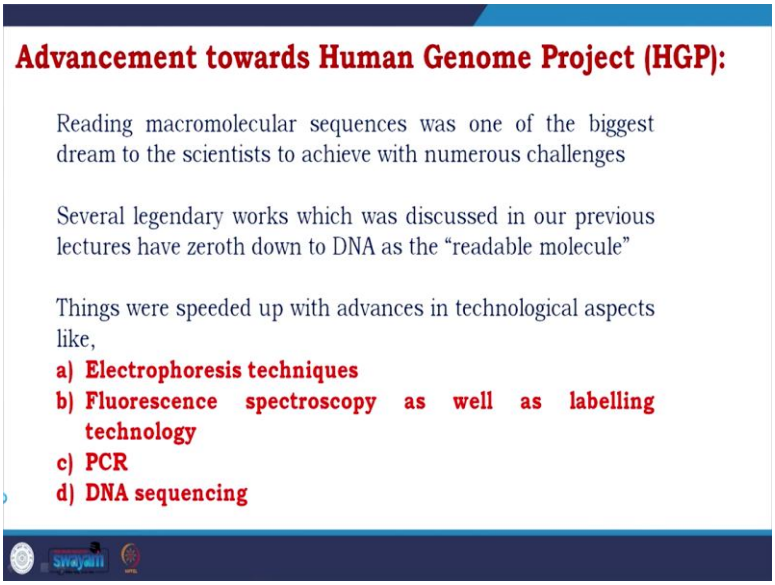
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology – Roorkee

Lecture 05
Introduction: Post Genomic Era

Hi, welcome again to the course on structural Biology. We are still going to the introduction module, and today is the last lecture of this module. Today we will talk about the post-genomic Era. We have already talked about how gene sequencing convinced us and realized that we need to fill Biology by reading at least one macromolecule of the DNA. And then, it leads to the great revolution of Biology,

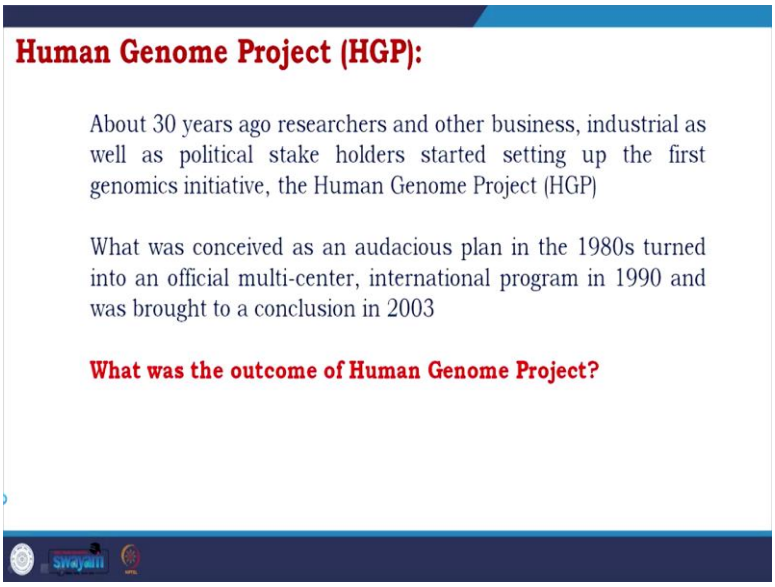
the Human Genome Project.

(Refer Slide Time: 01:01)



So how the advancement happens first, as I told macromolecular reading sequences was one of the biggest dreams to the scientist to achieve with numerous challenges. The previous lectures discussed several legendary works, which zeroth down to DNA as the readable molecule. When it was zeroing to DNA, it was heated up with advances like the electrophoresis technique in DNA sequencing. In the beginning, electrophoresis technique separation was a challenge which was then become advanced, became capillary, become automated with laser technology, Spectroscopy. Labelling Technology has become advanced. PCR, invented by Carol Moulin's is one of the greatest advancements and, last but not the least, the DNA sequence Technique by Sanger and Maxam Gilbert.

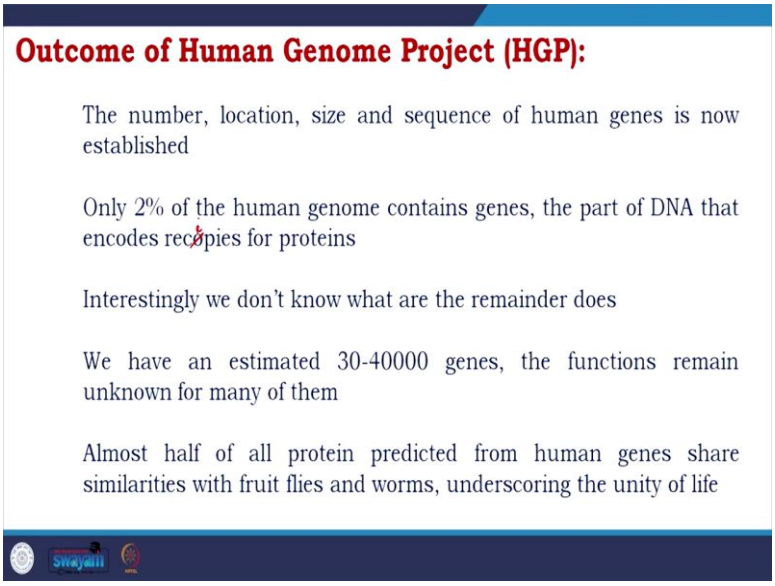
(Refer Slide Time: 02:25)



About 30 years ago, the researchers in 1980 were around started in 1990. Other business industry and political stakeholders started setting up the first genomic initiative, the Human Genome Project (HGP). What was conceived as an audacious plan in the '80s turned into an official multicenter International program in 1990 and was concluded in 2003.

Though we discussed, I still want to summarise. What was the outcome of this Human Genome Project?

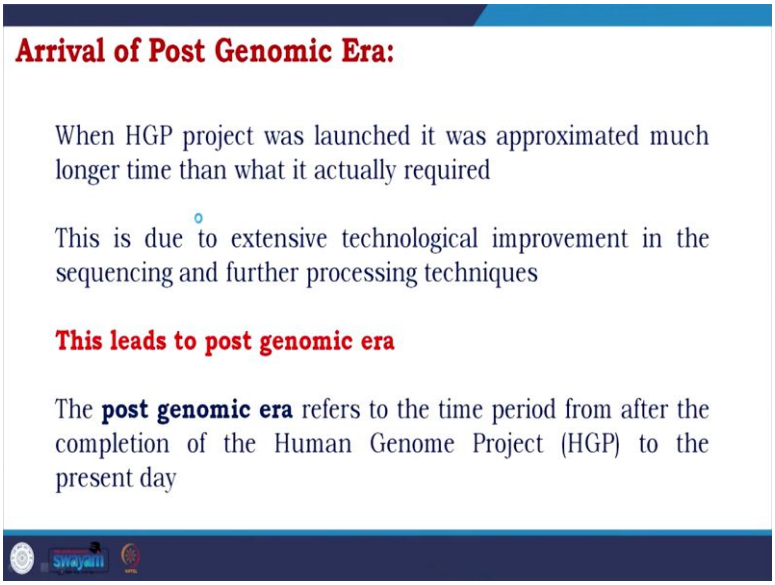
(Refer Slide Time: 03:05)



The number, location, size, and sequence of human genes are now established. We know that only 2% of the Human Genome contains the part of DNA that encodes protein recipes. The rest of that, we know junk. Interestingly, we do not know what the remainder does. So we know that 2% gives us protein interestingly with such a long history of research in the field of

