

**Organic Chemistry In Biology And Drug Development**  
**Prof. Amit Basak**  
**Department of Chemistry**  
**Indian Institute of Technology, Kharagpur**

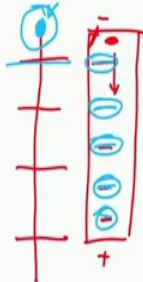
**Lecture - 24**  
**DNA Sequencing Method**

(Refer Slide Time: 00:31)

There are five steps to the Maxam-Gilbert method of DNA sequencing:

The first problem in DNA sequencing is to cleave the enormous DNA chain at predictable points to produce smaller, more manageable pieces, a task accomplished by the use of enzymes called **restriction endonucleases**. Each different restriction enzyme, of which more than 200 are available, cleaves a DNA molecule at well-defined points in the chain wherever a specific base sequence occurs. For example, the restriction enzyme *AluI* cleaves between G and C in the four-base sequence AG-CT (Figure 28.12). Note that the sequence is a *palindrome*, meaning that it reads the same from left to right and right to left; that is, the sequence (5')-AG-CT(3') is identical to its complement, (3')-TC-GA(5'). The same is true for other restriction endonucleases.

If the original DNA molecule is cut with another restriction enzyme having a different specificity for cleavage, still other segments are produced whose sequences partially overlap those produced by the first enzyme. Sequencing of all the segments, followed by identification of the overlapping sequences, then allows complete DNA sequencing.

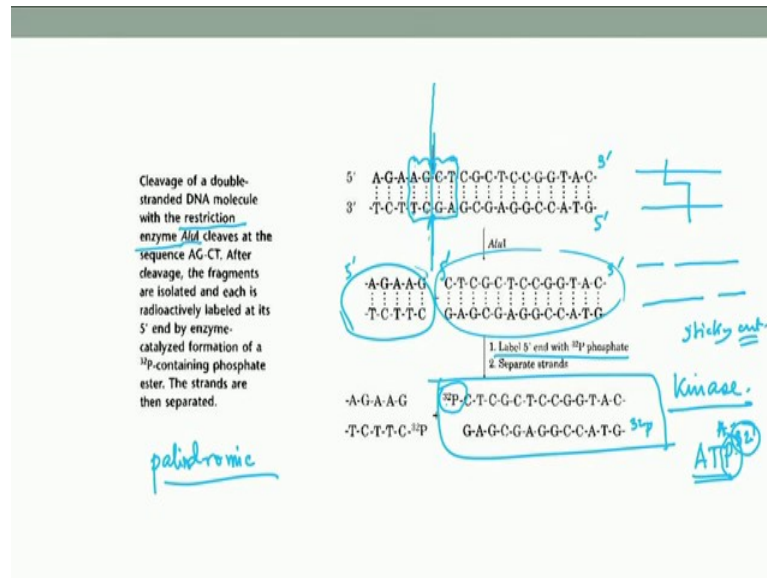


Sanger's Dideoxy method.

Welcome. Now, in this the last session, we discussed the melting temperature of DNA and its dependence on various factors. Then we started the very important aspect of determining the sequence of bases in DNA; that signifies that how the bases are arranged one after another, this feature distinguishes one DNA from the other DNA. There are two methods which were adopted when it was discovered in the late 1970s. There were two groups of scientists working on this field; one is Maxam and Gilbert and the other is Sanger's method. It is Gilbert who got the Nobel prize and Sanger got his second Nobel prize for discovering how to know the sequence of bases.

Maxam Gilbert's method is a chemical method; that means, he used a chemical reagent to know the sequence of bases. On the other hand Sanger's Dideoxy method is an enzymatic method; he used an enzyme to know the base sequence in a DNA strand. Now we will discuss the Maxam Gilbert method first.

(Refer Slide Time: 01:51)



You need to understand that since DNA is a very large molecule, so for this very large system, it is very difficult to sequence the whole molecule. In proteins also you have a large molecule; what you did there was that you chopped it into pieces. And what are these choppers? The choppers are basically the enzymes and then, you isolate those small and manageable pieces, and then do the sequence and finally, you do the overlapping sequence and get the full sequence of the protein molecule. In DNA also, very similar thing is done.

