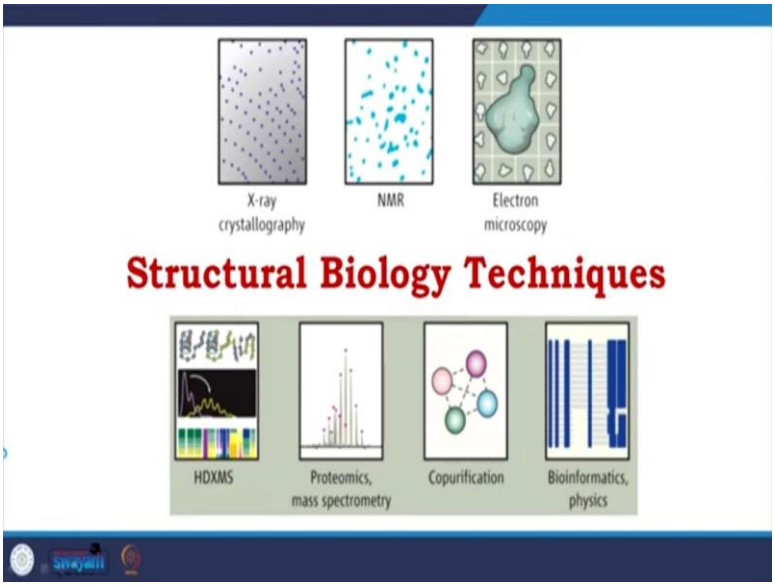
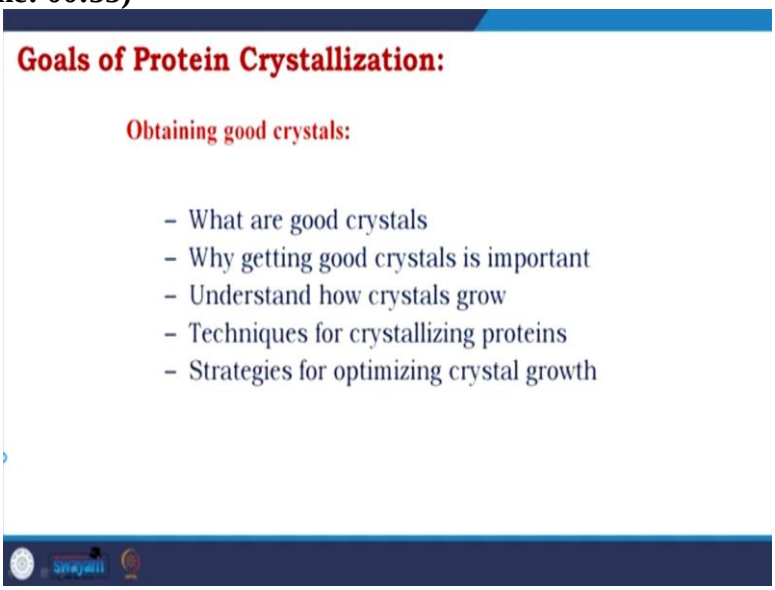
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology, Roorkee

Lecture - 13
X-ray Crystallography – Crystallization, Part 1



Hi everyone, again welcome to the structural biology course. It is the continuing module where we are discussing structural biology techniques, now we are continuing with the discussion of X-ray crystallography. I have talked about the basic the background and like starting of protein crystallization.

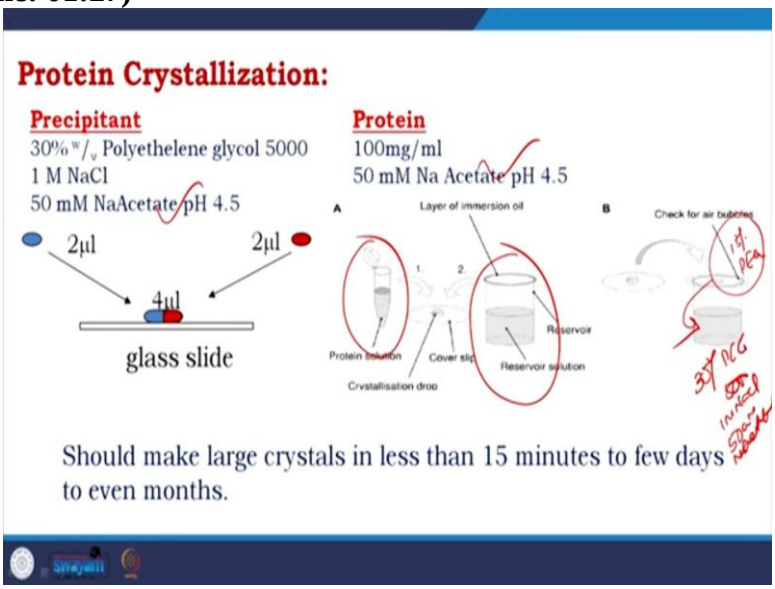
(Refer Slide Time: 00:55)



Now in this class we will start with the goal of protein crystallization, what is actually a goal definitely to obtain good crystals? So we will discuss what are good crystals in terms of

science, is it the crystals which look good or good? Is it the crystals which are actually good? We will look at them. Why getting good crystal is important? Understand how crystals grow? Techniques for crystallizing proteins and strategies for optimizing crystal growth.

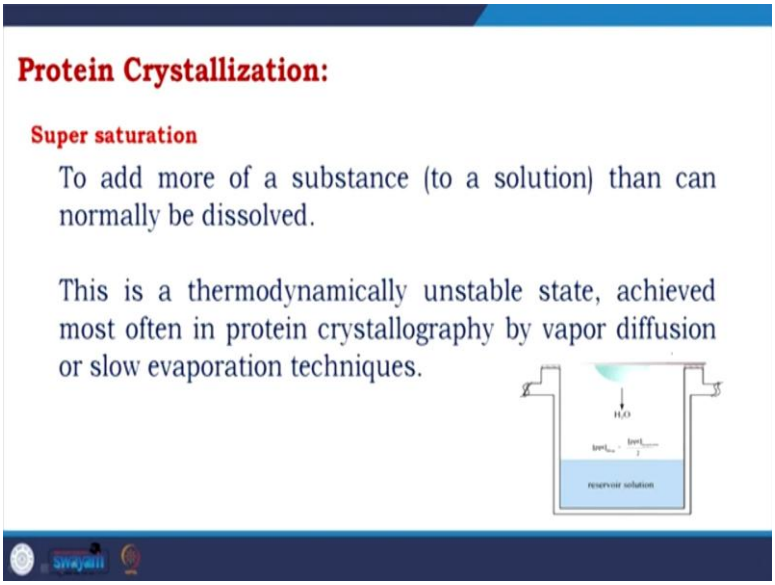
(Refer Slide Time: 01:27)



So protein crystallization is a process which you could call a development of a thermodynamically closed complex closed system. We have the protein solution and we have the reservoir solution, the reservoir is made of precipitant which actually boosts the protein towards precipitation. When it goes through that change of phase from liquid protein solution to solid precipitation, there is a small unique phase point, phase juncture where the protein crystallizes.

