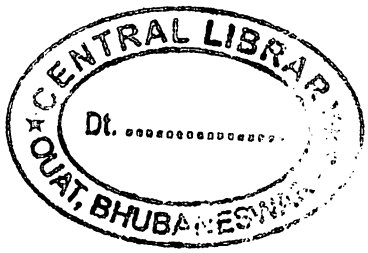
CRYSTALLOGRAPHIC ANALYSIS AND STRUCTURE SOLUTION OF HUMAN SEMINAL PLASMA PROTEIN PSP 94

A
Thesis submitted to the
Orissa University of Agriculture and Technology, Bhubaneswar
for partial fulfillment of the requirements for the Degree of
Master of Science in Bioinformatics.

By
Ashutosh Nanda
Adm. No: 25BI/07
B.Sc Zoology(Hons)



DEPARTMENT OF BIOINFORMATICS

CENTRE FOR POST-GRADUATE STUDIES

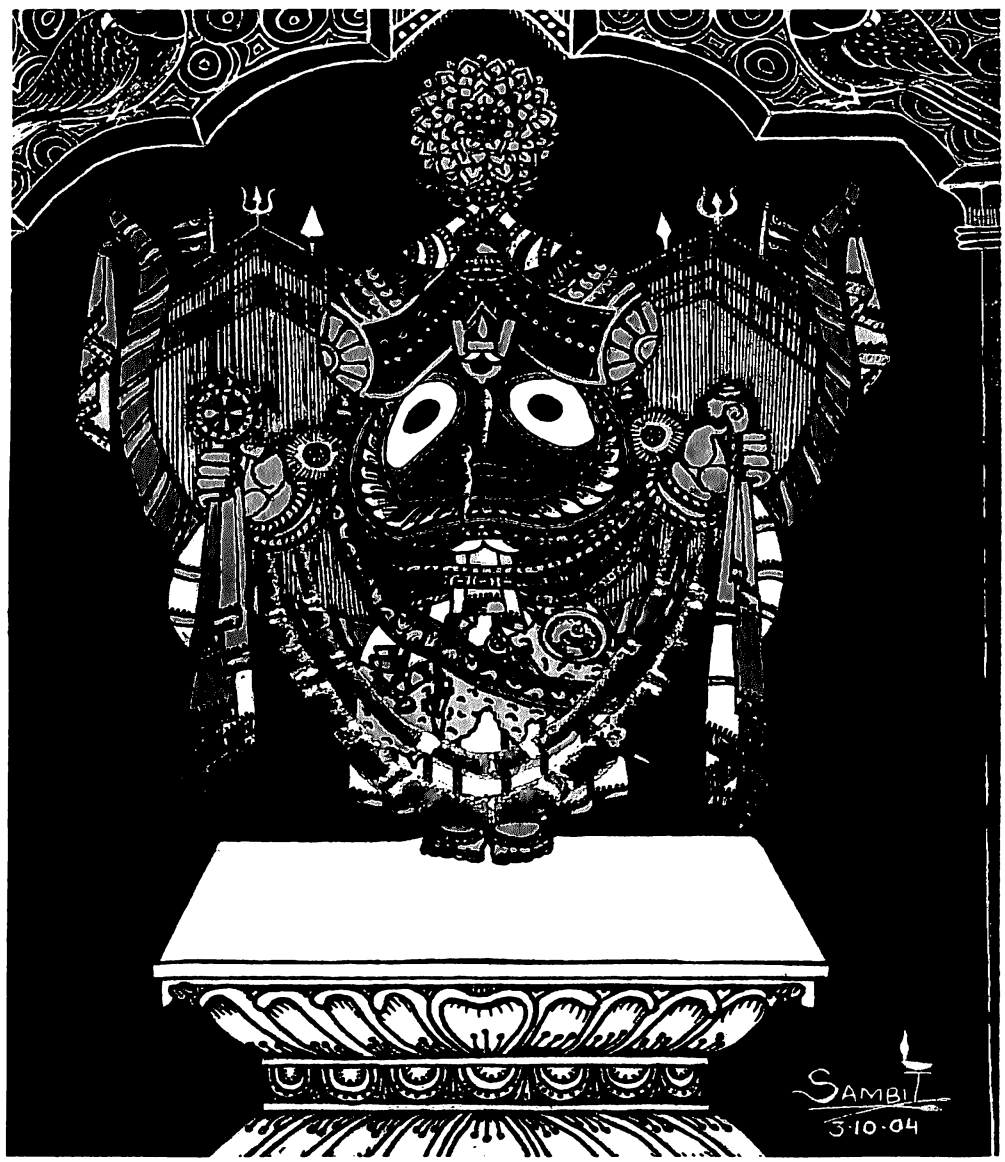
ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

Bhubaneswar-751003,ORISSA

2009

Thesis Advisor:
Dr.P.N.Jagadev

Dedicated to :



Respected H.O.D. Sir, Jeje Maa, Bapa, Nana, Maa.

ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
DEPARTMENT OF BIOINFORMATICS CENTRE FOR POST GRADUATE STUDIES BHUBANESWAR, ORISSA

Dr.P.N.Jagadev
Head P. G. Dept. of Bioinformatics

CERTIFICATE – I

This is to certify that the thesis entitled "***CRYSTALLOGRAPHIC ANALYSIS AND STRUCTURE SOLUTION OF HUMAN SEMINAL PLASMA PROTEIN PSP 94***" submitted for the award of degree of **Master of Science** in the subject of Bioinformatics Of the Orissa University of Agriculture and Technology, Bhubaneswar embodies a faithful record of *bona fide* and original research work carried out by **Ashutosh Nanda (Adm. No: 25BI/07)** under my guidance and supervision. No part of this thesis has been submitted for any other degree or diploma.

The assistance and help received during the course of investigation has been fully acknowledged.

Place – Bhubaneswar
Date - 16/7/09

16/7/09
Dr.P.N.Jagadev
Chairman,
Advisory Committee

CERTIFICATE – II

This is to certify that the thesis entitled "***CRYSTALLOGRAPHIC ANALYSIS AND STRUCTURE SOLUTION OF HUMAN SEMINAL PLASMA PROTEIN PSP 94***" submitted by **Ashutosh Nanda (Adm. No: 25BI/07)** to Orissa University of Agriculture and Technology, Bhubaneswar in the partial fulfilment of the requirements for the award of the **Degree of Master Of Science in Bioinformatics** has been approved by the students advisory committee after an oral examination on the same in collaboration with the an External Examiner.

ADVISORY COMMITTEE

1. Dr.P.N.Jagadev — **Chairman** 22/8/09

HOD, Bioinformatics

2 Dr. Sashikala Beura — **Member** 22.8.09

Assoc. Prof. Horticulture

3. Mr.sukanta kumar pradhan — **Member** skpradhan 22/8/09

Asst. Prof. bioinformatics

External Examiner

Signature:

Date: 22/8/09

ACKNOWLEDGEMENT

I fall in short of words to acknowledge those who were with me for the successful completion of my dissertation.

First and foremost I would take this opportunity to express my deep gratitude and indebtedness to my advisor and guide ***Dr.P.N.Jagadev*** *the Chairman of the Advisory Committee and Head, Department Of Bioinformatics, Orissa University of Agriculture and Technology, Bhubaneswar,*

I express my heartiest thanks to ***Mr. Suryanarayan Rath*** *Asst. Prof. Bioinformatics,* ***Mr. Sukanta ku. pradhan*** *Asst. Prof. Bioinformatics and* ***Dr. Sashikala Beura*** *Assoc. Prof. Horticulture, for their valuable guidance, constant supervision, steady encouragement, whole-hearted co-operation and painstaking efforts during the entire period of investigation and throughout the preparation of this project.*

I extend my hearty thanks and profound gratitude to Dr.H.K.Senapati Dean PGF-cum-DRI for his support during my entire course of action.

I express my heartfelt gratitude to ***Dr. Mukesh Kumar S/O(F)****, BHABHA ATOMIC RESEARCH CENTER, MUMBAI, for his guidance, support, kind help, whole-hearted co-operation and encouragement to carry out my work successfully.*

I wish to acknowledge from the core of heart to ***Dr .M.V. Hosur S/O(H+)*** *head, protein crystallography section, BARC, MUMBAI,* ***Mr. Vishal Prashar S/O(D), Mr. Amit Das S/O(D), Mr. Subhas Ch. Bihani S/O(D),*** *protein crystallography section, BARC, MUMBAI, for their sustained interest, whole-hearted support and valuable suggestions and providing the necessary facilities and information to carry out my research work and also to Mr. S. R. Jadhav for his valuable technical help. I am grateful to all of them for their proper guidance, needful suggestions, encouragement and kind support throughout the period of my research work.*

I express my courteous gratitude to my faculty members ***Ms.Sushma Rani Martha, Ms.Sucharita Balbantray, Ms.Sudipta Mohanty and Ms. Leepika Panda*** *for their kind support and valuable suggestions during my research work.*

I especially thank my friends Bikash, Yangya, Sabyasachi, Pullak, Gopal for their support, deep concern and selfless help, advice, encouragement during the entire course and research.

I bow down before my beloved parents, my brother Milu and sister, other family members for being there for me in all my good and bad times with their unbound love, constant support, immeasurable moral and blessing to build my career.

I am also oblized to the Orissa University of Agriculture and Technology for the kind help and support in providing the necessary instruments to carry out my research work. At last I express my thanks to all the persons who have helped me directly or indirectly whose name could not find a separate place as duly acknowledged.

At the end, I bow my head to the almighty whose omnipresence has always guided me and made me a better person.

Department of Bioinformatics,
Center for Post Graduate Studies,
O.U.A.T, Bhubaneswar.

Ashutosh Nanda
Ashutosh Nanda

Name of the student	:	**Ashutosh Nanda**
Admission No	:	**25BI/07**
Title of Thesis	:	***CRYSTALLOGRAPHIC ANALYSIS AND STRUCTURE SOLUTION OF HUMAN SEMINAL PLASMA PROTEIN PSP 94***
Degree for which Thesis submitted	:	**Master of Science in Bioinformatics.**
Name of the Department College & University	:	**Dept. of Bioinformatics Centre for Post Graduate Studies Orissa University of Agriculture & Technology, Bhubaneswar-751003**
Year of Submission	:	**2009**
Name of Advisor	:	**Dr.P.N.Jagadev**

ABSTRACT

The human seminal plasma protein PSP 94 is a small protein of 94 residues that contains ten cysteines. Since its discovery about 25 years ago, several potential biological functions have been reported for this protein. Many PSP 94 homologues have also been identified since then from various species, but no crystal structure has been determined to date. PSP 94 has been purified from human seminal plasma and crystallized. These crystals diffracted to ~2.2 A resolution. Then the collected data was modified by using ccp4i and structure solution was done by heavy atom method. Model building and refinement was done by coot and phenix.

Ashutosh Nanda

Dr.P.N.Jagadev

Advisor

Ashutosh Nanda

Author

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

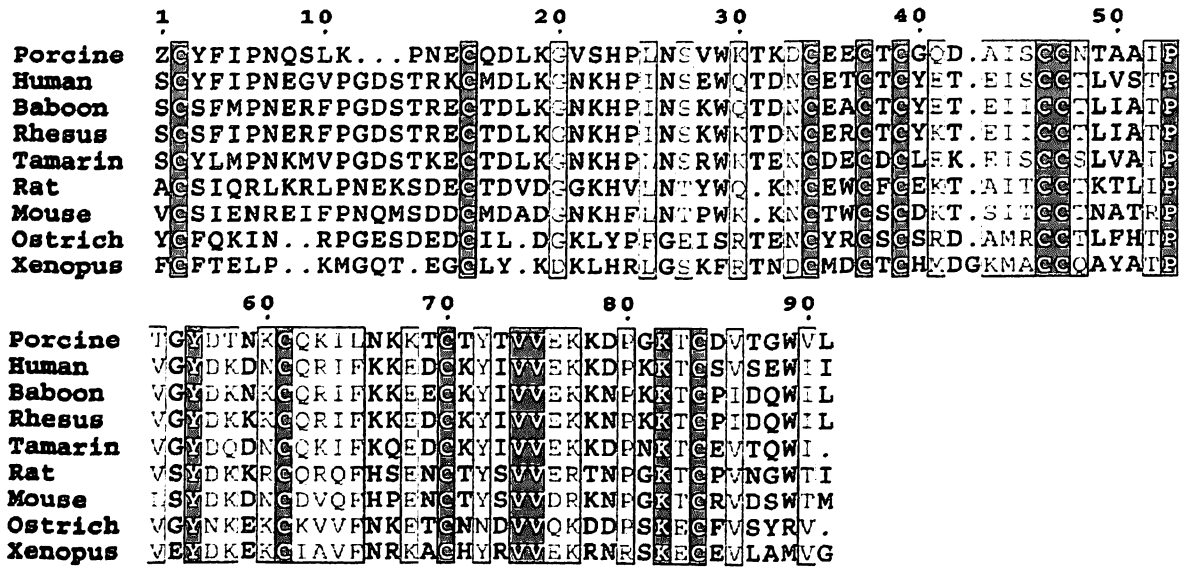
PSP 94	-	Prostate Secretory Protein of 94 Amino Acids
Phenix	-	Software for Model Building
Inhibin	-	Prostate Secretory Protein of 94 Amino Acids
MSMB	-	Prostate Secretory Protein of 94 Amino Acids
MSP	-	Prostate Secretory Protein of 94 Amino Acids
MatLyLu-PTHrP	-	Parathyroid Hormone Related Protein
PCK 3145	-	Synthetic Peptide Corresponding to Amino Acid 31 to 45
Coot	-	Software for Model Building
O	-	Software for Model Building
CCP 4 i	-	Software for Integrating, Indexing and Scaling
MOSFLM	-	Programme for X-ray Data Collection
SCALA	-	Programme for X-ray Data Collection
Synchrotron	-	High Intensity X-ray Source
CyBio-HTPC	-	Crystallization Robot
HPLC System	-	Used For Reverse Phase Column

INTRODUCTION

1.1. -Introduction

Human seminal plasma, the non-cellular component of semen, provides the environment to maintain the viability and quality of the spermatozoa to ensure the success of fertilization. Apart from the low molecular-weight constituents such as inorganic ions, sugars, lipids, steroid, polyamines, and nitrogenous bases, human seminal plasma contains several proteins that carry out many crucial functions [1]. These proteins are secreted by various male sex organs into seminal plasma [5]. Three abundant proteins secreted by the prostate into the seminal plasma are prostate specific antigen (PSA), prostatic acid phosphatase (PAP) and prostatic inhibin (PI) [2]. Crystal structure of the first two proteins (PSA and PAP) has already been solved, but no crystal structure is available till date for the third protein (PI) [3]. This report describes the work carried out by the author towards crystal structure determination of the third protein (PI). Here in this report, this protein is referred as Prostate Secretary Protein of 94 residues (PSP94), although it has been called by many other names like inhibin, human seminal plasma inhibin (HSPI), β-microseminoprotein (β-MSP), sperm motility inhibitor (SMI) and immunoglobulin binding factor (IgBF) [1, 3, 4, 7]. It is secreted by the epithelial cells of the prostate gland [8]. It is a hydrophilic, nonglycosilated, protein with a molecular mass of 10 kDa. This small protein (of 94 residues) has ten cystein and all these are paired to form five intra-molecular disulfide bonds [9].

Apart from humans, PSP94 homologues have been identified from several other species including baboon, rhesus monkey, rat, pig, mouse, cotton-top tomarin, chicken, ostrich and recently from the serum of reptiles [4]. The amino acid sequences from these species show low sequence similarity (<30%) with other proteins present in genbank suggesting that PSP94 homologous form a separate unique protein family. Furthermore, among the PSP94 homologues themselves, there is a large degree of sequence diversity from different species. Except for the ten cysteine residues, only very few amino acid residues are totally conserved in all these species as shown in Fig (1). The number of amino acid residues also varies between species. Conservation of cystein residues suggests that the overall fold of this family of proteins should remain the same [4].