Suppose you have a double stranded DNA which is quite big. So, first thing that you have to do is that you have to cut the DNA into small pieces. Now, whenever there is discussion about this cutting of the DNA into small pieces, then enzymes come because chemically that is very difficult to cut at very precise places. But fortunately, there are enzymes which can cut cleave the DNA at particular points, where there are some sequences which the enzyme recognizes; like chymotrypsin recognizes aromatic amino acids and breaks the peptide bond.

Similarly, there are enzymes which can recognize certain sequences. Chymotrypsin recognizes only one amino acid; but here in case of DNA, some sequences of bases are recognized and then once it finds that, it will cleave there. So, these enzymes are called endonuclease. It is nuclease because they are breaking nucleic acids. They are called endonuclease as they are breaking the nucleic acid inside the chain.

Now these endonucleases are known as restriction enzymes. They are very interesting; there are possibly more than 200 restriction enzymes available. Like this restriction enzyme which is called *AluI*, there are different names *BamHI*, then *EcoRI*. So, what is the function of this enzyme? It cleaves the nucleic acid here and there; that means, it cuts the whole thing like this. This is what is called a blunt cut, which means you straightaway cut the whole thing. Because there are different types of cuts possible. One is that you can cut like this; that means, this point is here that point is there and this is a bigger one and that is the smaller one. This is called that is the blunt cut and this is the sticky cut.

So, all these cuts are possible, but anyway, here *AluI*; does a blunt cut, a straight cut. So, it cuts it in two pieces. So, there are two DNA pieces now you have here and there. Suppose you want to you are interested to know the sequence of bases here. Now, it also says something, this restriction enzyme *AluI* recognizes this AGCT sequence. Wherever there is AGCT sequence, it recognizes that and does a blunt cut between G and C.

Now, interestingly for this AGCT, if you look at the complementary strand, it is the same AGCT because you have to read from the 5' end. This is the 5' end now and this is the 3' end. So, you are reading from here; AGCT you are reading from there, AGCT; this is what is called palindromic sequence. So, all these restriction enzymes work upon palindromic sequence. So, they recognize palindromic sequence. Different palindromic sequences are available. So, different restriction enzymes are available. So, here you are using this *AluI* which recognizes AGCT; it is a palindromic sequence of 4 bases and so there is a blunt cut. Now, you have a manageable DNA double helix. Now, as I said before, you have to put a level (a reporter system) into this molecule in order to identify later when you do the gel electrophoresis.

So, what you do? Here there is this 5' end which is free because it starts from 5' and here is the 3'. There is an enzyme called the kinase enzyme which does phosphorylation of the free OH and there are different types of kinases. Some kinase does phosphorylation only on the 5' OH and do not touch the 3' OH; whereas, some kinases may be there which will do 3' OH phosphorylation and not the 5' OH.

Let us now discuss the method followed by Maxam Gilbert. They took a kinase which does phosphorylation of the 5' OH; this phosphorylation agent is nothing but ATP; in

this case ATP (a triphosphate) has the phosphorus at the gamma position labeled with the radioactive phosphorus which is called  $P^{32}$ , where the atomic mass number is 32. So, that is radioactive.

So, now, if use that kinase and add the ATP with a gamma radioactive phosphate so that phosphate will be attached to the 5' end of the DNA. So, this end will now be attached to the  $^{32}P$  DNA. Why only this end? This is because this kinase is specific for this 5' OH. Enzymes are very specific it will not go to the 3' end. So obviously, this 5' end is also  $^{32}P$  which is not written here. So, now, you have this piece where the DNA is now labeled with  $^{32}P$  phosphorus which is radioactive phosphorus and you have this double helix. Now, you heat it and separate the strands and you isolate only one strand, one of these strands with  $^{32}P$ .

Since you know these are complementary strands; so if you know the sequence of one, you know the sequence of the other. So, you have this CT; I am not writing everything because of time constrains. So, you have this C T up to A C. So, now, you have isolated a DNA strand which is labeled with  $^{32}P$  at the 5' end; this is the starting point of Gilbert Maxam sequencing method..

