no extended secondary structure); the domains of amaranthin consist of six strands of antiparallel β-sheet capped by three β-hairpins into a β-barrel. The two domains are linked by a 3_{10} helix. Two monomers associate head-to-tail to form a dimer by extensive non-covalent contacts between both monomers. In this dimeric organization, the *N*-domain of one monomer faces the *C*-domain of the other monomer. The dimer exhibits two surface-exposed carbohydrate-binding sites that appear as two shallow depressions formed at the interface between the *N*- and *C*-domains of the two facing monomers (Fig. 1.6).

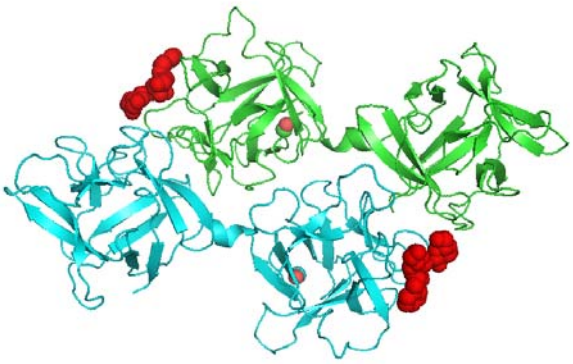


Fig. 1.6. Three-dimensional structure of amaranthin from *Amaranthus caudatus* (PDB code 1JLX) in complex with T-antigen disaccharide (shown as space filling models).

Amaranthus caudatus agglutinin is inhibited by GalNAc, however it has a much higher affinity for the disaccharide Gal-β-1,3-GalNAc (Rinderle *et al*., 1989). Considering this, these lectins could be involved in the plant's defense against seed predators.

VII. Cucurbitaceae phloem lectins

This is a small family of chitin binding lectins which occur in the phloem exudate of some plants from Cucurbitaceae family. They are not related to the seed lectins of Cucurbitaceae plants (which are mostly type 2 RIPs) and also do not contain the hevein domain. Cucurbitaceae phloem lectins (also called PP2) are abundant proteins and so far have been identified in phloem exudates of *Cucurbita*, *Citrullus*, *Cucumis*, *Sechium*, *Luffa,* and *Coccinia* species.

These lectins consist of unglycosylated subunits of about 25 kDa and are dimeric in solution (Bostwick *et al.*, 1992; Wang *et al.*, 1994). The two subunits of the *Cucurbita maxima* lectin are covalently linked through two interchain disulfide bonds (Read and Northcote, 1983). They show specificity toward oligomers of GlcNAc. The three-dimensional structure of none of these lectins has been reported so far.

Most probably, the Cucurbitaceae phloem lectins are involved in the plant defense mechanism. If the phloem vessels get injured, these lectins react with another abundant phloem protein, PP1, to form a rigid gel which blocks the cut vessels and prevents microbial infection (Read and Northcote, 1983).

1.3. Applications of plant lectins

Due to their ability to recognize and bind selectively to carbohydrates, plant lectins find applications in several areas of basic and medical sciences. According to Rüdiger and Gabius (2001), common applications of lectins in various disciplines include:

A. Biochemistry

I. Detection of defined carbohydrate epitopes of glycoconjugates in blots or on thin-layer chromatography plates.

II. Purification of lectin-reactive glycoconjugates by affinity chromatography.

III. Glycan characterization by serial lectin affinity chromatography.

IV. Glycome analysis (glycomics).

V. Quantification of lectin-reactive glycoconjugates in enzyme-linked lectin-binding assays (ELLA).

VI. Quantification of activities of glycosyltransferases/glycosidases by lectin-based detection of products of enzymatic reaction

B. Cell biology

I. Characterization of cell surface presentation of glycoconjugates and their preceding intracellular assembly and routing in normal and genetically engineered cells.

II. Analysis of mechanisms involved in correct glycosylation by lectin-resistant cell variants. Fractionation of cell populations.

III. Modulation of proliferation and activation status of cells.

IV. Model substratum for study of cell aggregation and adhesion.

C. Medicine

I. Detection of disease-related alterations of glycan synthesis.

II. Blood group typing and definition of secretor status.

III. Quantification of aberrations of cell surface glycan presentation, for example in malignancy.

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

1.4. Plant lectins with complex sugar specificity

Lectins were in general thought to be specific towards mono- or oligosaccharides which could inhibit their hemagglutination activity. This view is also reflected in the current definition of lectins (Peumans and Van Damme, 1995). However, with advances in lectin studies, it was observed that many lectins do not show specificity towards simple sugars and are inhibited by glycoproteins like fetuin, asialofetuin, thyroglobulin, fibrinogen, ovalbumin etc. and their corresponding glycopeptides. Such complex-sugar specific lectins have been reported to occur in the plants from several families, such as leguminosae (seed lectins from *Phaseolus vulgaris* (Kamemura *et al.*, 1993; Kaneda *et al.*, 2002) and *Acacia constricta* (Guzmán-Partida *et al.*, 2004)), Moraceae (seed lectins from *Ficus cunia* (Ray *et al.*, 1992) and *Ficus bengalensis* (Singha *et al.*, 2007)), Lamiaceae (seed lectins from *Salvia sclarea* (Piller *et al.*, 1986); *Salvia bogotensis* (Vega *et al.*, 2006); *Glechoma hederacea* (Singh *et al.*, 2006)); Araceae (tuber lectins from *Gonatanthus pumilus* (Dhuna *et al.*, 2007); *Arisaema helleborifolium* (Kaur *et al.*, 2006b); *Sauromatum venosum* (Singh Bains *et al.*, 2005)), Amaranthaceae (*Amaranthus viridis* seed lectin (Kaur *et al.*, 2006c)).

Being specific towards various glycoproteins, some of these lectins show anti-insect (Kaur *et al.*, 2006a, b) as well as anti-fungal (Kaur *et al.*, 2006c) properties, suggesting their possible use in protecting crop plants from the attack of insects and fungal pathogens. Some of these lectins have also shown mitogenic and/or anti-proliferative effect against human cancer cell lines (Singh *et al.*, 2004; Dhuna *et al.*, 2007), hence are considered to be useful in cancer research and therapy.

While screening seeds of several plants for the presence of hemagglutination activity, the crude extracts of *Cicer arietinum* and *Moringa oleifera* seeds were found to agglutinate pronase treated and untreated human/rabbit erythrocytes, respectively. Hemagglutination inhibition assays indicated that these two hemagglutinins do not show

specificity towards any simple sugars and their hemagglutination activity is inhibited only with certain glycoproteins. X-ray crystallographic as well as biochemical and biophysical studies were undertaken to elucidate the three-dimensional structure and to understand the structural basis of their complex-sugar specificities.

1.5. *Cicer arietinum* (Chickpea)

Classification:

Kingdom: Plantae

Division: Magnoliophyta

Phylum: Streptophyta

Order: Fabales

Family: Fabaceae (Leguminosae)

Subfamily: Papilionoideae

Tribe: Cicereae

Genus: *Cicer*

Chickpea (*Cicer arietinum*) is a very versatile legume and forms an important part of the diet in the Indian sub-continent as well as many Middle-Eastern countries. Originated in the Middle East (most probably southeastern Turkey (van der Mäsen, 1987)), it is now cultivated in the tropical, sub-tropical and temperate regions. There are two varieties in chickpea, namely Desi (also known as Bengal gram or kala chana) and Kabuli. Seeds of Desi type (Fig. 1.7. (A)) are small, dark with a rough coat while Kabuli type seeds are light-colored, larger and have a smoother coat. Kabuli type is cultivated in temperate regions while the Desi type is cultivated in the semi-arid tropics (Mühlbauer and Singh, 1987; Malhotra *et al.*, 1987).

Chickpea seeds are very nutritive with high protein content (25.3-28.9%; Hulse, 1991) and hence are considered an important part of the vegetarian diet. Apart from this,

they are also a good source of soluble and insoluble fiber, carbohydrates, vitamins and minerals. Soluble fibers are known to control the blood cholesterol level and hence reduce the risk of heart diseases. Insoluble fiber helps preventing digestive disorders. Seeds contain low amount of fats most of which are polyunsaturated, which is also helpful in preventing cardiovascular diseases.

Chickpea plant as a whole is also consumed as a green vegetable. Chickpea seeds are consumed in many ways. While fresh seeds are eaten as raw snack, mature seeds are cooked and used in several delicacies. Seed flour (known as *besan*) is used in many Indian recipes.

X-ray crystallographic studies on *Cicer arietinum* lectin (CAL) to determine its three-dimensional structure are reported in this thesis. The lectin was isolated from the mature seeds of Desi variety (in particular, cultivar BDN 9-3) of chickpea. During the initial screening, it was observed that the crude extract of chickpea seeds agglutinated pronase treated human or rabbit erythrocytes. When seeds of several varieties of chickpea were studied for the presence of hemagglutination activity, the variety BDN 9-3 showed maximum hemagglutination titre. Hence the purification of the lectin was attempted from seeds of this variety.

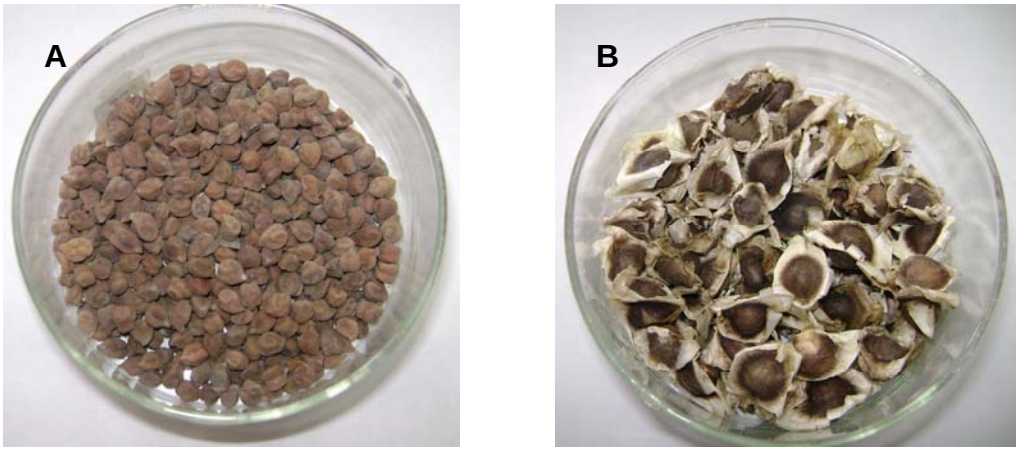


Fig. 1.7. (A) Mature seeds of *Cicer arietinum* Desi variety.
(B) Mature seeds of *Moringa oleifera*.

1.6. *Moringa oleifera* (drumstick)

Kingdom: Plantae

Division: Magnoliophyta

Phylum: Streptophyta

Class: Magnoliopsida

Order: Brassicales

Family: Moringaceae

Genus: *Moringa*

Moringa oleifera is a multipurpose tree belonging to family *Moringaceae*, which is a single genus family of shrubs and trees. Native to the Northern India, it is now cultivated in most of the tropical regions (Fuglie, 2001). It is small, deciduous, perennial tree of 2.5-10 m in height, the timber of which is of low quality (Fahey, 2005). It tolerates poor soil including coastal areas and grows best in dry sandy soils. It is a fast growing and draught resistant plant and is known to be non-toxic to humans and animals (Maikokera and Kwaambwa, 2007).

Almost all parts of *Moringa oleifera* are used for food or are having some beneficial property; hence it is considered to be one of the world's most useful trees. The immature green pods, called as drumsticks are commonly consumed in India as a delicious vegetable. Mature seeds (Fig. 1.7. (B)) are eaten like peas or roasted like nuts. Flowers can also be eaten when cooked. Leaves are used as green leafy vegetable, and are highly nutritious (Nesamani, 1999). Shredded and dried roots have a flavor which resembles to horseradish hence are used as a condiment. However, roots also contain the alkaloid spirochin, which is a potentially fatal nerve paralyzing agent (Morton, 1991). Seeds contain 38-40% edible oil, also called Ben oil, which can be used in cooking, cosmetics, and lubrication. It has physical and chemical properties equivalent to that of olive oil and contains a large quantity of tocopherols (Tsaknis *et al.*, 1999).

All parts and their products of *Moringa oleifera* tree: bark, sap, roots, leaves, seeds, oil and flowers are used in traditional medicines in several countries. The root extract is known to be anti-inflammatory (Ezeamuzle *et al.*, 1996). The leaf extract is hepatoprotective (Pari and Kumar, 2002), hypotensive (Faizi *et al.*, 1995) and has antitumour activity (Murakami *et al.*, 1998). The seeds have strong coagulative and antimicrobial properties (Eilert *et al.*, 1981). In Taiwan and China, these seeds are used to treat athlete's foot and *Tinea* infections (Chuang *et al.*, 2007). Anti-fungal activity of crude extracts of leaves and seeds as well as of the essential oil from seeds has also been demonstrated by Chuang *et al.* (2007).

According to Fuglie (2001) the many uses for *Moringa oleifera* include: alley cropping (biomass production), animal forage (leaves and treated seed-cake), biogas (from leaves), domestic cleaning agent (crushed leaves), blue dye (wood), fencing (living trees), fertilizer (seed-cake), foliar nutrient (juice expressed from the leaves), green manure (from leaves), gum (from tree trunks), honey- and sugar cane juice-clarifier (powdered seeds), honey (flower nectar), medicine (all plant parts), ornamental plantings, biopesticide (soil incorporation of leaves to prevent seedling damping off), pulp (wood), rope (bark), tannin for tanning hides (bark and gum) and water purification (powdered seeds).

Purification and characterization of the hemagglutinating protein (*Moringa oleifera* lectin; MoL) revealed that it is the same fluocculating protein described by Gassenschmidt *et al.* (1995) and Ndabigengesere *et al.* (1995). However, so far, the hemagglutinating activity of the protein has not been described nor the structure-function relationship for this protein has been studied. Hence, biochemical and biophysical characterizations as well as crystallization of the protein were undertaken in order to finally solve the three-dimensional structure of the protein.

Both these proteins (CAL and MoL) agglutinate pronase treated/untreated RBCs, hence can be termed as hemagglutinins. Since they are inhibited only by certain glycoproteins and not by any simple sugars, their specificities are unknown and hence they do not comply with the classic definition of lectins. However, in this thesis, for the sake of simplicity, they have been treated as lectins.

1.7. Crystal polymorphism in lectins

Crystal polymorphism is defined as the ability of a solid material to exist in more than one crystal form. The material may crystallize in different space groups or in the same space group with different unit cell dimensions. Polymorphism can be of different kinds. “Packing polymorphism” refers to the polymorphism which exists as a result of difference in the crystal packing, whereas the existence of different conformers of the same molecule leads to the “conformational polymorphism”. Differences in the hydration or solvation can give rise to different crystal forms which is termed as “pseudopolymorphism”.

Polymorphism in organic crystals is a well-known phenomenon, being reported for the first time as early as in 1832 (Wöhler and Liebig, 1832) in the crystals of benzamide. Subsequently, polymorphism has been documented in many other organic molecule crystals, to such an extent that it is supposed to be a frequent occurrence for organic solids. In a famous sentence, Walter McCrone stated that the number of polymorphs is proportional to the time and effort spent in their search (Gavezzotti, 2007).

Polymorphism affects the packing of the molecules in the crystals and their intermolecular contacts. It may also result in difference in the properties of the same molecule in different crystal forms. However, the solid-state properties (elastic tensors, bulk modulus, solubility, resistance to permeation, color or electric and magnetic properties) of organic crystal polymorphs remain mostly similar (Gavezzotti, 2007). In

pharmaceuticals, crystal polymorphism may play an important role in determining the properties of a particular drug. For those medicines which are orally administered as crystalline solids, the dissolution rates depend on the exact crystal form of a polymorph.

Although the phenomenon of polymorphism is not always well understood, certain factors are known to be associated with occurrence of polymorphic crystals. For example, temperature (Siegrist *et al.*, 1995), pressure (Moggach *et al.*, 2006), solvent (Weissbuch *et al.*, 2005), additives (Thallapally *et al.*, 2004), crystallization conditions etc. Sometimes, a metastable crystal form changes to another stable form, as it happens in case of the benzamide crystals (Wöhler and Liebig, 1832).

Polymorphism has also been observed in case of crystals of macromolecules like nucleic acids and proteins. However, protein crystal polymorphism is not as well-studied as the polymorphism in the organic crystals. In protein crystals, the polymorphism usually arises due to mutations, change in crystallization conditions like pH, temperature, precipitant etc. In this thesis, an attempt has been made to study the polymorphism in the case of certain plant lectins for which crystal structures are available in several space groups.

1.8. Tetrameric association of lectins and other proteins

Most of the proteins in the modern cell are oligomeric in nature consisting of two or more subunits. While soluble and membrane-bound functional proteins typically have 2 or more, finite number of identical subunits, almost all structural proteins are built from hundreds to millions of subunits. Symmetric oligomeric protein complexes are favored over monomeric proteins for several reasons. In case of enzymes, allosteric regulation and multivalent binding can be achieved with a multimeric protein. The oligomeric nature of lectins is responsible for their hemagglutination activity. Large protein complexes are more stable against denaturation and have a reduced surface area exposed to the solvent (Goodsell and Olson, 2000).

Although several oligomerization states are seen in the functional protein molecules, dimers and tetramers are the most common ones. While the dimeric arrangement shows a C2 (2-fold rotational) symmetry, tetrameric proteins can show several types of arrangements. The symmetry in the molecule may also influence the space group in which the protein crystallizes, as well as the packing interactions between protein molecules in the crystal, which could be related to their functions. In the present thesis, crystal structures of various homotetrameric proteins were studied with respect to their space groups and association of subunits so as to identify common patterns and features of tetrameric association in them.

Chapter 2

Crystallization and X-ray Crystallographic Studies on a Hemagglutinin from *Cicer arietinum* (Chickpea)

2.1. Summary

A hemagglutinin with complex sugar specificity was isolated and purified by two successive ion exchange chromatography steps from the mature seeds of *Cicer arietinum* (chickpea). The protein (15 mg ml^{-1} in deionised water) was crystallized using the hanging-drop vapor-diffusion technique in the condition consisting of 0.1 M sodium cacodylate buffer pH 6.5, 0.2 M sodium acetate and 30 % w/v polyethylene glycol 8000. The quality of the crystals was further improved by using a precipitant solution containing 0.1 M sodium cacodylate buffer at pH 6.5, 0.2 M sodium acetate and 10-15 % polyethylene glycol 8000. The protein crystallizes in two different space groups, trigonal (*P*3) with unit cell dimensions *a* = *b* = 80.21, *c* = 69.14 Å and orthorhombic ($P2_12_12$) with unit cell dimensions *a* = 70.93, *b* = 73.32, *c* = 86.98 Å. Determination of the three-dimensional structure of the protein was attempted using multiple isomorphous replacement method, since no suitable structural model could be found for obtaining a solution through molecular replacement method. The protein also crystallized in the presence of various heavy atom salts, namely KI, $HAuCl_4$, $Pb(NO_3)_2$, *p*-hydroxy mercury benzoate and dichloro(ethylenediamine)platinum (II). Only iodine derivative could be characterized for the presence of heavy atom.

2.2. Introduction

X-ray crystallography is the most powerful technique available so far to determine accurately the three-dimensional structure of biological macromolecules. The other techniques include nuclear magnetic resonance (NMR) and cryoelectron microscopy. Although NMR can be used to study the protein structure in solution and does not require protein crystals, it can be used to determine the structures of proteins having molecular weights less than ~50 kDa. On the other hand, the cryoelectron microscopy can be used only to study very large protein and/or RNA complexes like

ribosomes (Frank, 2003), proteasomes (Nickell *et al.*, 2007), or viruses (Tang and Johnson, 2002).

The hemagglutinin from *Cicer arietinum* (chickpea), a member of the family leguminosae, is chosen for the present study. It agglutinates pronase-treated rabbit and human erythrocytes and its hemagglutination activity is inhibited by fetuin and desialated fetuin but not by simple mono- or oligosaccharides. The purified lectin is a dimer of molecular weight 43000 Da composed of two identical subunits (MW 21500 Da), as confirmed by SDS-PAGE and gel chromatography using HPLC. The secondary structural elements of the protein as determined by circular dichroism (CD) experiments are *α* helix, 34%, *β* sheet, 28% and random coil, 38%. This is quite unusual for a legume lectin which is typically expected to have high amount of *β* sheet and *β* turns as secondary structure (Swamy *et al.*, 1985). The quaternary structure of most of the legume lectins consists of four identical subunits, or alternatively forms from two light and two heavy chains (Loris *et al.*, 1998). They agglutinate untreated RBCs and are inhibited by simple sugars. In contrast to most lectins in legumes, *Cicer arietinum* lectin (CAL) does not require Mn^{2+} or Ca^{2+} for agglutination. The sequence of only the first 25 amino acids of the *N*-terminal side of CAL has been determined (Kolberg *et al.*, 1983); a search using this sequence showed significant sequence similarity with the *N*-terminal sequence of a major seed albumin (PA-2) from *Pisum sativum* and no similarity with the sequences of other legume lectins. Thus, considering its molecular weight, amino acid sequence, quaternary structure, and physicochemical requirements for agglutination, CAL is a lectin different from those commonly found in legumes. It is likely that the protein is an albumin with carbohydrate-binding property. CAL has also been shown to bind hemin in 1:1 ratio, although with relatively low affinity (Pedroche *et al.*, 2005).

The structural studies of this lectin, to elucidate its three-dimensional structure, were undertaken to fully characterize it by getting more information on its molecular architecture and its carbohydrate binding property.

2.3. Materials

Cicer arietinum seeds of cultivar BDN 9-3 were obtained from the Agriculture Research Station, Badnapur, Maharashtra, India. Tris base, NaCl, ammonium sulphate and sodium acetate were purchased from SRL, India, while DEAE-cellulose, SP-sephadex, sodium cacodylate, polyethylene glycol 8000 (PEG 8000), $HAuCl_4$, $Pt(NO_3)_2Cl_2$, $Pb(NO_3)_2$, *p*-Hydroxy mercury benzoate and dichloro(ethylenediamine)platinum (II) were purchased from Sigma, USA. $HgCl_2$, $NiCl_2$, $CoCl_2$, $RuCl_3$, KI were from SRL, India and Merck, India.

Crystal Screen 1, glass capillaries (Φ = 1 mm) and nylon loops used to mount the crystals were purchased from Hampton Research, USA. The 24-well tissue culture trays used for hanging-drop vapor-diffusion crystallization were from Axygen and Corning. The circular coverslips (Φ = 19 mm) were of Blue Star make. The Cu-*K*α radiation was generated using a rotating anode X-ray generator from Rigaku-MSC, USA which was equipped with a confocal mirror focusing system. X-ray diffraction data was collected on an *R*-AXIS IV^{++} image plate. To collect diffraction data at low temperature, the crystal was flash cooled in a nitrogen stream produced by X-stream (Rigaku-MSC, USA).

Different crystallographic softwares were used to process the data, like DENZO and MOSFLM (to integrate the diffraction images) and SCALEPACK or SCALA (to scale the data). A Silicon Graphics workstation (Octane) with Irix 6.5 as the operating system and an IBM PC with Fedora Core 5 were used to run these programs.

2.4. Methods

2.4.1 Purification of *Cicer arietinum* lectin

The first and foremost step before attempting to crystallize a protein is to obtain it in a pure form, either from its natural source or from an organism in which the gene coding for the protein is cloned and expressed. Several purification steps might be needed to purify the required protein from the source.

The *Cicer arietinum* lectin (CAL) was isolated from the mature seeds of chickpea (*Cicer arietinum*), purified and biochemically characterized elsewhere (Katre *et al.*, 2005) and made available to the candidate in a purified form for crystallization. All the purification steps were carried out at 277 K unless otherwise mentioned. Briefly, 100 g dry seeds of *C. arietinum* were finely powdered in a mixer-grinder and soaked in 500 ml 10 mM Tris-HCl and 150 mM NaCl at pH 7.2. The suspension was stirred for 16 h on a magnetic stirrer and filtered through a muslin cloth. The filtrate was centrifuged in an SS-34 rotor at 10000 *g* for 20 min. To the supernatant, ammonium sulfate (AS) was added slowly while stirring the solution continuously on a magnetic stirrer to obtain a final saturation of 80% AS. The protein was allowed to precipitate for 16 h after which the solution was centrifuged at 10000 *g* for 20 min in an SS-34 rotor. The precipitate thus obtained was dissolved in 20 mM Tris-HCl buffer pH 7.2 and dialysed against the same buffer for 20 h with three changes of buffer. The dialysate was centrifuged at 10000 *g* for 10 min in microfuge tubes and the clear supernatant was loaded onto a DEAE-cellulose column pre-equilibrated with 20 mM Tris-HCl buffer pH 7.2. The column was washed with the same buffer until the wash showed no absorbance at 280 nm. The lectin eluted with the unadsorbed portion. Fractions (2 ml) of the wash with OD greater than 0.2 and possessing hemagglutination activity were pooled together, concentrated and dialyzed against 20 mM acetate buffer pH 5.0. This sample was loaded onto a SP-Sephadex column (4 x 20 cm) pre-equilibrated with 20 mM acetate buffer pH 5.0 and washed with

the same buffer. Finally, the bound protein was eluted with acetate buffer pH 5.0 containing NaCl in a stepwise gradient of 0.1-0.5 M NaCl in steps of 0.1 M. The fractions with OD_{280} greater than 0.2 and possessing hemagglutination activity were pooled and dialyzed against deionized water. The final yield of the lectin was 150 mg from 100 g dried seeds, with a specific activity of 5×10^4 U mg^{-1}.

The purified protein has a tendency to form aggregates in solutions of low ionic strength and high protein concentration, especially when stored at temperature 273 K or below. Hence it was maintained at 277 K in the form of a dilute solution (protein concentration ~ 0.2-0.3 mg ml^{-1}) in 1X TBS or PBS and 0.02 % sodium azide was added to it to avoid bacterial/fungal growth in the solution. The protein solution was dialysed against deionized water and concentrated to 10-15 mg ml^{-1} just before setting up crystallization experiments.

2.4.2 Crystallization of the purified *Cicer arietinum* lectin

A crystal is a solid in which the constituent atoms, ions or molecules are packed in a regularly ordered, repeating pattern. The growth of diffraction quality single crystals from the purified protein solution is often a crucial and rate-limiting step in determining the protein structure by X-ray crystallography. To crystallize a given protein, its solution has to be slowly brought to a condition of supersaturation. In such a case instead of forming an amorphous precipitate, the protein molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation.

Crystallization is one of the several means (including non-specific aggregation or precipitation) by which a metastable supersaturated solution can reach stable lower energy state by a reduction in the solute concentration (Weber, 1991). The general processes by which substances crystallize are similar for molecules of both small molecules (salts and small organic molecules) and macromolecules like proteins, DNA

and RNA. Crystallization is known to lower the free energy of proteins by ~3-6 kcal/mole relative to the solution state (Drenth and Haas, 1998). There are three stages observed in the crystallization of the given molecule: nucleation, growth of nuclei and cessation of growth. In the nucleation stage, molecules freely moving in the solution come together to arrange systematically and produce a crystalline aggregate, which is a thermodynamically stable arrangement in a repeating lattice. This does not grow till it exceeds the critical size (which is defined by the ratio of the surface area of the aggregate to its volume (Feher and Kam, 1985; Boistelle and Astier, 1988)) above which it is capable of further growth to produce a well-sized crystal. Cessation of crystal growth could be due to several causes, like depletion of macromolecules from the surrounding media, growth defects, poisoning of crystal faces or ageing of the molecules (Ducruix and Giege, 1999).

Protein crystallization is generally considered to be a trial and error procedure and more as an “art” rather than a “science”. Several factors are known to affect the crystallization and the quality of crystals obtained, namely temperature, pH, ionic strength of the solution, presence of impurities, additives etc. Along with such known factors, several unidentified factors may also play a role in crystallization. In general, in crystallization experiments, a precipitant such as polyethylene glycol (available in a range of molecular weights), salt (like ammonium sulfate, lithium sulfate, tri-sodium citrate etc.) or organic solvent (such as ethanol, 2-methyl-2,4-pentanediol (MPD), acetone, isopropanol) is allowed to diffuse into the protein solution maintained at a specific pH, ionic strength and temperature (Drenth and Haas, 1998). To find out the favorable pH, temperature and concentration of the precipitant at which the given protein will crystallize one has to screen several variations of these factors. A variety of crystallization screening kits are available nowadays, most of which are sparse matrix

screens, which involve an intentional bias towards combinations of conditions that have been found successful in many instances.

Out of the several techniques used to setup protein crystallization experiments, the vapor diffusion techniques are the most popular (Chayen, 1998). In these techniques, usually equivolume amounts (0.5-10 µl) of protein and precipitant solutions are mixed and allowed to equilibrate against a relatively large volume (0.25 – 1 ml) of the reservoir solution of precipitant in a closed system. Initially, the droplet of protein solution contains a concentration of precipitant lesser than that of reservoir solution, which increases later as the water evaporates from the drop and equilibrates with the reservoir, to a level optimum for the crystallization of the protein.

Vapor diffusion technique can mainly be practiced using two methods: hanging drop and sitting drop. While sitting drop method is most suitable in high-throughput screening for optimum crystallization conditions and can be mechanized, hanging drop method is used to reproduce the results and to get good diffraction quality crystals. Crystallization of CAL reported here has been achieved using hanging-drop vapor-diffusion technique.

Initial crystallization trials for CAL were carried out using the sparse-matrix screen Crystal Screen I supplied by Hampton Research. 0.5 ml of each of the screen solutions were dispensed in the wells of multiwell trays, the edges of which were coated with silicone grease. 1 µl of the protein solution was mixed with 1 µl of reservoir solution on a siliconised cover slip, which was then kept inverted over the respective well. The plates were kept at 295 K and checked periodically for the appearance of crystals. The drops were viewed at 10 X zoom through an Olympus microscope, equipped also with a digital camera, to check and record the presence of crystals.

The protein was found to crystallize in the condition numbers 28, 40 and 46 of the Crystal Screen. The crystals grown from condition no. 28 (0.2 M sodium acetate, 0.1

M sodium cacodylate pH 6.5, 30% (w/v) polyethylene glycol (PEG) 8000) were found to be suitable for X-ray diffraction data collection. The quality of the crystals was further improved by refining the above condition, particularly with regard to buffer choice and PEG concentration.

To collect the data at low temperature (~ 100 K), the crystal was cryoprotected using 25-30% glycerol or PEG 200 incorporated in the mother liquor. The crystal was briefly soaked (1-2 seconds) in the cryoprotectant solution, before transferring it in the liquid nitrogen jet from X-stream, in front of the X-ray beam.

2.4.3 Data collection

Once a good quality protein crystal is obtained, the next step is to carry out the X-ray diffraction experiment to measure the intensities of Bragg reflections. In this experiment, the crystal is placed in an intense X-ray beam, mounted in a glass capillary or a nylon loop on a goniometer which can centre the crystal and rotate it. The X-rays diffracted from the crystal are then measured with the help of a detector.

Various types of X-ray generators are available to produce X-rays. Since protein crystals diffract relatively poorly compared to the small molecule crystals of the same size, a high intensity X-ray radiation is needed to get good quality data from them, as well as a highly sensitive detector is needed to record the diffraction. A typical laboratory X-ray source is of the sealed tube or of the rotating anode type, which can produce X-rays of a fixed wavelength, characteristic of the anode material. Prior to rotating anode, sealed tube type of anodes were in use, which produced X-rays with relatively low intensity due to the heating of the anode at the focal spot. The introduction of rotating anode X-ray tubes in the 1960's brought about a 10-fold increase in X-ray intensity. For protein crystallography, a copper anode is used which produces X-rays having a wavelength of 1.5418 Å. Alternatively particle accelerators like synchrotrons are used as

X-ray sources. The intensity of X-rays produced by synchrotron sources is much higher compared to the conventional sources (~ 10^{18} photons/s/mm^2/mrad^2 for a typical 3rd generation synchrotron source). Another advantage of synchrotron radiation is that it is tunable, unlike the conventional sources. Hence any suitable wavelength in the spectral range can be selected with a monochromator. This property is used in multiple wavelength anomalous dispersion and for Laue diffraction studies. Due to the enhanced brightness of the radiation and variable wavelength, synchrotron radiation can be used to study crystals which are too small or have a unit cell too large to be studied at home source.

Although the higher intensity of the source is advantageous, it causes serious radiation damage to the crystal, which will be reflected in the quality of the data collected. The problem of radiation damage can be restricted if the data is collected at a temperature as low as 100 K. This can be achieved by cooling the crystal in a jet of vaporizing liquid nitrogen (cryostream). To prevent formation of ice in the protein crystal at such a low temperature, the crystal needs to be first soaked briefly in a cryoprotectant solution and freezed rapidly in the cryostream. Entire data collection is then carried out at low temperature. Low-temperature data collection is useful even at home sources. The advantages of low-temperature data collection include: reduction in the radiation damage of the crystal on exposure to X-rays, improvement in the limit of resolution, decrease in the thermal parameters, storage and reuse of crystals, overcoming the scaling problem by enabling the completion of entire data collection using only one crystal. However, introduction of cryoprotectant in the protein crystal can disrupt the crystal lattice which may lead to increase in mosaicity (Garman and Schneider, 1997). The technique of mounting crystals in a loop for flash cooling was first introduced by Teng (Teng, 1990) which was a major advancement in cryocrystallography. The loops are made of fine fibers like nylon which do not diffract X-rays. The crystal is scooped

from the cryosolution and held within the loop suspended by a thin film of the solution. The loop is supported by a pin, which is itself attached to a steel base used for placing the assembly on a magnetic cap on the goniometer.

To detect the X-rays diffracted by the crystal, several types of detectors can be used. At home sources, the X-ray storage-phosphor image plate (IP) is most commonly used. They are at least 10 times more sensitive than X-ray films and their dynamic range is much wider (Drenth, 1994). Area detectors based on a charge-coupled device (CCD) are even more sensitive and fast enough to record the data at synchrotron sources.

The data collection of CAL crystals was carried out at the Macromolecular X-ray Diffraction Facility at National Chemical Laboratory, Pune, India. For collecting the data at room temperature (295 K), a single crystal was mounted in a thin walled glass capillary having a diameter of 1 mm. To collect the data at low temperature (100 K), the crystals were momentarily soaked in the mother liquor containing 20-30 % glycerol or PEG 200 as cryoprotectants. X-ray diffraction data were collected on an *R*-AXIS IV^{++} image plate using Cu *K*α radiation generated by a Rigaku rotating-anode X-ray generator operated at 50 kV and 100 mA and equipped with a confocal mirror focusing system. The crystal alignment was done by viewing the image captured using a CCD camera in a TV monitor. The different processes such as exposure, data collection, readout and storage of data were carried out automatically via the CrystalClear program supplied by Rigaku MSC. Crystal-to-detector distance was chosen based on the longest unit cell dimension and mosaic spread, so that the diffraction spots were well resolved, without overlap of intensities of the spots. In this case a distance between 150 and 200 mm is sufficient to achieve reasonable resolution of the data. During data collection, the crystals were oscillated through an angle of 0.5^0 about an axis perpendicular to the direction of the X-ray beam. The exposure time per oscillation frame depended upon the quality of crystal. For the room temperature data collection, an exposure time of 120-180

seconds per frame was sufficient, while a 600 seconds exposure per frame was used to collect data at low temperature. A total of 180-360 diffraction images were collected in each dataset.

2.4.4 Data Processing

The processing of the X-ray diffraction data of a crystal consists of the following steps:

1. Indexing of the diffraction pattern and determining of the crystal orientation.
2. Refinement of the crystal and detector parameters.
3. Integration of the diffraction intensities of reflections.
4. Refining the relative scale factors between equivalent measurements.
5. Precise refinement of crystal parameters using all the data.
6. Merging and statistical analysis of the symmetry related reflections.

The oscillation images were displayed using the program XDisplayF. The first three steps in the data processing were carried out using the program DENZO and steps 4-6 were done using the program SCALEPACK. All these three programs are part of the HKL package (Otwinowski and Minor, 1997) and were run on an Octane workstation from Silicon Graphics, Inc. DENZO provides numerical analysis of each oscillation image, whereas SCALEPACK provides overall statistics for the whole data set. DENZO accepts peaks for autoindexing only from a single oscillation image and makes a complete search for all possible indices of all reflections. The program calculates the distortion index for all 14 Bravais lattices; user is given the choice of selecting the appropriate lattice and space group. The interactive mode of indexing is also an option in DENZO. Our processing procedure involved interactive mode. The peaks were selected by defining a non-overlapping box. The program SCALEPACK performs the scaling and merging of data from all images and does a global refinement of crystal parameters.

Alternatively, for some datasets, the program MOSFLM (Powell, 1999) was used to display images and for integrating the intensities, which were scaled using the program SCALA (Collaborative Computational Project, Number 4, 1994). These programs were run on an IBM PC with Fedora Core 5 as the linux operating system.

2.4.5 Assessment of the quality of the diffraction data

One of the major criteria to assess the quality of data is the resolution to which the crystal diffracts. There are two ways to determine the high resolution limit of diffraction. The first is the ratio of the intensity to the error in the measurement of intensity (I/σ(I)). Another commonly used indicator of the quality of X-ray diffraction data is the symmetry *R*-factor (R_{sym}), same as the merging *R*-factor (R_{merge}), which is the sum of the differences between the intensity measurements of the same or symmetry related reflections and the average value of the measurements divided by the sum of all the measurements and is given by the following formula:

$$R_{merge} = \Sigma_{hkl} \Sigma_i | I_i(hkl) - I(\overline{hkl})| / \Sigma_{hkl} \Sigma_i I_i(hkl) \quad (2.1)$$

Where $I_i(hkl)$ is the intensity of ith observation of (*h k l*) reflection, and $I(\overline{hkl})$ is the mean intensity of all measured symmetry equivalents.

From a statistical point of view, I/σ(I) is a superior criterion as by definition it is the signal to noise ratio of the data, whereas R_{merge} is not directly related to the signal-to-noise ratio. It has been argued that the R_{merge} is seriously flawed as it has an implicit dependence on the redundancy of the data (Diederichs and Karplus, 1997). Secondly, the σ(I) assigned to each intensity derives its validity from the χ^2's (or the goodness-of-fit). While scaling the data, the χ^2 is brought closer to 1.0 by adjusting the parameters of

the error model. *R*merge, on the other hand, is an unweighted statistical parameter which is independent of the error model. It is sensitive to both intentional and unintentional manipulation of the data used to calculate it, and may not directly correlate with the quality of the data.

2.4.6 Matthews number

Once the space group and unit cell dimensions of the crystal are known, it is possible to estimate the number of molecules in the crystallographic asymmetric unit and the solvent content of the protein crystals with the knowledge of the molecular weight of protein. The following equations (2.1 and 2.2) are used (Matthews, 1968).

$$V_m = (V \times z) / (MW \times n) \quad (2.2)$$

$$V_{solv} = 1- (1.23/ V_m) \quad (2.3)$$

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

For both the crystal forms of CAL, namely trigonal and orthorhombic, the number of molecules in the crystallographic asymmetric unit was estimated using Matthews Probability Calculator, with the resolution as an additional input (Kantardjieff and Rupp, 2003).

2.4.7 Sequence analysis of CAL and a search for homologous protein structure

Both the amplitudes and phases of reflections are required for determining the structure from X-ray diffraction data. In the diffraction pattern, however, only the intensities are recorded, from which only the amplitudes of the reflections can be measured. Phase information from the reflection data is lost. This is termed as the phase problem in X-ray crystallography. In macromolecular crystallography, *ab initio* phase

information is obtained by using isomorphous replacement and anomalous dispersion methods. Alternatively, if the structure of a homologous protein is known, it can be used as a search model to determine the orientation and position of the molecules within the unit cell. This method is known as Molecular Replacement (MR). MR method is the simplest of all structure solution methods and it depends on the presence of related structure and is particularly relevant in the case of homologous proteins with closely similar structures. It is a method of choice when a suitable homologous protein structure is available, as this does not require preparation of heavy atom derivatives of the crystals, which can be time consuming and often has comparatively low success rate. The pioneering studies of Rossman and Blow (1962) laid the foundation for successful use of MR method. As the number of available protein structures in the PDB is rapidly increasing, molecular replacement has become the most popular method for determining the structure of the unknown protein. The success of the MR method depends on the sequence and structural similarity between the search model and the unknown structure. The sequence of the unknown protein is used to find out a suitable structural model based on the sequence similarity.

N-terminal sequencing of CAL was carried out at the National Institute of Immunology, New Delhi. It has confirmed that the sequence of this protein is same as the previously reported sequence of the first 25 amino acids (Kolberg *et al.*, 1983). Hence, this 25-amino acid sequence was used to carry out a BLAST search (Altschul *et al.*, 1990) against a non-redundant database as well as with the sequences of structures deposited in the Protein Data Bank (PDB). The sequence hits obtained were aligned using ClustalW v. 1.82 (Thompson *et al.*, 1994). The structures of the proteins which showed the maximum sequence similarities with the CAL sequence were then used in the molecular replacement calculations. MR was tried with CCP4-based programs

AMoRe (Navaza and Saludjian, 1997), MolRep (Vagin and Teplyakov, 1997) and PHASER (McCoy *et al.*, 2005).

2.4.8 Preparation of heavy atom derivatives of CAL crystals

Out of the three methods used to calculate the phases of the unknown protein structure, anomalous dispersion and isomorphous replacement are the *ab initio* methods, which do not require any previous knowledge about the protein to phase the reflections. Based on the number of derivatives and/or the number of wavelengths used to get anomalous signal, the phase determination methods are termed as single isomorphous replacement (SIR) (Blow and Rossmann, 1961), single wavelength anomalous dispersion (SAD) (Hendrickson and Teeter, 1981), multiple isomorphous replacement (MIR) (Blow and Crick, 1959), multiwavelength anomalous dispersion (MAD) (Hendrickson, 1991), single isomorphous replacement and anomalous scattering (SIRAS), and multiple isomorphous replacement and anomalous scattering (MIRAS) (North, 1965).

Multiple isomorphous replacements

The method of multiple isomorphous replacements has a central role in determining the structure of an unknown protein or of a protein for which no homologous structure is available. It was first exploited successfully by Perutz and his coworkers to solve the structure of hemoglobin (Green *et al.*, 1954). This method involves the addition of heavy atom compounds (such as that of lead, mercury or platinum etc.) to the crystals, without disturbing the unit cell of the crystal. The heavy atom can bind at specific sites on protein molecules throughout the crystal. The binding of the heavy atom to protein results in changes in the intensities of the reflections, ideally without changing the diffraction pattern itself.

The change in the amplitude as a result of the heavy atom contribution to each of the reflections is given by

$$F_{PH} - F_P \approx F_H \cos(\alpha_{PH} - \alpha_H) \quad (2.4)$$

Where F_H is the amplitude contribution by the heavy atom to a reflection *hkl*, F_{PH} is the amplitude of the reflection *hkl* from heavy atom derivative, α_{PH} is the phase angle of the reflection *hkl* from heavy atom derivative, F_P is the amplitude of the reflection *hkl* from native protein and α_P is the phase angle of the reflection *hkl* from native protein.

$(F_{PH} - F_P)^2$ is proportional to the difference in intensities of the reflection *hkl* between heavy atom derivative and native crystal, that is, $(I_{PH} - I_P)$. If a diffraction pattern is calculated where the amplitude of each reflection is $(F_{PH} - F_P)^2$, it gives the diffraction pattern of the heavy atom (s) alone in the unit cell. From this diffraction pattern, the location of heavy atom (s) in the unit cell can be calculated (discussed in section 2.4.9).

The structure factor equation is:

$$\mathbf{F}_{hkl} = \sum_{j=1}^{n} f_j \exp\left(2\pi i \left(hx_j + ky_j + lz_j\right)\right) \quad (2.5)$$

Where f_j is the scattering factor of atom *j*, x_j, y_j and z_j are its fractional coordinates and *hkl* are the Miller indices of the reflection.

If $\mathbf{F}_P$ is the structure factor of a native reflection, $\mathbf{F}_{PH}$ is the structure factor of the corresponding derivative reflection and $\mathbf{F}_H$ is the structure factor for the heavy atom, then the relationship between these three vectors is given by:

$$\mathbf{F}_{PH} = \mathbf{F}_H + \mathbf{F}_P \quad (2.6)$$

Hence, $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$ (2.7)

$\mathbf{F}_H$ can be obtained if the heavy atom is located in the unit cell. Solving the above complex equation for $\mathbf{F}_P$, its phase angle (α_P) can be calculated (Fig. 2.1 A and B).

Anomalous dispersion

In this method, an anomalous scatterer (like selenium) is incorporated in the protein itself or the crystal and the amplitudes of the diffraction pattern are determined at different wavelengths of X-rays using a synchrotron source. When the incident X-ray wavelength is near the absorption edge of the heavy atom, a fraction of the radiation is absorbed by it and re-emitted with altered phase. This introduces a dispersive and an absorptive term to the atomic scattering factor and results in a violation of Friedel's law (Fig. 2.1 C). As a result, the intensities of reflections *hkl* and *–h-k-l* are no longer the same.

Getting the phases

In isomorphous replacement method, one heavy atom derivative is enough to determine the phase unambiguously for a centrosymmetric crystal. For non-centrosymmetric protein crystals, there is an ambiguity in phase determination. This can be represented using a construction devised by Harker (1956) (Fig. 2.2 A). This phase ambiguity can be resolved by using a minimum of two derivatives (Fig. 2.2 B). Although two derivatives are theoretically sufficient to resolve the phase ambiguity, it is useful to include phase information from additional derivatives to overcome the poor phases resulting from the experimental errors in the measurements of $\mathbf{F}_{PH}$, $\mathbf{F}_{P}$, $\mathbf{F}_{H}$ and α_{H}.

In the case of anomalous dispersion, the differences in the intensities of Friedel pairs are used to locate the anomalous scatterers (Fig. 2.2 C). The phase ambiguity can be overcome by using isomorphous replacement as well as anomalous data (Fig. 2.2 D). The strength of the anomalous signal depends on the wavelength of data collection (Hendrickson *et al.*, 1989; Hendrickson, 1991).

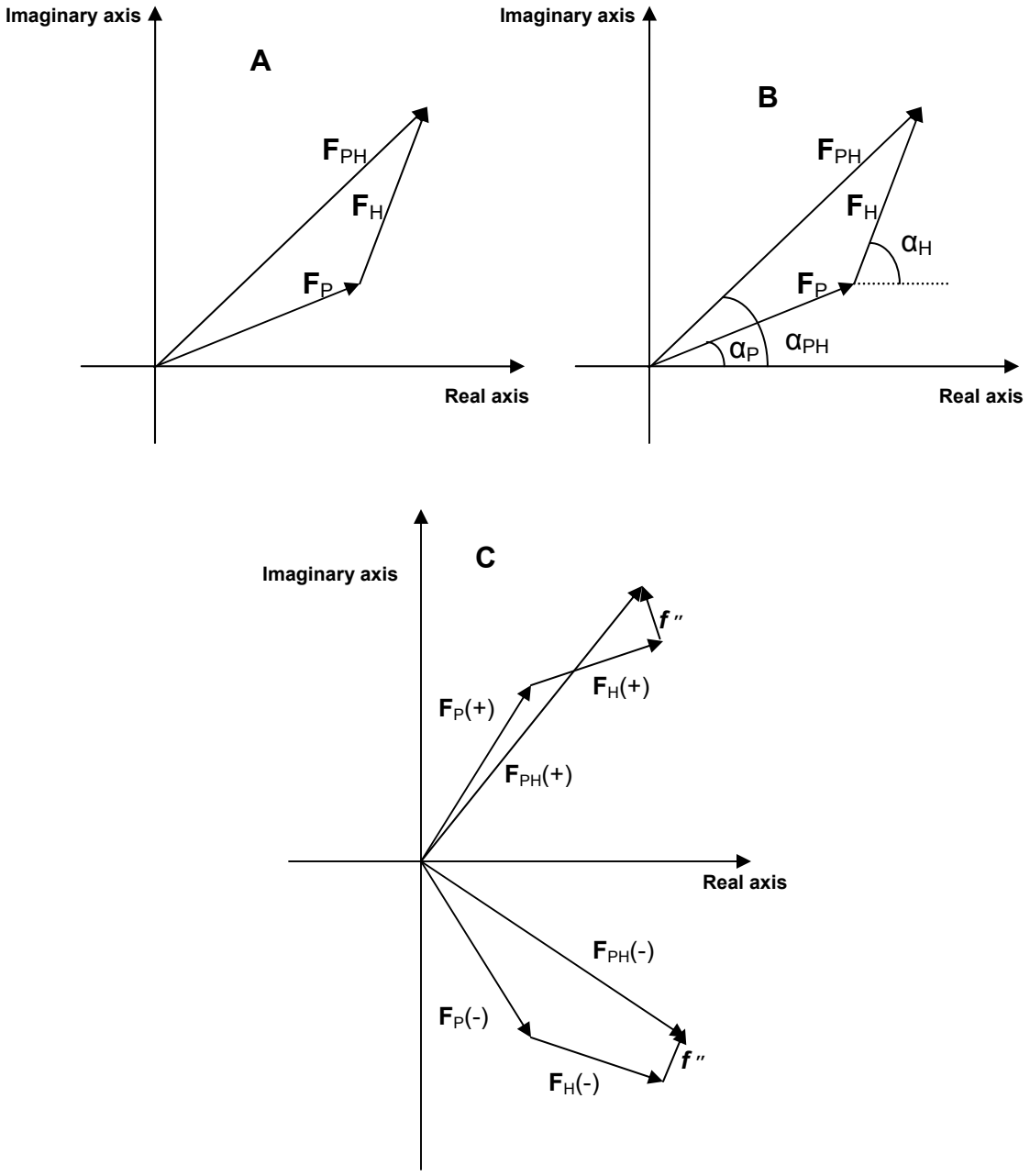
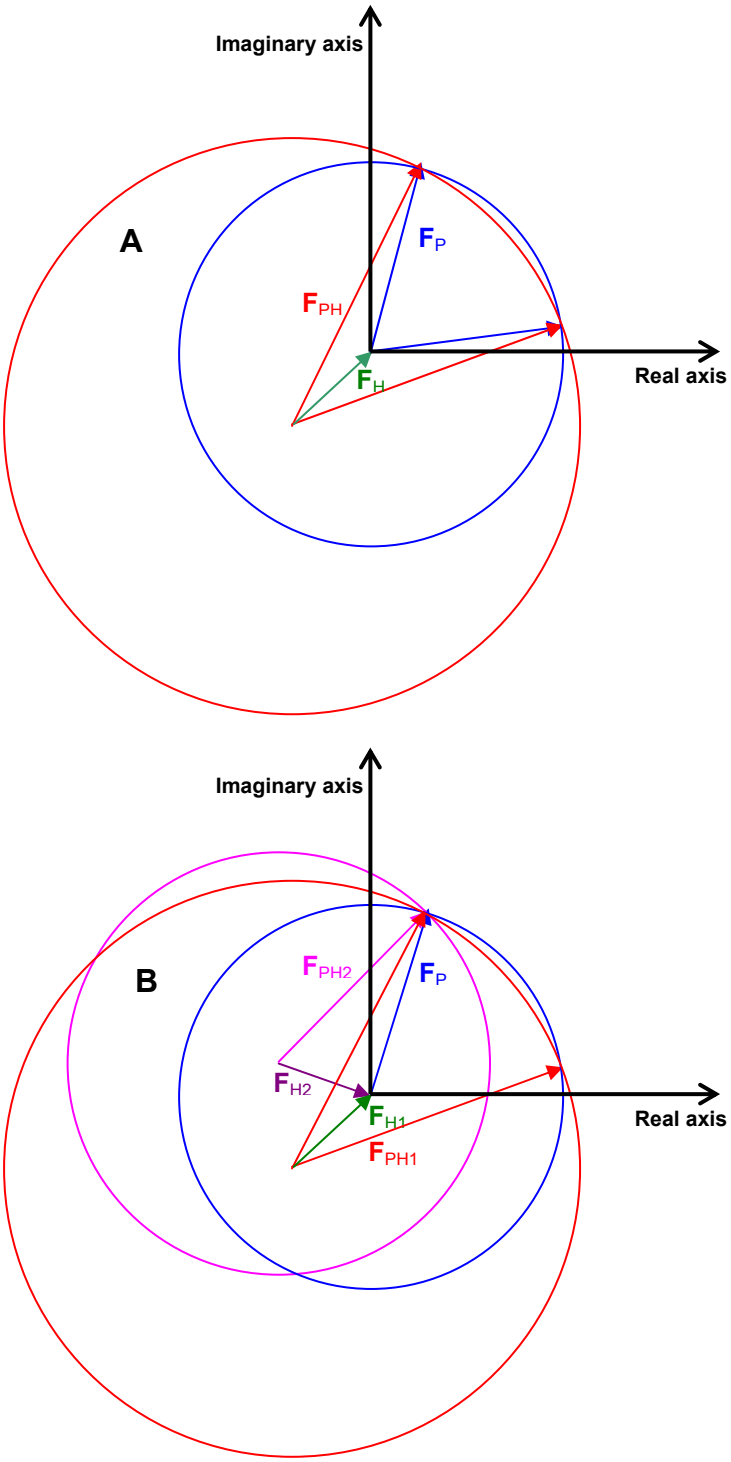


Fig. 2.1. (A) An Argand diagram showing the native protein ($\mathbf{F}_P$) and heavy atom ($\mathbf{F}_H$) contribution to the structure factor ($\mathbf{F}_{PH}$) of the heavy atom derivative of the protein. **(B)** Corresponding structure factor amplitudes and phases. **(C)** An Argand diagram showing breakdown on Friedel's law in the presence of an anomalous scatterer. $\mathbf{F}_{PH}(-)$ and $\mathbf{F}_{PH}(+)$,$\mathbf{F}_P(-)$ and $\mathbf{F}_P(+)$, and $\mathbf{F}_H(-)$ and $\mathbf{F}_H(+)$ are the pairs of Friedel-related amplitudes for protein, derivative and heavy atom respectively.


Imaginary axis
A
F_P
F_{PH}
F_H
Real axis
Imaginary axis
B
F_{PH2}
F_P
F_{H2}
F_{H1}
F_{PH1}
Real axis

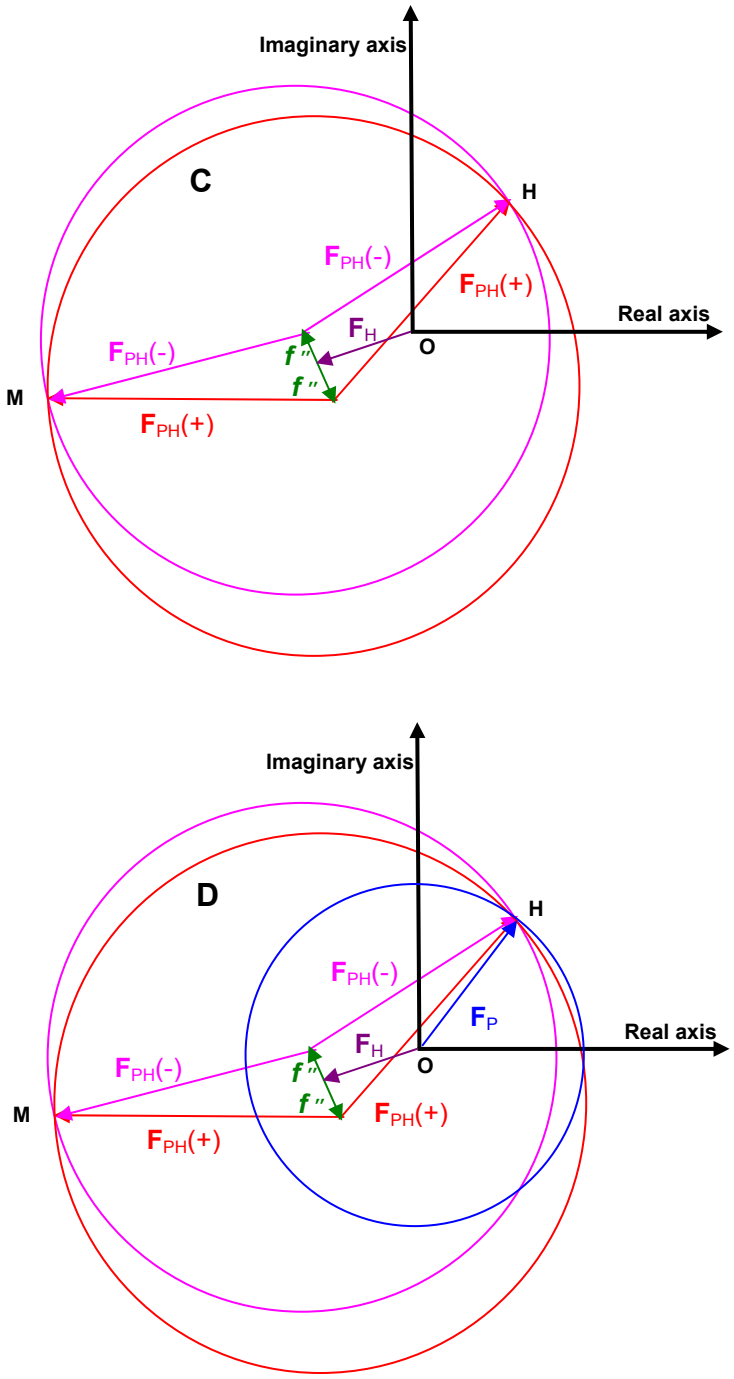


Fig. 2.2 Harker constructions illustrating **(A)** Phase ambiguity with a single derivative (SIR) **(B)** Overcoming the phase ambiguity using another derivative (MIR). **(C)** Phase determination using single-wavelength anomalous scattering (SAD). Two possibilities of $\mathbf{F}_P$ are given by OH and OM. **(D)** Phase determination using single isomorphous replacement and anomalous scattering (SIRAS).

The electron density of the protein structure can be calculated using inverse Fourier transform of the structure factors, which is given by the following equation:

$$\rho(x,y,z) = (1/V) \sum_{h} \sum_{k} \sum_{l} F_{hkl} \exp[-2\pi i (hx + ky + lz)] \quad (2.8)$$

Where V is the volume of the unit cell and the summation is from -∞ to ∞.

Since no model protein structure gave a solution in MR (discussed in detail in section 2.5.4), only MIR or MAD could be the alternative to solve the structure of CAL.

Heavy atom derivative of a protein crystal can be prepared either by controlled soaking of a pre-formed protein crystal in the mother liquor containing low concentrations (typically 1-5 mM) of the required heavy atom compound, or by co-crystallizing the protein along with the heavy atom compound. Both the methods were tried for preparing the heavy atom derivatives of CAL. Salts of various heavy metals, namely $HgCl_2$, $NiCl_2$, $CoCl_2$, $HAuCl_4$, $Pt(NO_3)_2Cl_2$, $RuCl_3$, $Pb(NO_3)_2$, KI, *p*-Hydroxy mercury benzoate, dichloro(ethylenediamine)platinum (II) were used in the crystallization solution at final concentrations of 1-5 mM. These salts were dissolved in water or buffer separately and added to the crystallization drop directly. For soaking experiments, the salts were dissolved in the mother liquor at a final concentration of 1-5 mM. The time of soak was varied from one to several hours.

2.4.9 Determination and refinement of heavy atom sites

Coordinates of heavy atoms or anomalous scatterer can be determined by direct methods (Schneider and Sheldrick, 2002) or Patterson based method (Patterson, 1934; Harker, 1956).

In Patterson based methods, the Patterson function $P(u,v,w)$ is used which needs only intensities and is given by the following formula:

$$P(u,v,w) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}|^2 \exp [-2\pi i (hu + kv + lw)] \quad (2.9)$$

Where V is the volume of the unit cell and the summation is from -∞ to ∞.

The Patterson function will have maximum values when u, v, w correspond to inter atomic vectors.

A difference Patterson in which amplitudes are squares of the difference between heavy atom and protein amplitudes can be used to detect heavy atom positions, as follows:

$$\Delta P(u,v,w) = (1/V) \sum_h \sum_k \sum_l |\Delta \mathbf{F}_{hkl}|^2 \exp [-2\pi i (hu + kv + lw)] \quad (2.10)$$

Where $(\Delta \mathbf{F})^2 = (|\mathbf{F}_{PH}| - |\mathbf{F}_P|)^2$

A difference Patterson map, which is a contour map of $\Delta P(u,v,w)$ displays peaks at locations corresponding to position vectors of heavy atoms. The Patterson map is much more complicated than an electron density map; however, for a simple structure like one or a few heavy atoms in the unit cell, the Patterson map is simple enough to locate the positions of heavy atoms.

Unit cell symmetry simplifies the search for peaks in a three-dimensional Patterson map. In a unit cell with a screw axis, vectors between symmetry related atoms all lie in Harker sections or Harker planes.

In order to locate the heavy atoms in the crystals of isomorphous derivative, several steps were followed, making use of different programs in the CCP4 suite for calculations (Fig. 2.3).

1. The intensities from Scalepack output were converted to structure factor amplitudes using the program Scalepack2mtz.

2. The datasets of the native and derivative crystals were merged together using the program CAD.
3. These datasets were scaled together using Scaleit to bring all data on the same scale and the heavy atom data were analyzed to estimate the strength of the heavy atom and anomalous signals.
4. The heavy atom vectors were found out using Patterson search (difference map as well as anomalous map). The Harker sections were plotted to check the peaks of vectors.
5. The vectors given by the Patterson search were analyzed to find the fractional coordinates of at least one heavy atom site. This site was then refined using the program MLPHARE, which refined the coordinates as well as the occupancy of the site.
6. The program MLPHARE was rerun by including new heavy atom sites, located using difference Fourier and phases calculated from the known sites.
7. The inter-atomic vectors between all the symmetry related heavy atoms were calculated and cross-checked in the Patterson map.

