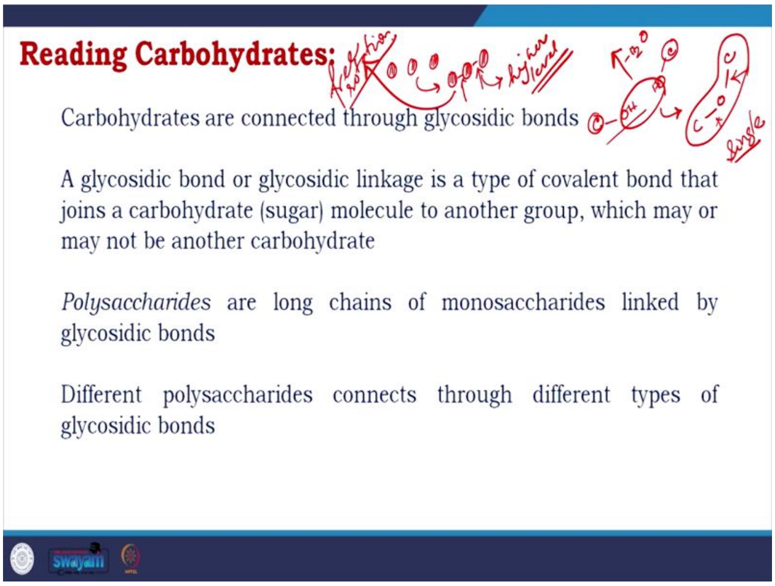
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology – Roorkee

Lecture 03
Introduction: Decoding Biological Macromolecules

Hi everyone. Welcome to another class of the course, Structural Biology. We are going through the module of the introductory section and today, as told in the previous class at the end. We are going to discuss decoding biological macromolecules.

(Refer Slide Time: 00:47)



Reading Carbohydrates:

Carbohydrates are connected through glycosidic bonds

A glycosidic bond or glycosidic linkage is a type of covalent bond that joins a carbohydrate (sugar) molecule to another group, which may or may not be another carbohydrate

Polysaccharides are long chains of monosaccharides linked by glycosidic bonds

Different polysaccharides connects through different types of glycosidic bonds

So, how to decode biological macromolecules? The first and foremost requirement is to understand biological macromolecules is to read. You have to know about the basics. You have to read how the monomers are forming polymers. We are targeting three Polymers carbohydrate, protein, and nucleic acids, and then we are also interested in knowing about lipid and fat. So, how the primary sequence of those Polymers could be studied? What are the challenges in studying them? What are the possibilities of improvement in terms of instrumentation and technical development people are working on.

So let us start with reading carbohydrates. As I told you, carbohydrates are considered a very important molecule because of food sources. You will get the most accessible sources. They are very, very rich in carbohydrates. Carbohydrates are connected through glycosidic bonds. A glycosidic bond or glycosidic linkage is a covalent bond that joins a carbohydrate molecule to another molecule.

So basically, in the case of carbohydrate, you have a hydroxyl group connected to the carbon and another carbohydrate-containing the hydroxyl group. Now it will go through dehydration and loss of water, making a glycosidic bond. This is called a glycosidic linkage.

So when we are thinking about the architectural hierarchies, they would go on the higher level of structure. But if there are connected by single bonds in between any confirmation, remember I told you about three secretes of covalent bonds: Chirality, configuration, and conformation. So, here they would adopt any confirmation because free rotation is possible, which is why the prediction would be very difficult. Polysaccharides are a long chain of monosaccharides linked by glycosidic bonds. Different polysaccharides connect through the different types of glycosidic bonds.