(Refer Slide Time: 10:20)

After cleavage of the DNA into smaller pieces, called *restriction fragments*, each piece is radioactively tagged by enzymatically incorporating a  $^{32}P$ -labeled phosphate group onto the 5'-hydroxyl group of the terminal nucleotide. The double-stranded restriction fragments are then separated into single strands by heating, and the strands are isolated. Imagine, for example, that we now have a single-stranded DNA fragment with the following partial structure:

5'  $^{32}P$ -GATCAGCGAT... 3'

4- cleavage conditions

By working carefully, Maxam and Gilbert were able to find reaction conditions that are selective for cleavage at either A or G. (They found that G methylates five times as rapidly as A, but the hydrolytic breakdown of methylated A occurs more rapidly than the corresponding breakdown of methylated G if the product is first heated with dilute acid prior to base treatment.)

Breaking the DNA chain next to the pyrimidine nucleotides C and T is accomplished by treatment of DNA with hydrazine,  $H_2NNH_2$ , followed by heating with aqueous piperidine. Although no conditions have been found that are selective for cleavage next to T, a selective cleavage next to C can be accomplished by carrying out the hydrazine reaction in 5 M NaCl solution.

Now, suppose this is my ultimate DNA strand that I will have a  $^{32}P$  here at the 5' end and suppose we take some arbitrary sequence GATCAGCGAT. I do not know the sequence; I want to develop a method by which I know the sequence of bases here.

As I told you, it was very difficult to find reaction conditions which will only recognize G and then break that phosphodiester bond between the two base sugar hybrids (nucleosides). But the group under Maxam and Gilbert were so brilliant that they could find out reaction conditions by which they can selectively selectively cleave these phosphodiester linkages. So, what they did? They developed four cleavage conditions.

(Refer Slide Time: 11:43)

Cleave DNA strand at specific positions using chemical reactions. For example, we can use one of two chemicals followed by piperidine.

Dimethyl sulphate selectively attacks purine (A and G), while hydrazine selectively attacks pyrimidines (C and T).

The chemical treatments outlined in Maxam-Gilbert's paper cleaved at G, A+G, C and C+T. A+G means that it cleaves at A, but occasionally at G as well.

For A+G: DMS, then dil H<sup>+</sup>, then piperidine

For G only: DMS, then piperidine

For C+T: NH<sub>2</sub>NH<sub>2</sub>, piperidine

For C only: NH<sub>2</sub>NH<sub>2</sub>/aq NaCl, then piperidine

The diagram shows a DNA strand with the sequence 32P-C-A-T-G-C-A-G. A blue circle highlights the 'A' in the second position. A blue arrow points from the 'A' to the right, indicating cleavage. The resulting fragments are shown as 32P-C and 32P-ATGC. A handwritten note '(CH<sub>3</sub>)<sub>2</sub>SO<sub>4</sub>' is written next to the 'A' in the original sequence.

Under one condition, wherever there is A and wherever there is G, then you get a cleavage. So, suppose you have a sequence like this C A T G C A G and this is your <sup>32</sup>P here. If you utilize this condition, (we will discuss the condition little later) wherever there is A, that will be cleaved. Now A is here; so the question is which side of A will be cleaved? In proteins, you know that for chymotrypsin, always hydrolyses the peptide end between the aromatic amino acid and the other amino acid from the C-terminal. It will not cleave the peptide bond from the N-terminal. But here the reactions are such that it will actually cleave both sides of A.

If both sides of A are cleaved, what you will get? You will get <sup>32</sup>P and C, which will be one of the pieces that you will get; the other piece will be T G A C A G; but this second piece does not have the radioactive phosphorus in it. So, this will not be visible. So, the part on the right side can be neglected because that will not be visible in the electrophoresis gel.

So, we are only interested in the pieces which contain radioactive phosphorus, but this may not be the only one if there is another A. So, you can get a breakage here on both sides of A if that happens, then you will get  $^{32}\text{P}$  C A T G C. This method is for A plus G; that means, again you have to see where are the Gs and then what are the cleavage products. So, this is one cleavage product; you will get another cleavage product, you will get  $^{32}\text{P}$  C A T then there will be cleavage of all this G.