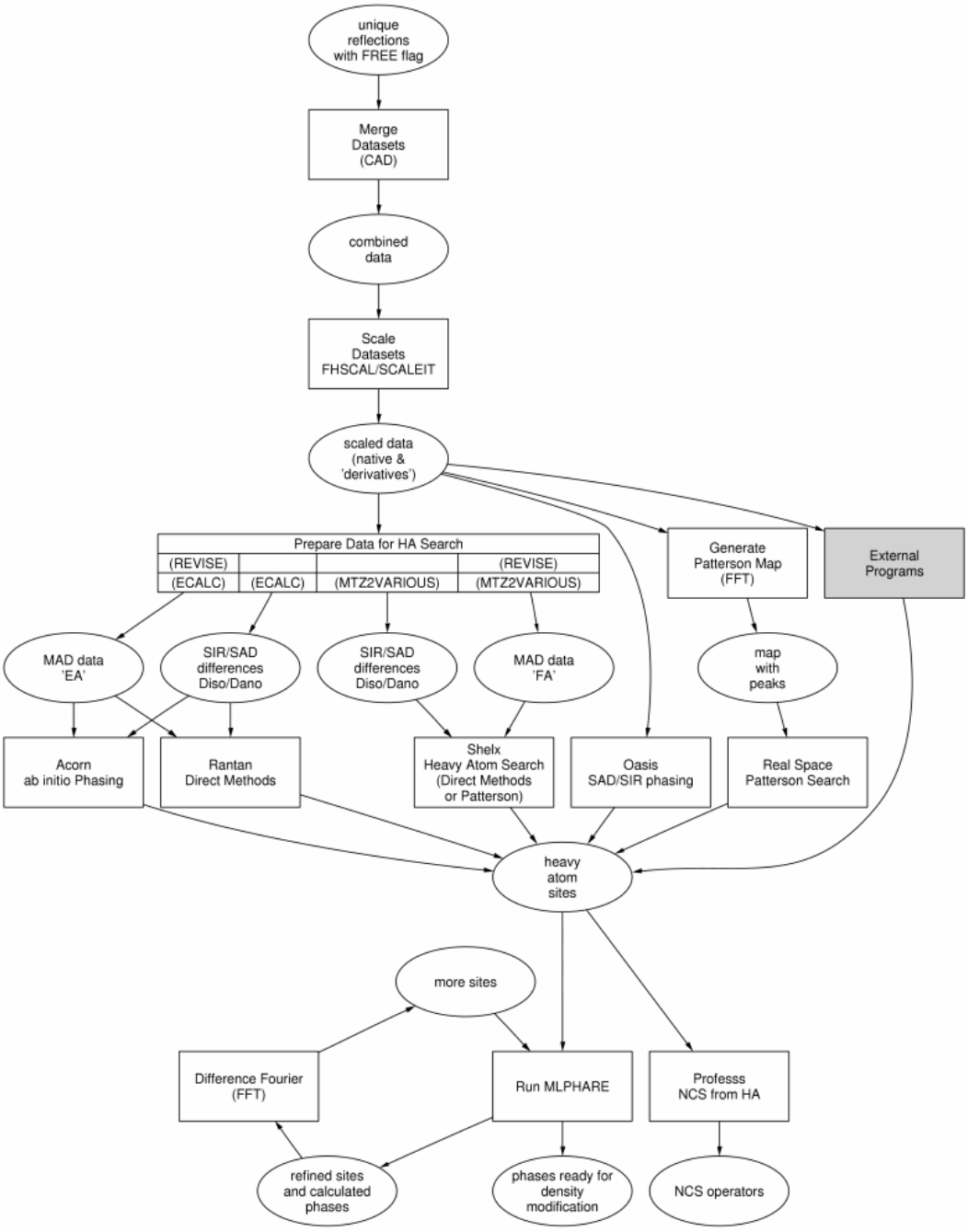


Fig. 2.3. Roadmap for experimental phasing in CCP4i (Reproduced from http://www.ccp4.ac.uk/dist/examples/tutorial/html/heavy-doc.html)

2.5. Results and Discussion

2.5.1 Crystallization of *Cicer arietinum* lectin

CAL was concentrated to 10-15 mg ml^{-1} in deionized water before setting up crystallization experiments. In the initial crystallization trials using the Crystal Screen 1 from Hampton Research, it was found to crystallize in the condition numbers 28 [consisting of 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 30%(w/v) polyethylene glycol (PEG) 8000], 40 [consisting of 0.1 M sodium citrate tribasic dihydrate pH 5.6, 20% v/v 2-propanol and 20% w/v polyethylene glycol 4,000] and 46 [consisting of 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 18% w/v polyethylene glycol 8,000].

The crystals grown in condition number 28 could be used to collect X-ray diffraction data. The crystals were of triangular or rhombohedral shape and grew to full size (0.6-0.7 mm in maximum dimension) within 6 days. These crystals belonged to the rhombohedral space group *R*3 (twinned) or trigonal space group *P*3. Later on, another crystal form of similar size was found to grow in the same condition, which had a parallelepiped shape. These crystals belonged to the orthorhombic space group $P2_12_12$. Sometimes, crystals belonging to both the above space groups were found to grow in the same drop.

Replacing the sodium cacodylate buffer with sodium phosphate buffer, maintaining the same molarity and pH, did not have any effect on crystal morphology and/or quality, but later it was found that phosphate ions interfered with most of the heavy atom salts causing them to precipitate. Hence crystallization experiments were continued with the sodium cacodylate buffer system. Change in pH and/or molarity of the buffer did not improve the quality of the crystal. Similarly, change in molarity of sodium acetate deteriorated the crystal quality. Reducing the concentration of polyethylene glycol from 30 % to 10-15 % improved the crystal quality as well as reduced the number

of crystals appearing in a drop. Concentrating the protein in the presence of salt and/or buffer yielded poor quality crystals.

It was observed that the protein has a tendency to aggregate in concentrated solutions and in the absence of salt. Hence a concentrated protein could not be used once concentrated, after 24 h for crystallization purposes. Every time, when a fresh batch of protein was dialyzed against deionized water, it had to be concentrated and used for crystallization within 24 h. To rule out the possibility of aggregation due to oxidation of cysteine residues, reducing agent such as dithiothreitol (DTT) or 2-mercaptoethanol (BME) was introduced in the crystallization solution at a final concentration of 1-5 mM. However, the presence of reducing agent led to the immediate and heavy precipitation of the protein instead of helping to overcome the precipitation of the protein.

2.5.2 X-ray characterization of CAL crystals

A. Rhombohedral crystals

The crystals obtained in condition number 28 of the Crystal Screen 1 of Hampton Research (Fig. 2.4) diffracted up to 2.3 Å. Crystal-to-detector distance was kept at 150 mm and 180 frames were collected with an oscillation angle of 0.5^0 per frame. These crystals belonged to the space group *R*3 (in hexagonal setting; also termed as *H*3) with unit-cell parameters $a = b = 81.2$, $c = 69.4$ Å and $\alpha = \beta = 90$, $\gamma = 120^0$. The merged data was 93.8 % complete with an R_{merge} of 10.3 %. No data beyond this resolution could be included as there were no visible reflections beyond this resolution limit, as well as further extending the resolution (to 2 Å) sharply increased the R_{merge} and correspondingly reduced the signal-to-noise ratio.

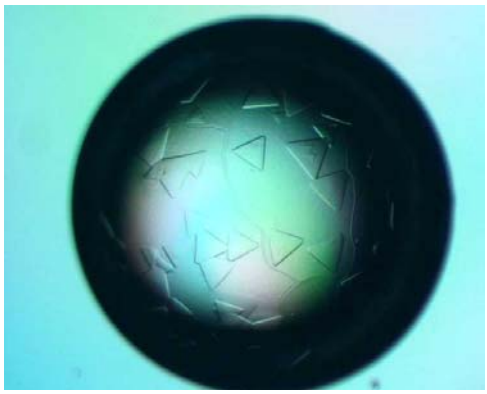


Fig. 2.4. Crystals of CAL grown in condition no. 28 (0.1 M sodium cacodylate buffer pH 6.5, 0.2 M sodium acetate and 30 % w/v polyethylene glycol 8000) of Crystal Screen of Hampton Research. The protein concentration was 15 mg ml^{-1}. The crystals belonged to the space group *R*3 and were found to be twinned.

Unfortunately, it was later found that this data set was perfectly twinned. Twinning is a special case of crystal intergrowth, where two or more crystalline domains grow in different orientation in the same crystal. Several types of twinning are known, classified according to the types of domains and their relative proportion in the crystal.

1. Non merohedral or epitaxial twinning

In this case, domain overlapping is observed in less than three dimensions. The diffraction pattern of such twinned crystal is an interlaced pattern of the two domains (Yeates, 1997). This type of twinning can be recognized from the diffraction pattern itself.

2. Merohedral twinning

In this type of twinning, the twin domains overlap in all three dimensions. This kind of twinning is observed in certain space groups where the crystal symmetries are less than that of the lattice symmetry. Depending on the twinning fraction, each diffraction intensity is the sum of the intensities from twin domains. Twinning fraction is the relative proportion of the volumes of each of the domains. A twin fraction of 0.5

corresponds to a case of perfect twinning, while twin fractions less than half is termed as partial twinning (Yeates, 1997). Table 2.1 enlists space groups that permit merohedral twinning and their apparent symmetries after twinning.

Table 2.1. Space groups which show merohedral twinning.

Space groups	Twin laws	Apparent symmetry*
*P*4, *P*4$_1$, *P*4$_2$, *P*4$_3$, *I*4, *I*4$_1$	k h -l	*P*4$_X$2$_X$2, *I*4 $_X$22
*P*3, *P*3$_1$, *P*3$_2$	-h –k l, k h –l, -k –h l	*P*6 $_X$, *P*3 $_X$21, *P*3 $_X$12
*R*3	k h –l	*R*32
*P*321, *P*3$_1$21, *P*3$_2$21	-h –k l	*P*6 $_X$22
*P*312, *P*3$_1$12, *P*3$_2$12	-h –k l	*P*6 $_X$22
*P*6, *P*6$_1$, *P*6$_2$, *P*6$_3$, *P*6$_4$, *P*6$_5$	k h –l	*P*6 $_X$22
*P*23, *P*2$_1$3, *I*23, *I*2$_1$3, *F*23	k –h l	*P*4$_X$32, *F*4 $_X$32, *I*4 $_X$32

* Apparent symmetries occur in case of perfect twinning. All possibilities of screw axes are denoted by subscript x.

The most common type of merohedral twinning for macromolecules is hemihedral twinning, where there are two distinct domains within the twinned crystal.

3. Pseudomerohedral twinning

This type of twinning occurs in space groups with fortuitous unit cell parameters (Parsons, 2003), for example a monoclinic cell with $\beta \approx 90^0$ or an orthorhombic cell with almost equal cell lengths.

4. Tetartohedral twinning

This is a rare case of twinning in protein crystals, where four domains exist in a single crystal (Gayathri *et al.*, 2007).

The presence of a twinned crystal can be detected by the statistical properties of the collected data as well as by other symptoms of twinning like:

I. Packing density

The assignment of wrong space group for a twinned crystal might lead to abnormally low values of Matthews coefficient due to apparently higher symmetry.

II. Intensity statistics

The intensities of reflections from a twinned crystal are the result of the sum of intensities from two crystal domains, hence an averaging effect is observed. Hence, the number of weak reflections and strong reflections will be less in the data from a twinned crystal, leading to a sigmoidal appearance of the cumulative intensity distribution rather than hyperbolic. The expected values of the *k*th moments of normalized intensities for a twinned crystal and an untwined crystal also differ. The ratio $\langle I^2 \rangle / \langle I \rangle^2$ for acentric reflections in narrow resolution bins is expected to be 2.0 for twinned crystals and 1.5 for untwined crystals.

Using these parameters, twinning was detected in the CAL crystals belonging to the space group *R*3. Cumulative intensity distribution for observed acentric reflections is distinctly sigmoidal (Fig. 2.5). The moments of the distribution were calculated from the second to the fourth order with the usual analysis programs (*TRUNCATE*; Collaborative Computational Project, Number 4, 1994), and the experimental values were compared with the expected theoretical values. In case of this dataset, the values of second, third and fourth moments were found to be exactly matching with those for the data from twinned crystals. A comparison of the data of twinned crystals with that of untwined crystals is shown in Fig. 2.6, 2.7 and 2.8. To obtain an accurate estimate of the twinning fraction, Britton plot (Britton, 1972) has been calculated (Fig. 2.9) using the program DETWIN (Collaborative Computational Project, Number 4, 1994). This plot shows the number of negative intensities after de-twinning as a function of the twinning fraction. Extrapolation of the linear part of the plot gives the twinning fraction, which was found out to be 0.36.

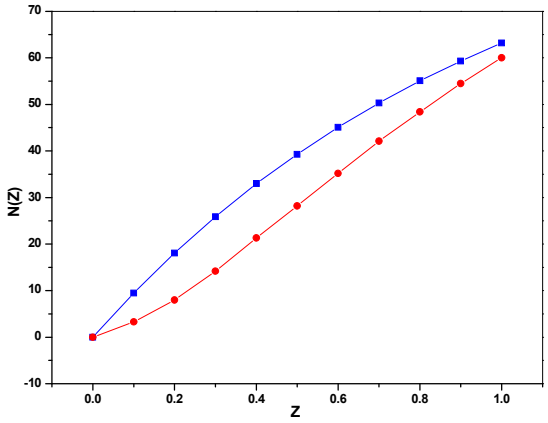


Fig. 2.5. Cumulative intensity plot for the twinned dataset belonging to the space group *R*3. -■- acentric theoretical (untwined), -●- acentric observed indicating twinning in the crystal.

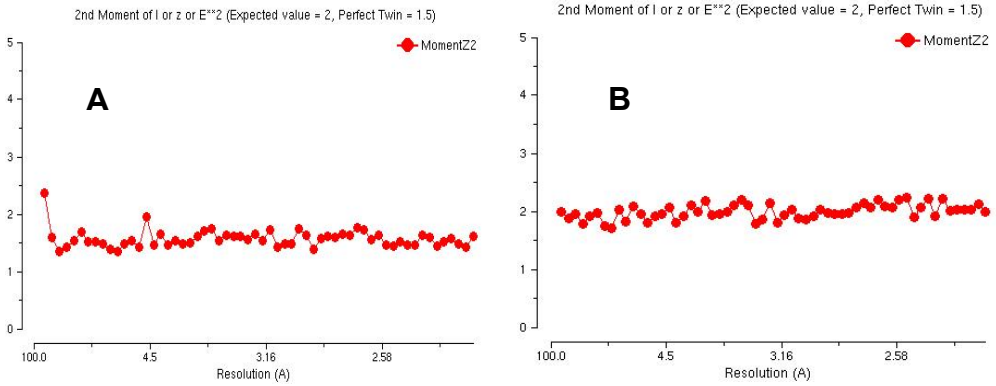


Fig. 2.6. 2nd moment of I or Z or E^2. **(A)** For a twinned crystal belonging to space group *R*3. **(B)** For an untwined crystal belonging to space group *P*3.

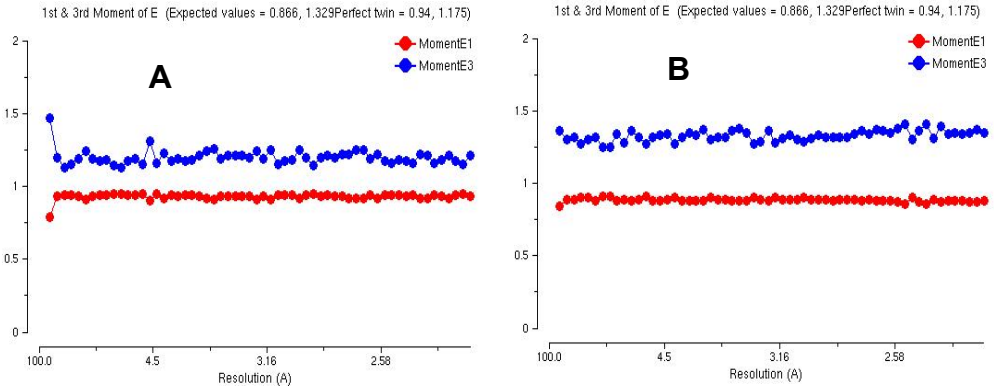


Fig. 2.7. 1st and 3rd moments of E. **(A)** For a twinned crystal belonging to space group *R*3. **(B)** For an untwined crystal belonging to space group *P*3.

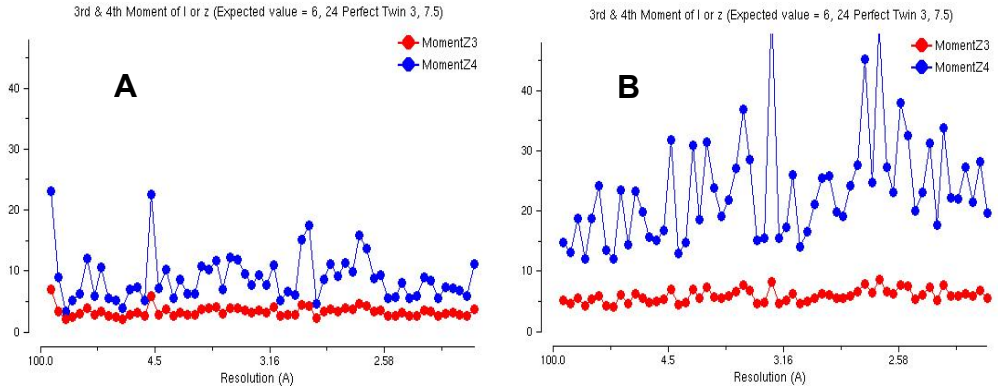


Fig. 2.8. 3rd and 4th moments of I or Z. **(A)** For a twinned crystal belonging to space group *R*3. **(B)** For an untwined crystal belonging to space group *P*3.

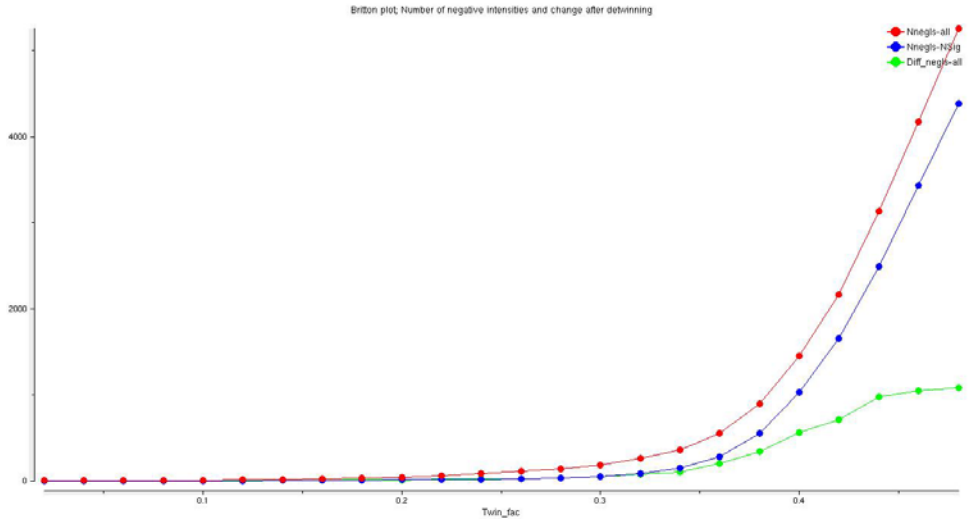


Fig. 2.9. Britton plot for the twinned crystal belonging to the space group *R*3.

Twinning observed in CAL crystals belonging to space group *R*3 might not be typical merohedral twinning, as untwined crystals belong to the space group *P*3 and perfect merohedral twinning would have given rise to the space groups $P6_X$, $P3_X21$, $P3_X12$. It does not seem to be psuedomerohedral twinning as well. It could be a case of epitaxial twinning. So far, no report has been found on conversion of *P*3 space group to *R*3 or *H*3 due to twinning.

Detection of twinning can help in solving the structure with descent *R* factors. However, if the twinning remains undetected, the structure may be solved and refined, although the *R* factors will remain at unacceptably high values. A search was carried out to locate such structures in the Protein Data Bank, belonging to the space group *H*3 which have high R values. Table 2.2 enlists some such structures, which may have a potential problem of twinning.

Table 2.2. A list of structures in Protein Data Bank belonging to space group *H*3 (*R*3 in hexagonal settings) with *R* factor and/or R_{free} higher than 0.3.

PDB ID	a	b	c	Resolution (Å)	*R* value	R_{free}	Reference
1DOA	83.9	83.9	191.2	2.60	0.257	0.320	Hoffman *et al*., 2000.
1GH7	185.7	185.7	103.3	3.00	0.267	0.307	Carr *et al.*, 2001.
1IF1	84.8	84.8	203.7	3.00	0.242	0.309	Escalante *et al.*, 1991
1JA3	91.7	91.7	89.6	3.00	0.333	0.308	Dimasi *et al.*, 2002.
1M57	340.7	340.7	89.8	3.00	0.293	0.329	Svensson-Ek *et al.*, 2002
1MG1	102.9	102.9	118.0	2.50	0.223	0.305	Kobe *et al.*, 1999.
1OM7	182.9	182.9	37.6	2.80	0.237	0.303	Ravaud *et al.*, 2003
1QZU	124.8	124.8	153.5	2.91	0.293	0.342	Manoj and Ealick, 2003.
1TJR	123.0	123.0	98.8	2.30	0.285	0.338	Kulik *et al.*, 2005
1TQQ	265.3	265.3	96.3	2.75	0.266	0.305	Higgins *et al.*, 2004
1VI7	148.8	148.8	35.6	2.80	0.261	0.332	Park *et al.*, 2004
1YQ8	52.2	52.2	234.8	2.60	0.261	0.300	Merckel *et al.*, 2005
2D5G	125.6	125.6	115.4	3.20	0.238	0.314	-
2PK2	203.8	203.8	124.8	2.67	0.272	0.306	Anand *et al.*, 2007

B. Trigonal crystals

By reducing the concentration of PEG 8000 in the crystallization condition from 30% to 10-15% the experiment yielded normal crystals which belonged to the space group *P*3. The crystals had a triangular or hexagonal shape (Fig. 2.10).

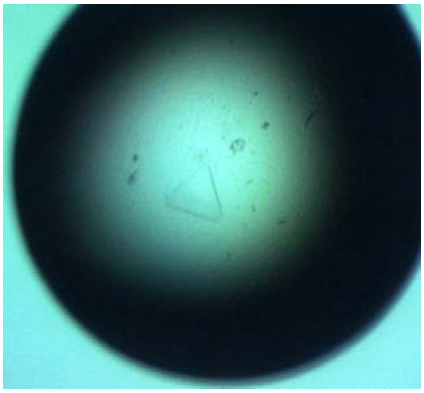


Fig. 2.10. Triangular crystal of CAL grown in 0.1 M sodium cacodylate buffer pH 6.5, 0.2 M sodium acetate and 13 % w/v polyethylene glycol 8000. The protein concentration was 10 mg ml^{-1}. The crystal belonged to the space group *P*3.

C. Orthorhombic crystals

Crystals of CAL which belonged to the orthorhombic space group $P2_12_12$ were found to grow in the same conditions where crystals belonging to the space group *P*3 used to grow. These crystals could be distinguished from trigonal ones based on the morphology and were rectangular/rhombus shaped (Fig. 2.11). The data collection statistics for the rhombohedral, trigonal and orthorhombic crystals are shown in Table 2.3.

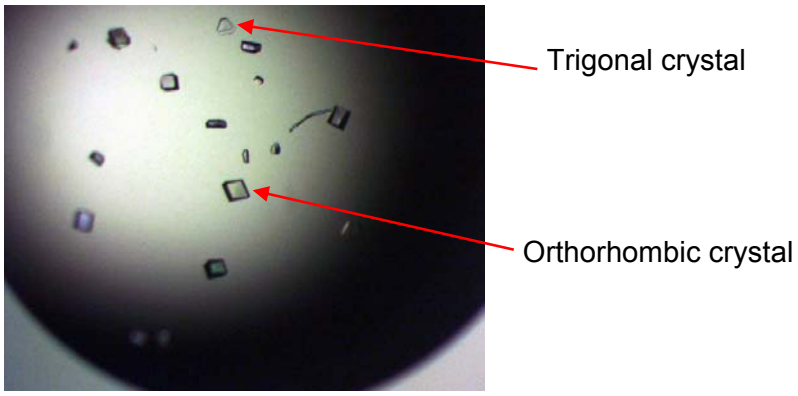


Fig. 2.11. Orthorhombic (rectangular/rhombus shaped) and trigonal (triangular shaped) crystals of CAL grown in the same drop, containing 0.1 M sodium cacodylate buffer pH 6.5, 0.2 M sodium acetate and 12 % w/v polyethylene glycol 8000. The protein concentration was 10 mg ml^{-1}. The orthorhombic crystals belonged to the space group $P2_12_12$.

Table 2.3. Data collection statistics for the native crystals of CAL

Temp ^{o}C	~ 22	-146	~ 22
Crystal-to-detector Distance mm	150	150	200
No. of frames	1-180	1-240	1-300
Space group	***R*3**	***P*3**	***P*2$_1$2$_1$2**
Unit cell *a* Å	81.20	80.21	70.93
Unit cell *b* Å	81.20	80.21	73.32
Unit cell *c* Å	69.40	69.14	86.98
Unit cell α^{o}	90	90	90
Unit cell β^{o}	90	90	90
Unit cell γ^{o}	120	120	90
Unit cell volume Å^3	396613.0	385265.3	452331.3
Resolution range Å	30.00-2.30 (2.38-2.30)	30.00-2.30 (2.38-2.30)	20.00-2.60 (2.69-2.60)
Reflections, total	126672	425948	556973
Reflections, unique	7121	22003	14456
Completeness %	93.8 (97.9)	86.7 (87.9)	99.9 (99.7)
R_{factor}	0.10 (0.19)	0.06 (0.26)	0.11 (0.33)
Average I/ σ<I>	8.8 (4.3)	9.33 (2.32)	8.97 (2.84)
Matthews Coefficient Å^3/Da	2.05	2.99	2.63
No. of monomers per unit cell (*Z*)	9	6	8
Solvent content %	39.9	58.8	53.2

2.5.3 Sequence analysis of CAL using BLAST and ClustalW and a search for molecular replacement model

The *N*-terminal sequence of CAL determined is T-N-F-G-Y-I-N-A-A-F. The sequence reported by Kolberg *et al.* (1983) is T-N-F-G-Y-I-N-A-A-F-R-S-S-X-N-N-E-A-Y-L-F-I-N-G-K. Since the sequence determined by us matched well with this, the 25 amino acid sequence of CAL was used to carry out the BLAST search. Initially, the BLAST search carried out against the entire non-redundant protein database of NCBI disclosed a match at 90% identity with the *N*-terminal sequence of a major seed albumin (PA-2) from *Pisum sativum* (National Center for Biotechnology Information (NCBI) accession no. P08688; Higgins *et al.*, 1987) (Fig. 2.12).

```
CLUSTAL W (1.83) multiple sequence alignment

PA-2      MTKTGYINAAFRSSQTNEAYLFINDKYVLLDYAPGTSNDKVLYGPTPVRDGFKSLNQTVF 60
CAL       -TNFGYINAAFRSSXTNEAYLFINGK---------------------------------- 25
           *: ********** *********.*

PA-2      GSYGVDCSFDTDNDEAFIFYEKFCALIDYAPHSNKDKIILGPKKIADMFPFFEGTVFENG 120
CAL       ------------------------------------------------------------

PA-2      IDAAYRSTRGKEVYLFKGDQYARIDYETNSMVNKEIKSIRNGFPCFRNTIFESGTDAAFA 180
CAL       ------------------------------------------------------------

PA-2      SHKTNEVYFFKGDYYARVTVTPGATDDQIMDGVRKTLDYWPSLRGIIPLEN 231
CAL       -------------------------------------------------
```

Fig. 2.12. Alignment of *C. arietinum* lectin (CAL) *N*-terminal sequence with that of albumin 2 (PA-2) from *P. sativum* (NCBI accession no. P08688). (The identical amino acids are highlighted in red and marked with a star and the conservatively substituted amino acids are in blue and marked with a colon). The alignment was produced using *ClustalW* v.1.82 (Thompson *et al.*, 1994).

Since no three-dimensional structure of this albumin is available, its full sequence was used to compare with the sequences of structures deposited in the Protein Data Bank (PDB) to search for a model structure. In the BLAST search, putative conserved

domain (CD) (cd00094, HX, Hemopexin-like repeats) was detected to be present in the PA-2 sequence (Marchler-Bauer and Bryant, 2004). This CD contains four instances of the repeat. Hemopexin is a serum glycoprotein that binds heme and transports it to the liver for breakdown and iron recovery. Hemopexin-like repeats occur in vitronectin and some matrix metalloproteinases family (matrixins). The HX repeats of some matrixins bind tissue inhibitor of metalloproteinases (TIMPs). Presence of Hemopexin-like repeats might be conferring hemin binding property to CAL.

The BLAST search using PA-2 sequence against the PDB sequences database gave a match at 28% identity with the *C*-Terminal Domain (Haemopexin-Like Domain) of Gelatinase A (PDB code 1gen; Libson *et al.*, 1995) (Fig. 2.13) and Human Matrix Metalloproteinase-2 (PDB code 1rtg; Gohlke *et al.*, 1996)

When a BLAST search was carried out using the sequence of PA-2 against a non-redundant database, a considerable match was found with the sequence of Hemopexin and D-tyrosyl-tRNA (Tyr) deacylase of *Medicago truncatula*, a model organism used to study legume biology (Fig. 2.14). However, the three-dimensional structures of both these proteins are not available. Moreover, the structure of only one probable eukaryotic D-tyrosyl-tRNA (Tyr) deacylase from *Leishmania major* (PDB code 1TC5) is available, which does not show any significant sequence similarity with either PA-2 or D-tyrosyl-tRNA (Tyr) deacylase from *Medicago truncatula*, hence could not be used as a model for molecular replacement method.

Using the structures of 1RTG and 1GEN as models, molecular replacement was tried with CCP4-based programs AMoRe, MolRep and PHASER. None of these attempts yielded a clear molecular replacement solution.

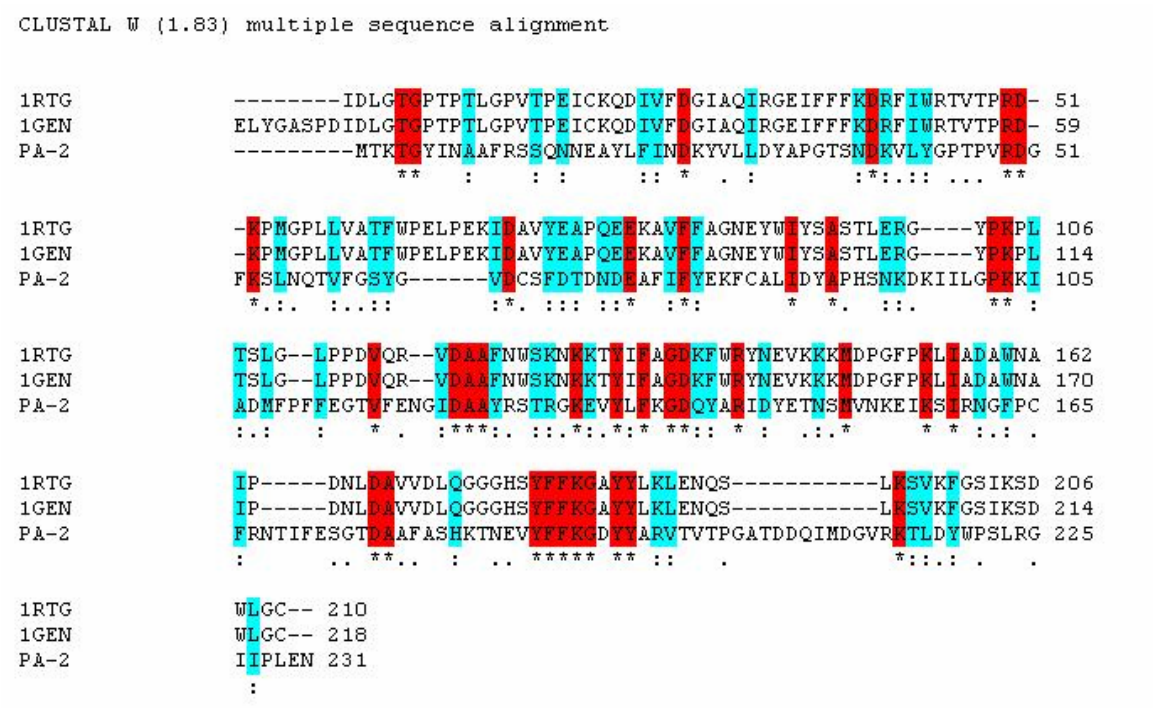


Fig. 2.13. Alignment of Pea albumin (PA-2) sequence with the sequences of Human Matrix Metalloproteinase-2 (PDB code 1RTG) and the *C*-terminal domain of gelatinase A (PDB code 1GEN). (The identical amino acids are highlighted in red and marked with a star and the conservatively substituted amino acids are in blue and marked with a colon in this figure as well as Fig. 10.). The alignment was produced using *ClustalW* v.1.82 (Thompson *et al.*, 1994).

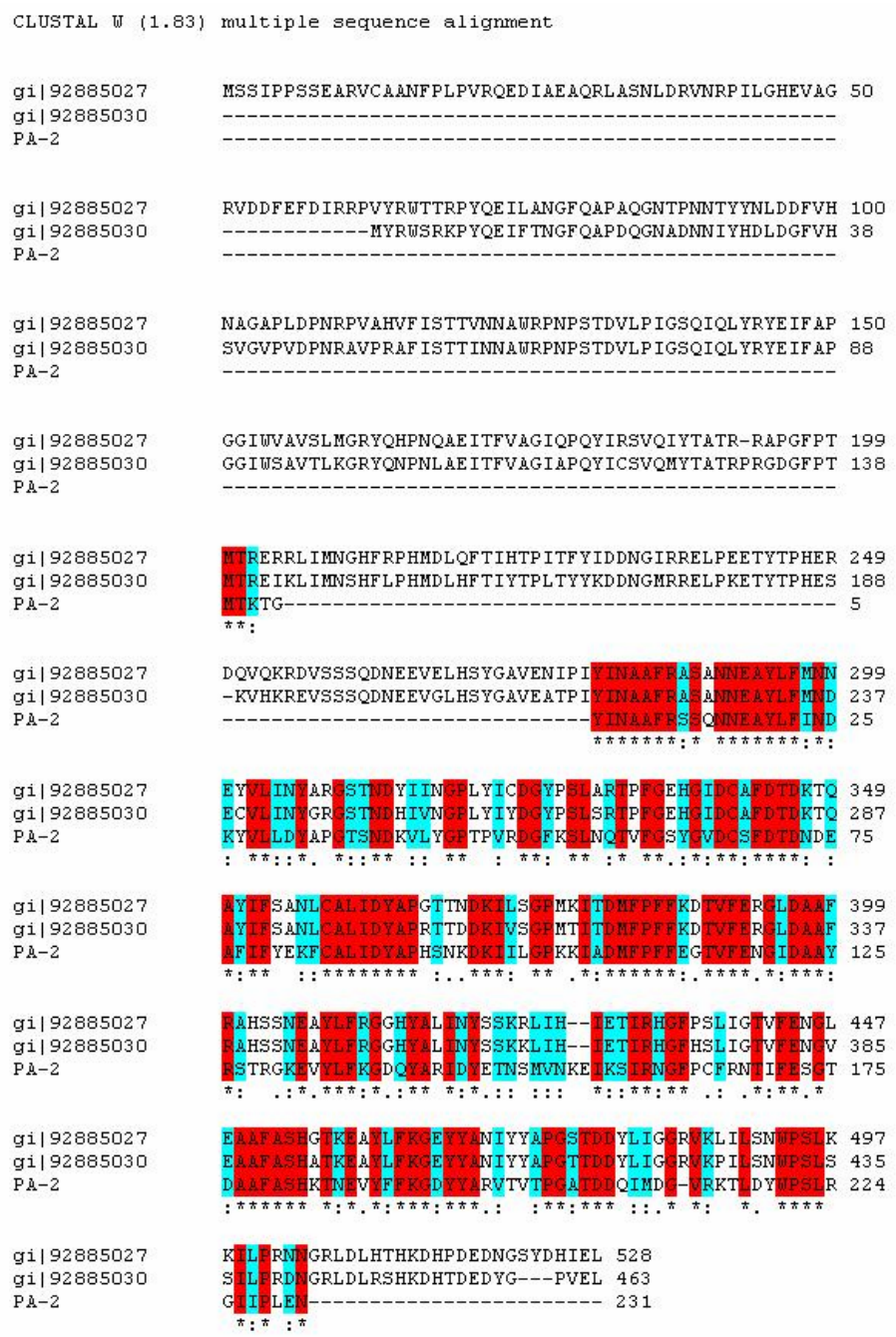


Fig. 2.14. Alignment of Pea albumin 2 (PA-2) with gi|92885027 (Hemopexin; D-tyrosyl-tRNA(Tyr) deacylase [*Medicago truncatula*]) and gi|92885030 (Hemopexin [*Medicago truncatula*]). The alignment was produced using *ClustalW* v.1.82 (Thompson *et al.*, 1994).

2.5.4 Heavy atom derivatives of CAL crystals

Since the molecular replacement method was not successful in the case of CAL, next option was to try MIR by preparing heavy atom derivatives. To prepare heavy atom derivatives of CAL crystals, either the heavy atom salt was included in the crystallization experiment (co-crystallization) or a pre-formed crystal was soaked in the mother liquor containing heavy atom salt (soaking). It was observed that co-crystallization experiments yielded better quality crystals than soaking experiments, soaking deteriorated the quality of the crystals. CAL crystallizes in the presence of 5 mM potassium iodide or 1-2 mM of *p*-hydroxy mercury benzoate, $HAuCl_4$, $Pt(NO_3)_2Cl_2$, $Pb(NO_3)_2$ and hemin. Only in one soaking experiment, where a single crystal of CAL was soaked in the mother liquor containing 1 mM $Pb(NO_3)_2$, a good diffraction data could be obtained. In all experiments where salts of Hg, Pt and Au were used, a quick deterioration of the crystals was the result.

The data collection statistics for CAL crystals soaked/grown in the presence of heavy atom compounds has been shown in Table 2.4 (trigonal crystals) and Table 2.5 (orthorhombic crystals).

Table 2.4. Data collection statistics for the CAL crystals soaked/grown in the presence of heavy atom compounds in the space group *P*3.

Heavy atom used	Pb (soak)	Pb (cocrystal)	Hg
Temp ^{0}C	~ 22	~22	-169
Crystal-to-detector Distance mm	150	150	200
No. of frames	1-360	1-300	1-360
Space group	***P*3**	***P*3**	***P*3**
Unit cell *a* Å	81.90	81.89	80.37
Unit cell *b* Å	81.90	81.89	80.37
Unit cell *c* Å	69.60	69.60	69.21
Unit cell $\alpha = \beta$ ^{O}	90	90	90
Unit cell γ ^{O}	120	120	120
Unit cell volume Å^3	404249.3	404120.9	387147.5
Resolution range Å	20.00-2.30 (2.38-2.30)	20.00-2.80 (2.90-2.80)	20.00-2.60 (2.69-2.60)
Reflections total	453192	229026	475529
Reflections unique	23190	12868	15370
Completeness %	99.5 (98.9)	100 (100)	98.5 (96.8)
R_{factor}	0.08 (0.19)	0.14 (0.32)	0.09 (0.32)
Average I/ σ<I>	9.72 (4.57)	9.25 (4.25)	8.02 (2.35)
Matthews Coefficient Å^3/Da	3.1	3.1	3.0
No. of molecules per unit cell (*Z*)	6	6	6
Solvent content %	61	61	59

Table 2.5: Data collection statistics for the CAL crystals grown in the presence of heavy atom compounds in the space group $P2_12_12$.

Heavy atom used	Iodine	Pt	Au
Temp ^{0}C	RT	RT	RT
Crystal-to-detector Distance mm	150	200	200
No. of frames	1-300	1-360	1-360
Space group	***P*2$_1$2$_1$2**	***P*2$_1$2$_1$2**	***P*2$_1$2$_1$2**
Unit cell *a* Å	71.25	70.96	71.12
Unit cell *b* Å	73.36	73.93	73.34
Unit cell *c* Å	87.17	87.11	86.92
Unit cell $\alpha = \beta = \gamma$ ^{O}	90	90	90
Unit cell volume Å^3	455621.6	456981.6	453357.3
Resolution range Å	20.00-2.20 (2.28-2.20)	20.00-2.80 (2.90-2.80)	20.00-2.90 (3.00-2.90)
Reflections total	598857	304376	303430
Reflections unique	23739	11752	10554
Completeness %	99.9 (99.6)	100 (100)	100 (100)
R_{factor}	0.10 (0.29)	0.10 (0.26)	0.12 (0.26)
Average I/σ<I>	8.31 (3.17)	10.72 (4.54)	9.01 (4.69)
Matthews Coefficient Å^3/Da	2.7	2.7	2.6
No. of monomers per unit cell (*Z*)	8	8	8
Solvent content %	53	54	53

2.5.5 Determination of heavy atom sites and refinement of heavy atom parameters

By calculating a difference Patterson using the program for performing Fast Fourier Transform (FFT), it was observed that only the iodine derivative of the CAL crystal has formed. From the vector peaks, the coordinates of the iodine atom were calculated. These coordinates were refined with the program MLPHARE which gave coordinates of all the possible iodine sites in the asymmetric unit.

The iodine derivative of CAL crystal belonged to the space group $P2_12_12$. The equivalent points for this space group are:

Atom 1: x, y, z

Atom 2: -x, -y, z

Atom 3: 0.5-x, 0.5+y, -z

Atom 4: 0.5+x, 0.5-y, -z

The Harker sections are given by:

Vector 1: Atom 1 → 2 = 2x, 2y, 0

Vector 2: Atom 1 → 3 = 2x-0.5, 0.5, 2z

Vector 3: Atom 1 → 4 = 0.5, 2y-0.5, 2z

Vector 4: Atom 2 → 3 = 0.5, 0.5-2y, 2z

Vector 5: Atom 2 → 4 = 0.5-2x, 0.5, 2z

In the difference Patterson map for the iodine derivative (Table 2.6), three vector peaks could be located within the limits of maps plotted.

Peak 1: 0.36, 0.44, 0

Peak 2: 0.14, 0.5, 0.12

Peak 3: 0.5, 0.06, 0.12

Table 2.6. Difference Patterson vectors for iodine derivative of CAL crystal from FFT output.

1	1	1	102.53	0	0	0	0.0000	0.0000	0.0000	0.00	0.00	0.00
2	17	17	24.86	30	36	44	0.4209	0.5000	0.5000	29.99	36.68	43.59
3	13	13	9.44	31	0	33	0.4367	0.0000	0.3738	31.11	0.00	32.59
4	10	10	9.39	36	5	11	0.5000	0.0652	0.1252	35.62	4.78	10.91
5	16	16	7.99	5	5	44	0.0650	0.0646	0.5000	4.63	4.74	43.59
6	9	9	6.11	10	36	11	0.1400	0.5000	0.1224	9.97	36.68	10.67
7	11	11	5.81	0	36	12	0.0000	0.5000	0.1336	0.00	36.68	11.65
8	8	8	5.59	36	31	0	0.5000	0.4345	0.0000	35.62	31.87	0.00
9	12	12	5.55	6	31	33	0.0826	0.4314	0.3750	5.88	31.65	32.69
10	4	4	5.37	5	0	0	0.0701	0.0000	0.0000	4.99	0.00	0.00
11	7	7	5.00	36	7	0	0.5000	0.0951	0.0000	35.62	6.97	0.00
12	5	5	4.50	26	31	0	0.3547	0.4360	0.0000	25.27	31.98	0.00
13	3	3	4.06	0	25	0	0.0000	0.3512	0.0000	0.00	25.76	0.00
14	18	0	3.94	36	6	44	0.5000	0.0833	0.5000	35.62	6.11	43.59
15	2	2	3.80	0	10	0	0.0000	0.1415	0.0000	0.00	10.38	0.00
16	6	6	3.17	34	0	0	0.4705	0.0000	0.0000	33.52	0.00	0.00
17	15	15	3.16	36	4	40	0.5000	0.0525	0.4515	35.62	3.85	39.36
18	14	14	3.05	26	36	38	0.3597	0.5000	0.4349	25.63	36.68	37.91

Assuming that Peak 1 (0.36, 0.44, 0) is equivalent to the first Harker vector (2x, 2y, 0), the x and y coordinates of the first iodine binding site were found out to be 0.18 and 0.22. From Peak 2 and 3, which are equivalent to the second and third Harker vectors respectively, the z coordinate of this site was found out to be 0.06. This first site was refined with the program MLPHARE. The occupancy of this site was refined around 0.32. Difference Fourier gave another possible site of bound iodine, (0.74, 0.22, 0.44). The occupancy of this site was refined to 0.34. The phasing statistics for the iodine derivative has been shown in Table 2.7.

Schematic representation of these two atoms and their equivalence points has been shown in Fig. 2.15. Since the asymmetric unit is a dimer the two heavy atom sites could be due to iodine binding at the same site of two monomers of the dimer. This also explains the nearly equal occupancies of the sites. Thus it may be concluded that CAL monomer has only one iodine binding site.

Table 2.7. Phasing statistics for CAL iodine derivative.

Parameter	Acentric	Centric
Isomorphous differences	25.5	36.3
Lack of closure	21.4	26.5
$^{\#}R_{cullis}$	0.84	0.73
* Phasing power	1.24	0.92
Figure of Merit (FOM)	0.23	0.39

$^{\#}\ R_{cullis} = \Sigma\ ||\mathbf{F}_{PH} \pm \mathbf{F}_{P}| - |\mathbf{F}_{H(calc)}|| \ / \ \Sigma\ |\mathbf{F}_{PH} \pm \mathbf{F}_{P}|$

* Phasing power = $[\Sigma\ \mathbf{F}_{H}^{2} / \Sigma\ (|\mathbf{F}_{PH(obs)}| - |\mathbf{F}_{PH(calc)}|)^{2}]^{1/2}$

$\mathbf{F}_{PH}$ and $\mathbf{F}_{H}$ are the structure factors for the derivative and heavy atom substructure, respectively.

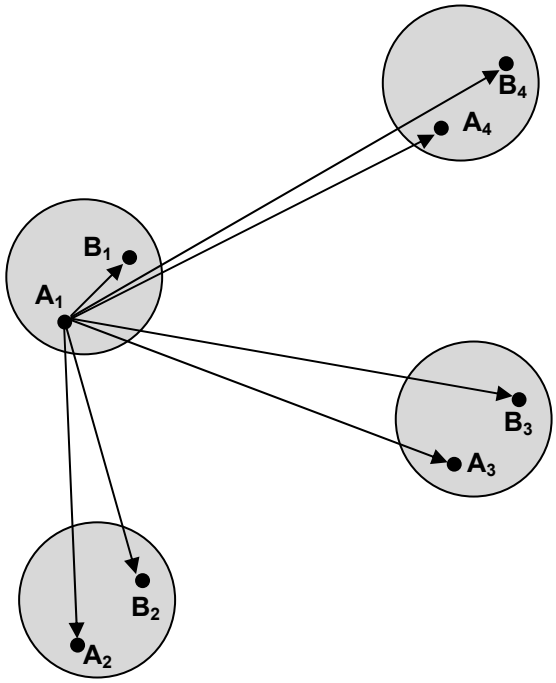


Fig. 2.15. Schematic representation of two iodine atoms bound to CAL molecules, termed as A and B. Each sphere represents a dimer. The equivalent points and vectors from atom A_1 to all other atoms are shown. Similar vectors can be drawn from each of the atom to rest of the atoms.

The equivalent points for these two atoms are:

	x	y	z
A_1	0.18	0.22	0.06
A_2	0.82	0.78	0.06
A_3	0.32	0.72	0.94
A_4	0.68	0.28	0.94

	x	y	z
B_1	0.74	0.22	0.44
B_2	0.26	0.78	0.44
B_3	0.76	0.72	0.56
B_4	0.24	0.26	0.56

Inter-atomic vectors between all these equivalent positions were calculated. Some of the vector peaks which lie on the Harker sections *u*=0.5, *v*=0.5 or *w*=0.0 could be located in the corresponding sections of Patterson maps (Fig. 2.16 A, B and C). Also, vectors occurring at *v*=0.0 and *w*=0.5 could be located (Fig. 2.16 D and E).

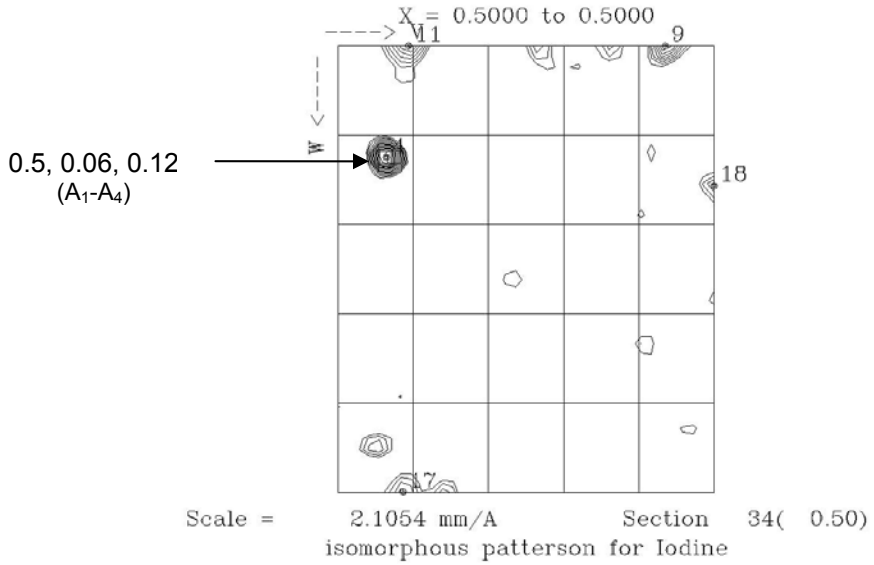


Fig. 2.16 A. Harker section at X = 0.5 of difference Patterson map of iodine derivative of CAL.

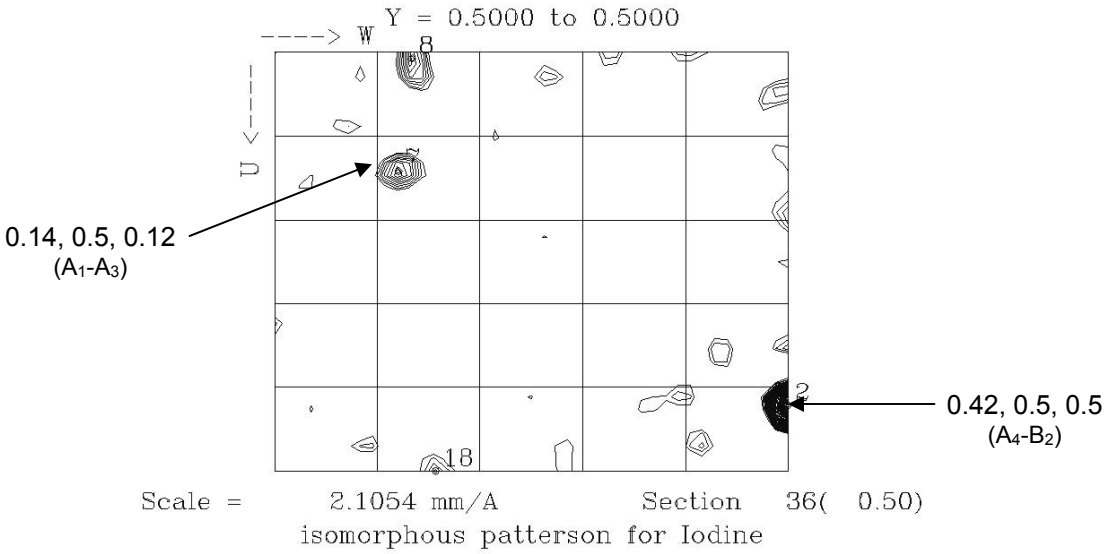


Fig. 2.16 B. Harker section at Y = 0.5 of difference Patterson map of iodine derivative of CAL.

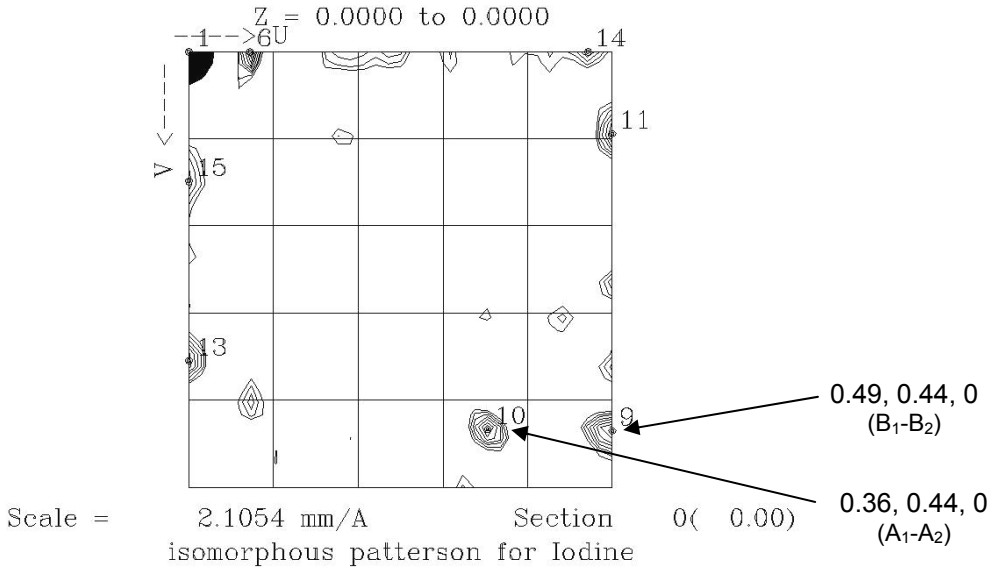


Fig. 2.16 C. Harker section at Z = 0.0 of difference Patterson map of iodine derivative of CAL.

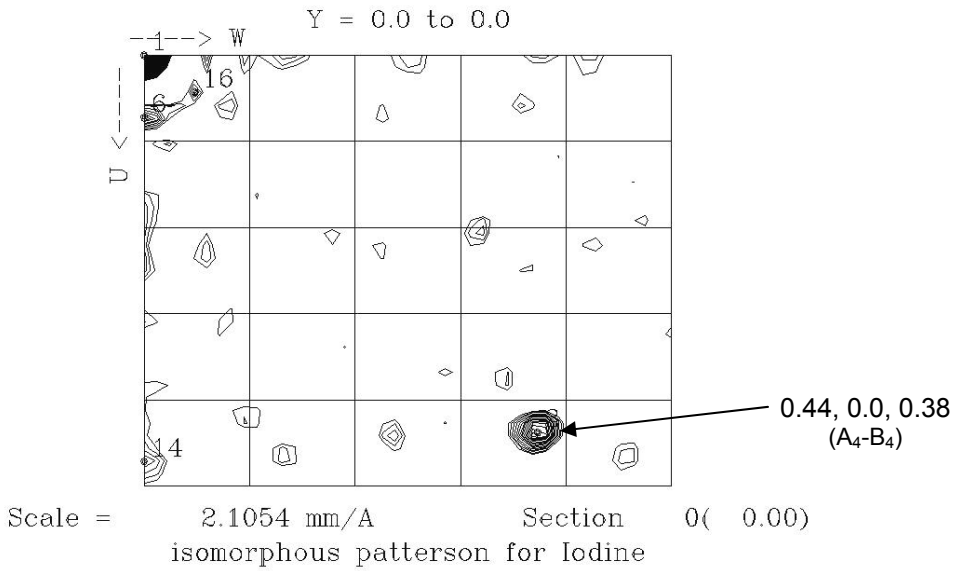


Fig. 2.16 D. Section at Y = 0.0 of difference Patterson map of iodine derivative of CAL.

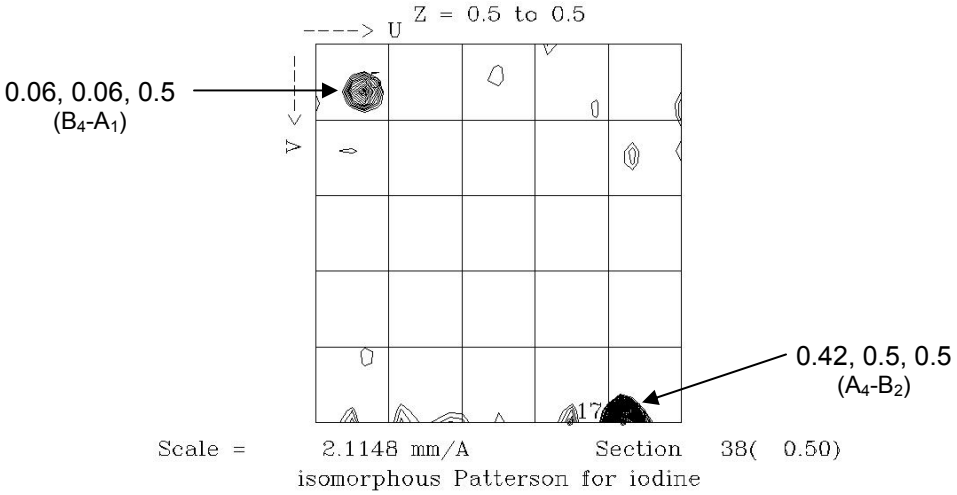


Fig. 2.16 E. Section at Z = 0.5 of difference Patterson map of iodine derivative of CAL.

2.5.6 Non-crystallographic symmetry in orthorhombic crystals of CAL

The symmetry present within the asymmetric unit of a crystal structure is termed as the non-crystallographic symmetry (NCS) (Glusker and Trueblood, 1985). Since proteins are often oligomeric molecules, NCS is commonly found in many protein crystals. Sometimes, the subunits in a protein are related by an NCS operator which is different than the crystallographic symmetry operators. NCS in proteins can be used as an aid in structure determination. Since the intensity of reflections results from many molecules in identical orientations diffracting identically, the diffraction pattern is the sum of diffraction patterns from all individual molecules. Thus, a large number of weak, noisy signals (each from a single molecule) are added to get a strong signal. In some cases, the strength of this signal can be increased further by averaging the signals from molecules related by NCS. It improves the signal-to-noise ratio of the data and helps to get a clearer image of the molecules (Rhodes, 2000). This method as such is not useful to determine the new structures directly; however it can be used to improve the poorly interpretable electron density and has found the major application in structure determination for the capsids of spherical viruses. It can also be used to cautiously extend the resolution limits (Blow, 2002).

To determine the NCS relation between the subunits of the same structure, self rotation function can be used. In this, a Patterson function is rotated upon itself over a volume *V* near the origin, for all possible rotations. Using the agreement of the Patterson with its rotated version, the rotational operators which relate the subunits to each other can be found out (Blow, 2002). Sometimes, when the NCS axis is parallel or nearly parallel to a crystallographic axis, self-rotation function does not reveal any new peaks. In such a case, the native Patterson map can be useful to get the peak associated with NCS (http://www.ruppweb.org/Xray/Patterson/Native_Patterson.htm). Alternatively, in

case of MIR, if two or more heavy atom sites are found out, they can be used to determine the NCS.

As orthorhombic (as well as trigonal) crystals of CAL have a dimer in the asymmetric unit, a study was carried out to find out whether the monomers in the dimer are related to each other by an NCS operator. From the two sites determined for the binding of iodine atoms to CAL, it was evident that their y-coordinates are same. Hence it may be concluded that the NCS axis is almost parallel to the crystallographic *b* axis. A self-rotation function calculated using the program MOLREP (Fig. 2.17) did not reveal any peaks other than those corresponding to the crystallographic symmetry operators, which was also indicative of an NCS axis parallel to a crystallographic axis.

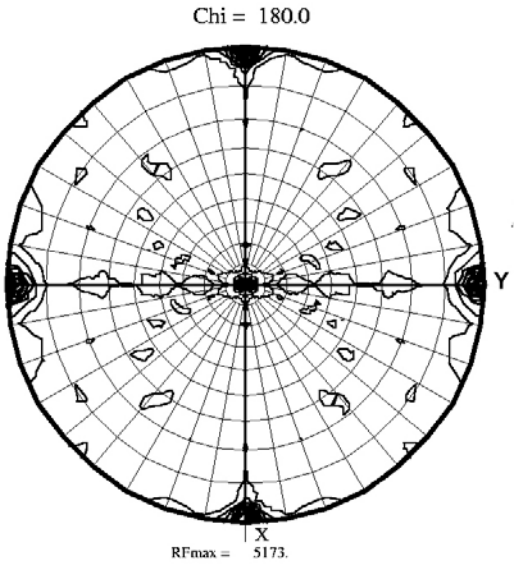


Fig.2.17. Self-rotation function for the native data of CAL crystal belonging to the space group $P2_12_12$ in the range 10.0-4.0 on the section $\chi=180^0$. The strong peaks on the horizontal and vertical axes appearing on the circumference are due to the space group symmetry ($P2_12_12$). Absence of any other strong peak indicates either absence of NCS or coincidence of NCS axis with one of the crystallographic axes.

