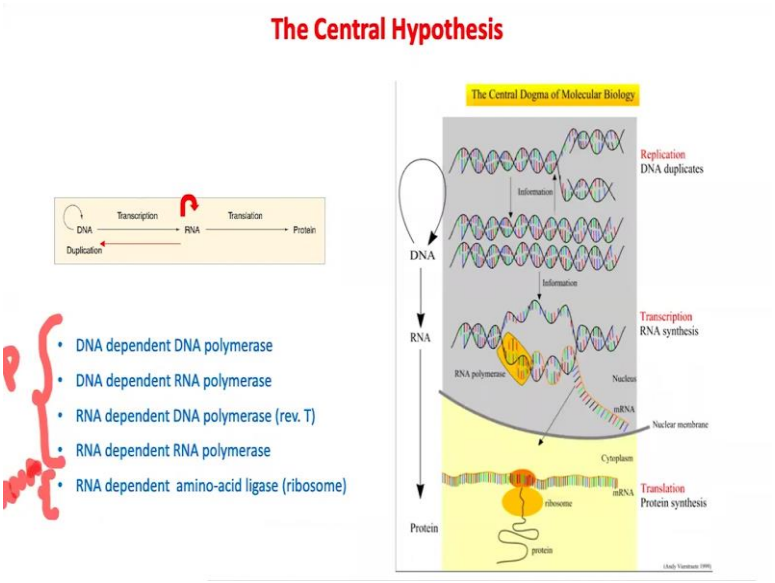
Introduction to Cell Biology
Professor Girish Ratnaparkhi and Professor Nagaraj Balasubramaniam
Department of Biology
Indian Institute of Science Education and Research, Pune
Central Dogma: The DNA structure Part 1

(Refer Slide Time: 0:16)



Today, I basically continue with where I left off. And as I told you in the last class that I am going to make the central dogma, which is actually a central hypothesis as a sort of a keystone in my lectures. So, I started with talking to you about the central dogma and using not only the idea of information transfer from a storage entity, all the way to a execute entity which is protein with an intermediate and between, which is RNA. I will take you to different molecules, I will explain to you initially proteins, and then I will move on to lipids, carbohydrates, and then move on to genetics, once the biochemistry part is over.

Now, in front of you is again, the flowchart of the central dogma with DNA being duplicated, which is what happens when one cell has to make two cells. So, DNA is first duplicated and split equally between two cells. There is a process of transcription, which is nothing but the process of copying the information on DNA into RNA. Now, as you will ask, those of you who have done biology will know and the others will realize that the copying of DNA to RNA is completely faithful, well, mostly faithful in the last few years, we have realized it is not completely faithful.

But the information in RNA is actually modified, and edited, which is a process of splicing. And then the spliced edited RNA is then translated into a completely different molecule,

which is also a linear molecule called a protein. So, duplication, transcription, translation, protein folding, these are the areas I will be covering in the next few classes.

Now, what I also told you is, as you know that DNA is the molecule of life, I touched upon the discovery or the understanding of the structure, which is basically at the molecular level. And, what I will also do is talk about scales in terms of size scales, which functional biology is happening at the molecular level, and I will get to that in a few slides.

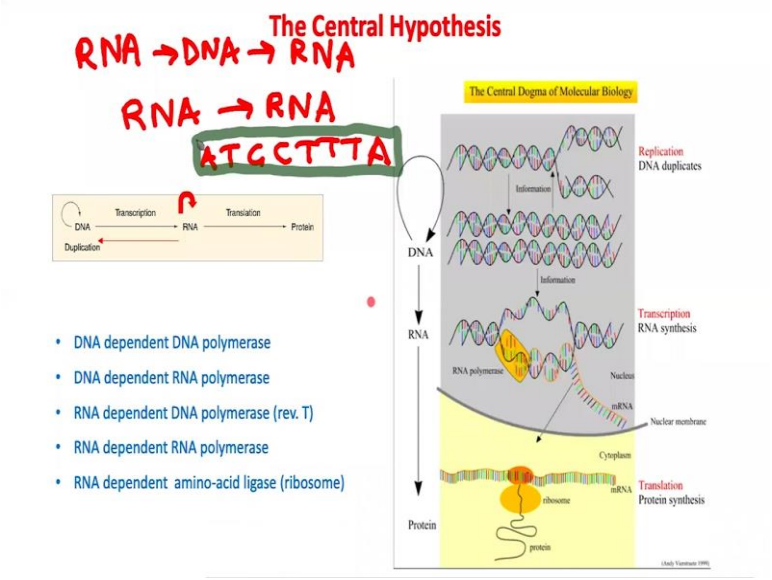
But more importantly, DNA is managed, manipulated and the process of transferring of information from DNA to protein is done by a series of molecular machines. Now, these molecular machines, for example, for DNA replication, is DNA dependent DNA polymerase, so I am giving it its full name, rather than the short form which most of us use, which is DNA polymerase.