(Refer Slide Time: 05:03)

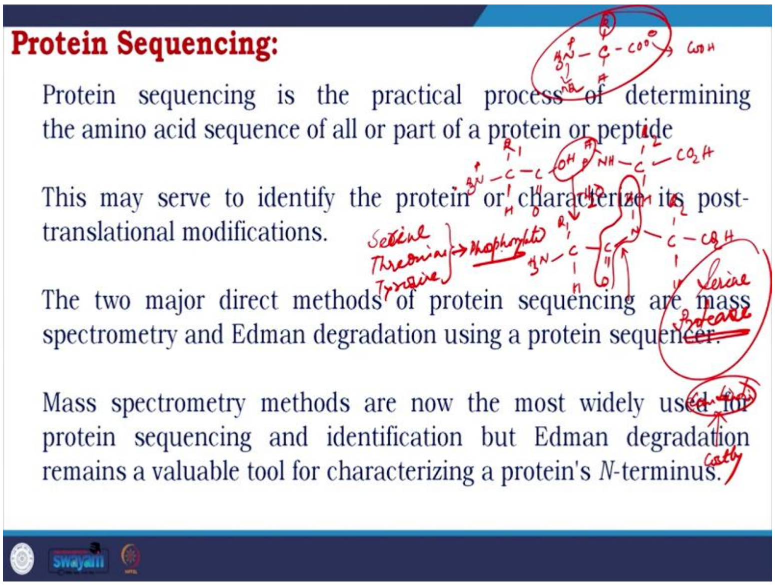
If you see α (1,4) glycosidic bond, these are present in amylose, and you see that in a straight chain. Now when you look at another glycosidic bond, this is in cellulose, which is β (1,4) be forming straight-chain. Now you see the branching where you see the presence of both α (1,4) and α (1,6) glycosidic bond.

α (1,6) bond is present in a polysaccharide called amylopectin. A branch chain and a single bond in the connectivity mix is nearly impossible to study carbohydrate. Also, when they are complicated for the same bond to get access, you need other enzymes, which is very interesting. Many biochemical engineers, enzymes engineers, polymer specialists are involved in utilizing and valorising them.

And one of the classic examples is cellulose, hemicelluloses, and lignin. So, a lot of research is currently on where people are trying to degrade Lignin. If you could degrade lignin, you could get many value added products like chlorogenic acid, sinapinic acid, and all.

So what I mean is all the DNA, RNA, and protein have a pre-made template, but in the case of carbohydrate, lipid, and fat, you do not have any template. When you are investigating a particular cell, you have information about the chromosome, and you know what type of DNA? But carbohydrate though we say that monosaccharide forms polysaccharide, or in case of fat, glycerol and fatty acid, there are a number of fatty acids number of monosaccharides which could be very different. There is no standard set for them. That is another problem, and they are very individualistic.

(Refer Slide Time: 13:43)



Protein Sequencing:

Protein sequencing is the practical process of determining the amino acid sequence of all or part of a protein or peptide

This may serve to identify the protein or characterize its post-translational modifications.

The two major direct methods of protein sequencing are mass spectrometry and Edman degradation using a protein sequencer.

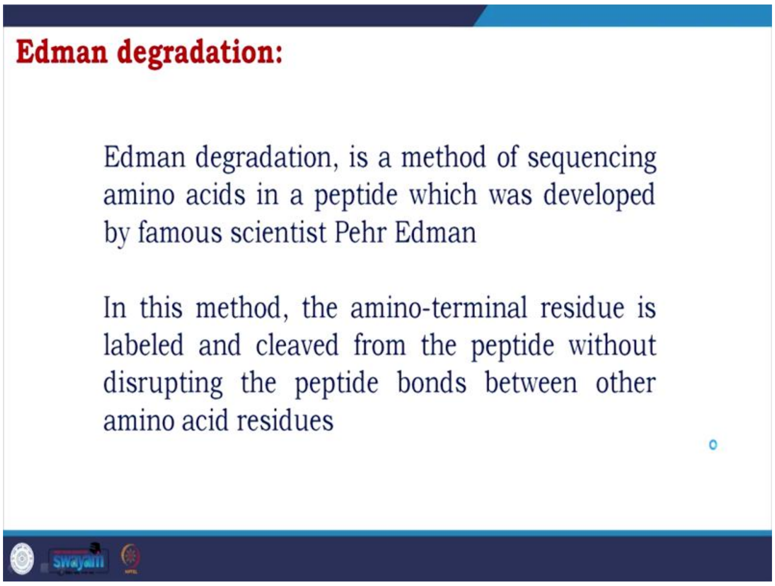
Mass spectrometry methods are now the most widely used for protein sequencing and identification but Edman degradation remains a valuable tool for characterizing a protein's *N*-terminus.

