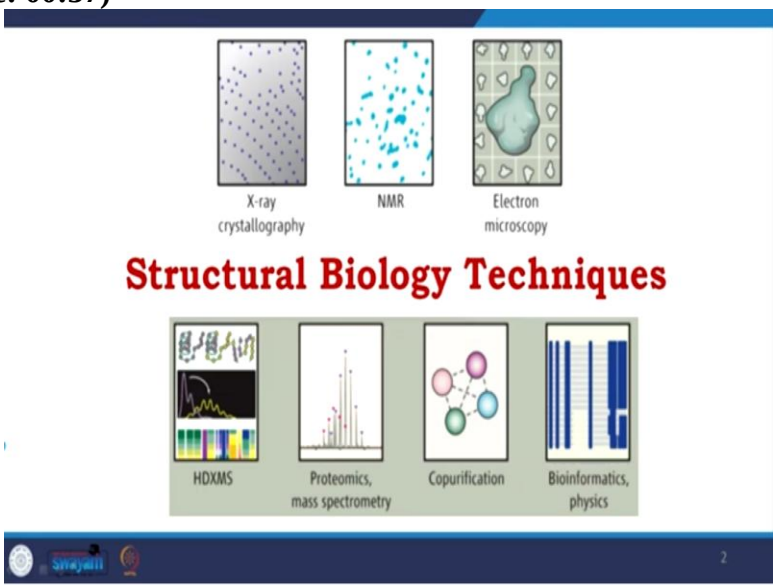
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology – Roorkee

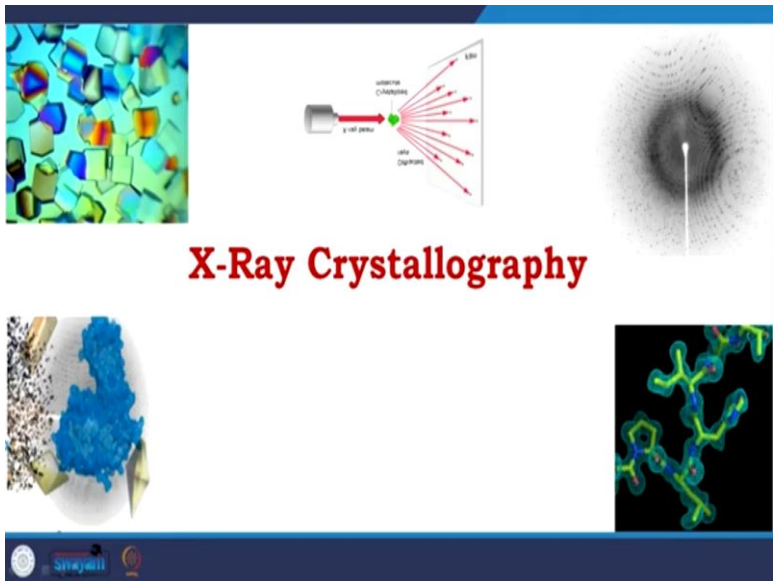
Lecture – 12
Introduction to Structural Biology Techniques, Part – II

Hi everyone, welcome to the next class of structural biology. And today we are going to talk about the next part of structural biology techniques where we have ended in the last class.

(Refer Slide Time: 00:37)

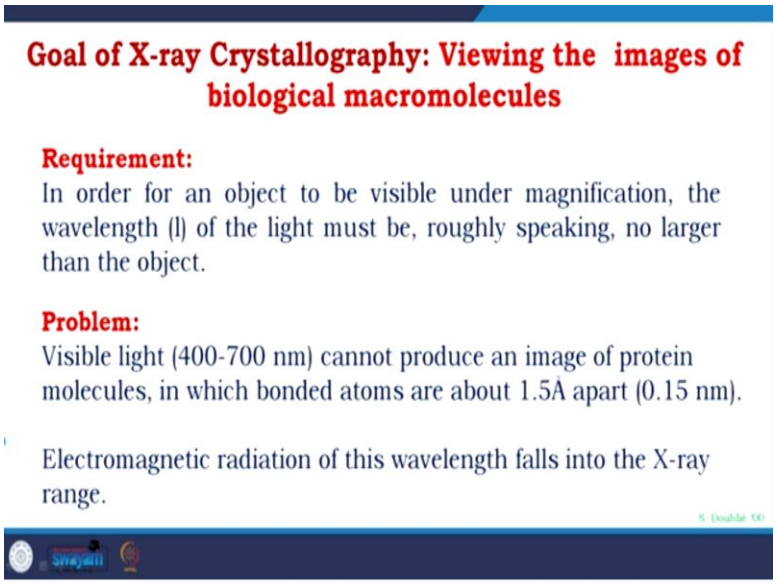


(Refer Slide Time: 00:51)



Today, we are continuing where we ended the discussing about the technique X-ray crystallography.

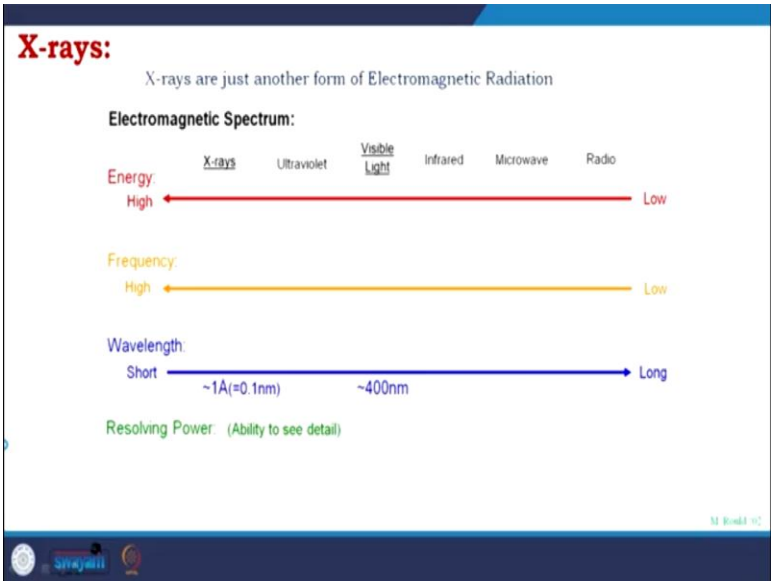
(Refer Slide Time: 00:59)



The goal of X-ray crystallography is viewing the images of biological macromolecules. So, what is the requirement in order for an object to be visible under magnification, the wavelength (λ) of the light must be, roughly speaking not larger than the object. So, what is the problem? The problem is visible light which is 400 to 700 nanometer cannot produce an image of protein molecule in which bonded atoms are about 1.5 angstrom apart.

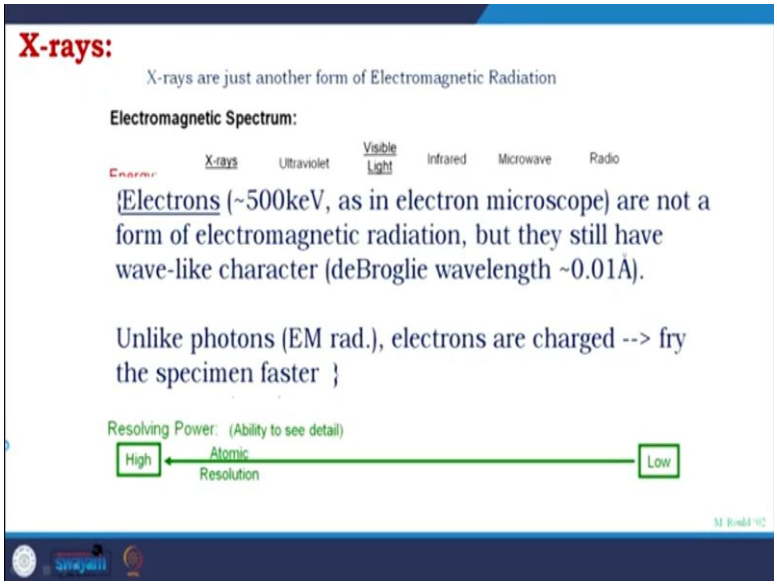
If you think a protein molecule is made up carbon, nitrogen, oxygen and bonded with sulphur and phosphorus, so, most of the bonds are ranging from 1.5 angstrom around and it is impossible for light microscopy to go there and help you visualize that. Electromagnetic radiation of this wavelength falls into the X-ray range.

