Chapter 1

INTRODUCTION

The human body contains thousands of proteins which play essential roles in maintaining life. A protein's structure determines the specific role that it plays in the human body. Crystallography, through the study of protein crystals, helps determine the three-dimensional molecular structures of proteins, some viruses, and other large molecules. With an improved understanding of the molecular structures and interactions of proteins, drug designers may be able to develop new drug treatments that target specific human, animal, and plant diseases.

To determine a protein's structure, it is necessary to first crystallize the protein and analyze the resulting crystals by X-ray diffraction. Precise measurements of thousands of diffracted intensities from each crystal are used to map the probable positions of the atoms within each protein molecule. The structures of many important proteins remain a mystery simply because it is very difficult to obtain crystals of high enough quality or large enough size. Generally, crystals must have dimensions of approximately 0.05 mm or larger, single, free of cracks and defects, and the protein molecules must be arranged in an orderly, repeating pattern.

1.1 BIOMACROMOLECULES : PROTEIN

Living organisms contain 70% water, 15% proteins, 7% nuclic acids, 5% polysacharids, lipids and precursors, 1% small molecules and 1% inorganic ions. The biomacromolecules crucial for life- proteins and nuclic acids, 22% of the total are produced, perform their functions and degrade in these aqueous solutions. They are the material components of the chemical life machinery. The properties of the macromolecules are determined by their three-dimensional (3D) structures which follow the molecule composition.

Each protein molecule is a specifically folded chain (or several chains) built of 20 amino acids. The amino acid is small organic molecule with a molecular weight of about

100 Da (Daltons). Each amino acid includes the carboxyl ($COOH^-$) and amino (NH_2^+) groups and a specific side chain. When the carboxyl group belonging to one amino acid reacts with the amino group of the other, they form the strong covalent-ionic peptide bond $CONH_2$, releasing water (H_2O). The protein chain is thus the polypeptide chain.

Among the 20 amino acids, 5 are acidic or basic, i.e. their side chains easily lose either the proton H^+ or the OH^- group and become charged. Therefore, they are highly hydrophilic. Less hydrophilic are the 7 amino acids with polar side chains. The remaining 8 amino acids contain hydrocarbon groups and are hydrophobic. Therefore, in a physiological solution, pH ≈7, the protein chain folds to form the molecular globulae, with the hydrophilic and polar amino acids sitting on the surface and hiding the hydrophobic amino acids in the inner core.

1.2 PROTEIN STRUCTURE

Proteins are an important class of biological macromolecules present in all organisms. All proteins are polymers of amino acids. Classified by their physical size, proteins are nanoparticles. Each protein polymer, also known as a polypeptide, consists of a sequence of 20 different L-α-amino acids, also referred to as residues. For chains under 40 residues the term peptide is frequently used instead of protein. To be able to perform their biological function, proteins fold into one or more specific spatial conformations, driven by a number of non-covalent interactions such as hydrogen bonding, ionic interactions, Van Der Waals forces, and hydrophobic packing. To understand the functions of proteins at a molecular level, it is often necessary to determine their three-dimensional structure.

A protein is made by arranging amino acids together in a specific sequence, which is different for each protein. These amino acids are held together by a special bond called a peptide bond. There are four distinct levels of protein structure (see Figure 1.1):

Primary structure The primary structure of a protein is determined by the number and order of amino acids within a polypeptide chain. A polypeptide is a sequence of two or more amino acids joined together by peptide bonds. The two ends of the polypeptide chain are referred to as the carboxyl terminus (C-terminus) and the amino terminus (N-terminus) based on the nature of the free group on each extremity. Counting of residues starts at the N-terminal end (NH2-group), which is the end where the amino group is not involved in a peptide bond.

Secondary structure Secondary structure refers to highly regular local substructures. Two main types of secondary structure are the α-helices and the β-strands or β-pleated sheets. These secondary structures are defined by patterns of hydrogen bonds between the main-chain peptide groups. **Tertiary structure** Tertiary structure refers to the three-dimensional structure of a single protein molecule. This structure of

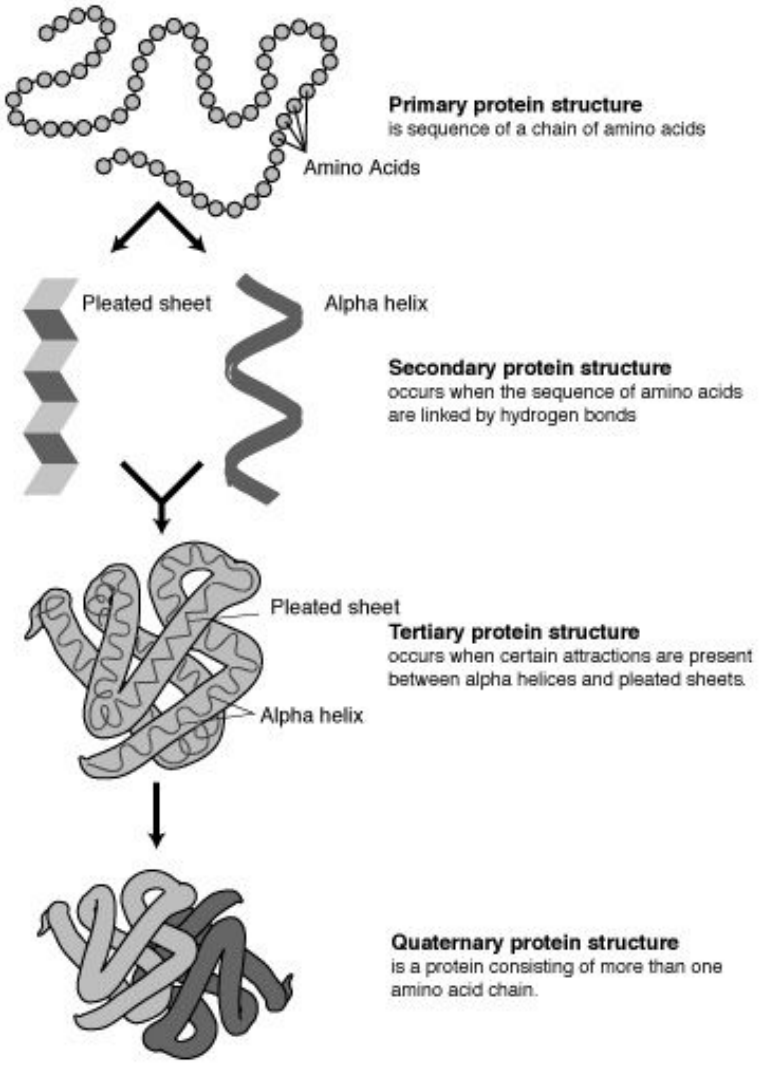


Figure 1.1: Different structures of protein. (*http://vls.wikipedia.org/wiki/Ofbeeldienge:Protein-structure.png*)

a protein is formed when the attractions of side chains and those of the secondary structure (alpha-helices and beta-sheets) are folded into a compact globule. This causes the amino acid chain to form a distinct and 3-dimensional structure which gives a protein its specific function.