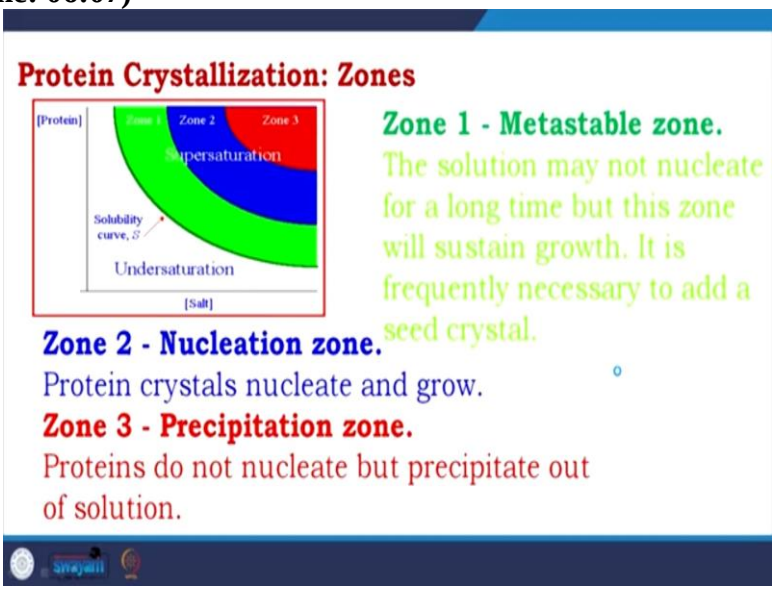
Let us say one precipitant condition 30% wet by volume polyethylene glycol 5000, 1 M NaCl, 50 mM sodium acetate ph 4.5, 100 mg/ml protein, 50 mM sodium acetate ph 4.5. Now we take 2 µl of the protein and 2 µl of the precipitant, and keep it on the glass slide, you could imagine copper slips which used in biology as a glass slide. Then put grease there and then we put back it on the reservoir place, where it produced the closed system. Then we check for air bubbles to ensure that this is a closed system.

(Refer Slide Time: 04:51)



What is super saturation? To add more of a substance (to a solution) than can normally be dissolved, it is called Super saturation. Super saturation is thermodynamically unstable state, achieved most often in protein crystallography by vapor diffusion or slow evaporation techniques. We will look at that, as I told you have the drop, where the protein and precipitants are there and water would be coming from there because the reservoir condition is half here.

(Refer Slide Time: 06:07)

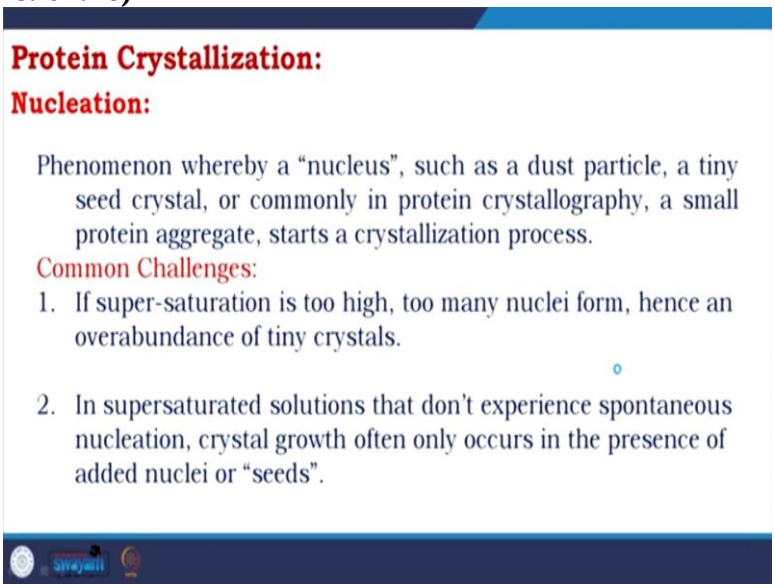


Zones are extremely important for protein crystallization. Zone one is metastable zone, the solution may not nucleate for a long time but this zone will sustain growth, it is frequently necessary to add a seed crystal. So after you set up the crystal drops, you have to keep checking under microscope. If you see clear zone, this is the zone one. Zone two is nucleation

zone, protein crystals nucleate and grow. Zone three is precipitation zone, protein do not nucleate but precipitate out of solution.

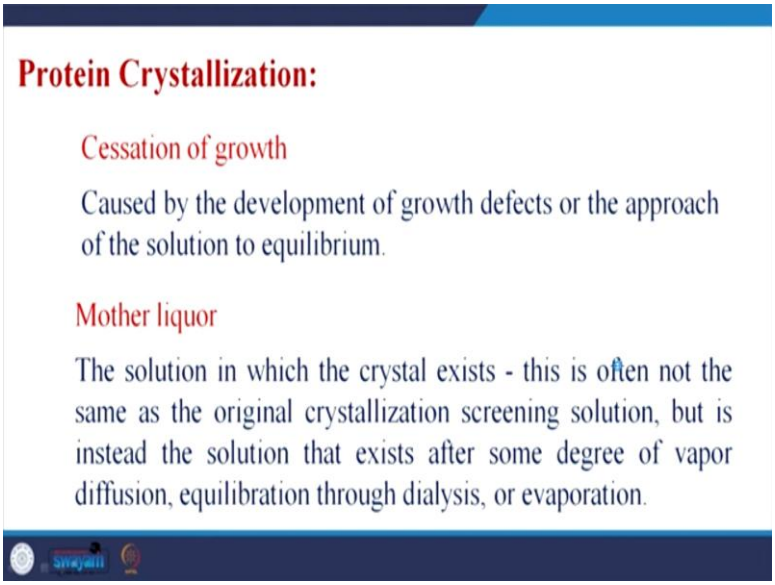
If you get zone one, and that remains there, after some time you have to start thinking different strategies like increasing concentration of the protein, increasing concentration of the precipitant and anything. If it goes to zone three then immediately you know that this condition is not fit for crystallization, you have to again think for reducing everything.

(Refer Slide Time: 07:16)



Nucleation, as I told is a very important term, the phenomenon whereby a nucleus such as a dust particle, a tiny seed crystal or commonly in protein crystallography, a small protein aggregate starts a crystallization process. I talked about, you have a protein solution, you add precipitant, and there is slow dehydration, so the concentration of protein is increasing on the drop. This slow dehydration cause a change of phase from liquid to solid, in between there is one unique phase point where the crystal would grow and you are very lucky if you get that. Common challenges: If the super saturation is too high, too many nuclei form, hence an overabundance of tiny crystals. You get tiny crystals, you get crystals but these crystals are not possible to diffract so you cannot proceed from there. In super saturated solution spontaneous nucleation crystal growth often only occurs in the presence of added nuclei. How to get micro crystal? There are many techniques; one which people commonly use is seeding. So we could do seeding using tiny crystals.

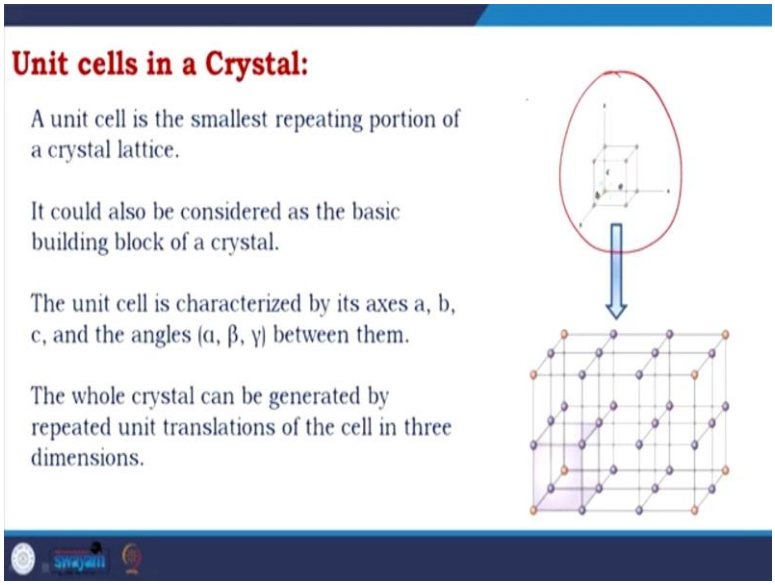
(Refer Slide Time: 08:56)



Protein crystallization: Cessation of growth caused by development of growth defects or the approach of the solution to equilibrium. Mother liquor, the solution in which the crystal exists, this is often not the same as the original crystallization screening solution but is instead the solution that exists after some degree of vapor diffusion equilibration through dialysis or evaporation.