(Refer Slide Time: 02:03)



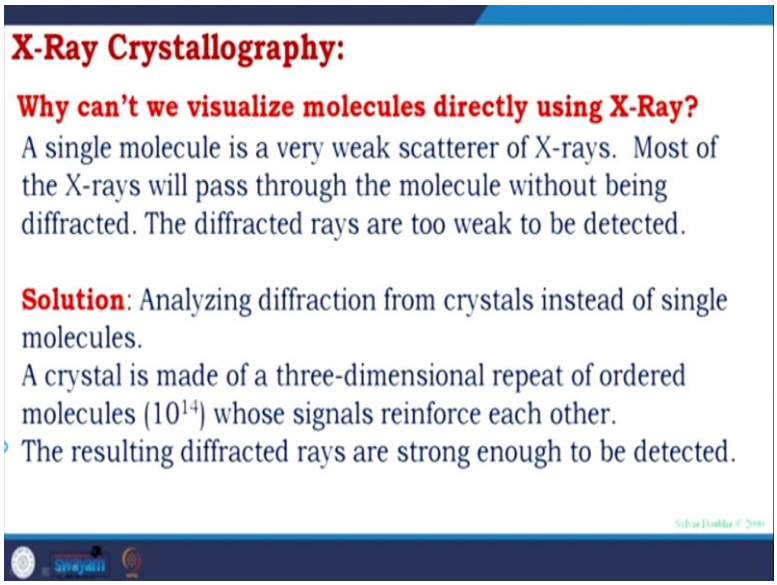
So, X-ray, it is another form of electromagnetic radiation. And if you look at the electromagnetic spectrum, from radio frequency, to microwave, to infrared, to visible light, to ultraviolet to X-ray, you will see that the energy from radio to X-ray it is going from low to high. The frequency in correlating to energy also going from low to high from radio to X-ray, the wavelength is inversely proportional. So, it is going from smaller short wavelength to long wavelength. And if you look at the visible light, it is 400 nm and for X-ray it is around 1 to 3 angstroms. So, electrons as an electron microscope are not form of electromagnetic radiation but they still have wave like character which is needed for this. And in de Broglie if you see that wavelength is 0.1 angstrom, unlike photons, electrons have charged, so, they fry the specimen faster. So, two things you are going to look here, the energy for X-rays high and X-ray could damage the protein. So, these are two negative points.

(Refer Slide Time: 03:34)



But now, coming to the resolving power, it is increasingly enhancing coming to radio frequency to microwave to infrared to visible to ultraviolet to X-ray and in X-ray as we talked about we could see in the atomic resolution.

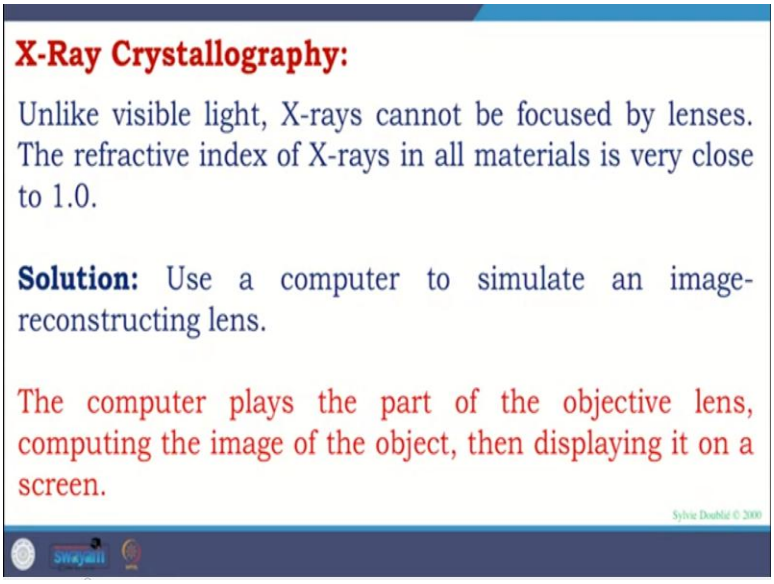
(Refer Slide Time: 03:52)



So, why cannot we visualize molecule directly using X-ray? There are people who always want to go for straightforward. So, they want to ask, we understand that lower wavelength is required, we understand that the radio frequency is required, it should be looked at the atomic resolution so, why not have an X-ray microscope? That is because a single molecule is very weak scatter of X-ray. If you look at the technique, you will see that the basis of X-ray crystallography technique is, the X-ray is coming, it is hitting carbon, nitrogen, oxygen in the crystal and it is diffracting.

The diffraction or the scattering is giving us information. When you consider direct viewing, you are looking at one single molecule. A single molecule is a very weak scatter of X-ray and most of the X-ray will pass through the molecule without being diffracted. So, the diffracted rays would be too weak to be detected. Is there any solution, analyzing diffractions from crystal instead of single molecule, that is the reason we are going from understanding a single molecule of protein to a crystal, we need a crystal, because a crystal have a lot of molecules. A crystal is made up three dimensional repeat of ordered molecules 10^{14} approximately, whose signal reinforce each other, the resulting diffracted rays are strong enough to be detected.

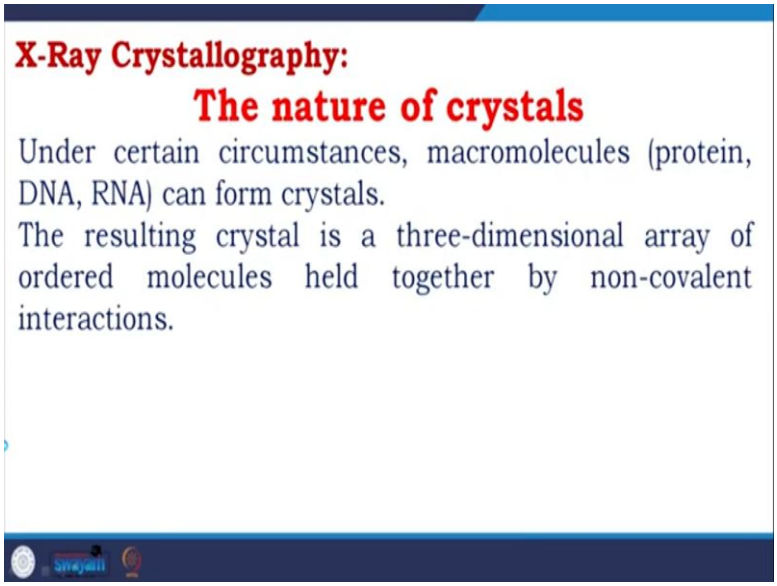
(Refer Slide Time: 06:07)



Unlike visible light, X-ray cannot be focused by lens; the refractive index of X-ray in all materials is very close to 1.0. We use computer to simulate an image reconstructing lens. So, instead of putting a lens in case of light microscope or electron microscope, we are using computer and the computer simulation through Fourier transformation is used as a lens here.

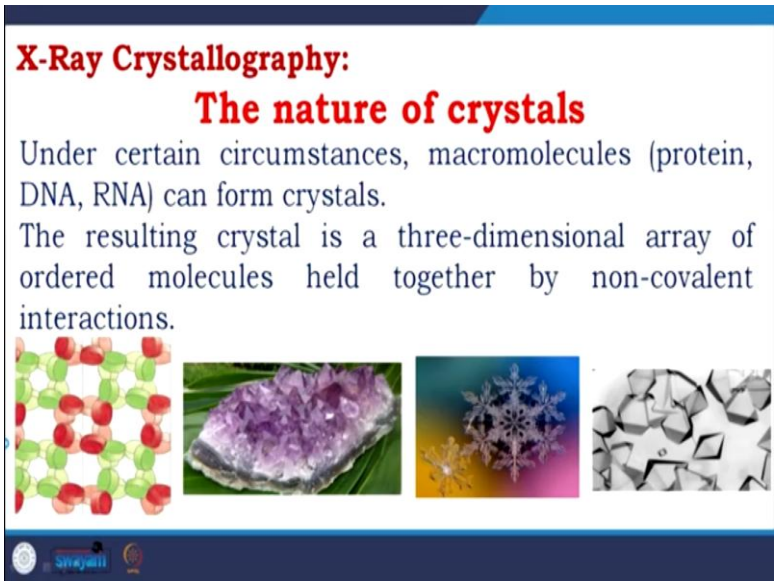
The computer plays the part of the objective lens, computing the image of the object, then displaying it on the screen. So, you put your data, and computer simulates it, and by simulating they are processing it like the lens is processing the light or the electromagnetic wave in light microscope or in electron microscope.