Quaternary structure Quaternary structure is a larger assembly of several protein molecules or polypeptide chains, usually called sub-units. The quaternary structure is stabilized by the same non-covalent interactions and disulfide bonds as the tertiary structure. Complexes of two or more polypeptides (i.e. multiple sub-units) are called multimers, for example, dimer, trimer, tetramer etc. Many proteins do not have the quaternary structure and function as monomers. Lysozyme serves as a model of protein

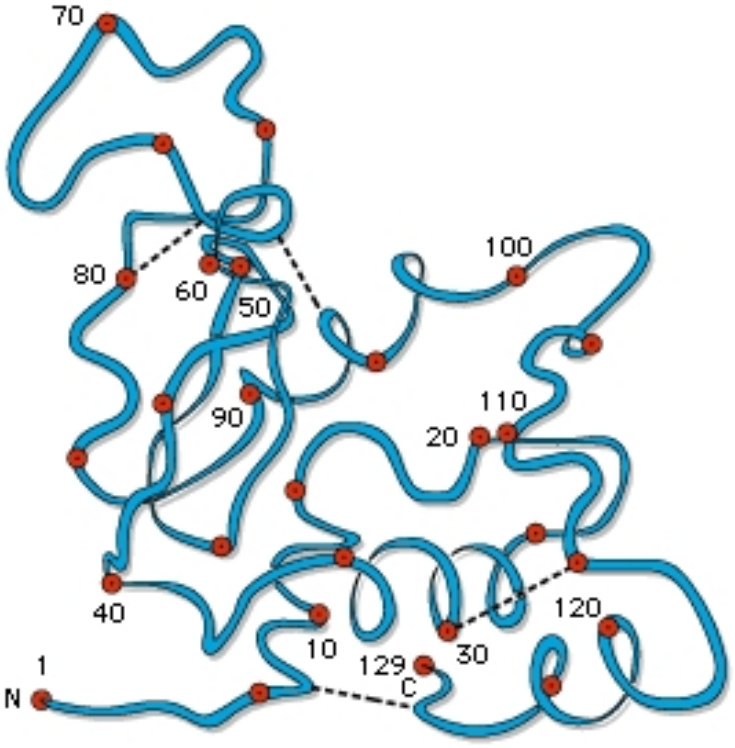


Figure 1.2: The simplified structure of lysozyme from hen's egg white has a simple peptide chain of 129 amino acids. Circles indicate every fifth residue; every tenth residue is numbered. Broken lines indicate the four disulfide bridges. Alpha-helices are visible in the ranges 25 to 35, 90 to 100, and 120 to 125. (*http://www.britannica.com/EBchecked/media/3207/Conformation-of-lysozyme*)

structure and function. Lysozyme or 1-4-N-acetylmuramidase is a single polypeptide of 129 amino acid residues enzyme, having a mass of 14.6 kDa and contains four disulfide bonds. Figure 1.2 shows the simplified structure of lysozyme protein. It destroys bacterial cell walls by catalyzing hydrolysis of the polysaccharide component of the cell wall. Lysozyme is a component of tears and mucous and is believed to serve a protective role in defending the body against productive bacterial infection. Most of the lysozyme

used for research comes from egg whites. This protein serves many of the current protein crystallization research, and a lot has been discovered about the structure and its functioning. When lysozyme is crystallized, it takes on tetragonal form.

These biomacromolecules are overwhelmingly crystallized from an aqueous solution. Their solubility decreases when the precipitant- an organic salt, polyethelene glycol or a buffer that changes pH- is added to the biomacomolecular solution.

1.3 CRYSTALS

Crystals are solids that show highest structural uniformity and purity. More specifically, a single crystal consists of atomic arrays that are periodic in three dimensions with equal spacing in a given direction. Real crystals are finite in size and contain defects. However a single crystal is a solid that is uniform in structure and free of defects, to the extent possible. An ideal crystal is one in which the surroundings of any atom would be exactly the same as the surroundings of every similar atom. The uniformity of single crystals can allow transmission without scattering of electromagnetic waves. This property makes it possible for crystals to be studied using x-ray diffraction technique. Research on crystal growth has intensified in recent years owing to their vivid physical and chemical properties. Large and defect-free crystals find applications in areas ranging from high power lasers (optical crystals) to biomedical applications, for example, protein crystals. Therefore, it can be said that a crystal is an ordered, three-dimensional periodic array, whose particular thermodynamic and kinetic properties are specific to the protein under study. Consider crystal grown from an aqeous solution. At supersaturation the solute partitions into the solid and the soluble phases. Although entropy is reduced, there are a high number of specific contacts made to make the free energy favorable. For small molecules, these properties may be predicted, but not for macromolecular systems. Therefore, the characterization of macromolecular crystal is still an empirical process.

At supersaturation, the solid phase can have either crystal or disordered amorphous state. Sometimes, there is an equilibrium between amorphous states to solubility that allow for redirection to the crystalline solid. The strategy to obtain crystals, is to slowly approach the supersaturation (limited solubility) conditions. Some additives may be added to encourage the formation of the maximum amount of favorable interactions. In the case of proteins, there are many possible approaches and hence the physiochemical properties of the protein are different in each case. Therefore, the space group (type of arrangement formed) is likely to be different in each condition.

In crystallography, the space group (crystallographic group or Fedorov group) of a crystal is defind as a description of the symmetry of the crystal. It can have one of 230

types.

Crystal systems are characterized by a three dimensional repeating unit composed of translations. These are further characterized by lengths of the edges and angles between the axes. There are seven fundamental types of unit cells, each of which defines a crystal system.

1. **Triclinic**: a ≠ b ≠ c and $\alpha \neq \beta \neq \gamma$
2. **Monoclinic**: a ≠ b ≠ c and $\alpha = \gamma, \beta \neq 90^o$
3. **Orthorhombic**: a ≠ b ≠ c and $\alpha = \beta = \gamma = 90^o$
4. **Rhombohedral**: a = b = c and $\alpha = \beta = \gamma < 120^o$ (Trigonal: a = b ≠ c and $\alpha = \beta = \gamma = 120^o$)
5. **Tetragonal**: a = b ≠ c and $\alpha = \beta = \gamma = 90^o$
6. **Hexagonal**: a = b ≠ c and $\alpha = \beta = 90^o, \gamma = 120^o$
7. **Cubic**: a = b = c and $\alpha = \beta = \gamma = 90^o$

Based on these fundamental types of unit cells, there are fourteen Bravais lattices.

1.4 SUPERSATURATION OF A SOLUTION

In order to grow crystals from their aqueous solution, the initial solute concentration should be more than the equilibrium concentration. Supersaturation is thus the driving force that governs the rate of crystal growth. The supersaturation of a system may be expressed in a number of ways. The degree of supersaturation of a solution is expressed in terms of the absolute supersaturation σ, defined as

$$\sigma = C - C_0 \tag{1.1}$$

where C is the concentration of the dissolved substance and C_o is its solubility limit. The degree of supersaturation can also be expressed in terms of the relative supersaturation γ, defined as

$$\gamma = \frac{C - C_0}{C_0} \tag{1.2}$$

If the concentration of a solution can be measured at a given temperature and the corresponding equilibrium saturation concentration is known, the extent of supersaturation can be obtained.