Then there is a DNA dependent RNA polymerase, which is basically doing the job of transcribing, which is making a copy of DNA on the genome to an RNA intermediate. Genomic DNA stays inside the nucleus, in sitting on chromosomes. And the RNA then comes out of the nucleus in a eukaryotic cell goes to the cytoplasm. In a prokaryotic cell, of course, there is no compartment between nucleus and cytoplasm.

And once it goes to the cytoplasm, it gets translated. Now, sometimes as shown by the red arrow, this was a discovery in the 60s and the 70s, that RNA itself can be copied back to DNA. And this is done by a protein or a molecular machine as I keep on calling all of these called reverse transcriptase. But its full name is RNA dependent DNA polymerase.

And RNA can also copy itself. And this happens quite a lot, for example, in viruses, and this would be done by a protein or a molecular machine called RNA dependent RNA polymerase. And finally, the process of translation, which is making of a protein from RNA is I have just given it a long name, we call it a ribosome. But it is basically a RNA dependent amino acid ligase. So different amino acids are stitched together. And the linear sequence of the stitching of these amino acids depends on the sequence of the RNA. Now, one of the things which I will not really talk about too much, but we will get to at some point is that all of these are basically proteins. So, the enzymatic activity of all these 1, 2, 3, 4 molecular machines are based on proteins, but this last molecular machine ribosome, turns out not to be a protein enzyme, it is actually an RNA enzyme. And there are implications for this which I will talk about at some later point. So, let us move on.

(Refer Slide Time: 5:01)



Student: Sir, where is this RNA dependent RNA polymerase used, the fourth point?

Professor: So, the RNA dependent RNA polymerase is mostly used by viruses. So, many viruses have RNA for a genome rather than DNA for a genome, and when they replicate, even if it is inside a eukaryotic or prokaryotic cell, they use this particular enzyme to duplicate its DNA.

Student: Okay, sir.

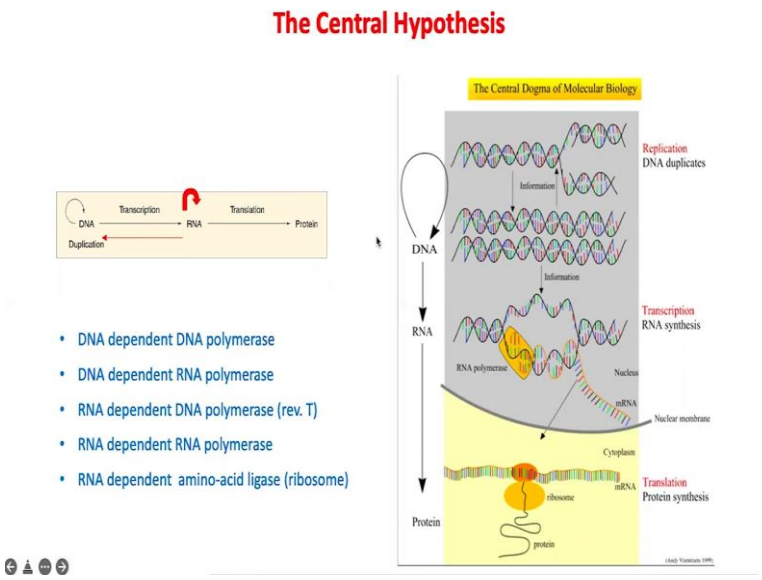
Professor: So, since you have asked this question, things are actually very simple, you just have to think simply, if RNA genome wants to duplicate itself, it can go to a DNA intermediate. And then it can make RNA again. Now, this will only work if there is no editing of information. Otherwise, if information is edited, during any step, then you will lose information. But RNA can directly make RNA, it does not need to go through a DNA intermediate. And many viruses do this also.

Student: Sir, the red arrow on RNA is describing this process.

Professor: That red arrow is basically describing the flow of information. But you literally are making a new molecule with a sequence. So, the sequence for example, in DNA is A, T, G, C, T, T, A, A, so you are just copying it. And that copying is what we are talking about. And I am emphasizing not on the chemical synthesis of putting nucleotides together.

