Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Module - 03
Basics of Biomolecules
Lecture-12 Biomolecules (Part 2: DNA Sequencing)

Hello everyone, this is Dr. Vishal Trivedi from Department of Biosciences and Bioengineering, IIT Guwahati. you are purifying the human genome or if you are purifying the human genome, it should not be the case that you actually isolated something else, right. And that information only you will get when you are actually going to do the sequencing of this particular DNA, right. And that is very, very relevant when you are actually working with us, you know, the recombinant clones, right. For example, you generated a recombinant clones, then it is important that you verify that DNA with the help of the sequencing reactions.

So first thing is how you can be able to determine the purity of the DNA, right. So the purity of the DNA can be determined by spectrophotometer. So you can you know that the DNA and RNA absorbs at 260 nanometer, whereas the protein absorbed at 280 nanometer. But if you see the curve, right.

So if you see the curve, if you what you will see here is that it 260 the DNA and RNA are going to absorb, but they also absorb at 280 nanometer, right. So this is for DNA and RNA. Whereas for the proteins, it is actually going to be like this. So this is for the protein, right. Now what you see here is that this is showing a lambda max for the nucleic acid, right, for the nucleic acid, whereas this is the lambda max for the protein.

But it does not mean that the protein is not contaminating or not, you know, having any absorbent at 260 nanometer. And that is why it is important that we should calculate a ratio of 260 to 280 ratio. So what you can do is you take the absorbance not only at 260 nanometer, but at 280 nanometer. And that is actually going to tell you the purity of your sample. For example, if you calculate the 260 to 280 ratio, and if it comes at 1.

8, right, then it is a pure DNA, right. If 260 to 280 ratio is less than 1.8, then there will be a contamination of the protein, which means this component is now increasing and this component remains the same, right. So, this means it is actually going to, the ratio will actually going to lower down, right. This means, see, what will be the, if it is a 50-50 contribution, right, then the ratio is going to be 1, right.

Because then the absorbent at 260 and 280 are actually going to be equal, right. So that

time the contamination would be 50%. But if it is less than 1.8, then there will be a contamination of protein. Now if 260 to 80 ratio is more than 2, right, which means you are actually going to have very high absorbance at 260 nanometer, that means there will be a contamination of RNA into the DNA pair.

Now this is very important to understand. You know that the DNA is double standard, which means you have two strands and the bases are inside. So, whereas the RNA is mostly been single standard, right, which means this is the RNA with the bases. Now, if you see the very carefully, the bases within the DNA are being protected within the DNA structure, right. And because they are not exposed to the outer environment, because one base is, you know, put next to each other, they are actually going to show the lesser quantum yield and lesser excitation to the light, okay.

And because of that, they are actually going to show you the lower absorbance compared to the RNA molecule because RNA the bases are exposed to the water and exposed to the outer environment. And because of that, they will actually going to make the absorbance more, right. And because of that, the RNA is actually going to show you the more absorbance compared to the DNA. So, even if the DNA is pure and if there is a RNA contamination, the RNA is actually going to show you the 260 reading more. So, if the 260 reading is more and 280 is same or equal, then the ratio will actually above 2, right and that is how it is actually going to give you the indication that there is a contamination of RNA species.

Now, this is all about the DNA purity, right. Now, the second point is about the sequencing. So, DNA sequencing, right. So, DNA sequencing historically, there are two methods of DNA sequencing with a similar principle of breaking the DNA into the small fragment followed by the separation and analyse them on a high resolution electrophoresis gel. So, if you want to sequence any biomolecule, you are supposed to do this, right.

For example, if this is the DNA, what we want to sequence, what we can do is we can just split this into small, small, small, small fraction, right. And then we can actually be able to sequence. So, it is actually the same rule as divide, right and sequence. Because it is easy to then manage these small fragments, you cannot actually manage a 3 KB DNA, but you can easily manage 100 base pair DNA because then it is easy to manage, right. So, what you can do is you amplify this particular sequence with a modified base.

So, when you amplify this with a modified base, wherever you are actually going to have the modified base, it is actually going to break, right. And that is how it is actually going to give you the small fragment. And you know where it is actually going to break.

For example, if I take the modified base for A, it is going to break here. It is going to break.