[Fig.-1 Sequence alignment results for PSP 94 in different species using Clustal W]

Although, PSP94 is present in large amount in seminal plasma, its presence, at much lower levels, has been reported in several other tissues like the respiratory (tracheal, bronchial and lung) tissues and the antrum part of the gastric mucosa, female reproductive tissues such as breast and ovaries and human serum[6].

A number of functions, both systemic and confined within the reproductive tract have been postulated or demonstrated for PSP 94, including a modulator of circulating FSH (follicle stimulating hormone) levels [7], a motility inhibitor of sperm [8], a binder of immunoglobulin in the female reproductive tract [9], a growth regulator and inducer of apoptosis in protease cancer cells in vitro and in vivo, a regulator of calcium levels during hypocalcaemia in malignancy.

It was first shown by Prof. A. R. Sheth that PSP94 suppresses the secretion of follicle stimulating hormone (FSH) from the pituitary gland. It was proposed that PSP 94 regulates the FSH secretion by binding to some receptors on the pituitary cell membrane. However such a binding did not affect the release of leutenizing hormone. By carrying out a peptide mapping it was further shown that C-terminal 28-residues (67-94) alone were responsible for suppression of FSH release. Due to this inhibitory effect, it was first called inhibin, although there is no similarity of this protein with the well-known ovarian inhibins. During last two decades, this activity has been debated by various research groups.

It was also shown that PSP94 suppresses the motility of the spermatozoa in a dose dependent manner. The suppression of motility was suggested by inhibiting the function of dynein which is located at the tail of spermatozoa and responsible for its motility. To support

the hypothesis, it was shown that PSP94 inhibits competitively the activity of Na^+,K^+-ATPase in a dose-dependent manner. The inhibitory effect could be reversed by the addition of ATP.

PSP94 has also been shown to be a marker of prostate cancer. The expression of PSP 94 has been shown to be differential depending on the stage of prostate cancer. The level of expression of PSP-94 in the normal prostate is high but progressively decreases as the prostate cancer advances from an early, low invasive, androgen dependent state to a late, highly invasive, androgen independent state cancer. This differential expression could allow the use of PSP-94 as a prognostic marker for prostate cancer in the clinical setting.

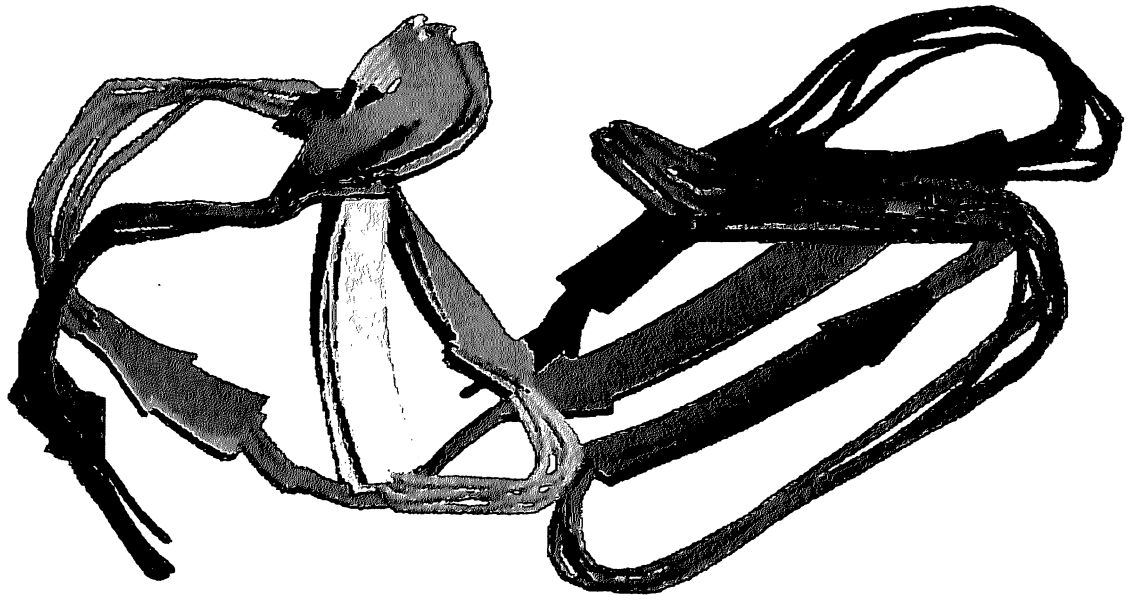
PSP94 also has utility in the treatment of prostate cancer. Adenocarcinoma of the prostate remains one of the most commonly diagnosed cancers in men and is the second leading cause of cancer death following lung cancer. Though an increased number of men are being diagnosed with prostate cancer, the prostate cancer mortality has remained unchanged due to the limited therapeutic strategies available. Surgery and radiotherapy is still remaining the most common approaches for early stage organ confined prostate cancer. However limited success has been obtained when treating late-stage hormone-independent prostate cancer. This is due to the fact that increased production of several growth factors and proteases by tumor cells and the surrounding stroma leads to increased tumor cell proliferation, decreased apoptosis and increased angiogenesis as well as inducing an increased propensity for osteoblastic skeletal metastases.

PSP 94 decreases the tumor growth and hypercalcemia of malignancy. When animals inoculated with theMatLyLu-PTHrP cells (MatLyLu prostate cancer cells were tranfected with full length cDNA encoding parathyroid hormone-related protein [PTHrP], which is known to be the major pathogenetic factor for malignancy-associated hypercalcemia) were infused with different doses of PSP 94, decreases the tumor growth and normalization of plasma calcium and reduction in PTHrP production. Where the decrease in plasma PTHrP and calcium level can serve as the useful markers for prostate cancer.[10]

It was also shown in later study that the synthetic peptide corresponding to amino acid 31-45(PCK 3145) of PSP 94 showed a significant reduction in tumor cell proliferation [11]. The major cause of prostate cancer morbidity is bone pain as a result of nerve impingement by skeletal metastatic tumor. The PCK3145 is the ability to block the development of experimental skeletal metastases and it also reduces the vertebral column perturbation and spinal cord impingement. It has also the ability to reduce plasma calcium and PTHrP levels. It also lowered the biochemical parameters of last stage prostate cancer[11].

Also PSP61 a modified isoform of human PSP 94 consisting of the N-terminal 61 residues was discovered in patients with benign prostatic hyperplasia (BPH) and suggested as a specific biological marker for diagnosis of BPH.

Recently several PSP94 homologues have been discovered in the serum of several reptiles [12]. One of the PSP94 homologue (SSP-2) from this reptile, interestingly, binds to a snake venom-derived Ca2+ channel blocker (triflin) imparting perhaps a self-protection to the snake from its venom.



[Fig-2 Solution structure of N- terminal domain of PSP BP]

During last three decades, a number of studies have been carried out on PSP94 and its homologues, the exact biological function still remains elusive. Two important recent findings are that PSP 94 binds both to PSP 94 binding protein (PSP BP) a 70/95 kda glycosylated protein (436 AA) residues) in human plasma and to cystein rich secretory protein (CRISP 3), a 28 kda protein, first isolated from granulocytes and shown to be present in saliva.

[Fig – 3 Solution structure of N- terminal domain of CRISP-3]

The N terminal part of PSP BP (residues 6-170) has a sequence similarity with a corresponding N-terminal part of CRISP-3 (residues19-183) indicating that this part of the two protein is responsible for the binding of PSP 94.Consequentely it is now possible to study the interaction of PSP94 with natural ligands motivating determinants of the 3D structure of PSP 94 and its many homologues. Although, no crystal structure is available till date, two laboratories [12, 4] have reported the NMR structure of PSP94, showing that this small protein has two domains bridged by a disulfide linkage. Interestingly, the secondary structures of the individual domains reported by the two laboratories are similar, but the orientation of the two domains is very different, giving rise to two distinct overall shapes for this small protein (fig-2). These differences reportedly stem from a different interpretation of ten specific inter-domain NOEs. It remains to be seen which one of the two structures is correct or whether the differences are genuine and is of biological relevance.

REVIEW OF LITERATURE

REVIEW OF LITERATURE

2.1. - Introduction

Proteins are the workhorses of the living cell that carry out the program of activities encoded by genes. They are the most diverse of all macromolecules, and each cell contains several thousand different proteins, which performs a wide variety of functions. Some proteins act as an enzyme, some are also act as hormones, some are antibodies and some are also acts as the carrier of live stock elements [13]. So the central importance of proteins in biological chemistry is indicated by their name, which is derived from the word "proteios" means "of the first rank".

Reflecting there numerous functions, proteins have many shapes and sizes. Proteins are formed from only 20 different monomers called amino acids. The unique shape of proteins arises from non-covalent interactions between regions in the linear sequence of amino acids[14]. Only when the protein is in its correct three-dimensional structure or conformation, it is able to function properly and efficiently. Only a limited set of building blocks can do so much is a nature's marvel. Hence proteins are true glory of the biological world. So the elucidation of three-dimensional structure of proteins is the key to understanding how they work.

2.2. - Tools to study 3-D structure of proteins

The necessity to determine the three dimensional structures of proteins has led to the development of various sophisticated tools based on sound scientific principles. Following four approaches are currently being used to study the structures at a resolution range where individual macromolecules and their internal structure could be visualized [15].

2.2.1. - X-ray Crystallography: X-ray diffraction is currently the predominant source of structural information and over 80% of the structures deposited in the PDB have been determined by this method. X-ray crystallography is historically the first method for structure determination of macromolecules. There is theoretically no limitation on the size of the macromolecules whose structures could be obtained by this method. Primary requirement of this method is the availability of suitable single crystal of macromolecules

that could give reasonably good diffraction pattern and this requirement could become a serious bottleneck. Inspite of this more than 22,000 protein structures have been solved by protein crystallographic methods. X-ray diffraction analysis usually cannot resolve the positions of hydrogen atoms or reliably distinguish between nitrogen, oxygen and carbon atoms. Position of hydrogen atom could be more precisely located by the technique of neutron diffraction, which is similar to x-ray diffraction in principle. Principles of structure determination by X-ray crystallography are described in the next section. The author has used X-ray crystallographic tools in due course of this work, as reported in this thesis.

2.2.2. - Nuclear Magnetic Resonance: NMR detects chemical shifts of atomic nuclei with nonzero spin (e.g. ^{1}H, ^{13}C and ^{15}N). The shifts depend on the electronic environments of the nuclei, namely, the identities and distances of nearby atoms. The result of NMR analysis is a set of estimates of **distances between specific pairs of atoms**, called "constraints". From these constrained, three-dimensional structure of proteins is derived as a set of 10-50 probable models. About 16% of the structures deposited in the PDB have been determined by NMR [16]. It allows the study of the proteins in solution avoiding the time-consuming search for optimal crystallization conditions as required for crystallographic work. However, in order to get NMR resonances sufficiently sharp for adequate resolution, the molecule must tumble rapidly. This typically limits the size of the molecule to about 30 kD.

2.2.3. - Cryo-electron microscopy: In this technique, the biological sample is flash frozen to liquid nitrogen temperatures and investigated in high vacuum by high-energy electrons. Electron micrographs thus obtained are 2-D projection of a 3-D object. A 3-D structure is reconstructed from these different 2-D projections of the molecule by various image reconstruction programs. Cryo-electron microscopy becomes most useful for large assemblies, which is where crystallography tends to become very difficult. However there are special problems with sample damage, which means that very low electron dose must be used to avoid destroying the sample, so that the images have extremely poor signal-to-noise ratio and must be averaged out.

2.2.4. - Atomic force microscopy: The AFM uses a cantilever usually made from silicon or silicon nitride with a very low spring constant, on the end of which a sharp tip (few nanometer) is fabricated using semi-conductor processing techniques. When the tip is brought close to a sample surface the interatomic forces between them cause the cantilever to bend and this motion is detected optically by a laser beam which is reflected off the back

of the cantilever. If the tip is scanned over the sample surface then the deflection of the cantilever can be recorded as an image, which in its simplest form, represents the three dimensional shape of the sample surface.

2.3. - Principles of X-ray crystallography

All of (macromolecular) crystallography relies on the basic theory of X-ray scattering from a crystal, which is a well-understood phenomenon. A single crystal can be thought of as a three-dimensional array of identical unit cells. This periodicity leads to the Laue diffraction conditions:

$$\begin{aligned} \vec{S} \bullet \vec{a} &= h \\ \vec{S} \bullet \vec{b} &= k \\ \vec{S} \bullet \vec{c} &= l \end{aligned} \qquad (1)$$

And to Bragg's law: $$2d(hkl)\sin\theta = \lambda \qquad (2)$$

Here d the distance between the lattice planes (hkl), θ is the reflecting angle and λ is the wavelength, $\vec{a}$, $\vec{b}$ and $\vec{c}$ are the crystal translation vectors, $\vec{S}$ is the vector normal to the reflecting plane with magnitude $2\sin\theta/\lambda$ and h, k, l are the Miller indices of the reflecting lattice plane. According to equations (1) and (2), crystals diffract X-ray beams constructively only in certain discrete directions.

The intensity of the diffracted beam (hkl) is proportional to the square of the amplitude of the structure factor $\vec{F}(hkl)$. The structure factor is a function of the electron density distribution in the unit cell:

$$\vec{F}(hkl) = V \int_{cell} \rho(xyz) e^{2\pi i(hx+ky+lz)} dxdydz \qquad (3)$$

with V the volume of the unit cell and ρ the electron density distribution

By inverse Fourier transformation of this function, the electron density distribution in the unit cell can be calculated, and by virtue of its periodicity expressed in a summation:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l \vec{F}(hkl) e^{-2\pi i(hx+ky+lz)} \qquad (4)$$

The structure factor of equation (3) is a complex quantity:

$$\vec{F}(hkl) = \left|\vec{F}(hkl)\right| e^{i\varphi(hkl)} \qquad (5)$$

While the amplitudes $\left|\vec{F}(hkl)\right|$ can be derived from the experimentally measured intensities of the diffracted beam (*hkl*), the phase angles $\varphi(hkl)$ cannot be obtained directly from thediffraction pattern. Without this essential piece of information equation (4) cannot be solved, and this causes the well known **'phase problem'** in crystallography.

Several methods have been developed to determine the phases indirectly that could be broadly classified into following five categories: [13]

- Multiple isomorphous replacement
- Multiple wavelength anomalous diffraction
- Molecular replacement
- Direct method
- Multiple beam diffraction method

In the present work multiple isomorphous replacement method is used.

2.3.1. - Multiple Isomorphous Replacement :

This involves the isomorphous attachment of heavy atoms to the protein molecules in the crystal. Since different atoms contribute to the scattered intensity in proportion to the square of the number of electrons they contain, a heavy atom like uranium that contains 15 times as many electrons as carbon will contribute to the intensity equivalent to that of 225 carbon atoms. As a result, the change in intensity from the addition of 1 uranium atom to a typical protein of 20kDa is easily measured.

If we have two crystals, one containing just the protein (native crystal) and one containing in addition bound heavy atoms (derivative crystal), we can measure diffraction data from both. The differences in scattered intensities will largely reflect the scattering contribution of the heavy atoms, and these differences can be used (for instance) to compute a Patterson map. Because there are only a few heavy atoms, such a Patterson map will be relatively simple and easy to deconvolute. (Alternatively, direct methods can be applied to the intensity differences.) Once the location of the heavy atoms in the crystal is known, their

contribution to the structure factors can be calculated. Here the assumption is made that the heavy atom doesn't change the rest of the protein structure and scattering from the protein atoms is unchanged by the addition of heavy atoms. Thus the structure factor for the derivative crystal $(\vec{F}_{PH})$ is equal to the sum of the protein structure factor $(\vec{F}_P)$ and the heavy atom structure factor $(\vec{F}_H)$ or

$$\vec{F}_{PH} = \vec{F}_P + \vec{F}_H \qquad (6)$$

In this equation only the lengths of $\vec{F}_{PH}$ and $\vec{F}_P$ are known, but $\vec{F}_H$ is known both in length and direction. Through Harker construction, it turns out that there are two ways to draw vector triangle that will satisfy above vector equation leading to two possible phases of $\vec{F}_P$. This twofold phase ambiguity can be removed by preparing a second derivative crystal with heavy atoms that bind at other sites. Only one phase choice of $\vec{F}_P$ will be consistent with all three observations. Hence, in principle, at least two heavy atom derivatives are required to get the phase information

2.4. - Steps involved in protein crystal structure determination

The process of determining a macromolecular crystal structure consists of:

- Sample extraction, purification and crystallization
- Diffraction data collection and processing
- Structure determination
- Model building and structure refinement

2.4.1. - Sample preparation

Major advances in sample preparation include the advent of recombinant DNA technology, which allows the production of large amounts of protein using the expression of the cloned gene in a microorganism. Recombinant DNA technology also provides the means to bring specific changes into a protein sequence to study the effects of mutation as well as to improve the crystallizability of a protein. However, crystallization of a biological macromolecule has remained the major stumbling block in protein crystallography. Robots have been developed to help automate the process of crystallization of proteins.