(Refer Slide Time: 07:05)



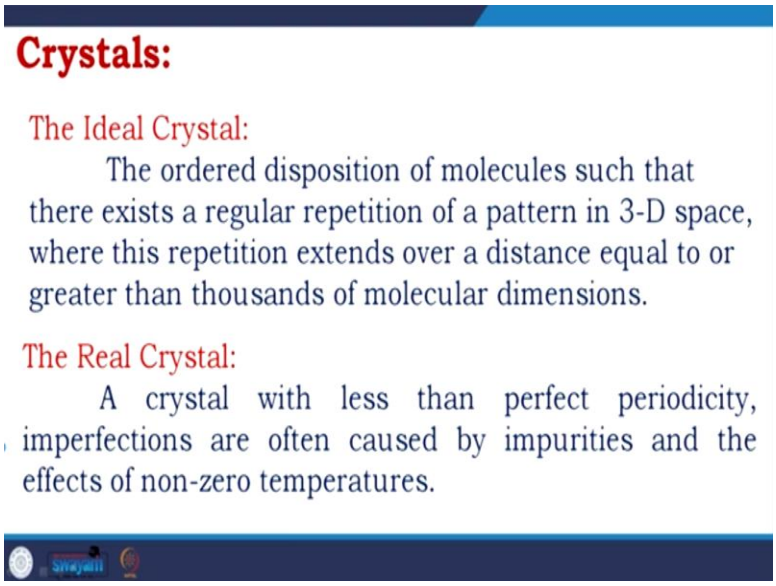
What are the natures of the crystals? Under certain circumstances, biological macromolecules like protein, DNA, RNA can form crystals. Forming crystal is the main factor to get that technology to be forwarded. In many cases, as statistics says only 8% cases when you are doing everything correct, only 8% cases you are finding lucky to get a crystal. The resulting crystal is a three dimensional array, held together by non covalent interactions.

(Refer Slide Time: 08:00)



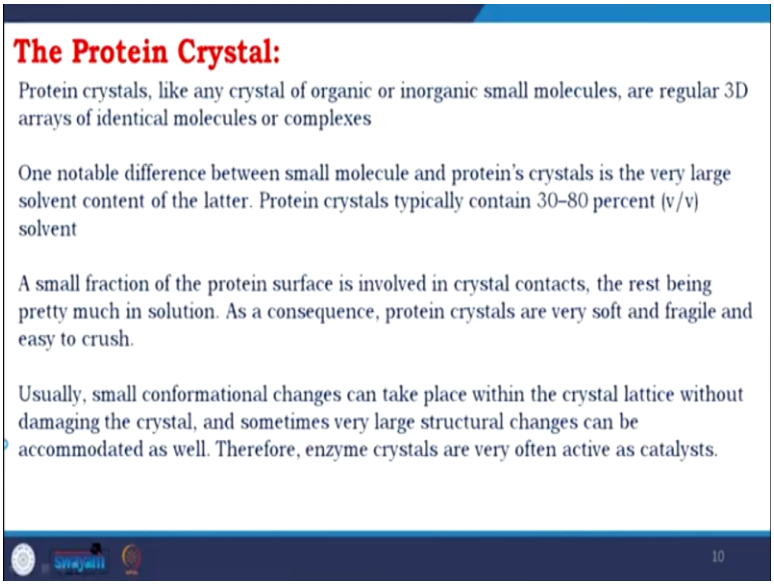
You could look this as a natural crystal, you could see that crystal of small molecule which are actually salts, this is an ice crystal this could be found American or European countries where temperature go below freezing temperature. And this is an example of protein crystal which will bring many times through the courses.

(Refer Slide Time: 08:29)



First, let us define an ideal crystal. An ideal crystal is the ordered deposition of molecules such that there exists a regular repetition of a pattern in 3-D space, where this repetition extend over a distance equal to or greater than 1000 of molecular dimension. A complete structure, the holistic picture, defines crystal and that is an ideal crystal. But the real crystal is not always ideal. The real crystal is less than perfect. The periodicity is not always good. Imperfections are often caused by impurities and the effect of temperature, higher temperature make perturbance disturbance and the ordered nature disturbed.

(Refer Slide Time: 09:35)

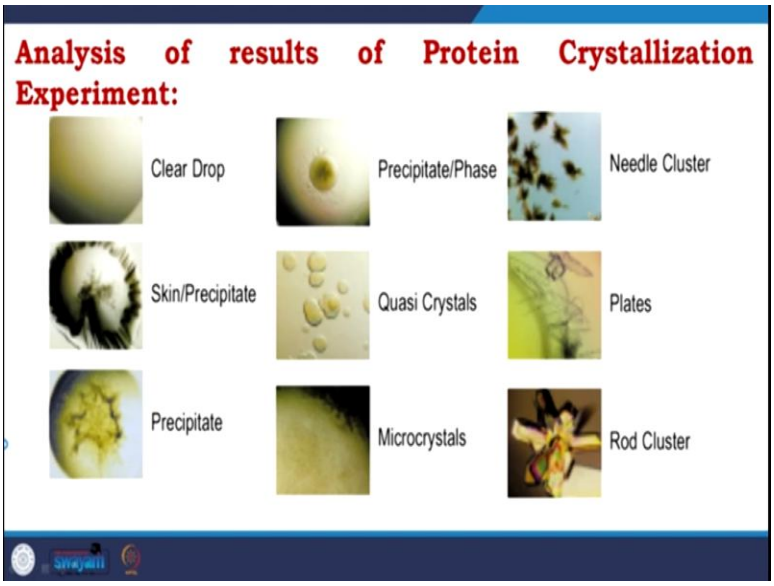


Now protein crystal: Like any other crystal of organic, inorganic small molecules, are regular 3-D arrays of identical molecules or complexes. One notable difference between small molecule and protein crystal is that very large solvent content of the latter. Protein crystals typically content 30 to 80% volume by volume solvent. You could imagine a crystal of sodium chloride, crystal of magnesium sulfate, and all they are very composed. But when you are thinking about protein crystal, they consist of water channels. And as you know, the protein exists and maintains their folding in aqueous solution. A small fraction of the protein surface is involved in crystal contacts, the rest being pretty much in solution. As a consequence, protein crystals are very soft, they are fragile, and they are easy to crush. Usually small conformational changes can take place within the crystal lattice without damaging the crystal. So, when you have a crystal of protein, if you add small molecule, if you add other protein and that goes into and form the crystal, there are small conformational changes observed. There might be very large structural changes which have seen in many cases by adding up small molecule or protein molecule. Therefore, enzyme crystals are very often active as catalysts. Many people think when the protein goes inside the crystal, it is a frozen condition because the outcome of the protein crystal is a static structure.

So, when you have a crystal of beta lactamase, if you add a molecule of beta lactamase inhibitor, you find the beta lactamase react with that compound and covalently linked there, that proves that enzymes are active even when they are in crystal. Anisotropic physical properties like Biorefringence due to an isotropic refraction indices. What is birefringence, if you put the crystal

inside the microscope and put polarized light, you see that color would be produced that is called biorefringence.