If the NCS rotation coinsides with crystallographic symmetry axes, a corresponding translation can be found (Evans, 2001). To detect this translation, a self Patterson map was calculated using FFT. This map showed a very strong peak at (0.42, 0.5, 0.5) (Fig. 2.18). confirming NCS rotation axis coinciding with crystallographic axis.

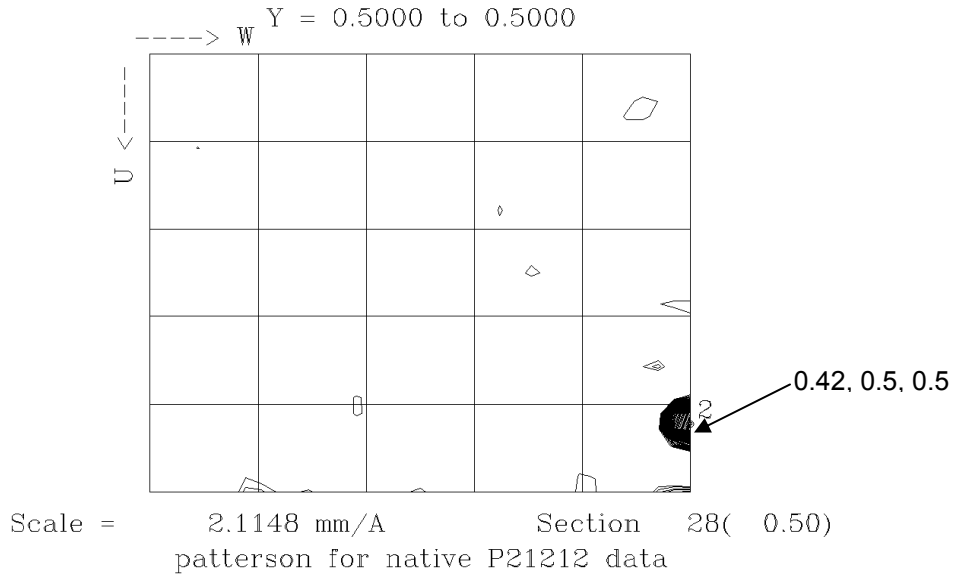


Fig. 2.18. Harker section at Y = 0.5 of Patterson map of a native dataset of CAL belonging to the space group $P2_12_12$ in the range 10-4.0.

It can be easily seen that this peak also occurs in the difference Patterson map calculated for the iodine derivative data. Since the occurrence of this peak is due to the NCS, the positions of the two subunits in the unit cell were determined as follows:

	x	y	z
A_1	1.18	0.22	0.06
A_2	0.82	0.78	0.06
A_3	0.32	0.72	-0.06
A_4	0.68	0.28	-0.06

	x	y	z
B_1	0.76	0.72	0.56
B_2	1.24	0.28	0.56
B_3	0.74	0.22	0.44
B_4	0.26	0.78	0.44

The vectors between all corresponding atoms ($A_1 \rightarrow B_1$, $A_2 \rightarrow B_2$ etc) give the peak (0.42, 0.5, 0.5) which appears very strongly in the Patterson map.

2.5.7 Determination of the three-dimensional structure of CAL

Crystallization of CAL in the initial stages yielded crystals of very good quality. The problem of twinning could also be overcome and native datasets up to a resolution of 2.3 Å could be collected. Initial crystals were obtained from the protein purified from the seeds of a particular cultivar of *Cicer arietinum*, namely BDN 9-3 (Katre *et al.*, 2005). Later on, due to the unavailability of seeds of this cultivar, the protein was being purified from the local varieties of chickpea. Also, initial purification protocol involved using of SP-sephadex as the cation exchange matrix. Due to the unavailability of this matrix, another anion exchange matrix (Q-sepharose) and hydrophobic interaction matrix (octyl sepharose) were used to purify the protein. Although this new batch of protein essentially showed the same biochemical characteristics as the previous one, as well as it gave the crystals with same unit cell dimensions and space group, the crystals were not of the same diffraction quality as the previous ones, no matter whatever efforts were input to improve the quality of the crystals. Often a large number of crystals were found to grow in a drop, which reduced the size of the crystals. The protein has a strong tendency to form large aggregates in a solution of low ionic strength and high protein concentration, which also might be responsible for the poorer crystal quality. Use of several detergents, namely octyl-β-D-glucopyranoside (n-octyl glucoside), Triton X-100 or Noidet at different concentrations did not help improving the quality of crystals. To prevent the aggregation of the protein due to possible formation of disulfide linkages, reducing agents (DTT and BME) were incorporated in the protein solution, even in the purification steps as well as crystallization conditions. However, this did not improve the crystal quality, on the contrary, resulted in precipitation of the protein in the crystallization drop. Several methods of seeding the crystallization drop with seeds of pre-formed crystals were tried. Formation of enormous number of crystal nuclei could not be

prevented by seeding. Molecular replacement trials with various chosen models were unsuccessful as discussed in the section 2.5.3.

Cryoprotection of CAL crystals with glycerol, PEG 200 or PEG 400 was often unsuccessful, which deteriorated the quality of crystals. Rarely the diffraction data could be collected at low temperatures. Due to this, data with high redundancy could never be collected. In case of iodine derivative crystal, anomalous signal could not be detected as the data did not have enough redundancy. The occupancy of the iodine atom in the crystal was found to be very low, ~0.3. To get the derivative crystals with higher occupancy of iodine, higher concentrations of iodine (5-10 mM) were tried. However, this resulted in formation of a lot of small crystals unsuitable for data collection.

The primary structure of the pea albumin (PA-2) showed presence of three cysteine and four methionine residues. Considering the sequence similarity between CAL and PA-2, the presence of seven sulfur atoms in the protein could have been used to solve the structure by in-house sulfur SAD phasing (Sarma and Karplus, 2006). In this method, the data can be collected at the Cu-*K*α wavelength (1.5418 Å); however, a highly redundant dataset preferably from a single crystal is necessary for this purpose, which could not be obtained for CAL crystals.

Owing to all these difficulties, we could not proceed with the three-dimensional structure determination of CAL. Efforts are being made to get the seeds of cultivar BDN 9-3 from which protein that gives better quality crystals could be purified. Cloning of CAL gene into a eukaryotic system is in progress, which can be used to prepare selenomethionine substituted protein. Possibility of collecting data at a synchrotron radiation source is simultaneously being explored. Since these efforts are bound to take more time, the final structure of CAL could not be reported in this thesis.

Chapter 3

Purification and biochemical characterization of a hemagglutinin from *Moringa oleifera* (drumstick)

3.1. Summary

A hemagglutinin from the mature seeds of *Moringa oleifera* (MoL) was purified to homogeneity by ammonium sulfate precipitation (90% saturation), followed by two successive ion exchange chromatography steps using DEAE cellulose and CM Sephadex C50. The yield of the protein was about 800 mg per 100 g of the seeds having a final specific activity of 160 $HUmg^{-1}$. The native protein is a homodimer of molecular mass ~14 kDa and subunit molecular weight ~7 kDa. MoL has a pI of 10 and is stable in the entire pH range 1-12, up to 24 hours at 25 ^{0}C as well as up to 30 minutes at 85 ^{0}C and is inhibited only by glycoproteins such as thyroglobulin, fetuin and holotransferin. The protein loses its hemagglutination activity in the presence of 0.25 M GdnHCl or more than 3 M concentration of urea. The *N*-terminal sequence of MoL determined is APGIMYRVQR. Only a single tryptophan per monomer is detected in the protein by titration with NBS, in both native and denatured conditions. Although no free cysteines could be detected, the three disulfide linkages detected per dimer indicated presence of 3 cysteine residues in each monomer. Chemical modification studies targeted individually at residues such as arginine, histidine, aspartic acid, glutamic acid, cysteine, tyrosine, tryptophan or serine showed no major role for them in the hemagglutination activity of MoL.

3.2. Introduction

Moringa oleifera (drumstick or horse radish tree) is a tropical plant belonging to the family *Moringaceae*, which is widely cultivated in India to use its leaves, fruits, and roots as food and for medicinal purposes. The *Moringa oleifera* seed flour is traditionally used as a coagulant in water treatment (Gassenschmidt *et al.*, 1995). Use of these seeds for softening the hard water has also been reported (Muyibi and Evison, 1995). The seeds contain up to 40% w/w of quality edible oil (with > 80% unsaturated fatty acid

content (Mohammed *et al.*, 2003)). Seeds and oil free pancakes yield proteins which are capable of acting as essential coagulants in water and waste water treatment (Bhuptawat *et al.*, 2007).

A cationic protein of molecular weight 6.5 kDa and having isoelectric point above 10.0, isolated from the seeds of this plant has been reported and its flocculating activity has been described by Gassenschmidt *et al.* (1995). It has been shown that the active agents in the aqueous extracts of *Moringa oleifera* seeds are dimeric cationic proteins, having molecular weight of 13 kDa and isoelectric points between 10 and 11 (Ndabigengesere *et al.*, 1995). Also, a water soluble acidic lectin having MW 20 kDa and an antioxidant component have been reported from the seeds of *Moringa oleifera* (Santos *et al.*, 2005).

In the present study, isolation and characterization of a lectin with complex sugar specificity from the mature seeds of *Moringa oleifera* (*Moringa oleifera* lectin; MoL) has been reported. Based on its molecular weight and pI, this lectin can be the same flocculating protein described by Gassenschmidt *et al.* (1995) and by Ghebremichael (2005) but different from the lectin described by Santos (2005).

Purification of MoL was undertaken to crystallize it and determine its three-dimensional structure by X-ray crystallography. However, despite several attempts, MoL could not be crystallized, an indication that the protein has some unusual structural properties. Further biochemical and biophysical characterization of the protein has been undertaken to get more information about these properties and to improve upon the crystallization strategy based on them.

3.3. Materials

Dry mature seeds of *Moringa oleifera* were purchased from local market. Ammonium sulfate, potassium dihydrogen monophosphate, dipotassium hydrogen

phosphate, sodium chloride, citric acid, phenol, sulfuric acid were purchased from Merck, India or SRL, India. DEAE cellulose, CM sephadex C50, Sephadex G-100, molecular weight markers for gel filtration, ampholytes (pH range 3.0-10.0), guanidium hydrochloride, urea, phenylglyoxal, diethyl pyrocarbonate, *N*-bromosuccinimide, Woodward's reagent K, 5,5'-dithiobis(2-nitrobenzoic acid), *N*-acetyl imidazole, sodium borohydride, glucose, galactose, mannose, methyl α-glucose, methyl α-galactose, methyl α-mannose, fucose, LacNAc, ManNAc, Gal-β-1,3-GalNAc were purchased from Sigma, USA. Low range molecular weight markers for SDS PAGE were procured from Bangalore Genei. The vertical gel electrophoresis apparatus and the power pack were from Tarsons.

3.4. Methods

3.4.1 Purification of *Moringa oleifera* lectin

All purification steps were performed at 4 ^{0}C unless otherwise stated. 25 g of mature seeds of *Moringa oleifera* were soaked overnight at 25 ^{0}C in deionised water and then crushed in phosphate buffered saline (PBS, containing 10 mM potassium phosphate buffer pH 7.2 and 150 mM NaCl). The suspension was kept stirred for 5 h, then filtered through muslin cloth and centrifuged at 10000 *g* for 20 min. The supernatant was filtered through Whatman filter paper no. 1 to remove the lipid layer. To this filtrate, $(NH_4)_2SO_4$ was added slowly to achieve a final saturation of 90% with constant stirring, while keeping the container in ice bath. The solution was allowed to stand for about 12 h for completing the precipitation. The precipitate was collected by centrifugation at 10000 *g* for 20 minutes, dissolved in 10 mM potassium phosphate buffer (pH 7.2) and dialyzed extensively against the same buffer. After centrifuging the dialysate at 8000 *g* for 5 minutes, the clear supernatant was loaded on a DEAE-cellulose column (10 X 2 cm) preequilibrated with the same buffer at a constant flow rate of 25 ml hr^{-1}. The column

was washed with the same buffer and washings were collected as fractions of 5 ml. The lectin elutes from the column unbound as tested by absorbance at 280 nm and hemagglutination test. Fractions having OD_{280} > 0.05 and showing hemagglutination activity were pooled together and dialyzed against 10 mM citrate-phosphate buffer (pH 6.0). The dialysate was centrifuged at 8000 *g* for 5 min and loaded on CM sephadex C50 column (10 X 2 cm) equilibrated with the same buffer at a constant flow rate of 12 ml hr^{-1}. The column was washed with the same buffer until the OD_{280} was <0.05, then washed with 0.5 M NaCl and finally with 1.0 M NaCl. MoL binds very tightly to the column and is eluted only with 1.0 M NaCl as tested by absorbance at 280 nm and hemagglutination test. The pure protein was dialyzed against deionized water and stored at -20 ^{0}C until further use.

3.4.2 Determination of protein concentration

Protein concentrations were determined according to the method of Lowry *et al.* (1951) using bovine serum albumin (BSA) as the standard.

3.4.3 Erythrocyte preparation

Human or rabbit RBCs were washed 5 times with 1X PBS (10 mM potassium phosphate buffer pH 7.2 containing 150 mM NaCl). A 3% (v/v) suspension of the erythrocytes in the same buffer was prepared.

3.4.4 Hemagglutination and hemagglutination inhibition assays

Hemagglutination tests were performed in standard microtitre plates by 2-fold serial dilution method. A 50 µl aliquot of the erythrocytes suspension was mixed with 50 µl of serially diluted lectin and agglutination was examined visually after incubation for 30 min at ~25 ^{0}C. The unit of hemagglutination activity (HU) was expressed as the

reciprocal of the highest dilution (titre) of the lectin that showed complete agglutination. The specific activity of the lectin is defined as the number of hemagglutination units per milligram of the protein (HU mg^{-1}).

Hemagglutination inhibition assays were performed similarly, except that the serial dilutions of sugar/glycoprotein solutions (25 µl) were pre-incubated for 30 min at ~25 ^{0}C with 25 µl of the lectin (~8 HU). RBC suspension (50 µl) was added, mixed and the plates read after 30 min.

3.4.5 Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) was performed at pH 8.8 as described by Laemmli (1970) to check the purity of the sample as well as to determine the subunit molecular mass of the protein. A 15% (w/v) gel of 1 mM thickness was prepared and the protein samples (5-10 µg) were electrophoresed at 25 ^{0}C by applying a voltage of 100 V and current 60-70 mA, until the gel tracking dye (bromophenol blue) reached till the end of the gel.

The isoelectric point of MoL was determined by tube gel electrophoresis (Vesterberg, 1977). The protein (~100 µg) was electrophoresed at 100 V for 16 h, in a 7.5% (w/v) polyacrylamide gel using wide range ampholytes (pH range 3.0-10.0).

In both the cases, protein bands were visualized by staining the gels with 0.25% Coomassie Brilliant Blue R-250 in 40% (v/v) methanol and 10% (v/v) glacial acetic acid, and destaining with a solution consisting of 40% (v/v) methanol and 10% (v/v) glacial acetic acid.

3.4.6 Carbohydrate content

The carbohydrate content of MoL was determined by phenol sulfuric acid method. The protein (200 µg in 200 µl water) was incubated with 200 µl of 5% (w/v)

phenol for 10 min at 25 ^{0}C. 1 ml of H_2SO_4 was added to it and the reaction mixture was allowed to cool for 20 min. The total carbohydrate was then estimated at 490 nm, using D-mannose as standard (DuBois *et al*., 1956).

3.4.7 Determination of molecular mass

The subunit molecular mass of the purified lectin was determined by SDS-PAGE under reducing as well as non-reducing conditions using low range molecular weight markers from Bangalore Genei, consisting of ovalbumin (43 kDa), carbonic anhydrase (29.9 kDa), soyabean trypsin inhibitor (20.1 kDa), lysozyme (14.3 kDa), aprotinin (6.5 kDa) and insulin (α and β chains) (3.5 kDa). The gel was photographed using the gel documentation system by Alpha-Innotech. The molecular weight of the protein was determined using the program provided by Alpha-Innotech, which calculates the relative mobilities (R_f values) of all the protein bands, plots a graph of log(MW) against R_f values and estimates the molecular weight of the unknown protein based on its R_f value.

The molecular mass of the native protein was determined by gel filtration chromatography. 3.0 mg of the purified protein was loaded on Sephadex G-100 column (60 X 1.5 cm) pre-equilibrated with 25 mM potassium phosphate buffer at pH 7.2 and was eluted with the same buffer at a flow rate of 12 ml hr^{-1}. The column was calibrated with Bovine Serum Albumin (66 kDa), carbonic anhydrase (29.9 kDa), lysozyme (14.3 kDa) and cytochrome C (12.4 kDa). A graph of V_E/V_0 against log(MW) of the protein markers was plotted; V_E being the elution volume of the protein and V_0 the void volume of the column and using estimated V_E/V_0 value of MoL, its native molecular weight was calculated.

3.4.8 Temperature and pH stability

Stability of MoL at various temperatures was studied by incubating the protein (0.2 mg ml^{-1}) at temperatures ranging from 30 to 100 ^{0}C and checking the residual activity by hemagglutination test after time intervals of 30 minutes.

Effect of pH on the stability of the protein was monitored by incubating the protein (0.2 mg ml^{-1}) in presence of 0.1 M buffers of pH range 1 to 12 for upto 24 hours at 25 ^{0}C. Buffers used were: pH 1.0, 2.0 and 3.0: glycine-HCl; pH 4.0 and 5.0: citrate-phosphate; pH 6.0 and 7.0: phosphate; pH 8.0 and 9.0: Tris-HCl; pH 10.0, 11.0 and 12.0: glycine-NaOH. The residual activity was checked by hemagglutination test after time intervals of 6, 12 and 24 hrs.

3.4.9 Chemical modification studies

In all the chemical modification studies, the residual activity of the modified protein was determined by hemagglutination assays. Lectin samples incubated in the absence of modifying reagent served as a positive control. A negative control was the modifying reagent in the absence of the lectin, in the same concentration as used for the modification reaction.

A. Modification of Arginine residues with phenylglyoxal

The reagent was prepared in methanol. The lectin (200 µg), in 50 mM phosphate buffer, pH 8.0, was treated with varying concentrations of phenylglyoxal (0.5-3.0 mM) for 30 min at 25 °C. Excess reagent was then removed by dialysis, and the residual hemagglutination activity determined (Takahashi, 1968). The methanol concentration in the reaction mixture did not exceed 2% (v/v) and had no effect on the activity and stability of the lectin during the incubation period.

B. Modification of Histidine residues with diethyl pyrocarbonate (DEPC)

The reagent was prepared in absolute ethanol just prior to use. The lectin (200 µg) in 50 mM phosphate buffer, pH 7.2 was treated with 1-10 mM DEPC for 30 min. Excess reagent was then removed by dialysis, and the residual hemagglutination activity determined (Ovadi *et al.*, 1967). The concentration of ethanol in the reaction mixture did not exceed 2% (v/v) and had no effect on the activity and stability of the lectin during the incubation period.

C. Modification and estimation of tryptophan residues with *N*-Bromosuccinimide (NBS)

1 ml of (~400 µg) MoL in 10 mM CPB pH 5.0 was titrated against 5 mM *N*-Bromosuccinimide. The reaction was monitored using the reduction in OD at 280 nm. The number of tryptophan residues modified was determined by assuming a molar absorption coefficient of 5500 M^{-1} cm^{-1} (Spande and Witkop, 1967). The residual activity was determined by hemagglutination assay. To determine the number of tryptophans in unfolded state, the reaction was carried out for the protein incubated in the presence of 6 M Urea for 16 hours.

D. Modification of carboxylate residues with Woodward's reagent K (WRK)

The protein (200 µg) was incubated in 50 mM CPB pH 6.0, with different concentrations (5-20 mM) of WRK. Aliquots were removed after every 15 min. The reaction mixture was dialyzed to remove excess reagent and the residual activity determined (Sinha and Brewer, 1985).

E. Modification of cysteine residues with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB)

Determination of free cysteine (s):

The reagent was prepared in 0.1 M PB pH 7.2. To 100 µg of protein at pH 7.2, 0.2 mM DTNB was added and incubated at 25 ^{0}C for 30 minutes. The reaction was

monitored by checking absorbance at 412 nm at different time intervals as well as by hemagglutination test. The number of sulfhydryl groups modified was calculated using the following formula and assuming a molar absorption coefficient of 13,600 $M^{-1}cm^{-1}$ (Habeeb, 1972).

$$\text{No. of cysteines} = \frac{\text{MW X Change in } OD_{412} \text{ X final volume (ml)}}{\text{13600 X Amount of protein in mg}}$$

Determination of disulphide bonds

To the lectin (50-150 µg), 0.72 g solid urea, 50 µl 0.1 M Na-EDTA and 500 µl of freshly prepared 2.5% sodium borohydride were added. The volume was adjusted to 1.5 ml with water. The reaction mixture was incubated at 37 ^{0}C for 45 min. Then 250 µl of 1 M KH_2PO_4 containing 0.2 M HCl was added. Nitrogen was passed in the solution for 5 min and 1 ml acetone was added after which nitrogen was passed for another 5 minutes. Then 250 µl of 10 mM DTNB was added and the total volume was adjusted to 3 ml with water. Nitrogen was bubbled for 2 minutes and reaction mixture was kept at 25 ^{0}C for 15 minutes before determining the number of sulfhydryl groups modified as mentioned in the section on determination of free cysteine(s) (Cavallini *et al.*, 1966).

F. Modification of tyrosine with *N*-acetyl imidazole (NAI)

To 1000 µg protein at pH 7.2, *N*-acetyl imidazole (10 mM) was added serially and the reaction was monitored by change in absorbance at 278 nm as well as by hemagglutination assay after each addition. The tyrosine residues modified were determined spectrophotometrically, using a molar absorption coefficient of 1160 $M^{-1}cm^{-1}$ at 278 nm (Riordan *et al.*, 1965).

G. Modification of serine with Phenylmethylsulfonyl fluoride (PMSF)

The lectin (100 µg) in 50 mM Tris-HCl buffer, pH 8.0 was incubated with 5 mM PMSF, at 27±1 ^{0}C, for 60 min. Aliquots were removed at 15 min intervals, the excess reagent removed by dialysis and residual activity determined (Gold and Fahrney, 1964).

H. Modification of lysine

With Trinitrobenzenesulphonic acid (TNBS):

The reagent was dissolved in 0.1 M Tris buffer, pH 8.0 to make a stock solution of 0.1 M. 200 µg of protein in 0.1 M tris buffer, pH 8.0 was allowed to react for 1 hr with 10 mM TNBS. The excess reagent was removed by dialysis and the residual activity was determined (Habeeb, 1966).

With acetic anhydride and citraconic anhydride:

Both these reagents were diluted in acetone to get a concentration of 100 mM. 200 µg of MoL in 0.1 M tris buffer, pH 8.0 was incubated with varying concentrations (1-10 mM) of both these reagents for 1 hr (Fraenkel-Conrat, 1957). Aliquots were removed after every 15 minutes and the residual activity was determined.

3.4.10 Amino acid analysis and *N*-terminal sequence determination

Salt-free lyophilized MoL sample (200 µg) was hydrolyzed in 6N HCl in vacuum for 20h at 110°C. Following hydrolysis, the sample was again lyophilized and dissolved in 100 µl of 20 mM HCl. 5 µl of this sample was mixed with 60 µl buffer and 20 µl flur reagent provided in the AccQ-Tag kit by WATERS. Total volume was made to 100 µl. 5 µl of this sample was loaded on the column. Amino acids were detected as fluorescent derivatives of 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC).

N-terminal sequencing of MoL was carried out using the automated protein sequencer at National Institute of Immunology, New Delhi, India.

3.4.11 Crystallization trials on purified MoL

Purified MoL was dialyzed against deionised water and concentrated upto 60-100 mg ml^{-1}. Crystallization trials were conducted using hanging-drop vapor-diffusion method as described in Chapter 2. Several commercially available crystallization kits as well as some procedures designed in our lab, enlisted below, were used to screen the suitable crystallization condition(s) for MoL.

1. Crystal Screen I and II Hampton Research.
2. Clear Strategy Screen (CSS I and II) from Molecular Dimensions Limited, at pH 6.5 and 8.5.
3. 30-80% saturated $(NH_4)_2SO_4$ in the pH range 4.0 to 9.0.
4. 5-30% polyethylene glycol 8000 in the pH range 4.0 to 9.0.
5. 20% polyethylene glycols 2000, 4000, 6000, 8000, 10000 and 20000 at pH 6.5 and 8.5.

3.5. Results and discussion

3.5.1. Characterization of MoL

MoL was purified using two successive ion exchange chromatography steps. The highly basic lectin eluted as an unbound protein from the anion exchanger DEAE-cellulose column, whereas bound very tightly to the cation exchanger CM Sephadex column and could be eluted only with 1M NaCl. It was purified approximately 2-fold with final recovery of 33% (Table 3.1). The final yield of the purified protein was 800 mg per 100 g of dried seeds with a specific agglutination activity of 160 $HUmg^{-1}$ for the purified protein.

MoL showed a single band on SDS-PAGE in the presence of 2-mercaptoethanol (Fig. 3.1 A) indicating its homogeneity. Isoelectric focusing (Fig. 3.1. B) showed that the isoelectric point of the protein is 10.0. It is a glycoprotein containing 1.5% neutral sugar.

Table 3.1: Purification of *Moringa oleifera* lectin

Purification Step[a]	**Total activity**[b] **(HU)**	**Total protein (mg)**	**Specific activity**[c] **($HUmg^{-1}$)**	**Fold purification**	**Recovery (%)**
Crude Extract	384000	4800	80	1	100
0–90% $(NH_4)_2SO_4$ precipitation	230400	2700	85.33	1.07	60
DEAE cellulose unadsorbed fraction	153600	1560	98.46	1.23	40
CM sephadex adsorbed, eluted with 1M NaCl	128000	800	160	2	33

[a] Starting with 100 g dry seeds.

[b] The reciprocal of the highest dilution (titre) of the lectin that showed complete agglutination was expressed as a unit of hemagglutinating activity.

[c] The specific activity of the lectin is defined as units of the hemagglutinating activity per milligram of lectin.

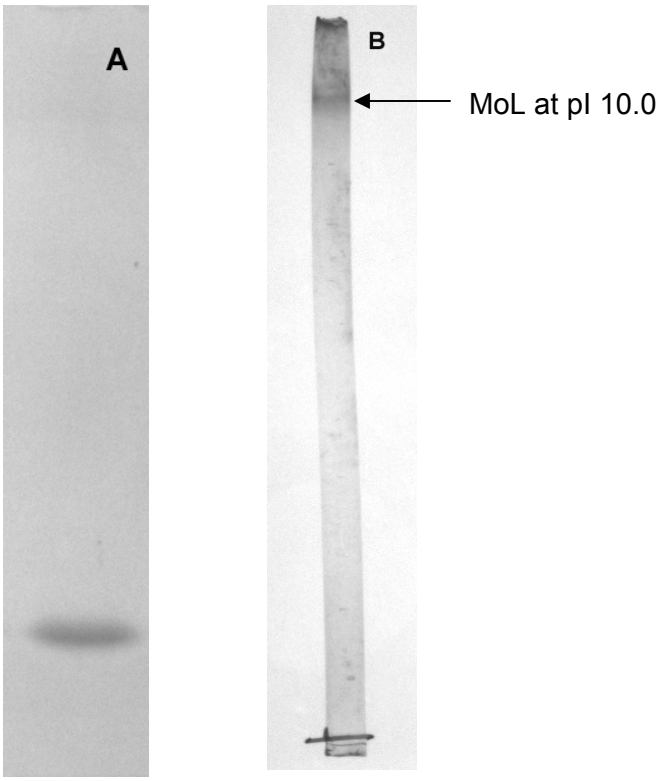


Fig. 3.1. Electrophoresis of purified lectin.
A. SDS-PAGE of MoL at pH 8.8 in presence of 2-mercaptoethanol.
B. IEF-PAGE of MoL. 100 µg of purified lectin was loaded and carrier ampholytes of the range 3-10 were used.

The molecular mass of the protein determined by gel filtration chromatography using sephadex G-100 was 14 kDa (Fig. 3.2). The molecular mass of the protein determined by SDS-PAGE in presence of 2-mercaptoethanol is 7.1 kDa, indicating that the native protein exists in solution as a homodimer linked by disulfide linkage(s). However, in the absence of 2-mercaptoethanol two bands corresponding to 13.6 kDa and 27.1 kDa were observed (Fig. 3.3) which indicated that a tetrameric species was getting formed due to aggregation in the presence of SDS. The dimer seems to be linked by intersubunit disulfide linkage(s). Presence of disulfide linkages in the dimer is in agreement with the observation of Ndabigengesere *et al.* (1995).

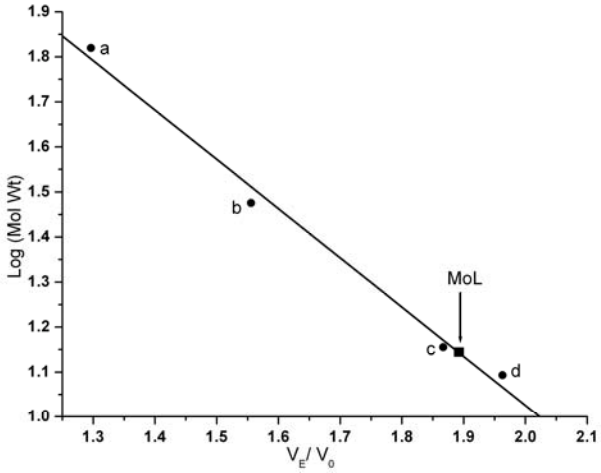


Figure 3.2. Molecular mass determination of MoL by gel filtration chromatography on Sephadex G-100 column. (a) Bovine Serum Albumin (66 kDa), (b) carbonic anhydrase (29.9 kDa), (c) lysozyme (14.3 kDa) and (d) cytochrome C (12.4 kDa). ■ represents MoL (14 kDa).

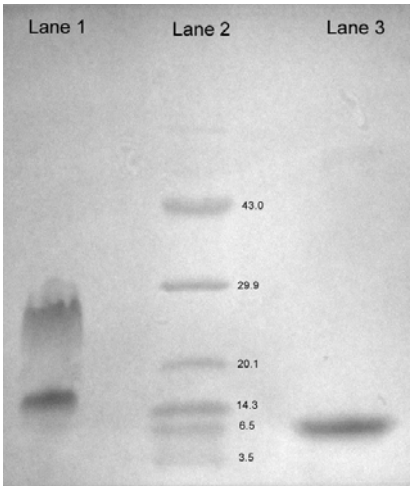


Fig. 3.3. Subunit molecular mass determination of MoL by SDS-PAGE. Lane 1: purified MoL without 2-mercaptoethanol. Lane 2: low range molecular weight markers from Bangalore Genei. The molecular weights of the markers in kDa are shown beside the bands. Lane 3: purified MoL with 2-mercaptoethanol. The gel was stained with Coomassie Brilliant Blue R-250.

3.5.2. Temperature and pH stability

MoL is a highly stable protein, which retains the hemagglutination activity in the pH range of 1-12, up to 24 hours at 25 ^{0}C. The protein shows hemagglutination activity even after incubating at 85 ^{0}C up to 30 minutes at pH 7.2 (Fig. 3.4). The activity of MoL heated at 100 ^{0}C for 30 min could be partly restored after incubating the protein at 25 ^{0}C for 1 h.

Although quite a few plant lectins are known to retain their hemagglutination activity after heating upto 55-60 ^{0}C, like *Arundo donax* rhizome lectin (Kaur *et al.*, 2005b), *Helianthus tuberosus* tuber lectin (Suseelan *et al.*, 2002), *Saraca indica* seed integument lectin (Ray and Chatterjee, 1995), not many are stable up to such high temperature as 85 ^{0}C. The hemagglutinin with comparable thermostability is from *Trichosanthes dioica* (Dharkar *et al.*, 2006), which is considered to be a type II RIP.

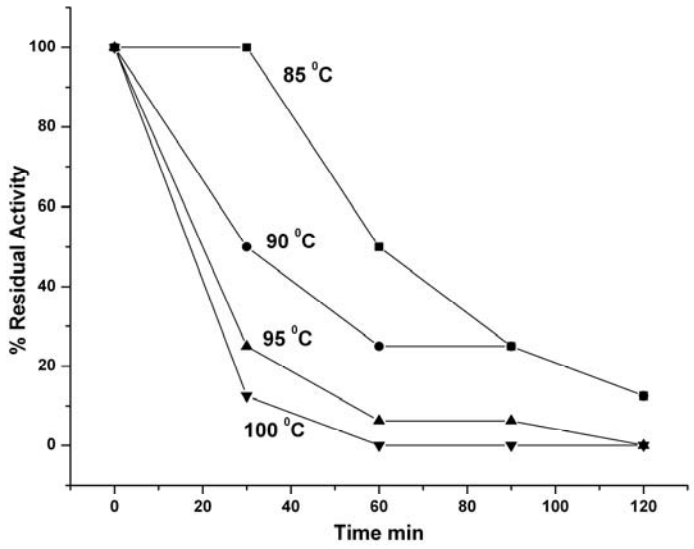


Figure 3.4. Thermal stability of MoL. 0.4 mg ml^{-1} protein was heated at various temperatures for different time intervals and the hemagglutinin activity was checked.

3.5.3. Carbohydrate binding specificity

MoL agglutinates human as well as rabbit erythrocytes; the affinity for the later is almost 250 times more than that for the former. This difference in the agglutination may be due to the nature of the glycoproteins protruding on the cell surface of both these cells. The agglutination is not blood group specific, in case of human blood. Treatment of erythrocytes with pronase did not increase the hemagglutination titre. Simple sugars like glucose, galactose, mannose, fucose, fructose, sucrose, *N*-acetyl-lactosamine, *N*-acetyl-mannosamine and Gal-β-1,3-GalNAc failed to inhibit the hemagglutination activity of MoL, whereas 0.04 mg/ml of fetuin and thyroglobulin or 0.08 mg/ml of holotransferin were found to inhibit the hemagglutination activity, indicating the requirement of clustering effect of multiple Gal-β-1,3-GalNAc or LacNAc chains for binding. Fibrinogen, a glycoprotein with bianternnary oligosaccharide chains did not inhibit the hemagglutination activity. Similarly, Glucuronic acid and Galacturonic acid failed to inhibit the hemagglutination activity, which rules out the charge based hemagglutination by MoL.

3.5.4. Chemical modification studies

Treatment of MoL with group specific reagents did not have any effect on the hemagglutinating activity of MoL, suggesting that none of the residues such as arginine, histidine, aspartic or glutamic acid, cysteine, tyrosine, tryptophan or serine are individually important for the hemagglutination activity. The saccharide-binding site could be shallow and broad involving a number of amino acids.

3.5.5. Estimation of tryptophan and tyrosine residues

The titration of MoL with NBS indicated that only a single tryptophan residue per monomer in the protein gets modified both in the native and denatured conditions. In the

native condition, modification of the tryptophan residue did not affect the hemagglutinating activity of the protein.

A single tyrosine residue per monomer of MoL was found to be getting modified, in both native and denatured conditions, when MoL was treated with NAI, without any effect on the activity.

3.5.6. Determination of cysteines and disulfide linkages

Free cysteine residues were not detected in the native or denatured protein. After reduction of the protein, six cysteine residues per dimer were detected by modification with DTNB indicating the presence of three disulfide linkages. The thermostability of the protein could be due to these disulfide bonds. Since the number of intra- and inter-subunit disulfide bonds have not been determined separately, it is possible that all the three disulfide bonds in the dimer are inter-subunit, or one of them is inter-subunit and the other two are intra-subunit.

3.5.7. Amino acid analysis and *N*-terminal sequence determination

The amino acid composition of MoL showed that it contained high amounts of arginine and glutamic acid/glutamine (Glx) residues (Table 3.2). The presence of 11 residues of arginine might be conferring the predominant positive charge on the protein. Lysine is completely absent in the protein.

N-terminal sequence of MoL is APGIMYRVQR. A BLAST search conducted using this sequence against the non-redundant database of NCBI did not yield any significantly matching sequence.

Table 3.2. Amino acid composition of MoL

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

Determined spectrophotometrically according to:

[a] Cavallini *et al.* (1966); [b] Spande and Witkop (1967).

3.5.8. Crystallization of MoL

Characterization of MoL was undertaken to study its structure-function relationship. To elucidate the three-dimensional structure of MoL, X-ray crystallographic studies were undertaken.

MoL exhibits very high solubility in water even in the absence of any buffer or ions. It could be concentrated up to ~100 mg ml^{-1} in deionised water without precipitation. When crystallization trials were conducted, crystals were observed only in condition numbers 1 and 24 of Crystal Screen from Hampton Research. Both these

conditions contain calcium chloride (0.02 M and 0.2 M, respectively), 0.1 M sodium acetate trihydrate buffer at pH 4.6 and organic solvents as precipitants (30% 2-Methyl-2,4-pentanediol (MPD) and 20% isopropanol, respectively). When several crystallization experiments were set by varying pH and concentrations of precipitants, it was observed that long needle shaped crystals used to appear within minutes in the crystallization drop after mixing the protein with the precipitant solution (Fig. 3.5). Unfortunately, these crystals turned out to be salt crystals when tested in the X-ray beam. Apparently, calcium chloride was being driven out of the solution by high concentration of the protein, as no crystals used to appear in the absence of the protein or calcium chloride.

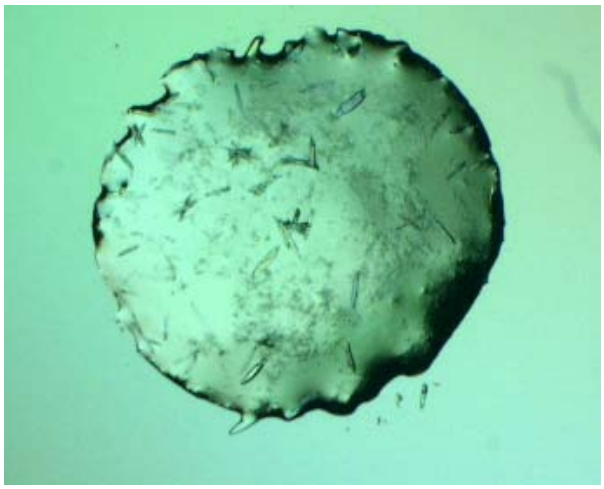


Fig. 3.5. Long needle shaped crystals obtained in condition number 24 (0.1M sodium acetate buffer pH 4.6, 0.2 M calcium chloride and 20 % v/v isopropanol) of Crystal Screen of Hampton Research. The concentration of MoL was ~80 mg ml^{-1}. The crystals were of salt as tested by X-ray diffraction.

Presence of high amounts of charged residues like glutamate, aspartate, arginine and lysine is known to confer high water solubility to the protein as well as render it recalcitrant to crystallization (Li and Lawson, 1995). Although MoL doesn't contain any lysine residue, presence of 11 arginines and 15 glutamate/glutamine residues might be

making it highly water soluble and difficult to crystallize. In such a case, surface engineering of the protein by mutating some of these residues possessing long side chains to shorter ones like alanine could help to reduce solubility of the protein and improve crystallizability (Derewenda, 2004; Makabe *et al.*, 2006). These surface mutations are expected not to significantly alter the protein conformation.

Another possible reason for difficulty in crystallization could be the presence of highly flexible or even completely unfolded fragments in MoL, that interfere with crystallization. Although this is observed commonly in case of large multidomain signaling proteins, MoL also can contain such unordered regions which make it difficult to crystallize. If such regions are present as the ligand binding sites, they can gain an ordered structure after binding to the ligand. Some proteins like *Artocarpus hirsuta* lectin (Rao *et al.*, 2004) are known to crystallize only in the presence of the ligand. Search for a suitable ligand for MoL and attempt of crystallization in the presence of ligand are necessary before ruling out this possibility.

3.6. Conclusions

The hemagglutinin property of a protein occurring in the *Moringa oleifera* seeds has been characterized. The lectin exists as a dimer in the solution, formed from monomers linked by disulfide bond(s). It is thermostable, retains activity in the entire pH range, highly basic and has complex sugar binding specificity. The disulfide linkages seem to hold the active conformation of the protein. Ionic interactions are also important for the hemagglutinating activity of MoL. The protein contains high amounts of arginine and glutamic acid/glutamine (Glx) residues. No particular amino acid residue was found to be important for the hemagglutination activity. The protein could not be crystallized despite several attempts, owing mainly to its high solubility and presence of high amounts of charged residues, possibly on the surface.

Chapter 4

Structural studies on hemagglutinin from *Moringa oleifera* using fluorimetric methods and circular dichroism (CD)

4.1. Summary

The intrinsic fluorescence of the protein was quenched upon titrating with sugars LacNAc (K_a = 1380M^{-1}) and fructose (K_a = 878M^{-1}), although these sugars did not inhibit the hemagglutination activity of MoL. To study the environment of the single tryptophan residue in MoL, solute quenching studies were carried out at pH 1.0, 7.2 and 10.0 and in the presence of 6 M urea using quenchers such as acrylamide, potassium iodide and cesium chloride. The results indicated that tryptophan is highly exposed to the solvent and present in a strongly electropositive environment. The study of the binding of hydrophobic dye 8-anilino-1-naphthalene sulfonic acid (ANS) using fluorescence spectroscopy showed exposed hydrophobic patches in protein that get further exposed at extreme acidic or alkaline pH but get buried in the interior in the presence of 1M GdnHCl or urea. MoL also binds adenine (K_a = 7.76 x 10^3 M^{-1}). Time resolved fluorescence of the native protein showed two lifetimes indicating two different conformers of tryptophan exist which get merged into a single one after quenching with 0.15M acrylamide.

Analysis of the far-UV CD spectrum of MoL by CDPro showed secondary structure composition of *α*R (regular *α*-helix) 16%, *α*D (distorted *α*-helix) 12%, *β*R (regular *β*-sheet) 14%, *β*D (distorted *β*-sheet) 9%, turn 20% and unordered 28%. The secondary structure was not affected by extreme acidic or alkaline conditions; however, it was drastically affected by the presence of a reducing agent like dithiothreitol (DTT) (1mM) at and above pH 7.0. The far UV CD spectra of the protein incubated in the presence of different concentrations of GdnHCl showed no change in the ellipticity in the region 215 to 225 nm, although it was affected at 210 nm and below indicating an increase in the unordered structural element. The secondary structure is stable up to 3M urea concentration above which it starts changing.

4.2. Introduction

Three-dimensional structure of a protein, which is responsible for its function, is held together by several factors like hydrogen bonds, ionic, hydrophobic and van der Waal's interactions as well as covalent interactions like disulfide linkages. The stability of the protein is dependant on these interactions. The conditions which interfere with these interactions lead to loss of the three-dimensional structure as well as activity of the protein. Increase in temperature and changes in pH, denaturation by chemical agents like GdnHCl and urea, reduction of disulfide linkages by DTT are some of such conditions. Changes in the native conformation of the protein in such conditions can be followed by using fluorescence and circular dichroism spectroscopic techniques. Fluorescence spectroscopy can also be used to quantitate the binding of ligands to a protein.

Moringa oleifera lectin (MoL) is a low molecular weight, dimeric protein with a highly basic pI of 10.0 (section 3.5.1). It is stable at comparatively high temperatures as well as in the entire pH range (section 3.5.2), and is inhibited only by glycoproteins (section 3.5.3). The studies on the conformational stability of this protein in various conditions by making use of fluorescence spectroscopy and circular dichroism have been described in this chapter. Saccharide binding as well as the accessibility of tryptophan residue of the lectin has also been studied.

4.3. Materials

Acrylamide, potassium iodide, cesium chloride, 8-anilino-1-naphthalene sulfonic acid (ANS), adenine, guanidium hydrochloride, urea, glucose, galactose, mannose, methyl α-glucose, methyl α-galactose, methyl α-mannose, fucose, LacNAc, ManNAc, Gal-β-1,3-GalNAc were purchased from Sigma, USA. All other chemicals used were of analytical grade.

Quencher solutions (acrylamide, potassium iodide, and cesium chloride) were prepared as 5 M stocks in deionised water. Potassium iodide solution was made in 200 µM sodium thiosulfate solution to prevent the formation of triiodide (I^{-3}) ions. Sugar solutions were prepared as 100 mM stocks. Buffers were prepared as 1 M stocks and pH was adjusted at 25 ^{0}C. All these solutions were filtered through 0.45 µ filter from Sartorius. Protein solutions were centrifuged at 8000 *g* for 5 min before use.

The software Microcal Origin 6.1 was used to plot the graphs and do the related calculations.

4.4. Methods

4.4.1. Purification of MoL

MoL was purified as described in section 3.4.1. The concentration of the protein was determined according to the method of Lowry *et al.* (1951) using BSA as standard.

4.4.2. Fluorimetric measurements

The aromatic amino acid residues in the protein, namely tryptophan, tyrosine and phenylalanine, get excited on absorption of light energy (photons). Since the residue is unstable in the excited state, it returns to its ground state which is accompanied by the dissipation of excess energy through fluorescence. Tryptophan has much stronger fluorescence than tyrosine and phenylalanine. The intensity and the wavelength of maximum fluorescence emission of tryptophan is highly dependant on the polarity of the environment surrounding it. Hence by studying the tryptophan fluorescence, the conformational changes in the protein can be monitored.

The intrinsic fluorescence of the protein was measured using a PerkinElmer Life Sciences LS50 fluorescence spectrophotometer. The samples were kept in a quartz cuvette, at a constant temperature (± 0.1 ^{0}C) using a Julabo F20 circulating water bath.

To eliminate the background emission, the signal produced by either buffer solution, or buffer containing the appropriate quantity of denaturants was subtracted. The protein solution (~0.025 mg ml^{-1}) was excited at 295 nm and the emission was recorded in the range of wavelengths 300-400 nm. Each spectrum was an average of 5 accumulations. Both the excitation and emission spectra were obtained setting the slit-width at 5 nm, and speed 100 nm min^{-1}.

4.4.3. Sugar binding studies

The binding of sugars to MoL was studied by titrating the protein against sugar solutions and monitoring the fluorescence. The temperature was maintained at 25 ^{0}C throughout the experiment using the waterbath. To 2 ml of lectin sample (concentration 0.04 mg ml^{-1}) in 50 mM phosphate buffer, pH 7.2, 3-10 µl aliquots of the sugar solutions (of concentration 10-100 mM) were added. The samples were excited at 295 nm and the fluorescence emission intensity at 351 nm (λ_{max} of the protein) was monitored. Corrections were also made to compensate the dilution effect upon addition of sugar to lectin. At the highest concentration of the saccharide, the volume change was less than 5 % of the solution in the cuvette. Each data point was an average of three independent sets of experiments with standard deviation (SD) less than 5%.

The following equation was used to determine the association constant (K_a) (Chipman *et al.*, 1967).

$$\log [C]_f = -\log[K_a] + \log [(F_0-F_C)/(F_C-F_\infty)] \qquad \text{(Eq. 4.1)}$$

From the ordinate intercept of the double reciprocal plot of $F_0/(F_0-F_C)$ versus $1/[C]$, where F_0 and F_C are the fluorescence intensities of the free protein and of the protein at a sugar concentration $[C]$, F_∞, the fluorescence intensity upon saturation of all the sugar binding sites is obtained. In the plot of $\log[(F_0-F_C)/(F_C-F_\infty)]$ versus $\log[C]$, the

abscissa intercept yielded the K_d value (the dissociation constant) for the protein-sugar interactions, which is the reciprocal of K_a (the association constant).

The free energy changes of the association (ΔG) were calculated by using the equation:

$$-\Delta G = RT \ln(K_a) \qquad \text{(Eq. 4.2)}$$

4.4.4. Solute quenching studies

Interaction between a fluorophore and a molecule induces perturbation or modification in the fluorescence parameters like intensity, quantum yield and lifetime. Two types of fluorescence quenching can be observed when the interaction takes place between the fluorophore and the quencher molecule. Collisional quenching occurs when the fluorophore and another molecule diffuse in the solution and collide with each other. In this case, the two molecules do not form a complex. In static quenching, on the other hand, two molecules bind one to the other forming a complex (Albani, 2004).

The most commonly used quenching molecules are acrylamide, succinimide (neutral), iodide (cationic) and cesium (anionic). Acrylamide, being a small uncharged molecule, can diffuse within a protein and can quench the fluorescence of buried tryptophans as well. Iodide and cesium ions, on the other hand, quench the fluorescence of tryptophans present at or near the surface of the protein. Being charged ions, their quenching efficiency also depends on the charge surrounding the tryptophan (Albani, 2004).

Fluorescence quenching experiments on MoL were carried out at 27 ^{0}C. Protein samples (0.05 mg ml^{-1}) incubated for 16 h in 50 mM buffers (pH 1.0, 7.2 and 10.0) as well as in 6M urea were titrated with small aliquots (5-10 µl) of 5M quencher solution (acrylamide, potassium iodide or cesium chloride). The samples were excited at 295 nm

and the emission spectra were recorded in the range of 300-400 nm. Each spectrum was an average of 5 accumulations. Quenching of the fluorescence was monitored at 351 nm (emission maximum of MoL) and corrections were made to compensate the dilution effect upon addition of quencher solutions.

Quenching data for all these quenchers were analyzed by the Stern-Volmer (4.3) equation (Lehrer, 1971):

$$F_0/F_C = 1 + K_{sv}[Q] \qquad \text{(Eq. 4.3)}$$

Where F_0 and F_C are the relative fluorescence intensities, corrected for dilution, in the absence and presence of quencher respectively, [*Q*] is the resultant concentration of the quencher and K_{sv} is the Stern-Volmer constant for the given quencher. Slopes of Stern-Volmer plots yield K_{sv} values.

4.4.5. Lifetime measurement of fluorescence decay

Lifetime measurements were carried out on an FLS920 spectrometer supplied by Edinburgh Instruments. A xenon flash lamp of pulse width 1 ns was used for excitation and a single photon counting photomultiplier was used for detection of fluorescence. The diluted Ludox solution was used for measuring Instrument Response Function (IRF). The samples (1 mg ml^{-1}) were excited at 295 nm and emission was recorded at 343 nm. Slit widths of 15 nm each were used on the excitation and emission monochromators. The resultant decay curves were analyzed by a multiexponential iterative fitting program provided by Edinburgh Instruments.

4.4.6. Binding of ANS and adenine to MoL

8-anilino-1-naphthalene sulfonic acid (ANS) is a fluorescent molecule which can bind to the exposed hydrophobic patches on the protein surface. ANS alone shows a weak fluorescence at 515 nm, when excited at 375 nm. When this dye binds to the

exposed hydrophobic patches of the protein, its fluorescence intensity increases considerably with a blue shift in the fluorescence maximum to 495 nm. This property can be used to estimate the exposed hydrophobic patches on the protein under various conditions (Daniel and Weber, 1966; Cardamon and Puri, 1992). Since the hydrophobic patches are usually buried in the interior of the protein and do not get exposed until the protein is unfolded by thermal/pH denaturation, ANS binding studies provide valuable information about the unfolding behavior of the protein, as well as about the intermediates formed during unfolding and refolding of the protein.

The binding of ANS to MoL was analyzed by measuring the fluorescence of MoL incubated in various conditions of temperature, pH and denaturants on the fluorescence spectrophotometer. 15 µl of 20 mM ANS was mixed with 2 ml of protein (0.05 mg ml^{-1}) which was then excited at 375 nm and the emission recorded between 450-550 nm. Reference spectrum of the buffer alone with same amount of ANS added in each of the condition was subtracted from the spectrum of the sample.

The binding of adenine to MoL was studied by intrinsic fluorescence titrations, in a way similar to that used to study the binding of sugars (section 4.4.3).

4.4.7. Circular Dichroism analysis

Circular dichroism spectroscopy measures the differences in the absorption of left- and right-handed polarized light by molecules which arise due to structural asymmetry in them. This property is exhibited by all optically active molecules like sugars and amino acids. Secondary structure elements in proteins and nucleic acids (α helix, β sheet, double helix etc.) also give rise to characteristic CD spectra. Using CD spectra of proteins in the far-ultraviolet region (180-250 nm), the secondary structure elements present in the protein can be estimated. At these wavelengths, peptide bonds in the protein act as chromophores and the signal produced by them is characteristic of

the secondary structure. Far UV CD spectra can also be used to follow the conformational changes in the protein.

In the near UV region (250-300 nm), the aromatic amino acids and the disulfide bonds act as chromophores. Phenylalanine gives signals in the region from 250-270 nm, tyrosine gives signals in the region from 270-290 nm and signals from 280-300 nm are given by tryptophan. Disulfide bonds give weak but broad signals throughout the near-UV spectrum. The CD signals in near UV region are sensitive to the overall tertiary structure of the protein.

CD spectra of the lectin samples were recorded on a JASCO-715 spectropolarimeter, at 25 ^{0}C, in the range of wavelengths 190–260 nm at a scan speed of 100 nm min^{-1} with a response time of 1 s and slit width 1 nm. The sensitivity was 20 mdeg. A rectangular quartz cell of 1 mm path length was used. All measurements were recorded at a lectin concentration of 0.08 mg ml^{-1}. 6 successive scans were collected for each spectrum, and their average was used for further analysis. Measurements were made in 25 mM buffers of pH 2.0, 4.0, 6.0, 7.2, 8.0, 10.0 and 12.0 and buffer scans recorded under the same conditions were subtracted from the protein spectra before further analysis.

Mean residue ellipticity (MRE) was calculated as:

MRE = (100 X θ X MW) / (l X c X N) deg cm^2 dmol^{-1} (Eq. 4.5)

Where θ is ellipticity in mdeg, MW is the molecular weight of the protein in daltons, l is the length of the light path in cm, c is the concentration of the protein in mg ml^{-1} and N is the number of amino acids in the protein.

To monitor the tertiary structure of the protein (0.8 mg ml^{-1} at pH 1.0, 7.2 and 10.0), CD spectra were recorded in the range of 250-300nm using a cuvette of path length 1 cm.

Effect of temperature on the secondary structure of MoL was studied by heating the protein at pH 7.2 from 30 to 90 ^{0}C (1 ^{0}C per min) and monitoring the ellipticity at 222 nm. Similarly, renaturation of MoL was studied by cooling the protein from 90 to 30 ^{0}C (1 ^{0}C per minute) and monitoring the ellipticity at 222 nm.

To study the effect of denaturants on the secondary structure of MoL, the protein (0.08 mg ml^{-1}) was incubated in the presence of 0.25-6 M GdnHCl or 0.25-8 M urea at pH 7.2 for 4 h at 25 ^{0}C and the scans were recorded in the range of 210-250 nm. The data at wavelengths lower than 210 nm could not be recorded due to increase in the noise and high tension. The effect of reducing agent dithiothreitol (DTT) on the secondary structure of MoL was studied by incubating the protein (0.08 mg ml^{-1}) at pH 2, 4, 6, 7.2, 8, 10 and 12 in the presence of 1 mM DTT for 4 h and recording the scans in the wavelength range 200-250 nm.

4.5. Results and discussion

MoL has a single tryptophan per monomer, as estimated by titration of the protein with NBS. When excited at 295 nm, the fluorescence emission spectrum of MoL showed maximum intensity at 351 nm, indicating that the tryptophans in the protein are in polar environment and are exposed to the solvent. Tryptophans fully exposed to the solvents show an emission maximum at 356 nm.

The far-UV CD spectrum of the native lectin was analyzed to derive quantitative information regarding secondary structural elements of the lectin by using three different methods, viz. CDSSTR, CONTINLL, and SELCON3 (Sreerama and Woody, 2000) available at http://lamar.colostate.edu/~sreeram/cdpro/main.html. A basis set containing spectra of 43 proteins was used as data for fitting the experimental spectrum. CDSSTR for the best fit values estimated that the lectin contained αR (regular α-helix) 16%, αD

(distorted α-helix) 12%, βR (regular β-sheet) 14%, βD (distorted β-sheet) 9%, turn 20% and unordered structure 28% (NRMSD= 0.069). Thus MoL is an alpha-beta protein.

4.5.1. Thermal stability of MoL

Effects of temperature on the conformation of MoL at pH 7.2 were studied by monitoring the changes in the intrinsic fluorescence, ANS binding as well as changes in the ellipticity at 222 nm. Increase in the temperature reduced the fluorescence intensity of the protein linearly (Fig. 4.1 A), without any change in the λ_{max} (Fig. 4.1 B). The native protein shows binding to the hydrophobic dye ANS at 30 ^{0}C, which reduces considerably with increasing temperatures (Fig. 4.1 C). The negative ellitpicity at 222 nm does not change much with temperature, only a marginal decrease of 1 mdeg (from -12.5 to -11.5) is observed which gets reversed by cooling (Fig. 4.1 D).

The decrease in the fluorescence intensity at higher temperatures could be due to the thermal deactivation of the fluorophore. Reduced ANS binding to MoL with increase in temperature shows that the hydrophobic patches are less accessible at higher temperature. The secondary structure of MoL remains stable at higher temperatures.

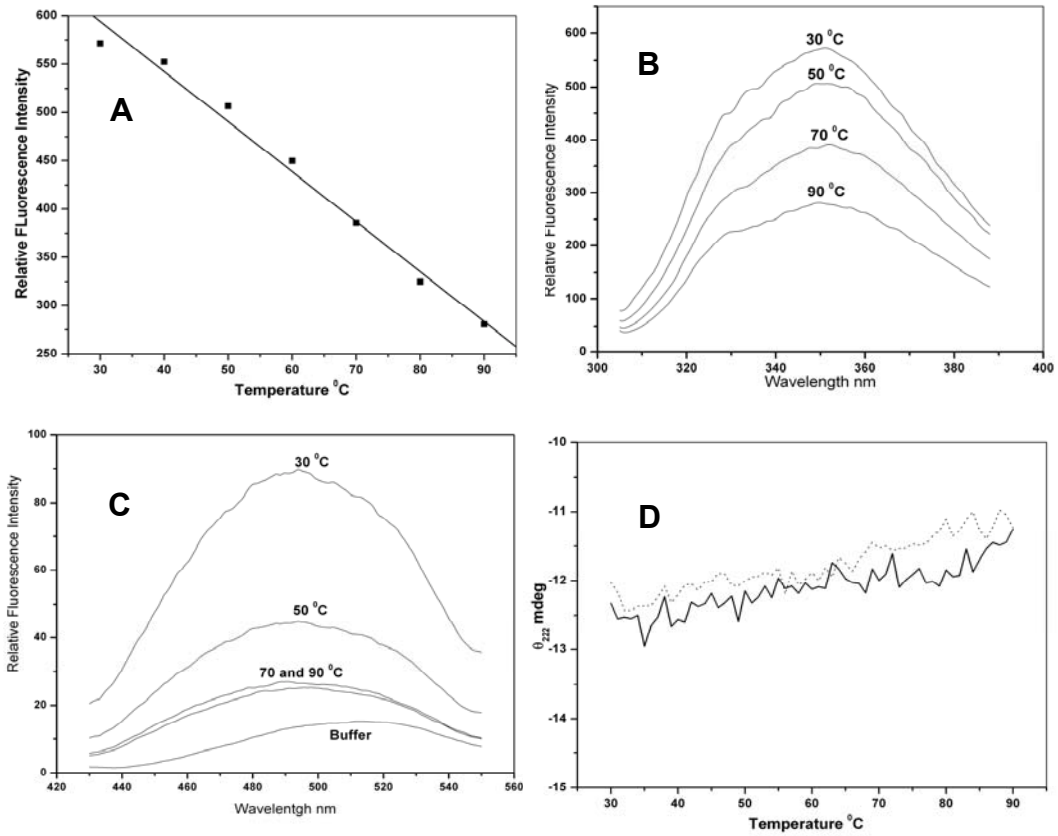


Fig. 4.1. Temperature stability of MoL

(A) Fluorescence intensity at 350 nm of MoL (0.04 mg ml^{-1}) as a function of temperature.

(B) Fluorescence emission spectra of MoL (0.04 mg ml^{-1}) incubated at various temperatures for 15 minutes. The temperatures are indicated on the spectra.

(C) Binding of ANS to MoL (0.05 mg ml^{-1}) at various temperatures. The temperatures are indicated on the spectra.

(D) Ellipticity of MoL (0.08 mg ml^{-1}) at 222 nm as a function of temperature. Solid line represents the denaturation, whereas the dashed line represents the renaturation of the protein by cooling.

4.5.2. Effect of pH on stability of MoL

Conformational stability of MoL at various pH was studied by fluorimetry as well as circular dichroism. The fluorescence λ_{max} of MoL does not change with pH. However, at acidic pH and at pH 8, MoL shows maximum fluorescence intensity while at extreme alkaline pH it reduces considerably (Fig. 4.2 A). When ANS binds to the native protein it shows a blue shift in the λ_{max} from 520 to 480 nm and an increase in the fluorescence intensity (Fig. 4.2 B). Compared to the binding at pH 7.2, there is a four times enhancement in the binding at pH 1.0 and 2.5 times enhancement at pH 10.0. The secondary structure of the protein is not affected by extreme acidic and alkaline conditions as seen in the far UV CD spectra (Fig. 4.2 C). The near UV CD spectrum (Fig. 4.2 D) of native protein shows that MoL possesses ordered tertiary structure. At alkaline and extreme acidic pH it gets affected only with respect to tryptophan and tyrosine environment (280-290 nm).

All these studies indicate that variation in pH in a wide range has little effect on the structural features of MoL. It was already shown that MoL retained its hemagglutination activity even at extreme pH values (section 3.5.2). The working pH for hemagglutination assay is 7.2; hence it could be possible that although MoL underwent structural changes at various pH, it regained its native structure when the pH was readjusted to 7.2 in the assay condition. However, conformational studies using fluorimetry and CD confirm that there are no major structural changes occurring in MoL even under extreme pH conditions.