Again I repeat that I am not bothered about the right side because on right side, whatever will be released, that will not have any radioactive phosphorus. So, you can ignore that as it will not show up in the gel. What will show up is this piece is  $^{32}\text{P}$  C A T where both sides of G are cleaved. There is another G here. Both sides of G are cleaved so, you will get  $^{32}\text{P}$  C A T G C A. So, these are the 4 pieces that you will get, if you adopt this condition. Under these conditions, you take this piece of DNA and then add DMS (dimethyl sulfate) and then you add dilute H plus (usually formic acid) and then aqueous piperidine.

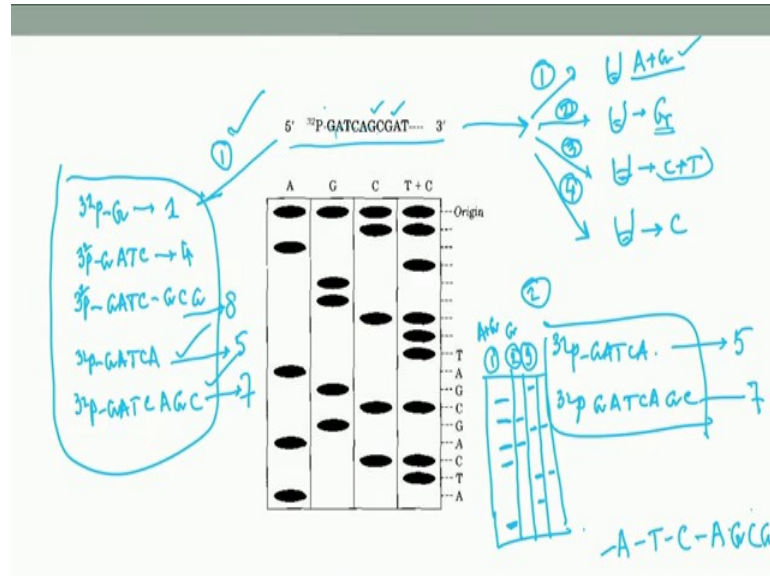
So, if you subject the DNA to this condition, you will see that there will be cleavage wherever there is A in the strand and wherever there is G in the strand. Then they developed another condition where only G will be cleaved and the condition is they just drop this dilute H plus treatment. So, if you take dimethylsulfate and then add aqueous piperidine, then you will get only the cleavage of the G; that means, you will only get this and this, you will not get these two pieces because there is no cleavage involving A.

And then they got another two conditions; one condition which cleaves both C and T. So, wherever there is C, there will be cleavage on both sides of C. So, you will get accordingly the number of cleavage products containing radioactive phosphorus. This is achieved by treatment with hydrazine, then aqua sodium chloride, and then aqueous piperidine. They also developed a method for C specific cleavage; which was achieved through omission of sodium chloride from the above mentioned condition.

So, basically what have they done? They have identified four reaction conditions; in one reaction condition, both A and G are cleaved; wherever there is A, wherever there is G that will be cleaved. In another condition only Gs are cleaved, wherever there is G, that will be cleaved and you will get the piece. Another one is C plus T; so you get C and T both cleaved; wherever there is C or wherever there is T, you will get cleavage; and in

the fourth condition, wherever there is C, that will only be cleaved. So, these are the conditions.

(Refer Slide Time: 17:20)



Now, let me give you an example; I think that will give you much better concept. So, suppose we started with this 5' <sup>32</sup>P phosphorus, then G A T C A G C G A T; I do not know what the sequence was. But I adopt those four conditions or I subject the DNA to these four conditions. Basically what I do? I take small test small tubes which are called eppendorf tubes and I take the DNA in every eppendorf tube; in one tube, I subject that to a condition where both A plus G will be cleaved; in another one, only G specific cleavage condition is employed; in another one, C plus T and the fourth one, only C specific cleavage conditions are employed. So, let us inspect this. A+G is the number 1 reaction, only G is the number 2, C+T is number 3, only C is number 4 reaction.

So, if I do the number 1 reaction how many pieces will I get? In number one, both A and G are cleaved. So, you have to see where are these A and G; first let us consider A. So, there is the A here. So, I will get <sup>32</sup>P and then G that is my first one, the second one is there is another G. So, I will get <sup>32</sup>P G A T C . One valid point that I missed is that I am considering here that one piece of DNA gets attacked only by one molecule of the reagent or it gets attacked only once by the reagent.