(Refer Slide Time: 14:06)

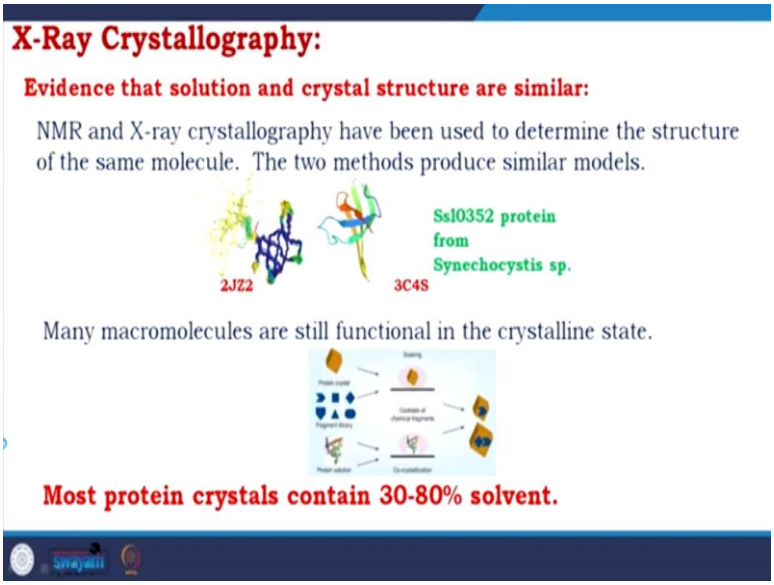


These are some conditions, some results of protein crystallization experiment, you see clear drop, you see precipitate, you see needle clusters, you see skin like precipitation, quasi crystals plates, micro crystals, Rod cluster and what not.

(Refer Slide Time: 14:53)

Whenever we look under the microscope and see a single crystal, we say Eureka we got it.

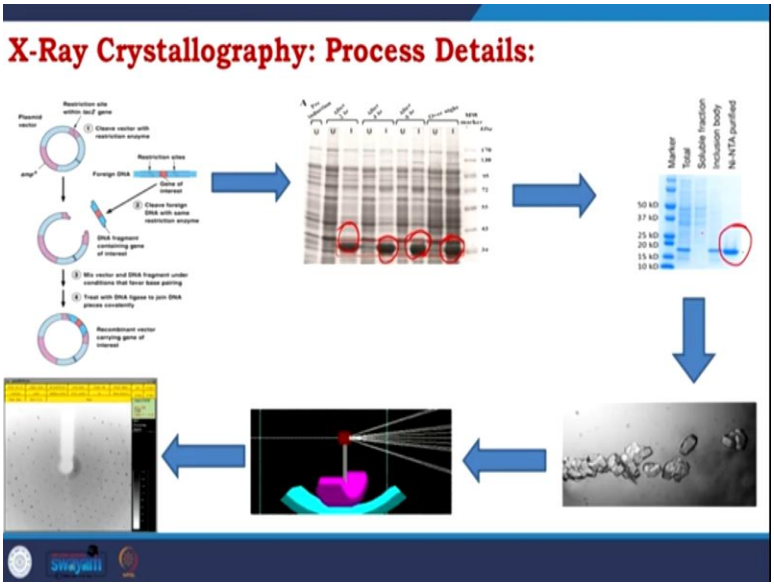
(Refer Slide Time: 15:00)



Evidence that solution and crystal structures are similar when people have done structure solution using NMR and X-ray. The two techniques X-ray as I already told make it in frozen condition, whereas NMR deals protein in solution. When people have picked two structures, the same molecule solved by two techniques, they have found that these structures are similar.

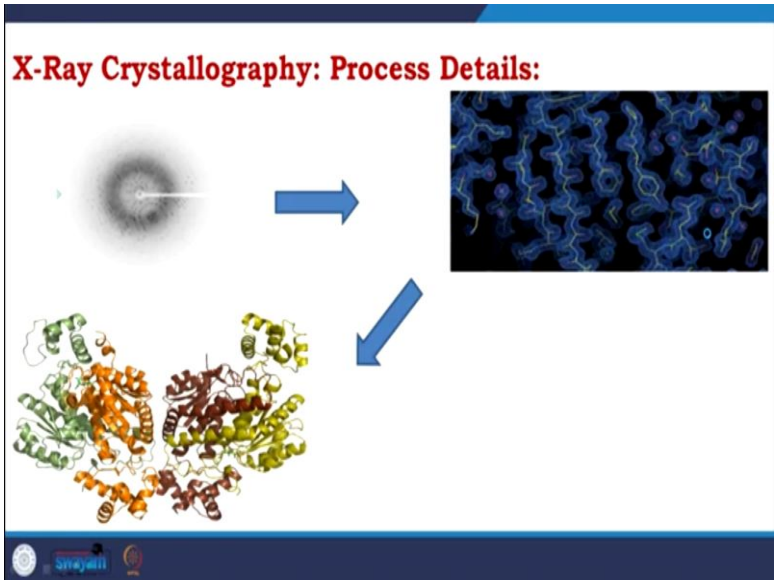
Here is an example, this one is a 2JZ2 is a structure solved by NMR technique and 3C4S is a structure which is solved by protein crystallography techniques, they are the same protein Ss10352 protein from Synechocystis species. When you compare, you will see similar accord in it is in both the structures. Many macromolecules are still functional in the crystalline state, this is what I was talking about, you do assay you will find that proteins are still behaving enzymatically active inside the crystal that means the protein is still functional. So, there are two types of experiments co-crystallization and soaking. When you do soaking if your protein is an enzyme which could act with the substrate and if there is any color production or some assay you will find that the protein is still active inside the crystal. More importantly, most protein crystals as I told earlier content 30 to 80% solvent. So, now we will go inside the process details of protein crystallography.

(Refer Slide Time: 17:07)



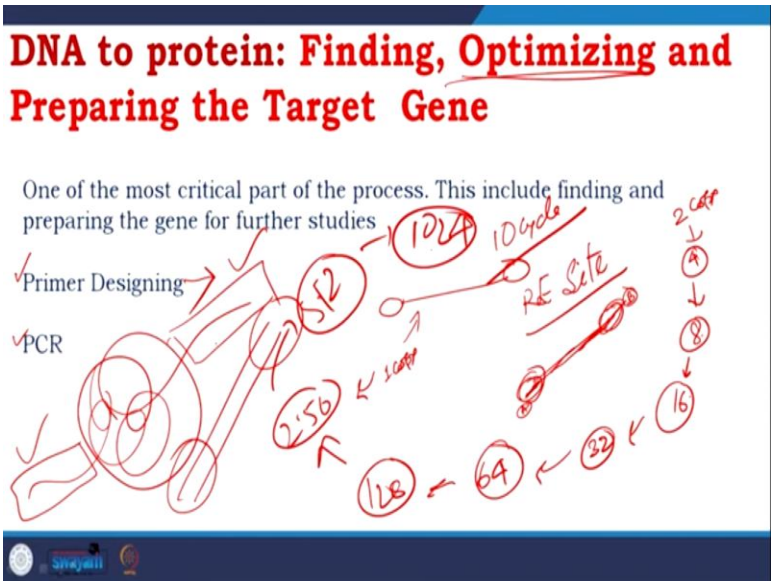
Initial phase is to start with DNA, we have to take a DNA and we have to go through the entire process of cloning, overexpression, and purification. You get a purified protein then you go for crystallization. Then if you get a single crystal, you will go for diffraction. And then when you get the diffraction data.