A typical solubility curve of a protein is shown in Figure 1.3. The precipitant - protein plane is divided by the solubility curve (solid) into two regions; unsaturated and supersaturated. The supersaturated region support crystal growth whereas the unsaturation region dissolves crystals. Precipitant and protein concentration data at which spontaneous crystallization occurs are represented by the broken curve. The diagram is now divided into four zones:

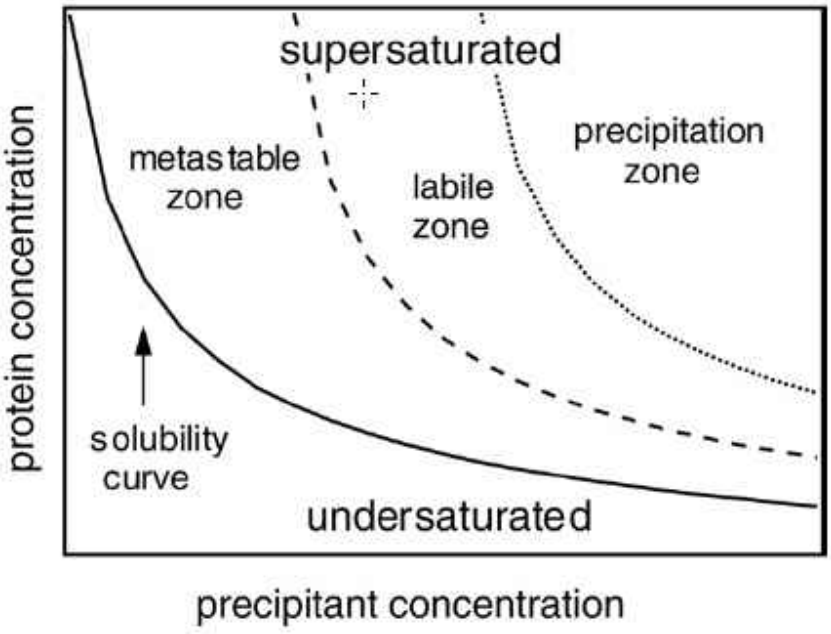


Figure 1.3: The solubility curve shows the regions that support crystal growth (supersaturated) from those that dissolve crystals (unsaturation) (Asherie, 2004).

1. The stable (unsaturated) zone, where crystallization is not possible.
2. The metastable zone, between the solubility and supersolubility curves, where spontaneous crystallization is improbable and a seed crystal is essential to facilitate crystal growth.
3. The third region is the unstable or labile (supersaturation) zone, where spontaneous crystallization, nucleation and growth is more probable.
4. The fourth region is the precipitation region, where clusters are more probable.

When the crystallization of proteins is under vapour diffusion, protein content goes from unsaturated to saturated concentration. Crystals nucleate and the concentration of protein decreases to metastable region for crystal growth.

1.5 PROTEIN CRYSTALS

The most widely used means for protein structure studies is through x-rays, electron, or neutron diffraction patterns produced by protein crystals. The importance of protein crystallization is that it serves as the basis for X-ray crystallography, wherein a crystallized protein is used to determine the protein's three-dimensional structure via X-ray diffraction. To resolve atoms that are, typically, 1.5 - 2 Å apart, these diffraction methods require single crystals as large as several tenths of a millimeter in all the three dimensions, of low defect contents and high compositional and structural unifor-

mity. Large, well-ordered crystals of various proteins can be produced under controlled conditions in microgravity.

Crystal growth is a controlled phase transformation process either from the solid, the liquid or the gaseous state to the solid phase. The growth units viz. the atoms or molecules, diffuse to the growth site from the mother phase, when given sufficient time to be arranged on the lattice.

Proteins can be prompted to form crystals when placed in the appropriate conditions. In order to crystallize a protein, the purified protein undergoes slow precipitation from an aqueous solution. As a result, individual protein molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation. The crystalline "lattice" that forms is held together by non-covalent interactions.

The goal of crystallization is usually to produce a well-ordered crystal that is lacking in contaminants and large enough to provide a diffraction pattern when hit with x-ray which can then be analyzed to discern the protein's three-dimensional structure. Because of the fragile nature of protein crystals, their crystallization is difficult. Proteins have irregularly shaped surfaces, which results in the formation of large channels within any protein crystal. Therefore, the non-covalent bonds that hold together the lattice must often be formed through several layers of solvent molecules. In addition to overcoming the inherent fragility of protein crystals, the successful production of x-ray diffraction worthy crystals is dependent upon a number of environmental factors, because so much variation exists among proteins, with each individual protein requiring unique condition for successful crystallization. Therefore, attempting to crystallize a protein without a proven protocol can be very tedious process. Some factors that require consideration are protein purity, pH, and concentration of protein, temperature, and precipitants. In order for sufficient homogeneity, the protein should usually be at least 97% pure. pH conditions are also very important, as different pH values can result in different crystal structures. Buffers, such as Tris-HCl and sodium acetate, are often necessary for the maintenance of a particular pH. Precipitants, such as ammonium sulphate and sodium chloride, are compounds that cause the protein to precipitate out of solution. Polyethylene glycol is commonly used as the cryoprotectant. Sodium azide (NaN_3) is an anti-microbial preservative that is sometimes used to protect samples and crystallization reagents from microbial contamination. Sodium azide is toxic and should be handled with care. Typical sodium azide concentrations are between 0.02% and 0.1% w/v.

1.6 PROTEIN CRYSTALLIZATION METHODS

Protein crystallization is inherently difficult because of the fragile nature of protein crystals. The most commonly used method to grow the protein crystal is vapour diffusion method. There are several techniques for setting up crystallization experiments (often termed "trials") including sitting drop vapour diffusion, hanging drop vapour diffusion, sandwich drop, batch, microbatch, under oil, microdialysis, and counter diffusion or free interface diffusion. Here an overview of these techniques is provided.

1.6.1 Hanging Drop Crystallization

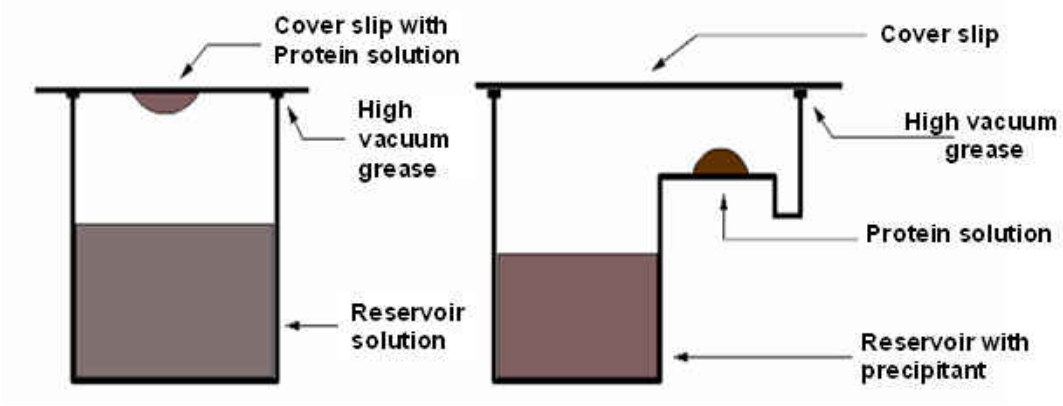


Figure 1.4: (a) Hanging drop and (b) Sitting drop vapour diffusion configuration.