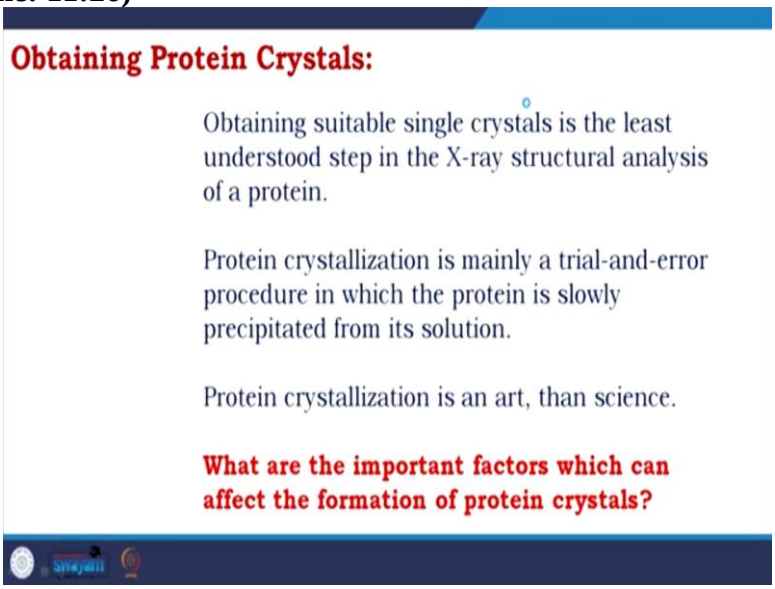
What this is talking about? When you set up a crystal, it is called mother liquor. What you have? If you have some concentration of the precipitant, it would become half, you develop a thermodynamically closed system, and water goes from the drop to the reservoir. So slowly the concentration of the protein as well as the precipitant would keep increasing in the drop. Drop which you think the mother liquor, you calculate the concentration where you initially started, it is changed down, because of the process of dehydration.

(Refer Slide Time: 10:07)



Unit cells in a crystal: Unit cell is very important; a unit cell is the smallest repeating portion of a crystal lattice. It could also be considered as the basic building block of the crystal. As we know million of cells organize the body, similarly million of unit cell build up the crystal. The whole crystal can be generated by repeated unit translation of the cell in three dimensions, unit cell is characterized by its axis a, b, c and angle alpha, beta, and gamma.

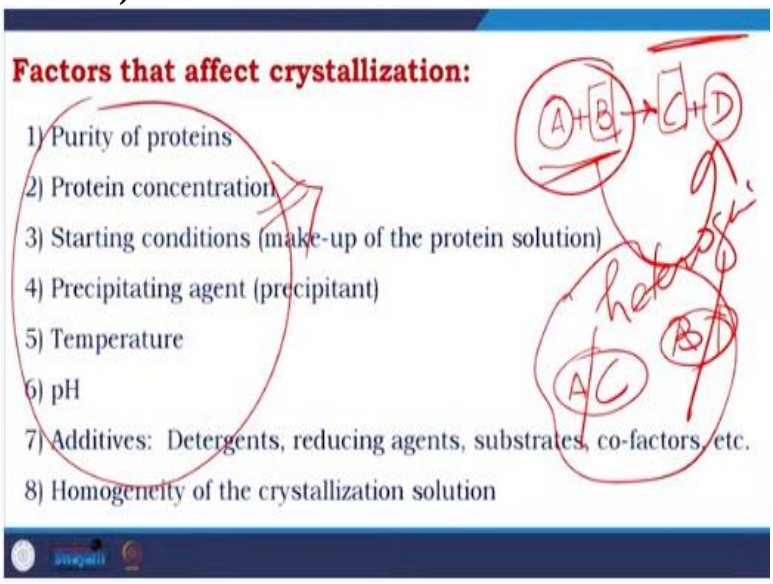
(Refer Slide Time: 11:26)



Obtaining suitable single crystal is the least understood step in x ray structural analysis of a protein. Process of crystallization is going through long time but still now, we did not understand the science behind. Protein crystallization is mainly a trial and error procedure in which the protein is slowly precipitated from its solution. You just do your work best, you do the targets selection, you do the optimization, you do the hosts choice, you do the cloning, you do the transformation, you do the over expression, you do the purification, get the protein

set the crystal then you go home, every day come shake and just hope that one sunny day you will find a crystal. That is the part we crystallographers hate about crystallography. People say crystallization is an art, and crystallography is a science

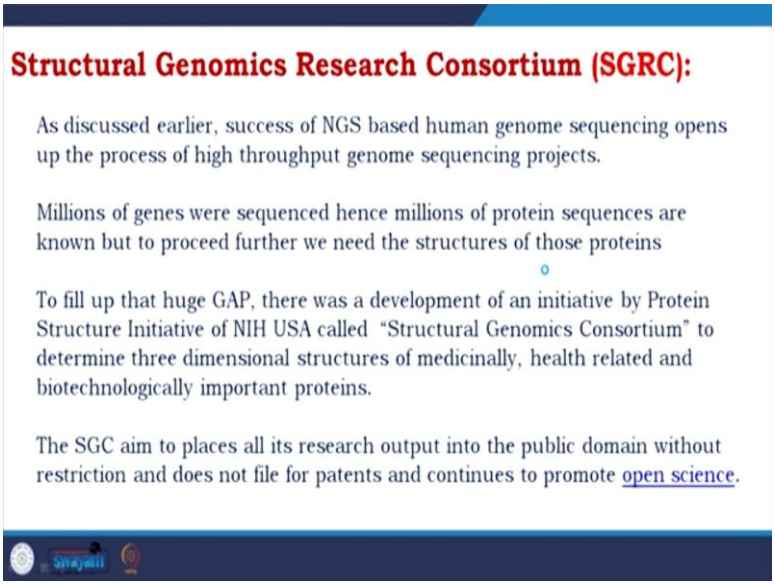
(Refer Slide Time: 13:09)



Factors that affect crystallization: Purity of protein, purity of protein is one of the most important and critical thing to get a crystal. When we used to do kinetics or spectroscopy experiments, we do not care if our protein is 90% or 95% pure. But when it comes to the crystallization, we try our best, we at least go for two round of purification, we do nickel NTA followed by size exclusion. Protein concentration, you have to optimize protein concentration and there is no parameter to optimize. So you have to do trial and error. But if you keep high concentration, there is a good chance.

Starting condition makeup of the protein solution, So, this would be heterogeneous condition and heterogeneous condition never support crystallization. Precipitating agent, like sample screening, nothing is known to be working for everything there is no rule. So you have to do more and more and you have to do optimization. Temperature, temperature plays a lot of role. So this is advisable that when you are setting your crystal for the first time, you set it in at least in 4 degrees, 12 degrees, 20 degrees and 37 degrees, pH you have to mix screening with different pH, it is better that you know where your protein is stable in which pH. Additives, detergents, reducing agents, substrates, cofactors, with all our experience of 180 years we have no idea, who are the important factor that effort crystallization, it is depending on your protein, you have to find it out and homogeneity in the crystallization solution which I have already talked about.