2.4.2. - Diffraction data collection

Firstly we need to ascertain the crystal symmetry, the unit cell parameters, the crystal orientation and the resolution limit. Armed with this information we derive a data collection strategy, which will maximize both the resolution and completeness of the data set. The method we use is to rotate the crystal through a small angle, typically 1 degree, and record the X-ray diffraction pattern. If the diffraction pattern is very crowded, then the rotation angle should be reduced so that each spot can be resolved on the image. This is repeated until the crystal has moved through at least 30 degrees and sometimes as much as 180 degrees depending on the crystal symmetry. The lower the symmetry, then more data are required. A typical medium resolution data set may take up to 3 days using an 'in house' X-ray source. For high resolution data collection synchrotron is used where the X-ray intensity is greater and therefore data collection times are shorter - sometimes as fast as 10 minutes! This is the point where we become heavily dependent on computers - every spot on each image needs to be measured. This is a formidable task. Some of our nitrogenase data sets contain around 300 images with over 5000 spots per image.

2.4.2.1. - Data processing: The analysis and reduction of a single crystal raw diffraction data consists of seven major steps. These are :

1) Visualization and preliminary analysis of the original, unprocessed, detector data.
2) Indexing of the diffraction pattern
3) Refinement of the crystal and detector parameters,
4) Integration of the diffraction maxima,
5) Finding the relative scale factors between measurements,
6) Precise refinement of crystal parameters using whole data set, and
7) Merging and statistical analysis of the measurements related by space group symmetry.

Several computer programs are available to perform all these seven steps. Widely used data reduction programs are *MOSFLM* and related programs - *DPS*, *XDS -OSC - DENZO MADNES*, the San Diego programs *XENGEN* and *X-GEN* The theory behind the data reduction methods is complex enough that a series of European Economic Community workshops were dedicated to this task only,. For

the work presented in this thesis, the *MOSFLM* software suite was used for data reduction/processing.

2.4.3. - Structure determination

After obtaining the diffraction data merged and scaled in a particular space group, we need to determine the electron density distribution in the unit cell. For this purpose, we need to know the phases of each reflection apart from their intensity. As mentioned earlier this phase information cannot be measured directly from the diffraction pattern. Several methods have been developed to get the phases. If similar protein structure is known then that protein can be used as search model in the molecular replacement method and initial estimate of the phases could be calculated. However, if no similar protein structure is available, as is the case with PSP94, experimental phasing is carried out either by MAD/ SAD or by MIR/SIR method. Here some heavy atom derivative is made either by inserting selenomethionine in the protein by genetic engineering method in case of MAD/SAD method or by soaking the crystals in solution containg heavy atoms like lanthanides or actinides (uranium). These heavy atoms are located in the crystals first that helps to get the initial estimate of the phases for protein reflections. Using the initial estimate of the phases, electron density distribution in the entire unit cell could be calculated. The protein model is then fitted in the calculate electron density.

2.4.4. - Model building and refinement

Although fitting a model to electron density, refining it to agree with the diffraction data, and validating the end result might seem like very different processes, in fact they are closely intertwined. All of these processes have to cope with the poor observation: parameter ratio that generally plagues macromolecular structure determination.

2.5. - Fitting the first model

2.5.1. - Simplified density maps

The problem of fitting an initial model is to see the forest for the trees. To get an overview of a density map, it needs to be simplified in some way. This is generally done through skeletonisation, or by making a "bones" representation of the map. In skeletonisation,

the map is reduced to a series of lines that run between the peaks of density, through regions of reasonably high density. With this representation, a lot of information can be presented quite simply.

2.5.2. - Tracing the chain

Once there is an idea of some landmarks in the structure, one can start to fit the polypeptide chain by choosing plausible alpha-carbon positions from the bones. It helps a lot to know that alpha-carbon atoms are always separated by about 3.8Å, that the peptide unit between them is planar, and that side-chains branch off from them. So guided by the bones we can look for the right kind of features to get started. In O and XtalView, there is a "baton" feature in which we can start from one alpha-carbon position and choose the next one at an appropriate point 3.8Å away, fix that, choose the next one, and so on.

Then once there is a trace of the main chain, the advantage of what we know about protein structure can be taken. It turns out that there are hardly any five residue stretches of a new protein that we haven't already seen many times before in other structures. So the alpha-carbon trace can be used to look in the database to find similar stretches of main chain. This packages can be found in coot also for model building.

2.5.3. - Fitting side chains

Once the main chain is traced, we have to work out where we are in the sequence. Some side chains, *e.g.* tryptophan, have characteristic shapes and are typically well-ordered. Sulfur atoms have about twice as many electrons as the other atoms in proteins, so disulfides and methionines often stand out. (In nucleic acids, phosphate groups serve as valuable signposts.) But "fuzzier" information can also be used. There are tools in programs like O that will match the sequence of the structure with descriptions of the side chains in terms of size, bulk and environment (inside or outside).

Then the side chains are fit into density, again using information from structural databases. Side chains are generally found in one of a few favoured conformations, called rotamers. Fitting programs will be presented with a number of choices for each side chain and may even make a tentative choice, based on fit to electron density. Again, the advantage is that the best fitting rotamer is more likely to be correct than an arbitrary conformation adjusted to fit the density. If the correct rotamer is chosen, even if it is not centered in density, the refinement program will be more likely to be able to adjust the local main chain conformation correctly.

2.5.4. - Refitting

Automated refinement, as discussed in the next section, will make the small adjustments necessary to make the model agree better with the data, which will have the effect of improving the fit to the electron density. As mentioned above, the phases computed from the model will generally become more accurate than any experimental phase information, so the maps will be getting clearer all the time. This will allow us to see, with confidence, the more ambiguous features, such as poorly-ordered side chains or water molecules. Regions of the structure that were interpreted incorrectly in earlier maps will become clearer. This happens in spite of the fact that they are contributing incorrect phase information to the map calculation, because the improvement of all the other atoms in the structure leads to a general improvement in phase accuracy.

The process of refitting differs from initial fitting only in the emphasis placed on different tools. It is no longer as important to obtain an overview of the structure

2.6. - Refinement

2.6.1- Restraints and constraints

As noted above, a big problem with macromolecular structure determination is the poor observation: parameter ratio at the resolutions to which crystals normally diffract. If we simply refine the positions and B-factors of all the atoms, the refinement will be poorly behaved, the data will be terribly over fit, and the resulting atomic model will probably be very poor.

The way we deal with this is either to add "observations" in the form of restraints, or reduce the number of parameters by constraining the model in some way. Typical restraints include bond lengths, bond angles and van der Waals contact distances. There are generally also restraints to keep planar groups planar, and to maintain the chirality of chiral centres. The restraints are entered as terms in the refinement target, and are weighted so that the deviations from ideal values match the deviations found in databases of high resolution structures.

Constraints are a bit different. When a structure is constrained, the parameterization is changed to reduce the total number of parameters. One example, is to allow only torsion angles to be adjusted, instead of all the individual x,y,z coordinates. Another very common case occurs when there are multiple copies of a molecule in the asymmetric unit of the

crystal. The model can be a single copy of the molecule, replicated by rotations and translations to create the other copies. So if, for instance, there are three copies in the asymmetric unit, the number of adjustable parameters can be reduced by a factor of three.

2.7. - Protein and crystals

2.7.1. - Crystal

A crystal is a three dimensional periodic assembly of individual molecules that align themselves in a repeating series of unit "cells". In á crystal lattice the molecules are arranged in a regular periodic fashion.

2.8. - In order for a protein to crystallize it needs to be

2.8.1. - Pure

For crystallization purpose the purity of the protein is important prerequisite. It shouldn't contain any contaminations.

2.8.2- Homogeneous

It should be homogeneous. i.e. shouldn't contain any truncated proteins, isomorphs, etc. There shouldn't any heterogeneity in posttranscriptional modification. If small tags are used for purification they need to stay attached or removed completely but the protein solution shouldn't contain a mixture.

2.8.3- Properly folded

Should check the activity of the target protein

2.8.4- Fresh

It should be fresh. Because proteins break down with time and the mixture becomes heterogeneous. So always to set up trials should made as soon as possible.

2.8.5- Monodispersed

It should be monodispersed. That means the protein exists in solution as a single oligomeric species e.g. Monomer or dimer. The protein should be free of non-specific

oligomers and aggregates. When a ligand is bound to them it should be added in excess so that all protein molecules have a ligand bound to them.

When there is a unsuccessful crystallization we should check the purity and homogeneity of the protein first.

2.9. - Methods for evaluation of protein before crystallization:

2.9.1. - SDS-PAGE

Sodium Dodycyl Sulphate –Poly Acrylamide Gel Electrophoresis is used to determine the overall estimation of protein size and purity. In this method we can differentiate the target protein from the other proteins. by knowing the molecular weight of that protein.

2.9.2. - Size exclusion chromatography

It is used to determine the size, purity and homogeneity.

2.9.3- Circular Dichroism

It is used to know the folding of the protein.

It shows the percentage of α-helix & β-sheets to check the correct folding.

2.9.4. - Mass spectrometry

It is an analytical technique for the determination of the elemental composition of a sample or protein. It is also used for elucidating the chemical structures of molecules, such as peptides and other chemical compound.

2.9.5. - Dynamic light scattering

It is used to determine the size distribution profile of protein in solution.

2.10. - Crystallization

Protein crystallization occurs when the concentration of protein in solution is greater than its limit of solubility and so the protein is in a supersaturated state.

Protein is to be crystallized for the determination of the structure. because in the amorphous solid state or in liquid state the molecules are not periodically arranged. so when x-ray will fall no result will produce. Because the energy of x-ray is very high. To get the scattered rays the molecule should be periodic arranged in a regular fashion. that's why we need to crystallize the protein, in which we can find the unit cells align themselves regularly or periodically to form a crystal lattice.

2.11. - Principle of crystallization

Protein crystallization occurs when the concentration of protein in solution is greater than its limit of solubility and so the protein is in a supersaturated state. The crystal stage of protein will appear only when the protein is in a zone over the saturation point .i.e. supersaturation. In supersaturation state nucleation centers will form. From this point the critical nuclei appear. Nucleation can either start with molecules themselves(unassisted nucleation), or with the help of some solid matter already in the solution (assisted nucleation).

Once a solution is saturated, or a melt nears the solidification point, solid material starts to form. If the molecules come together in a random arrangement, they do not occupy the closest packed space. However, if the molecules come together in an ordered array, they pack together in a much smaller space, like in a properly assembled jigsaw puzzle. The proper packing uses less space and is also of lowest energy, which is always the most stable condition. As it happens, this ordered array pattern repeats itself regularly in 3 dimensions, and the crystal is the macroscopic object. The nice faces of a crystal result from the fact that certain directions in this array are more accessible to the attachment of new molecules, so the crystals grow uniformly in these directions. However, the incoming molecules need a little time to align themselves properly at the surface of a growing crystal in order for the crystal to continue to grow nicely. Hence the need for slow crystallization. If the solution becomes over-saturated so that solid forms quickly, the incoming molecules do not have time to align properly with the result that one obtains small crystals that are usually poorly formed.

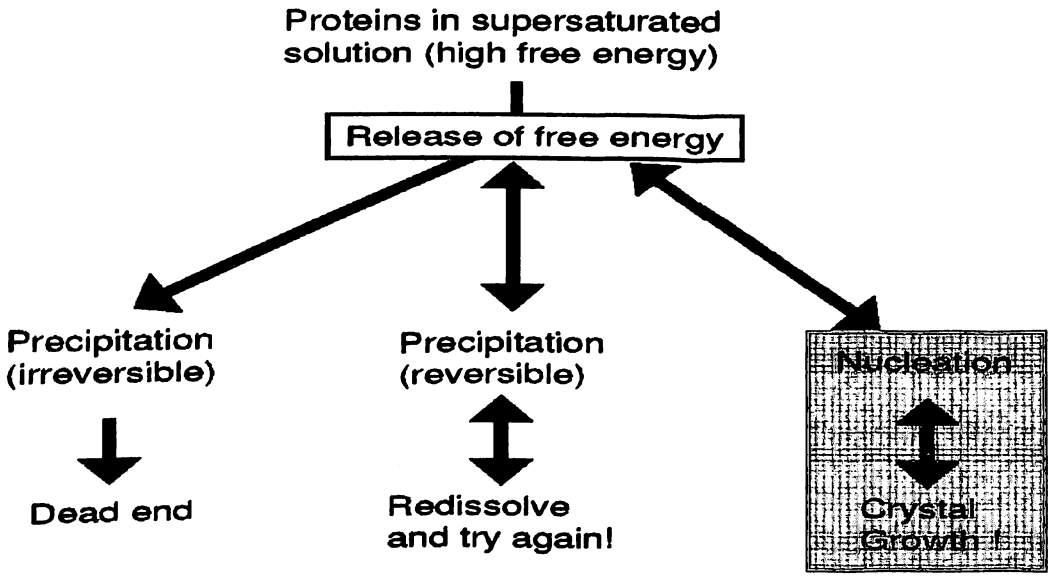
2.11.1. - Saturation

Saturation occurs when the rate of loss and gain of both the solid and solution phases of the protein are equal,and the system is in equilibrium.

2.11.2. - Supersaturation

To reach the supersaturation we have to concentrate the protein ,by adding different concentrations of precipitants like salts (e.g. ammonium sulphate, sodium chloride etc)organic solvents(e.g. ethanol, MPD...) PEG of different molecular weight.

2.11.3. - What can happens to a supersaturated solution?



[Fig. - 4]

When the protein is in supersaturated form it has high free energy. In this stage it releases free energy, which can leads to three ways.

- In the first way irreversible precipitate will form in which there is a dead end
- In second way precipitation will form, but protein is not still denatured. so we can rediscover the precipitate and try again.
- In the third way nucleation zone will appear where the crystal growth will start.

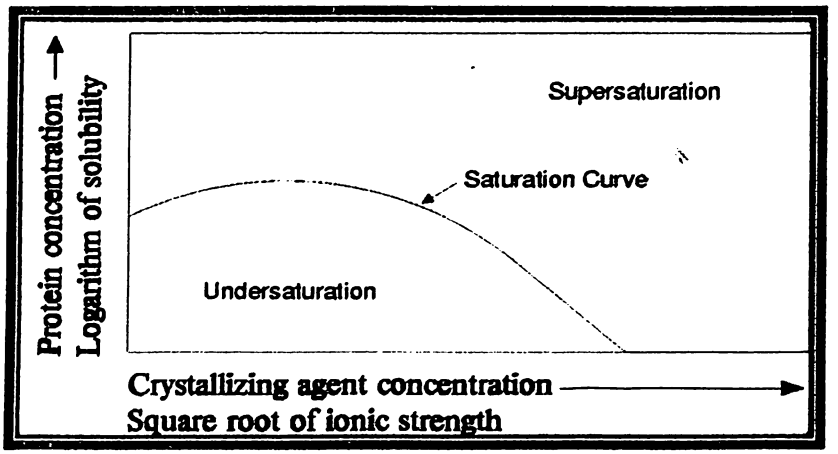
2.11.4. - Protein solubility

It is very difficult to find out the solubility of protein.