The hanging drop vapour diffusion technique is the most popular method for the crystallization of macromolecules. The principle of vapour diffusion is straightforward. In this technique (Figure 1.4(a)) a small (1 to 40 μl) droplet of the protein sample mixed with crystallization reagent (precipitant) and buffer is placed on a glass cover slide inverted over the reservoir in vapour equilibrium with the reagent. The system must be sealed off from the outside by using high-vacuum grease between glass surfaces. The initial reagent concentration in the droplet is less than that in the reservoir. To achieve equilibrium, water vapour leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop in a vapour phase for the reservoir and an equilibrium will exist between the drop and the reservoir. During this equilibration process the sample is also concentrated, increasing the relative supersaturation of the sample in the drop.

The advantages of the hanging drop technique include the ability to view the drop through glass without the optical interference from plastic, flexibility, reduced chance of crystals sticking to the hardware, and easy access to the drop. The disadvantage is that

a little extra time is required for set up.

1.6.2 Sitting Drop Method

The sitting drop vapour diffusion technique is a popular method for the crystallization of macromolecules. In sitting drop technique, (Figure 1.4(b)) a small (1 to 40 μl) droplet containing purified protein, buffer, and precipitant is allowed to equilibrate with a larger reservoir containing similar buffers and precipitants in higher concentrations. It is placed on a platform in vapour equilibrium with the reagent. The initial protein concentration in the droplet is less than that in the reservoir. The reservoir contains similar buffers and precipitants in higher concentrations. To achieve equilibrium, water vapour leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. During this equilibration process the sample is also concentrated, increasing the relative supersaturation of the sample in the drop. The system must be sealed off from the outside by using high-vacuum grease between glass surfaces.

The advantages of the sitting drop technique include speed and simplicity. It allows one to avoid greasing plates. The disadvantages are that crystals can sometimes adhere to the sitting drop surface making removal difficult. This disadvantage can turn into an advantage where occasionally the surface of the sitting drop can assist in nucleation. This method is suitable if the drop has very low surface tension, making it hard to turn upside down or if the drops need to be larger than 20μl.

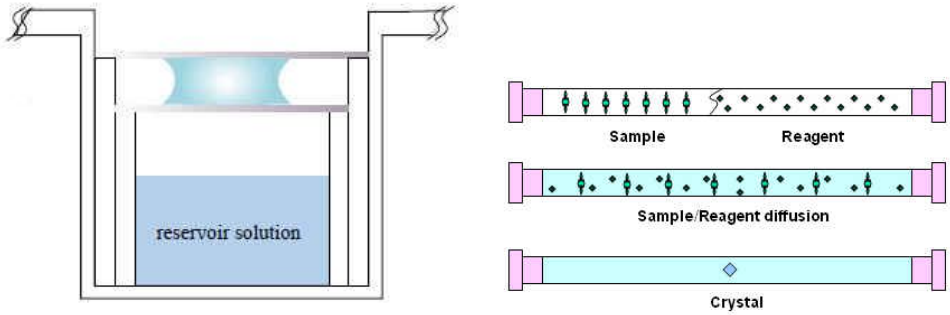
The hanging drop method differs from the sitting drop method in the vertical orientaion of the protein solution drop within the system. Both the methods require a closed system.

1.6.3 Sandwitch Drop Crystallization

For sandwich drop crystallization method, the protein solution mixed with the precipitant is placed in the middle of a lower siliconized glass cover slide followed by a larger siliconized glass cover slide in position along an upper edge. It allows for a small amount of space between the cover slides but is close enough such that the drop is sandwiched between the glass cover slides (Figure 1.5(a)).

1.6.4 Counter Diffusion or Free Interface Diffusion Method

Counter diffusion or free interface diffusion crystallization is one of the methods used by NASA in microgravity crystallization experiments. This method is based on the



(a) (b)

Figure 1.5: (a) Sandwitch drop crystallization and (b) Free interface diffusion configuration. Image courtesy of Hampton Research USA

diffusion of molecules of protein and precipitant when they are arranged to counter flow one against the other. The precipitating agent and protein solutions are placed together either in direct contact or separated by an intermediate chamber working as a physical buffer, either a gel or a membrane dialysis. To achieve this, the experiments are performed in capillaries of small diameter (≤0.1 mm) or alternatively, after gelling or increasing the viscosity of the protein solution (Figure 1.5(b)).

1.6.5 Batch Method

Batch crystallization technique is based on liquid-liquid diffusion phenomenon. In this method, the protein sample is mixed with the precipitant and appropriate buffer creating a homogeneous crystallization medium requiring no equilibration with a reservoir. A batch experiment can be readily performed in a capillary, small container, or plate with a small reservoir.

1.6.6 MicroBatch Process under Oil

In this technique a small drop of the sample combined with the crystallization reagent is pipetted under a layer of oil. For a true microbatch, the drop is placed under paraffin oil (Figure 1.6(a)), which allows little to no evaporation or concentration in the drop. A modified microbatch can be performed when the drop is placed under a mixture of paraffin oil and silicone oil, or straight silicone oil (Figure 1.6(b)). Such oils allow water vapour to permeate from the drop and allow sample and reagent concentration. Unless the drop is equilibrated with a reservoir, water will continue to leave the drop till only solids remain.

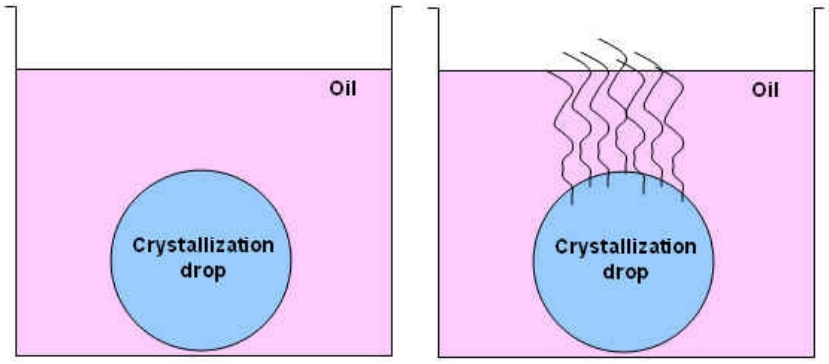


Figure 1.6: Microbatch under oil configuration. Image courtesy of Hampton Research USA

1.6.7 Microdialysis Crystallization

Dialysis crystallization involves placing the sample in a dialysis button which is sealed with a dialysis membrane. Water and some precipitants are then allowed to exchange while retaining the sample in the cell. The dialysis button is placed into a suitable container holding the precipitant or crystallization media (Figure 1.7). A sample requiring a high ionic strength for solubility against a solution of low ionic strength can be dialyzed. Equilibrium is attained faster in the dialysis method. It is controlled by the pore size of the tubing.