That means, if I have this molecule, you will not get cleavage of both the Gs in the same molecule because the concentration of the reagents are kept at such a level that statistically only one reaction is going to happen to one molecule.

So, that is why we are getting this piece because if this gets attacked by the reagent, this will be broken; there is another G and that is a separate molecule that G will be cleaved; that is why we are getting two different oligonucleotides.

Now, because it is the A plus G reaction, so more number of pieces will get  $^{32}\text{P}$ .

Again I repeat As are cleaved, so you get  $^{32}\text{P}$  G and this A is cleaved, you get  $^{32}\text{P}$  G A T C; but there is another A here. So, you will get,  $^{32}\text{P}$ GATCAAGCG. So, these are the three pieces from cleavage of A. Now we have to see what are the cleavage products of G? So, if the G is cleaved on both sides, basically there is no nucleotide on the left side. So, you can neglect the first one, the second one is this G. So, you will get what  $^{32}\text{P}$  GATCA.

There is one more G here. So, you will get  $^{32}\text{P}$  GATCAGC. So, these are the cleavage products, containing radioactive phosphorus that you will get in the first test tube. In the second test tube, only G specific cleavage will be obtained. So, in the second one, you will get  $^{32}\text{P}$ GATCA and then  $^{32}\text{P}$  GATCAGC. So, these are the two from the second test tube. And from the third test tube, where you have C plus T, you will get  $^{32}\text{P}$ GAT; and then you will get  $^{32}\text{P}$  GAT CAG.

So, you will get up to G. So, you will get two pieces for C and there are two Ts so, you will get two pieces from T. And then for C specific reaction, since there are two Cs, so you will get  $^{32}\text{P}$ GAT and then  $^{32}\text{P}$ GATCAG. Now we are not writing all that; what is important is that without going into details, we can summarize it very easily.

Let us consider this number one testtube where A and G are cleaved. You see the number of bases present in these pieces. So, the first cleaved fragment contains only 1 base, the second cleaved fragment contains 4 bases, the third cleaved fragment contains 8 bases, the fourth cleaved fragment contains 5 bases and the sixth cleaved fragment contains 7 bases. And in the second test tube, for the first cleaved fragment contains 5 bases and the second cleaved fragment will have 7, bases.



Now let us consider the results of doing an electrophoresis; forget about this diagram. If you do an electrophoresis, this is your first test tube which you are adding here. So, how many bands you will see when you take the radioactive photograph? You will see number 1 band, number 4 band, number 8 these are basically proportional to the molecular weight of the whole nucleotide; the more the number of bases, the more is the molecular weight. So, your number 1 (that will be the fastest moving), then you have number 4 here, you have 5 here that will be little bit slower and you have 7 here and you have 8 here.

So, these are the band. So, 5 bands will be visible. If you take the second test tube how, many bands you will see? Bands corresponding to 5 and 7. So, you will see a band here and you will see a band here. Similarly when you take the number 3 test tube, how many bands you will see? The number of bases is 3.

This you will get in the third lane. So, there will be a band here that is due to GAT because it is a C specific cleavage; and there is a T specific cleavage also because it is C plus T. So, you will get another one here.

And then there is one another C here. So, you will get a band corresponding to 6. 6 is somewhere here. There is a T here. So, you will get a band corresponding to 9 and that is the slowest one. And when you have only C, you will get only two bands. There are two Cs, that means one band is corresponding to 3 and the other will correspond to 6. 6 is somewhere here. So, this is the picture of the gel. So, that is the number 1 lane, this is the number 2 lane, this is the number 3 lane and that is the number 4 lane.

So, now what you do? This is your A plus G lane and the next one is your G lane. Now you want to know whether this one is only due to A or G. If this is only due to A, then in the G lane, you are not going to see any band; if it is due to G then you will see both in A plus G and G lane. Similarly for C plus T, if it is only T then you will only see in the C plus T lane, but you will not see that in the C lane. But if you see both; that means, it is actually C. This is C, and this is your G because they are present in both A plus G and G.