I am emphasizing on the fact that this sequence over here has is coding for very, very important information. And that information is pretty much what makes up all life. So, it is just that I am emphasizing it in a different way, then you would probably have heard in tenth standard or twelfth standard, is that clear?

(Refer Slide Time: 6:54)



Student: Sir, in which cases would reverse transcription happen?

Professor: I could not hear you clearly.

Student: Which cases would the third point happen?

Professor: You are talking about reverse transcriptase?

Student: Yes.

Professor: So, reverse transcriptase is used, for example, by the HIV virus. And many of the inhibitors you hear about, taken by people have AIDS are reverse transcriptase inhibitors. And that is just one example. So, all these processes are going on very routinely. They are not particularly special.

Of course, DNA replication and transcription and protein synthesis are the bulk of these processes, but reverse transcription is also happening in your genome. I do not know if RNA dependent RNA polymerase is active in your genome because your genome as far as I know does not need it, but viruses definitely need.

Student: Sir, if RNA dependent RNA polymerase, can be used to duplicate RNA directly within one step, why do we need RNA dependent DNA polymerase?

Professor: Why do we need reverse transcriptase? Because see, think of this as two forms, see that these are two forms of nucleic acids. Think of it as Android and iPhone, this is the best analogy I can come up with. Now, it is a very weird analogy. It is not exactly a correct analogy. So, both are two different operating systems.

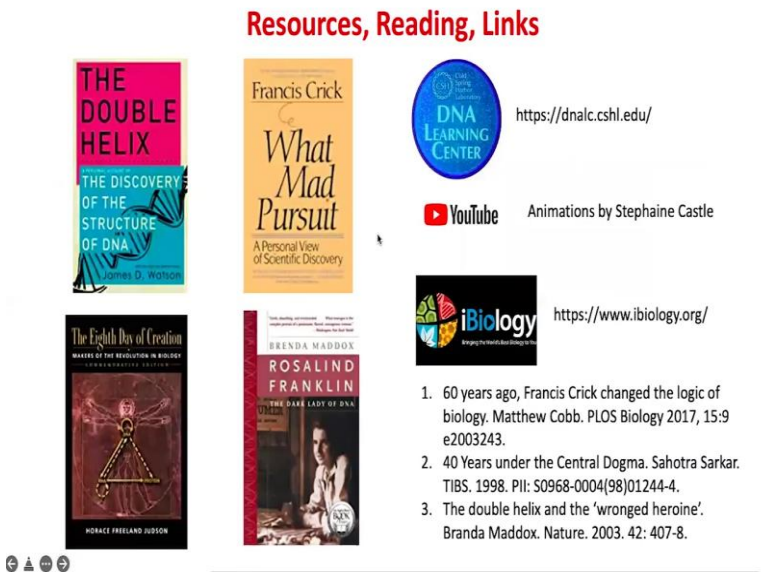
Some people say iPhones are let us say better at storage, but Android is much better at function. It is not true. But again, it is an imperfect analogy. And what we realize when we have studied these molecules now for over 100 years, is that RNA is very labile, in the sense that it is very easy to distort, whereas DNA is much more stable.

So, you literally have two forms of information. And over a couple of billion years of evolution, both forms have existed. So, they may be an alternate universe where there is only an RNA world, they may have an alternate universe where there is only a DNA world. But as of now, in life forms which are there on earth, both of these exist, both of them have their own pluses and minuses. And both of them are used. Both are nucleic acids.

Both are information storage devices. But long-term storage DNA seems to be better. RNA seems to be better for short-term storage. And also RNA, by the way, is a, many forms of RNA are enzymes, which is something you will realize as time goes by, so RNA is not only a storage of information molecule, it is also an executive molecule, but much of the executive functions in the cell are done by proteins. So, proteins are the action stars, RNA is partially action and DNA as far as we does not act. Was that answer reasonable enough?

Student: Yes, sir.

(Refer Slide Time: 10:00)



Professor: So, these are some of the resources and reading material, which I would recommend. Now, this is a very large class. And as you know and I have learned over the years, there is a mix of students from different sciences, some of you have not done biology. And the first point I always hear from some of you who have not done biology is that the terminology gets difficult. I understand completely terminology is a bit of a problem, biology is a whole new language. But it is not too difficult to learn, I would recommend that you apply your mind to it, make a small notebook with terminology written on it, you can cut and paste terminology and I am more than happy to help you along with the TAs to catch up as it would be.

Second thing is when I give you a resource like this to read, many of you are worried that unless you read these books, you are not going to pass in the biology exam. This is Jim Watson's book, which he wrote, which caused a controversy about double helix and Francis Crick's Crick's book, I do not mean that these are things which you have to read.