In the case of protein sequencing, it is the practical process of determining the amino acid sequence of all or part of the protein or peptide. This may serve to identify the protein or characterises post-translational modification. They could be phosphorylated, and in that way methylation, acetylation there is post-translational modification.

The two major direct methods of protein sequencing are mass spectrometry. Mass spectrometry methods are now most widely used for protein sequencing and identification.

But Edman degradation remains a valuable tool for characterizing a protein. Do you know a little sequence in N-terminus, and you could get to find out the protein done by Edman degradation.

(Refer Slide Time: 18:55)



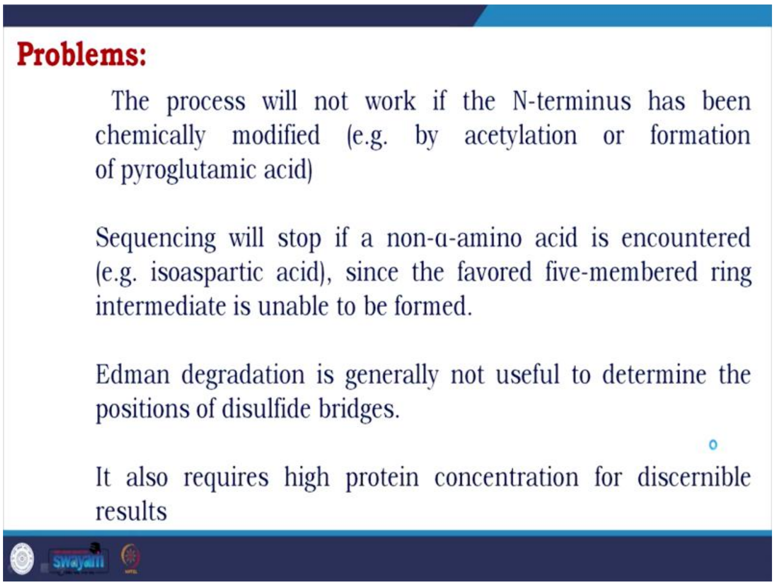
Edman degradation:

Edman degradation, is a method of sequencing amino acids in a peptide which was developed by famous scientist Pehr Edman

In this method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues

What is Edman degradation? Edman degradation is a method of Sequencing amino acids in a peptide that famous scientist Pehr Edman developed. In this method, the amino-terminal residue is labelled and cleaved without disrupting the peptide bonds between other amino acids. So you could do that one by one. That is the beauty of the method.

(Refer Slide Time: 19:26)



Problems:

The process will not work if the N-terminus has been chemically modified (e.g. by acetylation or formation of pyroglutamic acid)

Sequencing will stop if a non-α-amino acid is encountered (e.g. isoaspartic acid), since the favored five-membered ring intermediate is unable to be formed.

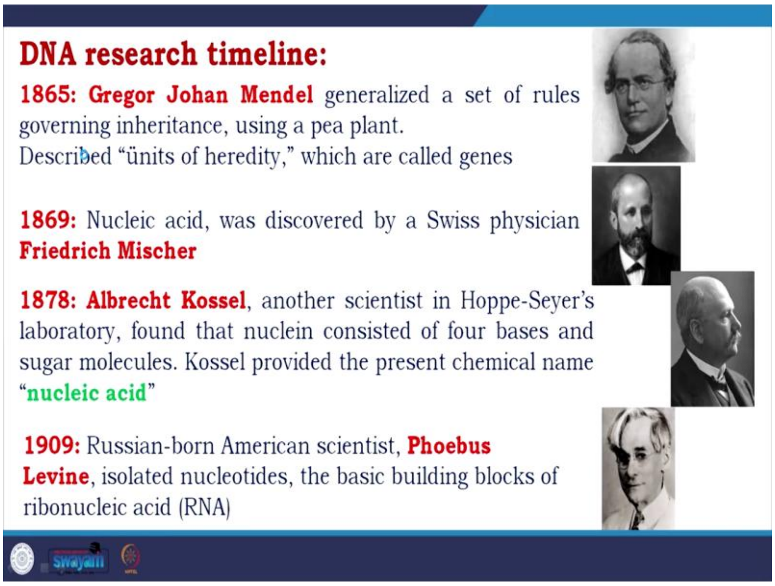
Edman degradation is generally not useful to determine the positions of disulfide bridges.