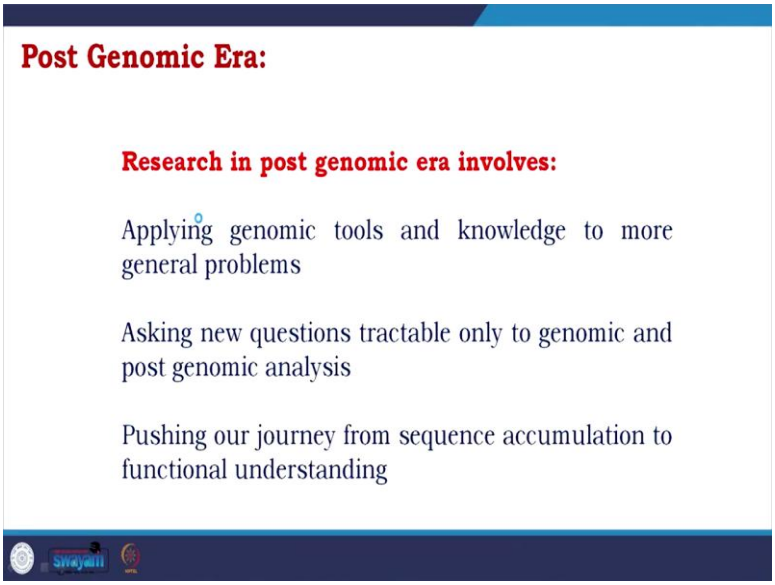
DNA in the field of the genome we still working on 2% is it not very interesting, so that we came to know. We have an estimated 30 to 40 thousand in the functions. Remains are unknown for many of them. And almost half of all proteins predicted from human genes share similarities with fruit flies and worms which tells us how united life is.

(Refer Slide Time: 04:31)



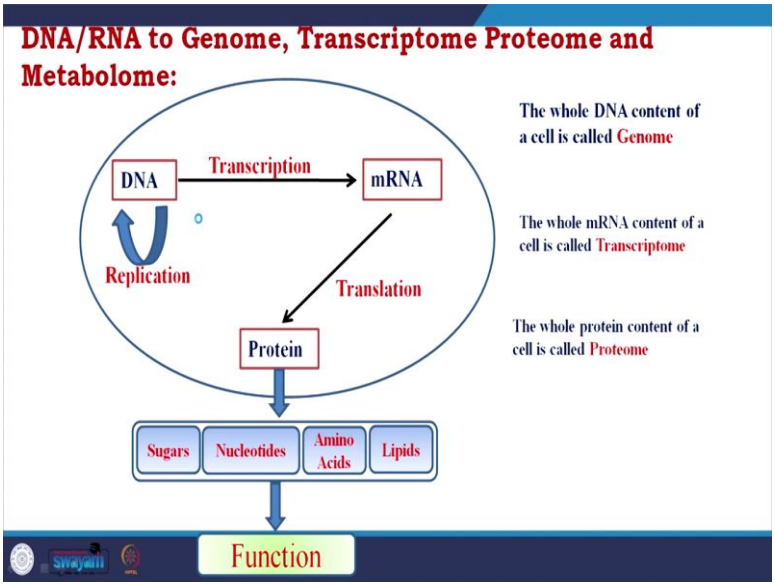
What is the post-genomic Era? When the Human Genome Project was launched, it was approximated much longer than required because many technical limitations restrict that time. This is due to extensive technological improvement in the sequencing and further processing technique. And as I told these the advancements though, we say it is because of Human Genome Project. We mean that when Human Genome Project was happening, it was a perfect combo of people from a basic science background and technological background coming together, working together, thinking together, sharing the thoughts that led to a lot of improvements and a lot of innovation. We have elevated ourselves from the genomic Era to the post-genomic Era. The post-genomic Era: Period after completing the Human Genome Project to even today.

(Refer Slide Time: 05:44)



In the post-genomic Era, the research applies genomic tools and knowledge to more general problems. So we solve the genomes, but we want to go further from genomes. We want to solve meaningful problems, asking new questions tractable only to genomic and proteomic analysis, pushing our journey from sequence accumulation to understanding functional inputs.

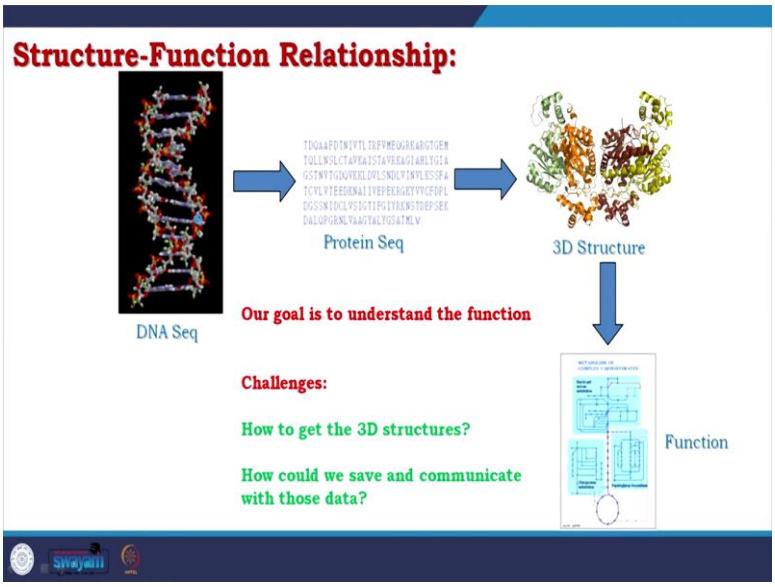
(Refer Slide Time: 07:35)



We already know that DNA gives DNA replication, RNA transcription, and mRNA gives protein through translation. But the following up from protein to sugar, nucleotides, amino acid with the small molecules which are metabolome which gives us actual track to the function. So the genome we get from sequencing, called genome sequencing and transcriptome sequencing, gives us the proteome.

And then, we do metabolic studies now through new technologies like mass spectrometry, like high-resolution NMR.

(Refer Slide Time: 08:43)

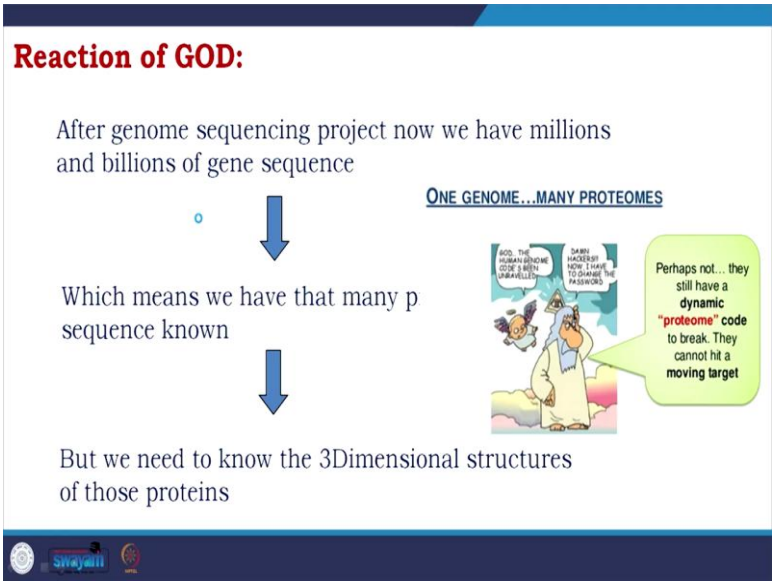


So as I told our journey is from DNA sequencing, and when we get the DNA sequence, we get the protein sequence, and you want to achieve function, so the greatest barrier here is to get the structure. We need to get the structures, which is the importance of structural Biology.