Yeah. And if I take the modified base like G, then it is going to break here. It is going to break here and something like that. And then I can analyze these sizes of the DNA on a high resolution electroverse gel and that is how it is actually going to give me the complete sequence. And if you follow this method where you are going to use the modified base, then this sequencing method is called as Sanger sequencing method. And for this sequencing method, the Sanger got the Nobel Prize.

Now other option is that you use the chemical reagents, right, you use the chemical reagents which are actually going to attack the bases. So, you can have the reactions for A, you can have the reaction for T, you can have the reaction for G, you can have the reaction for C. So, what will happen is it is going to break after A, T, G, C like that. And that is how you can actually be able to separate these fragments and that is how you are actually going to get the information about the sequencing. And if you use that method, then it is actually going to be called as Maxim-Glibert method.

So, let us first discuss the Sanger sequencing method and then we are actually going to discuss about the Maxim-Glibert method. So the Dideoxy chain termination method or the Sanger's method. This method is originally been developed by the Frederick Sanger in the year of 1977. In this method, a single standard DNA is used as a template to synthesize the complementary copy with the help of a polymerase in the presence of T nucleotides. The polymerase reaction contains a primer and the nucleotides, three normal nucleotides and a 2 prime, 3 prime dideoxy nucleotide triphosphate which is a modified nucleotides.

When the DNA polymerase utilizes the DDNTPs as nucleotides, it gets incorporated into the growing chain, but chain elongation stop at the dideoxy as the dideoxy due to the absence of 3 prime hydroxyl group. In the typical sequencing reaction, you are going to run the four different DDNPs are taken into the four separate reaction and analyze on the higher resolution polyacrylamide gels, the ratio of NTPs and the NTPs is adjusted so that the chain termination occurs at each position of the bases in the template. So when you do the dideoxy chain termination method, you can actually have this right. So this for example, this is the region which you want to sequence right. So you're actually going to have the primer.

So in the step one, a primer is added and annealed to the 3 prime of the DNA helix right in a template. In the step two, the radio labeled ATP is used to label the primer. So you are actually going to label the primer so that you know what will be the fragment. So you

can actually be able to identify this fragment onto the autoradiograms. Then the step three, the polymerase reaction is divided into the four reactions.

So you can have the four reactions, you can have the A reaction, you can have G reaction, you can have C reaction and you can have D reactions. So in the A reaction what you have, you have the A DD ATP actually, so dideoxy ATP. In the G reaction, you're going to have DD GTP. In the C reactions, you're going to have DD CTP right. And in the T reactions, you're going to have DD TTP right.

This is what it is showing here. So you can have the A reactions, you can have the D reactions, you can have G reactions, you can have C reactions and all the four reactions, you are actually going to be load onto the sequencing gel and then you are actually going to analyze them with the help of the rate autoradiogram. So in the step four, the DNA synthesis continue until the terminated by the incorporation of the specific dideoxy nucleotides right. Because the dideoxy nucleotides does not contain three prime hydroxyl groups. So it will actually going to terminate the chain elongations.

Then you are going to a chase of the polymerase reaction is performed in the presence of high concentration of NTPs to obtain the all non terminated sequence into the high molecular weight DNA. This high molecular weight sequence will not enter into the sequencing gel. So because the pore size what you are going to adjust in of the sequencing gel in such a way that these high molecular weight DNA is not going to enter because this high molecular weight DNA will not going to provide you any information about determinations and they are actually going to make the analysis more complex. So after this, you are going to have the four reactions, you can have the A reactions, you can have D reactions, you can have G reactions, then you can have C reactions. So the way it goes that you are actually going to have the A reactions right, then you can actually have the T reactions.

So from the A you are actually going to have the T right and from the T you are going to have the T. So you are actually going to read from the from the bottom okay. So for example, you are going to have A, T, T, A, G, then you are going to have A right, then you are going to have C, then you are going to like that. So if you go like this right, you go like this, then you are going like this, you go like this. So you have to read in the reverse direction.

So the smallest one you are going to put first. Then last you are going to do like this, third one like this, fourth one, fifth one, sixth one, seventh one, eighth one like that. Like you have to go from the bottom and you have to keep reading and keep putting the sequences like this and ultimately you are going to get the sequence of the DNA. Now

let us talk about the Max and Gilbert method. So the Max and Gilbert method actually relies on the chemistry part right.