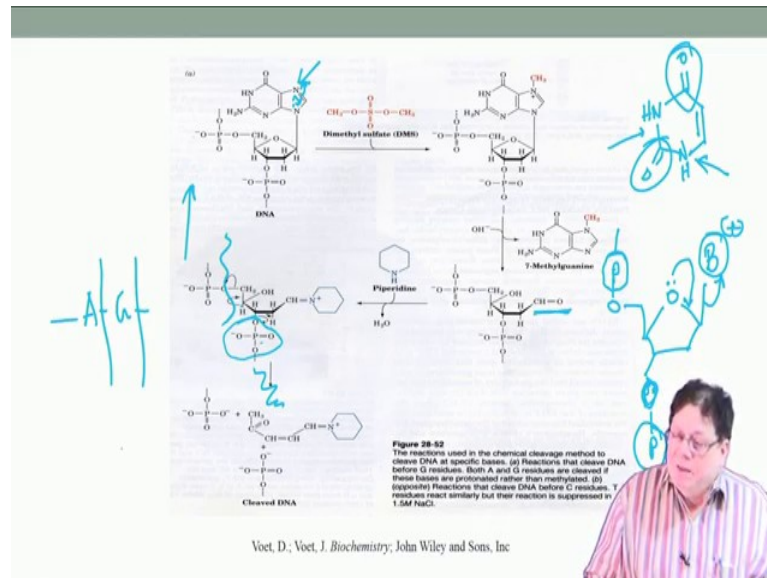
Now what you do? You read this gel from this side. So, you can write that first one is A, the second one one must be T, because you are not seeing it in the C lane. The third one must be C because you are seeing in both the lanes; I repeat, this needs careful screening. So, A then T then C then A, let us see whether it is matching. We are not getting the first

one. Actually for Maxam Gilbert, you do not get the first one since when G, is cleaved you do not get any nucleotide attached to the radioactive phosphorus.

So, first one you forget and then you read, ATCA then G then it will be C because both the bands are present then it will be G; just that way you can continue. We see that it is matching. ATCAG then C G and if you continue you will get the other two A T. So, that is how it is done. It is not very difficult, but the problem with Maxam Gilbert method is that you need good pair of hands; it needs lot of craftsmanship. I am not giving you the actual conditions that how many minutes we have to heat with hydrazine, how many minutes you have to treat with dimethyl sulfate, how many minutes you add sodium base or hydra pyridine aqueous pyridine? So, it is very person dependent process. So, your hands have to be extremely scrupulous.

So, that is one problem, otherwise when Maxam Gilbert method was discovered, it got wide attention and as you will see that Gilbert received the Nobel Prize for his work. But, this is a chemical method; he developed chemical conditions by which you can break the DNA strand at different sites depending on the base that is present. I hope this whole thing is clear. Sometimes problems will be given in the other way around that without giving this one, we give the gel picture. And then from the gel picture you may be asked to predict the sequence of the DNA. But reading the sequence is extremely easy; you just read from the bottom and go further up and you get from the 5'to 3' sequence.

(Refer Slide Time: 32:03)



Now, let us discuss the chemistry part in it. The chemistry part will address the issues like why there is specific cleavage? Or how was this cleavage of DNA strand accomplished? If you take a base, both the phosphodiester as bonds on the 3' end and the 5' end are cleaved.

So, some mechanisms are shown in the last two slides. Remember the pyrimidine based nucleobases are like this. If you look at the nitrogens, they are not that basic because the nitrogen lone pair of electrons are in conjugation with this carbonyl; hence they are not very basic. But, in the purine, forget about these nitrogens, but this is the nitrogen (N-7) which is extremely basic because of the flow of electrons like this.

So, this nitrogen is very basic and dimethyl sulfate is an alkylating agent. So, whenever you add an alkylating agent, wherever there is nucleophilic nitrogen, that is going to attack the dimethyl sulfate and it will undergo methylation. So, now, it is a general rule. I am not going into the detailed mechanism, but we can simplify the whole mechanism. This is the general structure of a DNA, this is the phosphate; without writing phosphate I made a circle showing it as a phosphate. One thing you must remember is that, for this base, if you can somehow put a positive charge on this base, then this N-glycoside bond becomes very weak; this oxygen lone pair flies here and the whole base is gone.