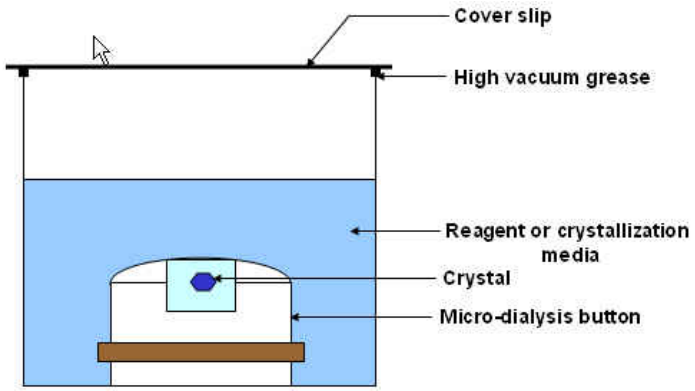


Figure 1.7: Microdialysis crystallization configuration. Image courtesy of Hampton Research USA

1.7 PROTEIN CRYSTALLIZATION PROCESS

Crystallization of biological macromolecules is done using a solution of the material in some solvent, usually from a super-saturated solution of the substance in water. The crystallization of proteins from their aqueous solutions is entirely analogous to that which governs the crystallization of small molecules such as sugar or salt. In each case

the solution is allowed to become supersaturated with solute (protein), and at some point, crystals of the solute appear and grow in size. The term supersaturation can be visualized as follows. In a cup of water, sugar can be dissolved in it up to a point at which no more sugar will dissolve, no matter how much stirring is given to the mixture. A solution of sugar in water, in contact with solid sugar is prepared. This system is said to be in equilibrium and neither crystallization nor dissolution takes place. This solution is said to be saturated, and at a given temperature the concentration of sugar in water is fixed. Normally, and as certainly as the case for sugar, a saturated solution at say 50^oC contains much more sugar in solution than does the same volume of solution saturated at 20^oC. If a solution of sugar saturated at 50^oC is cooled, usually at first nothing happens. There may be no crystallization even though the solution is cooled to 40 or 30^oC, even though there is now too much sugar in the solution for it to be stable. Such a solution is called supersaturated. At some point, crystals are able to form themselves upon some microscopic nucleus, and crystallization can recur very rapidly.

Proteins differ from small molecules only in the degree of supersaturation which is required to induce crystallization or to allow a useful rate of crystallization. The supersaturation of a saturated solution of a protein can be increased in several ways: 1) cool it down, 2) allow the water to evaporate, 3) add an ionic solute, e.g. salt, and 4) vary the pH. Using evaporation to form protein crystals is much like that observed when working with salt or sugar. As the water evaporates the solute (protein) starts to form crystals, when the supersaturation reaches a point that the amount of water available can no longer support the high concentrations of protein.

The commercially available 24 wells VDX plate or Limbro box is used to vary the conditions in each of the wells in order to determine the optimal conditions for growing crystals. The protein concentrations and pH can be varied simultaneously using theseplates Because different pH levels change the structural attractions of the amino acids that make up proteins, changes in pH can also cause a protein to crystallize.

Using an ionic solute such as salt creates another problem. Ionic materials, such as sodium chloride, or common salt, dissolve readily in water. Because of the strong electric field around these ions, Na^+ and Cl^-, a large number of water molecules are loosely bound around the ions in a sphere of hydration. This is because the water molecule itself is an electric dipole, with one end slightly positively charged and the other end negatively charged. The net effect of adding salt to a protein solution is to reduce the amount of water which is free to keep the protein molecules in solution. Thus, the degree of supersaturation of the protein solution is effectively increased, and crystallization becomes more likely.

1.8 IMPORTANT FACTORS IN CRYSTALLIZATION

There are some external factors which should be taken care of while crystallizing the proteins. Some of the important factors are listed below:

1. **Temperature**: The general temperature range is 0-40^{o}C usually 0, 14 or 25^{o}C is suitable temperatures for crystallization. Most proteins are more soluble at 25^{o}C than 4^{o}C at low ionic strength.
2. **Vibration**: Protein crystals are very fragile. Vibration can break crystals and make smaller crystals which increase the number of nuclei. Hence it should be avoided.
3. **Gravity**: Microgravity condition works better for protein crystal growth since it avoids the harmful convection currents.
4. **Pressure**: Crystallization can be done at 1 atm pressure.
5. **Time**: The crystallization time can be hours to months and sometimes over one year or more depending on the type of protein and crystallization conditions. High salt concentration usually takes short times to nucleate whereas it takes longer times to grow. Low ionic strength protein solution is faster in completion, but as a general rule slower is better.
6. **Buffer**: Buffers increase the activity of protein. The concentration of buffer should be kept as low as possible. Buffers, such as sodium acetate is often necessary for the maintaining a particular pH.
7. **Concentration**: The concentration of protein should be as high as possible. Concentration devices should be small so that dead volume of less than 10 ml may be used. It should be better to try fewer conditions than many at lower concentrations.
8. **pH**: pH conditions are very important to grow protein crystals as they can result in different packing orientations.
9. **Additives and salts**: Glycerol tends to make the protein too soluble for easy crystallization. Salt levels should be kept minimum, since these salts may oversolublize the protein.
10. **Purity of macromolecules**: The potential to grow exceptionally good, well ordered protein crystals is there only if very pure protein is available. Protein < 95% pure is not likely to produce very good crystals. The protein solution should contain only those chemicals necessary to maintain protein stability and at the lowest concentrations possible.

1.9 CRYSTAL DEFECTS

Crystalline solids have a very regular atomic structure. The local positions of atoms with respect to each other are repeated at the atomic scale. These arrangments form a crystal lattice, and their study is called crystallography. However, most crystalline materials are not perfect: the regular pattern of atomic arrangement is interrupted by crystal defects. Various types of defects are enumerated here.

1. **Point Defects**: Point defects are defects which are not extended in space in any dimension. The term point defect is used to mean defects which involve at most a few extra or missing atoms without an ordered structure of the defective positions. Larger defects in an ordered structure are usually considered dislocation loops. Vacancies are sites which are usually occupied by an atom but which are unoccupied. A vacancy (or pair of vacancies in an ionic solid) is sometimes called a Schottky defect.
2. **Line Defects**: Dislocations are linear defects around which some of the atoms of the crystal lattice are misaligned. There are two types of dislocations, the EDGE dislocation and the SCREW dislocation. Edge dislocations are caused by the termination of a plane of atoms in the middle of a crystal. The screw dislocation is more difficult to visualise, but basically comprises a structure in which a helical path is traced around the linear defect (dislocation line) by the atomic planes of atoms in the crystal lattice.
3. **Planar Defects**: Grain boundaries occur where the crystallographic direction of the lattice abruptly changes. This usually occurs when two crystals begin growing separately and then meet.

1.9.1 Crystal Twinning

When two separate crystals share some of the same crystal lattice points in a symmetrical manner, it is known as crystal twinning. Due to this effect an intergrowth of two separate crystals occurs and a variety of specific configurations are formed. A twin boundary or composition surface separates the two crystals. The type of twinning is a diagnostic tool in mineral identification. There are mainly three types of twinned crystals: contact twins, penetation twins and polysynthetic twins.

1. **Contact twins**: they share a single composition surface often appearing as mirror images across the boundary, for example, Quartz, gypsum.
2. **Penetration twins**: in this the individual crystals have the appearance of passing through each other in a symmetrical manner, for example- pyrite, fluorite.

3. **Multiple or repeated twins**: in this several twin crystal parts are aligned by the same twin law.
4. **Polysynthetic twins**: Here multiple twins are aligned in parallel, for example, pyrite, calcite etc.
5. **Cyclic twins**: Here, multiple twins are not parallel, for example, aragonite, cerussite.