(Refer Slide Time: 17:19)

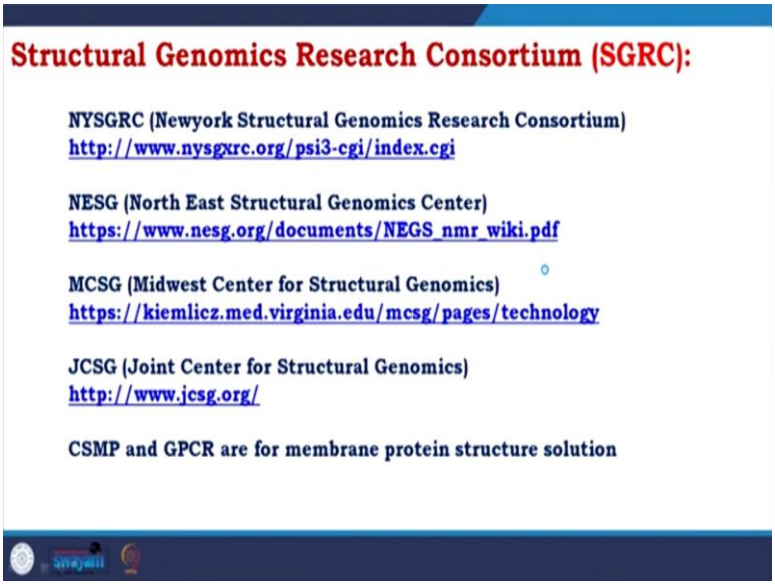


Now I will talk about a very interesting consortium it is called Structural Genomics Research Consortium. I talked about next generation sequencing, I talked about this is the technology which changed the spectrum how biology is looking at, after the success of next generation sequencing where they have done human genome sequencing. And through human genome sequencing they have, like, got the collaboration between good universities like MIT, Harvard, Stanford, Caltech, with big companies like Applied bio system, Illumina they are doing high throughput gene sequencing projects.

The high throughput gene sequencing projects are providing them millions of genes and that means they are millions of protein sequence. But to proceed further means, to go for biomedical problems, to go for disease identification, to go for drug development, they need more structures, more folded proteins. And there is a huge gap I have already discussed. To fill up that huge gap there was a development of an initiative by protein structure initiative of the National Institute of Health, US, called Structural Genomics Consortium.

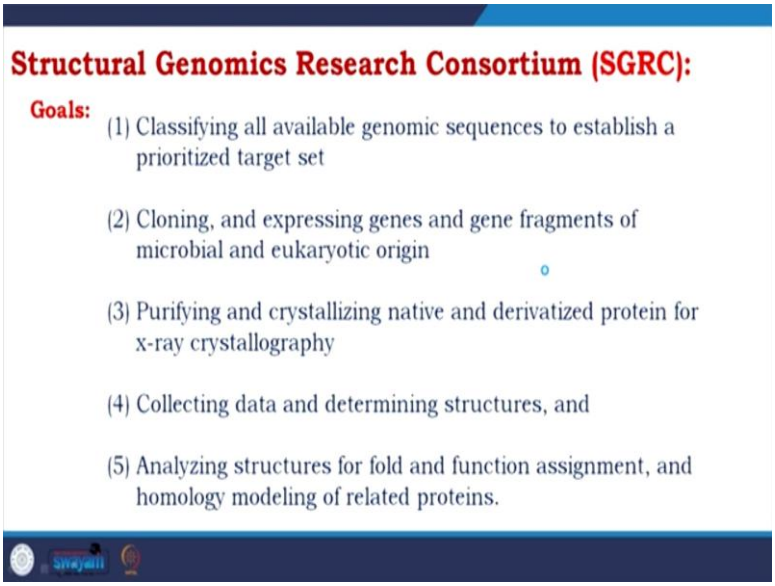
To determine the 3 dimensional structure of medicinally health related and bio technologically important proteins. So by making a consortium they got success that now they could sequence any genome. So they are thinking they would, their initiative is to develop further consortium which would solve high throughput way any structures specially which are important. So this consortium aims to places all its research output in the public domain without restriction and does not file for patents and continues to promote open science.

(Refer Slide Time: 19:29)



These have several groups, New York Structural Genomics Research group, North East Structural Genomics Centre, Midwest Centre for Structural Genomics, Joint Centre for Structural Genomics and CSMP and GPCR are 2 initiatives you know, GPCR that is G Protein Coupled Receptor and CSMP is thus, Consortium for Structure of Metalloproteinase. These are 2 rare consortium where they are trying to develop a high throughput initiative for membrane proteins which are very neglected.

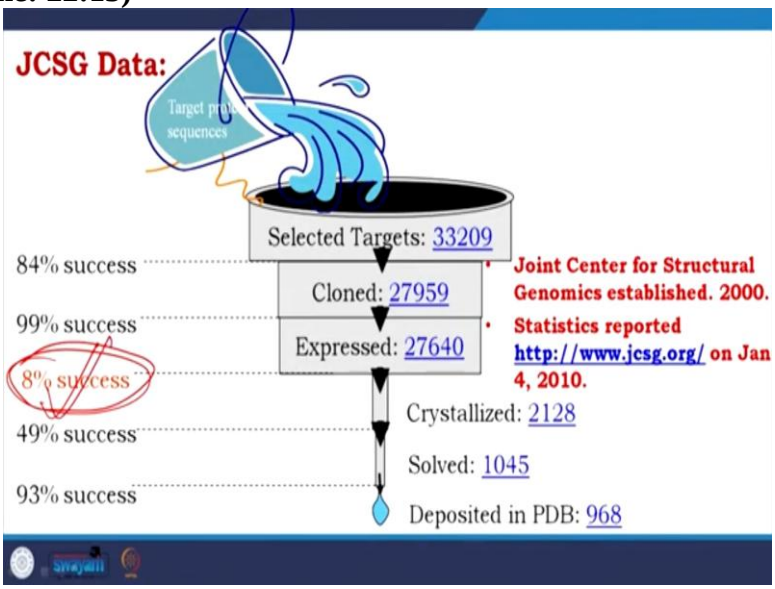
(Refer Slide Time: 20:51)



The goal of Structure Genomic Research Consortium is classifying all available genomic sequences to establish a prioritized target set. So they want to pick up the way I told you select genes, so they want to select a lot of genomes instead of genes, cloning, over expression, and purification, at a high throughput way. Purifying and crystallizing native and

derivatized protein for x ray crystallography. Collecting data and determined structures, analyzing structure for fold and function assignment and homology modeling of related proteins.

(Refer Slide Time: 22:13)



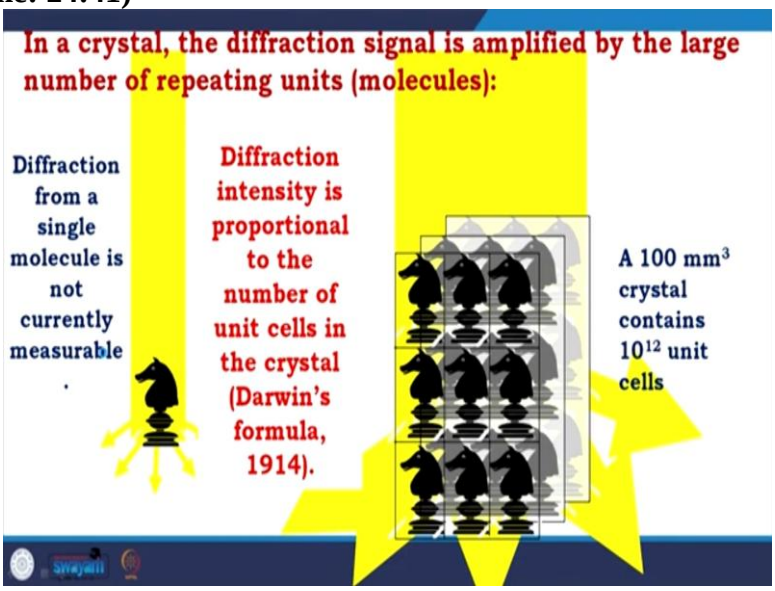
I am presenting the data from Joint Centre for Structural Genomics which was established at 2000 and this is the data after 10 years, 2010. Selected target of that consortium was 33,209. They have cloned 27,959 proteins. They have over expressed 27,640 proteins. They have crystallized 2128 proteins. They have solved the structure of 1045 proteins. And they have deposited in PDB 968. If you look at the percentage, from selected target to cloning, they have a success of 84%, from cloning to over expression, they have 99% success, from over expression to crystallization they have 8% success, from crystallization to solution they have 49% success, and from solution to deposition to PDP, they have 93% success. So you could clearly see which the blocking point is. And that is why from the introduction, I am giving so much importance to the process of crystallization.