Protein solubility is depends upon many factors, such as

- The protein it self i.e. the surface charges and structure

- The solution surrounding the protein i.e. the ph of the solution, the type of salts organics present etc.
- Temperature i.e. in different temperature the solubility will vary for different proteins.
- Concentration of protein.



[Fig - 5 Solubility curve]

2.11.5. - Salting out

It is a process in which there is a reduction of protein solubility as the concentration of salt increases. it is seen in the right hand side of the diagram.

2.11.6. - Salting in

It is seen in the left hand side of the diagram, where there is an increase in protein solubility as the concentration of salt increases.

2.12. - Methods of protein crystallization

There are many ways for protein crystallization

Commonly used techniques for crystallization

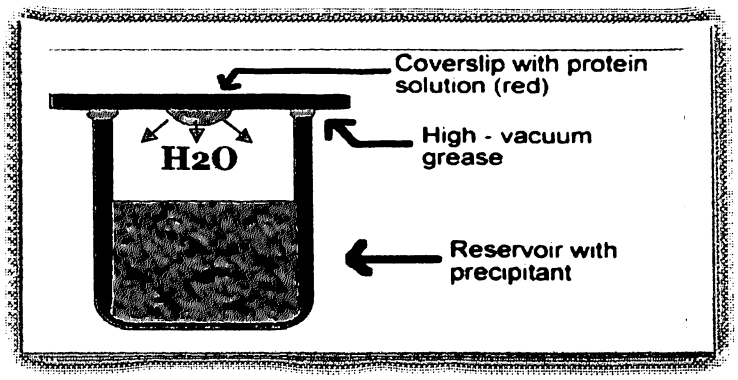
- Vapour diffusion
- Microbatch

- Microdialysis
- Counterdiffusion

2.12.1. - Vapour diffusion method

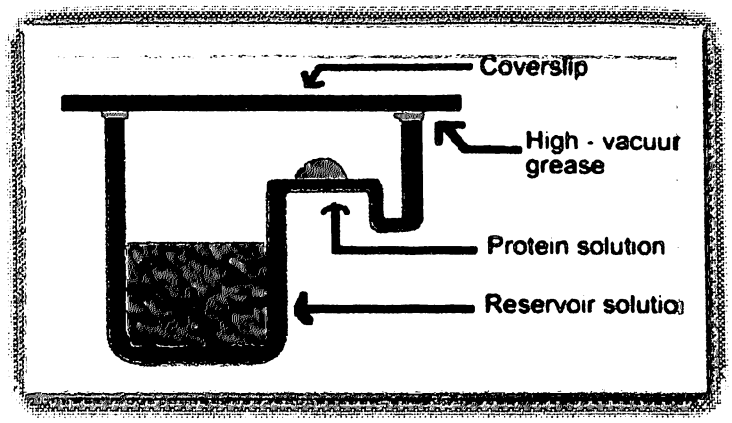
It is the most used and popular method. Because in this method there is a start and end point. in this method we can travel back through the phase diagram. so in this project this method was used. In a vapour diffusion experiment, small volumes of precipitant and protein mixed together and the drop equilibrated against a larger reservoir of solution containing precipitant or another dehydrating agent. In vapour diffusion method two types of set ups can be drawn.

2.12.1.1. - Hanging drop : The drop containing the protein hangs from the coverslip on to a sealed reservoir well as shown in the figure.



[Fig - 6 Hanging drop]

2.12.1.2. - Sitting drop : The drop containing the protein sits on an elevated platform within a sealed reservoir well as shown in the figure.



[Fig -7 Sitting drop]

In vapour diffusion set up small volumes of precipitant and protein mixed together and the drop is equilibrated against a larger reservoir solution containing precipitant or another dehydrating agent. Over time water will evaporated from the drop increasing the concentration of both protein and precipitant. When there is the right concentration and combination of chemicals, Ph and temperature, crystals will appear in the drop.

In this experiment protein concentration and crystallizing agent concentration both increase through water evaporation.

Nucleation happens at a higher saturation than crystal growth. As protein is used up to form the crystal, the concentration drops, and gets the clear metastable zone where the crystal growth continues.

2.12.2. - Microbatch

In microbatch set up crystallization the precipitant and protein are mixed directly under oil. supersaturation is reached immediately. Evaporation occurs depending upon the nature of the oil. If evaporation occurs we can start in stable part of phase diagram and move diagonally up, but it has no end point like the vapour diffusion method.

2.12.3. - Microdialysis

In a dialysis crystallization experiment, protein is equilibrated against a larger volume of precipitant through a dialysis membrane.

The concentration of protein becomes constant

2.12.4. - Counterdiffusion

In counter diffusion both protein and precipitant diffuse along capillary, generating many microconditions. One of them may be lead to crystal formation.in each region of capillary individual conditions form that have their own phase diagram.

2.13. - Features of protein crystal

- Protein molecules are larger in contrast to salt molecules, which are tightly packed. Protein molecules thousands of molecules where as salt crystals have two to ten atoms.

- Protein crystals are irregular in shape
- It has an irregular surface
- Protein crystals have less contact area between individual molecules.
- They are softer and fragile.
- The same protein can grow in to differently shaped crystals, using different intermolecular contact area.
- Have special requirements to PH, salts, and temperature to stay in native conformation.
- It shows weaker birefringence i.e. when polarized light passes through it rainbow color will not appear.
- It contains around 60% of water, which forms the channels.
- Should not dry out & be handled with care.
- It is more difficult to grow protein crystal than salt crystal

2.14. - Screening

Crystal screening is an important process in the determination of protein structures. Proteins are crystallized in order to determine their molecular structure by x-ray protein crystallography. However, the most time consuming process on the path to determining molecular structure is the protein crystallization step. This involves crystal screening a large number of buffer conditions until conditions are ideal to induce protein crystal growth.

In general terms:

Low polydispersity is good, high polydispersity is bad for high quality crystal growth.

It was found that results from light scattering instrumentation can be used to simplify the determination of the protein crystallization process. Light scattering may be used to screen protein samples quickly and with very little sample required.

2.15. - Screening methods

2.15.1. - Full Factorial

In full factorial screens, all elements of the matrix of parameters are sampled

2.15.2. - Incomplete Factorial

Incomplete factorial screening is a method of sampling parameter space evenly and efficiently. Factor levels are chosen randomly and then balanced to acheive uniform sampling.

All two-factor interactions are sampled as uniformly as possible.

2.15.3. - Random

Purely random sampling of all parameters, but it approximates incomplete factorial designs.

2.15.4. - Sparse Matrix

Sparse Matrix screens involve an intentional bias towards combinations of conditions that have worked previously.

2.16. - Seeding

Seeding is a way of manipulating the crystals.There three different methods of seeding

- Streakseeding
- Microseeding
- Macroseeding

And four different ways of using seeding

2.16.1. - Homogeneous seeding

The seeds are used to grow crystals of the same lattice and symmetry from an identical macromolecule

2.16.2. - Heterogeneous seeding (Cross-Seeding)

The seeds are used to grow crystals of the same lattice and symmetry from a different, but usually very similar macromolecule (such as a mutant)

2.16.3. - Homogeneous Epitaxial seeding

The seeds are used to grow crystals of a different lattice and symmetry from an identical macromolecule

2.16.4. - Heterogeneous Epitaxial Seeding

The seeds are used to grow crystals of a different lattice and symmetry from a different macromolecule e.g. the nucleation of protein crystals on cellulose fiber impurities in the drops.

2.17. - Soaking

Soaking is a process in which the crystal is taken out of the mother liquor and soaked in a solution containing glycerol or ligands/substrate or heavy atom solution. It is used to

- Protect crystal from damage in the x-ray beam by replacing water using cryoprotectant
- To determine the binding pocket of a ligands or substrate in the active site of an enzyme
- To produce heavy atom derivatives of the protein crystal, using the MIR method.

2.18. - Cryoprotection

It is process by which the crystal is protected from damage. Protein crystals contain many water channels. During time of freezing it can break the crystal. So when the crystal is cryoprotected the water channels become amorphous solid. So it protects the crystals.

X-ray radiation is very strong and can destroy the protein crystal. So to save the crystal from damage cryoprotectant is used.

Lastly crystal should freeze in liquid nitrogen.

MATERIALS & METHOD

MATERIALS & METHODS

[Sample Preparation]

PSP 94 was purified from human seminal plasma by Dr. Smita Mahale and her group at the National Institute of Research in Reproductive Health according to the published protocols [Thakur et al.1981;jagtap et al,2007]. Briefly, the seminal plasma was subjected to ammonium sulphate precipitation followed by hydrophobic interaction chromatography using a phenyl Sepharose column and the eluted PSP 94 fraction was further purified using a reverse-phase column on an HPLC system. The purified protein was then lyophilized and stored at 253 K until further use. Lyophilized powder was kindly provided by Dr. Mahale for our crystallographic work. For crystallization purposes, the lyophilized protein was dissolved in water (10 mg/ml). Initial screening for crystallization conditions was earlier performed using a CyBio-HTPC crystallization robot and several commercial screening kits from molecular dimensions (Structure screen 1and 2, JCSG+ and Memgold) and Jenabioscience (JBScreen basic 1,2,3 and 4). It was found that the protein crystallizes in 0.1M Sodium acetate, 0.1- 0.2M Lithium sulfate and 40-50% PEG 400. However, to get the diffraction quality crystals, this condition had to be optimized to get bigger and good quality crystals.

3.1. - Crystallization

The crystallization experiments were carried out in 24 well cell culture plates by hanging drop vapour diffusion method. It involves the following steps:

- Siliconization of cover slips
- Making the crystallization set-ups
- Seeding
- Soaking
- Freezing

3.2. - Materials required

- Coverslips
- Crystallization plates
- Grease
- Micropipette
- Tips
- pH meter
- PEG 400
- 1M Sodium Acetate
- 1M Lithium Sulphate
- Optical microscope
- Liq Nitrogen

3.3. - Methodology

3.3.1. - Siliconization

For making the crystallization set up by hanging drop vapor diffusion method, the glass cover slip needs to be siliconized. Siliconization makes the glass surface hydrophobic so that the protein drops will remain spherical in shape.

For siliconization, the coverslips were washed with methanol and then incubated in siliconization solution (5% dichloro dimethyl silane in toluene) for 30 minutes with gentle stirring. Siliconization solution was then removed and cover slips were rinsed with toluene for 5 minutes, followed by washing in methanol for 10 minutes. Finally, it was washed with water and incubated for 2 hours at 250°C.

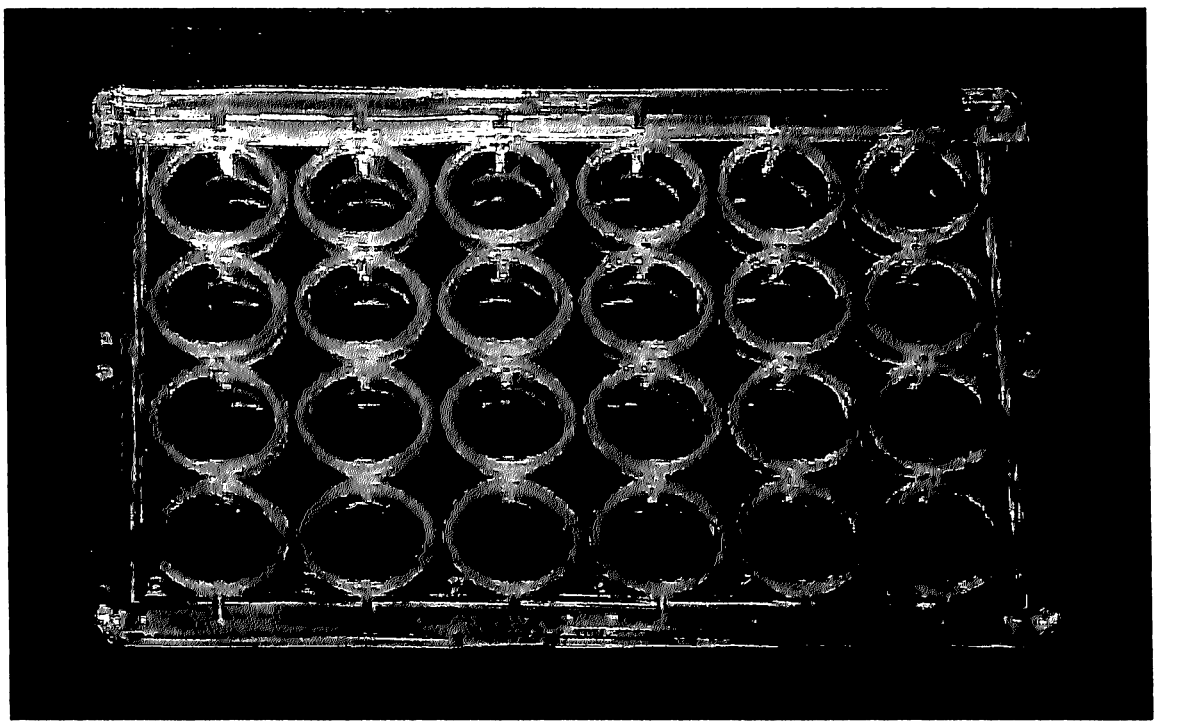
3.3.2. - Making setups

The crystallization experiments were carried out in 24 well culture plates as shown in figure. The crystallization plate was washed and rinsed with distilled water. The upper edge of each well was greased uniformly and filled with 1 ml of the crystallization solution containing different concentrations of salt, precipitant and buffer.

2µl of protein was pipetted onto the siliconized coverslips followed by addition of 2µl of crystallization solution from the respective well.

Each coverslip was then inverted and gently placed over the corresponding reservoir and sealed by gentle pressing.

The plate was kept at 25°c for under observation at regular interval. The crystals usually appeared after 2-3 days.



[Fig – 8: 24 Well Crystallization Plate]

3.3.3. - Seeding

For seeding purposes, usually one small needle shaped crystal was picked with the help of a loop and transferred to a clear drop. In the clear drop, the conditions were such that the protein didn't reach the nucleation zone but it were in a clear metastable zone where if the nucleation centres are provided, they would grow. As discussed in results, this macroseeding method worked well for the needle shaped crystals and they grew thicker.

3.3.4. - Soaking

Crystals were picked with the help of a loop and put in a drop containing Zinc Acetate, uranyl nitrate. These solutions were prepared in the reservoir solution.

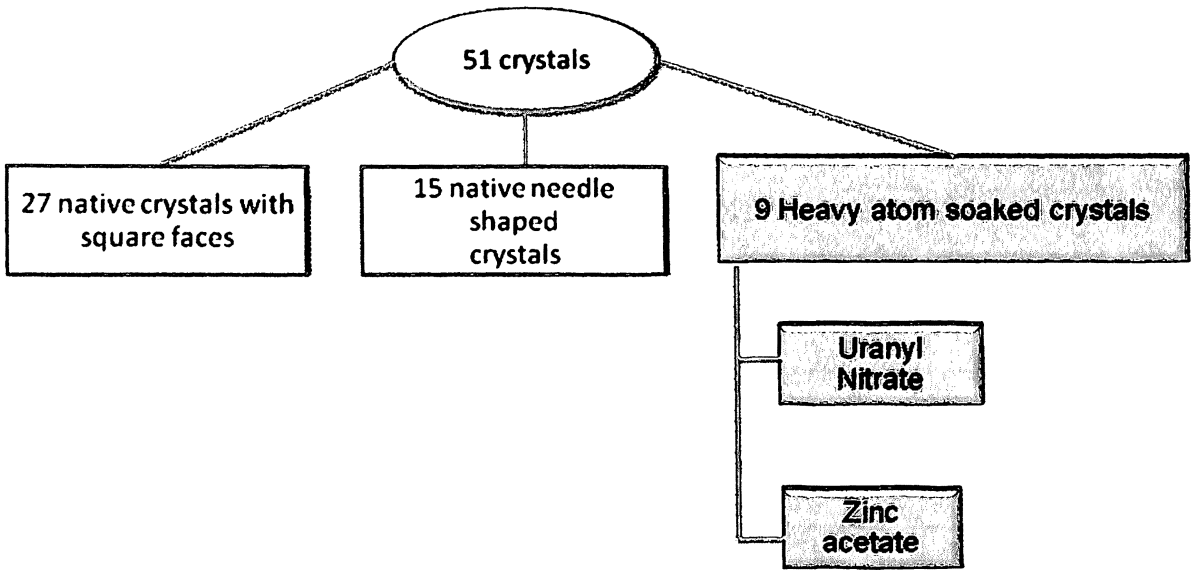
3.3.5. - Freezing

X-ray diffraction data is usually collected at 100K to avoid radiation damage to the crystals. Crystals were snap frozen in liquid nitrogen and stored in a dry shipper dewar, which was then sent to synchrotron facility for the X-ray diffraction data collection. No additional cryoprotectants were used as it was found that our crystallization solution containing ~40-50% PEG itself acted as cryoprotectant to the crystals.