Very few plant lectins are known to be stable in such a wide range of pH, some of them are: lectin from Pinto beans (pH range: 3-12) (Wong *et al.*, 2006); *Alocasia cucullata* lectin (pH range: 2-12) (Kaur *et al.*, 2005a) and *Arisaema tortuosum* lectin (pH range: 2-10) (Dhuna *et al.*, 2005).

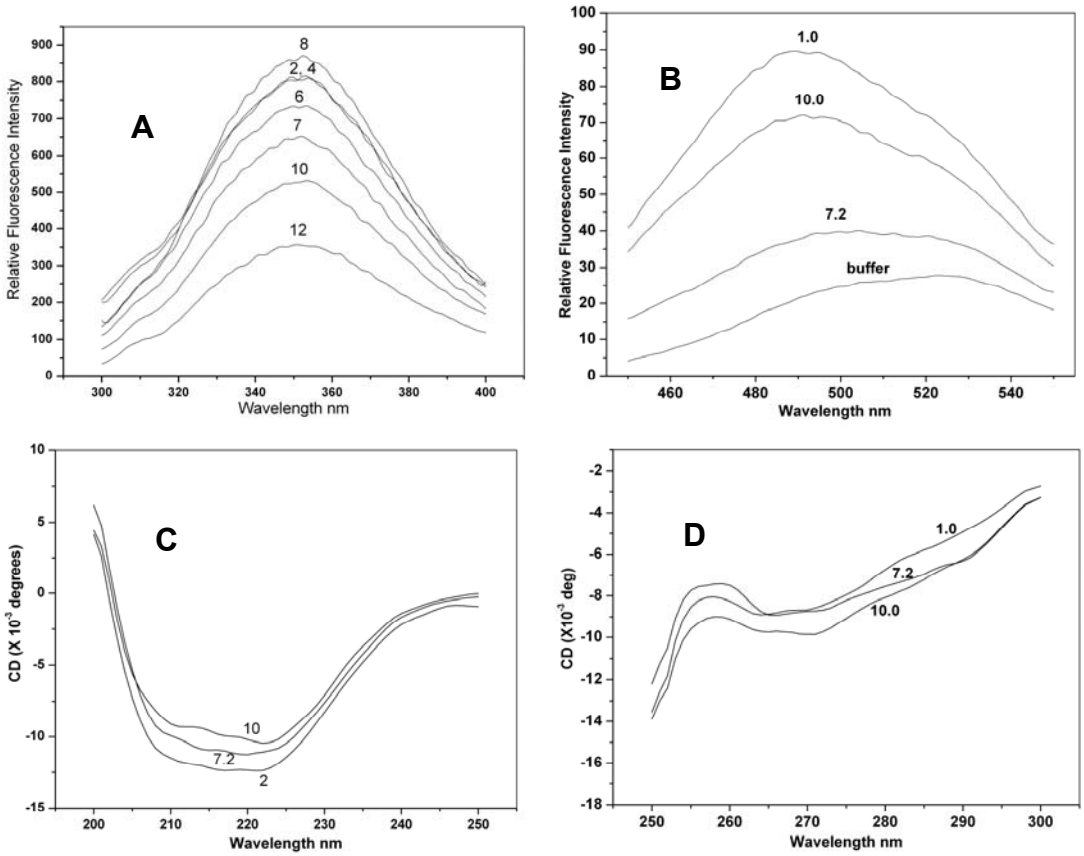


Fig. 4.2. Conformational stability of MoL at various pH. In each diagram, numbers on the spectra indicate the pH.

(A) Fluorescence emission spectra of MoL (0.04 mg ml^{-1}) incubated at various pH for 16 h.

(B) ANS Binding to MoL (0.05 mg ml^{-1}) at various pH.

(C) Far-UV CD spectra (protein concentration 0.08 mg ml^{-1}) at pH 2.0, 7.2, and 10.0.

(D) Near-UV CD spectra (protein concentration 0.8 mg ml^{-1}) at pH 1.0, 7.2 and 10.0.

4.5.3. Effect of denaturants and reducing agent on MoL

A. Guanidium hydrochloride (GdnHCl)

In the presence of GdnHCl, the fluorescence intensity of MoL increased with increase in GdnHCl concentration up to 1-2 M, along with slight red shift in λ_{max}, indicating increased exposure of the tryptophan to polar environment. However, with further increase in the concentration of GdnHCl, the intensity decreased and at 6 M concentration, a much broader spectrum was observed, indicating partial unfolding of the protein (Fig. 4.3 A). ANS binding decreases even at a low concentration (1M) of GdnHCl (Fig. 4.3 B) indicating that the minor variation in the ionic or hydrophobic interactions in the structure of the protein tend to modify the exposure of hydrophobic side chains. Thus the hydrophobic pockets in the protein seem to respond to even small changes in the environment. No wonder, MoL also loses hemagglutination activity in the presence of lower concentrations (0.25 M) of GdnHCl.

The CD spectra of the protein in the far UV region incubated in the presence of the different concentrations of GdnHCl are shown in Fig. 4.3 C. Although the ellipticity in the region between 215 to 225 nm appears unchanged, that at 210 nm or below (not shown due to noise) seems to be affected indicating the increase in the unordered structural element. Hemagglutinating activity of the protein might be getting abolished due to these structural changes.

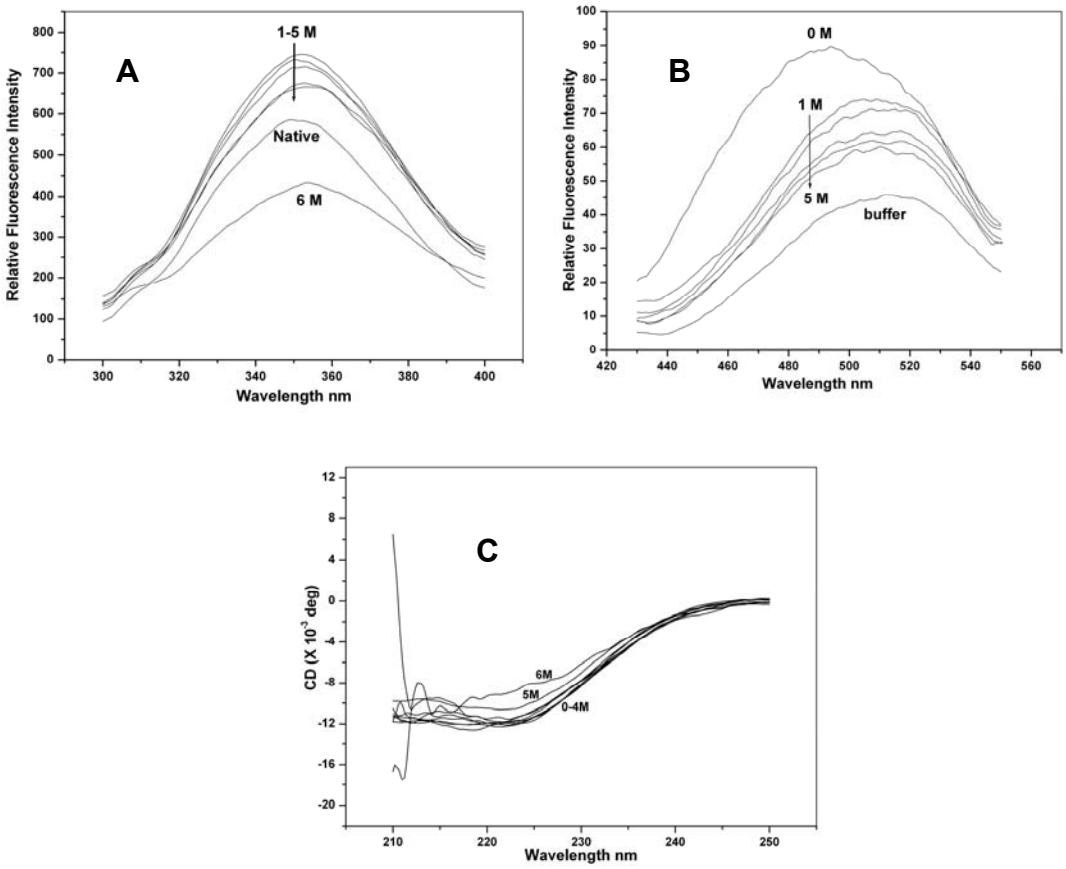


Fig. 4.3. Conformational stability of MoL in presence of various concentrations of GdnHCl. Concentrations of GdnHCl are indicated on the respective spectra.

(A) Fluorescence emission spectra of MoL (0.04 mg ml^{-1}) incubated for 16 h with different concentrations of GdnHCl.

(B) ANS Binding to MoL (0.05 mg ml^{-1}).

(C) Far-UV CD spectra (protein concentration 0.08 mg ml^{-1}) recorded after incubating the protein sample with different concentrations of GdnHCl for 4 h.

B. Urea

Fluorescence intensity of MoL remained constant up to 2M urea concentration, without any change in the λ_{max}. However, at 4 M concentration, there was a slight red shift in the emission spectrum with an increase in the intensity. After this, the intensity decreased further with any increase in urea concentration (Fig. 4.4 A). ANS failed to bind to protein above 2 M urea concentration (Fig. 4.4 B), whereas the secondary structure was found to be stable up to 3M urea concentration after which it started getting affected (Fig. 4.4 C). Above 6 M urea concentration, the secondary structure was drastically affected.

Urea, being a milder denaturant than GdnHCl, requires a higher concentration to be effective. It also explains, in this case, the retention of hemagglutinating activity of MoL below 3 M urea concentration.

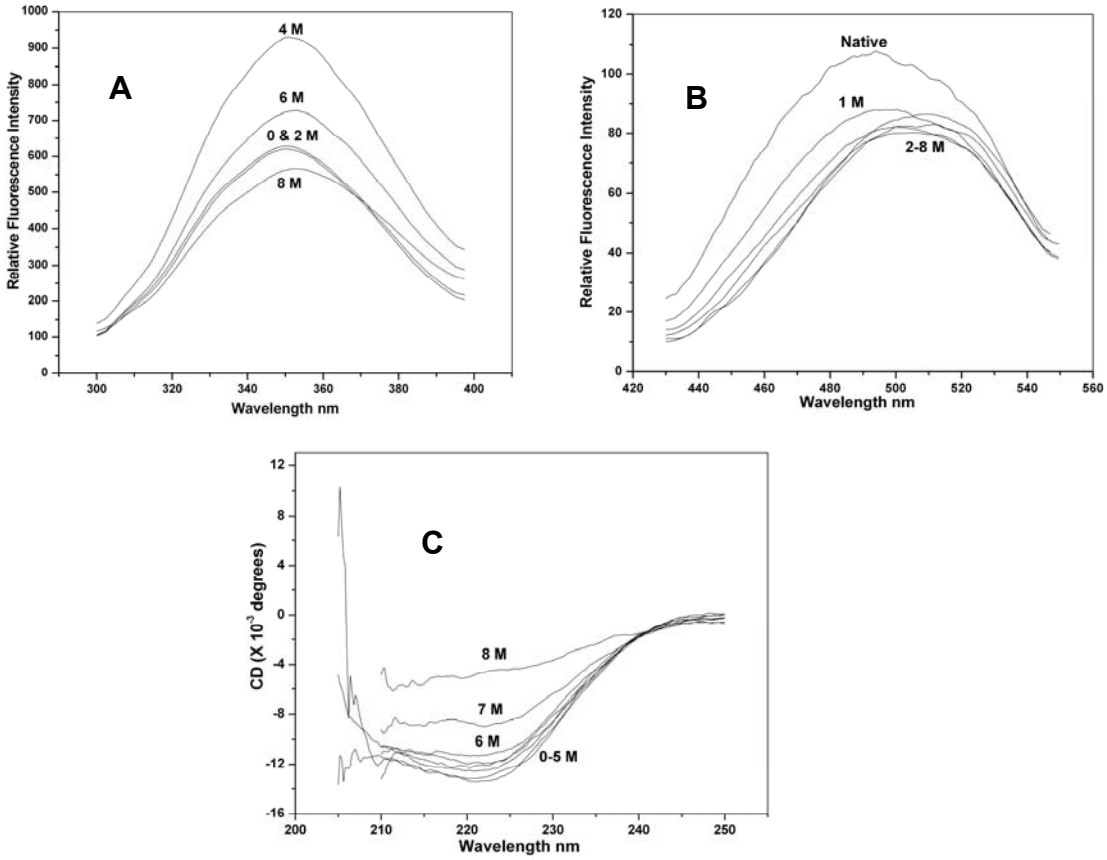


Fig. 4.3. Conformational stability of MoL in the presence of various concentrations of Urea. Concentrations of Urea have been indicated on the respective spectra.

(A) Fluorescence emission spectra of MoL (0.04 mg ml^{-1}) incubated for 16h with various concentrations of urea.

(B) ANS Binding to MoL (0.05 mg ml^{-1}).

(C) Far-UV CD spectra (protein concentration 0.08 mg ml^{-1}) recorded after incubating the protein sample with different concentrations of urea for 4 h.

C. Dithiothreitol

Chemical modification studies on MoL indicated that the native dimer contains three disulfide linkages. To study the role of these disulfide bonds on the structural stability of MoL, CD analysis was carried out in the presence of the reducing agent dithiothreitol (DTT). The secondary structure of MoL was drastically affected in the presence of 1 mM DTT at and above pH 7.2 (Fig. 4.4). The protection of the structure at lower pH values is merely due to the fact that DTT needs an alkaline environment to function as a reducing agent. This explains the loss of hemagglutination activity of MoL in the presence of DTT. The disulfide linkages present in the protein seem to hold the structure intact and offer the required conformational stability essential for the hemagglutination activity.

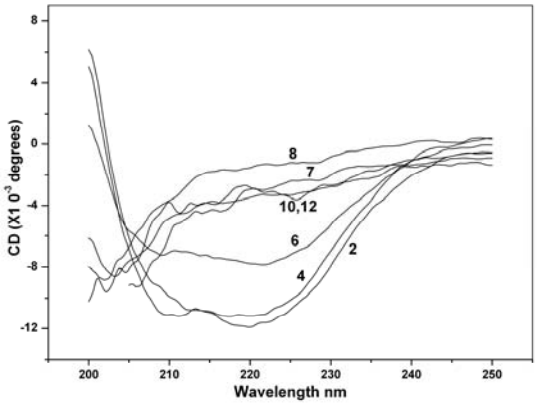


Fig. 4.4. Far-UV CD spectra of MoL (protein concentration 0.08 mg ml^{-1}) in the presence of 1 mM DTT recorded at various pH. Numbers on the spectra indicate the pH.

4.5.4. Fluorimetric analysis of sugar binding

Specific binding of the sugar to a lectin can change the environment of tryptophan either by enhancing (Gaikwad and Khan, 2006) or by quenching (Khan *et al.*, 2007) the fluorescence. This property can be used to calculate the binding affinity of the sugar towards protein.

Addition of glucose, galactose, lactose, sucrose, mannose, methyl α-glucose, methyl α-galactose, methyl α-mannose to MoL did not significantly change the fluorescence intensity of the protein. However, quenching was observed upon addition of *N*-acetyl-lactosamine (LacNAc) and fructose with no change in the emission maxima, although these sugars failed to inhibit the hemagglutination activity of MoL. Titration of MoL with LacNAc showed 9%, while titration with fructose showed 8 % quenching of the fluorescence.

The plots of $F_0/\Delta F$ vs $1/C$ for LacNAc and fructose are shown in Fig. 4.5 (A and B, respectively) from which the values of F_∞ have been determined and the double logarithmic plots have been shown in Fig. 4.5 (C and D, respectively). The affinities of MoL towards fructose and LacNAc are considerably low (K_a = 975 and 1380 M^{-1}, respectively).

The corresponding ΔG values for binding of fructose and LacNAc to MoL are -17.16 and -18.03 kJ mol^{-1}, respectively, which indicates the spontaneous nature of this binding.

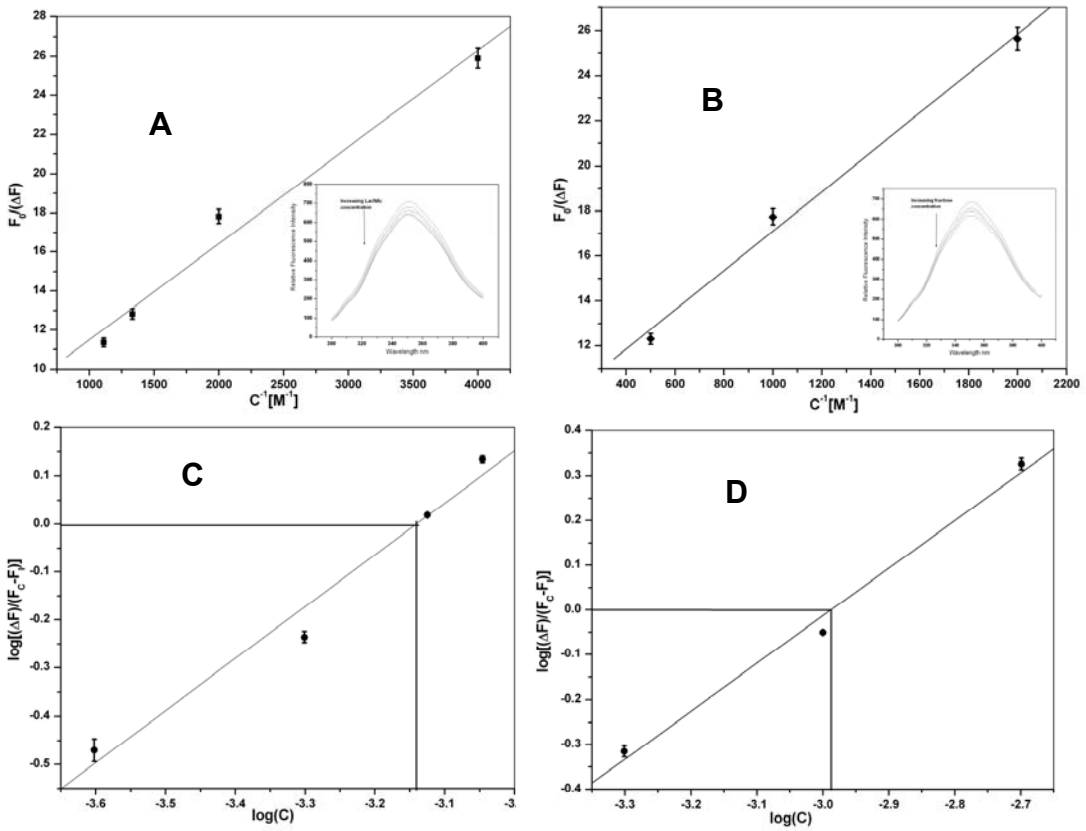


Fig. 4.5. Determination of association constant for the binding of sugars to MoL (0.04 mg/ml). Sugars used: LacNAc and Fructose, 10 mM stock each.

(A) and **(B)** Plots of ($F_0/\Delta F$) vs $[C]^{-1}$ for LacNAc and Fructose, respectively. Insets represent fluorescence quenching of MoL on addition of aliquots of sugars to the lectin solution.

(C) and **(D)** Plots of $\log(\Delta F/F_C\text{-}F_\infty)$ vs $\log(C)$ for LacNAc and Fructose, respectively.

4.5.5. Binding of adenine to MoL

The presence of exposed hydrophobic patches in the native lectin led us to check the binding of adenine to the protein. Addition of adenine resulted in quenching of the intrinsic fluorescence of tryptophan (Fig. 4.6 A inset) with K_a = 7.76 x 10^3 M^{-1}, which is significantly less than those estimated for other lectins reported to bind adenine (lima bean lectin; K_a = 8.3 x 10^4 M^{-1}, *Phaseolus vulgaris* erythroagglutinin; K_a = 1.2 x 10^5 M^{-1} and soybean agglutinin; K_a = 7.7 x 10^4 M^{-1}, *Dolichos biflorus* seed lectins; K_a = 7.31 x 10^5 M^{-1} and *Dolichos biflorus* stem lectin; K_a = 1.07 x 10^6 M^{-1} (Gegg *et al.*, 1992)). The ΔG calculated for adenine binding to MoL is –22.324 kJ mol^{-1}. The plots of $F_0/\Delta F$ vs $1/C$ and the double logarithmic plots for adenine binding are shown in Fig. 4.6 (B and C, respectively). Thus the adenine-binding site seems to be close to tryptophan. However, adenine does not inhibit the hemagglutination activity of MoL. This may be due to the known fact that the adenine binding sites can be distinctly different from the carbohydrate binding sites (Roberts and Goldstein, 1982; Hamelryck *et al.*, 1999).

The above mentioned legume lectins also showed binding to cytokinins, which are a group of plant hormones (Roberts and Goldstein, 1983b; Gegg *et al.*, 1992). Apart from legume lectins, wheat germ agglutinin (WGA) has also been shown to bind adenine and adenine-related phytohormones such as zeatin and kinetin as well as abscissic and gibberillic acids with affinities in the range of K_a = 1.6-2.3X10^6 M^{-1} (Bogoeva et al., 2004). MoL could also be binding in vivo cytokinins and other related plant hormones on the basis of observing adenine binding sites present in vitro.

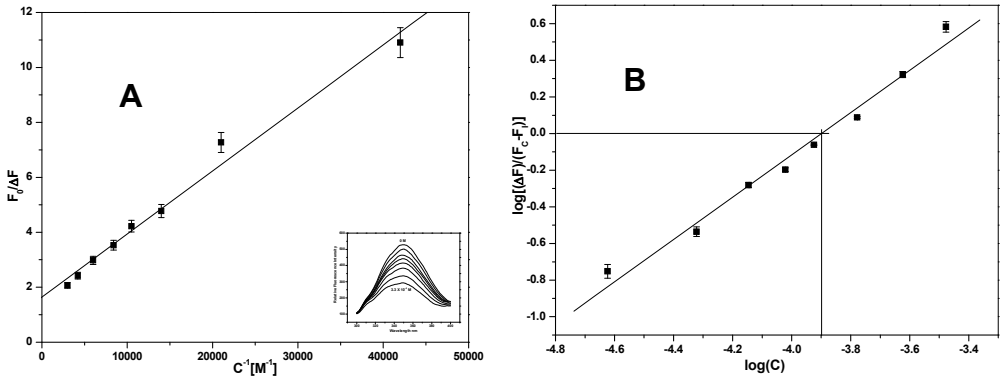


Fig. 4.6. Determination of association constant for the binding of adenine (10 mM stock in methanol) to MoL (0.04 mg/ml).

(A) The plot of ($F_0/\Delta F$) vs $[C]^{-1}$ for adenine. Inset represents fluorescence quenching of MoL on addition of aliquots of adenine to the lectin solution.

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

4.5.6. Solute quenching studies

Fluorescence spectra of native MoL recorded in the presence of increasing concentrations of acrylamide and KI are shown in Fig. 4.7 (A and B, respectively). Among the ionic quenchers used, CsCl failed to quench the fluorescence of MoL, whereas KI quenched almost 85% of the fluorescence in the native condition, indicating that the environment of the tryptophan residue is highly electropositive. The extent of fluorescence quenching achieved in each case is shown in Table 4.1.

Table 4.1 Extent of quenching of intrinsic fluorescence of MoL with acrylamide and KI in various conditions.

	Extent of quenching %	
Quencher	**Acrylamide (0.5 M)**	**KI (0.5 M)**
Native (pH 7.2)	82 ± 3	86 ± 3
pH 1.0	90 ± 3	95 ± 3
pH 10.0	89 ± 3	85 ± 3
In 6 M urea	88 ± 3	86 ± 3

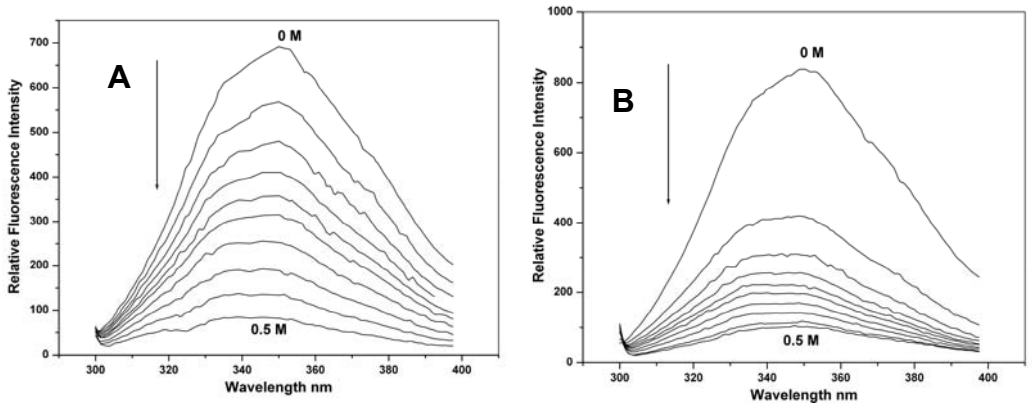


Fig. 4.7. Fluorescence emission spectra of MoL (0.05 mg/ml) in native condition in the absence and in the presence of quencher (**A**: acrylamide; **B**: potassium iodide). The quencher concentration for spectra is increased from 0 to 0.5 M, in the direction indicated by the arrow.

4.5.7. Stern-Volmer analysis of the quenching data

The Stern-Volmer plots (Eq. 4.3) obtained with acrylamide quenching showed a positive curvature (Fig. 4.8 A), which indicates that the quenching has both dynamic and static components. The static mechanism involves complex formation between tryptophan and acrylamide, while dynamic mechanism involves collisions with acrylamide during the lifetime of tryptophan in excited state. In such a case, the data can be analyzed by Eq. 4.6 by which the dynamic and the static components can be resolved (Lakowicz and Weber, 1973).

$$F_0/F_C = (1 + K_{sv}[Q])(1 + K_s[Q]) \qquad \text{(Eq. 4.6)}$$

Where K_{sv} is the Stern-Volmer (dynamic) quenching constant, K_s is the static quenching constant and [Q] is the quencher concentration. The dynamic quenching constant reflects the degree to which the quencher achieves the encounter distance from the fluorophore and can be determined by the fluorescence lifetime measurements according to the equation (Lakowicz and Weber, 1973)

$$\tau_0/\tau = (1 + K_{sv}[Q]) \qquad \text{(Eq. 4.7)}$$

Where τ_0 is the average lifetime in absence of the quencher and τ is the lifetime in the presence of a quencher at a concentration [*Q*]. Using the average lifetimes obtained from analysis of the time resolved fluorescence data, as described in the section 4.5.8, the value of K_{sv} obtained for acrylamide quenching of MoL was 0.216 M^{-1}. The low value of K_{sv} suggests low collision frequency. By substituting this value in Eq. 4.6 and plotting a graph of $(F_0/F_C)/(1+K_{sv}[Q])$ against [*Q*], the value of the static quenching constant (K_s) was obtained as 8.19 M^{-1}. The bimolecular quenching constant, k_q was calculated as $k_q = K_{sv}/\tau$, (Lehrer, 1971) and was found to be 7.55 x 10^{11} $M^{-1}s^{-1}$. Incorporating the values of K_{sv} and K_s in the expression $(1 + K_{sv}[Q])(1 + K_s[Q])$, the values obtained were plotted against [*Q*]. It was observed that the values of F_0/F_C and $(1 + K_{sv}[Q])(1 + K_s[Q])$ match very well (Fig. 4.8 B).

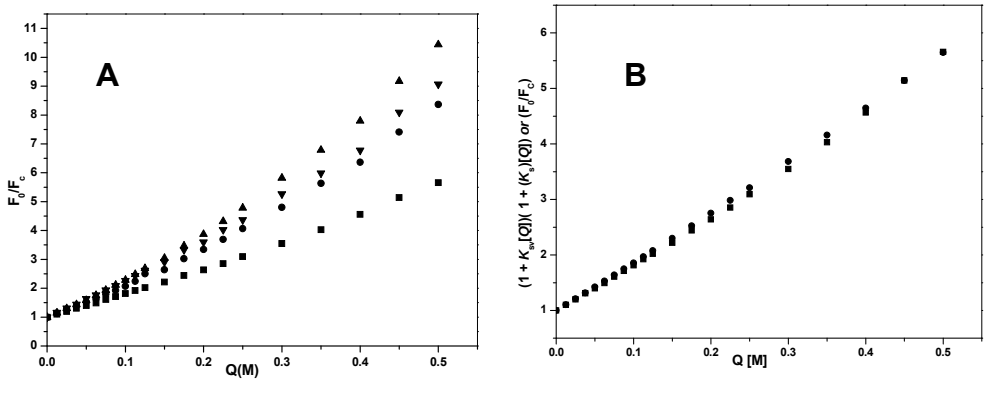


Fig. 4.8. (A) Stern Volmer analysis of fluorescence quenching of MoL. Plots of quenching profiles with acrylamide, in conditions
■ Native; ● 6 M Urea; ▼pH 10.0; ▲pH 1.0.
(B) The plot of F_0/F_C and $(1 + K_{sv}[Q])(1 + K_s[Q])$ against [*Q*] corresponding to the quenching of native MoL with acrylamide. ■ F_0/F_C, ● $(1 + K_{sv}[Q])(1 + K_s[Q])$.

4.5.8. Quenching of MoL fluorescence with iodide

Iodide exhibited a strong quenching effect on the fluorescence of MoL (85%), which was quite unusual. Stern-Volmer analysis showed a sharp downward curvature of the plot, which could not be resolved into linear components (Fig. 4.9). This cannot be correlated with the single tryptophan in the protein. Thus, apart from the strong positive charge around the tryptophan, there could be some non-specific binding of iodide to the protein. Iodide having a large ionic radius and being negatively charged, probably binds to the positively charged amino acid residues present in the neighborhood of the single tryptophan in the protein leading to affinity quenching of the fluorescence rather than collisional quenching.

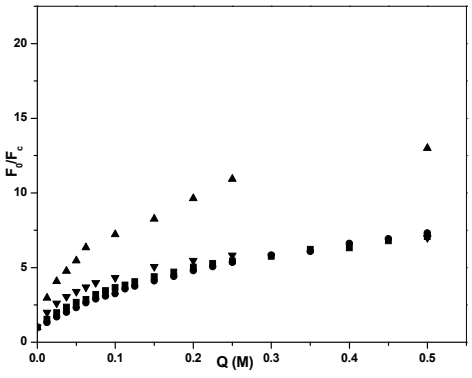


Fig. 4.9 Stern Volmer analysis of fluorescence quenching of MoL. Plots of quenching profiles with potassium iodide, in conditions ■ Native; ● 6 M Urea; ▼pH 10.0; ▲pH 1.0.

4.5.9. Lifetime measurements of fluorescence emission of MoL

The lifetime measurements of the fluorescence due to MoL from the decay curve (Fig. 4.10 A) was done by fitting it to a biexponential function (χ^2<1.005). From this fit two decay times τ_1 and τ_2 with their corresponding weight factors α_1 and α_2 were

obtained (Table 4.2). The native lectin showed two lifetimes, τ_1 (1.6 ns) and τ_2 (4.36 ns) with 54% and 46% contributions, respectively, indicating presence of two conformers of the single tryptophan (Martinho *et al.*, 2003). The shorter lifetime component contributed more to the fluorescence than the longer lifetime component. Normally the shorter lifetime component is due to fluorophors exposed on the surface while the longer one is due to fluorophors buried in the interior of the protein.

From the life time measurements of the quenching of the intrinsic fluorescence of MoL by acrylamide, the decay curve (Fig. 4.10 B) could be fitted to a bi-exponential function (χ^2<1.005) for concentrations of acrylamide below 0.1 M, above which they could be fitted well with a mono-exponential function, indicating that there is only one conformer of tryptophan remaining after a certain concentration of acrylamide is reached. Both τ_1 and τ_2 tend to decrease with increasing acrylamide till 0.1 M concentration of the acrylamide was reached, above which only one lifetime was observed which also decreased from 2.21 to 2.07 ns at 0.45 M concentration of acrylamide. The decrease in the average lifetime from 2.86 to 2.07 ns could be due to low collisional frequency.

The average lifetimes were calculated using the following equations (Inokuti and Hirayama, 1965; Grinvald and Steinberg, 1974):

$$\tau = \Sigma_i \alpha_i\tau_i / \Sigma_i \alpha_i \qquad \text{(Eq. 5)}$$

$$\langle\tau\rangle = \Sigma_i \alpha_i\tau_i^2 / \Sigma_i \alpha_i\tau_i \qquad \text{(Eq. 6)}$$

i=1, 2……

Where τ and $\langle\tau\rangle$ are the average life times obtained by two different approaches and α is the weighting factor.

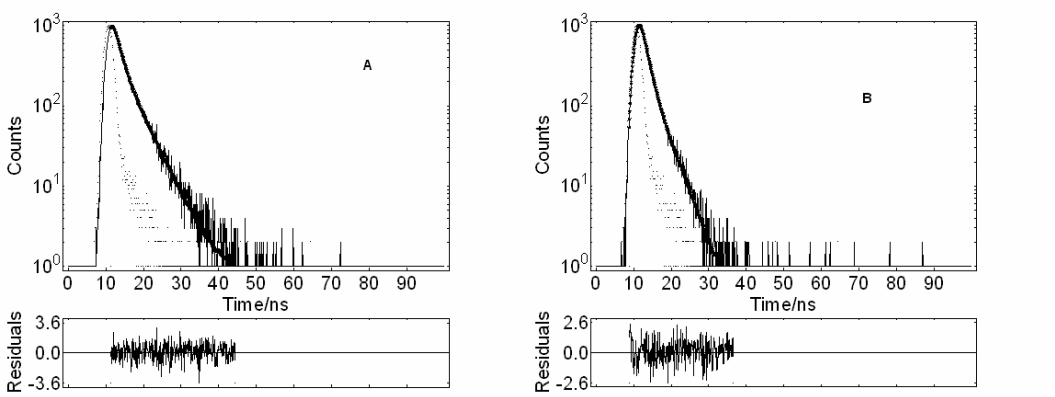


Figure 4.10: The fluorescence lifetime decay curves for native MoL **(A)** and MoL quenched with 0.5 M acrylamide **(B)**. The dotted lines correspond to the instrument response, the solid lines correspond to the experimental data and the starred lines correspond to the nonlinear biexponential fit of the experimental data to a biexponential function (in the case of native protein) or a monoexponential fit (in the case of quenched protein). The lower panels represent the corresponding residuals.

The plot of τ_0/τ for the quenching data of MoL with acrylamide in native condition is shown in Fig. 4.11, from which K_{sv} was calculated.

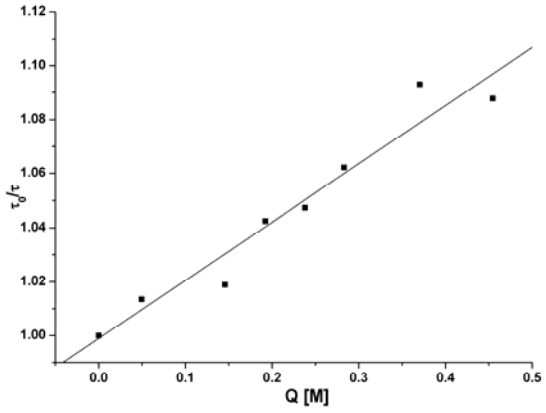


Figure 4.11: Plot of τ_0/τ for the quenching data of MoL with acrylamide in native condition.

Table 4.2. The lifetimes of fluorescence decay of MoL and the corresponding pre-exponential factors along with calculated average lifetimes for acrylamide quenching.

Q [M]	τ_1 (ns)	α_1	τ_2 (ns)	α_2	τ (ns)	$<\tau>$ (ns)	χ^2
0	1.60	0.065	4.37	0.020	2.25	2.86	1.002
0.050	1.66	0.061	3.65	0.024	2.22	2.58	1.002
0.098	1.32	0.058	2.94	0.034	1.92	2.25	1.003
0.146	2.21	0.079	0	0	2.21	2.21	1.002
0.192	2.16	0.079	0	0	2.16	2.16	1.003
0.238	2.15	0.079	0	0	2.15	2.15	1.002
0.283	2.12	0.083	0	0	2.12	2.12	1.004
0.370	2.06	0.083	0	0	2.06	2.06	1.001
0.455	2.07	0.091	0	0	2.07	2.07	1.005

4.6. Conclusions

Conformational studies carried out on MoL using fluorescence spectroscopy and circular dichroism supported the observed stability and the hemagglutination activity of MoL at high temperatures and extreme pH. The secondary structure of MoL hardly gets affected at higher temperatures and extremes of pH. Although the intensity of the intrinsic tryptophenyl fluorescence of the protein changes with temperature and pH, the λ_{max} of the fluorescence spectrum is not affected much, indicating no change in the overall conformation of the protein. The disulfide linkages in the protein seem to be essential for the active conformation of MoL, as both the activity and the secondary structure of MoL get severely affected by the presence of a reducing agent like DTT. Denaturants like GdnHCl and urea did affect the random coil element of the secondary structure causing some variations in the ionic or hydrophobic interactions in the structure of the protein which in turn affected the activity as well as ANS binding property of the protein.

MoL is a lectin exhibiting specificity solely for glycoproteins in hemagglutination experiments. However, in the sugar binding studies by fluorimetry, the lectin shows weak binding to sugars LacNAc and fructose. A clustering effect of multiple Gal-β-1,3-GalNAc or LacNAc chains might be necessary for binding, which has to be confirmed by studying the binding properties of glycopeptides of fetuin or thyroglobulin to MoL.

Binding of the hydrophobic dye ANS and adenine to the native protein indicates that MoL might have a physiological role in binding the phytohormones.

The single tryptophan per monomer in MoL is highly exposed to the solvent environment as well as surrounded by basic amino acid residues, as seen by fluorescence quenching studies. The native protein shows two fluorescence lifetimes, which get converged to a single one after quenching with acrylamide.

Chapter 5

Analysis of Polymorphism in the Crystal Structures of Plant Lectins

5.1. Summary

Many lectins, including *Cicer arietinum* hemagglutinin studied here, crystallize in different space groups or in the same space group with different unit cell dimensions. The analysis present here is carried out to understand the similarity and differences in the packing environments of molecules in the polymorphic crystals of plant lectins. A total of 300 plant lectin structures have been considered. Concanavalin A (ConA) is the lectin that shows maximum extent of polymorphism (17 polymorphs from 51 crystal structures), whereas *Pterocarpus angolensis* lectin shows least number (2 polymorphs from 30 crystal structures).

A thorough analysis of polymorphism in crystal structures of ConA was carried out. As expected the polymorphs with any common unit cell length have similar intermolecular contacts and packing arrangement along that particular cell axis. The intermolecular contacts that occur repeatedly in the lattice of different polymorphic crystal forms might have important role in its agglutination activity and oligomerisation in the solution. The tetrahedral arrangement of the subunits in the tetramer of ConA does not get distorted as much as a result of crystal packing and/or binding of ligand.

5.2. Introduction

Polymorphism in the crystals of small molecules has invited considerable attention during the last two decades because of its importance in drug research and pharmaceutical industry (Bernstein, 2002). Small organic molecules crystallize in different space groups owing to their inherent flexibility and different modes of interaction and sometimes due to presence solvent molecules of crystallization. However, when compared to the total number of crystal structures available in Cambridge structural database (CSD) (Allen *et al.*, 1979; 1991) only a small percentage of it is polymorphic crystal structures (Allen and Motherwell, 2002). Mostly, individual molecules crystallize in

same space group when crystallized under identical conditions. Even those molecules which show polymorphism in their crystals do not crystallize in all the crystal systems, implying that the available crystal structures are not evenly distributed under different space groups. Analysis of structural databases has shown that not all space groups are represented by crystal structures and an overwhelming preference is for a few space groups (Donohue, 1985; Wilson, 1988). An explanation for this large frequency of a few space groups among crystals is provided in terms of the limited packing preferences of molecules (Wilson, 1993; Brock and Dunitz, 1994).

Like crystals of organic molecules, proteins also show preference for a few space groups (Wukovitz and Yeates, 1995). However, proteins are more flexible compared to organic molecules and possess a variety of interacting surface groups and a higher solvent of crystallization present in their crystals. Hence, proteins are expected to show higher degree of polymorphism and pseudopolymorphism in their crystals compared to organic molecules. In protein crystallography, on the average, in the case of several proteins, reports of the preliminary characterization of different crystal forms of the same protein used to appear. However, all of them seldom are refined and deposited in protein data bank (RCSB PDB, Berman *et al*., 2000), presumably due to reasons such as poor diffraction or inconsistent reproducibility of some of the crystal forms (Kozak *et al*., 2002). In most applications the polymorphism in protein crystals poses lesser problems compared to organic molecules. Nevertheless, importance of polymorphism in protein crystals is recognized at least in the case of therapeutic proteins like insulin (Bernardo *et al*., 2005). It is becoming clear that a study of the intermolecular interactions in polymorphic crystal structures using crystallography, combined with study in solution using NMR, can help to understand the plasticity and dynamics of protein structures. (Kondrashov *et al*., 2007).

Many patches of protein surface can exhibit similar properties and involve in crystal packing interactions with equal binding energies. However, in crystals of a protein only some specific contacts are observed, depending on the lattice and/ or influenced by crystallization condition. Thus, sometimes mutant proteins crystallize in different space groups than wild type, due to changes in packing interactions resulting from specific mutations (Zhang *et al*., 1995; Jelsh *et al*., 1998). Similarly, influence of specific ions present in crystallization solution can determine a particular packing mode (Vaney *et al*., 2001; Harata and Akiba, 2007). Presence of organic molecules specifically recognized by protein can also selectively induce crystallization in a particular space group (Retailleau *et al*., 2005). Interestingly, in some cases, like those lectins which we have been studying, different crystal forms grow under the same crystallization condition (Rao *et al*., 2004; Chapter 2).

The crystallization of protein is a very crucial step in the study of three-dimensional structure. Conditions such as pH, type of buffer, temperature, precipitants, additives etc. influence the quality of crystals as well as packing interactions, and hence decide the space group and/ or unit cell (McPherson, 1999). Knowledge of these factors and the magnitude of their influence are important for obtaining crystals of heavy atom derivatives grown from crystallization solution containing heavy atom salt instead of soaking the pre-prepared crystals in heavy atom solution. In this case, a strict isomorphism is necessary and polymorphism is the most unwanted phenomenon. Although, proteins are individuals in terms of their properties, it will be very useful to understand the common rules and parameters that influence crystallization to help in selecting conditions where better quality crystals of desired form can be grown.

Studies of intermolecular interactions in crystal structures have used information from polymorphic crystals, particularly from ribonuclease, lysozyme and cutinase (Crosio *et al*., 1990; 1992; Zhang *et al*., 1995; Dasgupta *et al*., 1997; Jelsch *et al*., 1998;

Matsuura and Cheranov, 2003). It may be mentioned that in crystals of certain proteins non-isomerism due to large changes in one unit cell dimension as a result of local moments of flexible regions in proteins, without affecting the overall packing, has been reported (Swarnamukhi *et al.*, 2007; Redinbo *et al.*, 1999; Verschueren *et al.*, 1999; Randal and Kossiakoff, 2000). In our studies we have observed polymorphism in the crystals of *Artocarpus hirsuta* lectin (Rao *et al*., 2004), *Cicer arietinum* lectin (chapter 2) as well as in the reported jacalin crystals (Dhanraj *et al*., 1988; Banerjee *et al*., 1991). Concanavalin A (ConA), the first plant lectin for which three-dimensional structure has been reported (Hardman and Ainsworth, 1972), also happens to be the protein with the highest number of available polymorphic crystal structures. In the present study, an analysis of ConA in its different crystal forms has been carried out. The intermolecular interactions and packing arrangements have been studied in seventeen crystal forms of ConA span over four crystal systems. Also, an attempt has been made to list out the observed polymorphism in some of the other plant lectins present in the three-dimensional lectin database (http://www.cermav.cnrs.fr/lectines/).

5.3. Materials and Methods

All the computational work was carried out on a Silicon Graphics workstation (Octane) with Irix 6.5 as the operating system. The plant lectin structures enlisted in the lectin database (http://www.cermav.cnrs.fr/lectines/) were downloaded from the RCSB PDB site (http://www.rcsb.org/pdb). The crystal forms having less than 10% difference in all the unit cell parameters (*a*, *b*, *c* and *α*, *β*, *γ*) were considered to be isomorphous and belonging to the same crystal form. Wherever more than one structure were available in the same crystal form, the choice was made on the basis of maximum resolution and/or minimum overall and free R_{factor}. Sometimes, it was seen that different unit cells arose due to the non-conventional assignments of the unit cell parameters. In such cases, if

the unit cells were found to be convertible to a form already considered by interchanging the *a*, *b* or *c*, they were excluded from the analysis. Matthews coefficient and solvent content of each unique crystal form were calculated using the online "Matthews Coefficient (VM) Calculator" available on http://www.esrf.eu/exp_facilities/BM14/runexperiment/tools/matthews.html and assuming the monomeric molecular weight of 26000 Da.

The graphical software QUANTA (Accelrys, Inc) was used to visualize the molecules. Using the X-AUTOFIT/X-BUILD module of QUANTA (Accelrys, Inc.), the symmetry related molecules were generated and viewed. A pdb file was prepared for every symmetry related molecule which was seen to be in the direct contact with the original biological unit using the "write symmetry to pdb" tool in the X-AUTOFIT/X-BUILD module. The biological unit was generated wherever the asymmetric unit and biological unit were different, using the information in PDB or making use of symmetry related molecules. This biological unit was further used for studying the interactions with symmetry related molecules.

By selectively displaying the original biological unit and one of the symmetry related molecules at a time, the interactions at the interface between the two molecules were studied. Potential intermolecular hydrogen bonded contacts within a distance of 2.5-3.5 Å were generated using the CONTACT program of CCP4 program suite, as well as by using the "show/list neighbors" tool in QUANTA (Accelrys, Inc.). The distances between C^{α} trace of neighboring molecules were calculated by displaying only the C^{α} trace of both molecules and using the "distance" command in QUANTA (Accelrys, Inc.). Hydrogen bonded contacts and the C^{α} distances < 10 Å were manually analyzed. The water molecules were considered only if a water molecule at the interface of the protein molecules was found to be interacting with at least one residue of each protein. Also, an attempt was made to correlate the hydrogen bonded contacts with the crystallization

conditions like pH, precipitant and presence of ligand.

While preparing the diagrams of interactions, the two molecules were displayed as protein cartoons using the secondary structure information. Residues which interact with each other were shown in liquorices and labeled as per the residue name in three-letter code and residue number. The images were captured using the “snapshot” tool available in Irix 6.5. To prepare the packing diagrams, all the symmetry related molecules which lie in a plane were displayed with the original biological molecule at the center and the images were captured parallel to the particular plane. Interfaces between the neighboring molecules were labeled according to the interactions between them.

To calculate the centers of mass (COM) for each molecule and its subunits, the (centre-of-mass) command in Coot (Emsley and Cowtan, 2004) was used. The coordinates of COMs of all the four subunits as well as the tetramer were converted into a PDB file by putting a hypothetical water molecule at each of the five positions. This PDB file was displayed in Coot or PyMOL (DeLano, 2000) and the distances of COMs of each subunit from the COM of the tetramer as well as the angles between them were calculated.

In this analysis, we have not attempted to quantify the strength of interactions by estimating the solvent accessibility or energy of interactions.

5.4. Results and Discussion

5.4.1. Polymorphism in plant lectins

As of today, the 3-D lectin database (http://www.cermav.cnrs.fr/lectines/) contains about 300 three-dimensional structures of different plant lectins, belonging to five different families; namely, α-D-mannose-specific monocot lectins, agglutinins with hevein domain, β-prism plant lectins, β-trefoil lectins and legume lectins. A statistics of the structures available in each family is given in Table 5.1.

Table 5.1: *Statistics of available structures of lectins from five different families.

Type	Lectin family	No. of structures
A	α-D-mannose-specific monocot lectins	14
B	Agglutinins with hevein domain	26
C	β-prism plant lectins	32
D	β-trefoil lectins	24
E	Legume lectins	204
	Total	**300**

(*Reproduced from http://www.cermav.cnrs.fr/lectines/)

It can be easily seen that the legume lectin family is the one with largest number of structures reported. Out of all the available structures, only the ones solved by X-ray crystallography were considered for the analysis. Based on the space groups and/or unit cell dimensions, polymorphs were identified for each protein. **Table 5.2** enlists the lectins which have more than one crystal structure reported along with the total number of deposited pdb files and the number of polymorphs observed among them. The family of each lectin is indicated by a letter A, B, C, D or E, as per the Table 5.1.

Table 5.2: Number of crystal structures and polymorphs observed in different plant lectins.

Type	Name of the organism and protein	No. of PDB files of each protein	No. of Polymorphs	Biological unit
E	*Canavalia ensiformis* (Concanavalin A)	51	17	Tetramer
C	*Artocarpus integrifolia* (Artocarpin)	5	5	Tetramer
E	*Arachis hypogaea* (Peanut) agglutinin	23	5	Tetramer
C	*Artocarpus integrifolia* (Jacalin)	12	4	Tetramer*
E	*Erythrina cristagalli* lectin	5	4	Dimer
E	*Lens culinaris* lectin	4	4	Dimer
B	*Phytolacca americana*	4	4	Dimer
E	*Lathyrus ochrus* lectin 1	7	3	Dimer
E	*Pisum sativum* (pea) lectin	6	3	Dimer
B	*Urtica dioica* agglutinin	6	3	Dimer
A	*Galanthus nivalis* agglutinin	3	3	Tetramer
E	*Canavalia gladiata* lectin	5	3	Tetramer
D	*Ricinus communis* agglutinin (ricin B)	2	2	Tetramer
E	*Pterocarpus angolensis* lectin	30	2	Dimer
E	*Erythrina corallodendron* lectin	8	2	Dimer
E	Winged bean agglutinin I	7	2	Dimer
E	Soybean agglutinin SBA	5	2	Tetramer
B	Wheat germ agglutinin (WGA-1)	3	2	Dimer
E	*Dioclea grandiflora* lectin	3	2	Tetramer
A	*Gastrodia elata* lectin	2	2	Monomer
C	*Parkia platycephala* lectin	2	2	Dimer
E	*Cratylia floribunda* lectin	2	2	Tetramer
E	*Dioclea guianensis* lectin	2	2	Tetramer
E	*Griffonia simplicifolia* lectin	2	2	Tetramer
E	*Lathyrus ochrus* lectin 2 (LOL2)	2	2	Dimer
E	*Ulex europaeus* agglutinin (UEA-1)	2	2	Dimer
C	*Artocarpus hirsuta* lectin	2	2	Tetramer*
E	Winged bean agglutinin II	2	2	Tetramer
D	*Viscum album* lectin	9	1	Heterodimer
E	*Ulex europaeus* agglutinin UEA-2	5	1	Tetramer
D	*Sambucus ebulus* lectin (Ebulin)	4	1	Heterodimer
C	*Musa paradisiaca* lectin	4	1	Dimer
B	Wheat germ agglutinin WGA-3	4	1	Dimer
C	*Helianthus tuberosus* lectin	3	1	Tetramer
E	*Canavalia maritima* lectin	3	1	Tetramer
E	*Phaseolus vulgaris* lectin	2	1	Tetramer
E	*Robinia pseudoacacia* lectin	2	1	Tetramer
E	*Dolichos biflorus* lectin	2	1	Tetramer
D	*Amaranthus caudatus* agglutinin	2	1	Dimer
C	*Morus nigra* lectin	2	1	Tetramer
B	Wheat germ agglutinin WGA-2	2	1	Dimer
A	*Allium sativum* lectin	2	1	Dimer

* These lectins are hetero-octameric, each subunit consisting of an α and a β chain.

Although it was expected that the number of polymorphs occurring for a particular lectin should show a direct relationship with the total number of structures solved, not all lectins follow this suite. For example, the 30 reported crystal structures of the *Pterocarpus angolensis* lectin show presence of only two polymorphs that occur in the same space group and have only one unit cell dimension (*b*) slightly different. Peanut agglutinin has 5 polymorphs occurring in 23 structures. The largest number of structures has been reported for ConA and it also happens to be the lectin that shows the maximum number of polymorphs. This provides an unusually diverse example to compare structures of the same protein in different packing environments. It also enables to study the influence of pH and other crystallization conditions on the packing of the protein in different crystal forms and on the intermolecular interactions between symmetry related molecules. Using this information, it might be possible to gain some insight into the way by which a protein associates to form alternative contacts and crystal lattice. Hence ConA was chosen for further studies.

5.4.2. Polymorphism in ConA

At present, 51 different crystal structures of ConA are available in the protein data bank; the protein has been crystallized in presence/absence of divalent metal ions as well as various ligands wherever present. Table 5.3 enlists all the available crystal forms of ConA with the crystallization conditions and ligands. ConA binds not only carbohydrate ligands like derivatives of gluco- and mannopyranosides, but also hydrophobic molecules like porphyrin (Goel *et al.*, 2001) and oligopeptides (Jain *et al.*, 2001a; b) which mimic the carbohydrate ligands. In Table 5.4, space groups, unit cell parameters, maximum resolution and R_{factor}s of all these structures have been listed. The protein has been reported to crystallize in 4 different crystal systems, namely triclinic (*P*1), monoclinic (*P*2_1, *C*2), Orthorhombic (*P*2_12_12, *P*$2_12_12_1$, *C*222_1, *I*222, *F*222) and

cubic ($I2_13$). Table 5.5 enlists the different space groups and the statistics of ConA crystal structures occurring in each. The structures belonging to unique crystal forms, chosen for the present studies as per the parameters described in the section 5.3, have been listed in Table 5.6.

Table 5.3: Different crystallization conditions for ConA.

PDB ID	Space group	Period of crystal growth	Buffer	Precipitant/salt	Ligand	Reference
1VLN	*P*1	3 weeks	0.01 M Tris, pH 8.5	1 M $(NH_4)_2SO_4$	Ca, Mn	Kanellopoulos *et al.*, 1996b.
1CJP	$P12_11$	2-3 weeks	0.01 M Tris, pH 8.5	1 M $(NH_4)_2SO_4$	Ca, Mn, 4'-Methyl-umbelliferyl-α-D-glucopyranoside	Hamodrakas *et al*., 1997.
1CVN	$P12_11$	2 weeks	pH 9.0	20% PEG 6K	Ca, Mn, Trimannoside	Naismith and Field, 1996.
1ONA	$P12_11$	NA	0.04 M HEPES, pH 6.5	13 % PEG 5K MME	Ca, Mn, Methyl-3,6-di-O-(α-D-mannopyranosyl)-α-D-manno pyranoside	Loris *et al*., 1996
1QDO	$P12_11$	NA	pH 8.0	14% PEG 8000 and 35 mM potassium phosphate	Ca, Mn, Methyl-3-O-(α-D-Mannopyranosyl)-α-D-Mannopyranoside (Man(α1–3)Man(α1-O)Me)	Bouckaert *et al.*, 1999
1TEI	*C*121	NA	pH 6.0	15% PEG 6000	Ca, Mn, β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man- (1→6)]-Man	Moothoo and Naismith, 1998.
1APN	$P2_12_12$	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	-	Bouckaert *et al*., 1995.
1CES	$P2_12_12$	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	Zn	Bouckaert *et al*., 1996.
1ENS	$P2_12_12$	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	Co	Bouckaert *et al*., 1996.

1DQ2	$P2_12_12$	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	- (Unlocked)	Bouckaert *et al.*, 2000a.
1DQ4	$P2_12_12$	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	Mn (Unlocked)	Bouckaert *et al.*, 2000a.
1CN1	$P2_12_12$	NA	0.05 M sodium acetate, pH 5.0	1.35 M $(NH_4)_2SO_4$	-	Shoham *et al.*, 1979.
3ENR	$P2_12_12$	Several months	0.1 M sodium acetate, pH 6.15	1.8 M $(NH_4)_2SO_4$	Ca, Zn	Bouckaert *et al.*, 2000b.
1VAL	$P2_12_12$	2-3 weeks	0.01 M tris, pH 8.5	1 M $(NH_4)_2SO_4$	Ca, Mn, 4'-Nitrophenyl-α-D-glucopyranoside	Kanellopoulos *et al.*, 1996a.
1VAM	$P2_12_12$	2-3 weeks	0.01 M tris, pH 8.5	1 M $(NH_4)_2SO_4$	Ca, Mn, 4'-Nitro phenyl-α-D-Manno-pyranoside	Kanellopoulos *et al.*, 1996b.
1ENQ	$P2_12_12_1$	NA	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	Zn	Bouckaert *et al.*, 1996.
1QDC	$P2_12_12_1$	NA	0.025 M phosphat e buffer, pH 6.8	7% Jeffamine M-600, 10% PEG 8000, 50 mM ammonium acetate, 5 mM MgCl2	Ca, Cl, Mn, Man(α1–6)Man(α1-O)Me	Bouckaert *et al.*, 1999.
5CNA	$P2_12_12_1$	NA	0.05 M PIPES, pH 6.8	0.1 M $NaNO_3$, 1 mM $MnCl_2$, 1 mM $CaCl_2$, 2mg/ml NaN_3 and 0.1 M methyl-α-o-mannopyranosid e	Ca, Cl, methyl-α-o-mannopyranosid e	Naismith *et al.*, 1994.
1BXH	$P2_12_12_1$	2 weeks	0.1 M Tris pH 7.0	13.5% PEG 6000, 1.0 M LiCl	Ca, Mn, Methyl α1-2 Mannobioside	Moothoo *et al.*, 1999.
1GKB	$C222_1$	NA	0.1 M Tris-HCl pH 8.0	25%(v/v) ethanol	Ca, Mn, Mg	Kantardjieff *et al.*, 2002.
1JOJ	$C222_1$	NA	0.05 M Tris pH 9.0	$(NH_4)_2SO_4$	Ca, Mn, MYWYPY	Jain *et al.*, 2001a.
1JUI	$C222_1$	10-15 days	0.05 M Tris pH 7.5	Na_2SO_4	Ca, Mn, MYWYPYASGS	Jain *et al.*, 2001b.
1JYC	$C222_1$	10-15 days	0.05 M Tris pH 9.0	$(NH_4)_2SO_4$	Ca, Mn, RVWYPYGSYL TASGS	Jain *et al.*, 2001b.

1JYI	$C222_1$	NA	0.05 M Tris pH 9.0	1 M $(NH_4)_2SO_4$	Ca, Mn, DVFYPYPYASGS	Jain *et al.*, 2000.
1NXD	$C222_1$	10 days	0.1 M Tris HCl pH 8.0	12% (w/v) PEG 6000	Mn	López-Jaramillo *et al.*, 2004.
1QGL	$C222_1$	14 days	0.1 M citric acid pH 5.0	20% PEGI 6000	Ca, Mn, 1,3-di(n-propyloxy-α-mannopyranosyl)-carbomyl 5-methyazido-benzene	Dimick *et al.*, 1999.
1C57	*I*222	Several weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Mn	Habash *et al.*, 2000.
1CON	*I*222	2 weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Cd	Naismith *et al.*, 1993.
1I3H	*I*222	Several weeks	0.1 M citric acid pH 5.0	10% (w/v) PEG 6000	Ca, Mn, α1-2 mannobiose	Sanders *et al.*, 2001.
1NLS	*I*222	NOT REPORTED			Ca, Mn	Deacon *et al.*, 1997.
1QNY	*I*222	Several weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Mn	Habash *et al.*, 2000
1SCR	*I*222	2 weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ni, Ca	Emmerich *et al.*, 1994.
1SCS	*I*222	2 weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Co, Ca	Emmerich *et al.*, 1994.
1XQN (Neutron)	*I*222	Several weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Mn	Blakeley *et al.*, 2004.
2CNA	*I*222	NA	0.01 M sodium maleate, pH 6.80	0.1 M Na_2SO_4	Ca, Mn	Reeke *et al.*, 1975.
2CTV	*I*222	NOT REPORTED			Ca, Mn	Weisgerber and Helliwell, 1993.
2UU8	*I*222	NA	0.05 M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Ni	Ahmed *et al.*, 2007.
1DQ0	*I*222	Several months	0.1 M sodium acetate	2 M $(NH_4)_2SO_4$	- (Locked at pH 7.0)	Bouckaert *et al.*, 2000a.

1DQ1	*I*222	Several months	0.1 M sodium acetate	2 M $(NH_4)_2SO_4$	CaCa (Locked at pH 7.0)	Bouckaert *et al*., 2000a.
1DQ5	*I*222	Several months	0.1 M sodium acetate	2 M $(NH_4)_2SO_4$	MnMn (Locked at pH 7.0)	Bouckaert *et al*., 2000a.
1DQ6	*I*222	Several months	0.1 M sodium acetate	2 M $(NH_4)_2SO_4$	MnMn (Locked at pH 7.0)	Bouckaert *et al*., 2000a.
1ENR	*I*222	NA	0.2 M NaH_2PO_4	1.8 M $(NH_4)_2SO_4$	Ca, Zn	Bouckaert *et al*., 1996.
1HQW	*I*222	3-15 days	pH 5.45	2.2 mol/L sodium phosphate	Ca, Mn, YPY	Zhang *et al.*, 2001a.
1JBC	*I*222	2 weeks	0.05M Tris-acetate pH 6.5	0.1M $NaNO_3$	Ca, Mn	Parkin *et al.*, 1996.
1JW6	*I*222	3-15 days	pH 5.45	2.2 mol/L sodium phosphate	Ca, Mn, MYWYPY	Zhang *et al.*, 2001b.
2A7A	*I*222	2 months	DDW	34%(v/v) PEG 1500	Ca, Mn, Xe	Mueller-Dieckmann *et al.*, 2005.
2ENR	*I*222	Several months	0.1 M sodium acetate, pH 5.0	2 M $(NH_4)_2SO_4$	Cd	Bouckaert *et al*., 2000b.
2G4I	*I*222	2 months	DDW (pH7.0)	34%(v/v) PEG 1500	Ca, Cl, Mn, Na	Mueller-Dieckmann *et al.*, 2007.
3CNA	*I*222	NA	pH 5.9	2/4 M Na-phosphate	Ca, Mn	Hardman and Ainsworth, 1972.
1JN2	*F*222	3 weeks	0.01 M tris buffer, pH 7.4	1.25 M $(NH_4)_2SO_4$, 0.75 M NaCl	Ca, Mn, *meso*-tetrasulfonatophenyl porphyrin	Goel *et al.*, 2000.
1GIC	$I2_13$	Few days	pH 6.8	1.8 M sodium phosphate	Ca, Mn methyl-α-D-gluco-pyranoside	Harrop *et al.*, 1996.