The whole consortium is actually being closed because they think that the idea of high throughput would not be possible. They need personalized care to do some surface engineering and different strategies which I am talking about. And that is the only way to get success. So what was the realization here? When you were working with gene, you have a high throughput process; you could sequence all the genes. But from knowing the sequence, cloning, you probably have good success but from protein to structure, a big problem is this crystallization.

(Refer Slide Time: 24:25)

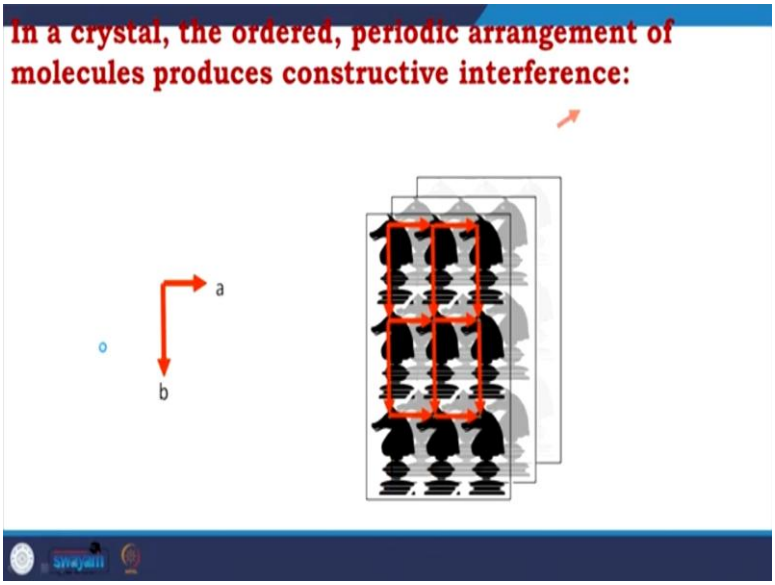
So growing crystal is clearly a hurdle. But why it is such a big hurdle? I would try to give you an explanation in terms of thermodynamics.

(Refer Slide Time: 24:41)



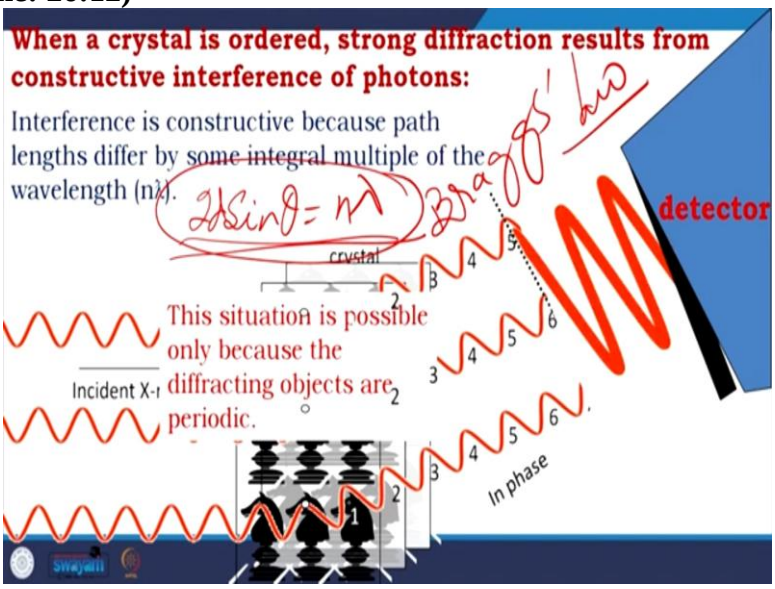
But before that I would like to show you the thing I talked in the last class, why single molecule is not good enough to get diffraction. So in a crystal the diffraction signal is amplified by the large number of repeating units. When you see diffraction from a single molecule, is not possible to measure, whereas when you have 100 mm^3 crystal, it contains 10^{12}unit cells. It is much more bigger mammoth and have a architecture where the diffraction intensity is much higher because diffraction intensity is proportional to the number of unit cell in the crystal.

(Refer Slide Time: 25:33)



In a crystal the ordered periodic arrangement of the molecule produces constructive interference. So if you look at the pattern closely, you will see that, in all the direction how they have developed the architecture of a three dimensional crystal.

(Refer Slide Time: 26:12)



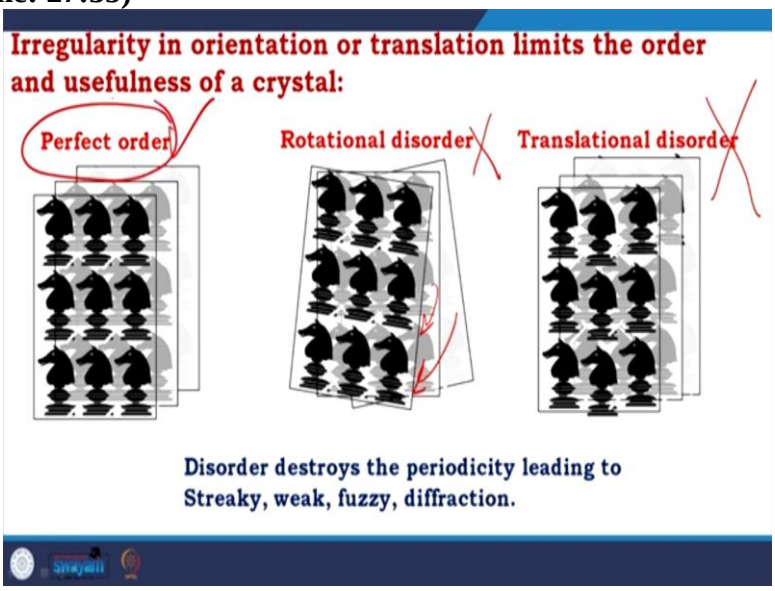
When a crystal is ordered, strong diffraction results from constructive interference of photons, you see the incident X ray which hit the crystal, interference is constructive because path lengths differ by some integral multiple of the wavelength. What I am talking about is a mathematical term which is called

$2d \sin\theta = n\lambda$

That this is the famous Braggs law we will talk about. This situation is possible only when the diffracting objects have periodic, to be periodic, all the planeshave to be in equal distance

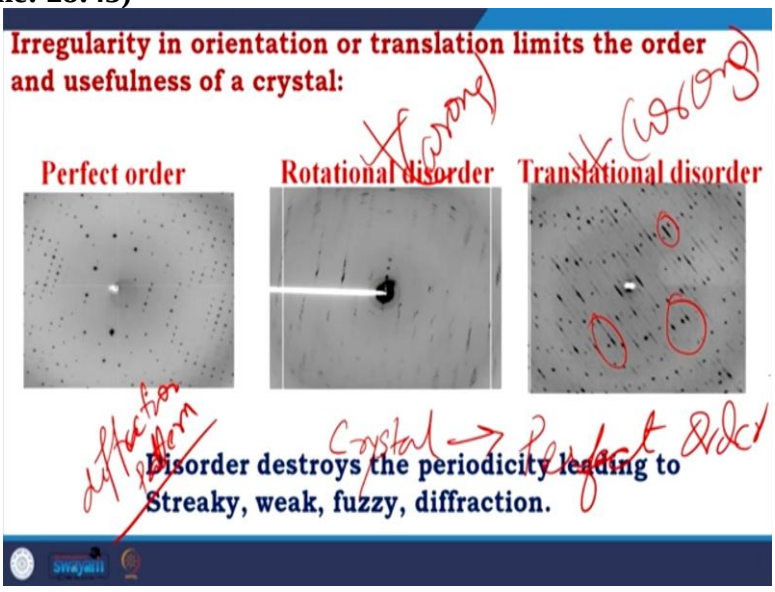
then it would be getting the periodic function and then it would be in phase so detector will get that.