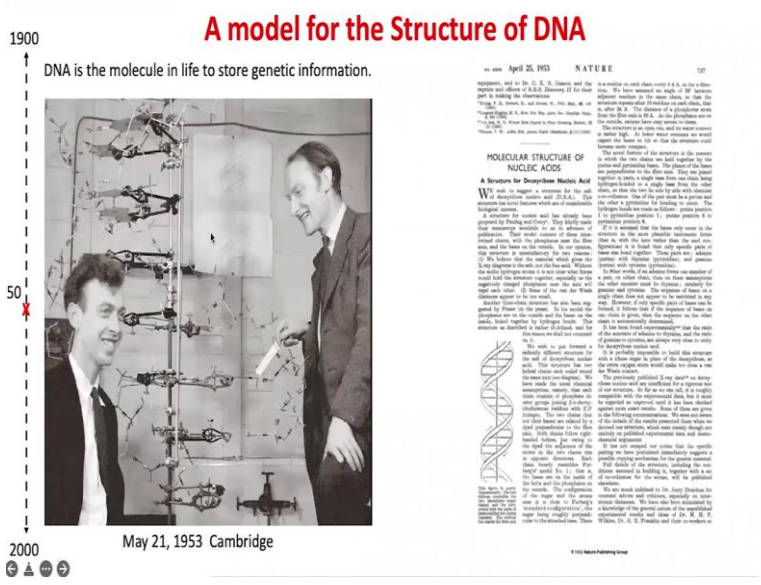
These are you can read this, if you are interested, you can be a student interested in mathematics, and yet read a general book about Jim Watson and his personal view of the discovery of the structure of DNA because it is basically English. It is like any other book. It is interesting, because you may be a chemist, you may be a mathematician, you may go into Earth Sciences. But DNA literally is a very important molecule.

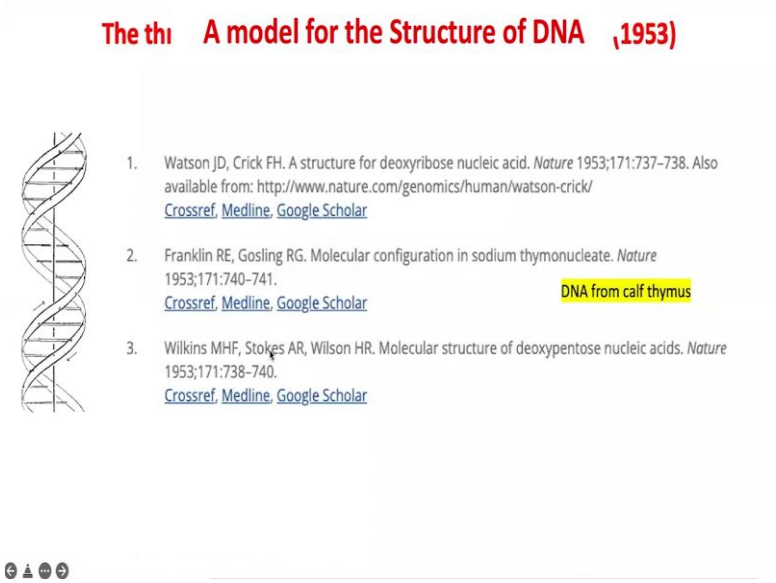
DNA is about life. So, there is no real reason for you to say, this is not this is beyond what my interests are, and I am not going to read about. So, these are resources for you to expand your horizons, not mandatory resources for you to read. I would recommend for those of you who like historical stories, to read The Eighth Day of Creation. This is where much of the material I am telling you comes from.

There is a book by Brenda Maddox on Rosalind Franklin, and I will talk about Rosalind Franklin a little bit. I also recommend that you go to the DNA Learning Center at Cold Spring Harbor and read about they have lots of resources, a lot of them are meant for school children in the six to eight standard.

So, I am pretty sure all of you will be able to appreciate it. I am using animations by Stephanie Castle. And as you well know, there is the Khan Academy, there are all kinds of videos on DNA, RNA, Central Dogma out there. I like these animations, because they are simple. And I would recommend them.