3.4. - X-ray data collection and refinement

The crystals of PSP94 were grown as described in the previous chapter. These crystals were flash frozen in liquid nitrogen and transferred to a dry shipper cryo dewar. The dewar was then shipped to Swiss Light Source (SLS), Switzerland for X-ray diffraction data collection. Diffraction data were collected on the X06DA (PXIII) beamline of SLS. The diffraction data were recorded at 100 K using a MAR Mosaic 225 mm CCD detector (MAResearch).

A total of 51 crystals were sent to SLS, out of which 27 were the native crystals with square faces, 15 were the native needle shaped crystals and 9 were needle shaped crystals soaked in either uranyl nitrate or Zinc Acetate. Schematically, the information is shown in Figure 1.



[Fig – 9 : Crystals for X-ray diffraction data collection]

The diffraction data were collected by oscillation method with 1° oscillation per image and 3 sec of exposure. The data were indexed, integrated and scaled using the programs MOSFLM and SCALA.

RESULTS

Results

The crystallization condition was optimized by varying the concentration of the salt (lithium sulfate) and the precipitant (PEG400) at different PHS. Each of these three parameters was systematically varied, one at a time, to look for its effect. Given below are some of the observations that were made during the set ups.

0.1 M Lithium sulphate, 0.1 sodium acetate PH-4.5

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table - 1

0.1 M Lithium sulphate, 0.1 sodium acetate PH-4.7

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table - 2

0.1 M Lithium sulphate, 0.1 sodium acetate PH-4.9

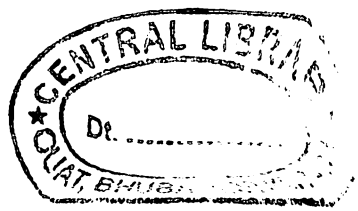
PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table - 3

0.1 M Lithium sulphate, 0.1 sodium acetate PH-5.1

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table - 4



0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal, precipitation
44%	Very small thin, **needle** shaped crystals, more in number
46%	No crystal
48%	No crystal
50%	No crystal

Table - 5

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5

PEG 400	OBSERVATION
42%	No crystal
42.4%	No crystal, precipitation
42.8%	No crystal
43.3%	Very small thin, **needle** shaped crystals, more in number
43.7%	No crystal
44%	No crystal

Table - 6

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.7

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal, precipitation
44%	No crystal
46%	No crystal
48%	No crystal
50%	Very small thin, **needle** shaped crystals, more in number

Table - 7

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.9

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal, precipitation
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table - 8

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-5.1

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	Small **square** shaped crystals with precipitation

Table - 9

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5

PEG 400	OBSERVATION
40%	No crystal
40.5%	No crystal
41%	No crystal
41.5%	No crystal
42%	No crystal
42.5%	No crystal

Table - 10

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5

PEG 400	OBSERVATION
42.5%	No crystal
42.7%	No crystal
42.9%	No crystal
43.1%	No crystal
43.3%	No crystal
43.5%	No crystal

Table - 11

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.9

PEG 400	OBSERVATION
48%	No crystal
48.4%	No crystal
48.8%	No crystal
49.2%	No crystal
49.6%	Small **square** shaped crystals more in number
50 %	Small **square** shaped crystals more in number

Table - 12

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-5.1

PEG 400	OBSERVATION
48%	No crystal
48.4%	No crystal
48.8%	No crystal
49.2%	Small **square** shaped crystals more in number
49.6%	Small **square** shaped crystals more in number
50 %	Small **square** shaped crystals more in number

Table - 13

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5

PEG 400	OBSERVATION
43%	Very small thin, **needle** shaped crystals, less in number
43.4%	No crystal
43.8%	No crystal
44.2%	Very small, thin, **needle** shaped crystals, more in number
44.6%	Very small, thin, **needle** shaped crystals, more in number
45 %	No crystal

Table - 14

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-5.1

PEG 400	OBSERVATION
49%	No crystal
49.2%	No crystal
49.4%	Small **square** shaped crystals more in number
49.6%	Small **square** shaped crystals more in number
49.8%	No crystal
50 %	Medium **square** shaped **good** crystals less in number

Table - 15

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-5.1
JBS screen + volatiles

PEG 400	OBSERVATION
50%	**Good square** shaped crystals less in number
50%	**Good square** shaped crystals less in number
50%	No crystal
50%	No crystal
50%	No crystal
50 %	No crystal

Table - 16

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5
JBS screen + volatiles

PEG 400	OBSERVATION
46%	No crystal
46.4%	Clear drop
46.8%	No crystal
47.2%	Very small, thin, **needle** shaped crystals, more in number
47.6%	Very small, thin, **needle** shaped crystals, more in number
48 %	Precipitation

Table - 17

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.5
JBS screen + C1 volatiles

PEG 400	OBSERVATION
46.2%	Clear drop
46.4%	Clear drop
46.6%	No crystal
46.8%	Precipitation
47%	Precipitation
47.2 %	Small **square** shaped crystals more in number

Table - 18

Observations

In these crystallization set ups, several different observations were also made. The meaning to these observations is as follows:

- **Clear drop:** The protein did not reach saturation or did not nucleate.

- **Matter:** Fibres, dirt, glass etc were sometimes seen in some of the setups.

- **Precipitate**: It indicated that the protein is supersaturated and came out of the solution quickly enough in amorphous form. To get the crystals, protein needs to come out of the solution slowly so that they can form an ordered lattice. A brown precipitate typically indicates that the protein denatured and wouldn't crystallize

- **Gel**: These were transparent, irregular regions of the drop.

- **Skin:** Process of formation is not understood. Crystals can grow before a skin forms but usually nothing happens afterwards

- **Phase separation:** Hundreds of small round droplets were sometimes seen in the set ups. Crystals can grow from phase separation. This usually happens when one precipitant becomes immiscible with another precipitant.

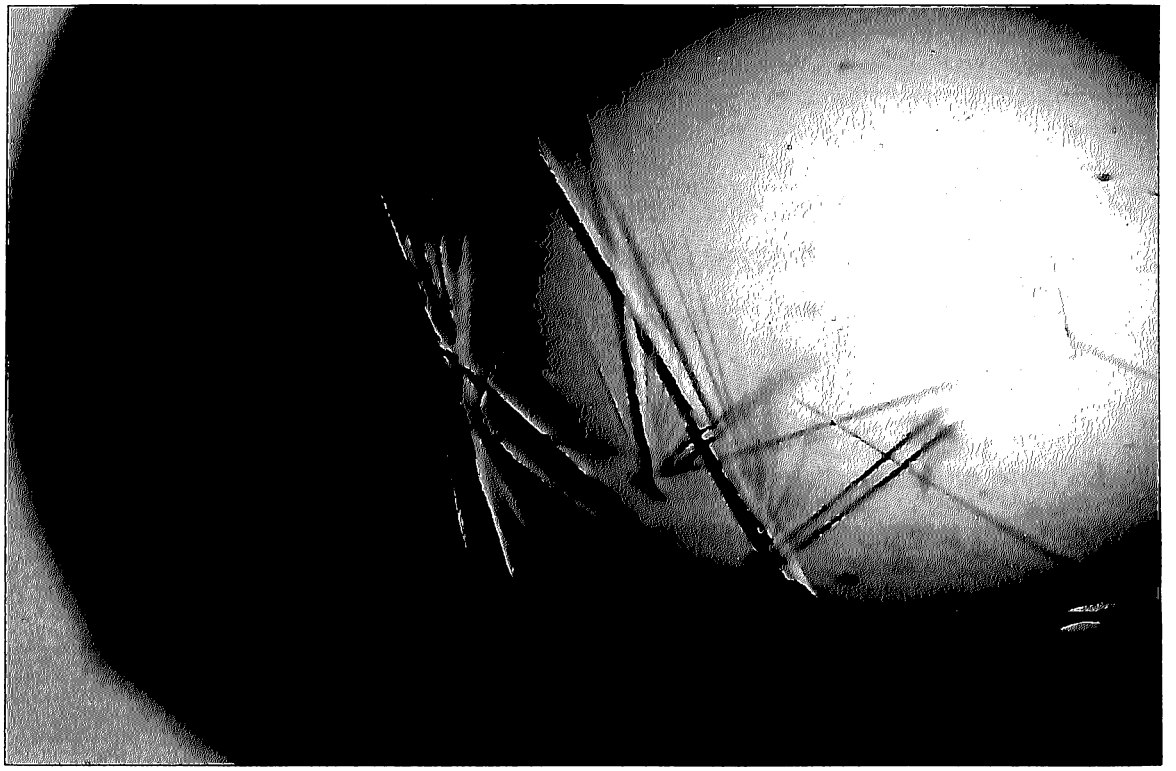
- **Oils:** Droplets are usually fewer and not as round as compared with phase separation. The protein has separated out from the aqueous phase

- **Spherulites:** Transparent birefringent clusters Can be difficult to distinguish from oils

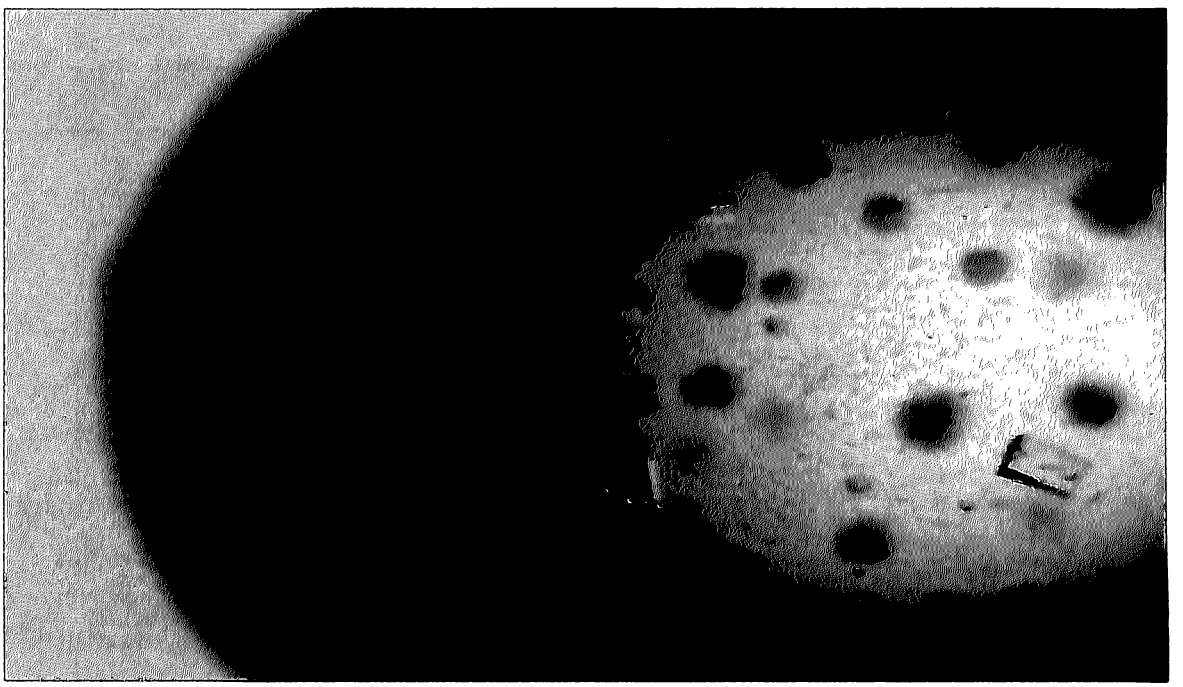
- **Crystals**: Looked like geometric object with transparent clean faces and straight edges

Types of crystals

Two types of crystals of PSP94 were obtained- the needle shaped and the square shaped. The needle shaped crystals grew in reservoir solution containing 0.1M Sodium acetate pH 4.5, 0.2 M Lithium sulfate and ~47% PEG400. Some better quality crystals were obtained with the addition of 0.4 µl additive solution (γ-butyrolactone) to the drop.



[Fig – 10: Needle shaped crystal]



[Fig - 11: Square shaped crystals]

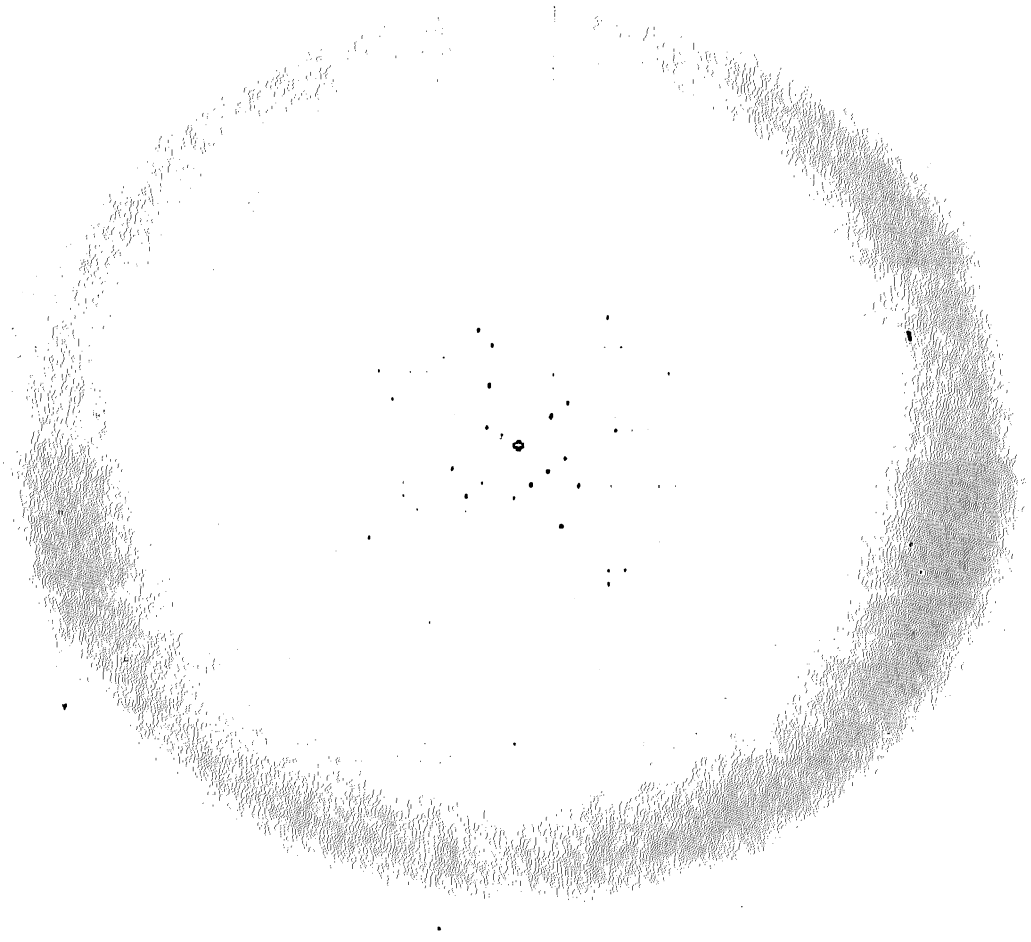
To grow the crystals bigger, macroseeding was used. These seeded crystals looked good and were thicker ($0.08X0.08X1mm^3$) as shown in figure - 10.

