Cell and Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Week 09
Central Dogma of Molecular Biology (Part 1)
Lecture - 32
Structure of DNA (Part 2)

Hello, everyone. This is Dr. Vishal Trivedi from the Department of Biosciences and Bioengineering at IIT Guwahati. So far, what we have discussed are the basic structures and basic properties of nucleic acids. So what we have discussed is how the nucleic acid is composed of the sugar, base, and phosphate groups. And within the base, you have purines and pyrimidines.

Purine always pairs with pyrimidine, and they are always present in definite compositions; the DNA strands run in an anti-parallel fashion to each other, so if the first strand runs from the five prime to three prime, the second strand runs from the three prime to five prime. And at the end, we have also discussed how you can isolate the nucleic acids or the genomic DNA from the cells. And at the end, we have also discussed how you can determine the concentration of the isolated DNA with the help of the absorbance at 260 nanometers or the colorimetric estimation using the difflamine reagents. So in today's lecture, we are going to discuss how you can determine the quality of this DNA, right? So, you have isolated the genomic DNA, haven't you? Or you isolated a plasmid, or you isolated any DNA, but how will you know that it is actually high-quality DNA? So if you want to do the quality assessment, you can actually assess the quality of any biomolecule, whether it is DNA, protein, RNA, or other substances.

Quality lies in two aspects: one aspect is the purity of the molecule, meaning how pure it is, and the other aspect is the sequence, which pertains to how authentic it is. Suppose you isolated DNA; then it is important that you know that. Apart from DNA, what are the other things you have purified while purifying cells? So, because when you are purifying from the cell, you are subtracting or removing the other biomolecules, right? You are only isolating the DNA but not the RNA, or you are not isolating the proteins along with it, or the lipids, right? So, purity analysis is actually going to tell you how pure your DNA is, how much protein contamination there is, what the contamination of RNA is, and what the composition of lipids is. These are the three major components that can actually contribute to DNA purity. Apart from that, you also want to know what the sequence of this DNA is, which you are actually going to sequence.

So the DNA sequence is important because if you are purifying the human genome, it should not be the case that you actually isolate something else, right? And that information will only be available 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, you know, the recombinant clones. Right? For example, if you generate recombinant clones, then it is important that you verify that DNA with the help of sequencing reactions. So the first thing is how you can determine. The purity of the DNA, right? So, the purity of the DNA can be determined by the spectrophotometric method. You know that DNA and RNA absorb at 260 nanometers, whereas proteins absorb at 280 nanometers.

But if you see the curve, right? So if you see the curve, what you will see here is that at 260, the DNA and RNA 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 shows a lambda max 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 does not have any absorbance at 260 nanometers, and that's why it is important that we calculate a ratio of 260 to 280. So what you can do is take the absorbance not only at 260 nanometers but also at 280 nanometers.

And that is actually going to tell you the purity of your sample. For example, if you calculate the 262 to 80 ratio and it comes out to be 1.8, then it is pure DNA, right? If the 260 to 280 ratio is less than 1.8, then there will be contamination of the protein, which means this component is now increasing while the other component remains the same, right? This means that the ratio is actually going to go down, right? This means to see what will happen if it is a 50/50 contribution; then the ratio is going to be 1, right? Because then the absorbance at 260 and 280 will actually be equal, right? At that time, the contamination would be 50 percent. But if it is less than 1.

8, there will be protein contamination. Now, if 260 to 280 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 Now, this is very important to understand. You know that DNA is double-stranded, which means it has two strands and the bases are inside. So, whereas the RNA is mostly single-stranded, this means it is RNA. With the bases now, if you look very carefully, the bases within the DNA are being protected within the DNA structure, right? Because they are not exposed to the outer environment, and because one base is placed next to another, they are actually going to show a lesser quantum yield and lesser excitation to the light.