Of the three common crystal structures, BCC, FCC and HCP, the HCP structure is the most likely to twin.

1.9.2 Crystal Mosaicity

Mosaicity is the angular measure of the degree of long-range order of the unit cells within a crystal. It is the width of the distribution of mis-orientation angles of all the unit cells in a crystal. Imperfect crystals are thought of as a heterogeneous combination (conglomeration) of tiny, perfect crystals which all have slightly different orientations. If the orientation of all these tiny perfect crystals are plotted into a large, (possibly 3D) histogram, then the width of this histogram would be called as mosaicity of the crystal. This number has units of degrees, and reflects the relative, average misorientation of the tiny perfect crystals (mosaic blocks or domains) in the crystal.

High mosaicity has the effect of broadening of spots in the diffraction pattern. That is, the spots would appear smeared tangentially (the direction perpendicular to the line between the spot and the beam center), as though the crystal had been rotated around the beam slightly during the exposure. There would also appear to be more spots than usual, because, for a mosaic crystal, any given rotation angle of the whole crystal, actually represents a range of rotation angles (one for each mosaic block), and each of these blocks diffracts independently of each other. Therefore, spots from highly mosaic crystals will stay in the diffraction condition longer than low-mosaic ones, and each spot will tend to get spread over a wider rotation range, and be more likely to overlap with other spots. Mosaicity can also smear spots radially (along a line drawn from the spot to the beam center). Weak diffraction and spot overlap are mainly due to the presence of mosaicity in a crystal. Lower mosaicity indicates better ordered crystals and hence better diffraction and vice-versa.

1.10 APPLICATIONS OF PROTEIN CRYSTALS

Protein crystal growth experiments aid the generation of computer models of carbohydrates, nucleic acids and proteins, and further advance the progress of biotechnology.

Understanding these results will lead to advances in manufacturing and biological processes, both in medicine and agriculture. Knowledge of protein structure may have potential benefits in many areas of biotechnology. These benefits include new information on basic biological process, development of food crops with higher protein content and increased resistance to disease, and basic research toward the development of more effective drugs. High quality single protein crystals are needed in a variety of crucially important medical applications. The crystallization of proteins currently has three major applications:

1. structural biology and drug design,
2. bioseparation, and
3. controlled drug delivery.

Structural biology and drug design: In this application, the protein crystals are used with techniques of protein crystallography to ascertain the three-dimensional structure of the molecule. This structure is indispensable for correctly determining the complex biological functions of these macromolecules. The design of drugs is related to it, and involves designing a molecule that can exactly fit into a binding site of a macromolecule and block its function in the disease pathway. Producing better quality crystals will result in more accurate 3D protein structures, which in turn means its biological function can be known more precisely, also resulting in improved drug design.

Many diseases involve proteins either directly or indirectly. These can be in the form of hormonal irregularities, toxins produced by invading organisms' proteins and their need to survive, prosper and replicate. Once the structure of a particular protein is known, it becomes much easier to think about how one might block its activity in the human body. This analysis requires a very high quality protein crystal for accurate mapping.

By understanding the structure of protein in the human body, scientists can then learn how these proteins fit into the overall biology of humans, and how the proteins work in the body.

Bio-separation: Bio-separations refer to the downstream processing of the products of fermentation. Crystallization is one of the commonly employed techniques for separating the protein from the biomass. It has the advantage of being a benign separations process, i.e. it does not cause the protein to unfold and lose its activity. The issues here are better prediction and control of the crystallization process to facilitate improved design of crystallization units.

An unrecognized application of protein crystallization is that these systems represent a convenient model for the crystallization phenomena that occur in a variety of

systems: from water freezing in clouds and oceans, through magma solidification in the earth's interior, to the pulling of 12 and 18 inch semiconductor dies, to casting of metals and alloys.

Controlled drug delivery: Most drugs when administered orally into the body system are reacted by several enzymes presented in the path of the destination. Hence before reaching the target, the level of the drug which is required by the target decreases. Side effects of the drugs are also high, since byproducts due to reaction with enzymes may be harmful, neutral or useful. To achieve a constant desired level of drug in any desease, which can be cured by a smaller dose, one requires a much larger amount of medicine to be prescribed by the physician.

When the drug is a protein administering the drug in the crystalline form shows promise of achieving controlled delivery. Therefore, the challenge is to produce crystals of relatively uniform sizes so that the dosage can be prescribed correctly. This can reduce the chance of both underdosing and overdosing, reduce the number of necessary administrations, provide localized and better use of the active agents, and increase patient compliance.

The slow dissolution rate of protein crystals is utilized to achieve sustained release of medications, such as insulin α- interferon. Steady release rates can be maintained longer if an administered medication dose consists of fewer, larger, equi-sized crystallites. To achieve such size distributions, crystal nucleation must be limited to a short time interval so that all crystals grow at the same decreasing supersaturation.

1.11 OPTICAL TECHNIQUES

Over the past decade, laser-based measurement techniques have become popular, though primarily as a flow visualization tool. Recent studies, however, have emphasized the possibility of quantitative measurements using optical methods. Possible reasons for this development are:

- Commercially available lasers have a high degree of coherence (both spatial and temporal) and are cost-effective.
- Optical images can be recorded conveniently through computers and can be processed as a string of numbers through numerical algorithms.

The coherence of the light source generates stable patterns that truly reflect the flow behavior. Computer programs based on mathematical models now simplify the data reduction steps. Image formation can be related to the patterns formed by solid particles suspended in the fluid, attenuation of radiation, scattering or the dependence

of reflectivity, and refractive index on temperature and/or concentration changes in the fluid medium. A majority of optical methods are field techniques in the sense that an entire cross-section of the physical region can be mapped. All of them require the medium to be transparent, and are thus suitable for measurements in fluids. These methods are based on the principle that a light beam passing through the test field is disturbed owing to the inhomogeneous distribution of the refractive index in the flowing fluid. As a consequence, (1) light is deflected from its original direction and (2) the phase of the disturbed light wave is shifted with respect to that of the undisturbed. Both phenomena can be used to visualize the flow as well as perform quantitative measurements.

Due to the non intrusive nature, optical techniques have advantages over traditional techniques as the flow field is not disturbed because of insertion of the probe into the working field. These techniques are inertia free leading to high speed of response for the study of dynamic phenomenon. Optical techniques employed in fluid-flow measurements can be classified into three major categories:

1. Direct visualization techniques
2. Scattering techniques
3. Refractive index based techniques

Direct visualization techniques: Here, a type of marker such as smoke, particle or dyes follow the fluid motion. Streamlines, streak lines, and path lines are used to describe the flow field. **Scattering techniques:** In scattering based techniques, there is direct interaction of light with matter. The distribution of the electromagnetic wave due to interaction is used as a measure of fluid motion. In inelastic scattering, the frequency of the light is changed with respect to the incident. From wavelength and intensity, valuable information about the properties of the flow field can be obtained. Some scattering based techniques are:

1. Laser Doppler Velocimetry (LDV)
2. Particle Image Velocimetry (PIV)
3. Laser Induced Fluorescence (LIF)

Refractive index based techniques: These techniques are based on refraction of light when it passes through a medium having an inhomogeneous distribution of refractive index. In these techniques refractive index or its spatial derivatives are used to obtain the properties of the field of interest. Three refractive index based optical techniques are:

1. Interferometry, where the image formation is related to changes in the optical path difference between the two superimposed light beams. It responds directly to the refractive index field of the flow system.