NA: Not Available.

Table 5.4: Unit cell parameters of ConA structures.

PDB ID	Space Group	A	B	C	α^0	β^0	γ^0	Res(Å)	R_{value}*	R_{free}*
Triclinic										
1VLN	*P*1	78.8	79.3	133.3	97.1	90.2	97.5	2.40	0.205	0.265
Monoclinic										
1CJP	$P12_11$	81.6	128.7	82.2	90	118.5	90	2.78	0.182	0.216
1CVN	$P12_11$	81.7	66.7	108.3	90	97.8	90	2.30	0.204	0.255
1ONA	$P12_11$	59.8	64.8	125.9	90	93.9	90	2.35	0.221	0.282
1QDO	$P12_11$	60.4	64.5	127.0	90	93.2	90	2.80	0.164	0.257
1TEI	*C*121	176.5	122.8	124.6	90	134.2	90	2.70	0.179	0.219
Orthorhombic										
1APN	$P2_12_12$	61.4	85.5	91.5	90	90	90	2.50	0.18	0.247
1CES	$P2_12_12$	61.3	86.2	91.2	90	90	90	2.70	0.204	0.249
1ENS	$P2_12_12$	61.4	86.3	91.5	90	90	90	2.80	0.207	0.259
1DQ2	$P2_12_12$	60.5	84.2	91.0	90	90	90	2.05	0.232	0.279
1DQ4	$P2_12_12$	60.9	85.0	91.6	90	90	90	2.90	0.200	0.246
1CN1	$P2_12_12$	84.3	91.2	61.1	90	90	90	3.20	0	0
3ENR	$P2_12_12$	60.1	96.5	87.4	90	90	90	2.40	0.215	0.283
1VAL	$P2_12_12$	134.7	155.7	71.4	90	90	90	3.00	0.186	0.274
1VAM	$P2_12_12$	135.2	155.4	71.3	90	90	90	2.75	0.185	0.260
1ENQ	$P2_12_12_1$	67.2	113.0	122.1	90	90	90	2.50	0.198	0.284
1QDC	$P2_12_12_1$	69.5	117.0	120.9	90	90	90	2.00	0.175	0.211
5CNA	$P2_12_12_1$	123.7	128.6	67.2	90	90	90	2.00	0.199	0
1BXH	$P2_12_12_1$	119.7	119.7	68.9	90	90	90	2.75	0.196	0.238
1GKB	$C222_1$	118.7	101.4	112.0	90	90	90	1.56	0.180	0.206
1JOJ	$C222_1$	102.7	118.4	253.6	90	90	90	3.00	0.186	0.231
1JUI	$C222_1$	103.0	118.3	253.6	90	90	90	2.75	0.188	0.222
1JYC	$C222_1$	103.1	118.6	255.0	90	90	90	2.75	0.195	0.240
1JYI	$C222_1$	102.6	118.3	252.6	90	90	90	2.75	0.192	0.265
1NXD	$C222_1$	101.3	118.0	249.5	90	90	90	1.90	0.184	0.201
1QGL	$C222_1$	99.1	127.4	118.9	90	90	90	2.66	0.176	0.200
1C57	*I*222	89.1	87.6	63.2	90	90	90	2.40	0.270	0.301
1CON	*I*222	88.7	86.5	62.5	90	90	90	2.00	0.171	0
1I3H	*I*222	90.9	86.4	65.4	90	90	90	1.20	0.170	0.190
1NLS	*I*222	89.6	86.5	62.1	90	90	90	0.94	0.127	0.154
1QNY	*I*222	89.1	87.6	63.2	90	90	90	1.80	0.168	0.198
1SCR	*I*222	89.4	87.2	63.0	90	90	90	2.00	0.159	0
1SCS	*I*222	88.7	86.5	62.5	90	90	90	1.60	0.178	0
1XQN	*I*222	89.2	86.1	61.5	90	90	90	2.50	0.266	0.320
2CNA	*I*222	89.9	87.2	63.1	90	90	90	2.00	0	0
2CTV	*I*222	88.7	86.5	62.5	90	90	90	1.95	0.153	0
2UU8	*I*222	88.9	86.1	61.5	90	90	90	0.94	0	0
1DQ0	*I*222	63.0	87.3	88.8	90	90	90	1.70	0.181	0.204
1DQ1	*I*222	63.0	87.3	88.9	90	90	90	2.15	0.182	0.276
1DQ5	*I*222	63.2	87.4	89.3	90	90	90	2.00	0.189	0.210
1DQ6	*I*222	63.2	87.4	89.3	90	90	90	1.90	0.186	0

1ENR	*I*222	63.1	87.1	89.1	90	90	90	1.83	0.176	0
1HQW	*I*222	62.6	86.6	89.3	90	90	90	2.40	0.175	0.237
1JBC	*I*222	62.0	86.1	89.1	90	90	90	1.15	0.142	0.167
1JW6	*I*222	63.3	86.4	89.3	90	90	90	1.93	0.210	0.250
2A7A	*I*222	62.0	85.6	88.4	90	90	90	1.75	0.184	0.206
2ENR	*I*222	63.2	87.7	89.2	90	90	90	2.35	0.173	0
2G4I	*I*222	61.6	86.1	88.5	90	90	90	2.40	0.193	0.260
3CNA	*I*222	63.2	86.9	89.3	90	90	90	2.40	0	0
1JN2	*F*222	106.0	117.3	126.0	90	90	90	1.90	0.195	0.232
Cubic										
1GIC	$I2_13$	167.8	167.8	167.8	90	90	90	2.00	0.170	0

***** '0' indicates unavailability of the values.

Table 5.5: A list of ConA structures occurring in different space groups and crystal forms. The PDB ids of the structures chosen for the final studies have been typed in *bold*.

Space group	No. of forms	No. of total structures	PDB codes of structures
*P*1	1	1	**1VLN**
$P12_11$	3	1+1+2=4	**1CJP**
			1CVN
			1ONA, 1QDO
*C*121	1	1	**1TEI**
$P2_12_12$	3	6+1+2=9	1APN, 1CES, 1ENS, **1DQ2**, 1DQ4, 1CN1
			3ENR
			1VAL, **1VAM**
$P2_12_12_1$	3	2+1+1=4	1ENQ, **1QDC**
			5CNA
			1BXH
$C222_1$	3	1+5+1=7	**1GKB**
			1JOJ, 1JUI, 1JYC, 1JYI, **1NXD**
			1QGL
*I*222	1	23	1C57, 1CON, 1I3H, **1NLS**, 1QNY, 1SCR, 1SCS, 1XQN, 2CNA, 2CTV, 2UU8, 1DQ0, 1DQ1, 1DQ5, 1DQ6, 1ENR, 1HQW, 1JBC, 1JW6, 2A7A, 2ENR, 2G4I, 3CNA
*F*222	1	1	**1JN2**
$I2_13$	1	1	**1GIC**
Total	**17**	**51**	

Table 5.6. Unique crystal forms of ConA.

No.	PDB ID	Space Group	A	B	C	α^0	β^0	γ^0	Res (Å)	Rvalue	Rfree	Asymmetric unit	Vm	Vs
1.	1VLN	*P*1	78.8	79.3	133.3	97.1	90.2	97.5	2.40	0.205	0.265	Octamer	3.9	69
2.	1CJP	$P12_11$:A	81.6	128.7	82.2	90	118.4	90	2.78	0.182	0.216	Tetramer	3.7	66
3.	1CVN	$P12_11$:B	81.7	66.7	108.3	90	97.8	90	2.30	0.204	0.255	Tetramer	2.8	56
4.	1ONA	$P12_11$:C	59.8	64.8	125.9	90	93.9	90	2.35	0.221	0.282	Tetramer	2.3	47
5.	1TEI	*C*121	176.5	122.8	124.6	90	134.2	90	2.70	0.179	0.219	Octamer	2.3	47
6.	1DQ2	$P2_12_12$:A	60.5	84.2	91.0	90	90	90	2.05	0.232	0.279	Dimer	2.2	45
7.	3ENR	$P2_12_12$:B	60.1	96.5	87.4	90	90	90	2.40	0.215	0.283	Dimer	2.4	50
8.	1VAM	$P2_12_12$:C	135.2	155.4	71.3	90	90	90	2.75	0.185	0.26	Tetramer	3.6	66
9.	1QDC	$P2_12_12_1$:A	69.5	117.0	120.9	90	90	90	2.00	0.175	0.211	Tetramer	2.4	48
10.	5CNA	$P2_12_12_1$:B	123.7	128.6	67.2	90	90	90	2.00	0.199	0	Tetramer	2.6	52
11.	1BXH	$P2_12_12_1$:C	119.7	119.7	68.9	90	90	90	2.75	0.196	0.238	Tetramer	2.4	48
12.	1GKB	$C222_1$:A	118.7	101.4	112.0	90	90	90	1.56	0.180	0.206	Dimer	3.2	62
13.	1NXD	$C222_1$:B	101.3	118.0	249.5	90	90	90	1.90	0.184	0.201	Tetramer	3.6	66
14.	1QGL	$C222_1$:C	99.1	127.4	118.9	90	90	90	2.66	0.176	0.2	Dimer	3.6	66
15.	1NLS	*I*222	89.6	86.5	62.1	90	90	90	0.94	0.127	0.154	Monomer	2.3	47
16.	1JN2	*F*222	106.0	117.3	126.0	90	90	90	1.90	0.195	0.232	Monomer	3.8	67
17.	1GIC	$I2_13$	167.8	167.8	167.8	90	90	90	2.00	0.17	0	Dimer	3.8	67

From Table 5.3, it is evident that the cubic crystal form takes shortest time to grow. ConA crystallizes in this form in presence of methyl-α-D-glucopyranoside, the crystals are reported to grow within a few days (Harrop *et al.*, 1996). The other crystal forms are known to grow generally in 10 days to few months (Kanellopoulos *et al.*, 1996a, b; Jain *et al.*, 2001a, b; Moothoo *et al.*, 1999; Dimick *et al.*, 1999). Demetallized ConA takes the maximum time to grow, up to several months (Bouckaert *et al*., 1995; 1996; 2000a). From Tables 5.4 and 5.5, it can be seen that almost 50 % of the reported structures of ConA belong to the space group *I*222. ConA crystallizes in this space group at slightly acidic pH, ranging from 5.0 to 6.8. In this space group, there are two types of unit cells reported: Unit cell 1 corresponds to $a \approx 89$; $b \approx 86$ and $c \approx 62$ Å whereas Unit cell 2 has $a \approx 63$; $b \approx 87$ and $c \approx 89$ Å. It is easily noticeable that these two forms arise due to different assignments of *a*, *b* and *c* and hence can be considered to be the same unit cell. It is interesting to note that this is the single form observed in the space group *I*222 despite a large number of structures reported in this space group.

There are only one crystal form with a single PDB entry in each reported for the space groups *P*1, *C*121, *F*222 and $I2_13$. The triclinic (*P*1) form and one of the orthorhombic forms ($P2_12_12$:C) have one of the unit cell lengths closer (*c* of *P*1 (133.3 Å) ≈ *a* of $P2_12_12$:C (135.2 Å)); whereas the third unit cell length (*b* of $P2_12_12$:C = 155.4 Å) is approximately double that of the *b* of *P*1 (79.3 Å).

Demetallized ConA crystallizes in the orthorhombic forms $P2_12_12$:A and B. These forms $P2_12_12$:A (1DQ2) and $P2_12_12$:B (3ENR) have almost similar unit cell dimensions: unit cell length *a* of both the forms is almost equal (~60 Å). Unit cell lengths *b* and *c* are, however, may seem as flipped. Since axis *b* is a screw axis and *c* is not, they have been considered as non-isomorphous unit cells.

In the case of the orthorhombic space group $C222_1$, crystal forms A (1GKB) and B (1NXD) have two of their unit cell lengths similar (*a* of A ≈ *b* of B ≈ 118 Å and *b* of A ≈

a of B ≈ 101 Å). The third unit cell length, *c* of form B (~249 Å) is approximately double of the *c* of form A (~112 Å). This might be the reason that form A has a dimer whereas form B has a tetramer in the asymmetric unit.

*C*222$_1$ forms A (1GKB) and C (1QGL) have similar unit cell lengths *a* and *c*, respectively, (~119 Å) as well as *b* and *a*, respectively, (~100 Å). The third unit cell length (*c* of form A (112.0 Å) and *b* of form C (127.4 Å) differ considerably. Although both the forms have dimers in the asymmetric unit, the dimer in the form C is made of two monomers belonging to different biological units which are connected to each other via a ligand.

5.4.1. Generation of symmetry related molecules

Although ConA exists in solution as tetramer at pH 7.0 and exclusively as dimer in the pH range 5.0-5.6 (Dimick *et al.*, 1999), it was observed that regardless of pH of crystallization solution, it crystallizes in the tetrameric form, although the asymmetric unit may comprise of a monomer, dimer or tetramer. To study the interactions between the neighboring molecules in the crystal lattice, it was necessary to use the tetrameric structure as a starting point. Eight of the seventeen different crystal forms of ConA have the functional tetramer in the asymmetric unit. This molecule was used to create the symmetry related molecules. Two structures, namely 1VLN and 1TEI, have octamers in their asymmetric units. These octamers were used to create packing diagrams and study interactions. For other molecules, which have monomers or dimers in the asymmetric units, the functional molecule was generated using the symmetry transformations as described below:

1DQ2 and 3ENR: These structures have a dimer in the asymmetric unit. Another dimer was generated using the symmetry operation (-X,-Y,Z) on the original molecule. These two dimers were combined to get the functional tetramer.

1GKB: This molecule also has a dimer in the asymmetric unit. Another dimer was generated using the symmetry operation (-X,Y,1/2-Z). The tetramer was generated combining the two dimers.

1NXD: This structure also has a dimer in the asymmetric unit. Another dimer was generated using the symmetry operation (-X+1,Y,1/2-Z). The tetramer was generated by combining the two dimers.

1QGL: Although, this crystal form also has a dimer in the asymmetric unit, it is made of two monomers from two different biological units connected through the ligand. These two monomers were separated first, one with the ligand and another without it, and the following symmetry operations were used to generate the four chains of the functional tetramer:

Chain A: Chain A of 1QGL

Chain C: Chain A of 1QGL, by applying symmetry transformation -X,Y,1/2-Z

Chain B: Chain B of 1QGL, by applying symmetry transformation X,-Y,-Z-1

Chain D: Chain B of 1QGL, by applying symmetry transformation -X-1/2,1/2-Y,Z+1/2

1NLS: This crystal form has a monomer in the asymmetric unit. The four chains of the functional tetramer were generated by applying the following symmetry transformations on the original molecule:

Chain A: Original molecule

Chain B: -X+1,Y,-Z applied to the original molecule.

Chain C: X,-Y+1,-Z applied to the original molecule.

Chain D: -X+1,-Y+1,Z applied to the original molecule.

1JN2: This structure also has a monomer in the asymmetric unit. The biological molecule was generated with following symmetry operations:

Chain A: Original molecule

Chain B: -X,Y,-Z applied to the original molecule.

Chain C: -X,-Y,Z applied to the original molecule.

Chain D: X,-Y,-Z applied to the original molecule.

1GIC: This structure has a dimer in the asymmetric unit and the tetramer was generated using the original molecule and its symmetry equivalent (Z,X-1,Y).

The symmetry operations used to generate the symmetry related molecules for each of the structures listed in the Table 5.6 are given in Table 5.7.

Table 5.7. Symmetry operations used to generate the symmetry related molecules of various ConA structures.

Mol. NO.	SYM
1VLN (*P*1)	
1	X,Y-1,Z
2	X-1,Y,Z
3	X+1,Y,Z
4	X,Y+1,Z
1CJP (*P*12$_1$1:A)	
1	X-1,Y,Z-1
2	X,Y,Z-1
3	-X,Y-1/2,-Z
4	-X+1,Y-1/2,-Z
5	X-1,Y,Z
6	-X,Y+1/2,-Z
7	X+1,Y,Z
8	-X+1,Y+1/2,-Z
9	-X+1,Y-1/2,-Z+1
10	X,Y,Z+1
11	X+1,Y,Z+1
12	-X+1,Y+1/2,-Z+1
1CVN (*P*12$_1$1:B)	
1	X,Y-1,Z
2	-X,Y-1/2,-Z
3	-X+1,Y-1/2,-Z
4	X-1,Y,Z
5	-X,Y+1/2,-Z
6	X+1,Y,Z
7	-X+1,Y+1/2,-Z
8	X,Y+1,Z
9	-X,Y-1/2,-Z+1
10	-X,Y-1/2,-Z+1
11	-X,Y+1/2,-Z+1
12	-X+1,Y+1/2,-Z+1
1ONA (*P*12$_1$1:C)	
1	X,Y-1,Z
2	-X,Y-1/2,-Z
3	-X,Y+1/2,-Z
4	-X,Y+1.5,-Z
5	-X,Y-1/2,-Z+1
6	-X,Y+1/2,-Z+1
1TEI (*C*121)	
1	-X,Y,-Z-1
2	-X+1,Y,-Z+1
3	1/2-X,Y-1/2,-Z
4	1/2-X,1/2+Y,-Z

Mol. NO.	SYM
1DQ2 (tetramer) (*P*2$_1$2$_1$2:A)	
1	-X-1/2, Y-1/2,-Z
2	1/2-X, Y-1/2,-Z
3	X-1,Y,Z
4	-X-1/2,1/2+Y,-Z
5	1/2-X,1/2+Y,-Z
6	X+1,Y,Z
7	-X-1/2,Y-1/2,-Z+1
8	1/2-X,Y-1/2,-Z+1
9	-X-1/2,1/2+Y,-Z+1
10	1/2-X,1/2+Y,-Z+1
3ENR (tetramer) (*P*2$_1$2$_1$2:B)	
1	X,Y,Z-1
2	-X-1/2,Y-1/2,-Z
3	X,Y-1,Z
4	1/2-X,Y-1/2,-Z
5	-X-1,-Y,Z
6	-X-1/2,1/2+Y,-Z
7	1/2-X,1/2+Y,-Z
8	X+1,Y,Z
9	X,Y+1,Z
10	-X-1/2,Y-1/2,-Z+1
11	1/2-X,Y-1/2,-Z+1
12	-X-1/2,1/2+Y,-Z+1
13	X,Y,Z+1
14	1/2-X,1/2+Y,-Z+1
1VAM (*P*2$_1$2$_1$2:C)	
1	X,Y,Z-1
2	X-1/2,1/2-Y,-Z+1
3	X,Y,Z+1
4	1/2+X,1/2-Y,-Z+1
5	-X,-Y,Z
6	-X,-Y+1,Z
7	1/2-X,Y-1/2,-Z+1
8	1/2-X,1/2+Y,-Z+1
1QDC (*P*2$_1$2$_1$2$_1$:A)	
1	-X,Y-1/2,1/2-Z
2	X-1,Y,Z
3	-X,1/2+Y,1/2-Z
4	X+1,Y,Z
5	X-1/2,1/2-Y,-Z
6	1/2+X,1/2-Y,-Z
7	X-1/2,1/2-Y,-Z+1
8	1/2+X,1/2-Y,-Z+1

Table 5.7. ***Continued***

Mol. NO.	SYM
5CNA (*P*$2_12_12_1$:B)	
1	X,Y,Z-1
2	1/2-X,-Y+1,Z-1/2
3	1/2-X,-Y+2,Z-1/2
4	1/2-X,-Y+1, Z-1/2
5	1/2-X,-Y+2,1/2+Z
6	X,Y,Z+1
1BXH (*P*$2_12_12_1$:C)	
1	X,Y,Z-1
2	1/2-X,-Y,Z-1/2
3	1/2-X,-Y+1,Z-1/2
4	-X,Y-1/2,1/2-Z
5	X-1/2,1/2-Y,-Z
2	1/2-X,-Y,1/2+Z
4	-X,1/2+Y,1/2-Z
6	1/2+X,1/2-Y,-Z
8	1/2-X,-Y+1,1/2+Z
10	-X,Y-1/2,-Z+1.5
12	X-1/2,1/2-Y,-Z+1
14	X,Y,Z+1
16	-X,1/2+Y,-Z+1.5
18	1/2+X,1/2-Y,-Z+1
1GKB (*C*222_1:A)	
1	-X,-Y,Z-1/2
2	-X,-Y,1/2+Z
3	-X,Y,1/2-Z
4	-X-1,Y-1,1/2-Z
5	-X,Y-1,1/2-Z
6	-X-1,Y,1/2-Z
1NXD (*C*222_1:B)	
1	-X-1,Y-1,1/2-Z
2	-X,Y-1,1/2-Z
3	-X-1,Y,1/2-Z
4	-X,Y,1/2-Z
1GIC (*I*2_13)	
1	1/2-Z,-X+1,Y-1/2
2	Z,X,Y
3	Y,Z,X
4	-Y+1,1/2+Z,1/2-X
5	1/2-Y,-Z+1,1/2+X
6	Z-1/2,1/2-X,-Y+1
7	-Z+1,1/2+X,-Y+1/2
8	Y-1/2,1.5-Z,-X+1

Mol. NO.	SYM
1QGL (*C*222_1:C)	
1	-X,-Y+1,Z-1/2
2	-X-1,Y,1/2-Z
3	-X,Y,1/2-Z
4	X,-Y,-Z
5	-X,-Y+1,1/2+Z
6	X-1,-Y,-Z-1
7	X,-Y,-Z-1
8	-X-1,Y-1,1/2-Z
9	-X,Y-1,1/2-Z
10	X-1,-Y,-Z
1NLS (*I*222)	
1	X,Y,Z-1
3	X,Y,Z+1
5	X-1,-Y-1,-Z-1
7	X,-Y-1,-Z-1
9	X-1,-Y,-Z-1
11	X,-Y,-Z-1
13	X-1,-Y-1,-Z
15	X,-Y-1,-Z
17	X-1,-Y,-Z
19	X,-Y,-Z
1JN2 (*F*222)	
1	X,Y-1/2,Z-1/2
2	X,1/2+Y,Z-1/2
3	X,Y-1/2,1/2+Z
4	X,1/2-Y,1/2-Z
5	-X-1/2,-Y,Z-1/2
6	1/2+X,Y,Z-1/2
7	X-1/2,Y,1/2+Z
8	1/2+X,Y,1/2+Z
9	+X-1/2,Y-1/2,Z
10	1/2+X,Y-1/2,Z
11	X-1/2,Y-1/2,Z
12	1/2+X,1/2+Y,Z

5.4.2. Crystal packing interactions in different structures of ConA

All the unique interactions between a ConA molecule and its symmetry equivalents have been listed in Table 5.8 along with the crystallization pH and the precipitant.

Table 5.8. Interactions between various ConA molecules and their symmetry equivalents.

Original molecule	Symmetry equivalent	Distance (Å)	Crystallization condition and ligand
1VLN (*P*1)			pH 8.5 1 M $(NH_4)_2SO_4$ Ca, Mn
OH B:100 TYR	NE2 A:205 HIS	2.569	
NH2 B:228 ARG	OG A:204 SER	3.021	
1CJP (*P*12$_1$1:A)			pH 8.5 1 M $(NH_4)_2SO_4$ Ca, Mn, 4'-Methylumbelliferyl-α-D-glucopyranoside
Interface 1			
OD1 D:237 ASN	NH2 A:33 ARG	3.345	
ND2 D:237 ASN	NH1 A:228 ARG	3.186	
ND2 D:237 ASN	NH2 A:228 ARG	2.949	
Interface 2			
OH D:100 TYR	OG1 C:15 THR	2.896	
OD2 D:16 ASP	O2 C:240 MUG	3.067	
Interface 3			
O2 D:240 MUG	OH B:12 TYR	3.034	
OG D:168 SER	O B:19 ASP	2.791	
OG1 D:226 THR	O B:15 THR	3.223	
1CVN (*P*12$_1$1:B)			pH 9.0 20% PEG 6K Ca, Mn, Trimannoside
Interface 1			
OG B:161 SER	O A:161 SER	3.227	
O C:161 SER	OG D:161 SER	2.949	
Interface 2			
OG C:168 SER	OG D:220 SER	2.948	
O2 K:240 MAN	OG D:223 SER	2.994	
OG C:223 SER	OD2 D:145 ASP	2.728	
OG C:223 SER	O D:168 SER	3.059	
OG1 C:226 THR	OG D:223 SER	2.951	
O C:226 THR	OG D:223 SER	3.316	
O C:224 GLY	N D:224 GLY	2.836	
Interface 3			
NH1 D:33 ARG	O A:16 ASP	3.056	
O4 L:242 MAN	O6 I:242 MAN	3.249	
O D:16 ASP	O A:15 THR	3.410	

O D:16 ASP	OG1 A:15 THR	2.936	
O D:16 ASP	N A:16 ASP	3.451	
Interface 4			
OD2 B:16 ASP	NZ D:36 LYS	3.081	
OD1 B:237 ASN	OG D:21 SER	2.772	
1ONA (*P*12$_1$1:C)			pH 6.5 13% PEG 5K MME Ca, Mn, Methyl-3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside
Interface 1			
No hydrogen-bond like interactions			
Interface 2			
OG A:223 SER	O2 B:240 MAN	2.492	
O A:168 SER	OG B:223 SER	2.716	
OD2 A:145 ASP	OG B:223 SER	2.639	
1TEI (*C*121)			pH 8.0, 14% PEG 8000 Ca, Mn, Methyl-3-O-(α-D-Mannopyranosyl)-α-D-Mannopyranoside (Man(α1-3)Man(α1-***O***)Me)
Interface 1 (between two tetramers in the asymmetric unit)			
O A:221 ILE	N G:1 ALA	2.870	
OD2 B:218 ASP	NH1 G:158 ARG	3.415	
Interface 2			
O A:69 ASN	NZ E:46 LYS	3.216	
OD1 A:71 ASP	OE1 E:43 GLN	3.298	
OD2 A:71 ASP	N E:71 ASP	3.379	
O A:202 PRO	NZ H:135 LYS	3.054	
OD1 G:83 ASN	OG A:21 SER	2.862	
OE1 G:87 GLU	O4 A:244 NAG	2.939	
OE1 G:87 GLU	NE2 A:205	3.394	
N G:87 GLU	O3 A:244 NAG	3.170	
NZ H:135 LYS	O A:202 PRO	3.054	
O3 A2:240 NAG	O1 A:242 MAN	2.725	
O1 A2:242 MAN	O3 A:240 NAG	2.725	
O E:21 SER	ND2 C:83 ASN	3.039	
O3 E2:244 NAG	O C:83 ASN	3.006	
NZ E:46 LYS	O A:69 ASN	3.216	
OE1 E:43 GLN	OD1 A:71 ASP	3.298	
Interface 3			
NZ G:46 LYS	OG1 B:15 THR	2.514	
NZ G:46 LYS	O4 B:243 MAN	3.436	
OE1 G:43 GLN	O7 B:244 NAG	3.428	
1DQ2 (*P*2$_1$2$_1$2:A)			pH 5.0 2 M $(NH_4)_2SO_4$ Unlocked at pH 5.0
Interface 1			
O C:82 ASP	OG D:204 SER	3.098	
OD1 C:82 ASP	NE2 D:205 HIS	2.915	
Interface 2			
OD1 B:203 ASP	ND2 A:83 ASN	3.318	

N B:204 SER	OD1 A:83 ASN	2.546	
Interface 3			
OD2 C:218 ASP	N B:1 ALA	2.964	
OD2 D:136 ASP	OG B:223 SER	2.737	
NZ D:138 LYS	OG B:223 SER	3.455	
O D:1 ALA	OD2 A:218 ASP	3.039	
3ENR (*P*2_12_12:B)			
Interface 1			
OD2 C:71 ASP	OD1 A:16 ASP	2.582	
OD2 C:71 ASP	OD2 A:16 ASP	2.681	
OD2 C:71 ASP	NH1 A:228 ARG	2.972	pH 6.15 1.8 M $(NH_4)_2SO_4$ Ca, Zn
OD2 C:71 ASP	NH2 A:228 ARG	2.764	
OD1 C:71 ASP	OD1 A:16 ASP	3.113	
OD1 B:83 ASN	ND1 A:205 HIS	3.203	
Interface 2			
N A:1 ALA	OD1 B:2 ASP	3.230	
OG A:223 SER	OD2 A:136 ASP	2.814	
1VAM (*P*2_12_12:C)			
Interface 1			
O8 D:240 PNA	NE2 B:132 GLN	2.809	
O7 D:240 PNA	NE2 B:132 GLN	3.188	
OG D:168 SER	O C:69 ASN	3.167	
OD1 D:16 ASP	OD1 B:151 ASP	2.976	
OG1 D:15 THR	OD2 B:151 ASP	2.876	
OH D:100 TYR	OG A:185 SER	3.139	
Interface 2			pH 8.5 1 M $(NH_4)_2SO_4$ Ca, Mn, 4'-Nitrophenyl-α-D- Mannopyranoside
OH C:100 TYR	NZ A:138 LYS	2.786	
OD1 C:16 ASP	OD2 B:218 ASP	2.977	
OD2 C:16 ASP	OD2 B:218 ASP	2.893	
NE C:228 ARG	OD2 B:218 ASP	3.148	
NH2 C:228 ARG	OD2 B:218 ASP	2.877	
NH1 C:228 ARG	OG B:220 SER	3.324	
OD1 C:237 ASN	OG B:223 SER	2.896	
Interface 3			
NH2 A:228 ARG	OD2 B:16 ASP	2.694	
NH1 A:228 ARG	OD2 B:16 ASP	3.262	
OD2 A:16 ASP	OD2 B:16 ASP	3.004	
1QDC (*P*$2_12_12_1$:A)			
Interface 1			pH 6.8, 7% Jeffamine M-600, 10% PEG 8000 Ca, Cl, Mn, Man(α1-6)Man(α1-O)Me
OH D:77 TYR	OD2 A:80 ASP	2.860	
OH D:77 TYR	OD1 A:83 ASN	3.160	
OD2 B:78 ASP	NZ C:36 LYS	3.087	
OH B:77 TYR	OH C:77 TYR	2.780	

Interface 2			
No hydrogen-bond like interactions			
Interface 3			
OD2 B:145 ASP	OG A:223 SER	2.539	
NH2 B:158 ARG	OG A:223 SER	2.889	
O B:168 SER	OG A:223 SER	3.284	
N B:224 GLY	O A:224 GLY	2.620	
OG B:223 SER	O A:226 THR	3.159	
OG B:220 SER	OG A:168 SER	2.790	
Interface 4			
OG C:168 SER	OD2 D:136 ASP	3.188	
OG C:168 SER	NZ D:138 LYS	2.491	
OD1 C:162 ASN	N C:1 ALA	2.824	
OG C:223 SER	NZ D:135 LYS	3.422	
ND2 C:237 ASN	NH1 D:158 ARG	3.132	
NH2 C:33 ARG	NH1 D:158 ARG	3.028	
NH2 C:33 ARG	NH2 D:158 ARG	2.547	
OD2 D:136 ASP	NZ D:138 LYS	2.858	
5CNA ($P2_12_12_1$:B)			
Interface 1			
No hydrogen-bond like interactions			
Interface 2			
OG C:204 SER	O A:83 ASN	2.914	pH 6.8 0.1 M $NaNO_3$ Ca, Cl, methyl-α-o-mannopyranoside
OH C:100 TYR	OD2 A:80 ASP	3.068	
OH C:100 TYR	OD2 A:82 ASP	2.961	
NE2 C:205 HIS	O A:82 ASP	2.742	
N C:204 SER	OD1 A:83 ASN	3.152	
Interface 3			
OH D:100 TYR	OG B:184 SER	3.404	
1BXH ($P2_12_12_1$:C)			
Interface 1			
O A:162 ASN	OG B:161 SER	3.181	pH 7.0 13.5% PEG 6000, 1.0 M LiCl Ca, Mn, Methyl α1-2 Mannobioside
OG D:161 SER	O C:162 ASN	3.428	
Interface 2			
O C:1 ALA	OH C:100 TYR	2.984	
Interface 3			
OG A:223 SER	O B:168 SER	2.846	
OG A:223 SER	OD2 B:145 ASP	3.062	
OG1 A:226 THR	OG B:223 SER	2.842	
OG A:168 SER	OG B:220 SER	2.478	

Interface 4			
OD1 D:83 ASN	OH A:77 TYR	2.683	
ND2 D:83 ASN	OH A:77 TYR	3.170	
ND2 D:83 ASN	O A:78 ASP	3.378	
NZ D:30 LYS	O A:83 ASN	2.190	
1GKB (*C*222$_1$:A)			
Interface 1			
NE2 B:205 HIS	OH A:100 TYR	2.789	pH 8.0 25%(v/v) ethanol Ca, Mn, Mg
OG B:21 SER	NZ A:101 LYS	3.393	
OG B:21 SER	OD2 A:203 ASP	3.420	
Interface 2			
OD2 A:218 ASP	NH1 C:90 ARG	2.858	
OD2 A:218 ASP	NH2 C:90 ARG	3.191	
N A:1 ALA	OH C:176 TYR	2.857	
OG A:220 SER	NE2 C:143 GLN	3.175	
1NXD (*C*222$_1$:B)			
Interface 1			pH 8.0 12% (w/v) PEG 6000 Mn
NE2 A:205 HIS	OH B:100 TYR	2.668	
OG A:21 SER	NZ B:101 LYS	3.066	
1QGL (*C*222$_1$:C)			
Interface 1			
OE1 A:166 GLN	OG B:164 SER	2.852	
Interface 2			pH 5.0 20% PEG 6000 Ca, Mn, 1,3-di(n-propyloxy-α-mannopyranosyl)-carbomyl 5-methyazido-benzene
ND2 C:14 ASN	OD2 A:16 ASP	3.071	
NE C:228 ARG	OD2 A:16 ASP	3.153	
NH2 C:228 ARG	OD2 A:16 ASP	2.893	
N C:16 ASP	OD1 A:16 ASP	3.042	
N C:15 THR	OD1 A:16 ASP	3.448	
O C:13 PRO	OG1 A:15 THR	2.592	
OD1 C:16 ASP	N A:16 ASP	3.042	
1NLS (*I*222)			
Interface 1			
No hydrogen-bond like interactions			
Interface 2			
OH B:12 TYR	O D:184 SER	2.762	Ca, Mn
OD2 B:16 ASP	OD1 D:118 ASN	2.971	
OG,A B:21 SER	NZ A:135 LYS	3.095	
N B:204 SER	OD1 D:83 ASN	2.841	
OG B:204 SER	OD1 D:83 ASN	3.301	
OH B:100 TYR	NE1 D:182 TRP	2.911	

1JN2 (*F*222)			pH 7.4 1.25 M $(NH_4)_2SO_4$ 0.75 M NaCl Ca, Mn, *meso*-tetrasulfonatophenyl porphyrin
Interface 1			
OD1 A:237 ASN	OD1 A:237 ASN	2.700	
ND2 A:237 ASN	OD1 A:237 ASN	2.779	
ND2 A:237 ASN	ND2 A:237 ASN	2.976	
O5 A:1001 SFP	NH1 A:228 ARG	3.280	
Interface 2			
No hydrogen-bond like interactions			
Interface 3			
NH2 A:158 ARG	OE1 C:166 GLN	3.350	
NH2 A:158 ARG	NE2 C:166 GLN	3.488	
NH1 A:158 ARG	OE1 C:166 GLN	3.484	
OG A:168 SER	O C:98 GLY	3.442	
OG A:168 SER	OG C:168 SER	3.077	
1GIC ($I2_13$)			pH 6.8 1.8 M sodium phosphate Ca, Mn, methyl-α-D-glucopyranoside.
Interface 1			
OH C:22 TYR	ND1 C:205 HIS	3.289	
OG1 C:37 THR	OH C:100 TYR	2.591	

5.4.3. Comparison of packing interactions in different ConA molecules

1VLN and 1VAM:

Although these two forms have one of the unit cell lengths similar, they do not share same packing or interactions in any of the directions.

1DQ2 and 3ENR:

These two molecules have same unit cell length *a* and have similar contacts and packing observed in *a*-direction, however, in the other two directions, the packing interactions are entirely different, except for the interaction between Asn 83 and His 205. In 1DQ2, the interfaces 1 and 2 are almost similar; however they involve slightly different residues. Involvement of Asp and Arg residues at the interface 1 of 3ENR could be due to its crystallization pH closer to neutral pH.

1GKB and 1NXD:

These two molecules share the interface 1, however, the interface 2 which is observed in 1GKB is absent in 1NXD. Molecules in the *c*-direction are not interacting in 1NXD, but have interactions in 1GKB. Thus, because of loose packing, the unit cell length in the *c*-direction is larger in 1NXD. Crystal form 1GKB is reported to crystallize in presence of ethanol. The reduction in the dielectric constant due to the presence of ethanol might be facilitating electrostatic interaction between Arg 90 and Asp 218.

The crystal packing and interactions between symmetry related molecules of all the unique crystal forms of ConA enlisted in the Table 5.6 have been shown sequentially in Figures 5.1 to 5.17.

1VLN

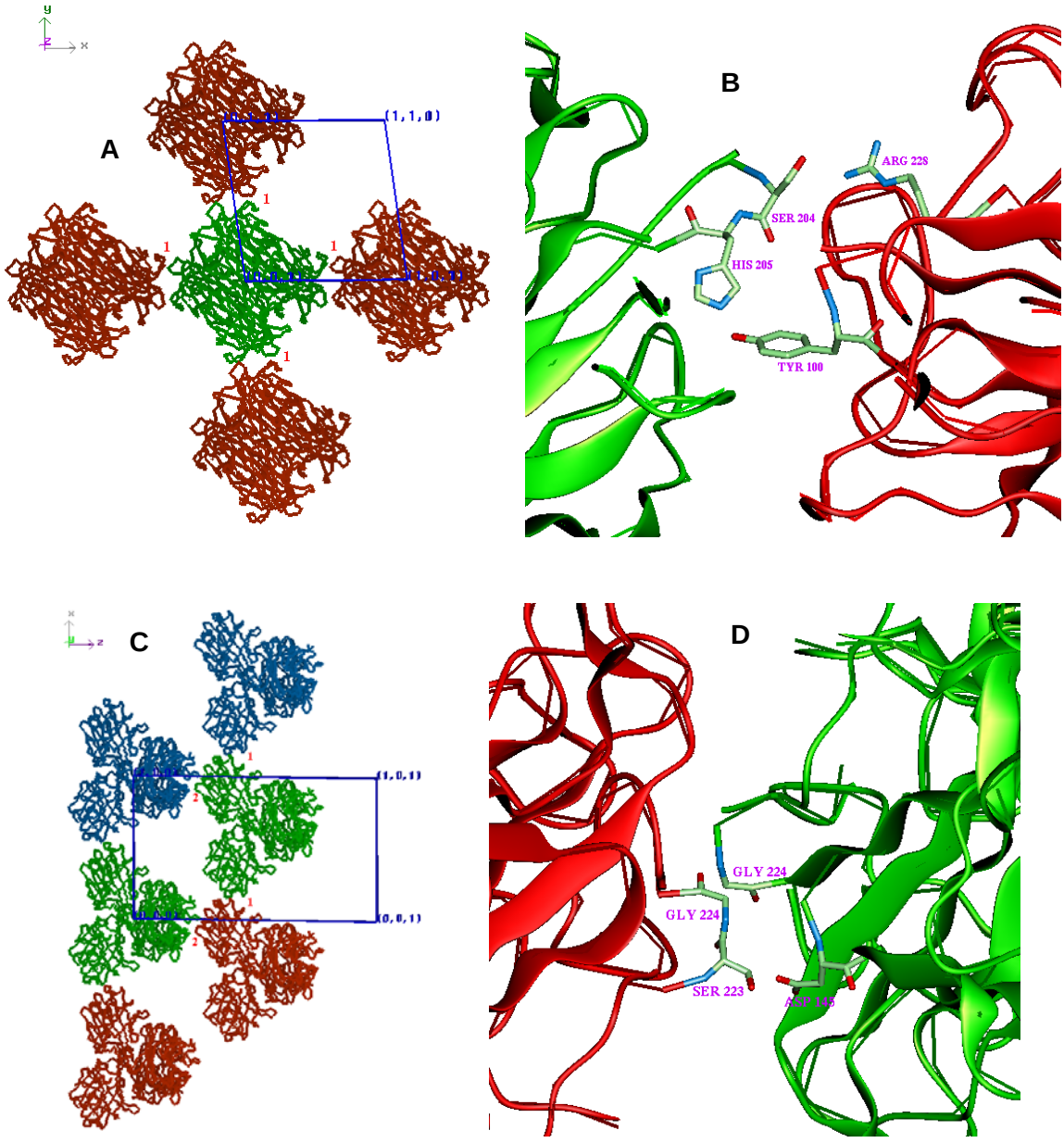


Fig. 5.1. (A) Packing of 1VLN (tetramer) as viewed in the *ab* plane. All the neighboring molecules have similar interface with the original molecule. **(B)** Packing interactions between 1VLN and symmetry related molecule at the interface 1. **(C)** Packing of 1VLN (octamer) as viewed in the *ac* plane. Both the interfaces, namely 1 and 2 are visible. Interactions at the interface 2 have been shown in **(D)**.

1CJP

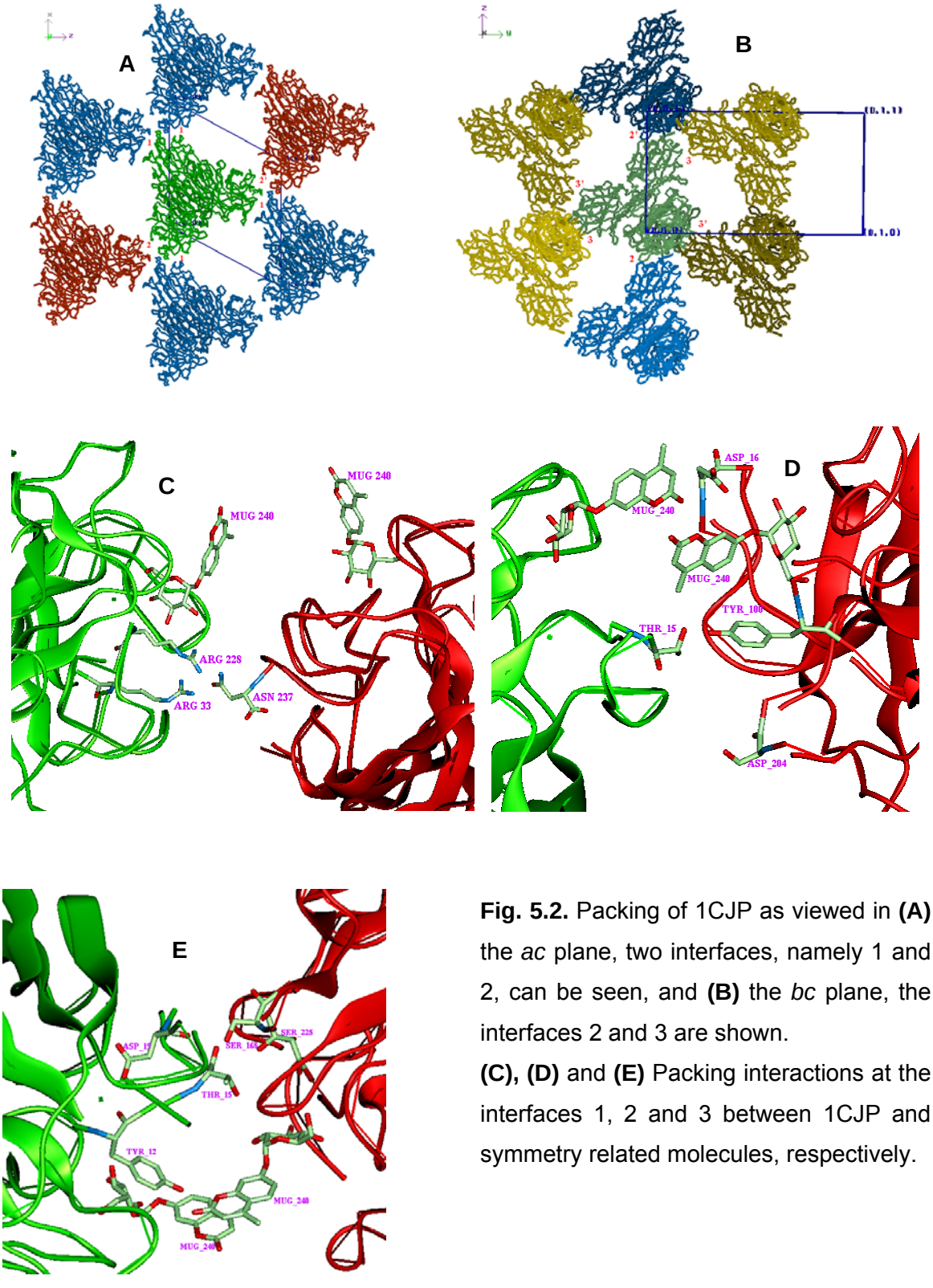


Fig. 5.2. Packing of 1CJP as viewed in **(A)** the *ac* plane, two interfaces, namely 1 and 2, can be seen, and **(B)** the *bc* plane, the interfaces 2 and 3 are shown.

(C), (D) and **(E)** Packing interactions at the interfaces 1, 2 and 3 between 1CJP and symmetry related molecules, respectively.

1CVN

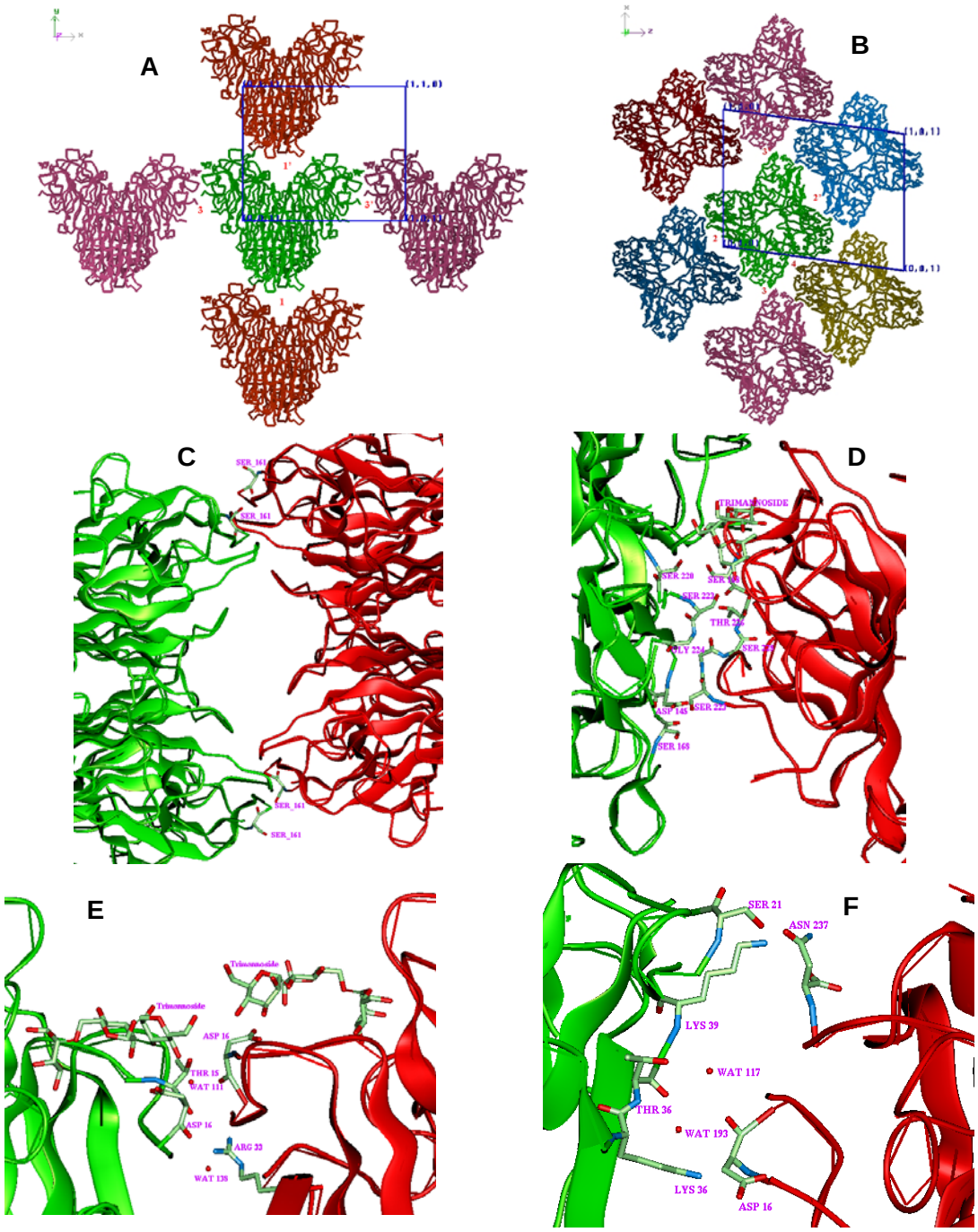


Fig. 5.2. (A) and **(B)** Packing of 1CVN as viewed in the *ab* and *ac* planes, respectively. Different interfaces have been labeled with numbers.

(C), (D), (E) and **(F).** Packing interactions between 1CVN and symmetry related molecule at the interfaces 1, 2, 3 and 4, respectively.

1ONA

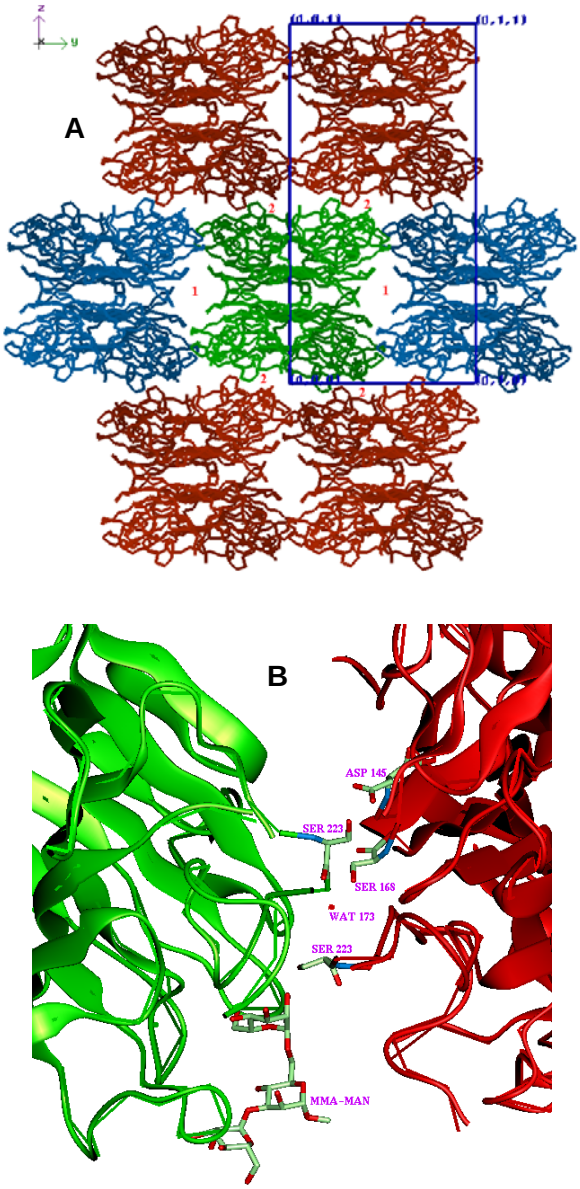


Fig. 5.4. (A) Packing of 1ONA as seen in the *bc* plane. Interfaces 1 and 2 are visible.

(B) Packing interactions at the interface 2 between 1ONA and symmetry related molecules, respectively. Interface 1 does not show any hydrogen-bond like interaction.

1TEI

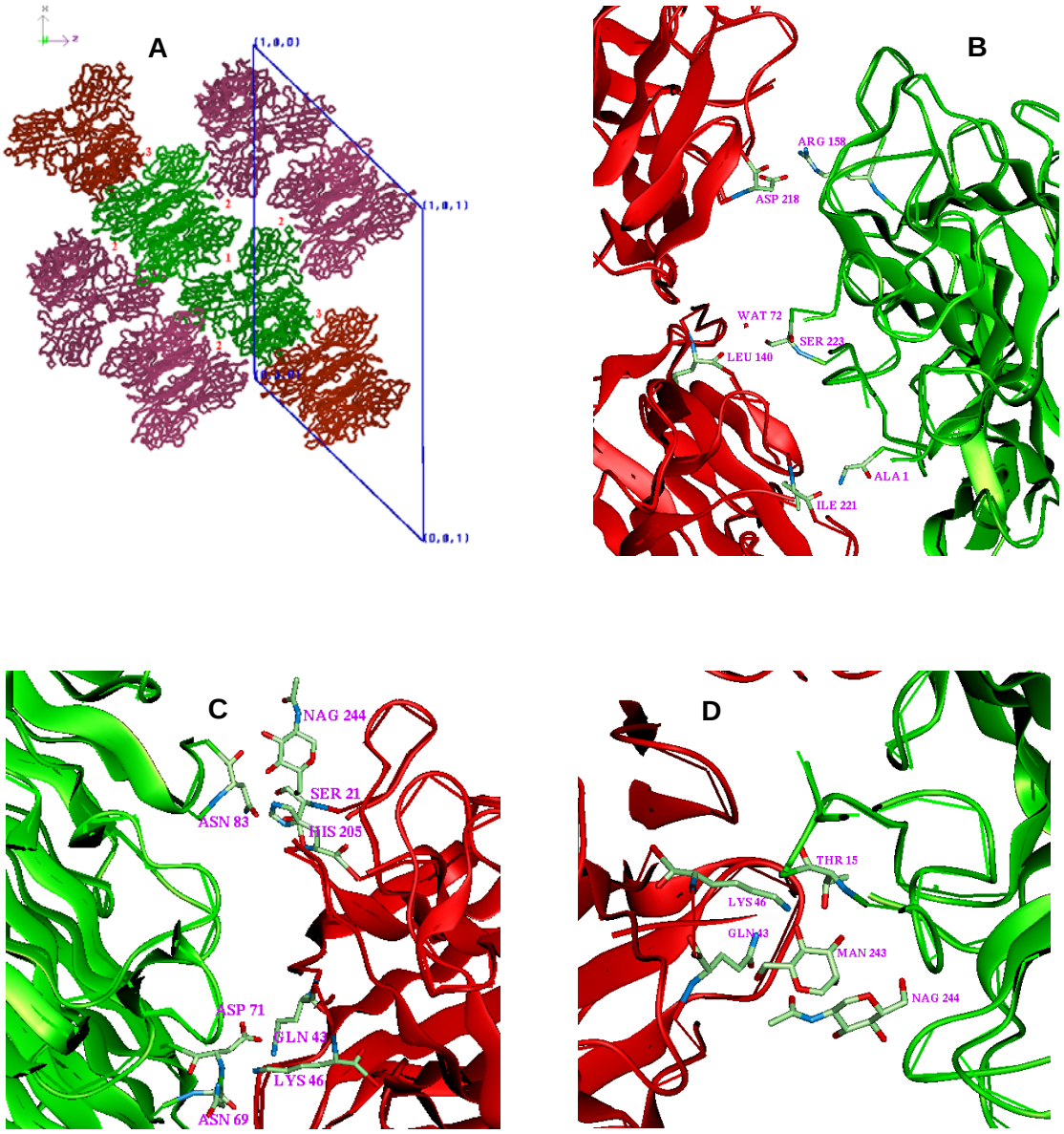


Fig. 5.5. (A) Packing of 1TEI octamer in the *ac* plane. Interfaces 1, 2 and 3 are shown. Interface 1 is between the two tetramers in the asymmetric unit. **(B)** Packing interactions between the two tetramers at the interface 1. **(C)** and **(D)** Packing interactions between 1TEI and the symmetry related molecules at the interfaces 2 and 3, respectively.

1DQ2

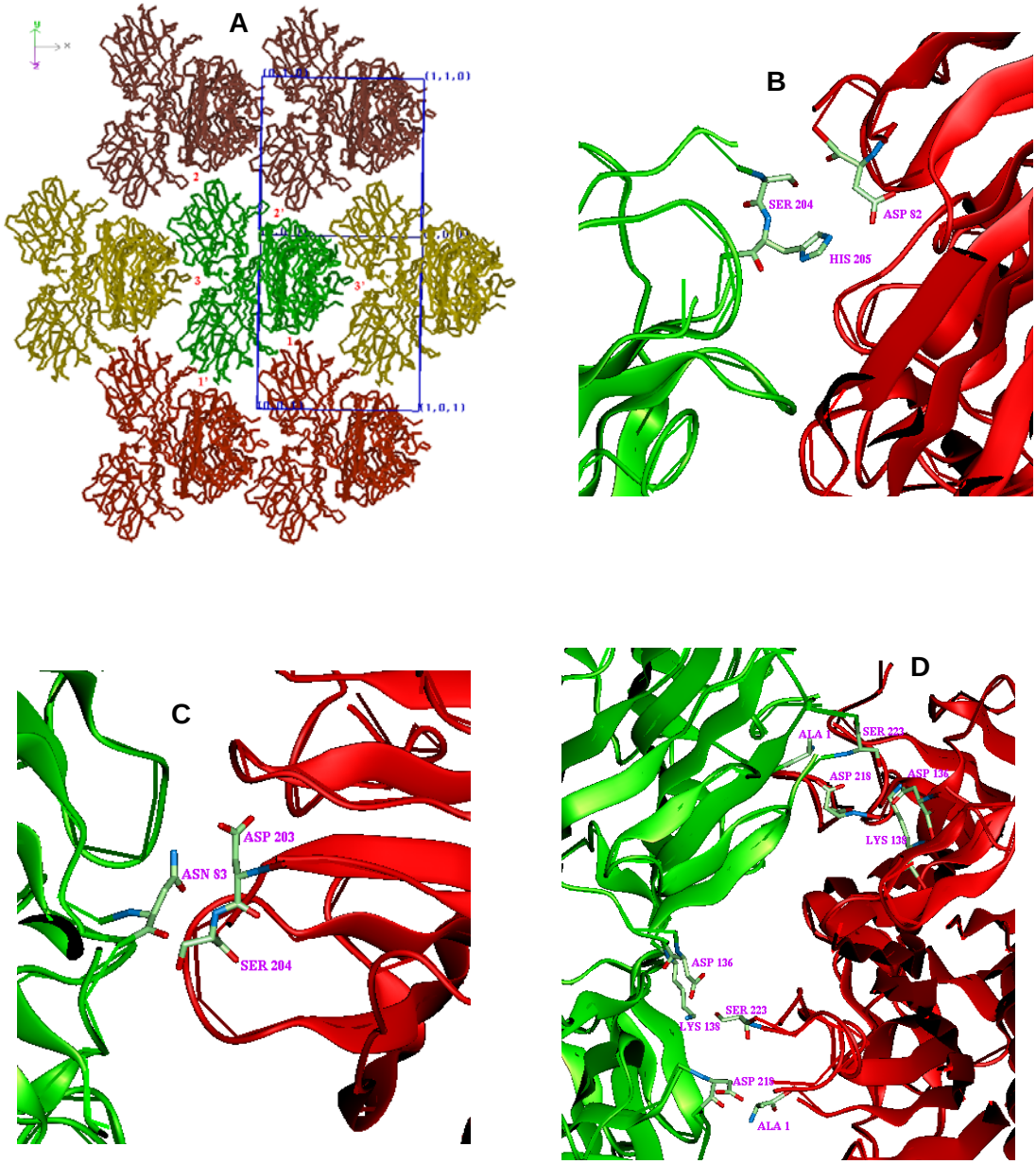


Fig. 5.6. (A) Packing of 1DQ2 in the *c*-direction. All the interacting interfaces, namely 1, 2 and 3 are shown. **(B), (C)** and **(D)** show packing interactions between 1DQ2 and symmetry related molecules at the interfaces 1, 2 and 3, respectively.