(Refer Slide Time: 17:58)



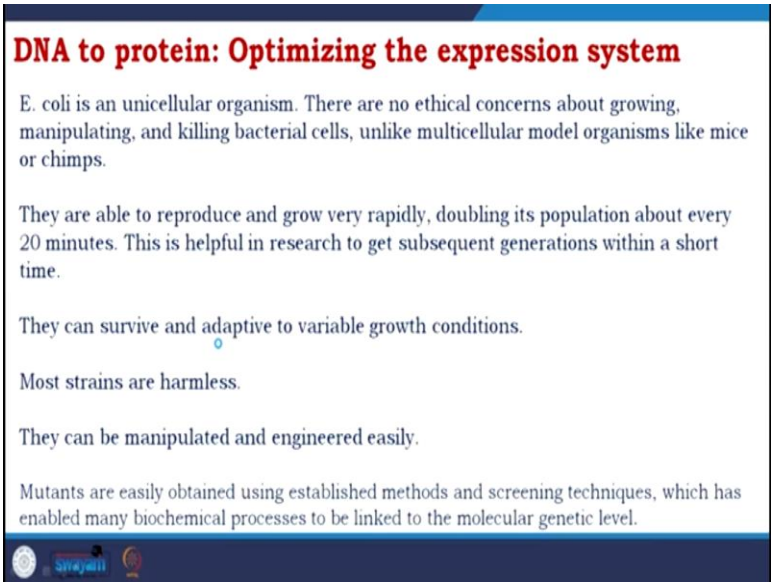
You will process the diffraction data, you will get electron density and then from the electron density, you will build up the model and the model is your final goal towards getting the structure solved by the technique of X-ray crystallography. So I will talk about the details because you know, a lot of people thinks the crystallography means starting from a protein to go for crystal behind getting the protein there is a lot of struggle there.

(Refer Slide Time: 18:50)



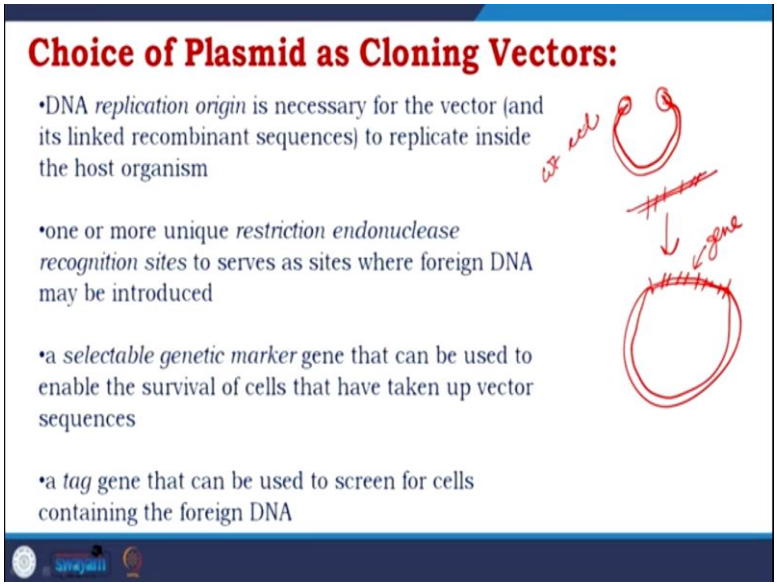
Finding, optimizing and preparing the target gene: Finding means you have to do background work, the first step is to choose or identify the gene. Then you have to go for primer designing, after designing the primer, you will do PCR. By PCR, single copy of your gene would be going to amplify. The cloning procedure also need some optimization, like if you see some of the portion of your protein are not well structured, you could identify that part, and those parts you could take out from your gene. By doing that, you make the structure more compact and that will ensure tha,t later when you are going to overexpress the protein you get more product, when you are going to crystallize the protein, you will get better chance of getting a crystal.

(Refer Slide Time: 24:00)



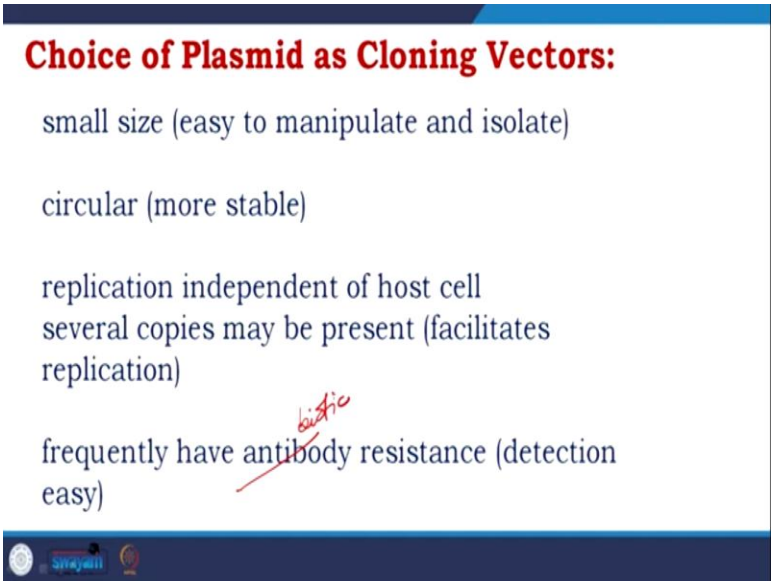
The next thing is to get an optimized expression system. When it is bacteria, the hand pulled choice is *E coli*. Because E-coli is an unicellular organism, pretty much well studied, there are no ethical concern about growing, manipulating, killing bacterial cell, unlike when you do it with mouse or chimpanzee or higher organisms. They are able to reproduce and grow very rapidly doubling time about every 20 minutes. They can survive and adaptive to variable growth conditions. They are well known, most of their strains are known, they are not harmful, and they can be manipulated and engineered easily. Mutants are easily obtained using established method and screening techniques, which has enabled many biochemical processes to be linked to the molecular genetic level.

(Refer Slide Time: 26:05)



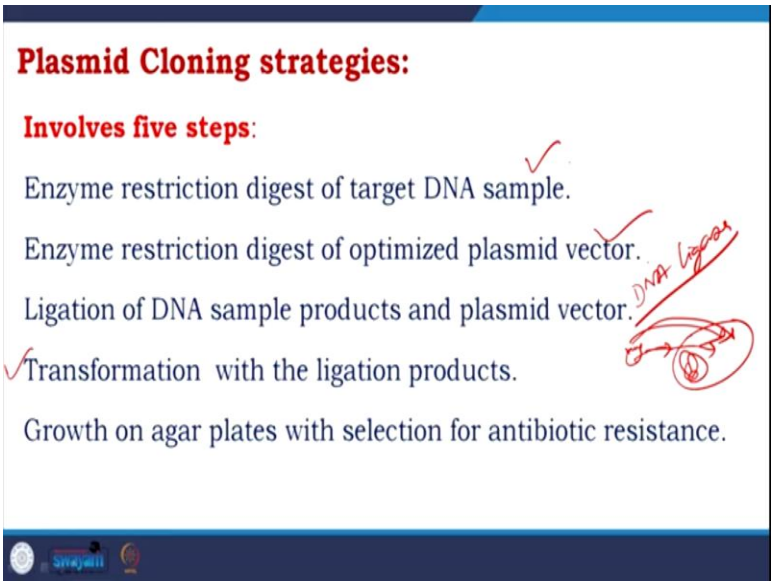
Then choice of plasmid as cloning vector, we could see if there is DNA replication origin, it is necessary for the vector and its linked recombinant sequence to replicate inside the host organism. One or more unique restriction endonuclease recognition site, a selectable genetic marker that can be used to enable the survival of cells that have taken up vector sequences. So, when you are going that E-coli you have to differentiate those E-coli from other E-coli. If you have a gene like beta lactamase that would save you from a beta lactam like penicillin. So, that is the selection. So, if you have a wild strain E-coli and you grow them in ampicillin, penicillin, the vector containing E-coli will survive whereas, the wild strain E-coli would not survive. A tag gene that can be used to screen cell containing the foreign DNA must be present.

(Refer Slide Time: 28:26)



Choice of plasmid as cloning vector: Small size, easy to manipulate and isolate, circular (more stable), generally easy to handle, replication independent of host cells, several copies may be present (facilitated replication), and frequently have antibiotic resistance (detection easy).