I also want to talk about different problems that have given a new field to explore to prosper in this post genomic era and how we could save and communicate with those data. So now you know in the Era before genomic sequencing happened.

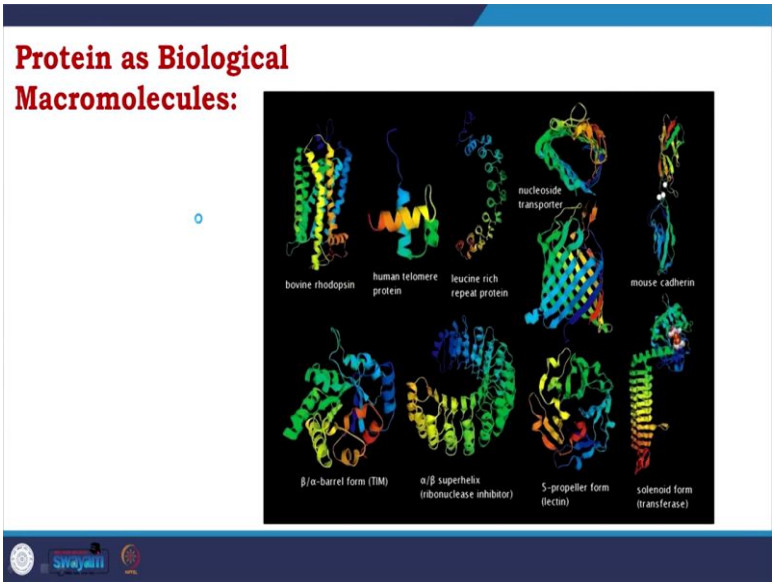
But genome sequencing is becoming easier day by day, and daily data is coming hugely. So that leads to the generation of a field where these data have to be kept in a very, very proper way. Then you come to protein which is different 20 amino acids. Then you come to the structure, which is coordinates. There are different types of profiles. You have to save them differently. You have to make them interact with each other. You have to make extract data from them, and that huge amount of data gives birth to a field that we call system Biology.

(Refer Slide Time: 12:03)



So after the genome sequencing project, now we have millions and billions of gene sequences, which means we have that many protein sequences known. But we need to know the 3-dimensional structures of the protein how they interact properly. How do the proteins interact with other proteins? How do the proteins interact with other macromolecules like DNA or RNA? How the protein interacts with small molecules like substrates and inhibitors.

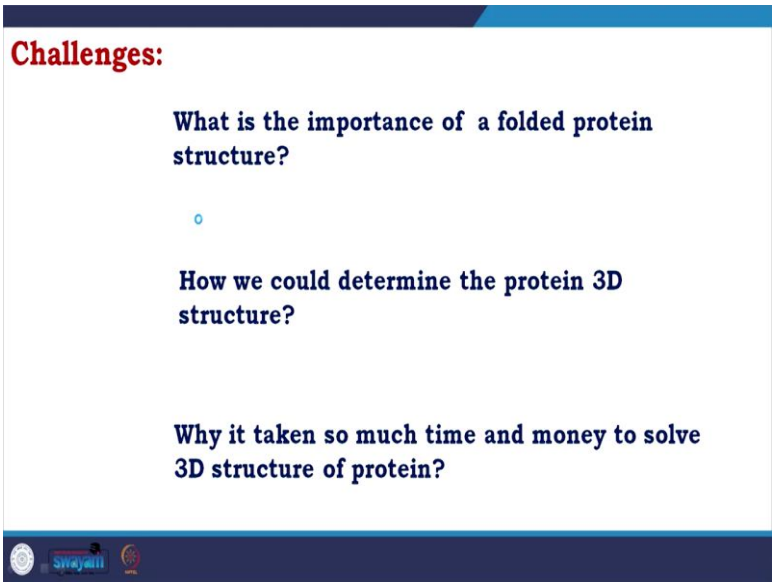
(Refer Slide Time: 13:02)



So who are the biological macromolecules? In a study done explored, DNA, RNA, and protein are all important. But protein is the most important macromolecule. Why? Because if you see proteins are functional, protein talks about what or how we work? How do we stay? More importantly, how the diseases are coming and protein function could only be understood through their 3D structure.

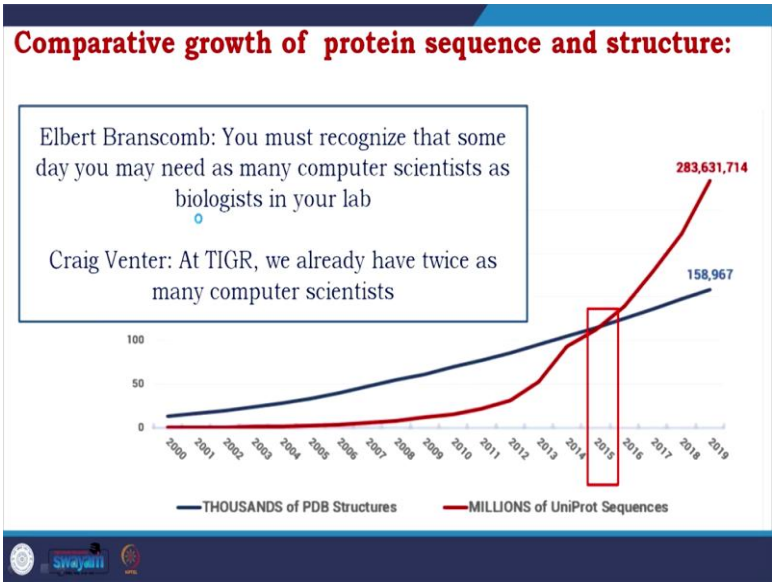
And if you look, this is a good example, where I am given different proteins. All the different proteins have different 3D structures, unique 3D structures. Whereas when you talk about DNA and RNA, they do not have such diversity in their structures. So, many other factors in the next module discussion will clearly show why protein structure could be the major focus of structural Biology.

(Refer Slide Time: 14:21)



So what are the challenges we are going to handle? What is the importance of folded protein structure? How we could determine the protein structure and why it took so much time and money to solve a protein's 3D structure.

(Refer Slide Time: 14:34)

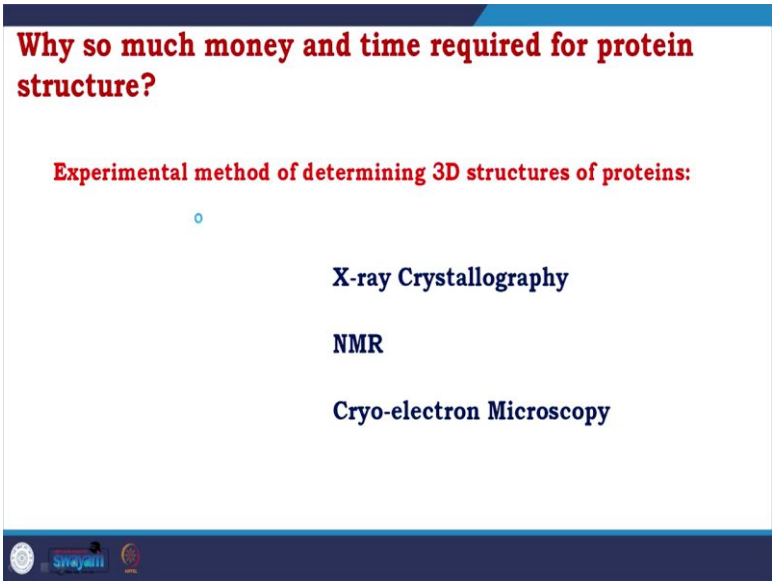


In around 2015, the journey of protein structure was passed by the genome or protein sequence determination, and then it rose. Now there is a huge difference and every day and

night the differences in enhancing. So you also want to understand why that much difference is happening between the determinations of protein structure versus the determination of protein sequence.