- Figure - 10 Needle shaped crystals
- Figure - 11 Square shaped crystals

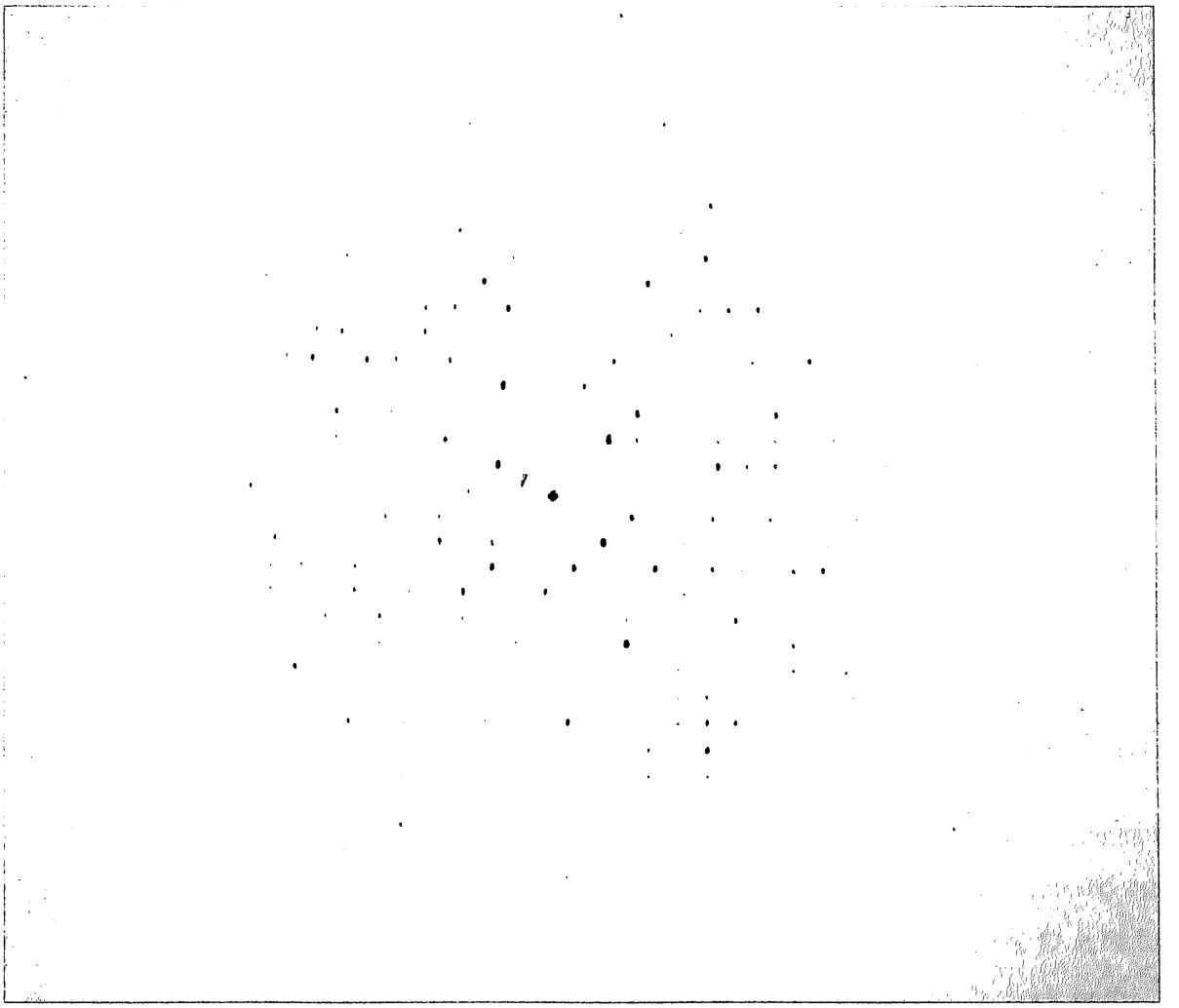
The square shaped crystals were grown with reservoir containing 0.1M Sodium acetate pH 5.0, 0.2 M Lithium sulfate and ~50% PEG400. The size of these crystals were about 0.1X0.1X0.1mm^3) and are shown in figure - 11.

Diffraction data from crystals with square faces:

Out of a total of 27 crystals six complete data sets could be obtained. A typical diffraction frame is shown in Figure 2. The data processing summary is shown in Table 19.



[Fig – 12: Typical diffraction image from crystals with square faces]



[Fig – 13: Magnified image of the diffraction pattern]

Table 19

Crystal number	5_2	5_3	7	8	25	26
Resolution (in Å)	5.5	5.3	5.23	4.5	4.7	4.25
% Completeness	98.2	99.1	99.3	96.7	99.2	93.7
R_merge (%)	8.6	7.9	8.4	11.1	16.3	12.5
I/σ	10.8	9.9	11.85	9.67	9.83	10.89

Mosaicity (o)	0.8	0.9	0.319	0.359	0.370	0.66
Sapce group	P41 21 2	P41 21 2	P41 21 2	P41 21 2	P41 21 2	P41 21 2
Cell dimensions (in Å)	a=b=102.9, c=193.8	a=b=102.3, c=193.9	a=b=102.2, c=193.6	a=b=102.7, c=195.2	a=b=102.3, c=194.1	a=b=103.0, c=196.0
No. of Unique reflections	6720	6933	7367	11436	10230	13329
No. of frames	90	90	90	90	151	90

These crystals belonged to space group P41 21 2 with unit cell dimensions a=b=103.0Å, c=196.0Å. Due to the large unit cell dimensions, these crystals diffracted only to a lower resolution. The best resolution obtained for these crystals were 4.25Å. Due to low resolution data, the structure solution by Molecular replacement method didn't work here.

Diffraction data from needle shaped crystals:

A total of 11 data sets were collected from the native needle shaped crystals. The data processing results are shown in Table 20 and 21.

Table - 20

Crystal number	28	29	30	31	32	33
Resolution (in Å)	2.9	2.74	2.69	3.2	3.01	2.12
% Completeness	98.7	99.8	99.4	99.5	99.4	99.9
R_merge (%)	12.3	9.6	12.6	17.0	11.3	9.7
I/σ	10.48	12.37	9.6	8.39	12.05	15.83

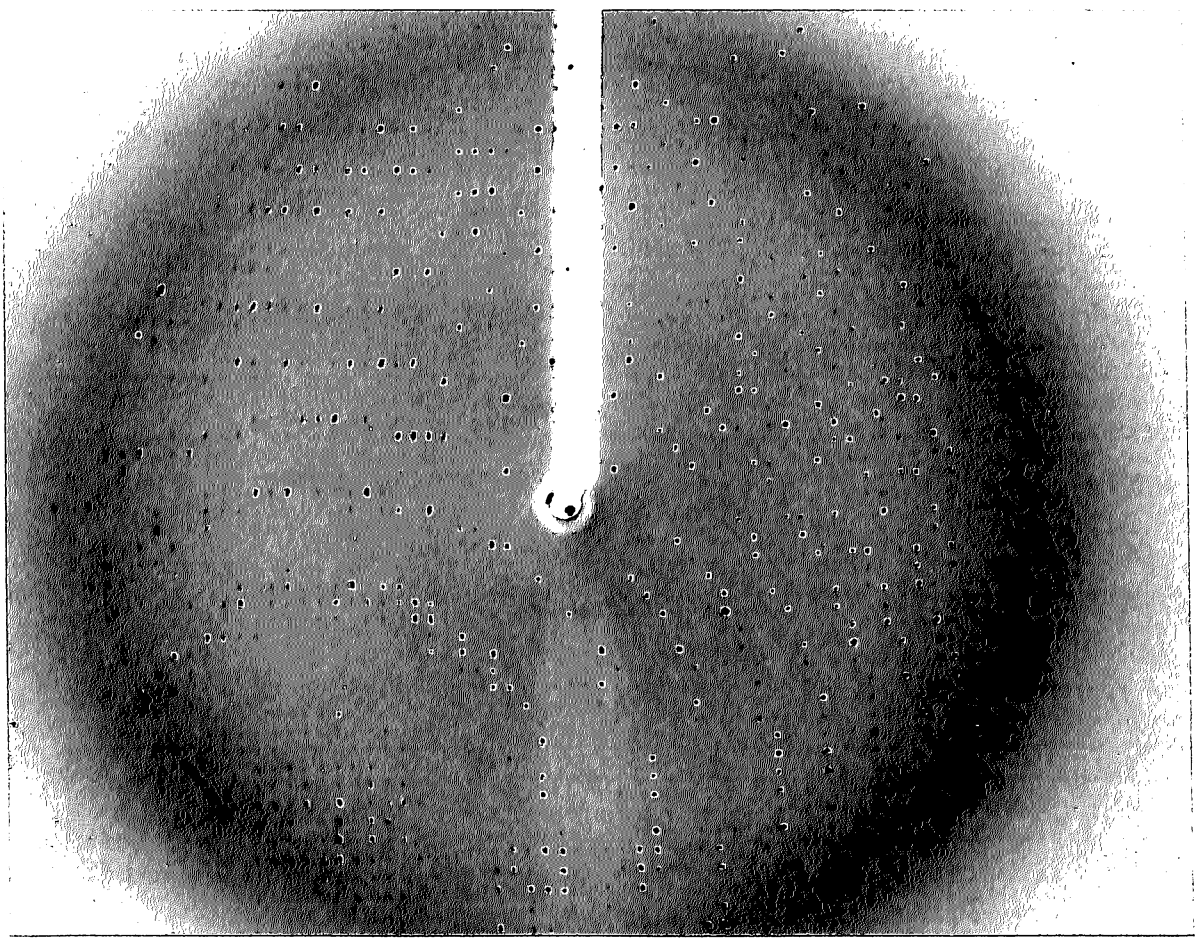
Mosaicity (o)	0.38	0.163	0.27	0.33	0.18	0.16
Sapce group	P41 21 2	P41 21 2	P41 21 2	P41 21 2	P41 21 2	P41 21 2
Cell dimensions (in Å)	a=b=108.1, c=92.3	a=b=107.3, c=91.4	a=b=107.5, c=92.2	a=b=108.1, c=92.1	a=b=107.4, c=91.6	a=b=107.7, c=91.9
No. of Unique reflections	22791	26834	28388	16972	20112	58623
No. of frames	90	90	90	90	90	186

Table - 21

Crystal number	34	38	41_4	41_9	42
Resolution (in Å)	2.24	2.62	Could not index	3.79	3.51
% Completeness	99.6	99.7		99.3	98.9
R_merge (%)	11.9	14.8		18.5	23.6
I/σ	20.94	13.61		13.27	9.01
Mosaicity (o)	0.185	0.15		0.127	0.157
Sapce group	P41 21 2	P41 21 2		P41 21 2	P41 21 2
Cell dimensions (in Å)	a=b=107.7, c=91.9	a=b=107.9, c=91.9		a=b=108.1, c=92.1	a=b=108.7, c=92.3

No. of Unique reflections	49763	31047		10277	13111
No. of frames	360	180	180	180	180

A typical diffraction image from the needle shaped crystal is shown in Figure 14.



[Figure 14. Typical diffraction image from needle shaped crystals.]

Diffraction data from needle shaped crystals soaked in heavy atom compound:

Multi wavelength anomalous diffraction (MAD) data were collected on these heavy atom soaked crystals. These crystals diffracted to ~3.5Å resolution. The data processing summary is shown in Table 22.

Table - 22

	Soaked in Uranyl nitrate			Soaked in Zinc at pH 7.0	
Crystal number	Peak (17.172keV) 43_3	Inflection (17.168keV) 43_4	Peak (17.172keV) 43_6	Peak 51_4	Inflection 51_5
Resolution (in Å)	3.5	3.5	3.2	3.5	3.5
% Completeness	99.5	99.5	99.6	99.6	99.6
R_merge (%)	15.0	16.3	13.7	16.0	15.1
I/σ	13.7	12.82	14.22	12.56	15.53
Mosaicity (o)	0.158	0.164	0.224	0.516	0.265
Sapce group	P41 21 2	P41 21 2		P41 21 2	P41 21 2
Cell dimensions (in Å)	a=b=107.2, c=91.2	a=b=107.3, c=91.2	a=b=107.4, c=91.3	a=b=107.9, c=92.1	a=b=108.3, c=92.4
No. of Unique reflections	12762	12786	16788	13110	13137
No. of frames	180	180	180	180	180

Structure solution:

The structure solution using MAD data from uranyl nitrate, unfortunately, didn't give any solution. Therefore, the structure was solved by molecular replacement using a model of PSP94 obtained in this laboratory previously. It was found that there are four molecules of PSP94 in the crystallographic asymmetric unit. Structure was refined using the software phenix (ref). The refinement statistics is shown in Table 23 and 24.

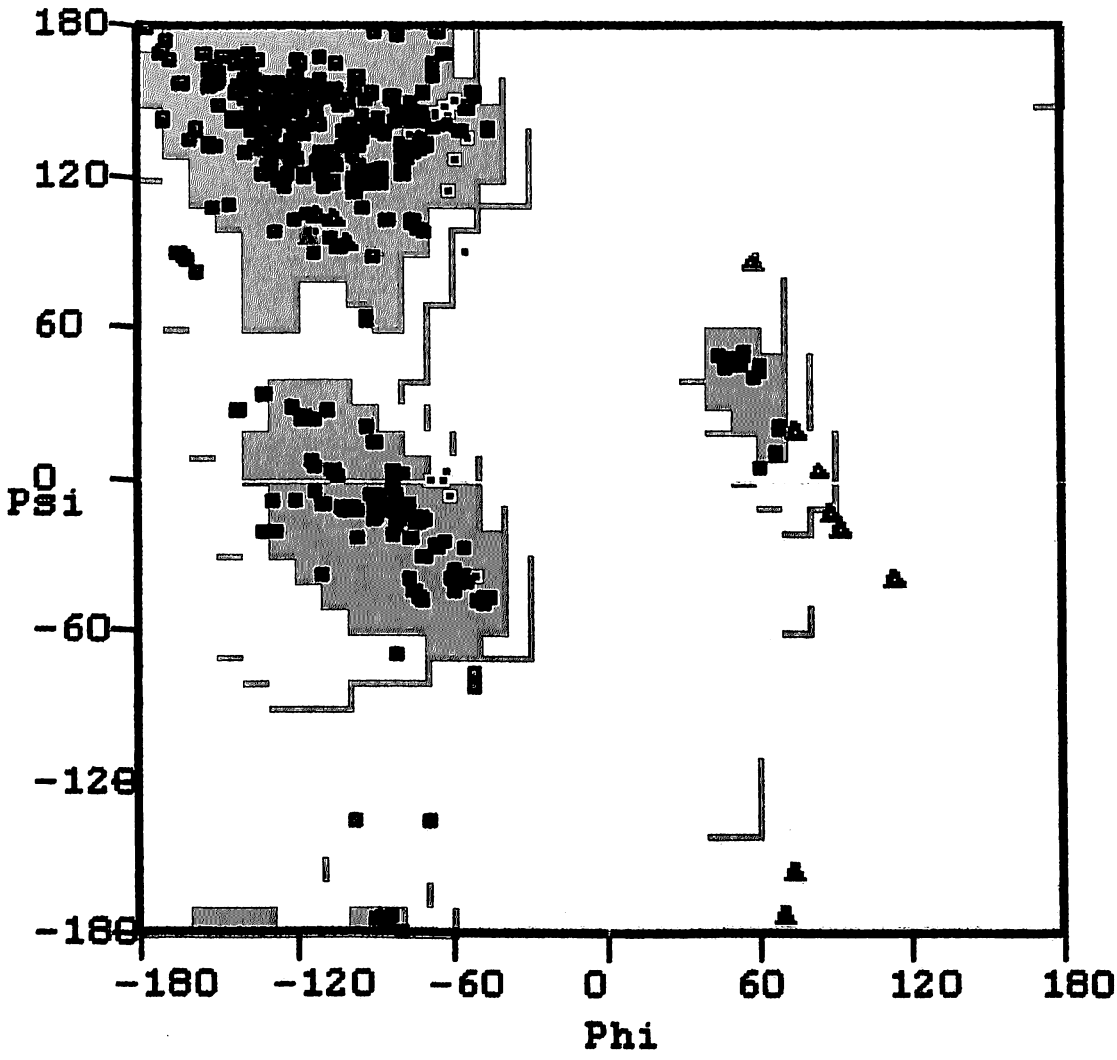
Table - 23

Resolution (Å)	**39.43-1.98**
R/R_{free}[+] **(all data)**	0.22/0.28
Reflections in test set (%)	5.21
No. of reflections	38400
Model statistics	
Non-H atoms	
Protein atoms	2955
Water atoms (full/partial)	363
Geometry: r.m.s. deviations from ideal values	
Bonds (Å)	0.009
Angles (°)	1.171
Planes (Å)	0.006
Chiral centers ($Å^3$)	0.076
Ramachandran plot	

Table - 24

Rsolution Range	Nwork	Nfree	Rwork	Rfree
39.4407 -4.2642	3872	213	0.2319	0.2647
4.2642 -3.3852	3698	203	0.1880	0.2597
3.3852 -2.9575	3656	201	0.2041	0.2691
2.9575 -2.6871	3633	200	0.2067	0.3067
2.6871 -2.4946	3607	198	0.2084	0.2533
2.4946 -2.3475	3597	197	0.2112	0.2767
2.3475 -2.2300	3604	198	0.2126	0.2757
2.2300 -2.1329	3583	197	0.2295	0.2840
2.1329 -2.0508	3568	196	0.2310	0.2906
2.0508 -1.9800	3582	197	0.2549	0.2949

Most of the residues were fitted in the electro density. However the density for residues 8-15 was poor. The ramachandran map, as shown in Figure 4, shows that 91.97% residues are in the most preferred region, whereas 6.37% residues are in the allowed region. There are six residues outside this region which needs to be fixed.