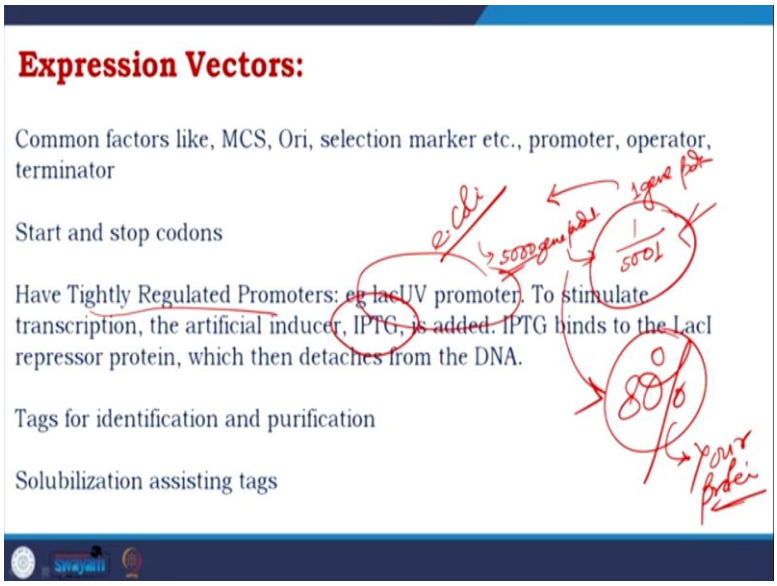
(Refer Slide Time: 28:53)



Cloning generally involves 5 steps. Enzyme restriction digest of target DNA sample, enzyme restrictions digest of optimize vector. Ligation of DNA sample products and plasmid vector using DNA ligase. Transformation of the ligation production. Now, you have the DNA content vector, you have to send it inside the cell of E-coli, Growth on agar plates with selection of antibiotic resistance. So, this is what I talked about, the growth on agar plate with the selection for antibiotic resistance, which will help in ensuring that the vector is there as well the gene is

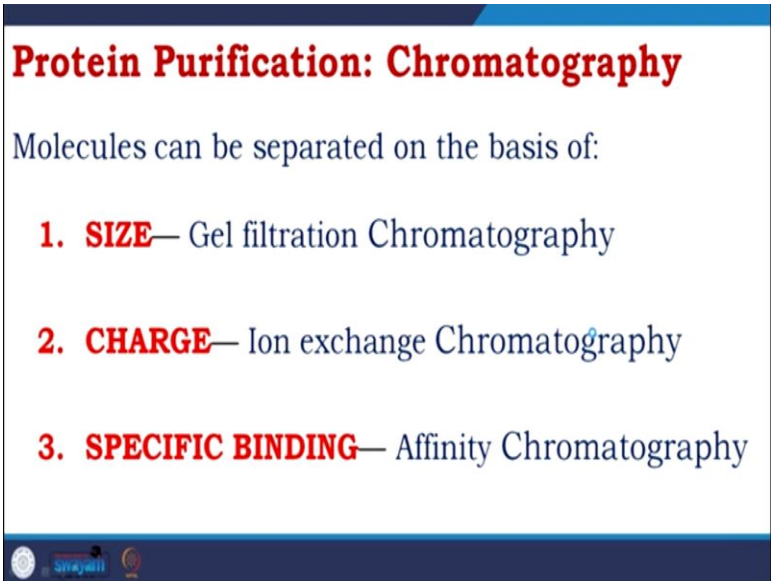
there. So we generally use two markers to ensure the presence of the vector, with one marker to ensure the present gene.

(Refer Slide Time: 30:11)



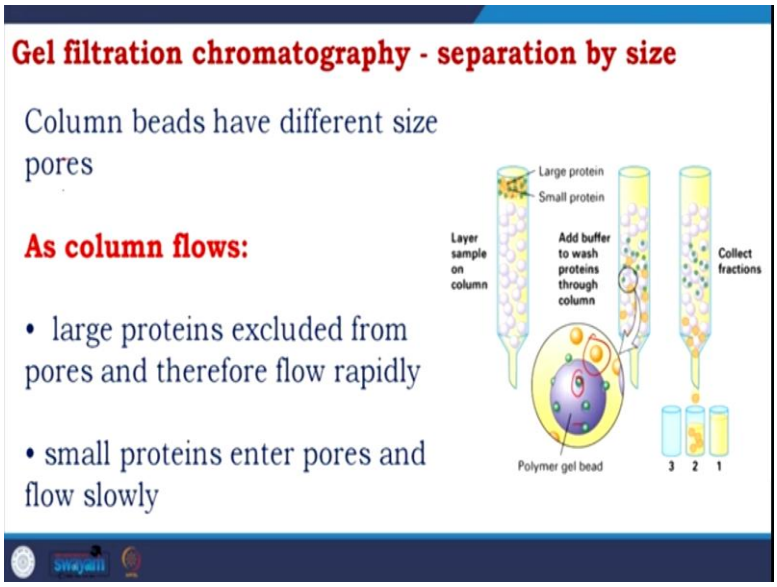
Expression vectors: Common factor like multiple cloning site, Origin of replication, gene selection marker, promoter, operator, terminator, start and stop codons. In addition for expression vector, they should have tightly regulated promoter like lacUV promoter to stimulate transcription, the artificial inducer like IPTG is there, which is a stable molecule that mimic allolactose, sit on the lac operon, it constitutively push the operon to produce the protein it is induced with. Tags for identification and purification, you need to have some engineered tag like his tag, GST tag. These engineering techniques require to further going through chromatography to purify your protein, because for crystallography, for NMR, for Cryo, you need very pure protein. Then solubilization assisting tags, there are some tags like sapiron, that helps your protein to solubilize. Solubilization is very important for optimum production.

(Refer Slide Time: 33:06)



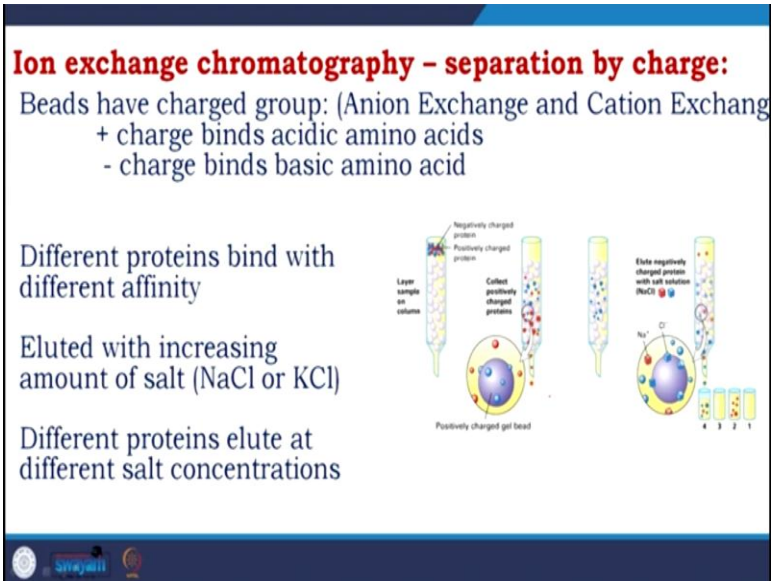
Now, going to the protein purification, there are chromatography techniques. For size based separation there is gel filtration chromatography or size exclusion chromatography, for charge based separation ion exchange chromatography, for specific binding related separation there is affinity chromatography.

(Refer Slide Time: 33:28)



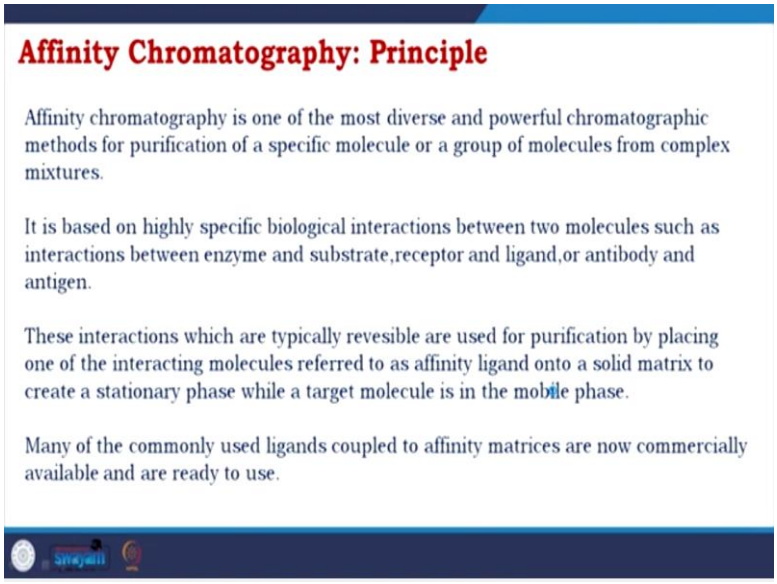
Gel filtration chromatography: In every chromatography column have several beads with different size and the size have the key here. The column beads have different size pores and as the column flows, you put the protein in the column it flows there, large protein excluded from pores and therefore flow rapidly and small protein enter pores that flow slowly.