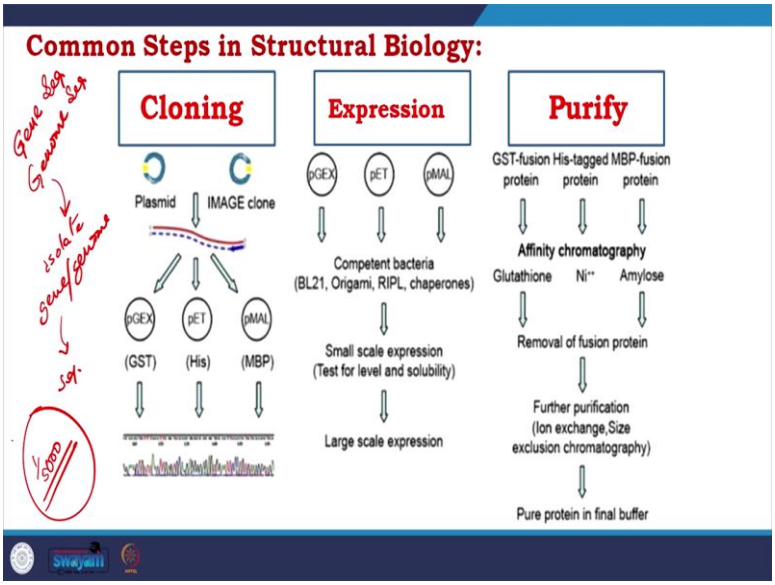
Because we understood the basics of genome and transcriptome sequencing, we know why it is becoming rapid and why it is becoming easier day by day. Still, we will look at the structural Biology technique in very basic in this module.

(Refer Slide Time: 16:43)



Why are so much money and time required for protein structure? So first, we have to know about the techniques used for determining high-resolution 3D structure: x-ray crystallography, NMR, and cryo-electron microscope.

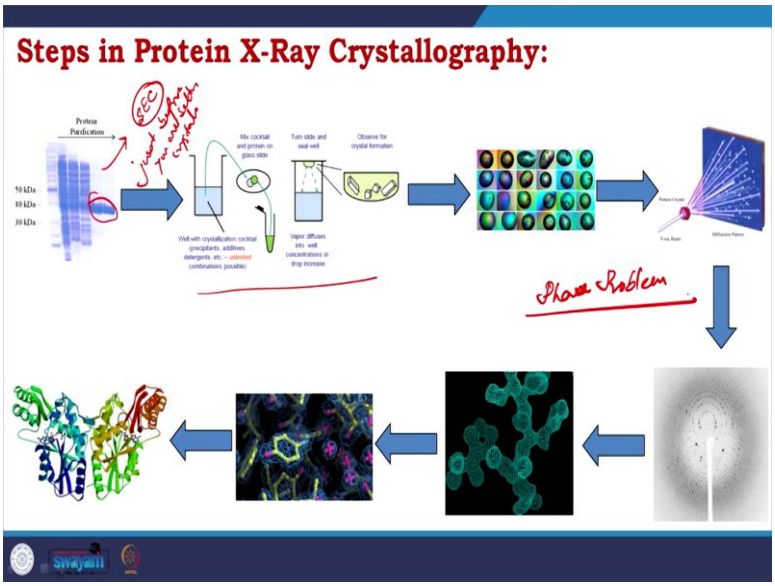
(Refer Slide Time: 16:58)



So, what are the common steps in structural Biology? First, you need to do clothing, so the first differentiation starts with gene sequencing or genome sequencing. You directly isolate the gene or genome and go for sequencing. Whereas when you want to work with protein first, you have to produce your protein exclusively. I will discuss this in detail, but the first part is that you have to go back to DNA and do a cloning process by which you could produce your protein. So cloning is a technique that helps you put your gene in some situations, which will help you control the further preceding the overexpression and the purification. These are most critical, and these are the differentiable factor that takes a lot of time

Purification of DNA is much easier, whereas purification of protein is the major thing is molecule linked by covalent bonds, right? Not here the only factor is a covalent bond. In protein, you have to Give the importance of non-covalent bonds like Hydrogen bonds, salt bridges, hydrophobic like stacking, metal. Cloning, expression, and purification would significantly differ in time and cost.

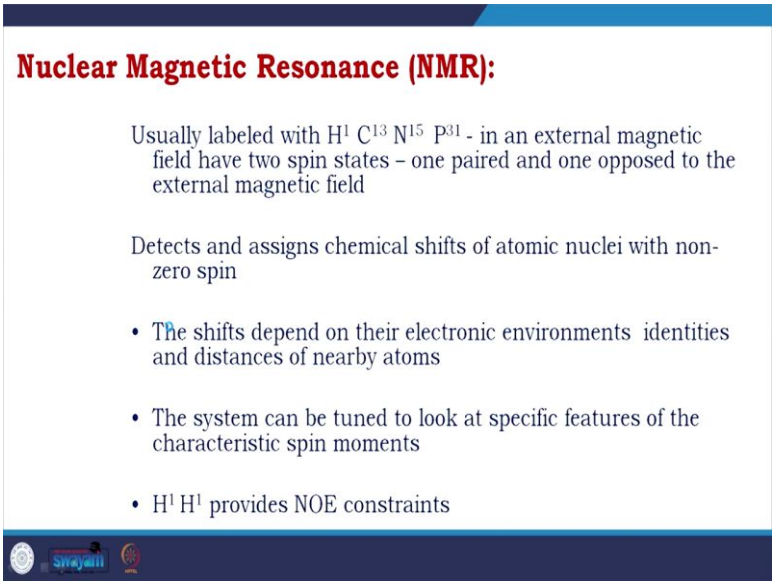
(Refer Slide Time: 21:37)



Now, if you look at the steps in protein x-ray crystallography, you get pure protein. You see, the pure single band's band will determine it is a pure protein. When you want to set up the drops, you should perform size exclusion chromatography. When you are freezing and thawing, there is a high probability that protein might have at least a slight partial denaturation. And this could be deadly towards the development of crystallization. Then you are going to set the crystals. There are many techniques I will discuss. But with all the techniques, everything being standardized, even the world's best crystallographer could not tell you the condition of crystallization. This is one very, very uncertain step in protein

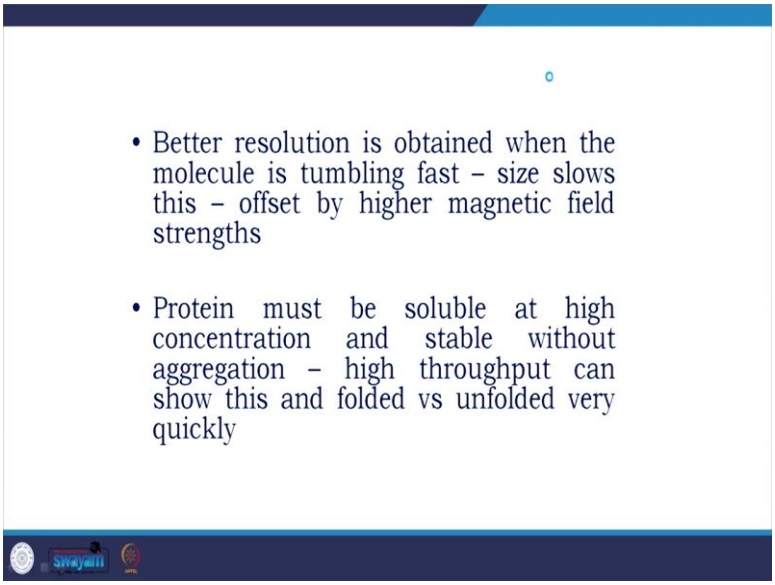
crystallography success. Crystallization is one of the most critical steps in protein crystallography success. Then you get crystals. You have to keep checking the crystal are deflectable or not. Then the crystals are diffracted you will get the diffraction pattern. Then you have many problems, but one of the significant problems is the phase problem in crystallography. If you pass through that, you will get electron density. Then you will go to build the model, and finally, get the refined model.