(Refer Slide Time: 12:30)



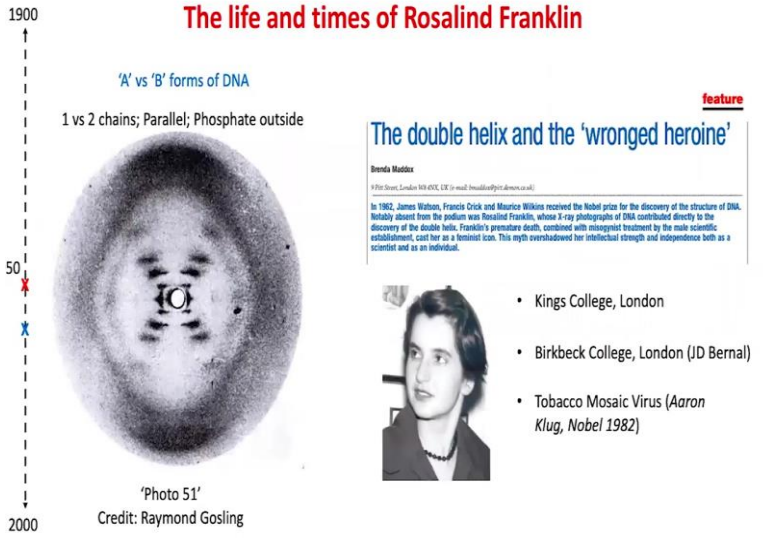


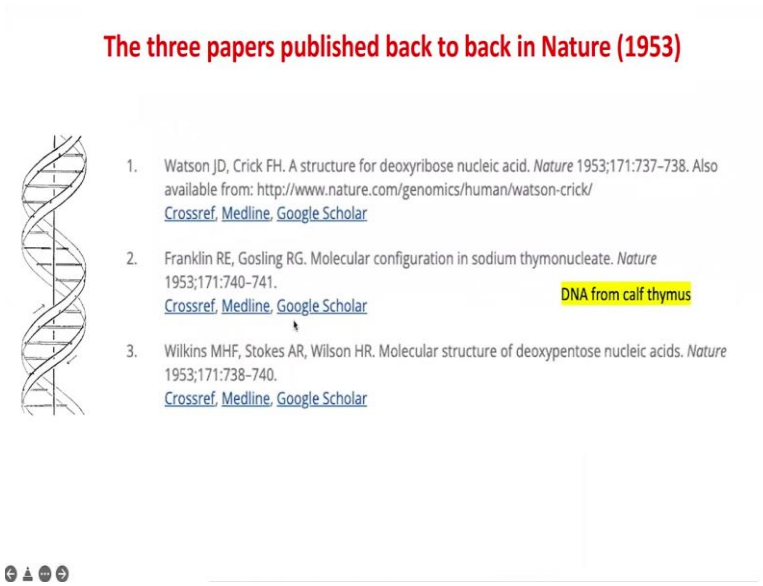
So, back to the science and the history of this very exciting molecule called DNA and as you know and this is a timeline I will be using on the left throughout my class, around 1953, which is halfway through the last century, Watson and Crick modeled DNA. They were not experimentalists, they never touched DNA. But by gathering information, thinking and building these models, after a few failures, they managed to solve the structure of DNA.

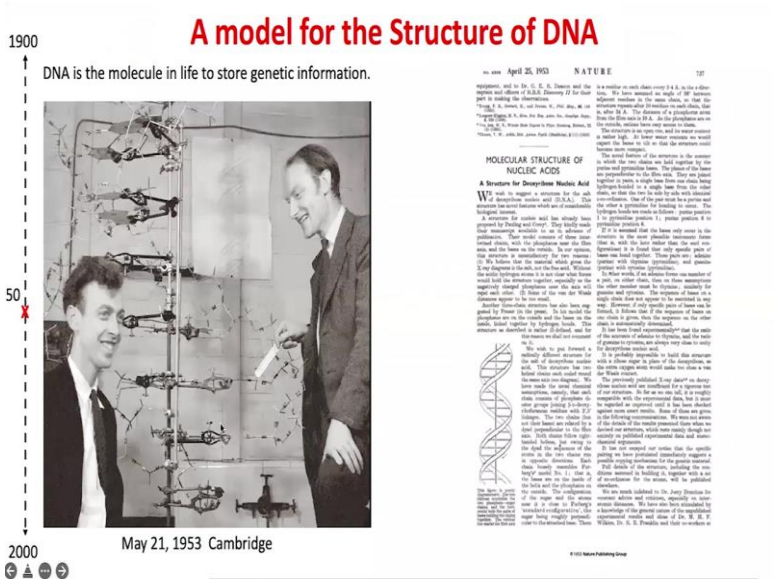
And they did all of this basically, after they met in Cambridge in the Cavendish Laboratory in a matter of two years, and the actual DNA model took them a little less than a month to finally make. Now, as I told you, this is a very simple paper very well written even today 70 years after the time this paper was written, it is readable.