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

So, is DNA sequencing correct? So, historically, there have been two methods of DNA sequencing with a similar principle of breaking the DNA into small fragments, followed by the separation and analysis of them on a high-resolution electrophoresis gel. So, if you want to sequence any biomolecule, you should do this right. For example, if this is the DNA that we want to sequence, what we can do is split this into a small production, right, and then we can actually sequence it. So, it is actually the same rule as division, right, and sequence. Because it is easy to manage these small fragments, You cannot actually, you know, manage a 3 KB DNA, but you can easily manage a 100 base pair DNA because it is easy to manage, right? So what you can do is amplify this particular sequence with a modified base.

So when you amplify this with a modified base, wherever you actually have the modified base, it is going to break, right? And that's 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. 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 DNA on a high-resolution electrophoresis gel, and that is how it will actually 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 Sanger sequencing.

And for this sequencing method, Sanger received the Nobel Prize. The other option is that you use the chemical reagents, isn't it? You use the chemical reagents that are actually going to attack the base. So you can have the reactions for A, T, G, and C. What will happen is that it will break after A, T, G, and C like that. And that's how you can actually separate these fragments.

And that's how you are actually going to get information about the sequence. And if you use that method, it is actually going to be called the maximum Gilbert method.

So, let us first discuss the Senghor sequencing method, and then we will actually discuss the maximum Gilbert method. So, the dideoxy chain termination method, or Sanger's method, is used. This method was originally developed by Frederick Sanger in 1977.

In this method, a single standard DNA molecule is used as a template to synthesize the complementary copy with the help of a polymerase in the presence of deoxynucleotides. The polymerization reaction contains primers and nucleotides. Three normal nucleotides and a two-prime, three-prime dideoxy nucleotide triphosphate, which is a modified nucleotide, are present. When the DNA polymerase utilizes the dNTPs as nucleotides, they get incorporated into the growing chain. But chain elongation stops at the dideoxy due to the absence of a three-prime hydroxyl group.

In the typical sequencing reaction, you will run the four different TTNPs, which are taken into the four separate reactions and analyzed on high-resolution polyacrylamide gels. The ratio of NTPs to dNTPs is adjusted so that 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, for example, this is the region that you want to sequence, right? So, you are actually going to have the primer? So in step one, a primer is added and annealed to the three-prime end of the DNA helix in a template. In step two, the radio labeled ATP is used to label the primers.

So you are actually going to label the primer so that you know what the fragment is. So you can actually identify this fragment on the autorhodograms. Then, in step three, the polymerase reaction is divided into four reactions. So you can have four reactions. You can have the A reaction, the G reaction, the C reaction, and the T reaction.

So in the A reaction, what do you have? You have the ADDATP; actually, so did deoxyATP. In the G reaction, you're going to have dGTP. In the C reactions, you're going to have DDCTP, right? And in the T reactions, you're going to have DDGTP. TTP, right? This is what it shows here. So, you can have the A reactions, the T reactions, the G reactions, the C reactions, and all four reactions are actually going to be loaded onto the sequencing gel.

And then you are actually going to analyze them with the help of the red autorhythmogram. So, in step 4, the DNA synthesis continues until it is terminated by the incorporation of the specific dideoxy nucleotides because the dideoxy

nucleotide does not contain a 3-prime hydroxyl group. So, it will actually terminate the chain elongation. Then you are going to chase the polymerization reaction performed in the presence of a high concentration of NTPs to extend all non-terminated sequences into high-molecular-weight DNA. This high molecular weight sequence will not enter the sequencing gel.

So, because the pore size is what you are going to adjust in the sequencing gel in such a way that this high molecular weight DNA does not enter, this high molecular weight DNA will not provide you with any information about determinations. And they are actually going to make the analysis more complex. So, in the Sanger sequencing method, what you have is target DNA that you have to sequence. So what we have, you know, if you want to do DNA sequencing using the Sanger sequencing method, you have two ways: either you can go with gel filtration chromatography and gel electrophoresis, or you can use capillary electrophoresis. So the first step is to take your DNA into the ampule, and then you have to add the primers.

You have to add these primers to the four reactions. If you remember, we said that you had to divide the reactions into four groups and then you had to add the DNA polymerase to each reaction. So, you have to add the reaction numbers 1, 2, 3, and 4. Once you add the DNA polymerase to the four reactions, you are going to add the nucleotides. You have to add all four nucleotides, such as C, T, A, and G, in all four reactions.