(Refer Slide Time: 25:10)



The second method, nuclear magnetic resonance: In nuclear magnetic resonance, you need labeling with odd numbers like H^1 C^{13} N^{15} P^{31}. It detects and assigns the chemical shifts of atomic nuclei with non-zero spin. The shift depends on their electronic environment identities and distances of nearby atoms.

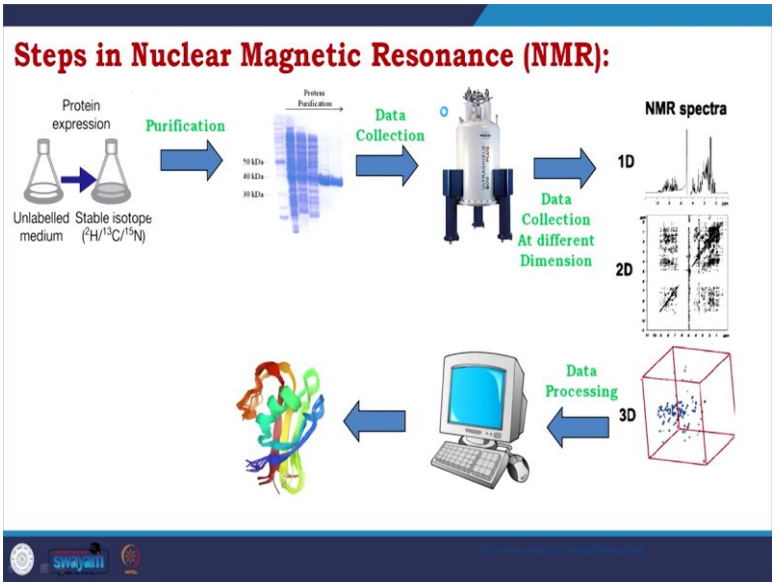
(Refer Slide Time: 27:38)



I will talk about the technological details, but let us determine what happens in protein. Better resolution is obtained from the molecule is tumbling fast. So the resolution depends on the movement. It is not a problem for small molecules because they are small, so they are in the solution that dribble fast.

First of all, it must be soluble at a very high concentration, and it could not be aggregate, and after that, there is a restriction of molecular weight. Increasing molecular weight tumbling would be slower; hence the Spectre starts overlapping. So NMR had a problem going into higher molecular weight protein structure determination. But it has certain advantages also, and we will discuss that later, but let us come to the steps.

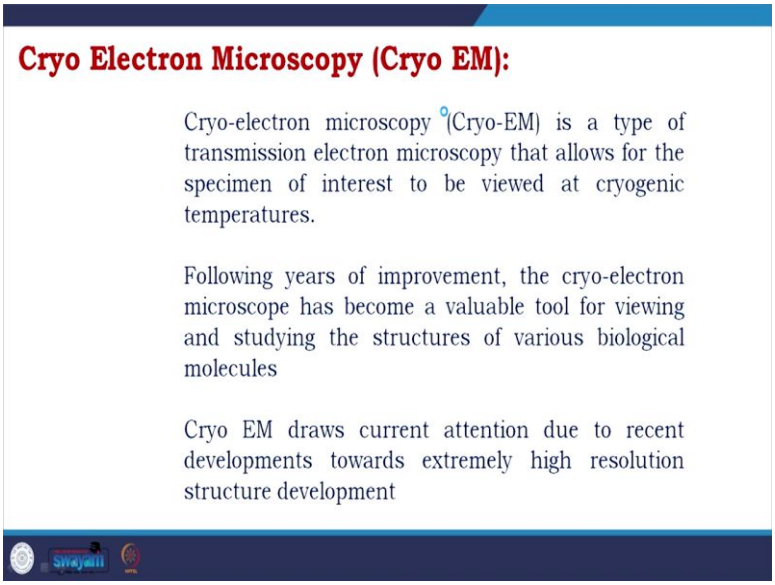
(Refer Slide Time: 28:53)



I talked about Protein Purification. Protein purification gets even more complicated because you have to use the isotopes here. You have to grow them in the presence of those isotopes.

And that makes certain things, one; it would be very costly because you have to grow them on a solution filled with those isotopes, And growing with those isotopes, the protein production might be reduced. So those are the extra challenges. Again you need the purified protein in very high concentration. Then you come to data collection with NMR. You will collect different dimensions of data 1D data, 2D data, 3D data. So, different dimension data collection, then data processing and the model.

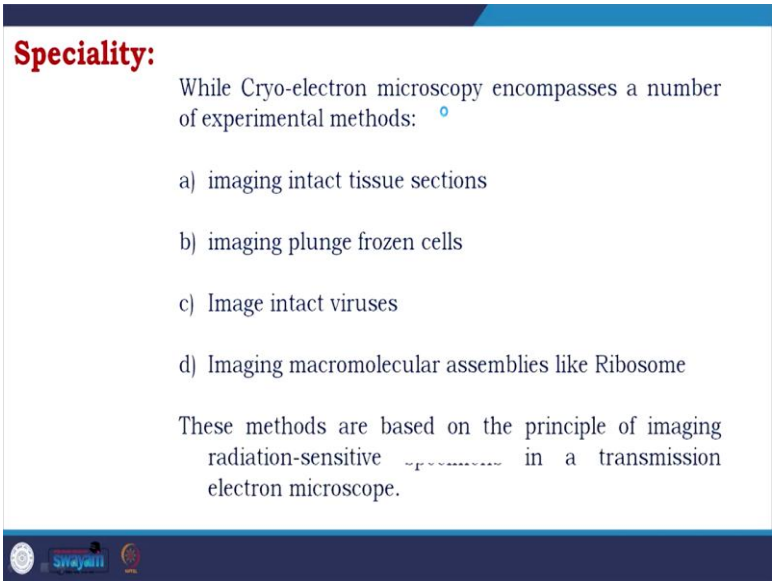
(Refer Slide Time: 30:33)



Cryo-Electron microscopy is a type of Transmission electron microscopy that allows the specimen of interest to be viewed at cryogenic temperature. X-ray diffraction, NMR Spectroscopy, Cryo-electron microscopy, these three techniques in different ways contribute to structural Biology. Cryo-electron started really with low resolution. But following years of improvement, Cryo-electron microscopy has become a valuable tool for viewing and storing the structures of various biological macromolecules. And especially it draws current attention due to recent developments. Now people are getting structures at extremely high resolution. In addition, the recent Nobel Prize has created a huge interest in Cryo-electron microscopy.

One of the biggest problems for Cryo-electron microscopy was that the machine was very costly. So, very few institutes in the world have developed Cryo-electron microscope facilities. And in India, currently, eight institutes have Cryo-electron microscope facilities.