Residues in	
Preferred region:	332 (91.97)
Allowed region:	23 (6.37%)
Outliers:	6 (1.66%

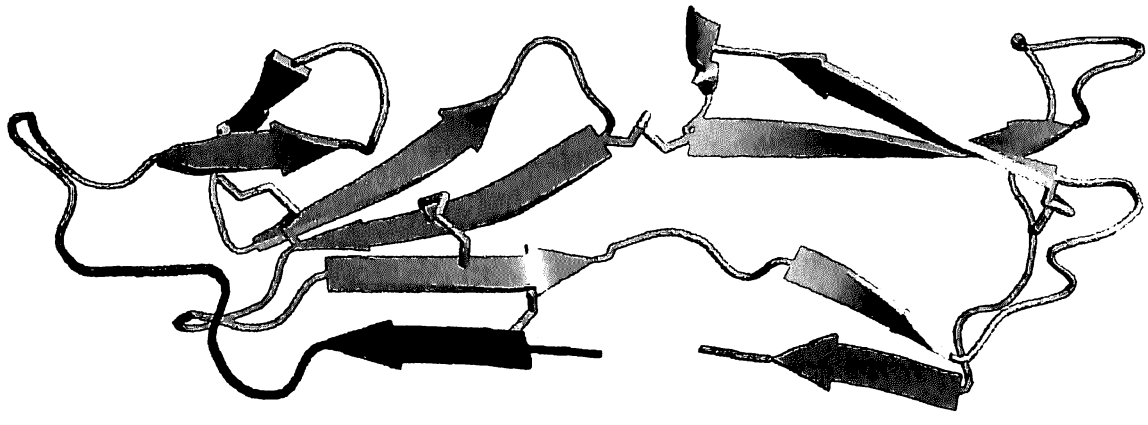
[Figure 15: Ramachandran map for PSP94 model]

DISCUSSION

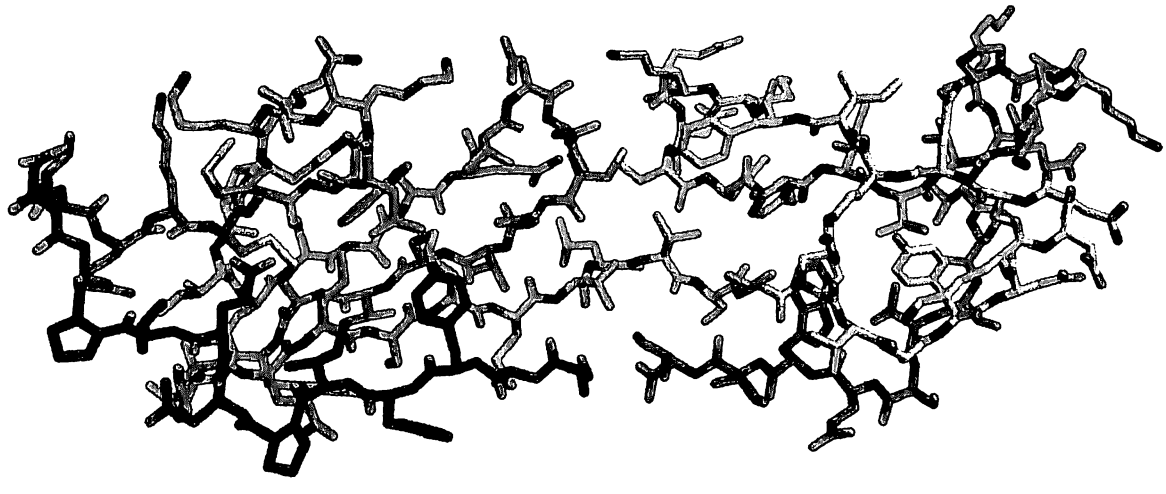
Discussion

5.1. - Structure description:

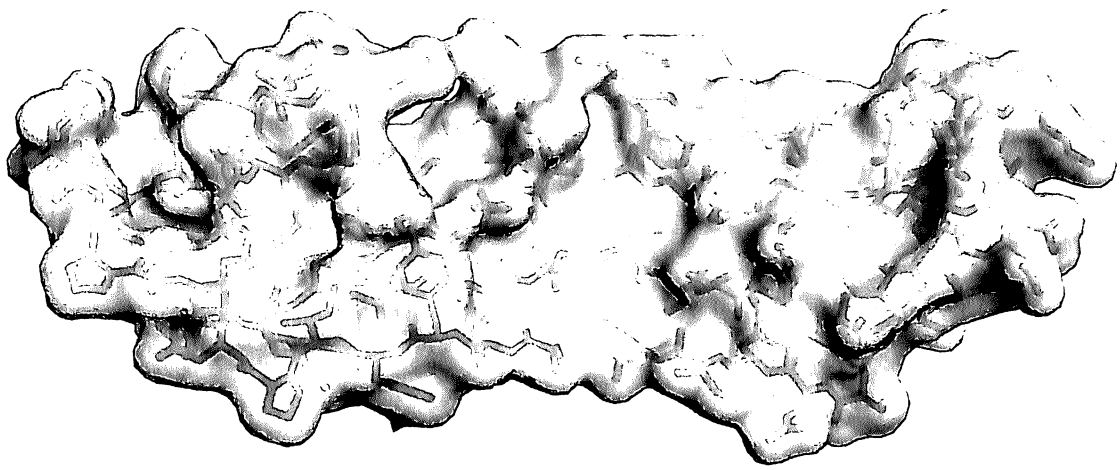
The crystal structure shows that PSP94 has two domains: the N-terminus domain from residues 1-50 and the C-terminus domain from residues 51-94 (Figure 16). The two domains are also held together by a disulfide bond between residues 37 and 74.



[Figure – 16(a): PSP94 structure cartoon representation]

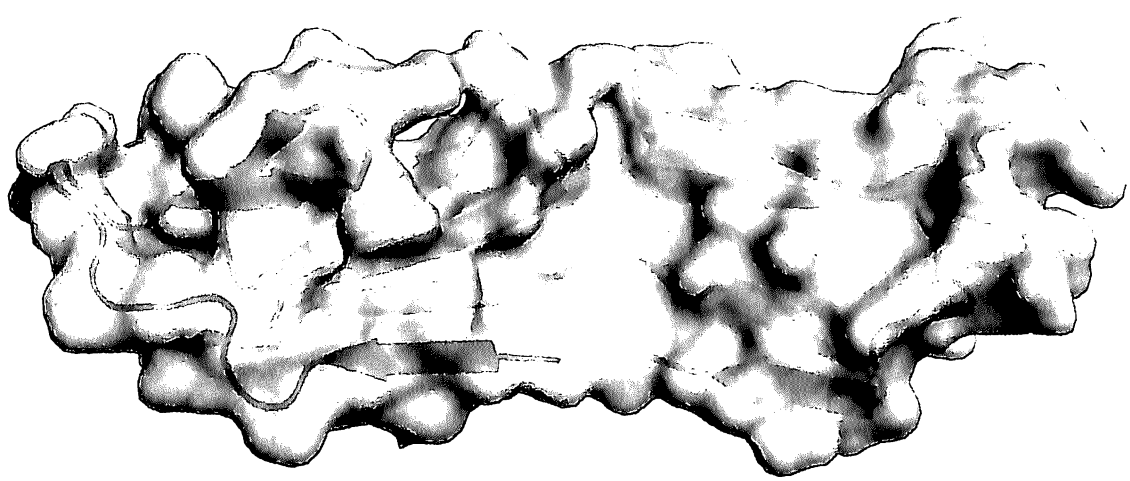


[Figure 16(b): PSP94 Stick representation]

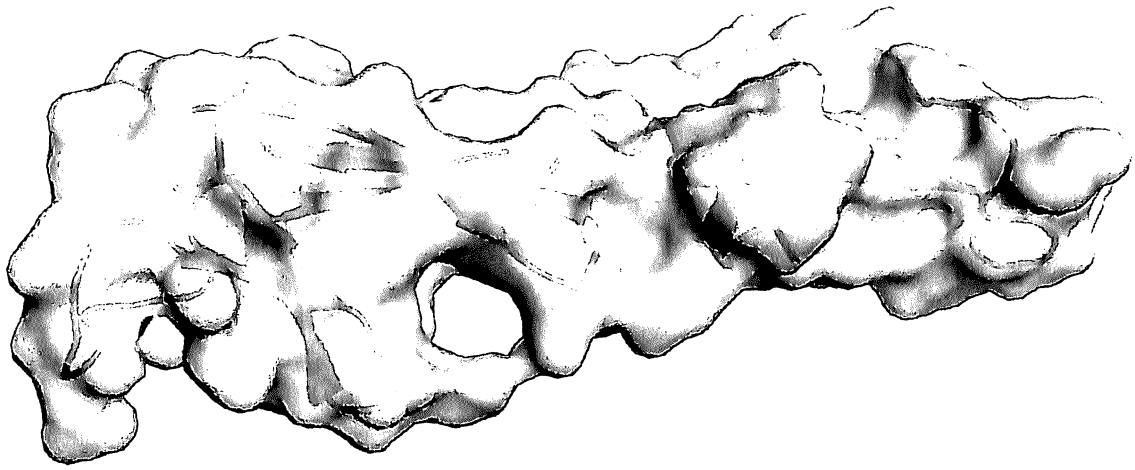


[Figure – 16(c): Surface representation]

The N-terminus domain has four antiparellel β-strand (β1: 1-6, β4: 26-33, β5:35-39 and β6:43-48) arranged in the form of Greek-key motif with two additional β-strand (β2: 17:22 and β3: 23-26) forming a flap on top of the Greek key motif. There are three disulfide bonds in the N-terminal domain. The two disulfides, between the strands β1 and β6 and between the strands β6 and β5 makes the Greek key structure rigid while the third disulfide bond between β5 and β2 orients the flaps onto the Greek key motif. The C-terminal domain has two double-stranded antiparallel β-sheet. The strands β7: 51-57 and β9:91-94 are arranged in tandem with the Greek key motif of the N-terminus domain of the other polypeptide chain in the dimer and forms an extended β-sheet. The other two β-strands β8: 66-72 and β10: 76-82 are longer and separated from the first pair. The lone disulfide between in the C-terminal domain brings the rigidity to the loop structure. Figure 17 shows cartoon representation of PSP94 along with the protein surface in two orthogonal views.



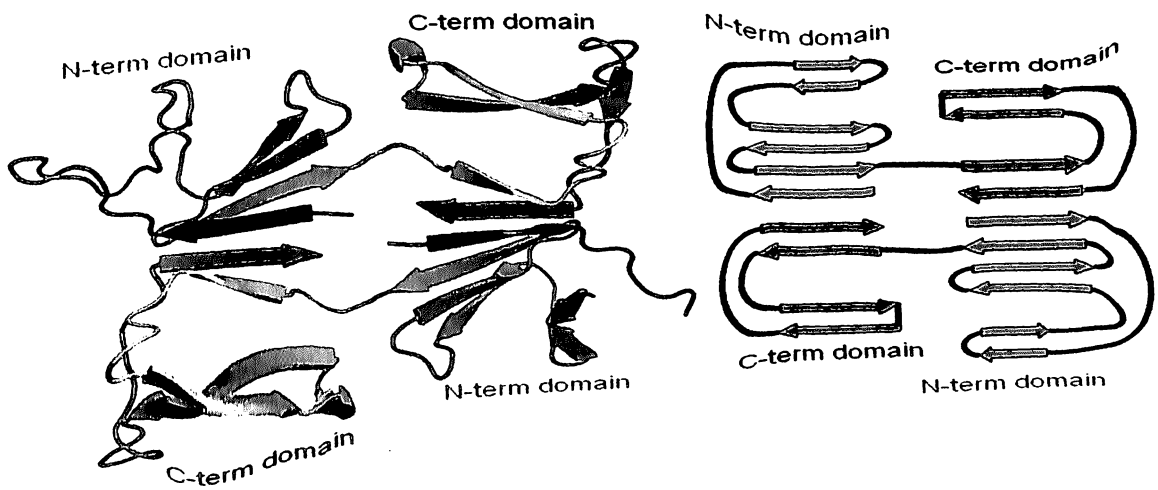
[Figure – 17(a): Cartoon representation of PSP94 inside a semi-transparent surface]



[Figure – 17(b): Cartoon representation of PSP94 inside a semi-transparent surface]

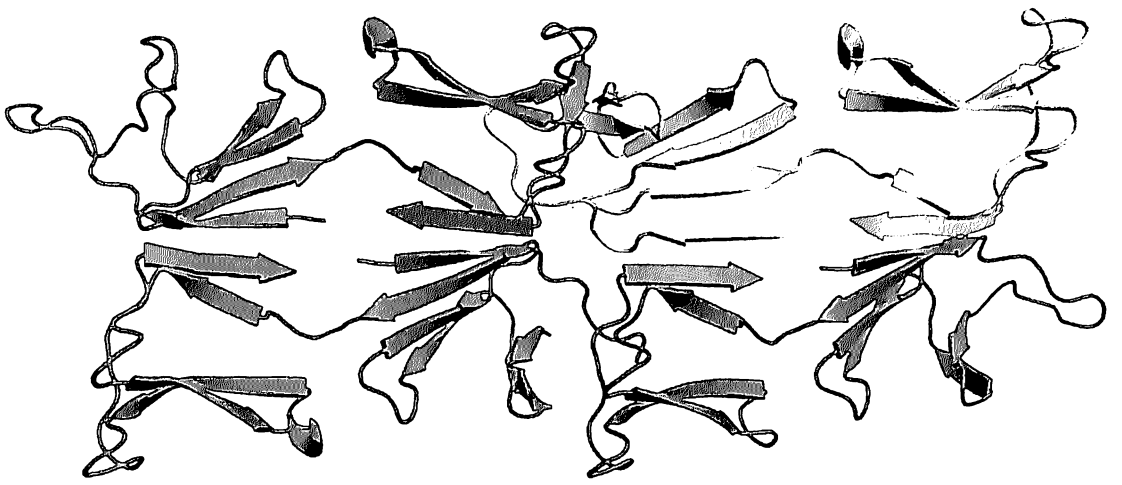
Figure 17: Two orthogonal views are shown in (a) and (b).