And in the subsequent step, you are going to add the dideoxynucleotides. So if you recall, you can have four different reactions: A reactions, D reactions, C reaction, and the G reaction. And in all of these, you have added the dideoxy nucleotide. What is the difference between normal and dideoxy nucleotides? The difference between a normal dNTP and dideoxynucleotides is that the former has a 5 prime phosphate and a hydroxyl group at the 3 prime, whereas the latter is missing this OH, which causes the synthesis to stop. Let us see how it is actually going to stop the synthesis.

So you can imagine that if there are dNTPs, they will actually form a bond through the phosphodiester linkage, and the OH is still there. So, that will continue the synthesis, whereas in the case of the dideoxy nucleotides, once the dideoxy nucleotide uses this phosphate and forms the phosphodiester linkage, since the OH is missing on this side, it will not allow the incoming nucleotide to bind. So, that is how it is actually going to stop the synthesis of DNA by DNA polymerase. Now what you have to do is take the four reactions and put them into the thermal cyclers where you have all the reagents. In the thermal cycler, you have different steps; for

the first step, you will perform the denaturation.

So in the denaturation step, you will increase the temperature of the thermal cycler, and once you increase the temperature of the reactions, the two strands of the DNA will detach. Then, you will add the primers, which are needed, and there will be an extension. But what will happen is that if there is a dideoxytherm in the NTPs, then it is actually going to terminate the reactions wherever the enzyme finds a dideoxy. So what you see here is that the termination is happening at every A. If it is going to find the A, it will give you the DNA of different reactions.

The same is going to happen for the dideoxy T reactions. So all the T, wherever you have it, is actually going to terminate the reactions. The same is true for the C reaction. So wherever you could find the C, it was actually going to terminate, like here; it was going to terminate, and so on. So the same is going to happen even for the G reactions; wherever they are found, the G is actually going to terminate.

So, for example, in this case, it found the G at the end. So it is going to terminate at this point; then it is going to terminate at this point, and that's how you see they are actually going to give you the different reactions. Now what you have to do is take these reactions out of the thermal cyclers and then resolve these samples on the gel electrophoresis. So you have to take all four reactions and load them into the four different wells, and you know that the DNA is negatively charged. So it is actually going to be resolved on the gel electrophoresis.

So you load the fourth reaction, then connect it to the power pack and turn it on. So when you turn on the DNA, it will run from the negative to the positive because the DNA is negatively charged. Because the DNA has a phosphate backbone, it actually gives a negative charge to the DNA. So, because of the negative charge, it goes toward the positive electrode in the gel electrophoresis. And you know that this migration is inversely proportional to the size of the DNA.

So the smaller DNA will run faster, and the larger DNA will run more slowly. Now what you have to do is, once you have resolved the DNA, you have two ways to visualize this DNA. You can either utilize the radio-labeled primers or use the labeled dNTPs. That labeling can be done with radioactivity. So, what you can do is use the P32-labeled DNA and labeled bases, which will actually label the DNA during the synthesis.

So irrespective of whether you use the labeled primer or the labeled ATPs, once you have the DNA resolved onto the agarose, all you have to do is visualize the DNA band

with the help of the autoradiogram. So what you have to do is take this gel, take the agarose gel, put it into the gel cassette, place the x-ray films in, close it, and let it expose overnight or for 72 hours. During that period, the radioactivity present on the gel will expose the X-ray film, and that's how you will get the bands of DNA. Now, as we discussed before, you have to read it in reverse orientation, which means you have to read it in this sequence first, then this sequence, then this sequence, then this sequence, then this sequence, and that's how you are going to get the sequences from each band.

What you have is your DNA sequence. So, what you have to do is take this sequence and these sequences, and that is actually going to be the DNA sequence that you are going to get from the Sanger sequence. So, this is all about the Sanger sequencing method that we discussed. So, after this, you are going to have the four reactions: you can have the A reactions, the T reactions, the G reactions, and then the C reactions. So the way it goes is that you are actually going to have the A reactions; then you can have the T reactions. So from the A, you are 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 bottom.