(Refer Slide Time: 32:22)

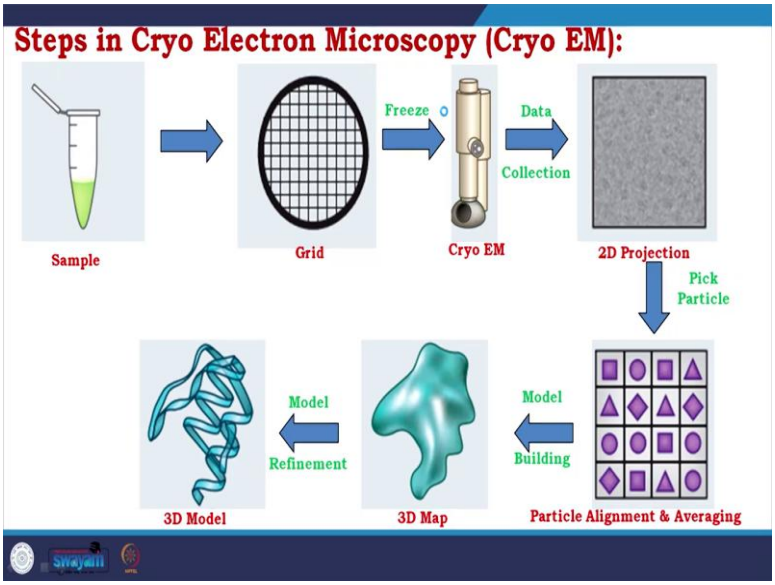


While Cryo-electron microscopy encompasses a number of the experimental method:

a) image intact tissue sections, b) imaging plunge frozen cells, c) intact image viruses, and d) image macromolecular assemblies like ribosomes. So as I told you when we are talking about protein crystallography, it is very efficient to provide high-resolution, but it is somehow restricted to the proteins. NMR has other applications, but it has restrictions of size.

Cryo-electron microscope is an instrument, a technique of the future. I will talk about the details.

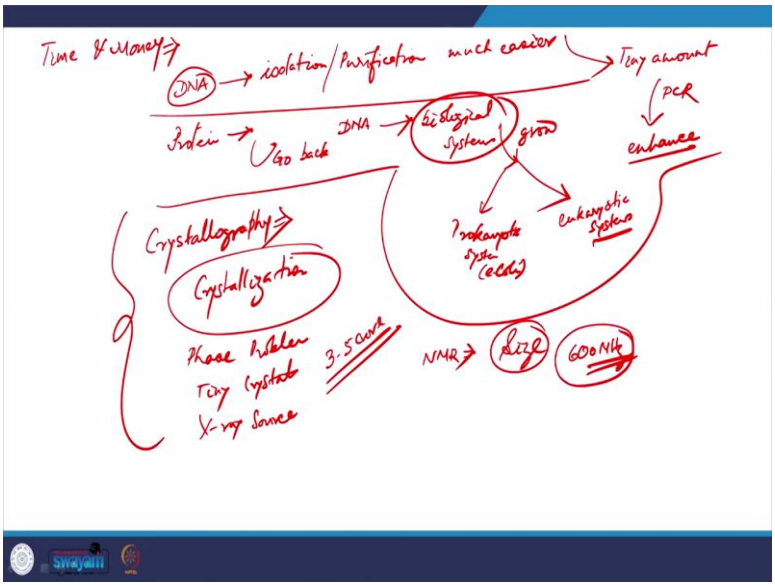
(Refer Slide Time: 33:34)



The steps are sample preparation. Then the sample will be going to grid preparation. The grids would be frozen, and then the grid will be coming under the microscope. And the thinner sections could be collected image, would be collected as a 2D image. And then, the

different 2D images would be matched, aligned, and averaged, giving a 3D map. And that 3D map would be refined to a 3D model.

(Refer Slide Time: 34:21)



Now come about the time and money. First of all, as I told, one of the main problems is when you are working on DNA, from isolation, purification much easier whereas when you are working in protein you have to go back to DNA you have to work on the DNA so that it could express because for DNA if you get a very tiny amount. You do not care because you have Gery Meullons PCR which will enhance the amount of DNA as much as you want. Here, you have to use biological systems which will help you to grow. The first choice would be a prokaryotic system like E. coli. Still, sometimes some proteins could not be expressed or show the required post-translational effects modifications when working with those proteins. You have to go to eukaryotic systems. Your challenge would be one more with the increasing price of the media and the maintenance of the cells. So, cloning, overexpression, and purification is big problem.

Then talking about crystallography, the first challenge, as I told you, is crystallization. A lot of research happened on them, and many statistical data people have explored. So crystallization is one of the major challenges, but there is also a phase problem, problem of tiny crystals. If you get a tiny crystal that would be diffracted and get data to solve it in a synchrotron source, you might not do that in the home sources. Also, with the new mutants, one change in the nucleotide changes the amino acid. There are possibilities of change in confirmation even if the same protein could have been in a different confirmation in its apo state with the addition of different ligands and different interactions.

As I told NMR has restrictions in size, so bigger proteins are not possible. You need a very high protein concentration. For protein, you need at least 600 megahertz. Megahertz is the magnet's power. A more powerful magnet provides you with high resolution.

Also as I told you earlier, you have to produce your protein using multiple isotopes. So that needs a separate setup. That needs you to pay much, much higher than normal protein production.

Coming to Cryo-electron microscopy, the first critical thing is the instrument. In Cryo-EM, you are already described to go for complete cell. But when you are working on a protein, you need a purified protein.

So those are the biggest challenge that makes the 3D structure determination difficult. Is there any other alternative? There are structure prediction methods, theoretical structure prediction, homology modeling, threading, and Abinitio. Now a lot of talking is happening on AI-based structure determination. Alpha fold has predicted structures nearly at the level of experimental structures

I will finish this module with think I talked about in the first lecture of this module. There is a need for a merger between basic Sciences and biological Sciences. And that initiative was coming when DNA research was going into advancement.

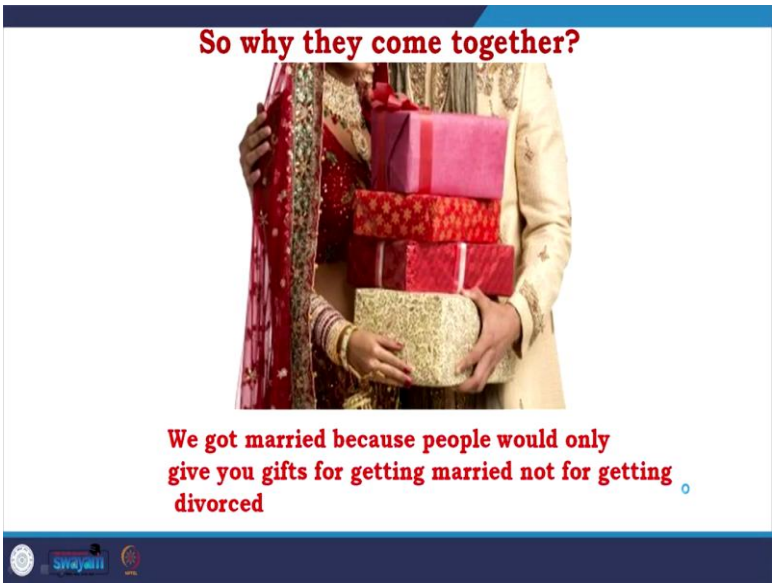
(Refer Slide Time: 44:54)

And I invited all of you to talk about these marriages, and I have my way of explaining the marriage. In reality, I consider this merger a marriage between basic Science and Biology.