One of you asked me yesterday what sodium thymonucleate was, and the answer I gave us, it was basically DNA. But I was curious about the terminology. So, I went back and read about it. And it turns out the only reason Rosalind Franklin called it sodium thymonucleate it was, of course, a sodium salt of DNA, but its origin was from calf thymus. And today, we know, it does not matter if it is from calf thymus, or its human DNA, or it is cockroach DNA. DNA is basically DNA. And as long as you have enough pure amounts of DNA, you can kind of work with it.

(Refer Slide Time: 14:03)







So, I am going to now basically stop the first part of the structure of DNA and the history of DNA by talking about Rosalind Franklin, and her role in our understanding of what DNA structure looks like. Now, let me remind you again that the model was solved by Watson and Crick, who were at the Cavendish Laboratory in Cambridge.

They went and talked to Maurice Wilkins, Maurice Wilkins was at King's College in London, Maurice Wilkins showed him his data and he also showed some of the pictures which Rosalind Franklin had taken in a lab and armed with all of this information, Watson and Crick fiddled around with their model, which is shown over here.

And they finally got a model which in their mind, fit all the available data. At the same time, this is something I told you in the last class, Linus Pauling who was one of the greatest chemists in the last few 100 years, came up with the idea of a triple standard helical structure. And he had just sent his paper and it was accepted in nature.

Both Watson and Crick had that paper in their hand, they had requested Linus Pauling to send them a preprint. They realize that Pauling structure was in all probability wrong. They came up with their own structure, and published it in nature. They talked to Rosalind Franklin, they talked to Maurice Wilkins, all three of them together, put their papers back-to-back. And today we know after so many years that that structure was indeed sound.

Now, Rosalind Franklin, Rosalind Franklin's role in this is kind of very interesting. She was trained as a crystallographer. And before she joined King's College London, working

broadly, in the big group of Maurice Wilkins, she had done much of her crystallography work, not on biological material, but on non-biological material, like coal, for example.

And one of the reasons why Watson and Crick made their breakthrough and Watson writes about this in his book, which is what caused the big controversy when he wrote his book was this very pretty picture of the fiber diffraction of DNA. Now, this picture was published in the paper, you will see it in Rosalind Franklin's paper.

And it was one of the long series of photographs which Rosalind Franklin and Gosling took together, this particular photo became very famous, it is a historical artifact, it was called Photo 51. And it turned out that Rosalind Franklin figured out that there was a lot of variation in the pictures she was getting with her graduate student who was Raymond Gosling.

And after a year or so, both of them realized that they were not dealing with one form of DNA, but two forms of DNA. And they call these as A versus B forms of DNA and instead of x and y, they said A and B. And this pretty much was the B form of DNA. And this was one of the clearest pictures which came out. It was this picture which Watson saw, which triggered his sudden understanding of what could possibly the model be.

And in his book, he actually says that if it was not for this picture, photo 51 he would not along with Crick been able to make a correct model of DNA, which we know today as the B form of DNA. Now, Rosalind Franklin herself was not very happy at King's College London, and it had to do more with the fact that she did not get along with Maurice Wilkins. All of this is in the books, which I have asked you to read.

And she did not even know that this data was actually shown to Watson and Crick. And at that point in time, she herself was thinking about the model of DNA. She had also made some small models, but she had not come to the conclusion which Watson and Crick did. She was still not very sure whether there were single chain or there was two chain or two chains.

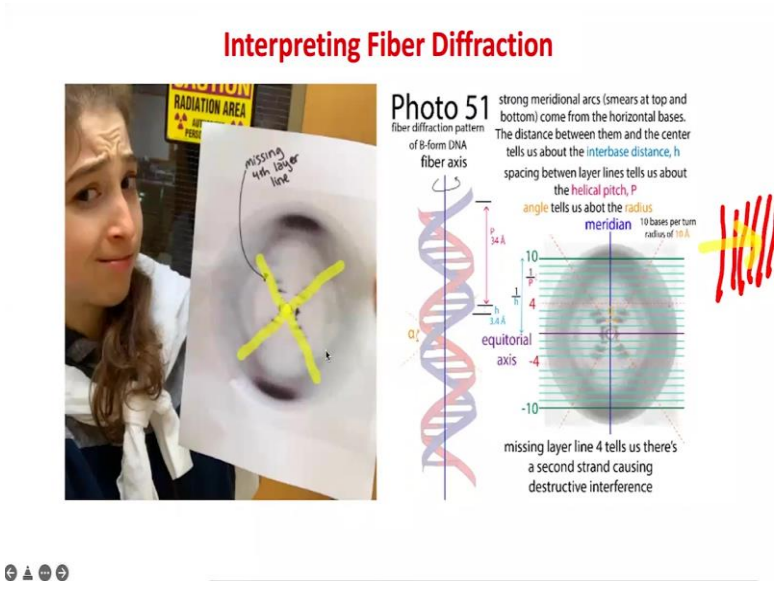
She really never thought about an antiparallel double helix. She was thinking pretty much always of a parallel double helix, if at all, there was a double helix. And she did know that the model probably had phosphate on the outside, which many of the previous models did not really contain. So, she had some inkling, she had excellent data.