(Refer Slide Time: 27:35)



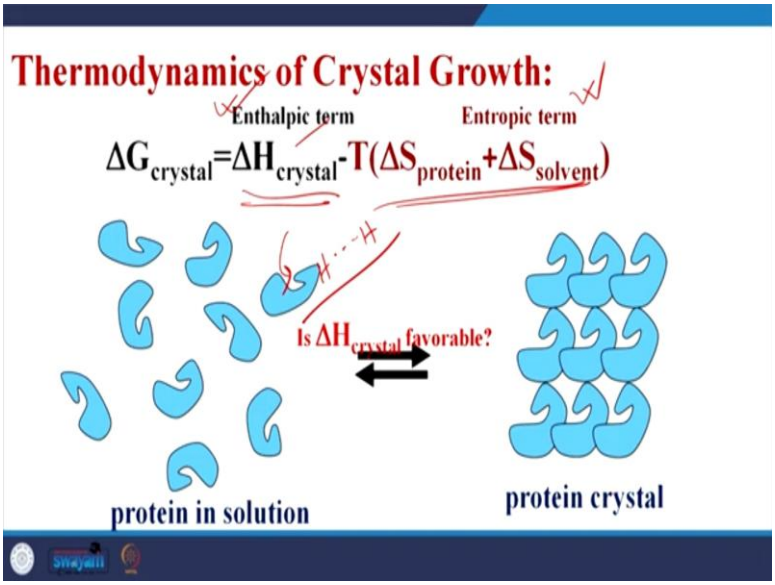
And you will get good signal. But if this is not periodic, if there is irregularity, you will see that the crystal is not giving good diffraction, so irregularity in orientation or translation limits the order and usefulness of a crystal. Disorder destroys the periodicity leading to streaky, weak, fuzzy diffraction. So now, you know you have to get a crystal but you also have to get the crystal in perfect order.

(Refer Slide Time: 28:43)



What is the effect? If it is perfect order, you see, that is the diffraction pattern. Here when there is a rotational problem, you see that, you do not get the spots properly, here you get the spots but if you look, you get the mix up of the spots.

(Refer Slide Time: 29:51)



So what makes crystallization such a difficult challenge? If we look at the thermodynamics, we know that

$\Delta G = \Delta H - T\Delta S$

This is the thermodynamic formula of anything happening spontaneously.

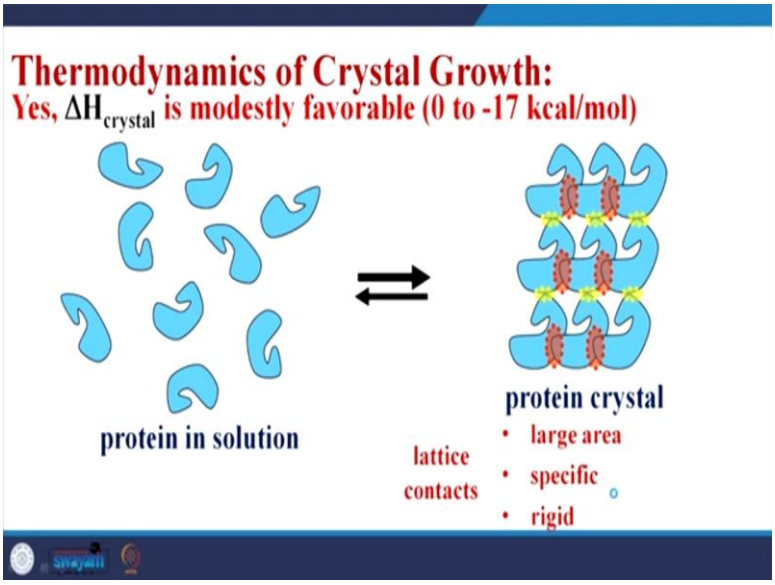
So for crystal growth,

$\Delta G_{crystal} = \Delta H_{crystal} - T(\Delta S_{protein} + \Delta S_{solvent})$

So this $\Delta S_{protein} + \Delta S_{solvent}$ is the entropic term and $\Delta H_{crystal}$ is the enthalpy term.

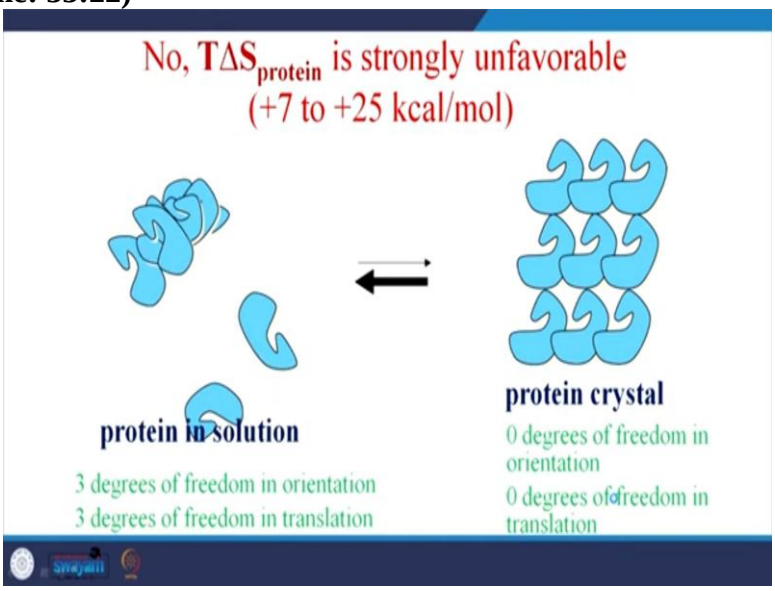
Presence of hydrogen bond includes the enthalpic term. What is entropic term? If you look at, there are proteins and there are solvents, when a protein grows, number of amino acids come together to form protein, let us say 250 amino acid form one protein. So formation of protein is entropically unfavorable. Entropy is increasing noise, entropy is disorderness. So this is the representation of protein in solution, and this is the ordered representation of protein in crystal.

(Refer Slide Time: 32:51)



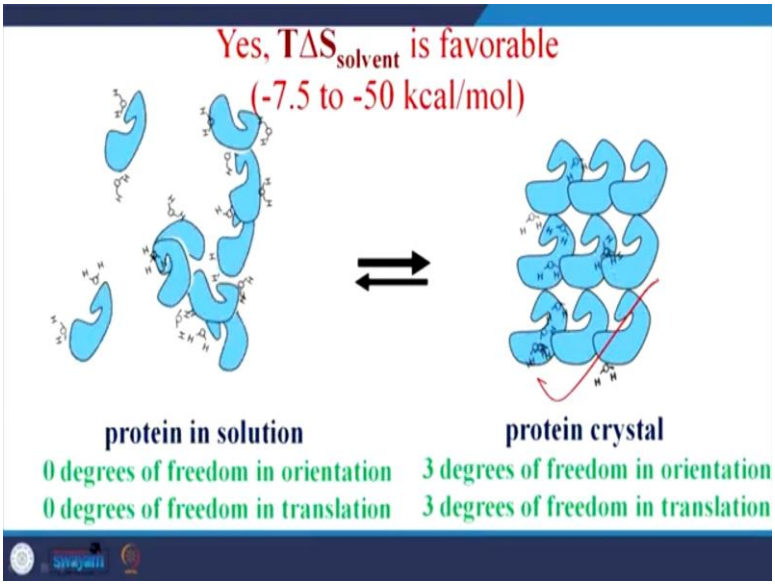
$\Delta H_{crystal}$ is modestly favorable, statistics says 0 to -17 kcal/mol, because if you see protein in solution, there is no interaction, whereas in protein crystal there are lattice contacts in large area, and the interaction makes it rigid. So this is favorable.