3ENR

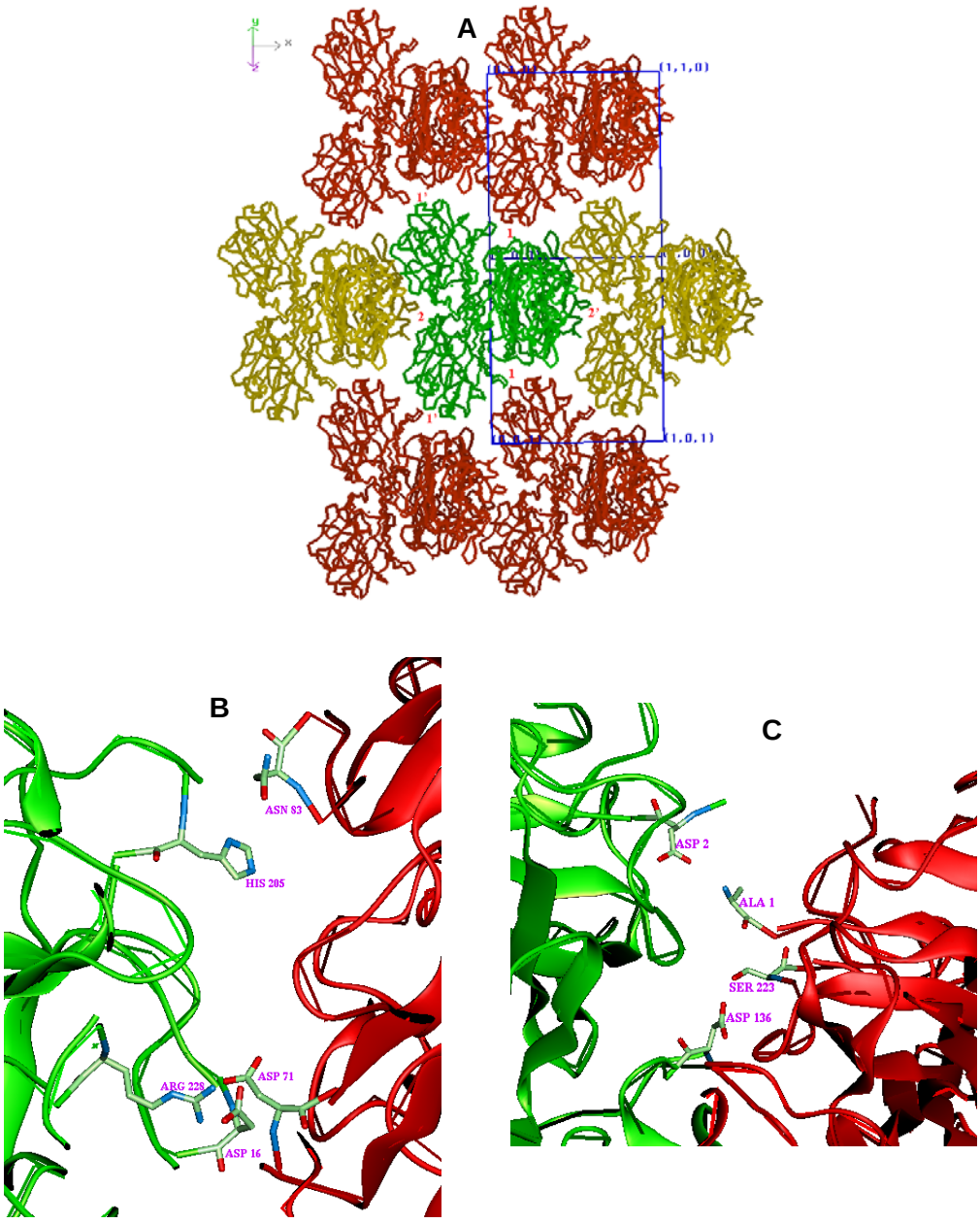


Fig. 5.7. (A) Packing of 3ENR in the *a*-direction. Both the interfaces, namely 1 and 2 are seen.

(B) and (C) Packing interactions between 3ENR and symmetry related molecules at the interfaces 1 and 2, respectively.

1VAM

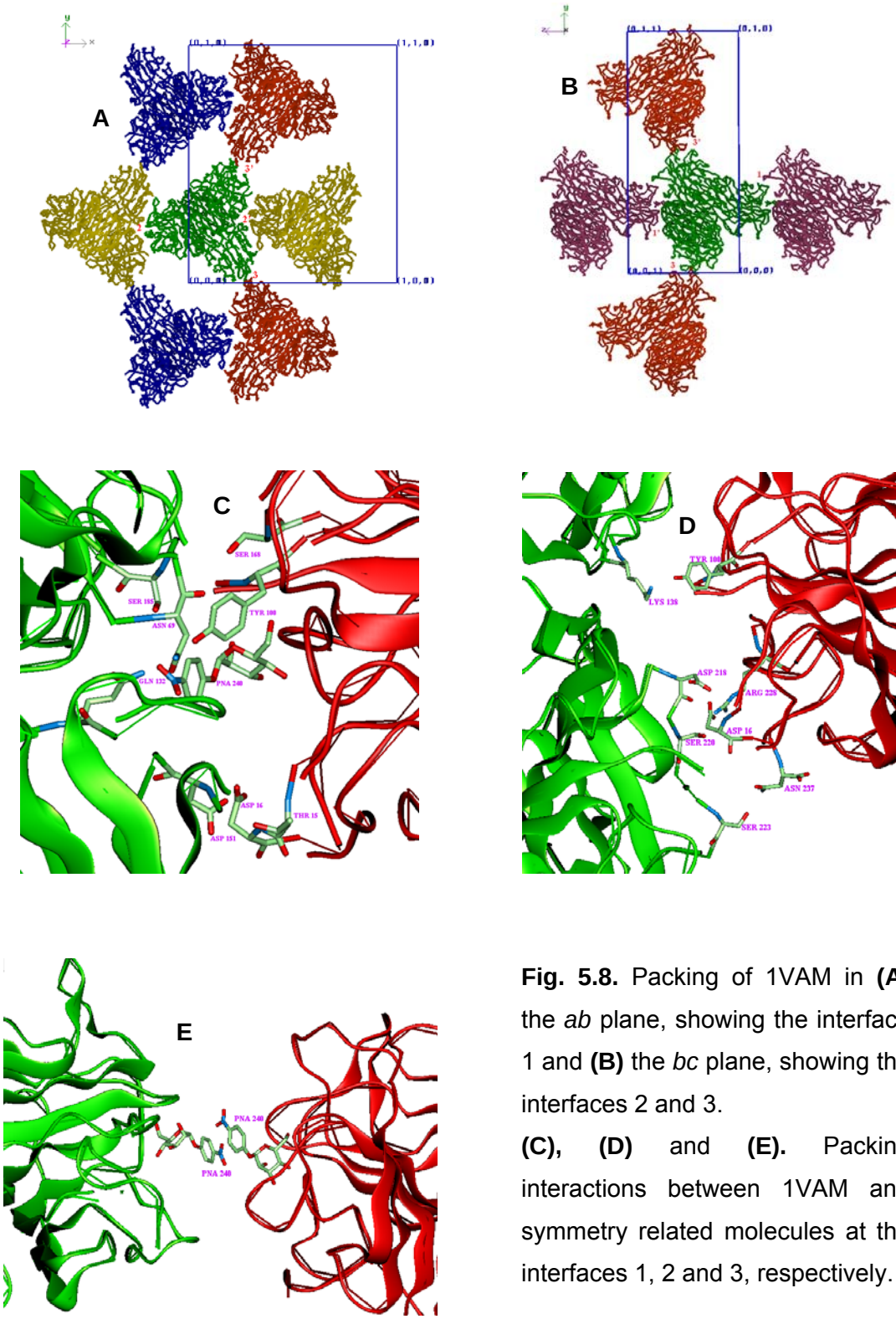


Fig. 5.8. Packing of 1VAM in **(A)** the *ab* plane, showing the interface 1 and **(B)** the *bc* plane, showing the interfaces 2 and 3.
(C), (D) and **(E).** Packing interactions between 1VAM and symmetry related molecules at the interfaces 1, 2 and 3, respectively.

1QDC

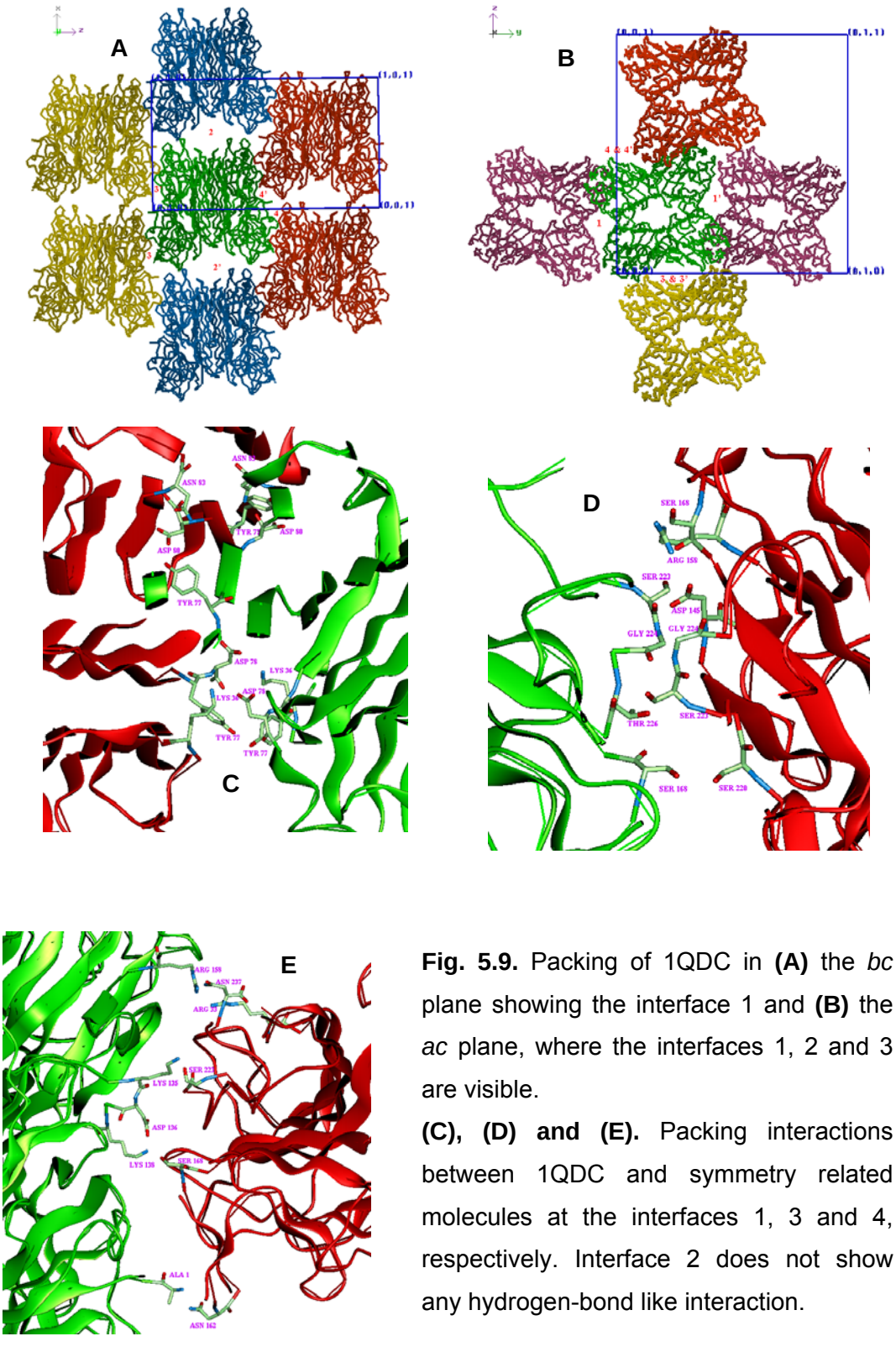


Fig. 5.9. Packing of 1QDC in **(A)** the *bc* plane showing the interface 1 and **(B)** the *ac* plane, where the interfaces 1, 2 and 3 are visible.

(C), (D) and (E). Packing interactions between 1QDC and symmetry related molecules at the interfaces 1, 3 and 4, respectively. Interface 2 does not show any hydrogen-bond like interaction.

5CNA

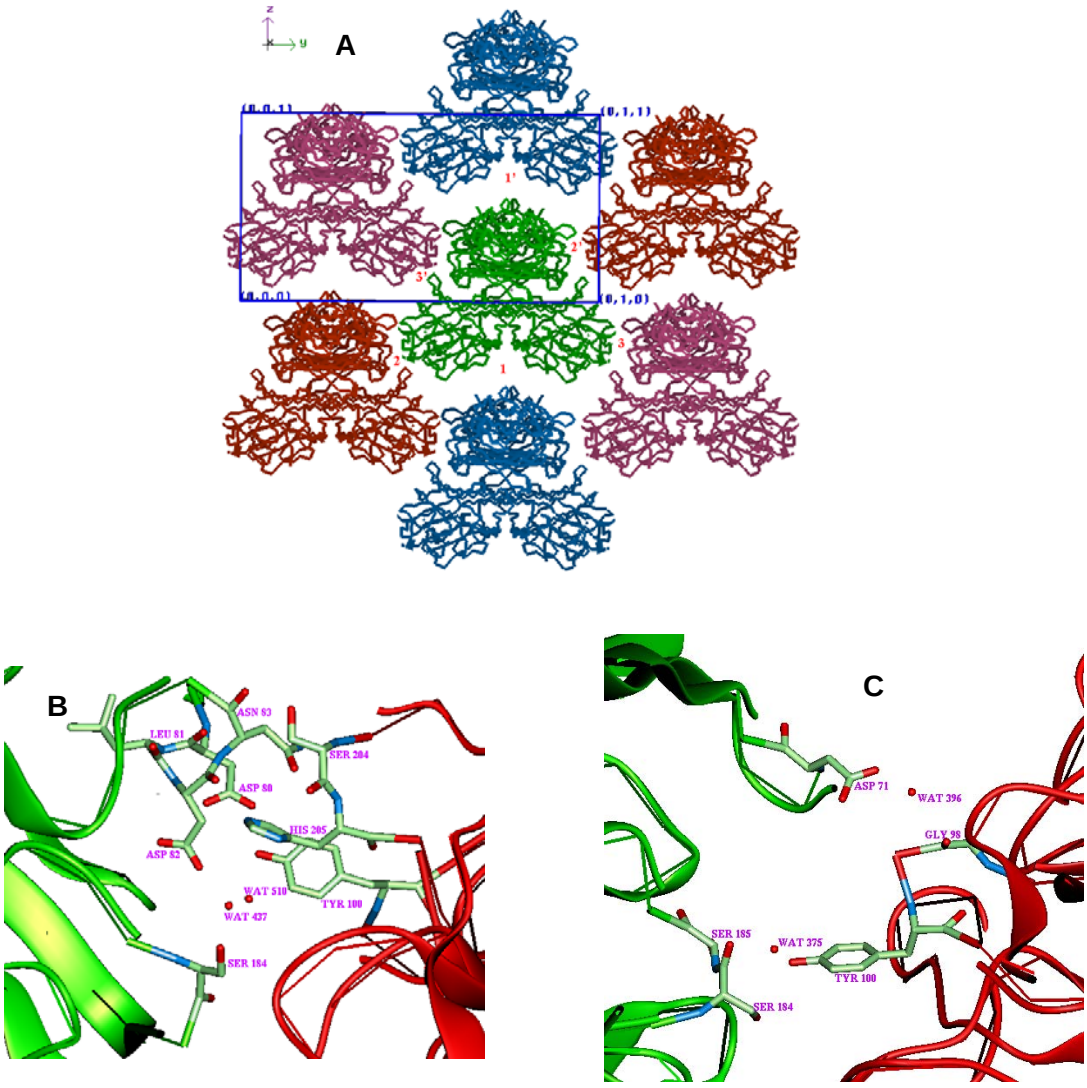


Fig. 5.10. (A) Packing of 5CNA in the YZ plane. The interfaces 1, 2 and 3 are shown. **(B)** and **(C)** show the packing interactions between 5CNA and symmetry related molecules at the interfaces 2 and 3, respectively. Interface 1 does not have any hydrogen-bond like interaction.

1BXH

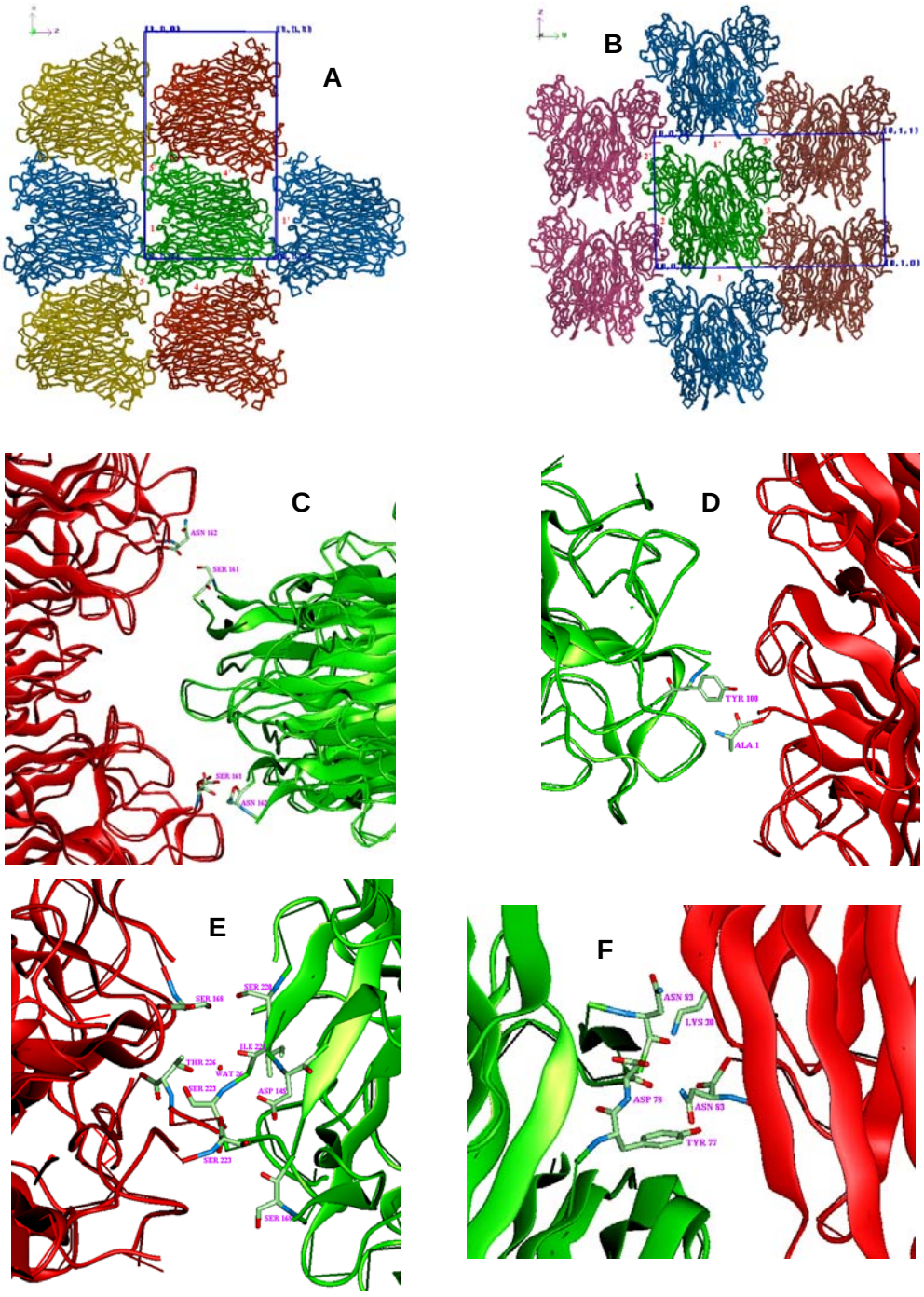


Fig. 5.11. Packing of 1BXH in **(A)** the *bc* plane, showing the interfaces 1, 2 and 3, and **(B)** the *ac* plane where the interfaces 4 and 5 are seen.

(C), (D), (E) and **(F)** show the packing interactions between 1BXH and symmetry related molecules at the interfaces 1, 2, 3 and 4, respectively. Interface 5 does not have any hydrogen-bond like interactions.

1GKB

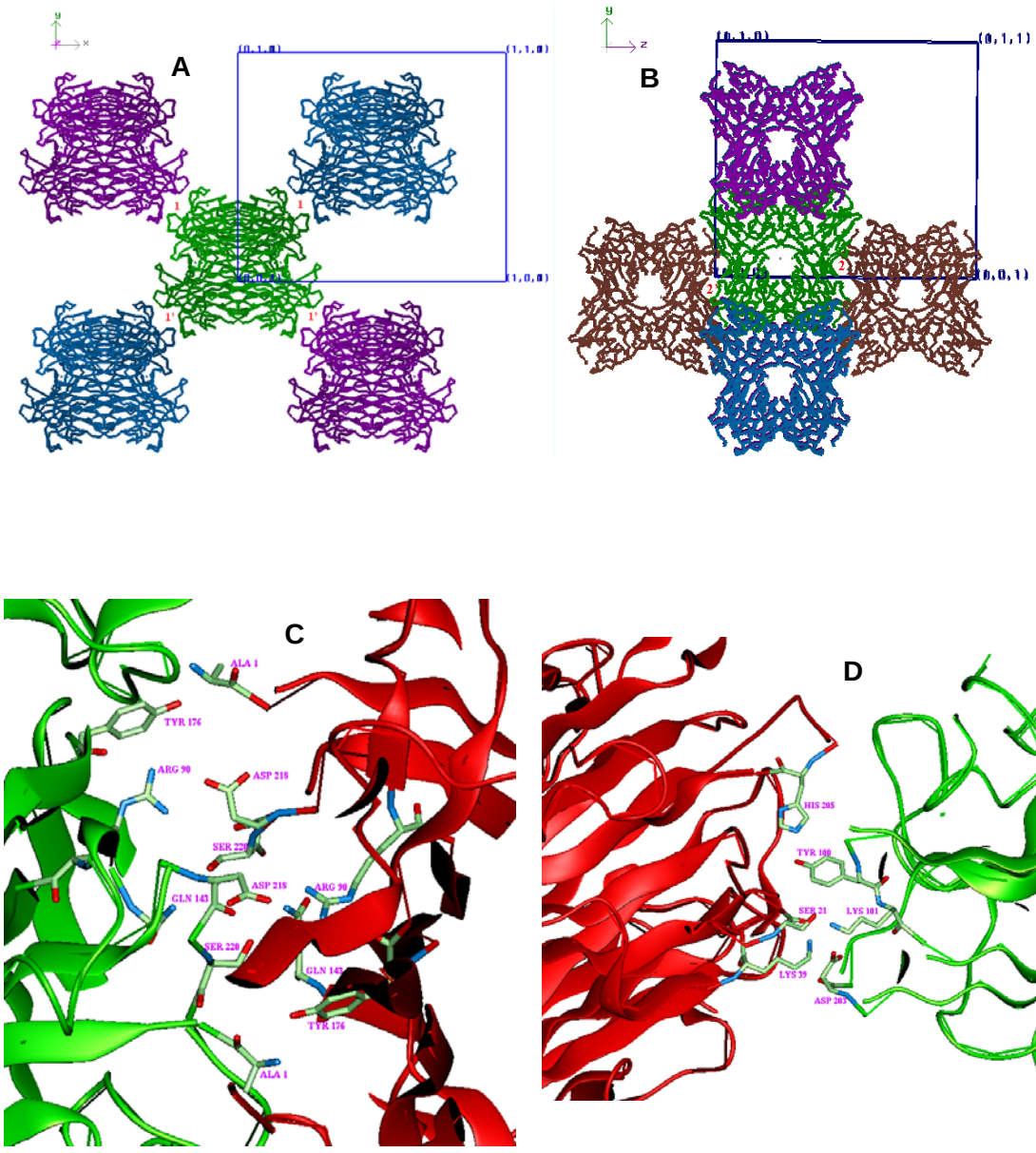


Fig. 5.12. Packing of 1GKB in **(A)** the *bc* plane showing the interface 1 and **(B)** the *ab* plane, where interface 2 is visible.

(C) and **(D)** show the packing interactions between 1GKB and symmetry related molecules at the interfaces 1 and 2, respectively.

1NXD

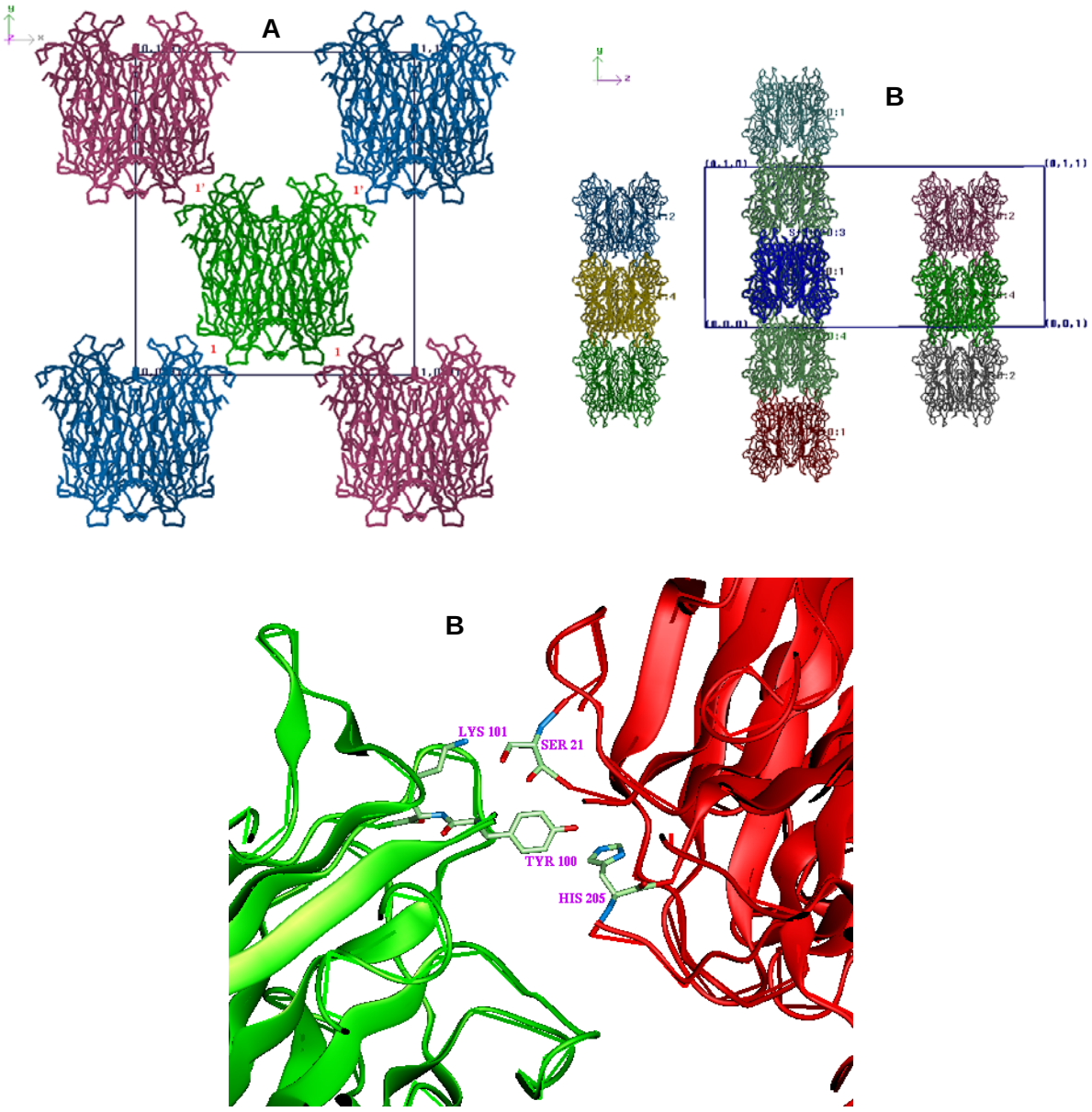


Fig. 5.13. Packing of 1NXD in **(A)** the *ab* plane and (B) the *bc* plane. Interface 1 is shown. The molecules which interact in 1GKB (**Fig. 5.12 B**) are placed far away in 1NXD, due to a large unit cell length in *c*-direction. **(C)** Packing interactions between 1NXD and symmetry related molecules at the interface 1.

1QGL

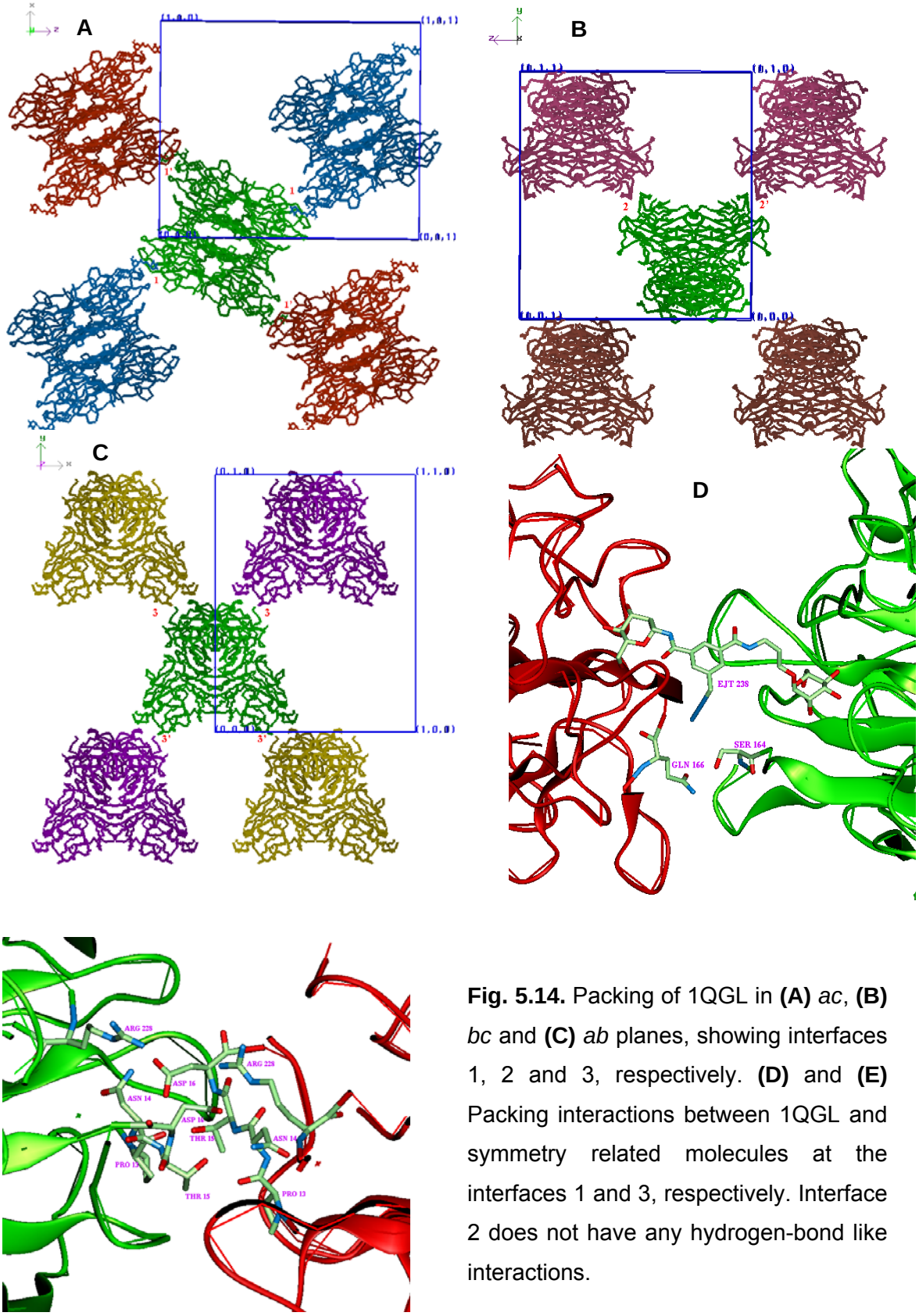


Fig. 5.14. Packing of 1QGL in **(A)** *ac*, **(B)** *bc* and **(C)** *ab* planes, showing interfaces 1, 2 and 3, respectively. **(D)** and **(E)** Packing interactions between 1QGL and symmetry related molecules at the interfaces 1 and 3, respectively. Interface 2 does not have any hydrogen-bond like interactions.

1NLS

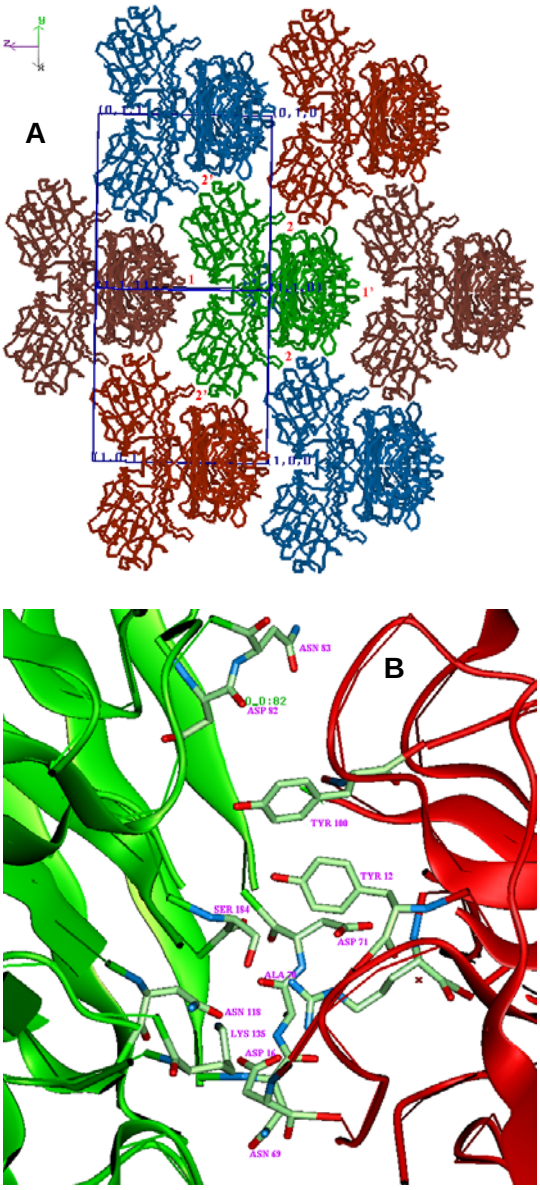


Fig. 5.15. (A) Packing of 1NLS viewed in the *a*-direction. Interfaces 1 and 2 are visible. **(B)** Packing interactions between 1NSL and symmetry related molecules at the interfaces 2. Interface 1 does not have any hydrogen-bond like

1JN2

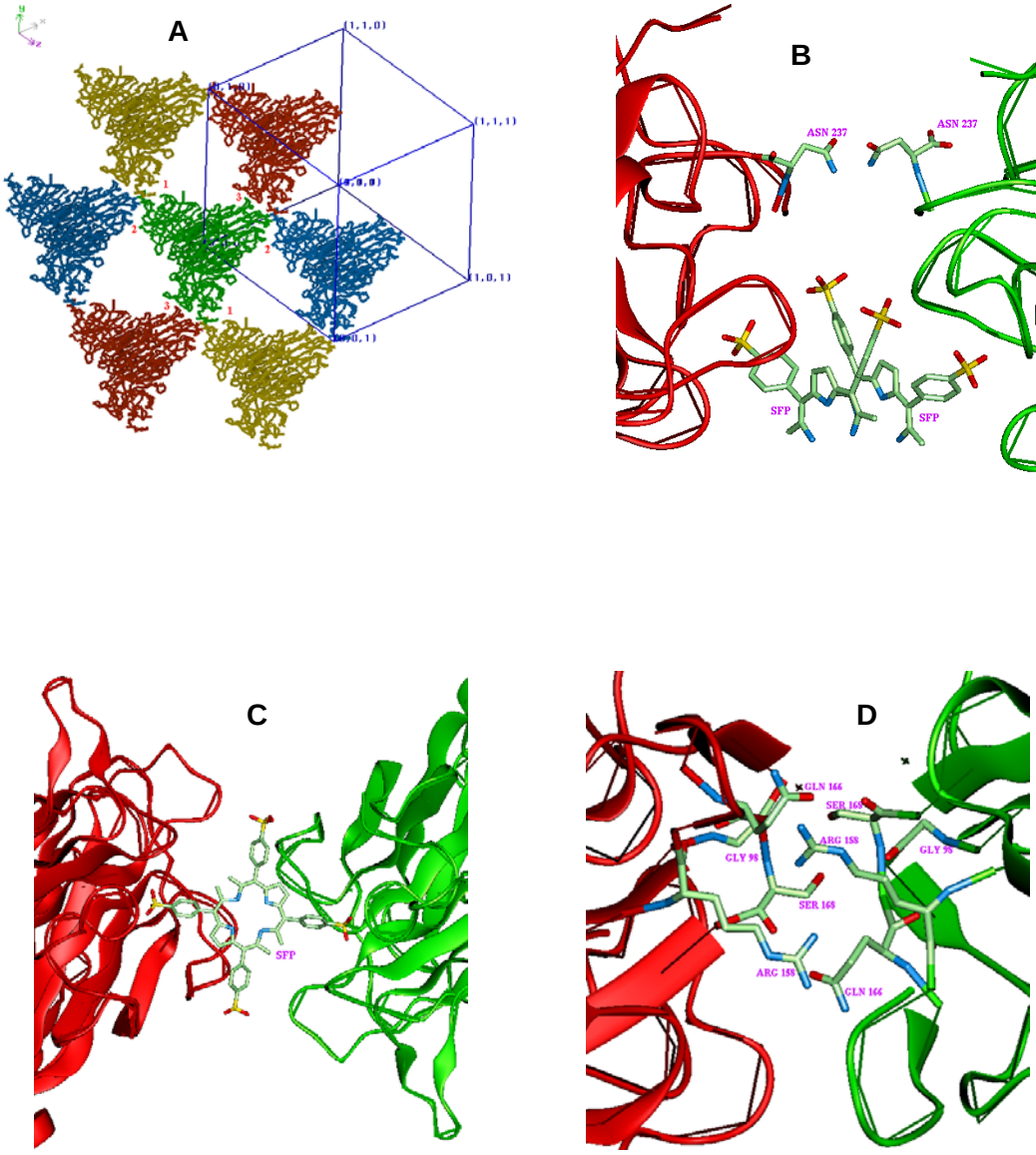


Fig. 5.16. (A) Packing of 1JN2. Interfaces 1, 2 and 3 are visible.
(B), (C) and **(D)** Packing interactions between 1JN2 and symmetry related molecules at the interfaces 1, 2 and 3, respectively.

1GIC

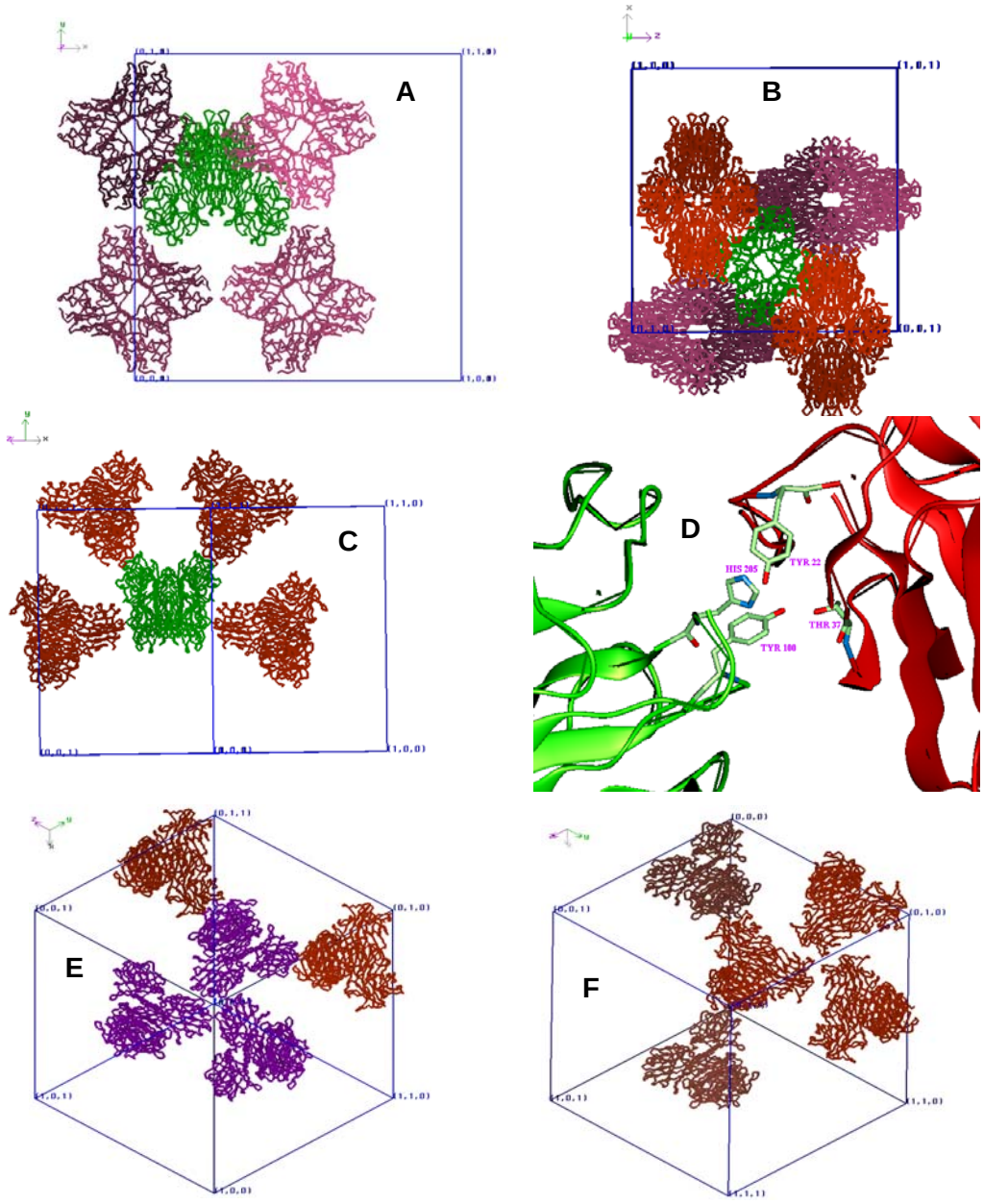
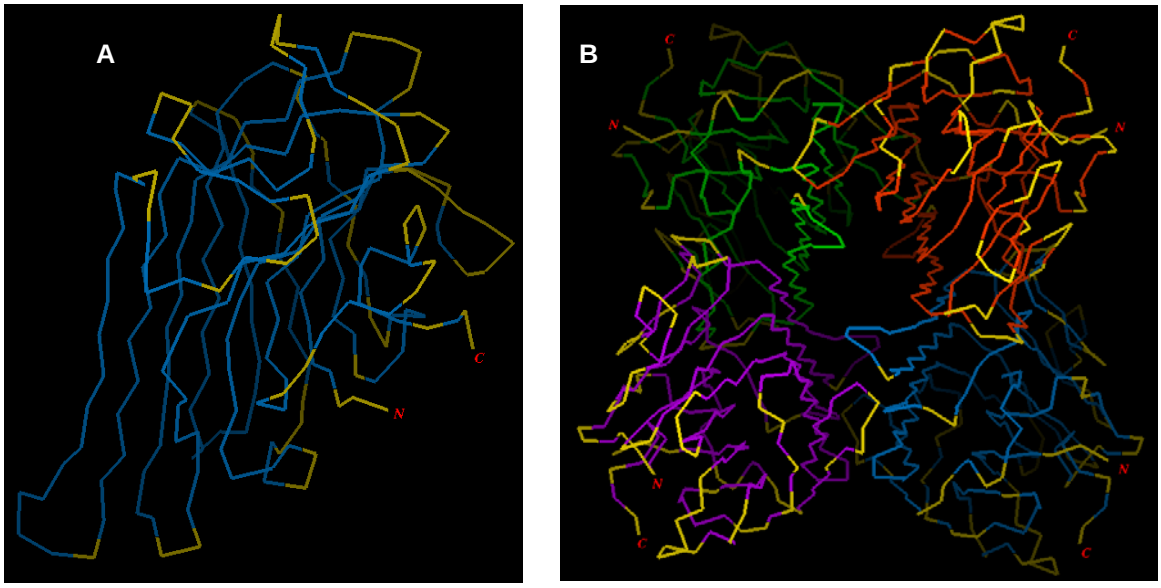


Fig. 5.17. (A), (B) and **(C)** show the packing of 1GIC in the *ab* and *ac* planes, as well as in the *b*-direction, respectively. Interface 1 is visible. **(D)** Packing interactions between 1GIC and symmetry related molecules at the interface 1. **(E)** Three-fold axis viewed down the body diagonal of the cube and **(F)** Three-fold axis parallel to a face diagonal of the cube.

5.4.4. Residues in ConA involved in crystal packing interactions

Fig. 5.18 represents the residues in the ConA at the interface which have their C^{α} atoms at less than 10 Å distance.



ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNMQDGKVGTAHIIYNSVDKRLSAVVSYPNA
DATSVSYDVDLNDVLPEWVRVGLSASTGLYKETNTILSWSFTSKLKSNSTHQTDALHFMFNQFSKDQKDL
ILQGDATTGTDGNLELTRVSSNGSPEGSSVGRALFYAPVHIWESSATVSAFEATFAFLIKSPDSHPADGI
AFFISNIDSSIPSGSTGRLLGLFPDAN

Fig. 5.18. (A) ConA monomer and **(B)** tetramer showing the C^{α} trace. C^{α} of amino acids which come within 10 Å distance from those in neighboring symmetry related molecules has been colored in yellow. *N* and *C* represent the *N*- and *C*- terminals, respectively. Below, a sequence of ConA is shown with the above mentioned amino acids colored in red. Residues which interact specifically with symmetry-related molecules however, if their C^{α}s do not come closer than 10 Å; are either colored in blue or underlined.

It can be seen that for residues with longer side chains, like Arg, Asp, Gln etc., the hydrogen-bond like interactions take place even if their C^{α}s do not come closer than 10 Å distance from those in neighboring symmetry related molecules.

5.4.5. Mutual disposition of subunits in polymorphic crystal structures

Here the subunits arrangement in tetramer is measured using the centre of mass of tetramers and the corresponding monomers. The tetrahedral shape of the ConA molecule is also depicted in the COM of all the four subunits. As position of COM depends on the position of all the atoms in the molecule, by calculating the distances and angles between the COM of each subunit and the tetramer, the distortion of the molecule can be measured. In the case of ConA, these distances and angles were calculated in order to study the expansion/contraction of the molecule which might occur due to binding of ligand or as a result of crystal packing. The definition of various distances and angles are shown in Fig. 5.19, for a tetramer in the representative structure, 1BXH. The distances and angles between COMs for various molecules in the 17 crystal structures of ConA have been listed in Table 5.9.

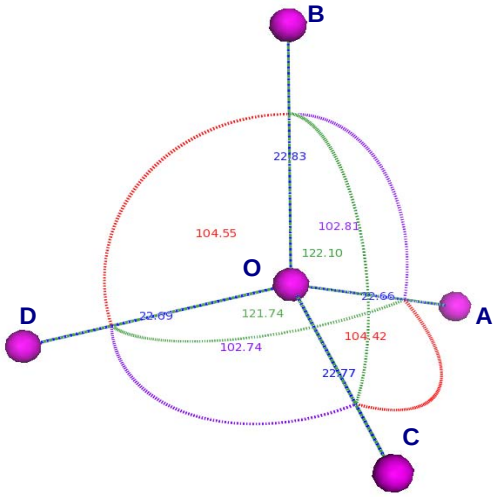


Fig. 5.19. A representative diagram showing the calculation of tetrahedral angles and distances for ConA molecules. Atom O represents the COM for the tetramer and atoms A, B, C and D represent the COM for subunits A, B, C and D, respectively. The distance between each COM and O is represented by a number colored blue. Angles in purple are the ones between the subunits of a canonical dimer; those in red are the angles between two adjacent subunits belonging to two canonical dimers and those in green are the ones between diagonally opposite subunits.
The diagram was prepared using PyMOL (DeLano, 2002).

Table 5.9. Distances and tetrahedral angles calculated for each molecule using COMs

PDB ID	Lengths				Mean Length	Angle 1		Mean angle 1	Angle 2		Mean angle 2	Angle 3		Mean angle 3
1VLN_1	23.2	23.1	23.0	23.1	23.1	101.8	101.0	101.4	104.3	102.5	103.4	126.4	123.1	124.8
1VLN_2	23.1	23.2	23.1	23.1	23.1	102.0	101.7	101.9	103.6	103.7	103.7	123.9	124.1	124.0
1CJP	22.9	22.9	22.8	23.0	22.9	102.0	102.0	102.0	103.7	104.2	104.0	123.5	123.2	123.4
1CVN	22.9	22.9	22.9	22.8	22.9	101.8	101.3	101.6	104.1	104.1	104.1	123.4	123.5	123.5
1ONA	23.9	23.4	23.6	23.4	23.6	102.4	102.4	102.4	104.0	104.1	104.1	122.7	123.1	122.9
1TEI_1	23.1	22.9	22.8	23.1	23.0	101.5	101.6	101.6	104.0	104.0	104.0	124.2	123.7	124.0
1TEI_2	22.8	23.1	23.0	23.0	23.0	101.3	101.4	101.4	102.7	103.4	103.1	125.0	125.5	125.3
1DQ2	22.0	21.7	22.0	21.7	21.9	101.1	101.1	101.1	104.2	104.2	104.2	124.4	124.4	124.4
3ENR	22.8	23.0	22.8	23.0	22.9	101.7	101.7	101.7	104.2	104.2	104.2	123.5	123.5	123.6
1VAM	23.0	23.0	23.0	23.0	23.0	102.3	102.2	102.3	103.2	103.2	103.2	124.0	124.0	124.0
1QDC	23.0	22.9	22.9	23.0	23.0	101.7	101.6	101.7	104.1	104.1	104.1	123.8	123.5	123.7
5CNA	22.9	22.9	22.9	22.9	22.9	102.1	102.0	102.1	104.4	104.4	104.4	122.9	122.7	122.8
1BXH	22.7	22.8	22.8	22.7	22.8	102.8	102.7	102.8	104.4	104.6	104.5	121.7	122.1	121.9
1GKB	22.9	22.8	22.9	22.8	22.9	101.7	101.7	101.7	104.7	104.7	104.7	122.9	122.9	122.9
1NXD	22.8	22.8	22.8	22.7	22.8	101.8	101.7	101.8	104.9	104.9	104.9	122.6	122.6	122.6
1QGL	22.9	22.9	22.9	22.9	22.9	101.8	101.8	101.8	104.2	104.2	104.2	123.4	123.4	123.4
1NLS	22.5	22.5	22.5	22.5	22.5	104.1	104.1	104.1	105.2	105.2	105.2	119.7	119.7	119.7
1JN2	23.4	23.4	23.4	23.4	23.4	101.8	101.8	101.8	103.1	103.1	103.1	124.7	124.7	124.7
1GIC	22.9	22.9	22.9	22.9	22.9	101.5	101.5	101.5	104.7	104.7	104.7	123.2	123.2	123.2

Since 1VLN and 1TEI have octamers in their asymmetric units, parameters for the two tetramers are separately listed; one containing the chains A, B, C and D; and the other one containing the chains E, F, G and H. These were individually used for the calculations.

The mean tetrahedral distance for each molecule varies from 21.9 (1DQ2, tetramer) to 23.6 (1ONA). Although a perfect tetrahedron has all of its six angles ~109.5^0, in the case of ConA three kinds of angles are observed for each molecule: "Angle 1" is the angle between the two subunits in the canonical dimer and has a value around 102^0; "Angle 2" is the angle between the neighboring subunits of different canonical dimers with a value around 104^0. In the case 1NLS, however, these two angles are quite close to each other (Angle 1 ≈ 104^0 and Angle 2 ≈ 105^0). "Angle 3" is the angle between the diagonally opposite subunits and has a value ranging from 119 to 126^0. Since four of the tetrahedral angles are reduced to ~101-105^0, the remaining two angles have expanded with larger values to compensate them.

When there is a monomer or dimer in the asymmetric unit and the tetramer is assembled symmetrically satisfying the crystallographic symmetry, the above mentioned lengths and angles will have same values, for example, 1NLS.

The low variation in these parameters indicates that the tetrahedral arrangement of ConA subunits does not get distorted much because of the crystal packing or binding of ligands.

5.4.1. Polymorphism in other plant lectins

Besides ConA, many other plant lectins also show significant degree of polymorphism. Table 5.10 enlists the unique crystal forms observed in other plant lectins. In this thesis, no attempt was made to bring out detailed analysis of these structures.

Table 5.10. Unique crystal forms of various plant lectins.

Lectin	PDB id	Space group	a	b	c	α	β	γ
Artocarpus integrifolia lectin (artocarpin) Polymorphs: 5, Total: 5	1J4S	$P12_11$	69.9	73.7	60.6	90	95.0	90
	1J4T	$P12_11$	87.7	72.2	92.6	90	101.0	90
	1VBO	$P12_11$	138.2	72.3	59.4	90	94.0	90
	1J4U	$P6_1$	129.2	129.2	78.6	90	90	120
	1VBP	$P4_132$	212.6	212.6	212.6	90	90	90
Urtica dioica lectin Polymorphs: 5, Total: 6	1EHH	$P12_11$	25.6	56.8	55.3	90	93	90
	1IQB	$P12_11$	30.7	42.1	62.6	90	102.0	90
	1EIS	$P2_12_12_1$	31.3	41.6	77.1	90	90	90
	1EN2	$P2_12_12_1$	31.8	39.6	63.6	90	90	90
	1ENM	$P2_12_12_1$	38.8	46.2	57.3	90	90	90
Artocarpus integrifolia lectin (jacalin) Polymorphs: 4, Total: 12	1M26	$P12_11$	58.9	78.0	67.9	90	101.0	90
	1WS4	$P2_12_12_1$	79.3	99.4	105.5	90	90	90
	1UGX	$I222$	43.3	100.6	102.4	90	90	90
	1JAC	$P6_522$	129.6	129.6	157.9	90	90	120
Erythrina cristagalli lectin Polymorphs: 4, Total: 5	1UZZ	$P1$	55.3	55.4	86.9	86.2	75.0	82.1
	1V00	$P12_11$	54.9	167.2	55.1	90	97.0	90
	1UZY	$P6_5$	134.0	134	81.6	90	90	120
	1GZC	$P4_32_12$	81.8	81.8	126.1	90	90	90
Peanut agglutinin Polymorphs: 4, Total: 22	1CQ9	$P1$	53.6	71.8	86.4	65.4	78.0	72.3
	1CR7	$P12_11$	128.3	126.8	85.6	90	116.0	90
	1V6I	$P2_12_12$	128.6	125.8	76.1	90	90	90
	1RIT	$P3_2$	94.9	94.9	144.1	90	90	120
Lentil lectin Polymorphs: 4, Total: 4	1LEN	$P12_11$	58.5	56.4	82.7	90	104.0	90
	1LES	$P12_11$	50.0	124.8	50.0	90	112.0	90
	2LAL	$P2_12_12_1$	75.8	125.5	56.5	90	90	90
	1LEM	$P6_522$	85.7	85.7	165.4	90	90	120
Phytolacca americana lectin Polymorphs: 3, Total: 4	1ULM	$C121$	98.6	26.3	65.2	90	110.0	90
	1ULN	$P2_12_12$	48.7	49.0	29.9	90	90	90
	1ULK	$H3$	104.1	104.1	69.7	90	90	120
Lathyrus ochrus lectin 1 Polymorphs: 3, Total: 7	1LOB	$P12_11$	56.3	139.8	62.7	90	91.0	90
	1LOF	$C121$	78.3	75.4	103.9	90	92.0	90
	1LOE	$P2_12_12$	135.8	63.1	54.5	90	90	90
Pisum sativum lectin Polymorphs: 3, Total: 6	1BQP	$P2_12_12_1$	73.9	104.1	64.8	90	90	90
	2LTN	$P2_12_12_1$	50.7	61.2	136.6	90	90	90
	2BQP	$P2_12_12_1$	62.8	135.3	54.8	90	90	90
Galanthus nivalis agglutinin Polymorphs: 3, Total: 3	1MSA	$P2_12_12_1$	138.8	64.11	62.1	90	90	90
	1JPC	$I4_122$	96.3	96.3	68.6	90	90	90
	1NIV	$I2_13$	138.2	138.2	138.2	90	90	90
Erythrina corallodendron lectin Polymorphs: 3, Total: 8	1AX0	$C121$	84.2	73.0	71.3	90	113.0	90
	1SFY	$C121$	87.2	144.9	127.7	90	93.0	90
	1FYU	$P6_5$	135.9	135.9	82.6	90	90	120

Lectin	PDB id	Space group	a	b	c	α	β	γ
Ricinus communis agglutinin Polymorphs: 2, Total: 2	2AAI	$P2_12_12_1$	72.7	78.5	114.3	90	90	90
	1RZO	$P3_2$	97.6	97.6	207.8	90	90	120
Pterocarpus angolensis lectin Polymorphs: 2, Total: 23	1UKG	$P2_12_12_1$	56.8	83.7	123.0	90	90	90
	2GME	$P2_12_12_1$	58.9	62.0	130.2	90	90	90
Winged bean agglutinin I Polymorphs: 2, Total: 7	1WBF	$C121$	99.16	75.16	95.77	90	107.0	90
	2D3S	$P2_12_12$	157.87	91.9	73.5	90	90	90
Soybean agglutinin Polymorphs: 2, Total: 5	1SBD	$P6_422$	144.6	144.6	107.2	90	90	120
	1SBF	$I4_122$	122.6	122.6	90.6	90	90	90
Wheat Germ Agglutinin 1 Polymorphs: 2, Total: 3	7WGA	$C121$	51.2	73.6	91.4	90	98.0	90
	2CWG	$P2_12_12$	111.0	50.4	63.4	90	90	90
Gastrodia elata lectin Polymorphs: 2 Total: 2	1XD5	$P2_12_12$	61.1	91.5	81.1	90	90	90
	1XD6	$C222_1$	78.1	97.5	36.1	90	90	90
Parkia platycephala lectin Polymorphs: 2, Total: 2	1ZGS	$P12_11$	80.2	114.2	80.3	90	120	90
	1ZGR	$P2_12_12_1$	63.6	68.5	208.5	90	90	90
Cratylia floribunda lectin Polymorphs: 2, Total: 2	2D3P	$P2_12_12_1$	71.4	106.9	119.6	90	90	90
	2D3R	$P2_12_12_1$	60.2	125.4	126.1	90	90	90
Dioclea guianensis lectin Polymorphs: 2, Total: 2	1H9P	$I4_122$	90.3	90.3	108.3	90	90	90
	1H9W	$P4_32_12$	90.2	90.2	106.8	90	90	90
Griffonia simplicifolia lectin Polymorphs: 2, Total: 2	1HQL	$P2_12_12$	111.2	51.3	77.1	90	90	90
	1GNZ	$P6_422$	75.9	75.9	190.6	90	90	120
Lathyrus ochrus lectin 1 Polymorphs: 2, Total: 2	1LGB	$P4_32_12$	63.5	63.5	251.9	90	90	90
	1LGC	$P4_32_12$	117.0	117.0	120.1	90	90	90
Ulex europaeus agglutinin 1 Polymorphs: 2, Total: 2	1JXN	$P12_11$	71.8	69.0	119.0	90	107.0	90
	1FX5	$C121$	79.6	70.7	122.2	90	109.0	90
Winged bean agglutinin II Polymorphs: 2, Total: 2	1FAY	$C121$	134.5	125.9	138.8	90	96	90
	1F9K	$H3$	182.1	182.1	45.0	90	90	120
Artocarpus hirsuta lectin Polymorphs: 2, Total: 2	1TOQ	$P2_12_12_1$	92.4	98.7	164.9	90	90	90
	1TP8	$P2_12_12_1$	89.9	121.9	131.6	90	90	90

5.4.2. Polymorphism observed in other proteins

Besides plant lectins, many other proteins are also known to exhibit polymorphism to various degrees in their reported crystal structures. Crystal forms of a few such proteins, namely bovine seminal ribonuclease, bovine pancreatic ribonuclease A, human insulin, lysozyme A, thymidylate synthases, bovine trypsin chain A have been identified from PDB and enlisted in Table 5.11. Because of time constraints, any further analysis of these structures is beyond the scope of this thesis.

Table 5.11. Crystal forms of some selected proteins exhibiting polymorphism.