2. Schlieren, where deflection of light in a variable refractive-index field is achieved. It handles the variation of first order derivative if the index of refraction is normal to the light beam.
3. Shadowgraph, where reduction in light intensity on beam divergence is employed. It measures the variation of the second order derivative of the index of refraction normal to the light beam.

In the context of protein crystal growth from its aqueous solution, a unique relationship can be established between the refractive index and the local density of the medium under study. If the bulk pressure of the region is practically constant, density in turn correlates with concentration, and the three methods become applicable for concentration and/or temperature field measurement in the fluid medium. The three techniques give the path integral of the information present in the region under study. Hence, they are well-suited to measurements in two-dimensional fields, where there is predominantly no change in the refractive index of the flow field along the light beam. Irrespective of the physical quantities being measured, the refractive index based optical techniques generate path integral of the information present in the field of study in the form of two-dimensional images.

A widely used technique to study the transport phenomena is colour schlieren deflectometry. The schlieren method is the name given to the gradient distribution of inhomogeneous transparent media. Index-of-refraction inhomogeneities in the test section cause the deflected rays to pass through different coloured strips of the filter. Therefore, the image (on the viewing screen) of all the points which deflect the light beam through an angle have the same colour.

Generally, human eyes and ordinary cameras cannot capture phase differences in a light beam. One can only see amplitude, colour contrast and the goal of an optical configuration (schlieren, shadowgraph or interfeometry) is to translate these phase differences into contrast and amplitude. A very good example is jellyfish in oceans, which is transparent. Neither its predators nor prey can see it. But small differences in refractive index between them and the surrounding cause them to cast a shadow and hence become detectable. Other examples are the twinkling of stars, disturbances in atmosphere, turbulence and thermal convection. Schlieren effect can be seen in solids, liquids and gases due to temperature changes, high speed flows and mixing of different materials.

1.12 TRANSPORT PROCESSES

The most common setup to grow protein crystals is by the hanging drop technique. A drop of few microliters of protein solution containing the precipitants is put on a glass

slide which covers the reservoir. As the protein/precipitant mixture in the drop is less concentrated than the reservoir solution, water evaporates from the drop into the reservoir. The precipitant reduces the vapour pressure over each of the air-liquid interfaces in proportional to its concentration. Thus a water vapour gradient is established in the air gap with a higher water vapour concentration present in drop-air interface. Water evaporates from the drop, thereby reducing the vapour pressure. Evaporation continues until chemical equlibrium is achieved. As a result the concentration of both protein and precipitant in the drop slowly increases, and crystals may form. Simultaneously, the evaporated water vapour condenses and settles on top of the air-reservoir interface creating a lower concentratred fluid on higher concentrated fluid atmosphere. Due to this, diffusion of lower concentrated fluid diffuses into the higher concentrated fluid. Hence, mainly evaporation, condensation, diffusion and crystallization processes have been taken place in hanging drop configuration of protein crystallization method.

1.12.1 Evaporation

Evaporation is the process by which molecules in a liquid state (e.g. water) spontaneously become gaseous (e.g. water vapour). It is the opposite of condensation. Generally, evaporation can be seen by the gradual disappearance of a liquid when exposed to a significant volume of gas. Evaporation is a type of vapourization of a liquid that occurs only on the surface.

On average, the molecules do not have enough heat energy to escape from the liquid, or else the liquid would turn into vapour quickly. When the molecules collide, they transfer energy to each other in varying degrees. Sometimes the transfer is so one-sided for a molecule near the surface that it ends up with enough energy to escape.

For molecules of a liquid to evaporate, they must be located near the surface, be moving in the proper direction, and have sufficient kinetic energy to overcome liquid-phase intermolecular forces. Only a small proportion of the molecules meet these criteria and the rate of evaporation is limited. Since the kinetic energy of a molecule is proportional to its temperature, evaporation proceeds more quickly at a higher temperature. As the faster-moving molecules escape, the remaining molecules have lower average kinetic energy, and the temperature of the liquid thus decreases. Evaporation also tends to proceed more quickly with higher flow rates between the gaseous and liquid phase and in liquids with higher vapour pressure. Three key parts to evaporation are heat, humidity and air movement.

If evaporation takes place in a closed vessel, the escaping molecules accumulate as vapour above the liquid. Many of the molecules return to the liquid, with returning molecules becoming more frequent as the density and pressure of the vapour increases.

When the process of escape and return reach equilibrium, the vapour is said to be saturated, and no further change in either vapour pressure and density, or liquid temperature will occur. For a system consisting of vapour and liquid of a pure substance, this equilibrium state is directly related to the vapour pressure of the substance, as given by the Clausius-Clapeyron relation:

$$\ln\left(\frac{P_2}{P_1}\right) = -\frac{\Delta H_{vap}}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{1.3}$$

where P_1, P_2 are the vapour pressures at temperatures T_1, T_2 respectively, ΔH_{vap} is the enthalpy of vapourization, and R is the universal gas constant. Evaporation is an essential part of the water cycle. Solar energy drives evaporation of water from oceans, lakes, moisture in the soil, and other sources of water.

1.12.2 Diffusion

When a lower concentration solution is placed on a higher concentrated solution, diffusion of molecules takes place. The molecules of lower concentration solution move to higher concentrated solution. A net movement of molecules having higher concentration occurs in the opposite direction. After sometime, the molecules of both the solution are equally distributed throughout the container (Figure 1.8).

Diffusion describes the spread of particles through random motion from regions of higher concentration to regions of lower concentration. The time dependence of the statistical distribution in space is given by the diffusion equation. The concept of diffusion is tied to that of mass transfer driven by a concentration gradient, but diffusion can still occur when there is no concentration gradient (but there will be no net flux).

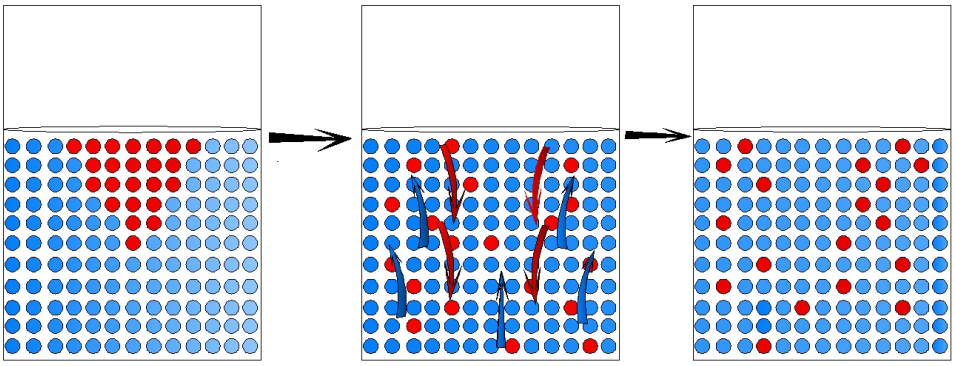


Figure 1.8: Schematic drawing showing the process of liquid liquid diffusion. Red dots: lower concentrated and Blue dots: higher concentrated fluid molecules

In molecular diffusion, the moving entities are small molecules which are self propelled by thermal energy and do not require a concentration gradient to spread out

through random motion. They move at random because they frequently collide. Diffusion is thus thermal motion of all (liquid and gas) molecules at temperatures above absolute zero. Diffusion rate is a function of only temperature, and is not affected by concentration.