(Refer Slide Time: 33:22)



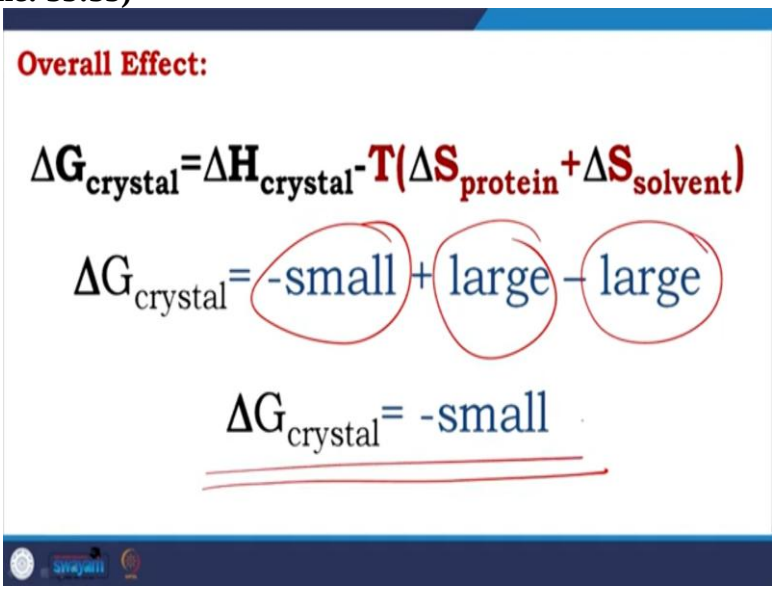
Is $T\Delta S_{protein}$ favorable? If you look at protein in solution and protein in crystal, you will see that protein in solution could move freely, whereas protein crystal could not move freely. So, in solution, there are 3 degrees of freedom in orientation and 3 degrees in translation. Whereas in protein crystal, 0 degrees of freedom in orientation and 0 degrees of freedom in translation. The formation of the crystal is entropically unfavorable. So, it is + 7 to + 25 kcal/mole when it is unfavorable.

(Refer Slide Time: 34:12)



The next question is, is $T\Delta S_{solvent}$ favorable in time of crystal growth? In the protein in solution, though the protein is moving, the water molecules are attached to the protein. So the water molecules are moving, but they do not have apparent movement. It is like you are moving in the car. The car goes from one place to another, but you are not moving. So the car is getting work. You are not doing anything. In protein crystals, proteins are not moving, allowing the solvent to move freely. And that is why here in case entropically winning. So for the solvent, 0 degrees of freedom of orientation and 0 degrees of freedom in translation, here it is 3 degrees of freedom in orientation and 3 degrees of freedom in translation. So it is entropically favorable - 7.5 to - 50 kcal/mol.

(Refer Slide Time: 35:35)

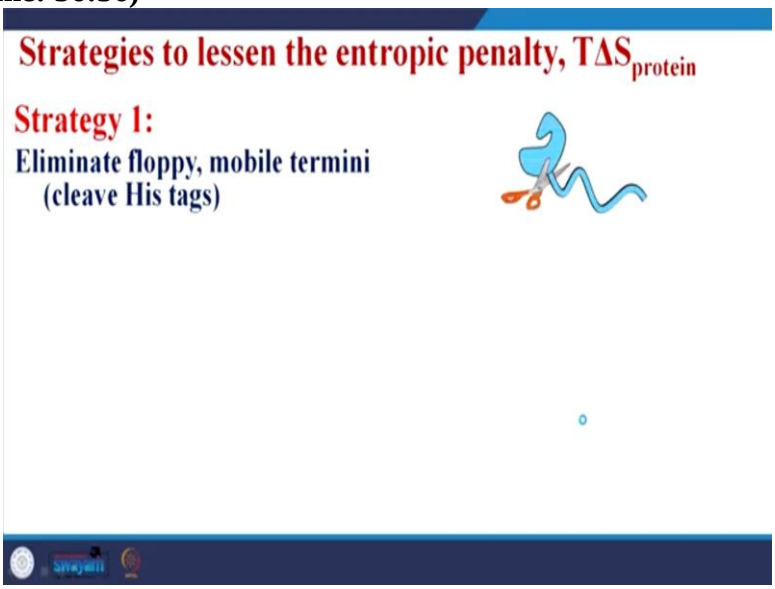


So if we look at the overall effect

$\Delta G_{crystal} = \Delta H_{crystal} - T(\Delta S_{protein} + \Delta S_{solvent})$

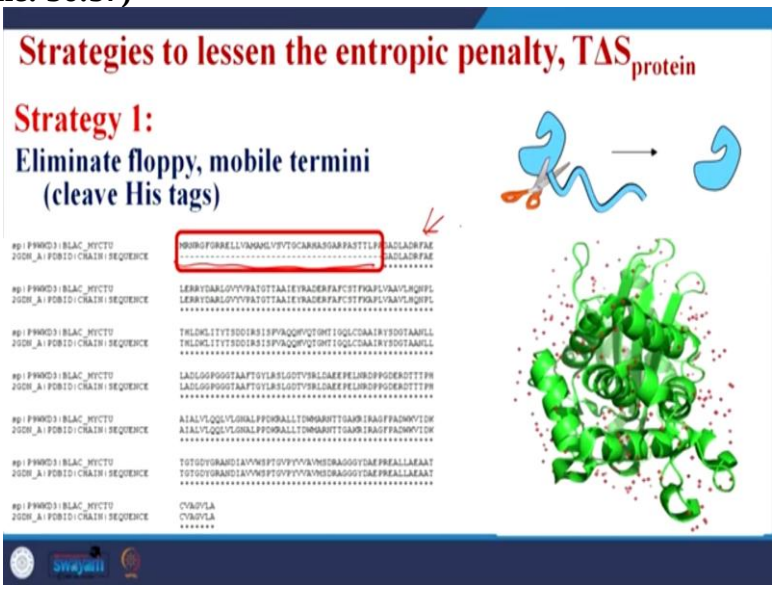
$\Delta G_{crystal}$ equal to negatively small favorable, largely unfavorable, and large favorable, so overall crystal growth is in thermodynamically small favorable. So that is why it is difficult to get a crystal, but there is a chance of around 8 to 10% if you do your other works perfectly.