PDB ID	Space Group	a	b	c	α^0	β^0	γ^0	Res (Å)	R_{value}	R_{free}
Bovine Seminal Ribonuclease										
11BA	$P12_11$	49.5	60.6	50.8	90	117.4	90	2.06	0.184	NA
1N3Z	$C121$	70.7	28.8	67.5	90	119.2	90	1.65	0.188	0.230
1TQ9	$P2_12_12$	41.5	69.7	110.8	90	90	90	2.00	0.210	0.269
1BSR	$P22_12$	36.5	66.7	107.5	90	90	90	1.90	0.177	NA
1R5D	$P2_12_12$	49.6	60.2	83.4	90	90	90	2.50	0.203	0.255
Bovine Pancreatic Ribonuclease A										
1RHA	$P12_11$	30.3	33.3	52.6	90	113.3	90	1.80	0.176	NA
1U1B	$P12_11$	31.1	75.9	51.8	90	106.3	90	2.00	0.218	0.258
1XPS	$P12_11$	33.5	106.2	30.6	90	102.0	90	1.80	0.175	0.248
1JVT	$C121$	101.5	33.3	73.5	90	90.3	90	2.05	0.180	0.240
3RN3	$P12_11$	30.5	38.4	53.2	90	106.0	90	1.45	0.223	NA
1RTA	$P2_12_12$	44.9	75.0	43.6	90	90	90	2.50	0.235	NA
1RSM	$P2_12_12$	37.1	41.3	75.6	90	90	90	2.00	0.184	NA
8RSA	$P2_12_12$	53.1	64.6	73.6	90	90	90	1.80	0.162	NA
1RCN	$P2_12_12_1$	71.9	43.2	43.8	90	90	90	2.32	0.172	NA
1RBN	$C222_1$	75.7	57.9	53.3	90	90	90	2.10	0.166	NA
1A2W	$P3_2$	57.0	57.0	81.4	90	90	120	2.10	0.192	0.256
1FS3	$P3_221$	64.1	64.1	63.4	90	90	120	1.40	0.217	0.255
Human Insulin										
1OS4	$P1$	33.9	49.4	49.3	115.1	104.0	102.7	2.25	0.243	0.295
1HTV	$P2_12_12_1$	49.8	51.5	100.6	90	90	90	1.90	0.196	0.240
2OMI	$P2_12_12_1$	60.5	61.8	86.0	90	90	90	2.24	0.203	0.268
2CEU	$I222$	58.2	58.2	54.8	90	90	90	1.80	0.174	0.211
2OMG	$P4_32_12$	61.7	61.7	85.5	90	90	90	1.52	0.184	0.209
1MSO	$H3$	81.3	81.3	33.7	90	90	120	1.00	0.183	0.201
Lysozyme A										
2VB1	$P1$	27.1	31.3	33.8	88.0	108	112.1	0.65	0.085	0.095
1HF4	$P12_11$	27.9	62.7	60.3	90	90.8	90	1.45	0.215	0.256
1B0D	$P4_32_12$	79.1	79.1	38.0	90	90	90	1.84	0.202	0.241
Thymidylate Synthases										
1TRG	$P3_121$	72.0	72.0	115.1	90	90	120	1.90	0.180	0.230
2A9W	$H3$	186.6	186.6	114.2	90	90	120	1.65	0.207	0.226
1JTU	$P6_3$	127.6	127.6	68.0	90	90	120	2.20	0.202	0.241
1JG0	$P2_12_12_1$	53.8	87.1	127.5	90	90	90	2.00	0.212	0.250
3TMS	$I2_13$	133.0	133.0	133.0	90	90	90	2.10	0.220	NA
Bovine Trypsin Chain A										
1TAW	$C121$	94.1	49.9	68.8	90	96.3	90	1.80	0.184	NA
1K1I	$P2_12_12_1$	54.9	58.9	67.2	90	90	90	2.20	0.165	NA
1Y3W	$P2_12_12_1$	63.1	69.3	64.0	90	90	90	1.80	0.177	0.237
2PLX	$P2_12_12_1$	113.6	41.5	51.2	90	90	90	1.56	0.145	0.18
1TYN	$P4_32_12$	71.7	71.7	88.8	90	90	90	2.00	0.169	0.222
1K1O	$P3_121$	55.2	55.2	110.0	90	90	120	2.00	0.165	NA
1AQ7	$P6_1$	48.3	48.3	145.2	90	90	120	2.20	0.164	0.237

NA: Not Available

5.1. Conclusions

An analysis of the polymorphism observed in crystal structures of various plant lectins was undertaken. Since Concanavalin A (ConA) is a lectin with maximum number of polymorphic crystal structures reported, a detailed study of various crystal structures of this protein was carried out. Packing of ConA molecules in different crystal structures and interactions between the symmetry related molecules were studied. A relation between the unit cell parameters and packing interactions could be established. It is still difficult to find direct correlation between pH and crystallization conditions based on this small dataset that leads to polymorphism and different packing interactions in ConA. Since most of the crystal structures have different ligand bound to the ConA molecules, it is possible that these ligands influence the polymorphism, although in very few cases ligands actually participate in the crystal packing interactions. The tetrahedral assembly of ConA monomers remains unaffected in different crystal environments analyzed here.

Chapter 6

Features of homotetrameric molecular association in the crystals of lectins and other proteins

6.1. Summary

The crystal structures of proteins with homotetrameric association, a common feature of many lectins, were analyzed to understand the characteristics of tetrameric association in terms of the arrangement of their subunits and its biological significance. The analysis could group the tetramer units into four categories:

1. Tetrahedral molecules, in which the four monomers form a nearly perfect tetrahedral arrangement. The angle between axes of any two monomers is ~ 109^0. Sometimes, the tetrahedral shape is distorted that is, the tetrahedral angle deviates from ideal value, which gives the molecule a twisted shape rather than a perfect tetrahedral shape.
2. Molecules that form sandwiched dimer of dimers where the two dimers are arranged perpendicular to each other, one upon the other.
3. Planar molecules where the four monomers lie in one plane and the corresponding sides of adjacent monomers face opposite directions (that makes diagonally opposite monomers to face the same direction). This can be considered a flattened tetrahedral shape.
4. Planar closed molecules where all the four monomers lie in one plane arranged in a head-to-tail fashion in a square.

The first group is the most commonly found arrangement. The importance of each arrangement for its biological function will be discussed. In addition, some unusual tetrameric assemblies were also observed, which were found to be relevant with the biological function of the particular protein.

6.2. Introduction

Many plant lectins, such as the *Artocarpus hirsuta* lectin reported from our group, concanavalin A analyzed for crystal polymorphism in the previous chapter of this thesis and other proteins such as penicillin V acylase and conjugated bile acid hydrolase also reported from our lab all showed a tetrameric association of their subunits. From the analysis of the structure of *Artocarpus hirsuta* lectin it was concluded that a pattern of subunit-association was formed to keep the functional sites of subunits at maximum distance away from each other. Obviously for a tetrameric association of single binding-site subunits a tetrahedral positioning is the most ideal one to minimize steric hindrance between ligands. This also helps to form networks of cells during agglutination of red blood cells. Based on this idea an analysis of the structures of proteins that are known to form tetramers was undertaken.

Researchers have approached the problem of protein assembly under different contexts. A variety of purposes were attributed to formation of oligomers and complexes. For example, in the case of virus assembly the genetic economy was identified as the advantage for subunit interactions and arrangement (Dokland, 2000; Phelps *et al.*, 2000). Transcription and association of several copies from the same gene can minimize errors or alternatively, can uniformly distribute mutations throughout the assembly. For protein like hemoglobin the requirement of tetramer was for the allosteric control of oxygen binding (Monod *et al.*, 1965). In certain membrane proteins subunit assembly helps to create an external hydrophobic and internal polar surface to help ion transport (Manting *et al.*, 2000; Sakaguchi *et al.*, 1997). Oligomerisation is proposed in situations where regulation of concentration levels of constituent subunits is required (Bray and Lay, 1997), which also results in increased stability and reduced surface area of the constituent molecules (Larsen *et al.*, 1997). In certain enzymes subunits assemble to form a symmetric substrate-binding cleft such as in the HIV protease (Wlodawer and

Erickson, 1993). Oligomerisation is identified as one of the ways to achieve thermostability of proteins in thermophilic organisms (Walden *et al.*, 2001). As may be expected, in most cases oligomerisation leads to higher order of complexity.

Many researchers have explored various aspects of protein oligomerisation and their relevance in biology (Ali and Imperiali, 2005). One study estimated more than one-third of the cellular proteins form oligomers (Goodsell and Olson, 2000). Although, oligomers can be composed of multiple subunits of the same polypeptide (homo-oligomer) or different polypeptides (hetero-oligomer), the preference seems to be for the formation of homo-oligomers in the cell (Goodsell, 1991). Similarly, even though oligomeric proteins can be formed from any number of subunits the average oligomeric state of cellular proteins estimated was tetrameric (Goodsell, 1991). In homo-oligomeric proteins, since the constituents of the assembly are identical, its formation can introduce simple point group symmetry. Goodsell and Olson (2000) have estimated that the point group symmetries of cyclic, dihedral and cubic are most frequently observed.

The stability of an oligomer will directly depend on the strength of association of the subunits, their affinity and duration. Thus, one with strong subunit-interactions can invariably be found as an oligomer, while the formation of oligomers by those with weaker interaction may depend on the concentration and other conditions like pH and temperature of the solution or in response to some other stimuli (Nooren and Thornton, 2003a, b).

Many researchers have attempted to rationalize and quantify protein-protein recognition, the type of interactions and the nature of interface involved in protein oligomerisation in a wider sense (Chothia and Janin, 1975; Miller *et al.*, 1987; Argos, 1988; Janin *et al.*, 1988; Miller, 1989; Jones and Thornton, 1996). Rationalization of quaternary structure formation in the case legume lectins is carried out in terms of the parameters of buried hydrophobic surface, interaction energy and shape

complementarity at the interface of subunits (Prabhu *et al.*, 1999). Similarly, there were attempts to study protein oligomerisation and stability in the case of legume lectins by analyzing the unfolding of their quaternary organisation (Srinivas *et al.*, 2001). Here the attempt is to analyze the quaternary structures of tetrameric lectins and other proteins with tetrameric association in terms of their symmetry of organization and biological relevance.

6.3. Materials and Methods

All the computational work was carried out on a Silicon Graphics workstation (Octane) with Irix 6.5 as the operating system as well an IBM PC with Fedora Core 6 as the operating system. Atomic coordinates of various homotetrameric proteins were downloaded from the Protein data bank (PDB, http://www.rcsb.org/pdb/) according to their space groups, at a cut-off of sequence identity of 70%. The information in the PDB file (REMARK 350) was used to decide whether the biological unit of the protein is a homotetramer or not. Although an attempt has been made to cover all unique tetrameric protein structures in the analysis, it is likely that some of the structures might have been excluded if they were not categorized as tetramers by PDB. Hetero-octameric or sometimes, even hetero-dodecameric molecules in which one subunit consisted of two or three different monomers, and four such subunits formed the biological molecule, were also considered for the analysis purpose.

Wherever the asymmetric unit differed from the biological unit, the coordinates of the biological unit were downloaded from the PDB website (*.pdb1 files), or, sometimes the biological unit was generated by calculating the coordinates of symmetry related subunits. The protein structures were visualized using the graphical software QUANTA (Accelrys, Inc.) and grouped into various classes. The secondary structure content of the protein was also roughly estimated, to check whether there is any dependence of the

overall assembly of the molecule on the secondary structure. Centers of mass of the tetramer and its subunits for selected structures were calculated as described in Chapter 5 (section 5.3). The diagrams of quaternary structures of proteins and that of centers of mass were prepared using PyMOL (DeLano, 2000).

6.4. Results and Discussion

At a sequence identity cut-off of 70%, about 700 unique homotetrameric protein structures were analyzed. The structures were grouped according to the space groups to find any trend in preferred quaternary structure in a particular space group. These protein structures could be grouped into four major categories.

6.4.1. Dihedral / Tetrahedral type assemblies

In this type, the four subunits of a protein are arranged pointing towards the four corners of an approximate tetrahedron. Concanavalin A (ConA), described in chapter 5, exhibits a near-perfect tetrahedral shape (Fig. 6.1. (A)), as shown by the measurements of the tetrahedral angles between the centers of mass of the four subunits. Other tetrameric legume lectins, for example *Dioclea grandiflora* lectin (PDB code 1DGL; Rozwarski *et al.*, 1998), *Phaseolus vulgaris* lectin (PDB code 1FAT; Hamelryck *et al.*, 1996) as well as hetero-octameric *Dolichos lablab* lectin (FRIL, PDB code 1QMO; Hamelryck *et al.*, 2000) also show a near-perfect tetrahedral arrangement of subunits. A variation of this kind of arrangement is observed in many other proteins, including enzymes and the arrangement gives an internal dihedral symmetry to the tetramer (point group 222).

Most of the tetrameric proteins show two steps of oligomerisation. First, two monomers associate to form a dimer and two dimers in turn associate to form a tetramer (Powers and Powers, 2003). Due to this, even in the near-perfect tetrahedral shaped

molecules, the angles between any two monomers deviate at least slightly from the normal tetrahedral angle of $109^0 28'$. The angles between two monomers of the same dimer have the least value, ~100^0, followed by the angles between adjacent monomers of two different dimers, which are at least 3-4^0 more than the angle between monomers of the same dimer. The angles between diagonally oppositely placed subunits have values of ~125^0, or sometimes even more, to compensate for the other four angles.

Not all homotetrameric proteins show such a near-perfect tetrahedral shape. Most of the proteins have the angles between their subunits deviating much more, distorting the tetrahedron, and this gives the molecule a twisted shape rather than a perfect tetrahedron.

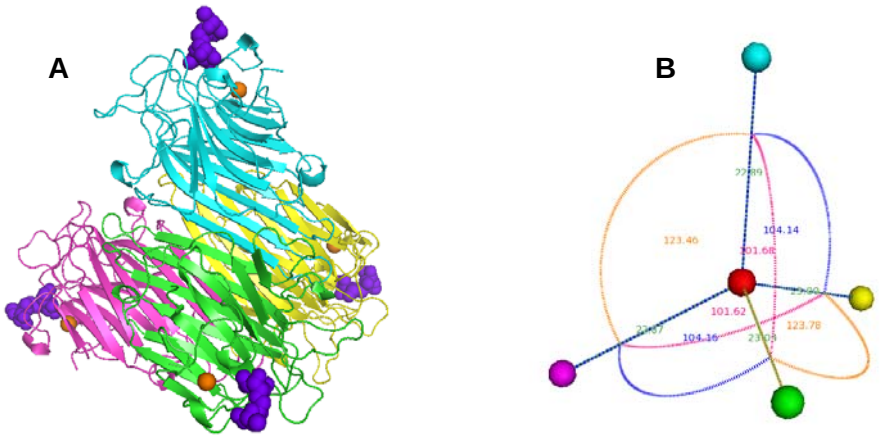


Fig. 6.1. (A) A perfect tetrahedral arrangement of subunits as observed in the tetramer of ConA (PDB code 1QDC). The carbohydrate ligand, man(α1-6)man(α1-O)methyl binds at the four corners of the tetrahedron. **(B)** Distances and angles between centers of mass of four subunits in the case of 1QDC. The distances are shown in green. Pink: angles between the subunits of the same dimer; blue: angles between the adjacent subunits of different dimers and orange: angles between the diagonally opposite subunits. Same color coding has been used for all subsequent diagrams of centers of mass.

6.4.2. Sandwiched dimer of dimers; two perpendicularly placed dimers

In this type also, two protein monomers associate to form a dimer and two dimers associate to form a tetramer. However, these dimers are placed roughly perpendicular to each other, that is, if each dimer is considered to be enclosed in a box; the two boxes appear perpendicular to each other. This was thought to be a distortion of the tetrahedral shape. The angles between monomers of the same dimer are reduced to ~80^0 and those between the adjacent monomers of two dimers range between 100-120^0. The angles between two oppositely placed monomers have values in the range of 140 to 160^0. An example of such arrangement is the indole-pyruvate decarboxylase from *Enterobacter cloacae* (PDB code 1OVM; Schutz *et al.*, 2003) shown in Fig. 6.2. (A), (B) and (C). However, in another such protein, SecB from *E. coli* (PDB code 1QYN; Dekker *et al.*, 2003), the angles between monomers of the same dimer are ~100^0 while those between the adjacent monomers of two dimers are ~80^0 (Fig. 6.2. (C), (D) and (E)). A schematic representation of this kind of arrangement is shown in Fig. 6.2. (G). This kind of arrangement also produces a 222 symmetry in the molecule, and it may be having the same biological significance as that of tetrahedral arrangement.

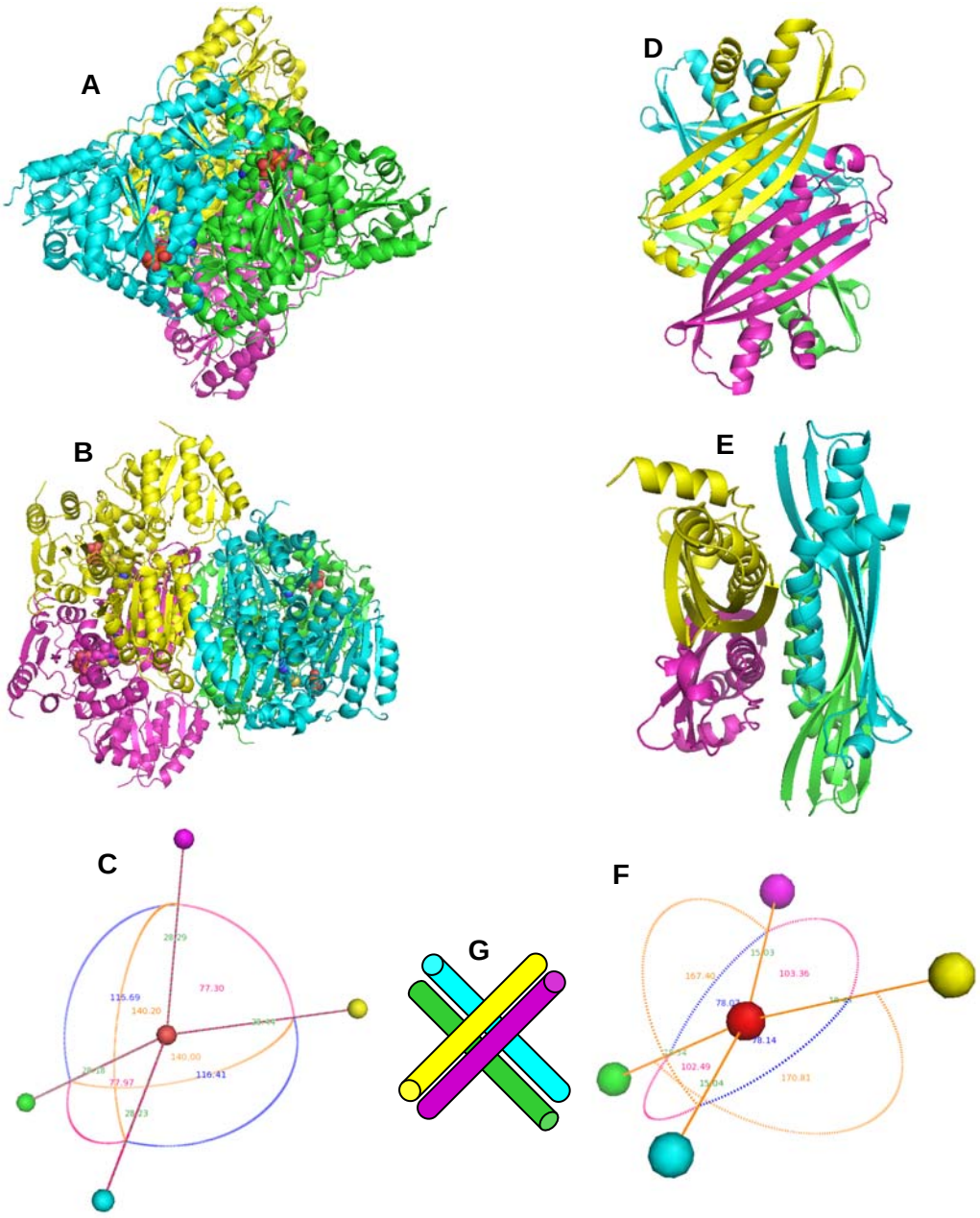


Fig. 6.2. (A) and **(B)** Indole-pyruvate decarboxylase from *Enterobacter cloacae* (PDB code 1OVM) and **(D)** and **(E)** SecB from *E. coli* (PDB code 1QYN) shown in two different orientations. These two proteins are dimer of dimers and the two dimers are roughly perpendicular to each other. Subunits colored in green and cyan form on dimer and those colored in magenta and yellow form another dimer. **(C)** and **(F)** Distances and angles between centers of mass of four subunits of 1OVM and 1QNY respectively. **(G)** Schematic representation of the dimers placed perpendicular to each other.

6.4.3. Planar assembly

This kind of arrangement is characterized by the presence of all four subunits of the protein in a single plane, with any two adjacent subunits facing in the opposite directions. This makes the diagonally oppositely placed subunits to face in the same direction. This can be considered as a flattened tetrahedral shape. From solvent accessibility calculations, it is assured that even in this type, the protein first dimerizes and the dimers associate to form tetramers. The angles between two monomers of the same dimer range from 70 to 80^0, while those between the adjacent monomers of different dimers have values of ~100^0. However, the angles between the diagonally opposite monomers have values very close to 180^0, which are responsible for the flattened or planar shape of the molecule. This type of arrangement also shows the point group 222 symmetry between the subunits. One such example is the enzyme Penicillin V acylase from *Bacillus sphaericus* (PDB code 2PVA, Suresh *et al.*, 1999), shown in Fig. 6.3. (A) and (B); (C) and (D) show similar kind of arrangement in TenA homolog from *Pyrobaculum aerophilum* (PDB code 2GM8), an all-alpha protein, whereas in (E), the distances and angles between centers of mass of four subunits have been shown and (F) displays a schematic representation of the molecules, showing the substrate binding sites face up.

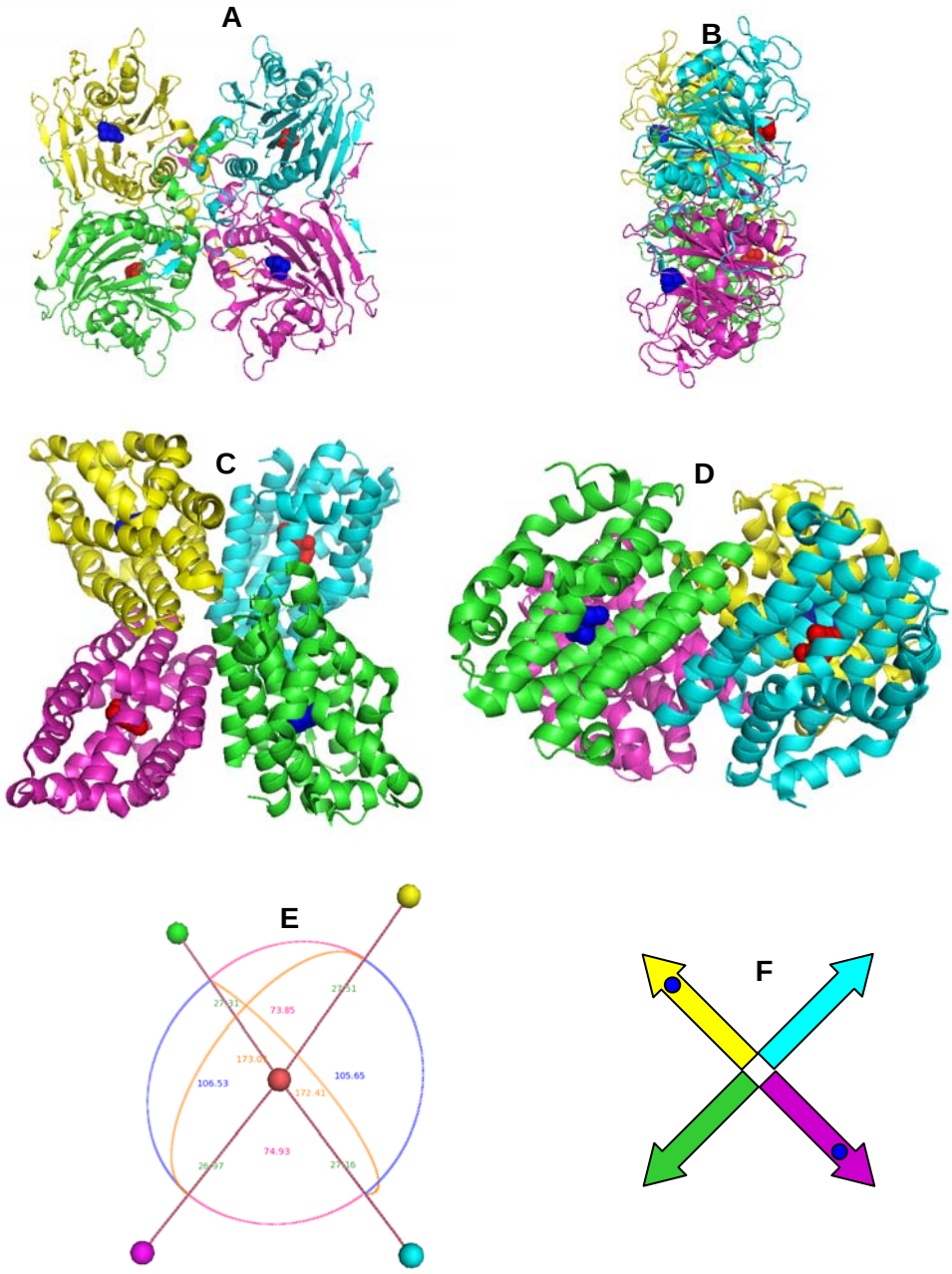


Fig. 6.3. (A) Top view and **(B)** side view of the enzyme Penicillin V acylase from *Bacillus sphaericus* (PDB code 2PVA). The bound dithiane diol molecules are shown as space filling models. The ligand-binding sites of any two adjacent subunits, which are same as substrate-binding sites, lie on the opposite faces.
(C) Top view and **(D)** side view of TenA homolog from *Pyrobaculum aerophilum* (PDB code 2GM8), an all-alpha protein. **(E)** Distances and angles between centers of mass of four subunits. **(F)** Schematic representation of the molecules, showing the substrate binding sites face up.

6.4.4. Planar closed molecules

The three types of subunit arrangements in tetrameric proteins described so far show point group 222 symmetry. The fourth kind of arrangement, in which all the four subunits lie in a plane and face in the same direction, shows a 4-fold symmetry between the subunits. Due to this, a "closed" tetramer is formed. The angles between any two adjacent monomers are nearly 90^0 and those between subunits oppositely placed are almost equal to the 180^0 and in two opposite directions. Many membrane-bound proteins, such as in ion channels and cell surface enzymes show this kind of arrangement, for example the potassium channel from *Streptomyces lividans* (PDB code 1BL8, Doyle *et al.*, 1998) which is shown in Fig. 6.4. (A) and (B); in (C) the distances and angles between centers of mass are shown; a schematic representation of this kind of arrangement is shown in (D). (E) shows a potassium channel embedded in the membrane lipid bilayer (reproduced from biop.ox.ac.uk/www/lj2001/sansom/sansom_1.jpg).

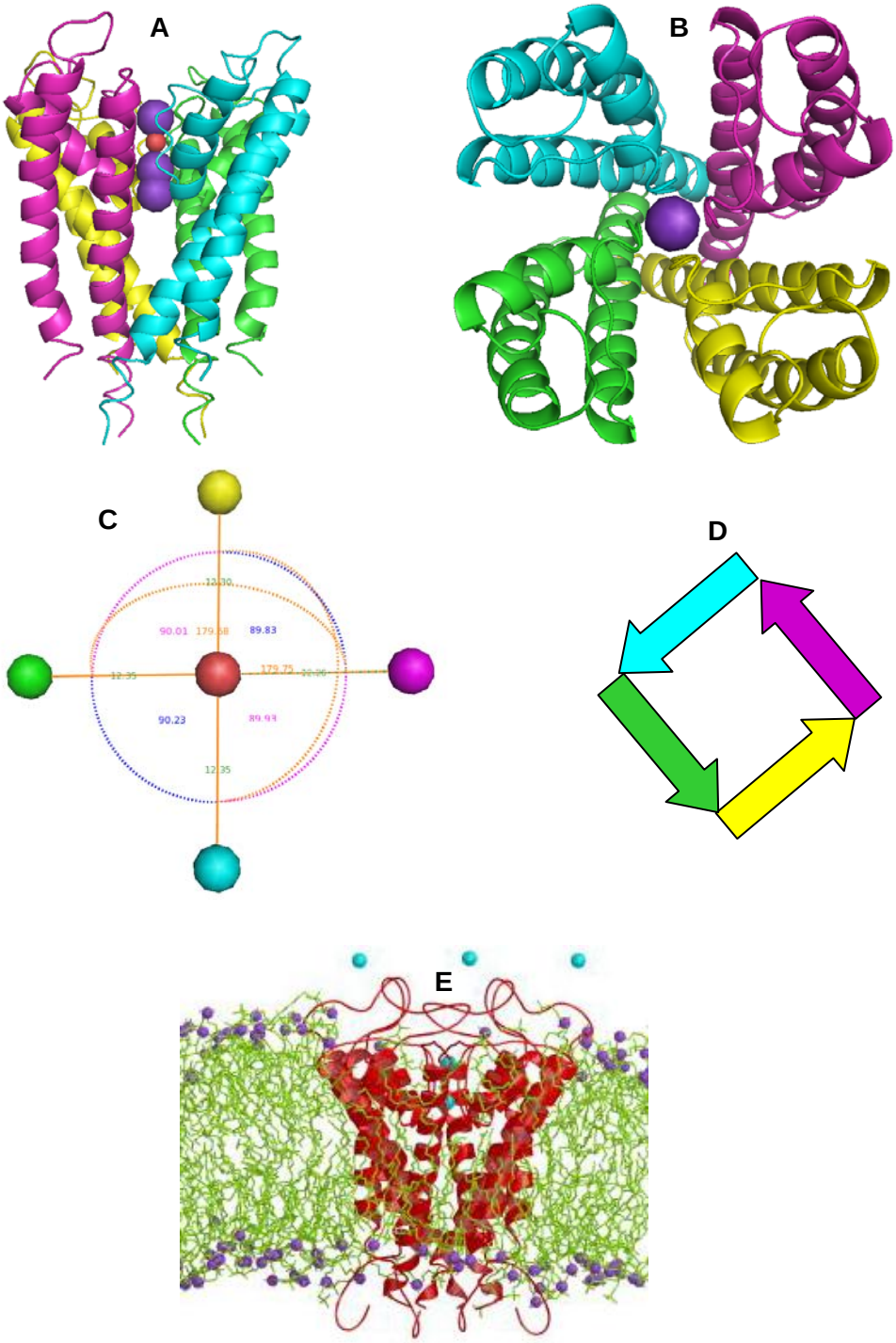


Fig. 6.4. (A) Side view and **(B)** Top view of potassium channel from *Streptomyces lividans* (PDB code 1BL8). The potassium ions and a water molecule have been shown as spheres. **(C)** Distances and angles between centers of mass of four subunits. **(D)** Schematic representation of the molecule exhibiting a four-fold symmetry. **(E)** Potassium channel in the membrane lipid bilayer. (reproduced from biop.ox.ac.uk/www/lj2001/sansom/sansom_1.jpg)

6.4.5. Tetrameric arrangements not belonging to the patterns described

Although most of the homotetrameric molecules could be grouped into one of the above mentioned four categories, some molecules could not be categorized. Most of such molecules, when analyzed in detail, were found to be wrongly labeled as homotetramers in their respective PDB files. Such molecules were discarded from further analysis. However, some of the molecules were found to be genuinely homotetrameric, still displayed a quaternary structure which could not be fitted into any of the above mentioned categories. They did not show any other recognizable pattern as well. The arrangements of subunits in some such molecules have been described below.

Peanut lectin **(PDB code 2PEL)**

This legume lectin, despite sharing sequence as well as secondary and tertiary structure similarity with other legume lectins, shows a very peculiar, "open" quaternary structure. It contains two identical dimers, each having a two-fold symmetry in its subunits, however, at the quaternary structure level; the molecule does not show any 222 or 4-fold symmetry (Fig. 6.5. (A) and (B)) (Banerjee *et al.*, 1994). This unusual structure was also responsible for the difficulty in solving its structure (Vijayan, 2007).

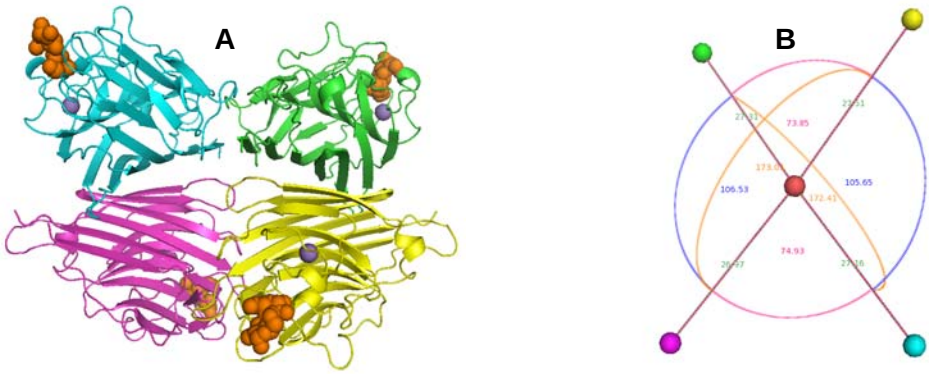


Fig. 6.5. (A) Quaternary structure of peanut lectin (pdb CODE: 2PEL).
(B) Distances and angles between centers of mass of four subunits.

DNA binding proteins

Several DNA binding proteins such as lactose operon repressor protein (LacR) from *E. coli* (PDB code 1TLF; Friedman *et al.*, 1995) form a very peculiar quaternary structure, which consists of two dyad-symmetric dimers which are nearly parallel to each other. Due to this, all four DNA binding domains of intact LacR are placed on the same side of the tetramer. This results in a deep, V-shaped cleft between the two dimers. An antiparallel four helix bundle which is formed from four C-terminal helices, one contributed by each monomer functions as a tetramerization domain (Fig. 6.6. (A) and (B)). While binding to DNA, the tethered dimers of this protein broaden by ~12^0 and the dimers twist by ~8^0 (Lewis *et al.*, 1996). Removing the *C*-terminal helix which assists in oligomerization of the protein, the affinity of the dimer remains the same (Brenowitz *et al.*, 1991); however, the induction ratios decrease (Oehler *et al.*, 1990).

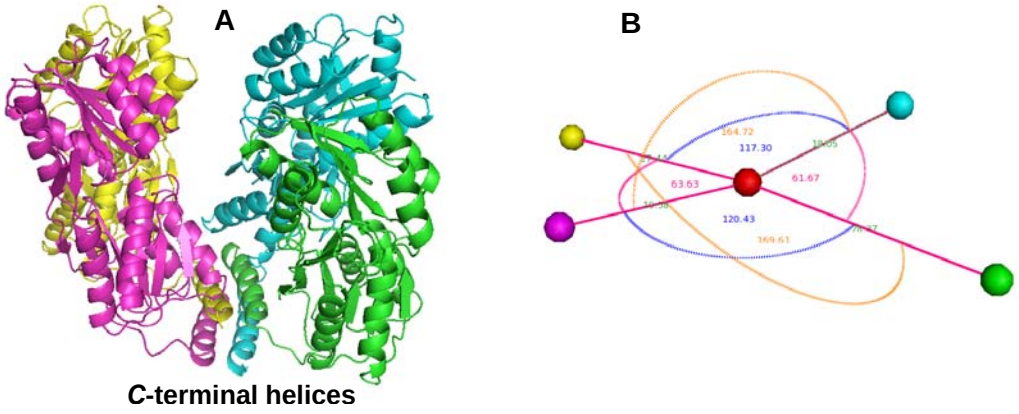


Fig. 6.6. (A) Quaternary structure of the lactose operon repressor protein (PDB code 1TLF). The *C*-terminal helices of all four monomers are involved in the tetramerisation of the molecule. **(B)** Distances and angles between centers of mass of four subunits.

***λ phage transcription activator protein CII* (PDB code 1XWR)**

This is another DNA-binding protein which binds to a unique direct-repeat sequence T-T-G-C-N6-T-T-G-C, observed in three phage promoters it activates. The tetramer is formed from dimers but does not exhibit any closed symmetry (Fig. 6.7. (A)). Arrangement of centers of mass of four subunits is also peculiar for this protein (Fig. 6.7. (B)). Here also, the tetramerization is achieved by formation of a four-helix bundle, each helix contributed by a monomer. The unusual quaternary structure of this protein allows it to place the helix-turn-helix motifs of two of the four CII subunits for interaction with successive major grooves of B-DNA, from one face of DNA and helps to identify a direct repeat DNA sequence rather than the inverted repeat sequence (Datta *et al.*, 2005).

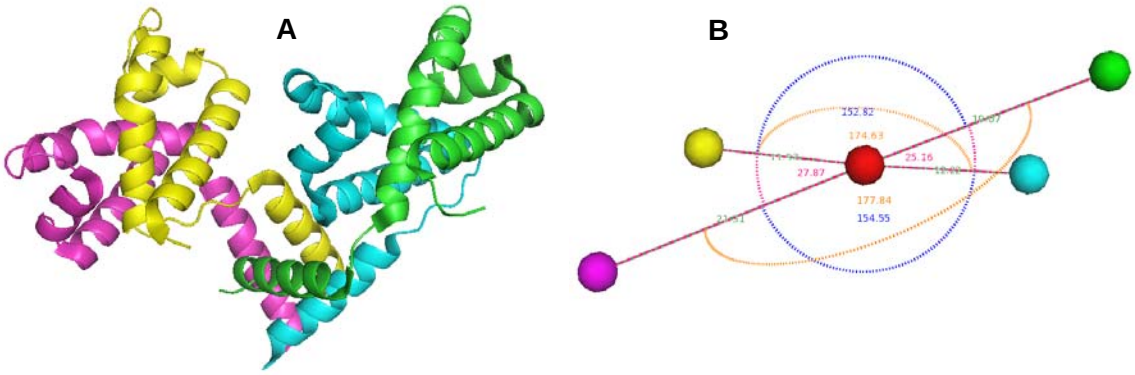


Fig. 6.7. (A) λ phage CII protein (PDB code 1XWR). This protein also has four helices involved in tetramerisation. **(B)** Distances and angles between centers of mass of four subunits, depicting their quite unusual arrangement.

***Mycobacterium tuberculosis* D-3-Phosphoglycerate Dehydrogenase (PGDH)**

The crystal structure of this enzyme (PDB code 1YGY, Dey *et al.*, 2005) consists of a dimeric asymmetric unit, made of two identical subunits, each consisting of four domains. However, in one of the two subunits there is a rotation of ~180^0 around a hinge region connecting two of the four domains. This introduces significant asymmetry in the dimer. Two such asymmetric units associate to form a biologically active tetramer (Fig. 6.6. (A)). Distances and angles between centers of mass of four subunits are shown in (B). This asymmetric arrangement leads to the formation of two different and distinct domain interfaces between identical domains in the asymmetric unit as well as introduces asymmetry in the substrate binding sites, which might have a role in the activity and regulation of the enzyme (Dey *et al.*, 2005).

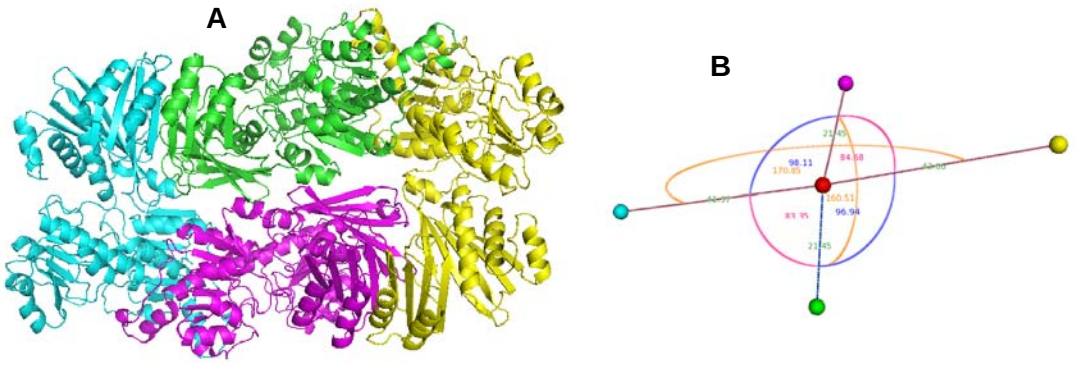


Fig. 6.6. (A) Unusual quaternary structure of *M. tuberculosis* PGDH. All four subunits have identical primary structure and consist of four domains each; however, in the subunits colored in cyan and yellow, two of the four domains are flipped by almost 180^0 around a hinge region. This introduces the asymmetry in the dimer.
(B) Distances and angles between centers of mass of four subunits, which form corners of an approximate rhombus.

6.4.6. Biological significance

In the case of most of the proteins exhibiting a tetrahedral shape, the binding sites are located at the four corners of the tetrahedron (Fig. 6.1. (A)). This reduces the steric hindrance between the ligand molecules and hence increases efficiency of the molecule. This could be the possible reason for the tetrahedral or the distorted tetrahedral shape being the most commonly observed feature of most of the homotetrameric molecules.

As seen in the case of molecules with tetrahedral shape, planar molecules which have their adjacent subunits facing in the opposite directions also have their binding sites placed at the maximum distance from each other, which reduces the steric hindrance between the ligand molecules. Enzymes belonging to this category have an additional benefit. Since their binding sites are placed on two opposite sides, substrate approaching from any side encounters the active site and hence the efficiency of the molecule increases.

The arrangement of monomers involving a four fold symmetry and hence closed planar pattern seems to be favoured by membrane bound proteins like aquaporins, plastocyanines and potassium channel proteins, or DNA binding proteins like RUVA. The reason could be that this kind of arrangement gives polarity to the molecule due to which all the hydrophobic part of the molecule gets buried in the membrane and the hydrophilic part remains exposed.

6.4.7. Correlation of subunit arrangement with the crystal systems

As the homotetrameric protein structures were also grouped according to their space groups, the prevalence of each type in particular space group was evident. As expected, the largest number of molecules displayed a tetrahedral or distorted tetrahedral shape in almost all of the crystal systems. Other types of quaternary arrangements which display the 222 symmetry, namely planar molecules with adjacent subunits facing in opposite directions were also observed almost in all crystal systems. Molecules that are sandwiched dimers of dimers, with both dimers roughly perpendicular to each other, are mainly observed in orthorhombic space groups..

In triclinic and monoclinic space groups, the asymmetric unit is always a tetramer, or sometimes, even two tetramers. In orthorhombic space groups, monomeric or dimeric asymmetric units are also observed which are located at special position and the crystallographic symmetry operations give rise to the functional tetramer. Occurrence of monomeric/dimeric asymmetric units is also common in space groups with two 2-fold axes belonging to tetragonal and hexagonal crystal systems, such as $P4_x22$ and $P6_x22$, where "x" denoted the screw axis. Trigonal space groups with a 2-fold axis, such as $P3_x21$ or $P3_x12$ may show presence of dimeric asymmetric unit placed at the special position. All these conditions give rise to a 222 symmetry in the molecule.

Planar closed arrangement of subunits involving a 4-fold symmetry is relatively rare. This type is mainly observed in the tetragonal or cubic space groups, where in most cases, the asymmetric unit is a monomer and the biological tetrameric molecule can be generated using the symmetry operations. Rarely, this kind of arrangement is seen in orthorhombic or monoclinic space groups as well, but then the asymmetric unit is always a tetramer.

Table 6.1 enlists all crystal systems with the prevalent types of quaternary arrangements observed in each.

Table 6.1. The seven crystal systems with different tetrameric arrangements observed in them.

Crystal System	Symmetry in tetramer	No. of molecules in asymmetric unit	Prevalent Type of subunit arrangement
Triclinic	Mostly 222	4	Tetrahedral/Distorted tetrahedral
Monoclinic	Mostly 222	4, 8	Tetrahedral/Distorted tetrahedral
			Planar 1 *
Orthorhombic	Mostly 222	1, 2, 4, 8	Tetrahedral/Distorted tetrahedral
			Planar 1
			Twisted perpendicular
Trigonal	222	4 in space groups without 2-fold axes, 2, 4 in space groups with 2-fold axes	Tetrahedral/Distorted tetrahedral
			Planar 1
Hexagonal	222	4 in space groups without 2-fold axes, 1, 2, 4 in space groups with 2-fold axes	Tetrahedral/Distorted tetrahedral
			Planar 1
Tetragonal	222 or 4-fold	1, 2, 4	Planar 2 #
			tetrahedral/distorted tetrahedral
			Planar 1
Cubic	222 or 4-fold	1, 2, 4	Planar 2
			Planar 1
			Tetrahedral/Distorted tetrahedral

*** Planar 1:** Planar molecules in which adjacent subunits face in the opposite directions; exhibiting 222 symmetry.

Planar 2: Closed planar molecules displaying 4-fold symmetry.

6.5. Conclusions

Study of unique homotetrameric protein structures reported in PDB revealed four types of prevalent arrangements of subunits in the tetramer. Three of these four types showed point group 222 symmetry in their subunits, while the fourth type exhibited a 4-fold symmetry. While the 222 symmetry is commonly observed in many proteins including lectins and most of the enzymes, the four-fold symmetry is restricted to most of the membrane bound proteins like ion channels and certain membrane bound enzymes. Some other unusual quaternary structures were also observed in this study, which did not form any group of their own. Such unusual arrangements could be correlated with the specific biological activity of the concerned protein.

References

The abbreviations used for journal full forms are from http://apps.isiknowledge.com/WoS/help/A_abrvjt.html.

- Ahmed HU, Blakeley MP, Cianci M, Cruickshank DWJ, Hubbard JA, Helliwell JR, 2007. The determination of protonation states in proteins. *Acta Cryst* D**63**:906-922.
- Albani JR, 2004. Structure and dynamics of macromolecules: absorption and fluorescence studies. Elsevier Science.
- Ali MH, Imperiali B, 2005. Protein oligomerization: How and why, *Bioorgan Med Chem* **13**:5013-5020.
- Allen AK, Bolwell GP, Brown DS, Sidebottom C, Slabas AR, 1996. Potato lectin: a three-domain glycoprotein with novel hydroxyproline-containing sequences and sequence similarities to wheat-germ agglutinin. *Int J Biochem Cell Biol* **28**:1285-1291.
- Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummeunk-Peters BG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG, 1979. The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information**.** *Acta Cryst* B**35**:2331-2339.
- Allen FH, Davies JE, Galloy JJ, Johnson O, Kennard O, Macrae CF, Mitchell EM, Mitchell GF, Smith JM, Watson DG, 1991. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J Chem Inf Comput Sci* **31**:187-204.
- Allen FH, Motherwell WDS, 2002. Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry. *Acta Cryst* B**58**:407-422.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
- Anand K, Schulte A, Fujinaga K, Scheffzek K, Geyer M, 2007. Cyclin Box Structure of the P-TEFb Subunit Cyclin T1 Derived from a Fusion Complex with EIAV Tat. *J Mol Biol* **370**:826-836.
- Argos P, 1988. An investigation of protein subunit and domain interfaces. *Protein Eng* **2**:101-113.

- Balzarini J, Neyts J, Schols D, Hosoya M, Van Damme E, Peumans W, De Clercq E, 1992. The mannose-specific plant lectins from *Cymbidium* hybrid and *Epipactis helleborine* and the (*N*-Acetylglucosamine)n-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication in vitro. *Antiviral Res* **18**:191-207.
- Balzarini J, Schols D, Neyts J, Van Damme E, Peumans W, and De Clercq E, 1991. α-(1–3)- and α-(1–6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob Agents Chemother* **35**:410-416.
- Banerjee R, Dhanaraj V, Mahanta SK, Surolia A, Vijayan M, 1991. Preparation and X-ray characterization of four new crystal forms of jacalin, a lectin from *Artocarpus integrifolia*. *J Mol Biol* **221**:773-776.
- Banerjee R, Mande SC, Ganesh V, Das K, Dhanaraj V, Mahanta SK, Suguna K, Surolia A, Vijayan M, 1994. Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *P Natl Acad Sci USA*. **91**:227–231.
- Barbieri S, Batelli G, Stirpe F, 1993. Ribosome inactivating proteins from plants. *Biochim Biophy Acta* **1154**:237-282.
- Barondes SH, 1988. Bifunctional properties of lectins: lectins redefined. *Trends Biochem Sci* **13**:480–482.
- Beintema JJ, 1994. Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett* **350**:159-163.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, 2000. The Protein Data Bank. *Nucleic Acids Res* **28**:235-242.
- Bernardo A, Calmanovici CE, Miranda EA, 2005. Observance of polymorphic behavior during dissolution of insulin and lysozyme. *Braz J Chem Eng* **22**:331-339.
- Bernstein J, 2002. Polymorphism in molecular crystals. Oxford University Press, New York.
- Bhuptawat H, Folkard GK, Chaudhari S, 2007. Innovative physico-chemical treatment of wastewater incorporating *Moringa oleifera* seed coagulant. *J Hazard Mater* **142**:477–482.

- Blakeley MP, Kalb-Gilboa AJ, Helliwell JR, Myles DAA, 2004. The 15-K neutron structure of saccharide-free concanavalin A. *P Natl Acad Sci USA* **101**:16405-16410.
- Blow D, 2002. Outline of Crystallography for Biologists. Oxford University Press.
- Blow DM, Crick FHC, 1959. The treatment of errors in the isomorphous replacement method. *Acta Cryst* **12**:794-802.
- Blow DM, Rossmann MG, 1961. The single isomorphous replacement method. *Acta Cryst* **14**:1195-1202.
- Bogoeva VP, Radeva MA, Atanasova LY, Stoitsova SR, Boteva RN, 2004. Fluorescence analysis of hormone binding activities of wheat germ agglutinin. *Biochim Biophys Acta* **1698**:213-218.
- Boistelle R, Astier JP, 1988. Crystallization mechanisms in solution. *J Cryst Growth* **90**:14-30.
- Bostwick DE, Dannehofer JM, Skaggs MI, Lister RM, Larkins BA, Thompson GA, 1992. Pumpkin phloem lectin genes are specifically expressed in companion cells. *Plant Cell* **4**:1539-1548.
- Bouckaert J, Dewallef Y, Poortmans F, Wyns L, Loris R, 2000a. The structural features of concanavalin A governing non-proline peptide isomerization. *J Biol Chem* **275**:19778-19787.
- Bouckaert J, Hamelryck TW, Wyns L, Loris R, 1999. The crystal structures of Man(α1-3)Man(α1-O)Me and Man(α1-6)Man(α1-O)Me in complex with concanavalin A. *J Biol Chem* **274**:29188-29195.
- Bouckaert J, Loris R, Poortmans F, Wyns L, 1995. Crystallographic structure of metal-free concanavalin A at 2.5 Å resolution. *Proteins* **23**:510-524.
- Bouckaert J, Loris R, Wyns L, 2000b. Zinc/calcium- and cadmium/cadmium-substituted concanavalin A: interplay of metal binding, pH and molecular packing. *Acta Cryst* D**56**:1569-1576.
- Bouckaert J, Poortmans F, Wyns L, Loris R, 1996. Sequential structural changes upon zinc and calcium binding to metal-free Concanavalin A. *J Biol Chem* **271**:16144-16150.

- Bourne Y, Roig-Zamboni V, Barre A, Peumans WJ, Astoul CH, Van Damme EJ, Rouge P, 2004. The crystal structure of the *Calystegia sepium* agglutinin reveals a novel quaternary arrangement of lectin subunits with a β-prism fold. *J Biol Chem* **279**:527-533.
- Bourne Y, Zamboni V, Barre A, Peumans WJ, Van Damme EJ, Rouge P, 1999. *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure* **7**:1473-1482.
- Boyd WC, Reguera RM, 1949. Studies on haemagglutinins present in seeds of some representatives of the family Leguminoseae. *J Immunol* **62**:333-339.
- Boyd WC, Shapleigh E, 1954. Specific precipitating activity of plant agglutinins (lectins). *Science* **119**:419.
- Bray D, Lay S, 1997. Computer-based analysis of the binding steps in protein complex formation. *P Natl Acad Sci USA* **94**:13493-13498.
- Brenowitz M, Mandal N, Pickar A, Jamison E, Adhya S, 1991. DNA-binding properties of a lac repressor mutant incapable of forming tetramers. *J Biol Chem* **266**:1281-1288.
- Brewin NJ, Kardailsky IV, 1997. Legume lectins and nodulation by *Rhizobium. Trends Plant Sci* **2**:92-98.
- Britton D, 1972. Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. *Acta Cryst* A**28**:296-297.
- Brock CP, Dunitz JD, 1994. Towards a Grammar of Crystal Packing. *Chem Mater* **6**:1118-1127.
- Broekaert WF, Marien W, Terras FRG, De Belle MFC, Proost P, Van Damme J, Dillen L, Claeys M, Rees SB, Vanderleyden J, Cammuet BPA, 1992. Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. *Biochemistry* **31**:4308-4314.
- Broekaert WF, Peumans WJ, 1986. Lectin release from seeds of *Datura stramonium* and interference of the *Datura stramonium* lectin with bacterial motility. **In***:* Lectins, Biology,

Biochemistry, Clinical Biochemistry. **5**:57–66, Bøg-Hansen TC, Van Driessche E, (Eds) W. De Gruyter, Berlin.

- Candy L, Van Damme EJM, Peumans WJ, Bouaouiche LM, Erard M, Rouge P, 2003. Structural and functional characterization of the GalNAc/Gal-specific lectin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* (Lib.) de Bary. *Biochem Biophys Res Co* **308**:396-402.

- Cardamone M, Puri NK, 1992. Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochem J* **282**:589-593.

- Carr PD, Gustin SE, Church AP, Murphy JM, Ford SC, Mann DA, Woltring DM, Walker I, Ollis DL, Young IG, 2001. Structure of the complete extracellular domain of the common β subunit of the human GM-CSF, IL-3, and IL-5 receptors reveals a novel dimer configuration. *Cell* **104**:291-300.

- Cavallini D, Graziani MT, Dupre S, 1966. Determination of disulphide groups in proteins. *Nature* **212**:294-295.

- Chayen NE, 1998. Comparative studies of protein crystallization by vapor diffusion and microbatch. *Acta Cryst* D**54**:8-15.

- Chipman DM, Grisaro V, Sharon N, 1967. The binding of oligosaccharides containing *N*-acetylglucosamine and *N*-acetylmuramic acid to lysozyme. *J Biol Chem* **242**:4388-4394.

- Chothia C, Janin J, 1975. Principles of protein-protein recognition. *Nature* **256**:705-708.

- Chrispeels MJ, Raikhel NV, 1991. Lectins, lectin genes, and their role in plant defense. *Plant Cell* **3**:1-9.

- Chuang P-H, Lee C-W, Chou J-Y, Murugan M, Shieh B-J, Chen H-M, 2007. Anti-fungal activity of crude extracts and essential oil of *Moringa oleifera* Lam. *Bioresource Technol* **98**:232–236.

- Collaborative Computational Project, Number 4, 1994. The *CCP*4 suite: programs for protein crystallography. *Acta Cryst* D**50**:760-763.

- Collinge DB, Kragh KM, Mikkelsen JD, Nielsen KK, Rasmussen U, Vad K, 1993. Plant chitinases. *Plant J* **3**:31-40.
- Crosio MP, Janin J, Jullien M, 1992. Crystal packing in six crystal forms of pancreatic ribonuclease. *J Mol Biol* **228**:243-251.
- Crosio MP, Rodier F, Jullien M, 1990. Packing forces in ribonuclease crystals. *FEBS Lett* **271**:152-156.
- Czapla TH, Lang BA, 1990. Effect of plant lectins on the larval development of European corn borer (Lepidoptera: Pyralidae) and southern corn rootworm (Coleoptera:Chrysomelidae). *J Econ Entomol* **83**:2480-2485.
- Daniel E, Weber G, 1966. Cooperative Effects in Binding by Bovine Serum Albumin. I. The Binding of 1-anilino-8-naphthalenesulfonate. Fluorimetric Titrations. *Biochemistry* **5**:1893-1900.
- Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA, 1997. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* **28**:494-514.
- Datta AB, Panjikar S, Weiss MS, Chakrabarti P, Parrack P, 2005. Structure of λ CII: Implications for recognition of direct-repeat DNA by an unusual tetrameric organization. *P Natl Acad Sci USA* **102**:11242-11247.
- Deacon A, Gleichmann T, Kalb(Gilboa) AJ, Price H, Raftery J, Bradbrook G, Yariv J, Helliwell JR, 1997. The structure of Concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94 Å resolution. *J Chem Soc Faraday Trans* **93**:4305-4312.
- Dekker C, de Kruijff B, Gros P, 2003. Crystal structure of SecB from *Escherichia coli*. *J Struct Biol* **144**:313-319.
- DeLano WL, 2002. The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org.
- Derewenda ZS, 2004. Rational protein crystallization by mutational surface engineering. *Structure* **12**:529–535.

- Dey S, Grant GA, Sacchettini JC, 2005. Crystal Structure of *Mycobacterium tuberculosis* D-3-Phosphoglycerate Dehydrogenase: EXTREME ASYMMETRY IN A TETRAMER OF IDENTICAL SUBUNITS. *J Biol Chem* **280**:14892-14899.
- Dhanaraj V, Patanjali SR, Surolia A, Vijayan M, 1988. Preparation and preliminary X-ray studies of two crystal forms of the anti-T lectin from jackfruit (*Artocarpus integrifolia*). *J Mol Biol* **203**:1135-1136.
- Dharkar PD, Anuradha P, Gaikwad SM, Suresh CG, 2006. Crystallization and preliminary characterization of a highly thermostable lectin from *Trichosanthes dioica* and comparison with other *Trichosanthes* lectins. *Acta Cryst* F**62**:205-209.
- Dhuna V, Bains JS, Kamboj SS, Singh J, Shanmugavel, Saxena AK, 2005. Purification and characterization of a lectin from *Arisaema tortuosum* Schott having in-vitro anticancer activity against human cancer cell lines. *J Biochem Mol Biol* **38**:526-532.
- Dhuna V, Kamboj SS, Kaur A, Saxena AK, Bhide SV, Shanmugavel, Singh J, 2007. Characterization of a lectin from *Gonatanthus pumilus* D. Don having anti-proliferative effect against human cancer cell lines. *Protein Pept Lett* **14**:71-78.
- Diaz C, Melchers LS, Hooykaas PJJ, Lugtenberg BJJ, Kijne JW, 1989. Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. *Nature* **338**:579-581.
- Diederichs K, Karplus PA, 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**:269-275.
- Dimasi N, Sawicki MW, Reineck LA, Li Y, Natarajan K, Margulies DH, Mariuzza RA, 2002. Crystal structure of the Ly49I natural killer cell receptor reveals variability in dimerization mode within the Ly49 family. *J Mol Biol* **320**:573-585.
- Dimick SM, Powell SC, Mcmahon SA, Moothoo DN, Naismith JH, Toone EJ, 1999. On the Meaning of Affinity: Cluster Glycoside Effects and Concanavalin A. *J Am Chem Soc* **121**:10286-10296.
- Dokland T, Freedom and restraint: themes in virus capsid assembly. *Structure* **8**:R157-R162.

- Donohue J, 1985. Revised space-group frequencies for organic compounds. *Acta Cryst* A**41**:203-204.
- Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo, A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R, 1998. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* **280**:69-77.
- Drenth J, 1994. Principles of protein X-ray crystallography. Springer-Verlag New York, Inc.
- Drenth J, Haas C, 1998. Nucleation in protein crystallization. *Acta Cryst* D**54**:867-872.
- DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F, 1956. Colorimetric method for determination of sugars and related substances. *Anal Chem* **28**:350–356.
- Ducruix A, Giege R, 1999. Crystallization of nucleic acids and proteins - A practical approach. Second edition, Oxford University Press, Oxford.
- Edelman GM, Cunningham BA, Reeke GN Jr, Becker JW, Waxdal MJ, Wang JL, 1972. The covalent and three-dimensional structure of concanavalin A. *P Natl Acad Sci USA* **69**:2580-2584.
- Eilert U, Wolters B, Nahrstedt A, 1981. The antibiotic principle of seeds of *Moringa oleifera* and *Moringa stenopetala*. *Planta Med* **42**:55-61.
- Elfstrand M, 1898. Über blutkörperchenagglutinierende Eiweisse. **In**: Görberdorfer Veröffentlichungen a. Band I. Kobert R, (Ed) Enke, Stuttgart, Germany.
- Emmerich C, Helliwell JR, Redshaw M, Naismith JH, Harrop SJ, Raftery J, Kalb AJ, Yariv J, Dauter Z, Wilson KS, 1994. High-resolution structures of single-metal-substituted concanavalin A: the Co,Ca-protein at 1.6 Å and the Ni,Ca-protein at 2.0 Å. *Acta Cryst* D**50**:749-756.
- Emsley P, Cowtan K, 2004. Coot: Model-Building Tools for Molecular Graphics. *Acta Cryst* D**60**:2126-2132.
- Endo Y, Mitsui K, Motizuki M, Tsurrugi K, 1987. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and characteristics of the modification in 28S ribosomal RNA caused by the toxins. *J Biol chem* **262**: 5908-5912.
- Escalante CR, Yie J, Thanos D, Aggarwal AK, 1998. Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature* **391**:103-106.

- Etzler ME, 1985. Plant lectins: molecular and biological aspects. *Ann Rev Plant Physiol* **36**:209-234.
- Etzler ME, 1986. Distribution and function of plant lectins. **In**: The Lectins: Properties, Functions and Applications in Biology and Medicine. 371–435. Liener IE, Sharon N, Goldstein IJ, (Eds), Academic Press, Orlando, FL.
- Etzler ME, 1992. Plant lectins: molecular biology, synthesis and function. In *Glycoconjugates: composition, structure and function*. Allen HJ, Kisailus EC, (Eds), Marcel Dekker, New York, pp. 521-539.
- Evans PR, 2001. Rotations and rotation matrices. *Acta Cryst* D**57**:1355-1359.
- Ezeamuzle IC, Ambadederomo AW, Shode FO, Ekwebelem SC, 1996. Antiinflammatory effects of *Moringa oleifera* root extract. *Int J Pharmacogn* **34**:207-212.
- Fahey JW, 2005. *Moringa oleifera*: A review of the medical evidence for its nutritional, therapeutic, and prophylactic properties. Part 1. Trees for Life Journal, 1:5.
- Faizi S, Siddiqui BS, Saleem R, Siddiqui S, Aftab K, Gilani AH, 1995. Fully acetylated carbonate and hypotensive thiocarbamate glycosides from *Moringa oleifera*. *Phytochemistry* **38**:957–963.
- Feher G, Kam Z, 1985. Nucleation and growth of protein crystals; general principles and assays. *Methods Enzymol* **114**:77-111.
- Fraenkel-Conrat H, 1957. Methods for investigating the essential groups for enzyme activity. *Methods Enzymol* **4**:247-269.
- Frank J, 2003. Electron Microscopy of Functional Ribosome Complexes. *Biopolymers* **68**:223-233.
- Friedman AM, Fischmann TO, Steitz TA, 1995. Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* **268**:1721-1727.
- Fuglie LJ (Ed.), 2001. The Miracle Tree- The multiple attributes of *Moringa*. Technical Centre for Agricultural and Rural Cooperation (CTA)/Church World Service (CWS), New York.