Further, the organization of the four molecules in the asymmetric unit shows that there are two dimer pairs related by a translation along the z-axis. There is a non crystallographic dyad axis between the dimers that brings the C-terminus domain of one monomer closer to the N-terminus domain of the other monomer in the dimer. The dimer is thus formed by extension of the β-sheets of the N- and the C-terminal domains by two monomer resulting in a symmetric arrangement of six-stranded β-sheet around the noncrystallographic dyad axis (Figure 18). Also, such an arrangement uniquely brings the N-terminal ends of the two polypeptide chains closer the their C-terminal ends.

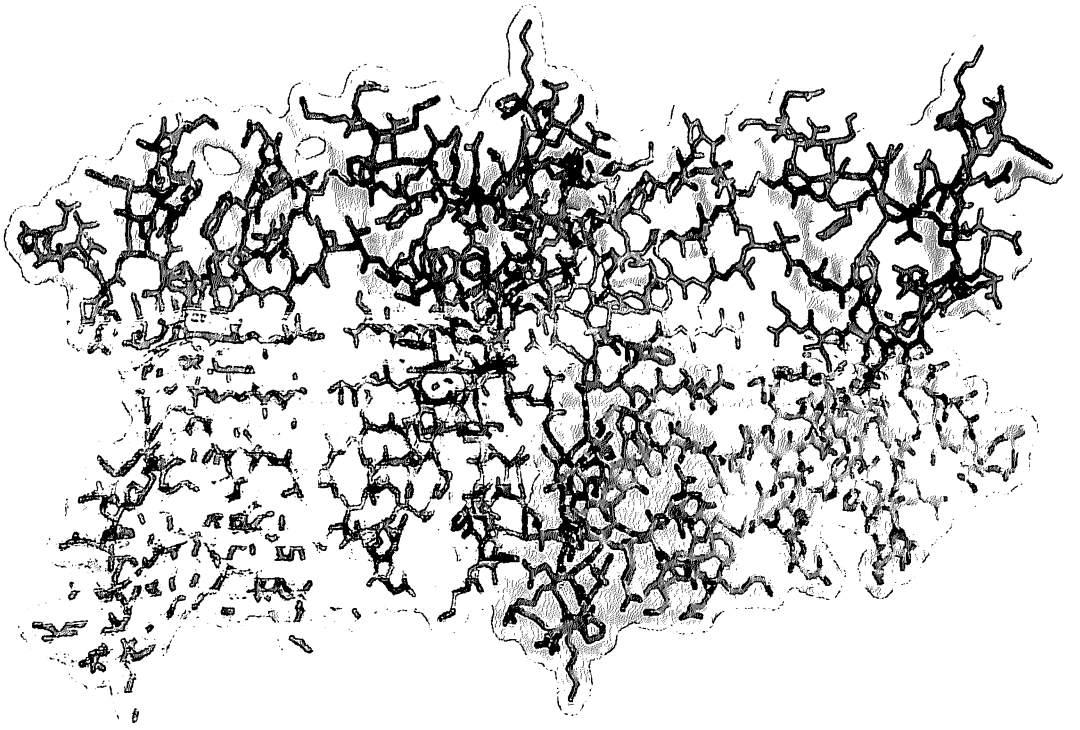


[Figure – 18: PSP94 Dimer and topology diagram]

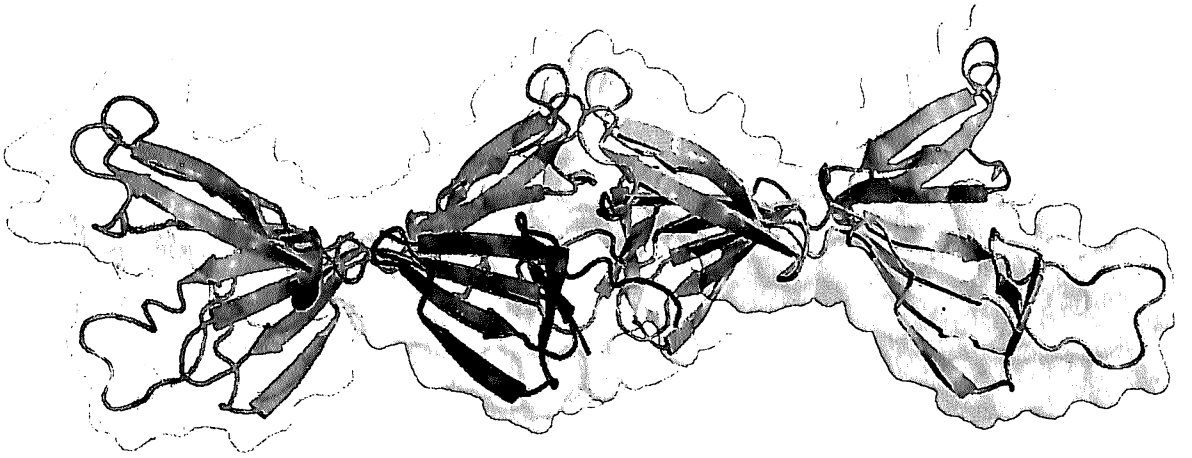
The four polypetdie chain in the asymmetric unit is shown in Figure 19 alongwith the surface representation in two orthogonal views.



[Figure – 19(a): Four molecules of PSP94 present in the asymmetric unit are colored differently]



[Figure – 19(b): Four molecules of PSP94 present in the asymmetric unit Surface representation]



[Figure – 19(c): Four molecules of PSP94 present in the asymmetric unit orthogonal view]

To look for the structural similarity of PSP94 with other known proteins Dali search was carried out. The top twenty matches from the Dali search are listed in Table 25. Structure superposition showed that the N-terminal domain of PSP94 has structural similarity with the fibronectin F1 module.

Table - 25. Dali search results.

	PDB id-Chain	Z-score	RMSD	Naligned	Nres	%id	Description
1.	1e8b-A	4.7	7.3	64	160	14	FIBRONECTIN;
2.	3ejh-B	4.7	4.4	52	87	10	FIBRONECTIN;
3.	1e88-A	4.6	1.7	38	160	21	FIBRONECTIN;
4.	2cg6-A	4.6	5.2	66	90	17	HUMAN FIBRONECTIN;
5.	1o9a-A	4.5	2.7	43	93	23	FIBRONECTIN;
6.	2rkz-F	4.4	2.1	40	89	20	FIBRONECTIN;
7.	2rkz-C	4.4	2.7	41	90	20	FIBRONECTIN;
8.	2cg7-A	4.3	2.4	40	90	20	FIBRONECTIN;
9.	3cal-A	4.3	2.9	43	88	19	FIBRONECTIN;
10.	3cal-C	4.3	3.0	43	87	19	FIBRONECTIN;

11.	2rkz-D	4.2	3.2	43	86	19	FIBRONECTIN;
12.	1tpn-A	4.1	2.9	44	50	16	TISSUE-TYPE PLASMINOGEN ACTIVATOR;
13.	1fbr-A	4.1	4.3	58	93	17	FIBRONECTIN;
14.	1tpn	4.1	2.9	44	50	16	TISSUE-TYPE PLASMINOGEN ACTIVATOR (TYPE 1 FIBRIN-BINDING
15.	2rkz-B	4.1	2.8	41	89	20	FIBRONECTIN;
16.	2rkz-A	4.0	2.8	41	89	20	FIBRONECTIN;
17.	1fbr	3.9	4.4	54	93	19	FIBRONECTIN;
18.	1tpm	3.9	2.7	44	50	16	TISSUE-TYPE PLASMINOGEN ACTIVATOR (TYPE 1 FIBRIN-BINDING
19.	1tpm-A	3.9	2.7	44	50	16	TISSUE-TYPE PLASMINOGEN ACTIVATOR;
20.	1wv3-A	3.8	3.0	50	186	6	SIMILAR TO DNA SEGREGATION ATPASE AND RELATED

The structure superposition of fibronectin F1 module on to the PSP94 N-term domain is shown in Figure 20.

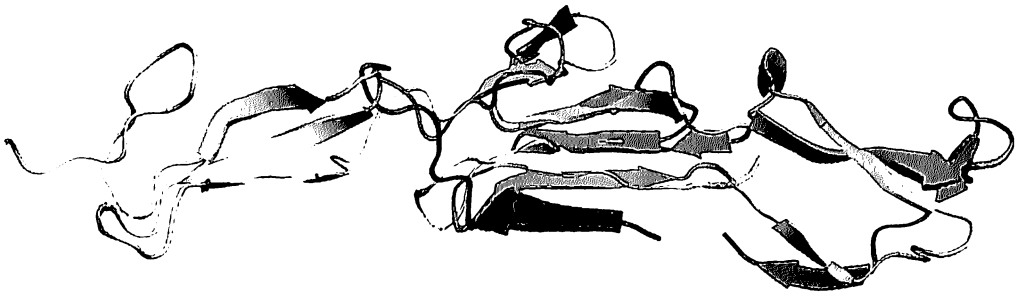


Figure – 20 (a)

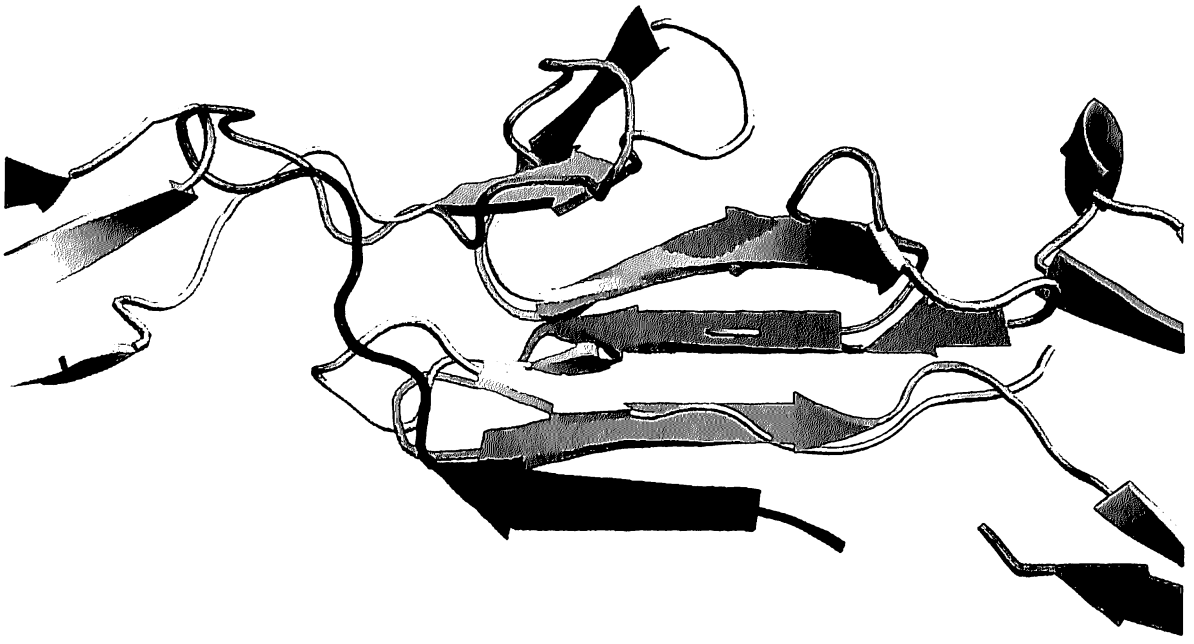


Figure – 20 (b)

Figure 20. (a) Superposition of PSP94 and fibronectin structure, (b) closer view of the superposition.

There is no match for the C-terminal domain of PSP94 with any known protein structure. In conclusion, PSP94 has a unique overall fold. There is only a partial structure similarity of the N-terminal domain with fibronectin F1 module.

SUMMARY

SUMMARY

Human Seminal Plasma Protein PSP94 is a Small Protein of 94 Amino Acids. Several potential Biological function reported by many researchers. It is also related to prostrate cancer, so for Structure based Drug Designing purpose the structure of the protein is very essential. Crystallization process is a very difficult process. The protein is crystallized by vapour diffusion method prior to which screening process was carried out by the CyBio Robot. For structure determination the X-ray radiation was performed at SLS, Switzerland on Synchrotron beam light. The data was collected indexed integrated and scaled by the program MOSFLM & SCALA. To solve the phase problem the heavy atom method is used. Structure solubility was done by using coot & Phenix. The crystal shows that PSP94 has two domains. The N terminus domain is from residue 1-50 and the C terminus domain is from 51-94. The two domains also held together by Di-Sulphide bonds. To look for the structural similarity of PSP94 with other known proteins Dali Search was carried out. The top 20 matches are from Dali. Structure superposition showed domain of PSP 94 similarity with fibronectine F1 module.

APPENDIX

0.1 M Lithium sulphate, 0.1 sodium acetate PH-4.5

PEG 400	OBSERVATION
40%	No crystal
40.5%	No crystal
41.0%	No crystal
41.5%	No crystal
42.0%	No crystal
42.5%	No crystal

Table-1

0.2 M Lithium sulphate, 0.1 sodium acetate PH-4.7

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	Very small thin, **needle** shaped crystals, more in number
46%	Very small thin, **needle** shaped crystals, more in number
48%	No crystal
50%	No crystal

Table-2

0.2 M Lithium sulphate, 0.1 sodium acetate PH-4.9

PEG 400	OBSERVATION
40%	Very small thin, **needle** shaped crystals, more in number
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	No crystal

Table-3

0.1 M Lithium sulphate, 0.1 sodium acetate PH-5.1

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	Small **square** shaped crystals with precipitation

Table-4

0.2M Lithium sulphate, 0.1 M sodium acetate PH-5.1

PEG 400	OBSERVATION
49%	No crystal
49.2%	No crystal, precipitation
49.4%	Small **square** shaped crystals with precipitation
49.6%	Small **square** shaped crystals with precipitation
49.8%	Small **square** shaped crystals with precipitation
50%	Small **square** shaped crystals with precipitation

Table-5

0.2M Lithium sulphate, 0.1 M sodium acetate PH-4.9
JBS screen + volatiles

PEG 400	OBSERVATION
49%	No crystal
49.2%	No crystal, precipitation
49.4%	No crystal
49.6%	No crystal
49.8%	No crystal
50%	No crystal

Table-6

0.2 M Lithium sulphate, 0.1 M sodium acetate PH-4.7

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal, precipitation
44%	No crystal
46%	No crystal
48%	No crystal
50%	Very small thin, **needle** shaped crystals, more in number

Table-7

0.2M Lithium sulphate, 0.1 M sodium acetate PH-4.9

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal, precipitation
44%	No crystal
46%	No crystal
48%	No crystal
50%	Small **square** shaped crystals with precipitation

Table-8

0.2M Lithium sulphate, 0.1 M sodium acetate PH-4.9
JBS screen + additives

PEG 400	OBSERVATION
40%	No crystal
42%	No crystal
44%	No crystal
46%	No crystal
48%	No crystal
50%	Small **square** shaped crystals with precipitation

Table-9

0.2M Lithium sulphate, 0.1 M sodium acetate PH-4.5
JBS Screen + additives

PEG 400	OBSERVATION
40%	No crystal
40.5%	No crystal
41%	No crystal
41.5%	No crystal
42%	No crystal
42.5%	No crystal

Table-10

0.2M Lithium sulphate, 0.1 M sodium acetate PH-4.5
JBS Screen + C_1 volatile.

PEG 400	OBSERVATION
42.5%	No crystal
42.7%	No crystal
42.9%	No crystal
43.1%	No crystal
43.3%	No crystal
43.5%	Very small , thin needle shaped crystals

Table- 11

0.3 M Lithium sulphate, 0.1 M sodium acetate PH-4.9
JBS Screen + C_1 volatile

PEG 400	OBSERVATION
48%	No crystal
48.4%	No crystal
48.8%	No crystal
49.2%	No crystal
49.6%	Small **square** shaped crystals more in number
50 %	Small **square** shaped crystals more in number

Table-12