But she could not make the final connection to the correct form, which was sort of a giant leap in terms of thinking about the model, which Watson and Crick could do so because Rosalind Franklin was not happy. She actually left King's College it was more to do with the place, and her interest in DNA.

And her interests actually moved to tobacco mosaic virus. She nearly really never went back to DNA. That was a project which was handed over to her to complete in King's College. She moved to Birkbeck College London, where she worked with JD Bernal, who was again, very famous crystallographer. And, in fact, one of her colleagues and somebody she trained Aaron Klug, won the Nobel in 1982 for tobacco mosaic virus.

Rosalind Franklin, of course, never really won the Nobel Prize, though she deserved it. And the reason was simply that she had cancer at the time she moved to Birkbeck College London and in about four to five years by 1960 she passed away and the Nobel was actually given in 1962. Two years too late for Rosalind Franklin.

(Refer Slide Time: 19:42)



Now, some of you had asked me questions about what fiber diffraction is and what crystallography is, I have actually never personally done fiber diffraction, but I have found some information which I will share with you. So, basically, if you take fibers of DNA, they are all stacked together, imagine, and I will try and draw it out, imagine DNA fibers all stacked together like this in a regular arrangement.

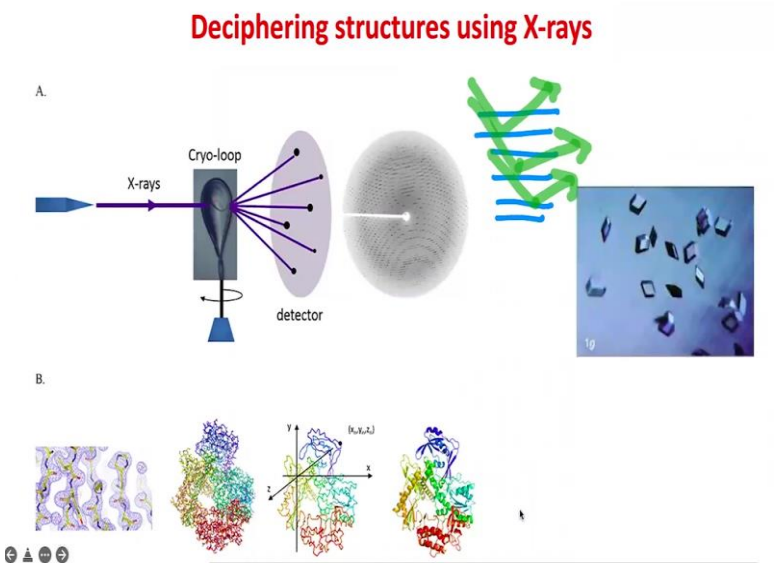
And then you can take a X-ray beam and point it towards these fibers. And what you see is a diffraction pattern of the way which is shown in photo 51 and these photographs, now, these are not very straightforward to interpret, I will just highlight a few facts. Now, effectively, one can figure out certain, let us say architectural features of DNA from these pictures, you cannot see the model of DNA at all, you cannot see things at atomic resolution.

But you can tell something about what the breadth of the helix is or the fiber is in this case, what the angle between these two strands is, how much is the gap between the nucleotides. So, all these features can be measured by just measuring distances, which are there between these spots. And one thing which is simply highlighted is if you count the number of spots in these in this x plane over here.

If you count them, you will realize that there are 1, 2, 3, and the fourth spot is missing. And there is a fifth spot over here. So, it turns out that the missing fourth spot, in fact, seems to say something fairly strong about the fact that there is a that it is a two stranded structure rather than a single stranded structure. So, again, this is not particularly useful.