- Gaastra W, Svennerholm A-M, 1996. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC) *Trends Microbiol*, **4**:444-452.
- Gaikwad SM, Khan MI, 2006. Binding of T-antigen disaccharides to *Artocarpus hirsuta* lectin and jacalin are energetically different. *Photochem Photobiol* **82**:1315-1318.
- Garman EF, Schnieder TR, 1997. Macromolecular Crystallography. *J Appl Crystallogr* **30**:211-237.
- Gassenschmidt U, Jany KD, Tauscher B, Niebergall H, 1995. Isolation and characterization of a flocculating protein from *Moringa oleifera* Lam. *Biochim Biophys Acta* **1243**:477-481.
- Gassenschmidt U, Jany KD, Tauscher B, Niebergall H, 1995. Isolation and characterization of a flocculating protein from *Moringa oleifera* Lam. *Biochim Biophys Acta* **1243**:477-481.
- Gavezzotti A, 2007. A Solid-State Chemist's View of the Crystal Polymorphism of Organic Compounds. *J Pharm Sci* **96**:2232-2241.
- Gayathri P, Banerjee M, Vijayalakshmi A, Azeez S, Balaram H, Balaram P, Murthy MRN, 2007. Crystal structure of triosephosphate isomerase (TIM) from *Methanocaldococcus jannaschii*. *Acta Cryst* D**63**:206-220.
- Gegg CV, Roberts DD, Sege IH, Etzler ME, 1992. Characterization of the Adenine Binding Sites of Two *Dolichos biflorus* Lectins. *Biochemistry* **31**: 6938-6942.
- Ghebremichael KA, Gunaratna KR, Henriksson H, Brumer H, Dalhammar G, 2005. A simple purification and activity assay of the coagulant protein from *Moringa oleifera* seeds. *Water Res* **39**:2338-2344.
- Glusker JP, Trueblood KN, 1985. Crystal Structure Analysis: A Primer. Second edition, Oxford University Press.
- Goel M, Damai RS, Kaur KJ, Maiya BG, Swamy MJ, Salunke DM, 2005. Crystal structures of the PNA-porphyrin complex in the presence and absence of lactose: mapping the conformational changes on lactose binding, interacting surfaces, and supramolecular aggregations. *Biochemistry* **44**:5588-5596.

- Goel M, Jain D, Kaur KJ, Kenoth R, Maiya BG, Swamy MJ, Salunke DM, 2001. Functional equality in the absence of structural similarity. An added dimension to molecular mimicry. *J Biol Chem* **276**:39277-39281.
- Gohlke U, Gomis-Ruth FX, Crabbe T, Murphy G, Docherty AJ, Bode W, 1996. The C-terminal (haemopexin-like) domain structure of human gelatinase A (MMP2): structural implications for its function. *FEBS Lett* **378**:126-130.
- Gold AM, Fahrney D, 1964. Sulfonyl Fluorides as Inhibitors of Esterases. 11. Formation and Reactions of Phenylmethanesulfonyl α-Chymotrypsin. *Biochemistry* **3**:783-791.
- Goldstein IJ, Hughes RC, Monsigny M, Osawa T, Sharon N, 1980. What should be called a lectin? *Nature* **285**:66.
- Goldstein IJ, Poretz RD, 1986. Isolation, physicochemical characterization, and carbohydrate- binding specificity of lectins. **In:** The Lectins, Properties, Functions, and Applications in Biology and Medicine. 33–247. Liener IE, Sharon N, Goldstein IJ, (Eds) Academic Press, Orlando, FL.
- Goodsell DS, 1991. Inside a living cell. *Trends Biochem Sci* **16**:203-206.
- Goodsell DS, Olson AJ, 2000. Structural symmetry and protein function. *Annu Rev Bioph Biom* **29**:105-153.
- Green DW, Ingram VM, Perutz MF, 1954. The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method. *P Roy Soc Lond A Mat* **225**:287-307.
- Grinvald A, Steinberg IZ, 1974. On the analysis of fluorescence decay kinetics by the method of least-squares. *Anal Biochem* **59**: 583-598.
- Guillot J, Konska G, 1997. Lectins in higher fungi. *Biochem Syst Eco* **25**:203-230.
- Gurjar MM, Gaikwad SM, Solakhe SG, Mukherjee S, Khan MI, 2000. Growth inhibition and total loss of reproductive potential in *Trilobium cattaneum* by *Artocarpus hisuta* lectin. *Invertebr Reprod Dev* **38**:95-98.
- Guzmán-Partida AM, Robles-Burgueño MR, Ortega-Nieblas M, Vázquez-Moreno I, 2004. Purification and characterization of complex carbohydrate specific isolectins from wild legume

seeds: *Acacia constricta* is (vinorama) highly homologous to *Phaseolus vulgaris* lectins. *Biochimie* **86**:335-342.

- Habash J, Raftery J, Nuttall R, Price HJ, Wilkinson C, Kalb AJ, Helliwell JR, 2000. Direct determination of the positions of the deuterium atoms of the bound water in concanavalin A by neutron Laue crystallography. *Acta Cryst* D**56**:541-550.

- Habeeb AFSA, 1966. Determination of free amino groups in proteins by trinitrobenzenesulfonic acid. *Anal Biochem* **14**:328-336.

- Habeeb AFSA, 1972. Reaction of protein sulfhydryl groups with Ellman's reagent. *Methods Enzymol* **25**:457-464.

- Hamelryck TW, Loris R, Bouchaert J, Dao-Thi MH, Strecker G, Imberty A, Fernandez E, Wyns L, Etzler ME, 1999. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J Mol Biol* **286**:1161-1177.

- Hamelryck TW, Dao-Thi MH, Poortmans F, Chrispeels MJ, Wyns L, Loris R, 1996. The crystallographic structure of phytohemagglutinin-L. *J Biol Chem* **271**:20479-20485.

- Hamelryck TW, Moore JG, Chrispeels MJ, Loris R, Wyns L, 2000. The role of weak protein-protein interactions in multivalent lectin-carbohydrate binding: crystal structure of cross-linked FRIL. *J Mol Biol* **299**:875-883.

- Hamodrakas SJ, Kanellopoulos PN, Pavlou K, Tucker PA, 1997. The crystal structure of the complex of concanavalin A with 4'-methylumbelliferyl-α-D-glucopyranoside. *J Struct Biol* **118**:23-30.

- Harata K, Akiba T, 2007. Effect of a sodium ion on the dehydration-induced phase transition of monoclinic lysozyme crystals *Acta Cryst* D**63**:1016-1021.

- Hardman KD, Ainsworth, CF, 1972. Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**:4910–4919.

- Harker D, 1956. The determination of the phases of the structure factors of noncentrosymmetric crystals by the method of double isomorphous replacement. *Acta Cryst* **9**:1-9.

- Harrop SJ, Helliwell JR, Wan TC, Kalb AJ, Tong L, Yariv J, 1996. Structure solution of a cubic crystal of concanavalin A complexed with methyl alpha-D-glucopyranoside. *Acta Cryst* D**52**:143-155.
- Hendrickson WA, 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**:51-58.
- Hendrickson WA, Horton JR, Murthy HM, Pahler A, Smith, JL, 1989. Multiwavelength anomalous diffraction as a direct phasing vehicle in macromolecular crystallography. *Basic Life Sci* **51**:317-324.
- Hendrickson WA, Teeter MM, 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**:107-113.
- Hester G, Kaku H, Goldstein IJ, Wright CS, 1995. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat Struct Biol* **2**: 472–479.
- Higgins MK, Eswaran J, Edwards PC, Schertler GF, Hughes C, Koronakis V, 2004.Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug effllux protein TolC. *J Mol Biol* **342**:697-702.
- Higgins TJV, Beach LR, Spencer D, Chandler PM, Randall PJ, Blagrove RJ, Kortt AA, Guthrie RE, 1987. cDNA and protein sequence of a major pea seed albumin (PA 2 : Mr≈26 000). *Plant Mol Biol* **8**:37-45.
- Hilder VA, Powell KS, Gatehouse AMR, Gatehouse JA, Gatehouse LN, Shi, Y, Hamilton WDO, Merryweather A, Newell C, Timans JC, Peumans WJ, Van Damme EJM, Boulter D, 1995. Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Res* **4**:18-25.
- Hoffman GR, Nassar N, Cerione RA, 2000. Structure of the Rho family GTP-binding protein Cdc42 in complex with the multifunctional regulator RhoGDI. *Cell* **100**:345-356.
- http://www.accelrys.com/products/quanta/
- http://www.cermav.cnrs.fr/lectines/
- http://www.imperial.ac.uk/research/animallectins/default.html

- http://www.rcsb.org/pdb
- http://www.ruppweb.org/Xray/Patterson/Native_Patterson.htm
- Huesing JE, Murdock LL, Shade RE, 1991. Effect of wheat germ isolectins on development of cowpea weevil. *Phytochemistry* **30**:785-788.
- Hulse JH, 1991. Nature, composition and utilization of grain legumes. 11-27. **In:** Uses of tropical Legumes: Proceedings of a Consultants' Meeting, 27-30 March 1989, ICRISAT Center. ICRISAT, Patancheru, A.P. 502 324, India.
- Inokuti M, Hirayama F, 1965. Influence of Energy Transfer by the Exchange Mechanism on Donor Luminescence. *J Chem Phys* **43**:1978-1989.
- Jain D, Kaur K, Sundaravadivel B, Salunke DM, 2000. Structural and functional consequences of peptide-carbohydrate mimicry. Crystal structure of a carbohydrate-mimicking peptide bound to concanavalin A. *J Biol Chem* **275**:16098-16102.
- Jain D, Kaur KJ, Salunke DM, 2001a. Enhanced binding of a rationally designed peptide ligand of concanavalin A arises from improved geometrical complementarity. *Biochemistry* **40**:12059-12066.
- Jain D, Kaur KJ, Salunke DM, 2001b. Plasticity in protein-peptide recognition: crystal structures of two different peptides bound to concanavalin A. *Biophys J* **80**:2912-2921.
- Janin J, Milles S, Chothia C, 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* **204**:155-164.
- Jelsch C, Longhi S, Cambillau C, 1998. Packing forces in nine crystal forms of cutinase. *Proteins* **31**:320-333.
- Jones S, Thornton JM, 1996. Principles of protein-protein interactions. *P Natl Acad Sci USA* **93**:13-20.
- Kamemura K, Furuichi Y, Umekawa H, Takahashi T, 1993. Purification and characterization of novel lectins from Great Northern bean, *Phaseolus vulgaris* L. *BBA-Gen Subjects* **1158**:181-188.

- Kaneda Y, Whittier RF, Yamanaka H, Carredano E, Gotoh M, Sota H, Hasegawa Y, Shinohara Y, 2002. The high specificities of *Phaseolus vulgaris* erythro- and leukoagglutinating lectins for bisecting GlcNAc or *β*1-6-linked branch structures, respectively, are attributable to loop B. *J Biol Chem* **277**:16928-16935.

- Kanellopoulos PN, Pavlou K, Perrakis A, Agianian B, Vorgias CE, Mavrommatis C, Soufi M, Tucker PA, Hamodrakas SJ, 1996a. The crystal structure of the complexes of concanavalin A with 4'-nitrophenyl-alpha-D-mannopyranoside and 4'-nitrophenyl-alpha-D-glucopyranoside. *J Struct Biol* **116**:345-355.

- Kanellopoulos PN, Tucker PA, 1996b. A Triclinic Crystal Form of the Lectin Concanavalin A. *J Struct Biol* **117**:16-23.

- Kantardjieff KA, Höchtl P, Segelke BW, Tao F-M, Rupp B, 2002. Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility. *Acta Cryst* D**58**:735-743.

- Kantardjieff KA, Rupp B, 2003. Matthew's coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**:1865-1871.

- Katre UV, Gaikwad SM, Bhagyawant SS, Deshpande UD, Khan MI, Suresh CG, 2005. Crystallization and preliminary X-ray characterization of a lectin from *Cicer arietinum* (chickpea). *Acta Cryst* F**61**:141-143.

- Katzin BJ, Collins EJ, Robertus JD, 1991. Structure of Ricin A-chain at 2.5 Å. *Proteins* **10**:251-259.

- Kaur A, Kamboj SS, Singh J, Saxena AK, Dhuna V, 2005a. Isolation of a novel *N*-acetyl-D-lactosamine specific lectin from *Alocasia cucullata* (Schott.) *Biotechnol Lett* **27**:1815-1820.

- Kaur A, Singh J, Kamboj SS, Sexana AK, Pandita RM, Shamnugavel M, 2005b. Isolation of an *N*-acetyl-D-glucosamine specific lectin from the rhizomes of *Arundo donax* with antiproliferative activity. *Phytochemistry* **66**:1933-1940.

- Kaur KJ, Khurana S, Salunke DM, 1997. Topological analysis of the functional mimicry between a peptide and a carbohydrate moiety. *J Biol Chem* **272**:5539-5543.

- Kaur M, Singh K, Rup PJ, Kamboj SS, Saxena AK, Sharma M, Bhagat M, Sood SK, Singh J, 2006a. A tuber lectin from *Arisaema jacquemontii* Blume with anti-insect and anti-proliferative properties. *J Biochem Mol Biol* **39**:432-440.

- Kaur M, Singh K, Rup PJ, Saxena AK, Khan RH, Ashraf MT, Kamboj SS, Singh J, 2006b. A tuber lectin from *Arisaema helleborifolium* Schott with anti-insect activity against melon fruit fly, *Bactrocera cucurbitae* (Coquillett) and anti-cancer effect on human cancer cell lines. *Arch Biochem Biophys* **445**:156-165.

- Kaur N, Dhuna V, Kamboj SS, Agrewala JN, Singh J, 2006c. A novel antiproliferative and antifungal lectin from *Amaranthus viridis* Linn seeds. *Protein Pept Lett* **13**:897-905.

- Khan F, Ahmad A, Khan MI, 2007. Purification and characterization of a lectin from endophytic fungus *Fusarium solani* having complex sugar specificity. *Arch Biochem Biophys* **457**:243-251.

- Kieliszewski MJ, Showalter AM, Leykam JF, 1994. Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). *Plant J* **5**:849-861.

- Kobe B, Center RJ, Kemp BE, Poumbourios P, 1999. Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *P Natl Acad Sci USA* **96**:4319-4324.

- Kobiler D, Mirelman D, 1981. Adhesion of *E. histolytica* trophozoites to monolayers of human cells. *J Infect Dis* **144**:539-546.

- Kocourek J, Horejsi V, 1983. A note on the recent discussion on definition of the term ‘lectin’. **In**: Lectins: Biology, Biochemistry and Clinical Biochemistry. **3**:3–6. Bøg- Hansen, T. C. and Spengler, G. A., Eds., Walter de Gruyter, Berlin, Germany.

- Kolberg J, Michaelsen TE, Sletten K, 1983. Properties of a lectin purified from the seeds of *Cicer arietinum*. *Hoppe-Seyler's Z Physiol Chem* **364**:655-664.

- Komath SS, Kavitha M, Swamy MJ, 2006. Beyond carbohydrate binding: new directions in plant lectin research. *Org Biomol Chem* **4**:973-988.

- Komath SS, Nadimpalli SK, Swamy MJ, 1996. Purification in high yield and characterisation of the galactose-specific lectin from the seeds of snake gourd (*Trichosanthes anguina*). *Biochem Mol Biol Int* **39**:243-252.
- Kondrashov DA, Zhang W, Aranda IV R, Stec B, Phillips Jr GN, 2007. Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. *Proteins* **70**:353-362.
- Konska G, 1985. *Lectin in higher fungi*. Thesis, Faculty of Pharmaceutical Sciences, Cracow.
- Konska G, 2006. Lectins of higher fungi (Macromycetes)-their occurrence, physiological role, and biological activity. *Int J Med Mushr* **8**:19-36.
- Kozak M, Borek D, Janowski R, Jaskólski M, 2002. Crystallization and preliminary crystallographic studies of five crystal forms of *Escherichia coli* L-asparaginase II (Asp90Glu mutant). *Acta Cryst* D**58**:130-132.
- Kulik V, Hartmann E, Weyand M, Frey M, Gierl A, Niks D, Dunn MF, Schlichting I, 2005. On the structural basis of the catalytic mechanism and the regulation of the alpha subunit of tryptophan synthase from *Salmonella typhimurium* and BX1 from maize, two evolutionarily related enzymes. *J Mol Biol* **352**:608-620.
- Kumar AM, Timm DE, Neet KE, Owen WG, Peumans WJ, Rao AJ, 1993. Characterization of the lectin from the bulbs of *Eranthis hyemalis* (winter aconite) as an inhibitor of protein synthesis. *J Biol Chem* **268**:25176-25183.
- Kumar SG, Appukuttan PS, Basu D, 1982. α-D-galactose specific lectin from jack fruit (*Artocarpus integrifolia*) seed. *J Biosci* **4**:257-261.
- Laemmli UK, 1970. Cleavage of structural proteins during the assembly of the Head of Bacteriophage T4. *Nature* **227**:680-685.
- Lakowicz EM, Weber G, 1973. Quenching of protein fluorescence by oxygen. Detection of structural fluctuations in proteins on the nanosecond time scale. *Biochemistry* **12**:4171-4179.
- Landsteiner K, Raubitschek H, 1907. Beobachtungen über Hämolyse und Hämagglutination. *Zentralbl Bakteriol Parasitenk Infektionskr Hyg Ab. 1: Orig* **45**:660–667.

- Larsen TA, Olson AJ, Goodsell DS, 1997. Morphology of protein-protein interfaces. *Structure* **6**:421-427.
- Lehrer SS, 1971. Solute perturbation of protein fluorescence. The quenching of tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry* **10**:3254-3263.
- Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P, 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**:1247-1254.
- Li H, Lawson CL, 1995. Crystallization and preliminary X-ray analysis of *Borrelia burgdorferi* outer surface protein A (OspA) complexed with a murine monoclonal antibody Fab fragment. *J Struct Biol* **115**:335-337.
- Libson AM, Gittis AG, Collier IE, Marmer BL, Goldberg GI, Lattman EE, 1995. Crystal structure of the haemopexin-like C-terminal domain of gelatinase A. *Nature Struct Biol* **2**:938-942.
- Lis H, Sharon N, 1998. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* **98**:637-674.
- López-Jaramillo FJ, González-Ramírez LA, Albert A, Santoyo-González F, Vargas-Berenguel A, Otálora F, 2004. Structure of concanavalin A at pH 8: bound solvent and crystal contacts. *Acta Cryst* D**60**:1048-1056.
- Lord JM, Roberts LM, Robertus JD, 1994. Ricin: Structure, mode of action, and some current applications. *FASEB J* **8**:201-208.
- Loris R, Hamelryck T, Bouckaert J, Wyns L, 1998. Legume lectin structure. *Biochim Biophys Acta* **1383**:9-36.
- Loris R, Maes D, Poortmans F, Wyns L, Bouckaert J, 1996. A structure of the complex between Concanavalin A and methyl-3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside reveals two binding modes. *J Biol Chem* **271**:30614-30618.
- Lowry OH, Rosenburgh NJ, Farr AL, Randall RJ, 1951. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**:265-276.

- Maikokera R, Kwaambwa HM, 2007. Interfacial properties and fluorescence of a coagulating protein extracted from *Moringa oleifera* seeds and its interaction with sodium dodecyl sulphate. *Colloid Surface B* **55**:173-178.
- Makabe K, Tereshko V, Gawlak G, Yan S, Koide S, 2006. Atomic-resolution crystal structure of *Borrelia burgdorferi* outer surface protein A via surface engineering. *Protein Sci* **15**:1907-1914.
- Malhotra RS, Pundir RPS, Slinkard AE, 1987. Genetic resources of chickpea. 67-81. **In:** The Chickpea. Saxena MC, Singh KB (Eds), C. A. B. International Cambrian News Ltd, Aberystwyth, UK.
- Manoj N, Ealick SE, 2003. Unusual space-group pseudosymmetry in crystals of human phosphopantothenoylcysteine decarboxylase. *Acta Cryst* D**59**:1762-1766.
- Manting EH, van Der Does C, Remigy H, Engel A, Driessen AJ, 2000. SecYEG assembles into a tetramer to form the active protein translocation channel. *EMBO J* **19**:852-861.
- Marchler-Bauer A, Bryant SH, 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**:W327-W331.
- Martinho JMG, Santos AM, Fedorov A, Baptista RP, Taipa MA, Cabral JMS, 2003. Fluorescence of the Single Tryptophan of Cutinase: Temperature and pH Effect on Protein Conformation and Dynamics. *Photochem Photobiol* **78**:15-22.
- Matsuura Y, Chernov AA, 2003. Morphology and the strength of intermolecular contacts in protein crystals *Acta Cryst* D**59**:1347-1356.
- Matthews BW, 1968. The Solvent Content of Protein Crystals. *J Mol Biol* **33**:491-497.
- McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ, 2005. Likelihood-enhanced fast translation functions. *Acta Cryst* D**61**:458-464.
- McPherson A, 1999. Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press.
- Merckel MC, Huiskonen JT, Bamford DH, Goldman A, Tuma R, 2005. The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* **18**:161-170.

- Miller S, 1989. The structure of interfaces between subunits of dimeric and tetrameric proteins. *Protein Eng* **3**:77-83.
- Miller S, Lesk AM, Janin J, Chothia C, 1987. The accessible surface area and stability of oligomeric proteins. *Nature* **328**:834-836.
- Moggach SA, Allan DR, Clark SJ, Gutmann MJ, Parsons S, Pulham CR, Sawyer L, 2006. High-pressure polymorphism in L-cysteine: the crystal structures of L-cysteine-III and L-cysteine-IV. *Acta Cryst* B**62**:296-309.
- Mohammed AS, Lai OM, Muhammad SKS, Long K, Ghazali HM, 2003. *Moringa oleifera*, potentially a new source of oleic acid-type oil for Malaysia. **3:**137–140. **In:** Investing in Innovation 2003, Bioscience and Biotechnology, Hassan MA (Ed), Universiti Putra Malaysia Press, Serdang Press, Selongor, Malaysia.
- Monod J, Wyman J, Changeux J-P, 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol* **12**:88-118.
- Moothoo DN, Canan B, Field RA, Naismith JH, 1999. Man α1-2 Man α-OMe-Concanavalin A complex reveals a balance of forces involved in carbohydrate recognition. *Glycobiology* **9**:539-545.
- Moothoo DN, Naismith JH, 1998. Concanavalin A distorts the β-GlcNAc-(1→2)-Man linkage of β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man- (1→6)]-Man upon binding. *Glycobiology* **8**:173-181.
- Morton JF, 1991. The horseradish tree, *Moringa pterygosperma* (Moringaceae) - a boon to arid lands? *Econ Bot* **45**:318-333.
- Mueller-Dieckmann C, Panjikar S, Schmidt A, Mueller S, Kuper J, Geerlof A, Wilmanns M, Singh RK, Tucker PA, Weiss MS, 2007. On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Cryst* D**63**:366-380.

- Mueller-Dieckmann C, Panjikar S, Tucker PA, Weiss MS, 2005. On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. *Acta Cryst* D**61**:1263-1272.

- Mühlbauer FJ, Singh KB, 1987. Genetics of chickpea. 99-125. **In:** The Chickpea. Saxena MC, Singh KB, (Eds). CAB. International, Wallingford, Oxon, OX10 8DE, UK.

- Murakami A, Kitazono Y, Jiwajinda S, Koshimizu K, Ohigashi H, 1998. Niaziminin, a thiocarbamate from the leaves of *Moringa oleifera*, holds a strict structural requirement for inhibition of tumorpromoter- induced Epstein–Barr virus activation. *Planta Med* **64**:319-323.

- Murdock LL, Huesing JE, Nielsen SS, Pratt RC, Shade RE, 1990. Biological effects of plant lectins on the cowpea weevil. *Phytochemistry* **29**:85-89.

- Murzin AG, Lesk AM, Chothia C, 1992. β-trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors, interleukins-1β, and 1α and fibroblast growth factors. *J Mol Biol* **223**:531–543.

- Muyibi SA, Evison LM, 1995. *Moringa oleifera* seeds for softening hardwater. *Water Res* **29**:1099-1105.

- Naismith JH, Field RA, 1996. Structural basis of trimannoside recognition by concanavalin A. *J Biol Chem* **271**:972-976.

- Naismith JH, Emmerich C, Habash J, Harrop SJ, Helliwell JR, Hunter WN, Raftery J, Kalb AJ, Yariv J, 1994. Refined structure of concanavalin A complexed with methyl alpha-D-mannopyranoside at 2.0 A resolution and comparison with the saccharide-free structure. *Acta Cryst* D**50**:847-858.

- Naismith JH, Habash J, Harrop S, Helliwell JR, Hunter WN, Wan TC, Weisgerber S, Kalb AJ, Yariv J, 1993. Refined structure of cadmium-substituted concanavalin A at 2.0 Å resolution. *Acta Cryst* D**49**:561-571.

- Navaza J, Saludjian P, 1997. AMoRe: An automated molecular replacement program package. *Methods Enzymol* **276**:581-594.

- Ndabigengesere A, Narasiah KS, Talbot BG, 1995. Active agents and mechanism of coagulation of turbid waters using *Moringa oleifera. Water Res* **29**:703-710.
- Nesamani S, 1999. Medicinal Plants (vol. I). State Institute of Languages, Thiruvananthapuram, Kerala, India. 425.
- Nickell S, Beck F, Korinek A, Mihalache O, Baumeister W, Plitzko JM, 2007. Automated cryoelectron microscopy of ‘‘single particles’’ applied to the 26S proteasome. *FEBS Lett* **581**:2751-2756.
- Nooren IM, Thornton JM, 2003a. Diversity of protein-protein interactions. *EMBO J* **22**:3486-3492.
- Nooren IM, Thornton JM, 2003b. Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* **325**:991-1018.
- North A, 1965. The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. *Acta Cryst* **18**:212-216.
- Oehler S, Eismann ER, Krämer H, Müller-Hill B, 1990. The three operators of the lac operon cooperate in repression. *EMBO J* **9**:973-979.
- Ofek I, Doyle RJ, 1994. Bacterial Adhesion to Cells and Tissues; Chapman and Hall: London. 578.
- Ofek I, Sharon N, 1990. Adhesins as lectins: specificity and role in infection. *Curr Top Microbiol Immunol* **152**:91-113.
- Otwinowski Z, Minor W, 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**:307-326.
- Ovadi J, Libor S, Elodi P, 1967. Spectrophotometric determination of histidine in protein with diethylpyrocarbonate. *Acta Biochem Biophys* (Budapest) **2**:455-458.
- Padma P, Komath SS, Nadimpalli SK, Swamy MJ, 1999. Purification in high yield and characterisation of a new galactose-specific lectin from the seeds of *Trichosanthes cucumerina*. *Phytochemistry* **50**:363-371.

- Pari L, Kumar NA, 2002. Hepatoprotective activity of *Moringa oleifera* on antitubercular drug-induced liver damage in rats. *J Med Food* **5**:171-177.
- Park F, Gajiwala K, Eroshkina G, Furlong E, He D, Batiyenko Y, Romero R, Christopher J, Badger J, Hendle J. Lin J, Peat T, Buchanan S, 2004. Crystal structure of YIGZ, a conserved hypothetical protein from *Escherichia coli* k12 with a novel fold. *Proteins* **55**:775-777.
- Parkin S, Rupp B, Hope H, 1996. Atomic resolution structure of concanavalin A at 120 K. *Acta Cryst* D**52**:1161-1168.
- Parsons S, 2003. Introduction to twinning. *Acta Cryst D* **59**:1995-2003.
- Patterson AL, 1934. A Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review* **46**, 372.
- Paulson JC, 1987. in The Receptors, Conn PM, Ed. (Academic Press, New York, 1985), vol. 2, pp. 131-219; Wiley DC, Skehel JJ, 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem* **56**:365-394.
- Pedroche J, Yust MM, Lqari H, Megías C, Girón-Calle J, Alaiz M, Millán F, Vioque J, 2005. Chickpea pa2 albumin binds hemin. *Plant Sci* **168**:1109-1114.
- Petri WA Jr, Chapman MD, Snodgrass T, Mann BJ, Broman J, Ravdin JI, 1989. Subunit structure of the galactose and *N*-acetyl-D-galactosamine-inhibitable adherence lectin of *Entamoeba histolytica*. *J Biol Chem* **264**:3007-3012.
- Peumans WJ, De Ley M, Broekaert WF, 1984. An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. *FEBS Lett* **177**:99-103.
- Peumans WJ, Hause B, Van Damme EJM, 2000. The galactose-binding and mannose-binding jacalin-related lectins are located in different sub-cellular compartments. *FEBS Lett* **477**:186-192.
- Peumans WJ, Smeets K, Van Nerum K, Van Leuven F, Van Damme EJM, 1997a. Lectin and alliinase are the predominant proteins in the nectar from leek (*Allium porrum*) flowers. *Planta* **201**: 298–302.

- Peumans WJ, Van Damme EJM, 1995. Lectins as plant defense proteins. *Plant Physiol* **109**:347-352.
- Peumans WJ, Van Damme EJM, 1998. Plant lectins; Specific tools for the identification, isolation and characterization of O-linked glycans. *Critic Rev Biochem Mol*, **33**:209-258.
- Peumans WJ, Van Damme EJM, 1999. Seed lectins. In; *Seed proteins*. Shewry PR, Casey R, eds., Kluwer Academic publishers, The Netherlands, 657-683.
- Peumans WJ, Verhaert P, Pfuller U, Van Damme EJM, 1996. Isolation and characterization of a small chitin binding lectin from misteltoe (*Viscum album*). *FEBS Lett* **396**:261-265.
- Peumans WJ, Winter HC, Bemer V, Van Leuven F, Goldstein IJ, Truffa-Bachi P, Van Damme EJM, 1997b. Isolation of a novel plant lectin with an unusual specificity from *Calystegia sepium*. *Glycoconjugate J* **14**:259-265.
- Phelps DK, Speelman B, Post CB, 2000. Theoretical studies of viral capsid proteins. *Curr Opin Struc Biol* **10**:170-173.
- Piller V, Piller F, Cartron J-P, 1986. Isolation and characterization of an *N*-acetyl galactosamine specific lectin from *Salvia sclarea* seeds. *J Biol Chem* **261**:14069-14075.
- Powell HR, 1999. The Rossmann Fourier autoindexing algorithm in *MOSFLM. Acta Cryst* D**55**:1690-1695.
- Powers ET, Powers DL, 2003. A Perspective on Mechanisms of Protein Tetramer Formation. *Biophys J* **85**:3587-3599.
- Prabhu MM, Suguna K, Vijayan M, 1999. Variability in quaternary association of proteins with the same tertiary fold: A case study and rationalization involving legume lectins. *Proteins* **35**:58-69.
- QUANTA (Accelry, Inc.)
- Rahbé Y, Sauvion N, Febvay G, Peumans WJ, Gatehouse AMR, 1995. Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum. Entomol Exp Appl* **76**:143-155.

- Raikhel NV, Lee HI, Borekaert WF, 1993. Structure and function of chitin binding proteins. *Ann Rev Plant Phys* **44**:591-615.
- Randal M, Kossiakoff AA, 2000. The 2.0 Å structure of bovine interferon-c; assessment of the structural differences between Species. *Acta Cryst* D**56**:14-24.
- Rao KN, Suresh CG, Katre UV, Gaikwad SM, Khan MI, 2004. Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose. *Acta Cryst* D**60**: 1404-1412.
- Raszeja S, 1958. Über spezifische Phytagglutinine aus *Laccaria laccata* varietas *proxima*. *Z Arztl Fortbild* **14**:801-803.
- Ravaud S, Gouet P, Haser R, Aghajari N, 2003. Probing the role of divalent metal ions in a bacterial psychrophilic metalloprotease: binding studies of an enzyme in the crystalline state by x-ray crystallography. *J Bacteriol* **185**:4195-4203.
- Ray S, Ahmed H, Basu S, Chatterjee BP, 1993. Purification, characterization, and carbohydrate specificity of the lectin of *Ficus cunia*. *Carbohydr Res* **242**:247-263.
- Ray S, Chatterjee BP, 1995. Saracin: a lectin from *Saraca indica* seed integument recognizes complex carbohydrates. *Phytochemistry* **40**:643-649.
- Read SM, Northcote DH, 1983. Subunit structure and interactions of the phloem proteins of *Cucurbita maxima* (pumpkin). *Eur J Biochem* **134**:561-569.
- Redinbo MR, Stewart L, Champoux JJ, Hol WGJ, 1999. Structural flexibility in human topoisomerase I revealed in multiple non-isomorphous crystal structures. *J Mol Biol* **292**:685-696.
- Reeke Jr GN, Becker JW, Edelman GM, 1975. The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. *J Biol Chem* **250**:1525-1547.
- Renkonen KO, 1948. Studies on hemagglutinins present in seeds of some representatives of Leguminoseae. *Ann Med Exp Biol Fenn* **26**:66-72.

- Retailleau P, Colloc'h N, Vivarès D, Bonneté F, Castro B, Hajji ME, Prangé T, 2005. Urate oxidase from *Aspergillus flavus*: new crystal-packing contacts in relation to the content of the active site. *Acta Cryst* D**61**:218-229.
- Rhodes G, 2000. Crystallography Made Crystal Clear. Second edition, Academic Press.
- Rinderle SJ, Goldstein IJ, Matta KL, Ratcliffe RM, 1989. Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus caudatus*, that recognizes the T- (or cryptic T)- antigen. *J Biol Chem* **264**: 16123-16131.
- Rinderle SJ, Goldstein IJ, Remsen EE, 1990. Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds. *Biochemistry* **29**:10555-10561.
- Riordan JF, Wacker WEC, Vallee BL, 1965. *N*-Acetylimidazole: A Reagent for Determination of "Free" Tyrosyl Residues of Proteins. *Biochemistry* **4**:1758-1765.
- Roberts DD, Goldstein IJ, 1982. Hydrophobic Binding Properties of the Lectin from Lima Beans (*Phaseolus lunatus*) *J Biol Chemistry* **257**:11274-11277.
- Roberts DD, Goldstein IJ, 1983a. Binding of hydrophobic ligands to plant lectins: Titration with arylaminonaphthalenesulfonates. *Arch Biochem Biophys* **224**, 479-484.
- Roberts DD, Goldstein IJ, 1983b. Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J Biol Chem* **258**:13820-13824.
- Rossmann MG, Blow DM, 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst* **15**:24-31.
- Rozwarski DA, Swami BM, Brewer CF, Sacchettini JC, 1998. Crystal structure of the lectin from *Dioclea grandiflora* complexed with core trimannoside of asparagine-linked carbohydrates. *J Biol Chem* **273**:32818-32825.
- Rüdiger H, Gabius H-J, 2001. Plant lectins: Occurrence, biochemistry, functions and applications. *Glycoconjugate J* **18**:589-613.
- Rutenber E, Katzin BJ, Collins EJ, Mlsna D, Ready MP, Robertus JD, 1991. Crystallographic refinement of Ricin to 2.5 Å. *Proteins* **10**:240-250.

- Rutenber E, Robertus JD, 1991. Structure of Ricin B-chain at 2.5 Å resolution. *Proteins* **10**:260–269.
- Sakaguchi T, Tu Q, Pinto LH, Lamb RA, 1997. The active oligomeric state of the minimalistic influenza virus M2 ion channel is a tetramer. *P Natl Acad Sci USA* **94**:5000-5005.
- Sanders DA, Moothoo DN, Raftery J, Howard AJ, Helliwell JR, Naismith JH, 2001. The 1.2 Å resolution structure of the Con A-dimannose complex. *J Mol Biol* **310**:875-884.
- Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M, 1996. A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a β-prism fold. *Nat Struct Biol* **3**:596-603.
- Santos AFS, Argolo ACC, Coelho LCBB, Paiva PMG, 2005. Detection of water soluble lectin and antioxidant component from *Moringa oleifera* seeds. *Water Res* **39**:975-980.
- Sarma GN, Karplus PA, 2006. In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs. *Acta Cryst* D**62**:707-716.
- Sauerborn MK, Wright LM, Reynolds CD, Grossmann JG, Rizkallah PJ, 1999. Insights into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: the crystal structure at 2 Å resolution in complex with alpha1-3 mannobiose. *J Mol Biol* **290**:185-199.
- Schneider TR, Sheldrick GM, 2002. Substructure solution with SHELXD. *Acta Cryst* D**58**:1772-1779.
- Schultze B, Zimmer G, Herrlere G, 1993. Viral lectins for the detection of 9-O-Actylated Sialic acid on glycoproteins and glycolioids. **In:** Lectins and Glycobiology, 158-174. Gabius HJ, Gabius S, (Eds) Springer-Verlag Berlin, Heidelberg.
- Schutz A, Sandalova T, Ricagno S, Hubner G, Konig S, Schneider G, 2003. Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from *Enterobacter cloacae*, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. *Eur J Biochem* **270**:2312-2321.
- Sharma V, Surolia A, 1997. Analyses of Carbohydrate Recognition by Legume Lectins: Size of the Combining Site Loops and their Primary Specificity. *J Mol Biol* **267**:433-445.

- Sharon N, 1987. Bacterial lectins, cell-cell recognition and infectious disease. *FEBS Lett* **217**:145-157.
- Sharon N, Lis H, 1989. Lectins as cell recognition molecules. *Science* **246**:227-234.
- Sharon N, Lis H, 1990. Legume lectins-a large family of homologous proteins. *FASEB J* **4**:3198-3208.
- Shoham M, Yonath A, Sussman JL, Moult J, Traub W, Kalb AJ, 1979. Crystal structure of demetallized concanavalin A: the metal-binding region *J Mol Biol* **131**:137-155.
- Siegrist T, Fleming RM, Haddon RC, Laudise RA, Lovinger AJ, Katz HE, Bridenbaugh P, Davis DD, 1995. The crystal structure of the high-temperature polymorph of a–hexathienyl (a–6T/HT). *J Mater Res* **10**:2170-2173.
- Singh Bains J, Singh J, Kamboj SS, Nijjar KK, Agrewala JN, Kumar V, Kumar A, Saxena AK, 2005. Mitogenic and anti-proliferative activity of a lectin from the tubers of Voodoo lily (*Sauromatum venosum*). *Biochim Biophys Acta* **1723**:163-174.
- Singh J, Singh J, Kamboj SS, 2004. A novel mitogenic and antiproliferative lectin from a wild cobra lily, *Arisaema flavum*. *Biochem Bioph Res Co* **318**:1057–1065.
- Singh T, Wu JH, Peumans WJ, Rougé P, Van Damme EJM, Alvarez RA, Blixt O, Wu AM, 2006. Carbohydrate specificity of an insecticidal lectin isolated from the leaves of *Glechoma hederacea* (ground ivy) towards mammalian glycoconjugates. *Biochem J* **393**:331-341.
- Singha B, Adhya M, Chatterjee BP, 2007. Multivalent II [β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc] and Tα [β-D-Gal*p*-(1→3)-α-D-Gal*p*NAc] specific Moraceae family plant lectin from the seeds of *Ficus bengalensis* fruits. *Carbohydr Res* **342**:1034-1043.
- Sinha U, Brewer JM, 1985. A spectrophotometric method for quantitation of carboxyl group modification of proteins using Woodward's Reagent K. *Anal Biochem* **151**:327-333.
- Spande TF, Witkop B, 1967. Determination of the tryptophan content of proteins with *N*-bromosuccinimide. *Methods Enzymol* **11**:498-506.

- Sreerama N, Woody RW, 2000. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Anal Biochem* **287**:252-260.
- Srinivas VR, Bhanuprakash Reddy G, Ahmad N, Swaminathan CP, Mitra N, Surolia A, 2001. Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerisation. *Biochim Biophys Acta* **1527**:102-111.
- Stillmark H, 1888. Über Ricin ein giftiges Ferment aus den Samen von *Ricinus communis* L. und einige anderen Euphorbiaceen, *Inaugural Dissertation Dorpat*, Tartu.
- Sultan NAM, Kenoth R, Swamy MJ, 2004. Purification, physicochemical characterization, saccharide specificity, and chemical modification of a Gal/GalNAc specific lectin from the seeds of *Trichosanthes dioica*. *Arch Biochem Biophys* **432**:212-221.
- Sumner JB, Howell SF, 1936. The identification of the hemagglutinin of the Jack bean with concanavalin A. *J Bacteriol* **32**:227–237.
- Suresh CG, Pundle AV, SivaRaman H, Rao KN, Brannigan JA, McVey CE, Verma CS, Dauter Z, Dodson EJ, Dodson GG, 1999. Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat Struct Biol* **6**:414-416.
- Suseelan KN, Mitra R, Pandey R, Sainis KB, Krishna TG, 2002. Purification and characterization of a lectin from wild sunflower (*Helianthus tuberosus* L.) tubers. *Arch Biochem Biophys* **407**:241-247.
- Svensson-Ek M, Abramson J, Larsson G, Tornroth S, Brzezinski P, Iwata S, 2002. The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from *Rhodobacter sphaeroides*. *J Mol Biol* **321**:329-339.
- Swamy MJ, Sastry MVK, Surolia A, 1985. Prediction and comparison of the secondary structure of legume lectins. *J Biosci* **9**:203-212.
- Swarnamukhi PL, Sharma SK, Padala P, Surolia N, Surolia A, Suguna K, 2007. Packing and loop-structure variations in non-isomorphous crystals of FabZ from *Plasmodium falciparum*. *Acta Cryst* D**63**:458-464.

- Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY, 1995. Crystal structure of Abrin-A at 2.14 Å. *J Mol Biol* **250**:354-367.
- Takahashi K, 1968. The reaction of phenylglyoxal with arginine residues in proteins. *J Biol Chem* **243**:6171-6179.
- Tang L, Johnson JE, 2002. Structural biology of viruses by the combination of electron cryomicroscopy and X-ray Crystallography. *Biochemistry* **41**:11517-11524.
- Teng TY, 1990. Mounting crystals for macromolecular crystallography in a free-standing thin film. *J Appl Cryst* **23**:387-391.
- Thallapally PK, Jetti RKR, Katz AK, Carrell HL, Singh K, Lahiri K, Kotha S, Boese R, Desiraju GR, 2004. Polymorphism of 1,3,5-trinitrobenzene induced by a trisindane additive. *Angew Chem Int Edit* **43**:1149-1155.
- Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**:4673-4680.
- Transue TR, Smith AK, Mo H, Goldstein IJ, Saper MA, 1997. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* lectin. *Nat Struct Biol* **10**:779-783.
- Tronchin G, Esnault K, Sanchez M, Larcher G, Marot-Leblond A, Bouchara JP, 2002. Purification and partial characterization of a 32-kilodalton sialic acid-specific lectin from *Aspergillus fumigatus*. *Infect Immun* **70**:6891-6895.
- Tsaknis J, Lalas S, Gergis V, Dourtoglou V, Spilotis V, 1999. Characterization of *Moringa oleifera* variety Mbololo seed oil of Kenya. *J Agr Food Chem* **47**:4495-4499.
- Vagin A, Teplyakov A, 1997 MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30**:1022-1025.
- Van Damme EJM, Peumans WJ, Barre A, Rouge P, 1998a. Plant lectins: a composite of several distinct families of structurally and evolutionarily related proteins with diverse biological roles. *Crit Rev Plant Sci*, **17**:575-692.

- Van Damme EJM, Peumans WJ, Pusztai A, Bardocz S, 1998b. Handbook of Plant Lectins: Properties and Biomedical Applications. John Wiley & Sons, Chichester, UK.
- Van Damme EJM, Smeets K, Peumans WJ, 1995. The mannose-binding monocot lectins and their genes. **In**: Lectins: Biomedical Perspectives. 59-80. Pusztai A, Bardocz S, (Eds), Taylor and Francis, London, UK.
- Van der Mäsen LJG, 1987. *Cicer* L. Origin, history and taxonomy of chickpea. **In**: The Chickpea. 11-34. Saxena MC, Singh KB, (Eds), CAB International Cambrian News Ltd, Aberystwyth, UK.
- Van Parijs J, Broekaert WF, Goldstein IJ, Peumans WJ, 1991. Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. *Planta* **183**:258-262.
- Vaney MC, Broutin I, Retailleau P, Douangamath A, Lafont S, Hamiaux C, Prangé T, Ducruix A, Riès-Kautt M, 2001. Structural effects of monovalent anions on polymorphic lysozyme crystals. *Acta Cryst* D**57**:929-940.
- Vasta GR, Ahmed H, Odom EW, 2004. Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Curr Opin Struc Biol* **14**:617-630.
- Vega N, Pérez G, 2006. Isolation and characterisation of a *Salvia bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry* **67**:347-355.
- Verschueren KHG, Tyrrell R, Murshudov GN, Dodson EJ, Wilkinsona AJ, 1999. Solution of the structure of the cofactor-binding fragment of CysB: a struggle against non-isomorphism. *Acta Cryst* D**55**:369-378.
- Vesterberg O, 1972. Isoelectric focusing of proteins in polyacrylamide gels. *Biochim Biophys Acta* **257**:11-19.
- Vijayan M, 2007. Peanut lectin crystallography and macromolecular structural studies in India. *J Biosci* **32**:1059-1066.
- Vijayan M, Chandra N, 1999. Lectins. *Curr Opin Struc Biol* **9**:707-714.
- Vodkin LO, Rhodes PR, Goldberg RB, 1983. A lectin gene insertion has the structural features of a transposable element. *Cell* **34**:1023–1031.

- Walden H, Bell GS, Russel RJM, Siebers B, Hensel R, Taylor GL, 2001. Tiny TIM: A small tetrameric, hyperthermostable triosephosphate isomerase. *J Mol Biol* **306**:745-757.
- Waljuno K, Scholma RA, Beintema J, Mariono A, Hahn AM, 1975. Amino acid sequence of hevein. *Proc Int Rubber Conf, Kuala Lumpur* **2**: 518–531.
- Wang H, Ng TB, Liu Q, 2003. A novel lectin from the wild mushroom *Polyporus adusta*. *Biochem Biophys Res Co* **307**:535-539.
- Wang M-B, Boulter D, Gatehouse JA, 1994. Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of *Cucurbita pepo*. *Plant Mol Biol* **24**:159-170.
- Weber PC, 1991. Physical principles of protein crystallization. *Adv Prot Chem,* **41**:1-36.
- Weisgerber S, Helliwell JR, 1993. High resolution crystallographic studies of native concanavalin A using rapid Laue data collection methods and the introduction of a monochromatic large-angle oscillation technique (LOT). *J Chem Soc Faraday Trans* **89**:2667-2675.
- Weissbuch I, Torbeev VY, Leiserowitz L, Lahav M, 2005. Solvent effect on crystal polymorphism: why addition of methanol or ethanol to aqueous solutions induces the precipitation of the least stable beta form of glycine. *Angew Chem Int Edit* **44**:3226-3229.Wilson AJC, 1988. Space groups rare for organic structures. I. Triclinic, monoclinic and orthorhombic crystal classes. *Acta Cryst* A**44**:715-724.
- Wilson AJC, 1993. Kitajgorodskij's categories. *Acta Cryst* A**49**:210-212.
- Wlodawer A, Erickson JW, 1993. Structure-Based Inhibitors of HIV-1 Protease**.** *Annu Rev Biochem* **62**:543-585.
- Wöhler F, Liebig J, 1832. Untersuchungen über das Radikal der. Benzoesäure. *Ann Pharm* **3**:249-282.
- Wong JH, Wong CCT, Ng TB, 2006. Purification and characterization of a galactose-specific lectin with mitogenic activity from pinto beans. *Biochim Biophys Acta* **1760**:808-813.

- Wood SD, Wright LM, Reynolds CD, Rizkallah PJ, Allen AK, Peumans WJ, Van Damme EJ, 1999. Structure of the native (unligated) mannose-specific bulb lectin from *Scilla campanulata* (bluebell) at 1.7 Å resolution. *Acta Cryst* D**55**:1264-1272.

- Wright CS, 1990. 2.2 Å resolution structure analysis of two refined *N*-acetylneuraminyl-lactose-wheat germ agglutinin isolectin complexes. *J Mol Biol* **215**:635-651.

- Wright CS, Kaku H, Goldstein IJ, 1990. Crystallization and preliminary X-ray diffraction results of snowdrop (*Galanthus nivalis*) lectin. *J Biol Chem* **265**:1676-1677.

- Wright LM, Reynolds CD, Rizkallah PJ, Allen AK, Van Damme EJM, Donovan MJ, Peumans WJ, 2000. Structural characterisation of the native fetuin-binding protein *Scilla campanulata* agglutinin: a novel two-domain lectin. *FEBS Lett* **468**:19-22.

- Wukovitz SW, Yeates TO, 1995. Why protein crystals favour some space-groups over others. *Nat Struct Mol Biol* **2**:1062-1067.

- Yeates TO, 1997. Detecting and overcoming crystal twinning. *Methods Enzymol* **276**:344-358.

- Young NM, Oomen RP, 1992. Analysis of sequence variation among legume lectins. A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J Mol Biol* **228**:924-934.

- Zhang XJ, Wozniak JA, Matthews BW, 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* **250**:527-552.

- Zhang Z, Qian M, Huang Q, Jia Y, Tang Y, Wang K, Cui D, Li M, 2001a. Crystal structure of the complex of concanavalin A and tripeptide. *J Protein Chem* **20**:59-65.

- Zhang Z, Qian M, Huang Q, Jia Y, Tang Y, Wang K, Cui D, Li M, 2001b. Crystal structure of the complex of concanavalin A and hexapeptide. *J Protein Chem* **20**:423-429.

Appendix

Refinement of the structure of *Artocarpus hirsuta* lectin at 2.5 Å resolution

Seeds of *Artocarpus hirsuta*, a plant belonging to Moraceae family, contain a galactose specific lectin (Gurjar *et al.*, 2000) which is a member of the jacalin-type plant lectin family, showing the characteristic *β*-prism I fold. The lectin was crystallized in two orthorhombic and one hexagonal forms (Rao *et al.*, 1999) and the structure could be solved for the two orthorhombic forms (Rao, 2002).

This appendix describes the crystallization, data collection and refinement using improved resolution data for one of the orthorhombic forms of *A. hirsuta* lectin. Although the crystallization and data collection had already been reported by Rao *et al.* (1999), the data was available only up to 2.8 Å resolution. A fresh crystallization experiment set using the previously known conditions yielded a crystal which diffracted up to 2.5 Å resolution. This dataset with improved resolution was used to refine the structure.

Experimental details

A. hirsuta lectin (purified and concentrated up to 70 mg ml^{-1}) was kindly provided by Dr. S. M. Gaikwad. The protein was mixed with 0.5 M methyl-α-galactose at a proportion of 10:1 (v/v) and incubated for 1 h at 22 ^{0}C, prior to setting up the crystallization. Crystallization experiments were set using the hanging-drop vapor-diffusion method, in which the well solutions contained 0.2 M sodium phosphate buffers of pH range 6.0-7.0 and 30 or 35% saturated ammonium sulfate. Drops were prepared by mixing 2 µl protein solution with 2 µl well solution, on siliconized coverslips which were then kept inverted over wells containing 1 ml of the precipitant solution.

X-ray diffraction data were collected using an *R*-AXIS IV^{++} image-plate detector mounted on a Rigaku rotating-anode generator operating at 100 kV and 50 mA. The data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The structure was determined by molecular replacement method using the program AMoRe (Navaza and Saludjian, 1997). A jacalin dimer (chains EF and GH; PDB code 1JAC;

Sankaranarayanan *et al.*, 1996) was used as a model. The structure was refined using REFMAC (Murshudov *et al.*, 1997) and *QUANTA* (Accelrys, Inc.) was used for the display and model fitting of the structures. Programs from the CCP4 suite (Collaborative Computational Project, Number 4, 1994) were used for crystallographic and other calculations.

Results and Discussion

The protein crystallized in the orthorhombic unit cell ($P2_12_12_1$) of dimensions a = 92.4, b = 98.7 and c = 164.9 Å. The details of data collection and refinement statistics are given in the Table A-1. The refined structure of *A. hirsuta* lectin closely resembles to that of jacalin; exhibiting the characteristic "*β*-prism I fold". The quaternary arrangement of monomer subunits shows that the sugar-binding sites of the four subunits constituting the tetramer are placed in an approximately tetrahedral arrangement. This pseudo-tetrahedral arrangement may be relevant with the biological activity of this protein as a hemagglutinin. The crystal packing mediated by sugar molecules provides three different microscopic sugar-binding environments (Fig. A-1). Despite high sequence similarity with jacalin, *A. hirsuta* lectin shows reduced binding to disaccharides as well as T-antigen. This could be because of the fact that the steric effects of molecular association of symmetry-related molecules as well as the disordered *C*-terminus of *β*-chain were found to be obstructing the secondary sugar binding sites in this lectin.

Improved resolution helped in identifying a pocket similar to sugar binding site on the opposite face of *β*-prism where 5 water molecules are bound (Fig. A-2 (A) and (B)). This suggests that the jacalin type lectins might have evolved from a multivalent lectin losing this additional binding pocket to form a tetramer.

Table A-1. Data-collection and refinement statistics of orthorhombic crystal form I of *A. hirsuta* lectin.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	92.4
b	98.7
c	164.9
Z	4
Resolution range (Å)	15.0-2.5 (2.6-2.5)
No. observations	169292
Completeness (%)	96.2 (98.2)
R_{sym} (%)	7.2 (26.7)
Average $I/\sigma(I)$	11.5 (3.1)
Matthews coefficient (V_M; $Å^3$ Da^{-1})	6.5
Solvent content (%)	80.0
No. unique reflections	46666
No. reflections for R_{free}	2358
R value (%)	19.7
R_{free} value (%)	23.5
No. protein atoms	4568
No. sugar atoms	52
No. water molecules	110
Average *B* factors ($Å^2$)	
Protein	44.3
α-Chain	44.4
β-Chain	43.2
Sugar atoms	44.3
Water molecules	50.4
RMS deviations	
Bond lengths (Å)	0.009
Bond angles in distances (Å)	2.2
Residues in allowed regions (%)	85.4
Residues in additionally allowed regions (%)	14.6
Overall average *G* factor	-0.12

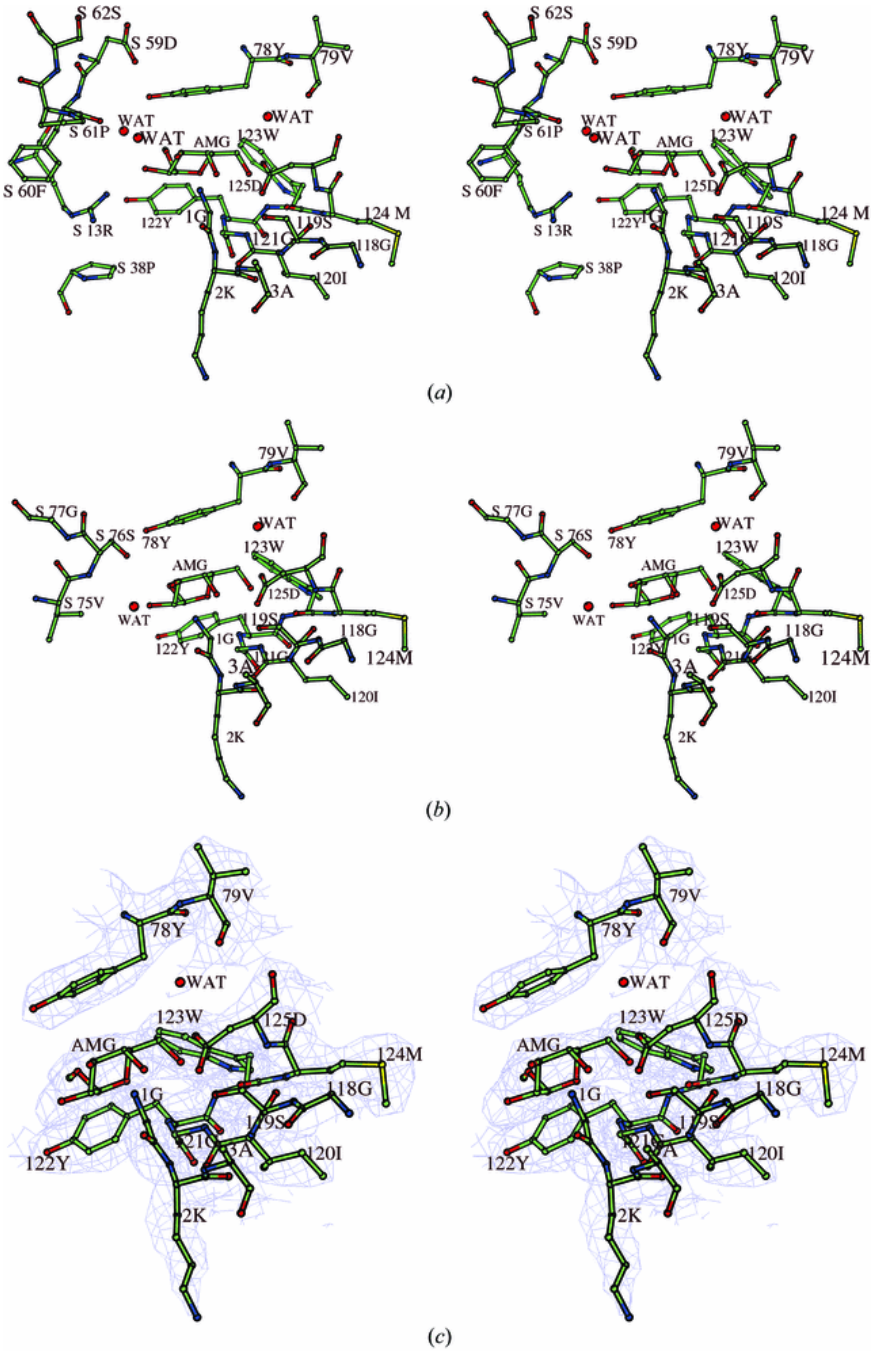


Fig. A-1. A stereoview of the carbohydrate-binding site of *A. hirsuta* lectin is shown along with the bound methyl-α-galactose (Gal), the amino acids and the solvent molecules around the recognition site. The three different microscopic environments found in the crystals of orthorhombic form I are shown separately. (*a*) The binding site as found in subunits III and IV of the tetramer. (*b*) The binding site in subunit I. (*c*) The sugar binding in subunit II. The ($2F_o$ - F_c) map is contoured at 1σ. This figure was prepared using *BOBSCRIPT* (Esnouf, 1999). The figure is taken from Rao *et al.*, 2004.

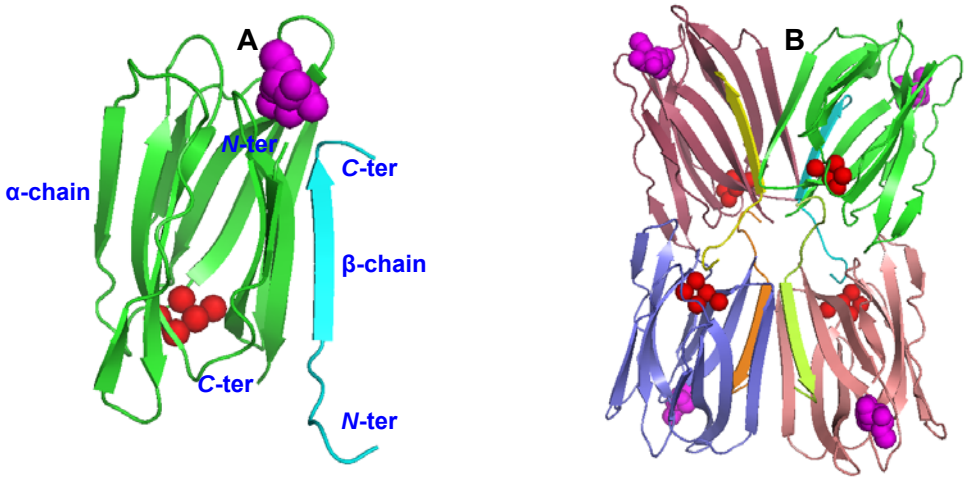


Fig. A-2. (A) *A. hirsuta* lectin monomer and **(B)** tetramer showing the bound sugar (methyl-α-galactose, colored in magenta) and a second binding site where five water molecules are found to be conservatively bound (colored in red). The second binding site is lost probably due to the tetramer formation.

The figure was prepared using PyMOL (DeLano, 2002).

References

- Collaborative Computational Project, Number 4, 1994. *Acta Cryst* D**50**:760-763.
- DeLano WL, 2002. The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org.
- Esnouf RM, 1999. Further additions to *MolScript* version 1.4, including reading and contouring of electron-density maps. *Acta Cryst* D**55**:938-940.
- Gurjar MM, Khan MI, Gaikwad SM, 1998. α-galactoside binding lectin from *Artocarpus hirsuta*: Characterization of the sugar specificity and binding site. *Biochem Biophys Acta* **1381**:256-264.
- Murshudov GN, Vagin AA, Dodson EJ, 1997. Refinement of macromolecular structures by the maximum likelihood method. *Acta Cryst* D**53**:240-255.
- Navaza J, Saludjian P, 1997. AMoRe: An automated molecular replacement program package. *Methods Enzymol* **276**:581-594.
- Otwinowski Z, Minor W, 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**:307-326.
- Rao KN, 2002. Studies on the Structure and Interactions of a Plant Lectin and a Plant Protease Inhibitor. Thesis, University of Pune.
- Rao KN, Gurjar MM, Gaikwad SM, Khan MI, Suresh CG, 1999. Crystallization and preliminary X-ray studies of the basic lectin from the seeds of *Artocarpus hirsuta*. *Acta Cryst* D**55**:1204-1205.
- Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M, 1996. A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a ß-prism fold. *Nat Struct Biol*, **3**:596-603.