So it is actually going to utilize the different types of reagents which are actually going to be basis specific. So you can have the A reaction, G reaction, C reaction, T reaction and so on. And that is how it is actually going to do the same thing what we are doing, what the Sanger has done with the help of the enzyme. But here you are actually using the different types of chemical reagents. So this method was discovered by the Max and Gilbert in the year of 1977 which is based on the chemical modification and subsequent cleavage.

In this method a 3 prime or 5 prime radionuclide DNA is treated with a basis specific chemical which is randomly cleaved the DNA at their specific target nucleotides. These fragments are analyzed on a high-resolution polyclomer gel and the autoradiogram is developed. The fragment with the terminal redolibut appear as a band in the gel. So the chemical reaction what are going to be performed in 2 steps. First you are going to have the basis specific reaction and the second step you are going to have the cleavage reaction.

So the basis specific reactions, first you are going to have the basis specific reaction. So different basis specific reactions are used to modify the target nucleotide. So reaction 1 you are going to have the dimethyl sulfate or DMS which is actually going to modify the N7 of the guanine and then open the ring between C8 and C9. This is going to be called as G reactions. Then in the reaction 2 you are going to use the formic acid and act on the purine nucleotides.

So it is actually going to act on G and A by attacking on to the glycosidic bond. Then you have a reaction 3 which is going where you are going to have the hydrazine and which is actually going to break the rings of the pyrimidines. So it is actually going to be not specific for a particular base but it is only going to be specific for pyrimidines. So it is going to be called as T plus C reactions. Then you are going to have reaction 4 where in the presence of salt it breaks the ring of the cytosine.

So it is going to called as C reaction. So basically you are going to have 4 reactions. One is called G reaction, other one is called as A plus G reactions. You are going to have C plus T reactions and you are going to have C reactions. So you are going to take the radiolabel DNA, you are going to add these reagents and that is how you are going to have the G reactions, A plus G reactions, C plus T reaction and C reactions. And then you are going to have the cleavage reactions.

So after the base reactions pyrimidine is added which will replace the modified bases and catalyze the cleavage of phosphodiester bond next to the modified nickel. So once you have done the reaction after the G, after the G, after G and C, after C it is going to be cleaved and that is how you are going to have the fragments. So the fragment, so you are going to have the G reactions, G plus A reactions, C plus T reaction and C reactions. So imagine that if this is the DNA sequence which you want to sequence.

And here also exactly the same way. You have to go in the reverse direction, but remember that when you have the bond band between G and G plus A then it is actually going to read as G. So it is for example here you have the 2 bands. One is in the G reaction, the other one is in the G plus A reaction. So it is not going to be read as A, it is actually going to be read as G.

So from here this is going to be read as G. Then you have this, so you actually will go in the reverse direction. So the fragment in J line is read as G whereas the fragment in the G plus A but absent in G is read as A. Similarly fragment in C is read as C whereas the fragment present in T plus C but absent in C is read as T. Same is true for this C. If the 2 fragments are present in same distance, of same size then it is actually going to read as C rather than T actually.

But if the fragment is absent in C but it is present in T then it is actually going to read as T. For example this one. So this one the fragment is absent in G but it is present in G plus A. So this is actually going to be read as A.

Same is true here. For example this is the T plus C. So this is actually going to be read as T and that is how if you go from the reverse direction you are actually going to be reduce the sequence at the end and that is what it is actually going to be the sequence of that particular DNA. So the way you are going to read this sequence is that you are actually going to read the lowest with the band and then you are actually going to go to the higher bands. So this is all about the assessment of the DNA quality. So what we have discussed, we have discussed about the purity of the DNA and we have also discussed about the DNA sequencing. While we were discussing about the purity of DNA we took the help of the spectrophotometer and when we were discussing about the DNA sequencing we discussed about the Sanger's method and we have also discussed about the Max and Gilbert method.

So with this I would like to conclude my lecture here. Our subsequent lecture we are going to discuss some more aspects related to the biomolecules which are relevant for the molecular biology. Thank you.