But just to give you an idea that the pattern, which you see on an X-ray film, this was done on X-ray films in those times. And the distances between the spots, allows you to figure out some architectural distances between the different between the length and the breadth, and the distance between bases of DNA, but it does not give you a sort of a molecular picture of DNA at all. So, you have to make a model. And the model distances, which you make should approximately agree with the distances in the spots which you see in the picture.

(Refer Slide Time: 22:12)



So, this is now a more traditional X-ray diffraction picture. And here also, you can see that they are spots, and they are in a very symmetrical manner. And unlike fiber diffraction, where you are working with fibrous things like collagen or DNA, this will be a this is a protein crystal. And if you purify a protein and you make a very high concentration of protein, under the right conditions, you can get that protein to crystallize.

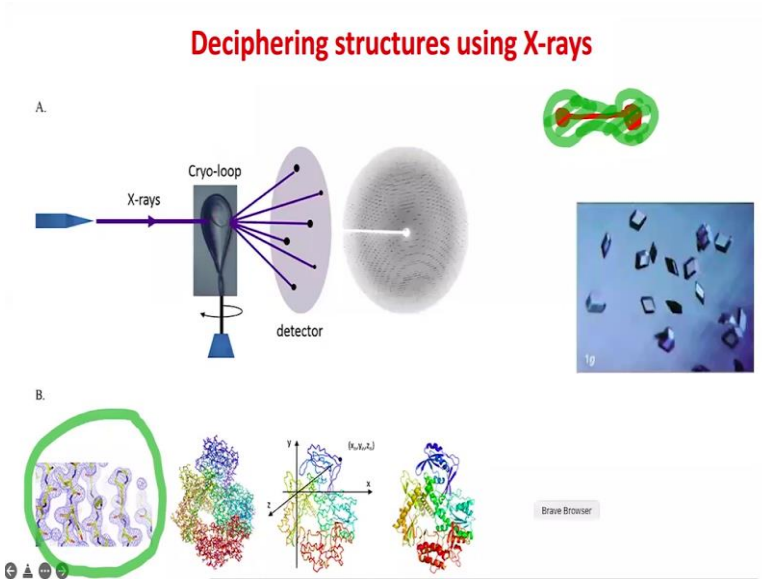
And this is what you see, these crystals are about a tenth of a millimeter or a twentieth of a millimeter, these are very small crystals. Now, you can pick these crystals up without damaging them, put them on a sort of a stand, and this stand will rotate 360 degrees, and you rotate, let us say every 10 minutes, you rotate by a degree, then you rotate by a degree again, you rotate by a degree again.

And at each rotation, you pass X-rays through the crystal and you take a snapshot. In the old days, this was done the detector was an X-ray film, but now we have electronic detectors with these X rays. And what you see is you see all these spots, and if you keep on rotating the crystal and you keep on taking snapshots, the position of these spots, the symmetry of these spots keep on changing.

And now the question is, why are you getting these spots, these spots are coming because the X-rays when they go through a crystalline lattice, which is a regular ordered arrangement, they see sort of planes of ordered arrangements and I will try and draw this I am sure it would not be a great so they see planes like this.

And effectively what ends up happening is that these X-rays, they bounce off these planes. So, there is interference created by these planes. And this can only happen if the proteins which are inside the crystal are organized in a crystalline array. And what you end up getting is these spots which are nothing but reflections of the crystalline lattice of the crystal. And these spots contain all the information you need minus one very important thing which I will talk about to calculate what is known as the electron density map.

(Refer Slide Time: 24:35)



You can take these spots, and for each spot, the intensity of the spot tells you something, the intensity of all the spots together leads you what is shown on the left-hand side over here, which is basically an electron density map. So, for example, if you have let me try and draw this, if you have a carbon atom over here, and you have a nitrogen atom over here, and there is a bond in between the electron density map which you draw will look something like this.

And you basically have to fit in this electron density map using a model, you still have to model the atoms of any molecule whose structure you are solving. The thing missing, which I talked about is that the X-ray experiment makes you lose what is called as the phase of each spot, you have the intensity and there is a mathematical equation called as a Fourier transform, which can be used to take these spots and calculate an electron density map, which finally leads you to the calculation of structure, shown the bottom over here is basically a protein structure solved using X-ray crystallography. But what is missing is the phase for each spot and that is called as the phase problem in crystallography.

And those of you who have a physics background, I think, many times physicists get extremely attracted to biology, simply because structural biology involves a lot of physics. And it is, at least in my opinion, much more fun to do. So, that is pretty much sort of a summary of how X-rays and crystals are used to solve structures of biomolecules.