(Refer Slide Time: 45:07)

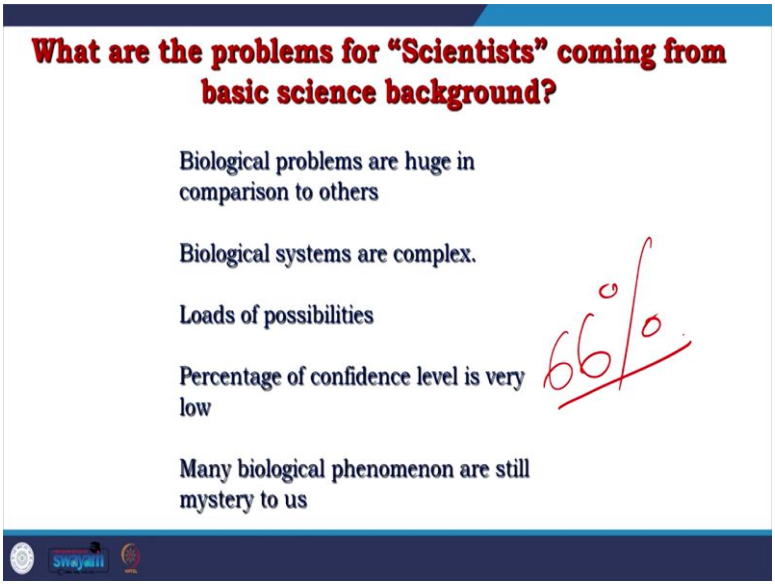
I considered this a typical Indian marriage because here, two fields were completely unknown to themselves. They used to hate each other in normal situations. Basic science has its ego, the biological Science of its conventions. So normal situation was not very friendly, and this was the first experience of living together with such an opposite entity.

(Refer Slide Time: 46:19)



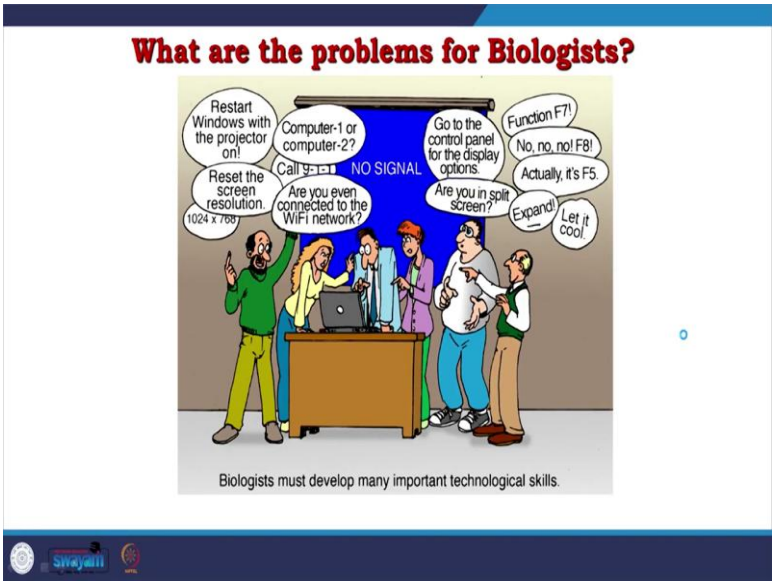
Why would you save why they come together if they have so many differences? They came together because it was an agreement on a common issue as no other better alternative way could be designed. So we got married because people would only give you a gift for getting married, not a divorce.

(Refer Slide Time: 47:31)



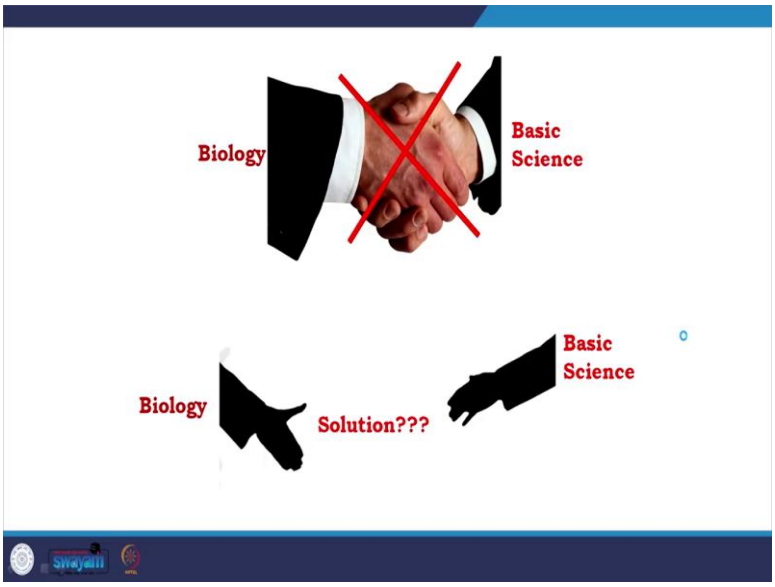
So what are the problems for scientists coming from basic science backgrounds? If you look at it, biological problems are huge compared to others. It isn't easy to work on them without understanding them in detail. The system is complex. Generally, there are lots of possibilities when you say the possibility is a binary, ternary. But if you look at biological problems, there are many possibilities. Percentage of confidence level is very low. We have been giving biological research for such a long time till we do not understand properly how a baby grows properly in the mother's womb. Even with those trillions of dollars we're spending on Cancer Research, and there are so many things, we are still here unable to understand the hundreds and thousands of people dying.

(Refer Slide Time: 49:45)

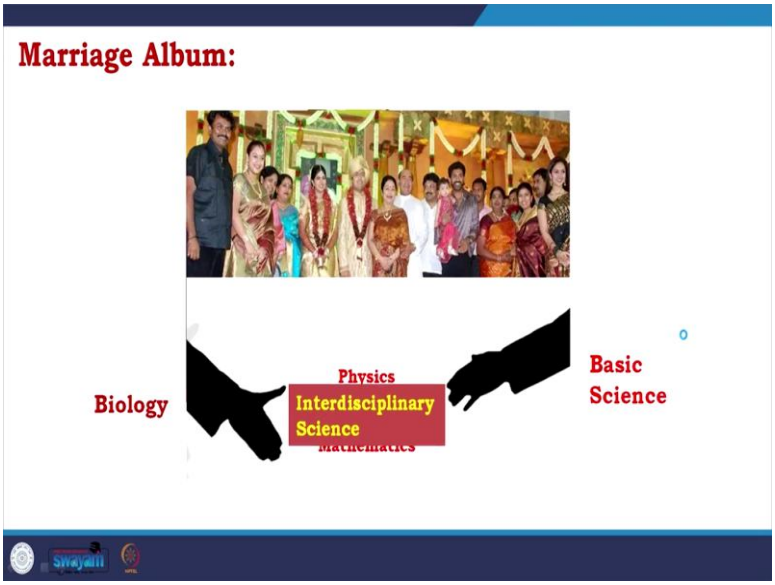


On the other hand, what is the problem for biologists? We should develop technological skills to come closer to defining the problems to people from basic science backgrounds.

(Refer Slide Time: 50:41)



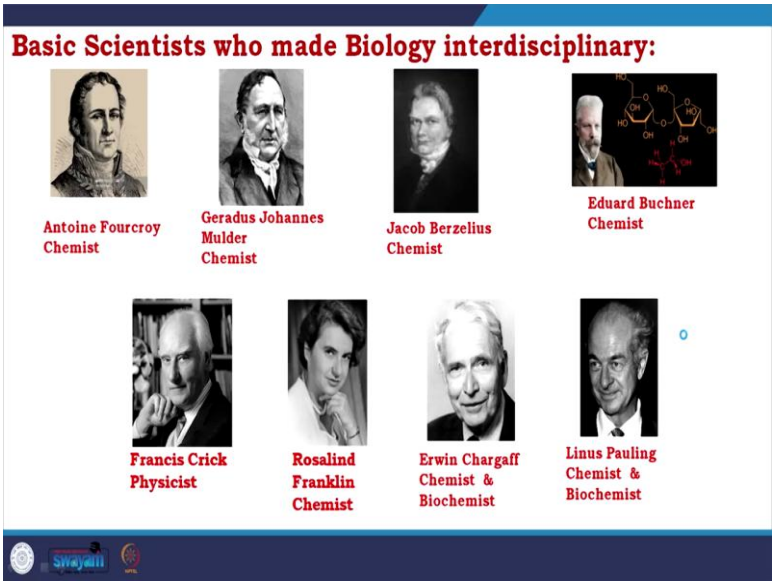
(Refer Slide Time: 51:06)



The photo is only complete when the family can accompany the bride and groom. So the family, in the case of Indian arranged marriage, the family solves a lot of problems. A lot of distance created a lot of misunderstanding would be solved by family involvement.