Basically this is called depurination or depyrimidination; because you are removing that purine base or the pyrimidine base. But, remember to do that, you need is to develop a

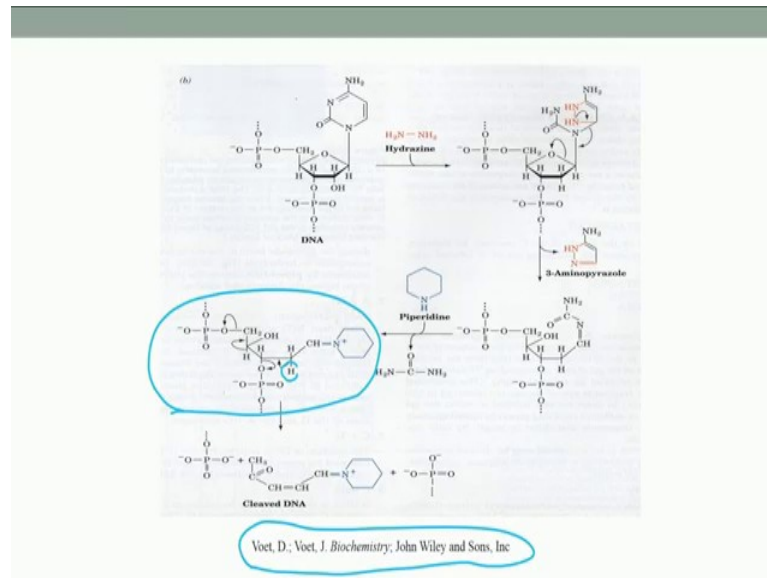
positive charge. This is the detailed mechanism that will be given in the slide, but the whole concept is based on this. So, when they added dimethylsulfate, this nitrogen (N- $\delta$ ) is alkylated; that means, this base has got positive charge. So, if you see that in the next step, immediately water comes and attacks here; before that can happen, this lone pair flies and breaks it down then water comes and attacks here.

So, your base is gone and it becomes ultimately an open chain. You know that sugars can be in open chain form. So, it becomes an aldehyde and alcohol. This aldehyde reacts with piperidine to form this piperidinium ion which is a Schiff's base. Due to this positive charge on nitrogen, these hydrogens at the 2' position become very acidic. So, now, there is a  $\beta$ -elimination;  $\beta$ -elimination results in a catastrophic outcome for the DNA; this phosphodiester now leaves; that means, whatever DNA chain is at the bottom, that goes off.

And as soon as there is a double bond here so, this hydrogen becomes acidic and that also does the same thing, it kicks out the phosphate. So, from the 5' end you have a breakage of phosphate and also from the 3' end, you have breakage of phosphate. Remember that in Maxam Gilbert method, there are breakages from both side (from the 3' as well as from the 5 prime end).

And by carefully adjusting the conditions, they successfully developed a condition where only G is affected; and A plus G is affected. They knew that if they do not need only A specific reaction, what they need is A plus G and G or A plus G and A, that will give you the information. So, they are clever enough to do that and then C plus T and C.

(Refer Slide Time: 36:46)



So, that is one mechanism which is given and the other mechanism is that involving hydrazine. You can remove the pyrimidine bases by putting the hydrazine; and this is given here. I think that reference is given here; you can go to the book on biochemistry by Voet and Voet for this mechanism of hydrazinolysis. This is called hydrazinolysis because you are cleaving the bond by hydrazine. Ultimately this is very similar like the earlier one; you get this Schiff's base, this increases the hydrogen acidity that undergoes  $\beta$ -elimination. So, you remove both the 3' and the 5' phosphate groups.

So, that ends the discussion on Maxam Gilbert method. Remember it is a chemical method and there are four reactions that has to be done and then, in the gel there are 4 lanes and the species runs according to the number of bases. That means their molecular weight. One important point is the first base that is attached at the 5' that cannot be located or that cannot be determined. Because, when there is a cleavage, there is no nucleotide which has got the radioactive phosphorus; it is only the radioactive phosphorus that is released. So, that is not a nucleotide, so you will not be able to see that. In the next session, we will do the Sanger's method.

Thank you.