(Refer Slide Time: 36:07)



Coming to ion exchange chromatography-separation by charge: if you see the charged beads they are positively charged or negatively charged. So, beads have charged group Anion exchange and Cation exchange. Positive charges bind acidic amino acids and negative charge bind basic amino acids. Different proteins bind with different affinity depending on their contents (positive or negative charge amino acids). Eluted with increasing amount of salt once the protein binds there, different proteins elute at different salt concentration and that is how they are differentiated.

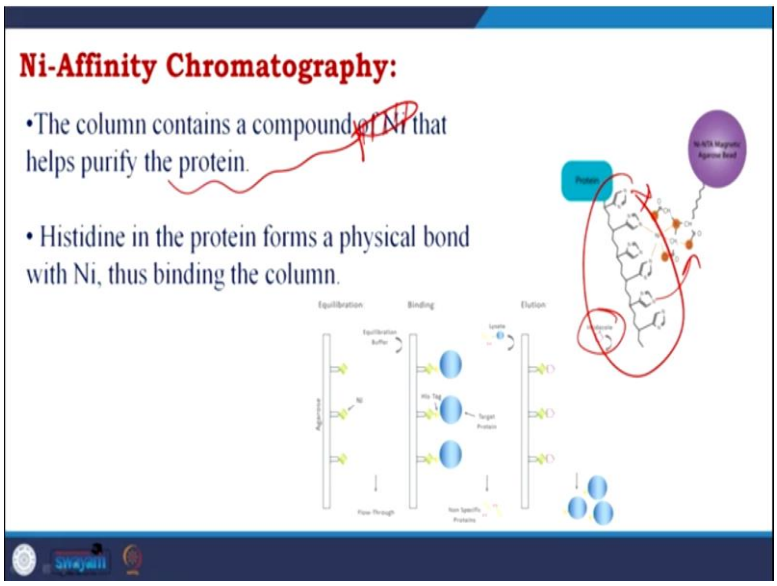
(Refer Slide Time: 36:55)



Affinity chromatography: Affinity chromatography is one of the most diverse and powerful chromatographic method for purification of a specific molecule or a group of molecule from complex mixture.

Today you find something which could be attaching to a specific protein then that would be used as an affinity chromatography column to purify that protein. That is the power of affinity chromatography. It is based on highly specific biological interaction between two molecules such as interaction between enzyme and substrate, receptor, ligand, antibody, antigen. The interactions which are typically reversible are used for purification by placing one of the interacting molecules referred to as affinity ligand into a solid matrix to create a stationary phase while a target molecule is the mobile phase, your protein is in the solution that is the mobile phase and the column is having agarose with the bead that is the stationary phase. Many of the commonly used ligand coupled to affinity matrices are now commercially available and are ready to use.

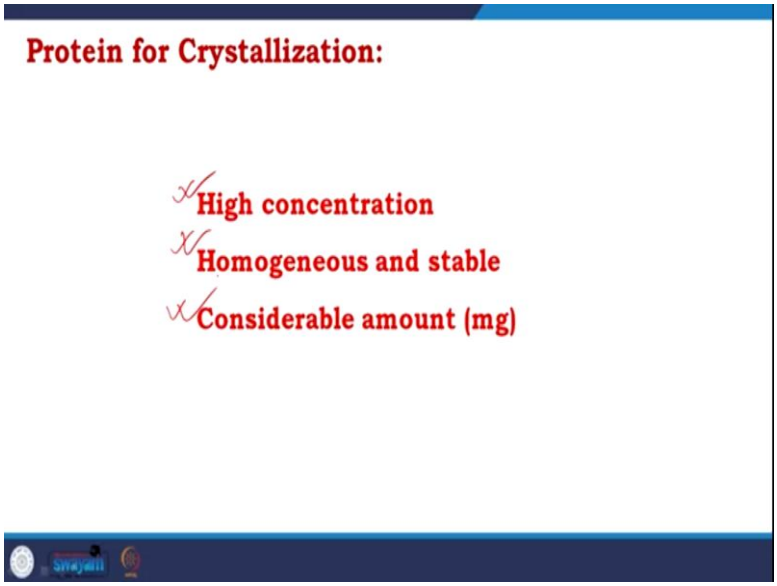
(Refer Slide Time: 38:24)



I will talk about one which is nickel affinity chromatography. As I told in the time of the genetic engineering, we include the sequence of 6 Histidine amino acids, so that these 6 Histidine amino acids will come at the end of the protein. Now, when we pass it through nickel affinity chromatography column, the Histidines would be attaching with the nickel so the Histidine bound protein would be specifically bound to this column, other protein would be flow through.

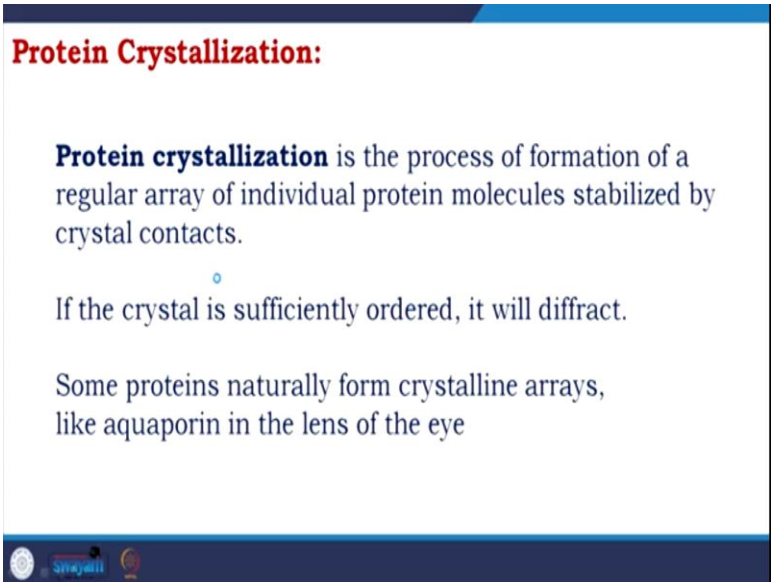
The column contents a compound of nickel that helps to purify the protein. Histidine in the protein forms a physical bond with nickel, thus binding the column. There are three stages. First, you have to equilibrate the column, and then you have to put the protein which will be binding, then you will use imidazole (mimicking the same chemistry like Histidine) to elute. Finally, the protein is ready for crystallization

(Refer Slide Time: 39:53)



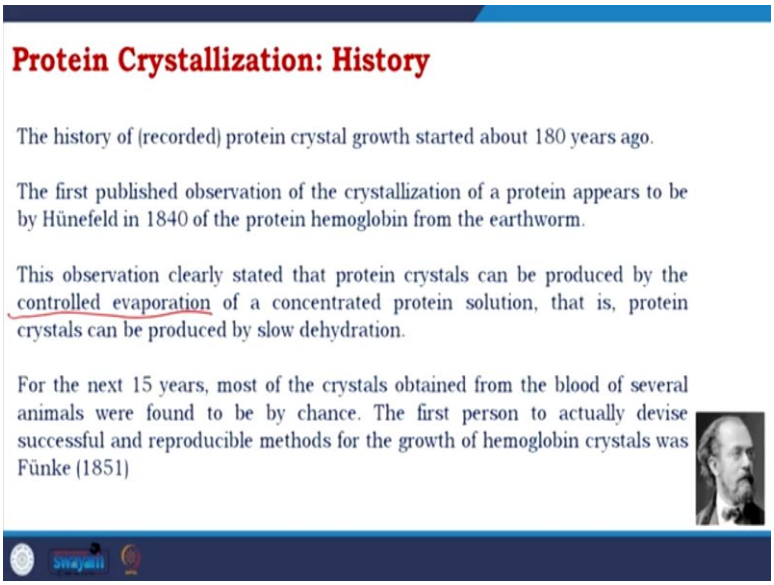
Three things is very important in protein crystallization. High concentration of protein is needed for crystallography, homogeneous and stable (protein have to be homogeneous, you cannot have impurity, if you have impurity or if you have second protein present your protein have a much less chance of crystallization), and you have a considerable amount because you have to make a screening, at least 500 to 1000 screen condition so you need good amount of protein.