Fick's first law relates the diffusive flux to the concentration, by postulating that the flux goes from regions of high concentration to regions of low concentration, with a magnitude that is proportional to the concentration gradient (spatial derivative).

Fick's laws of diffusion describe diffusion and can be used to solve for the diffusion coefficient, D. In one (spatial) dimension, this is

$$J = -D\frac{\partial C}{\partial x} \tag{1.4}$$

where J is the diffusion flux in $(\frac{\text{mol}}{\text{m}^2.\text{s}})$. J measures the amount of substance that will flow through a small area during a small time interval. D is the diffusion coefficient or diffusivity in $(\frac{\text{m}^2}{\text{s}})$. C (for ideal mixtures) is the concentration in $(\frac{\text{mol}}{\text{m}^3})$. x is the position in (m).

In two or more dimensions we must use ∇, the del or gradient operator, which generalises the first derivative, obtaining

$$J = -D\nabla C \tag{1.5}$$

Fick's second law predicts how diffusion causes the concentration to change with time:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{1.6}$$

where t is time. For the case of diffusion in two or more dimensions Fick's law becomes

$$\frac{\partial C}{\partial t} = D\nabla^2 C \tag{1.7}$$

This equation is analogous to the heat equation.

When water is deposited on the air-reservoir interface by condensation, water concentration is higher at this location. Water mixes with the reservoir solution and a net movement of water molecules occurs. A net movement of reservoir solution molecules occurs in the opposite direction. Eventually, water and the reservoir solution molecules are equally and evenly distributed throughout the reservoir cavity.

1.13 OBJECTIVES OF THE PRESENT WORK

In the present study, diffusion process during protein crystal growth in the hanging drop configuration has been considered. The diffusion field and reduction in drop radius have been explored experimentally using M-Z interferometry, monochrome and colour schlieren deflectometry and shadowgraph technique. The grown crystals have been visualized through M3Z stereomicroscope.

Convection has been found to be quality-limiting mechanism in the growth of protein crystals from their solution. Here, convection has the capability of distorting the effects of mechanisms such as surface tension and magnetic fields and thus influence the crystal growth process. Despite the importance of fluid convection, measurement of convection patterns has posed certain difficulties. Over the past decade, laser measurement techniques have become popular, though primarily as a flow visualization tool. Recent papers, however, have emphasized the possibility of quantitative measurements using optical methods. The specific objectives of the present work are stated as follows:

1. Protein crystal growth experiments in small circular wells,
2. Optical visualization of protein crystal growth process,
3. Parametric study of protein crystal growth process using colour schlieren technique,
4. Validation of colour schlieren technique (R-B convection and KDP crystal growth)

1.13.1 Protein Crystal Growth Experiments in Small Circular Wells

To achieve favourable conditions for protein crystallization, initial experiments have been conducted in small circular wells. Various crystallization conditions have been used to examine the conditions suitable for parametric study of protein crystal growth. The grown crystal have been visualized from top using stereo-microscope.

1.13.2 Comparison of optical techniques for Protein Crystal Growth Process

The literature on refractive index based optical methods shows that of the three methods, interferometry has been predominantly applied for qualitative as well as quantitative analysis of the thermal fields in buoyancy-driven convection problem. Schlieren finds potential applications towards qualitative visualization of the flow field and to a limited extent, it has been applied for quantitative studies. Shadowgraph has been extensively used for qualitative imaging. Protein crystal growth process in a rectangular cavity has been visualized by using refractive index based optical techniques namely, interferom-

etry, schlieren and shadowgraphy. Experiments have been conducted under practically identical conditions. The goal of this part is to explore the suitability of these measurement techniques to image, analyze and interpret the diffusion field during crystallization process. Also, these experiments have been conducted for deciding the optical technique to be used for parametric analysis of protein crystallization process.

1.13.3 Parametric Study of Protein Crystal Growth Process using Colour Schlieren Technique

The specific goals here are to study (a) reduction in drop size, (b) crystallization of protein crystal, (c) parametric study of the protein crystal growth, (d) to study water diffusion in reservoir solution after condensation, and (e) X-ray diffraction study of the grown crystals. Schlieren images have been analyzed to correlate the strength of diffusive currents with the concentration field and its gradients in the reservoir solution. The crystal quality has been gauged by examining the transparency of the crystal at the end of the experiment.

1.13.4 Validation of Colour Schlieren Technique

In the present work, the measurement of temperature distribution in buoyancy-driven convection in a rectangular cavity is visualized and analyzed using colour schlieren technique. The colour schlieren technique has been compared with other optical based techniques (interferometry, schlieren and shadowgraph) against this benchmark experimental configuration. The top and bottom walls of the cavity are maintained at uniform temperatures at all times in an unstably stratified configuration. Air is considered as the fluid medium. Temperature differences of 10 and 15 K for air have been employed in the experiments. For these cases the flow was seen to be steady.

The colour schlieren technique is employed to visualize the convection field around a growing KDP crystal from its aqueous solution in the free convection regime. Experiments have been conducted under practically identical conditions as in the literature. The process parameters studied are cooling rate of the solution and rotation rate of the growing crystal. The crystal size plays an influential role in determining the relative importance of buoyancy and rotation. An independent study of crystal size has also been presented. Crystal rotation is viewed in this study as a means of diminishing the impact of buoyancy. The goal of this part of the study is to explore the suitability of this measurement technique to image, analyze, interpret the convective field around a growing crystal and gain expertise in the colour schlieren technique.

1.14 THESIS ORGANIZATION

Subsequent chapters of the present thesis have been structured in the following manner. Chapter-2 is a brief review of published literature on optical techniques with a focus on colour schlieren deflectometry technique and its applications, literature on the studies of hen egg white lysozyme (HEWL), protein crystallization in hanging drop vapour diffusion method, other proteins, imaging of convection around KDP crystal growth and imaging of thermally bouyant fluid flow using optical techniques. The experimental setup and associated instrumentation with various optical techniques are discussed in Chapter-3 and Chapter-4 respectively. The procedure for data reduction along with uncertainty analysis involved in the technique are presented in Chapter-5. Chapter-6 deals with the optimum conditions obtained from VDX plate experimental results. In Chapter-7 results have been presented for the experiments conducted with protein crystal growth using optical techniques (interferometry, schlieren and shadowgraphy). Parametric study is discussed in Chapter-8. In this chapter characterization of protein crystal growth and experiments with colour schlieren technique are reported in which drop radius, reservoir height and drop concentration number of drops and visualization of three drops from different angles are studied. Major conclusions of the present study and scope for future work are summarized in Chapter 9. Appendices contain the details of melting ice experiments, benchmarking experiment of KDP crystal growth, filter generation (1D and 2D) and numerical simulation.