Okay, so for example, you're going to have ATTAG. Then you are going to have A, 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. You have to read in the reverse direction, so the smallest one should be put first, second to last, and you should do it like this: third one, fourth one, fifth one, sixth one, seventh one, eighth one, and so on. You have to go from the bottom and keep reading while putting the sequences like this.

And ultimately, you are going to get the DNA sequence. Now, let's talk about the Maxim-Gilbert method. So the Maxim Gilbert method actually relies on the chemistry part, right? So it is actually going to utilize different types of reagents, which are going to be basis specific. So you can have the A reaction, G reaction, C reaction, T reaction, and so on. And that's how it is actually going to do the same thing that Sanger did with the help of the enzyme. But here, you are actually using different types of chemical reagents.

So this method was discovered by Maxim and Wilbert in 1977, and it is based on chemical modification and subsequent cleavage. In this method, a 3-prime or 5-prime radionucleotide DNA is treated with a base-specific chemical, which randomly cleaves the DNA at its specific target nucleotides. These fragments are analyzed on a high-dilution polyacrylamide gel, and the autoradiogram is developed. The fragment

with the terminal radiolabel appears as a band on the gel. So, the chemical reaction is going to be performed in two steps.

First, you are going to have the base-specific reaction, and in the second step, you will have the cleavage reaction. So, based on specific reactions, first you are going to have the base-specific reaction. So, different base-specific reactions are used to modify the target nucleotides. In reaction one, you're going to have dimethyl sulfate (DMS), which is actually going to modify the N7 of guanine and then open the ring between C8 and C9; this is going to be called G reactions. Then, in reaction two, you are going to use formic acid and act on the purine nucleotide, so it is actually going to act on G and A by attacking the glycosidic bond.

Then you have a reaction 3, which is where you are going to have the hydrazine and which is actually going to break the rings of the pyrimidines. So, it is actually not going to be specific to a particular base, but it is only going to be specific to pyrimidines. So, it is going to be called T-plus-C reactions. Then you are going to have reaction 4, whereas in the presence of salt, it breaks the ring of cytosine.

So, it is going to be called the C Reaction. So, basically, you are going to have four reactions. One is called the G reaction, and the other is called the A plus G reaction. You are going to have C plus T reactions, and you are going to have the C reactions. So, you are going to take the radiolabeled DNA, 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 cleavage reactions. So, after the base reactions, piperidine is added, which replaces the modified bases and catalyzes the cleavage of the phosphodiester bond. next to the modified neater. So, once you have done the action after the G, after the G, after G and C, after C, it is going to be clicked, and that is how you are going to have the fragments. So, the fragment: you're going to have the G reactions, G plus A reactions, C plus T reactions, and C reactions.

So imagine that this is the DNA sequence you want to analyze. And here, also, in exactly the same way, you have to go in the reverse direction, but remember that when you have the band between G and G plus A, it is actually going to read as G, okay. So, for example, here you have the two bands: one is in the G reaction, and the other 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. From here, this is going to be read as G; then you have this.

So, you will actually go in the reverse direction. So, the fragment in JLN is read as G,

whereas the fragment in G plus A, but absent in G, is read as A. Similarly, the fragment in C is read as C, whereas the fragment present in T plus C, but absent in C, is read as T. The same is true for this C. If the two fragments are present at the same distance and of the same size, then they are actually going to read as C rather than T. But if the fragment is absent in C and present in T, then it is actually going to be read as T.

For example, this is one. So, in this case, the fragment is absent in G, but it is present in G plus A. Therefore, this is actually going to be read as A. The same is true here. For example, this is 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 deduce 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 band, and then you are actually going to go to the higher bands. So this is all about the assessment of DNA quality, right? So what we have discussed is the purity of the DNA, and we have also discussed the DNA sequencing. While we were discussing the purity of DNA, we took the help of the spectrophotometer, and when we were discussing DNA sequencing, we discussed the Sanger method and also the Maxam-Gilbert method. So, with this, I would like to conclude my lecture here. Thank you.