(Refer Slide Time: 40:38)



Protein crystallization is the process of formation of the regular array of individual protein molecule stabilized by crystal contacts. If the crystal is sufficiently ordered, it will diffract. Some proteins naturally form crystalline arrays like aquaporin (present in the lens part of the eye). It actually crystallized automatically, there is protein like myoglobin, lysozyme which are very crystallizable.

(Refer Slide Time: 41:15)



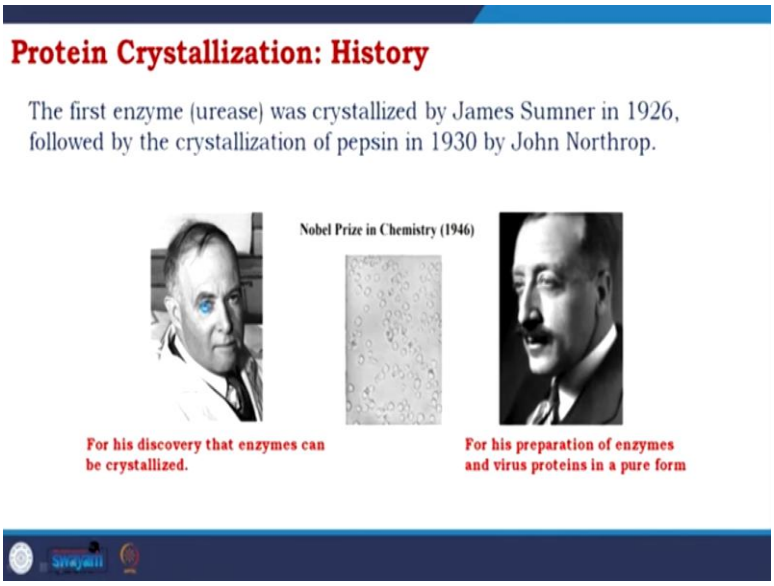
Anywhere we are going we should know about the history. The history of protein crystal growth started about 180 years ago, think about 180 years, so far is recorded. That might be there are other attempts which are at all, not at all successful, so not recorded. The first published

observation of the crystallization of the protein appears to be by Hunefeld in 1840 of the protein hemoglobin from the earthworm.

This observation clearly stated that protein crystals can be produced by the controlled evaporation of a concentrated protein solution. Remind the term controlled evaporation because that is what we are going to follow when we are going in depth and understanding the principle behind protein crystallization. That is protein crystal can be produced by slow dehydration. For the next 15 years, most of the crystal obtained from the blood of several animals was found to be by chance.

The first person who actually devise successful and reproducible method for growth of hemoglobin crystal was Funke, you could see him Funke and very interestingly, you might have a question in mind, why these guys were actually trying hemoglobin? Because that time people do not have protein purification techniques. What they know that if you get blood, you will get a lot of hemoglobin because of its red color, you could understand that the protein is there that is what the secret behind.

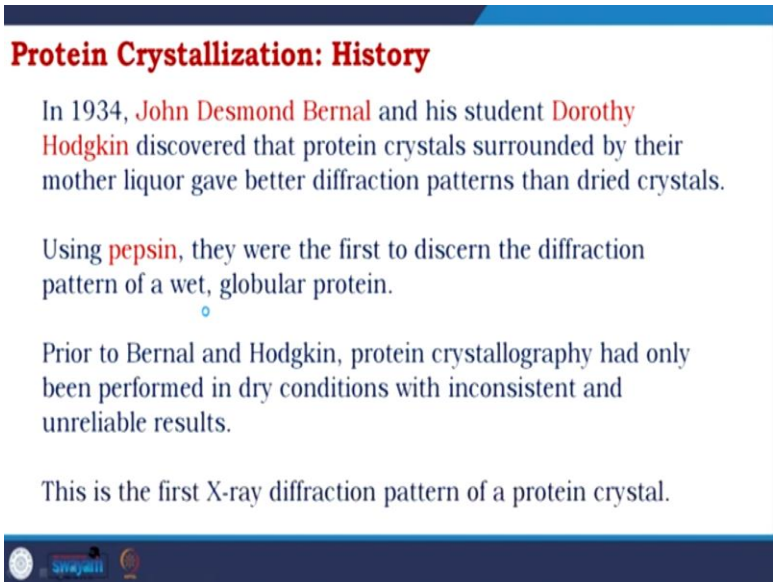
(Refer Slide Time: 43:12)



The first enzyme was crystallized by James Sumner in 1926 which was urease, followed by the crystallization of pepsin in 1930, by John Northrop and these 2 guys got Nobel Prize in Chemistry in 1946. For Sumner, it is for the discovery that enzymes can be crystallized. This is a

very interesting and revolutionary step towards the development of the technique which is protein crystallography. And then Northrop for the preparation of enzymes and more importantly, he has worked on viruses introduced the virus proteins in pure form.

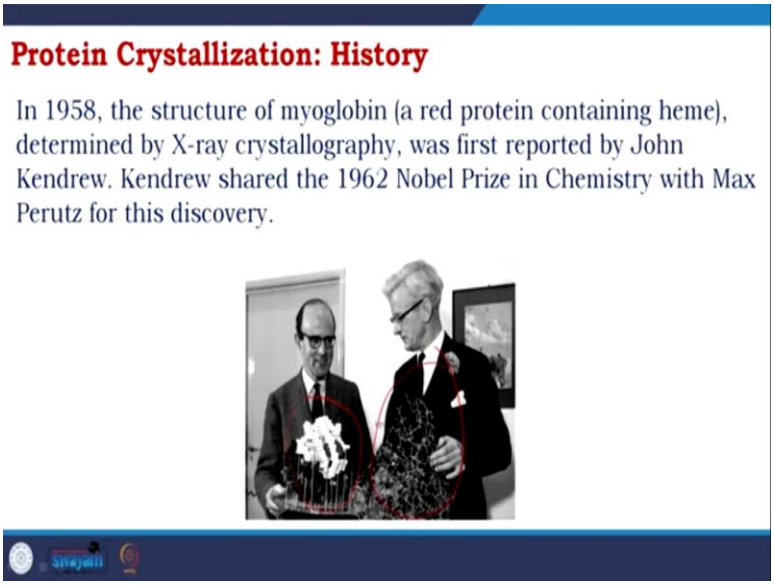
(Refer Slide Time: 43:58)



In 1934 John Desmond Bernal and his student Dorothy Hodgkin discovered the protein crystals surrounded by their mother liquors get better diffraction patterns than dried crystals. So up to now, people did not understand how to continue the process, they found that it is the mother liquor, I am talking about mother liquor further that would be helpful. Using pepsin they were the first to discern the diffraction pattern of a weird globular protein.

So this is the first attempt where they use X-ray to get a diffraction pattern, they could not solve it. Prior to Bernal and Hodgkin protein crystallography had only been performed in dry condition with inconsistent and unreliable results. This is the first X-ray diffraction pattern it is not the solution yet it is the first X-ray diffraction pattern of protein crystal.

(Refer Slide Time: 45:05)



And the start of the era the start of protein crystallography in 1958, the structure of myoglobin, a red protein containing heme again, I talked about hemoglobin, myoglobin is also another protein which having iron so it color red. And that is easy to identify between different proteins, determined by X-ray crystallography was first reported by John Kendrew. Kendrew shared the 1962 Nobel Prize in Chemistry with Max Perutz for this discovery.

And if you see here, this is what Kendrews model and this is what like Max Perutz model. So we will talk about this later but this could be the end of today's class. Thank you for listening. And I would continue from this to the in the next class. Thank you very much.