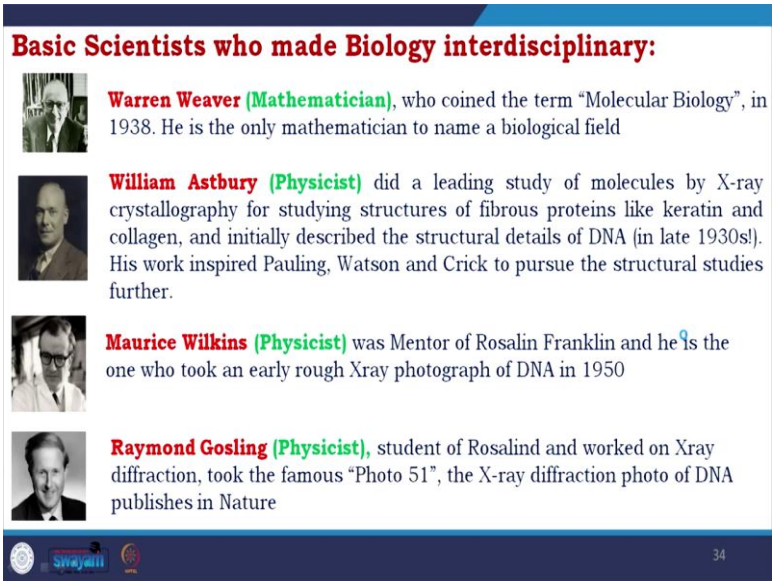
Similarly, the merger of biology and basic Science could be solved by introducing what we say in interdisciplinary Science. The introduction of interdisciplinary Science plays revolutionary ways to solve the distances to push them towards many path-breaking innovations.

(Refer Slide Time: 52:22)



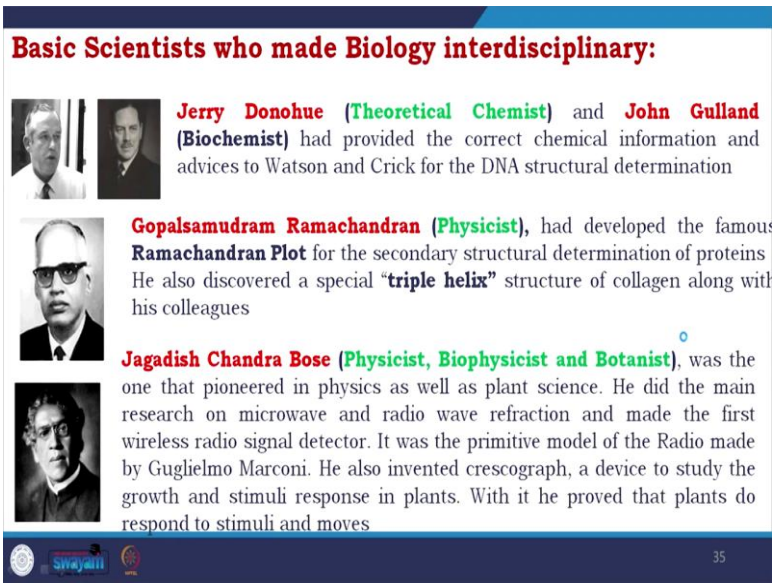
Let us look at the scientist who came from different backgrounds and contributed to Biology. I have already talked about Antoine Fourcroy, Gerardus Johannes, Jacob Berzelius, and Edward Buchner. They have contributed immensely to protein science. There is Francis Crick. We have already talked about the Watson and Crick model of DNA. Rosalin Franklin is considered a tragic hero of Science that she has contributed immensely but did not get her due, the first person to come up with the proper structural details of DNA. Erwin Chargaff and Linus Pauling have shown in-depth contributions in both the fields of chemistry and Biology.

(Refer Slide Time: 54:22)



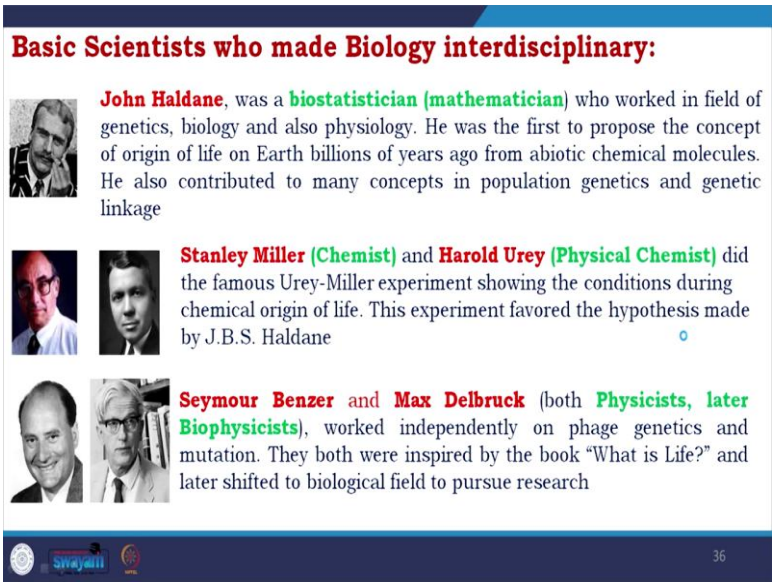
Warren Weaver is a mathematician who coined the term molecular Biology. Biology at the molecular level, and he is the one who has called the basic scientist to contribute in this field. William Astbury is a physicist. He did a leading study of macromolecules by x-ray crystallography for studying the structure of the fibrous protein, like keratin and collagen and initially described the structural details of the DNA in the late 1930s, which have inspired Pauling, Watson-Crick to come up with the structure. Maurice Wilkins was the mentor of Rosalin Franklin, and he is the one who took an early rough x-ray photograph in 1950. Raymond Gosling is a physicist. He is a student of Rosalin and worked on X-ray diffraction. Took the famous 'photo 51' which is the X-ray diffraction pattern of DNA published in nature.

(Refer Slide Time: 55:45)



Another two people, Jerry Donohue and John Gulland, was theoretical chemist, they provided the correct chemical information and advice to Watson and Crick DNA structure determination. I am very excited to talk about Gopal samudram Ramachandran, a physicist who has an immense contribution to the Indian scientist and Ramachandran's thought, which I will discuss in detail. How Ramachandran and his group have developed the Science, the hypothesis of protein structure just based on the dihedral angles would excite a young researcher towards, you know, higher Science. And you would be even amazed when you see the basic instrumentation they have used. I would talk about Gopal Samudram Ramachandran and Ramachandran's plot in detail. Also we cannot ignore the discovery of the triple helix structure of collagen by him. Another Indian scientist Jagdish Chandra Bose was a physicist, and he was the only physicist you could say who has contributed to the field of plant botanical Science. He was a Pioneer in physics as well as plants. Did the main research on Microwave and radio wave refraction and made the first Wireless radio signal detector?

(Refer Slide Time: 57:56)



Joh Helden is a biostatistician and mathematician who worked in genetics and has several hypotheses. Stanley Miller and Herald Urey have done the Urey-Miller experiment. This has shown us the origination of life. How life is originated could be explained by the Urey-Miller experiment. Seymour Benzer and Max Delbruck work independently on phage Genetics and mutation.

(Refer Slide Time: 58:57)