(Refer Slide Time: 36:36)



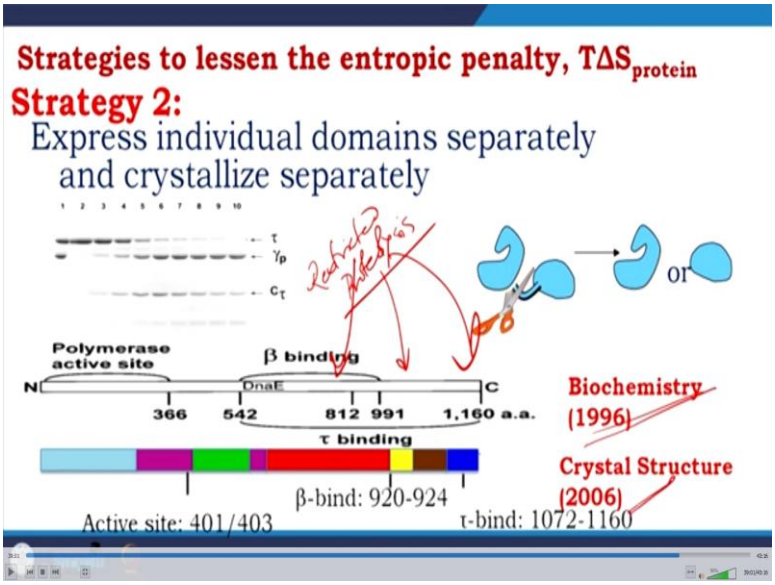
You could take personalized strategies to lessen the entropic penalty, the $T\Delta S_{protein}$, how? You could eliminate floppy, mobile termini. Like I talked about His tag, you could have cut the His-tag or cut other parts of the protein that does not contribute to the ordered structure.

(Refer Slide Time: 36:57)



So look at the protein we have taken this part and these proteins, the full protein was not giving us crystal whereas, after we have taken out this part we got the crystal. So the strategy to eliminate floppy, mobile harmony works here, and this strategy works for many cases.

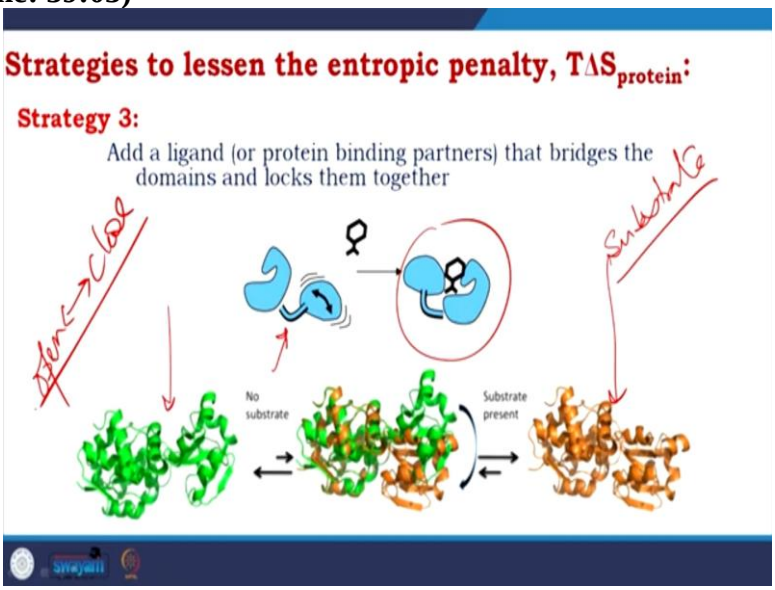
(Refer Slide Time: 37:26)



Then express individual domains separately and crystallize separately; what does that mean? So suppose you have a big protein. You do something we call restricted proteolysis; what is restricted proteolysis? Restricted proteolysis is when you get the protein and add protease from very small to increasing concentration. The protease will cut your proteins, but they will start chopping the protein more and more with increasing concentration.

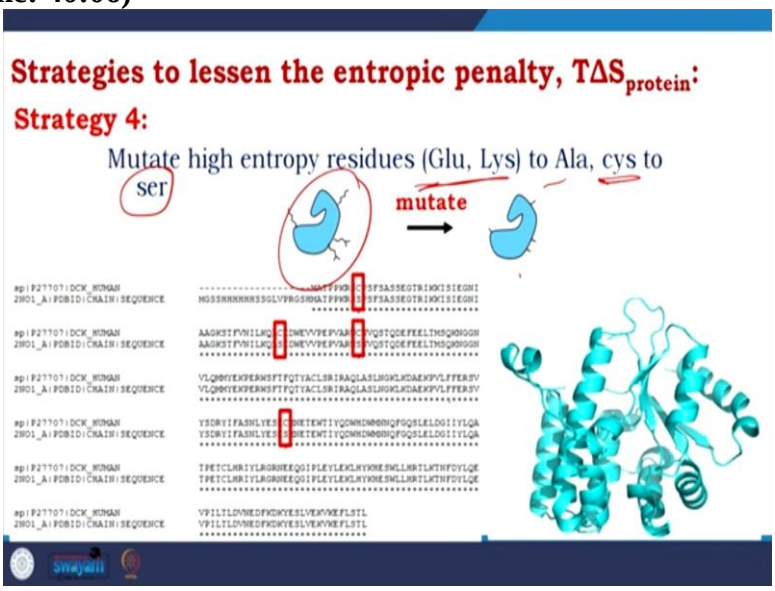
Now at different concentrations, you run the gel and you find that in which concentration of the protease you know, domain you get and then you go through purification techniques like size exclusion chromatography and then you separate the domains and crystallize them. So they get the biochemistry here, understand them, and then get the domains and publish. They did biochemistry in 1996. And it takes them 6, 10 years to get the crystal structure because this is a very big protein, and getting crystal is very difficult.

(Refer Slide Time: 39:03)



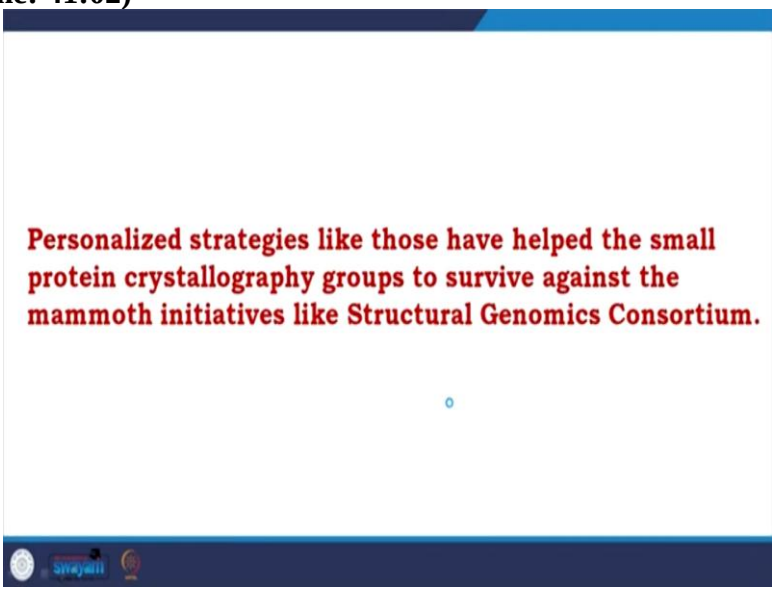
Add a ligand that bridges that domain and locks them together. Sometimes, you have 2 domains, and both are stable but moving based on the loop. The loop is floppy, so the proteins are stable, but they move against the loop. In that case, you add a small molecule, and you stable the whole system and crystallize it. Like here, you see that this is an open confirmation, and when the substrate is added, this goes to a closed conformation. Sometimes you get a crystal for both of them, and you study the open and close confirmation. You get to know the mechanism how it is working?

(Refer Slide Time: 40:06)



The fourth strategy is to lessen the entropic penalty to muted high entropy residues, bigger residues like glue or lysine, to alanine. You have these bigger residues, mute them to alanine, or mutate cysteine to serine; when you know there is no chance of disulfide bonding, then serine will introduce some interaction, it significantly improves the crystallizability.

(Refer Slide Time: 41:02)



So by doing that, by taking those personalized strategies, you could significantly improve the chance of crystallization of your protein. And that is what is missing in the Structural Genomics Consortium. So those personalized strategies would help the small protein crystallography groups. Suppose you compare the structural genomics initiative with 1000 pupils, including big crystallographers. In that case, with many cloning pupils, and a lot of high throughput things, a lot of money is working.