It also requires high protein concentration for discernible results

This process will not work if N-terminus has been chemically modified. For example, acetylated or pyroglutamic acid. Sequencing will stop if a non-alpha-amino acid is encountered like Isoaspartic acid, since the favoured five-membered ring intermediate, which is essential for the process of this reaction to go on, is unable to be formed.

Edman degradation is generally not useful to determine the position of disulfide bridges. Also, it requires high protein concentration, but besides the Edman degradation, the normal sequencing using mass spectrometry is not very useful because, as I told you, you have to digest. You need a combination of enzymes. Second, individually, the process is very expensive.

(Refer Slide Time: 21:28)



DNA research timeline:

1865: Gregor Johan Mendel generalized a set of rules governing inheritance, using a pea plant.
Described "ünits of heredity," which are called genes

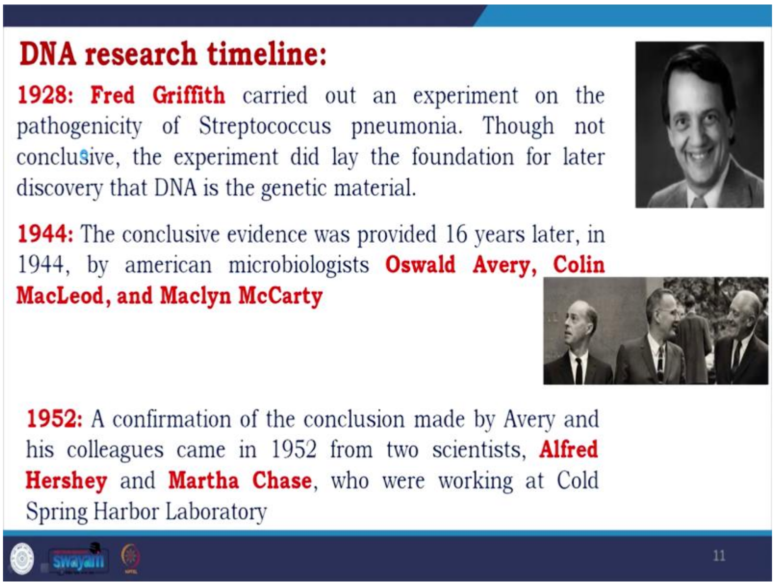
1869: Nucleic acid, was discovered by a Swiss physician **Friedrich Mischer**

1878: Albrecht Kossel, another scientist in Hoppe-Seyer's laboratory, found that nuclein consisted of four bases and sugar molecules. Kossel provided the present chemical name "**nucleic acid**"

1909: Russian-born American scientist, **Phoebus Levine**, isolated nucleotides, the basic building blocks of ribonucleic acid (RNA)

The first DNA research breakthrough was in 1865 with Gregor Johan Mendel. We know he is called the father of genetics, used the Pea's, did extensive hybridization, and developeded basic genetics laws. In 1869 Friedrich Mischer came up with the nucleic acid we call nucleum. In 1878 Albert Kossel from the same lab, Hoppe-Seyer's laboratory, came up with the name nucleic acid. He also identifies that this molecule consists of four bases and sugar molecules. In 1909, Russian-born American scientist Phoebus Levine isolated ribonucleic acid and characterised the building blocks.

(Refer Slide Time: 22:30)



DNA research timeline:

1928: Fred Griffith carried out an experiment on the pathogenicity of Streptococcus pneumonia. Though not conclusive, the experiment did lay the foundation for later discovery that DNA is the genetic material.

1944: The conclusive evidence was provided 16 years later, in 1944, by american microbiologists **Oswald Avery, Colin MacLeod, and Maclyn McCarty**

1952: A confirmation of the conclusion made by Avery and his colleagues came in 1952 from two scientists, **Alfred Hershey** and **Martha Chase**, who were working at Cold Spring Harbor Laboratory

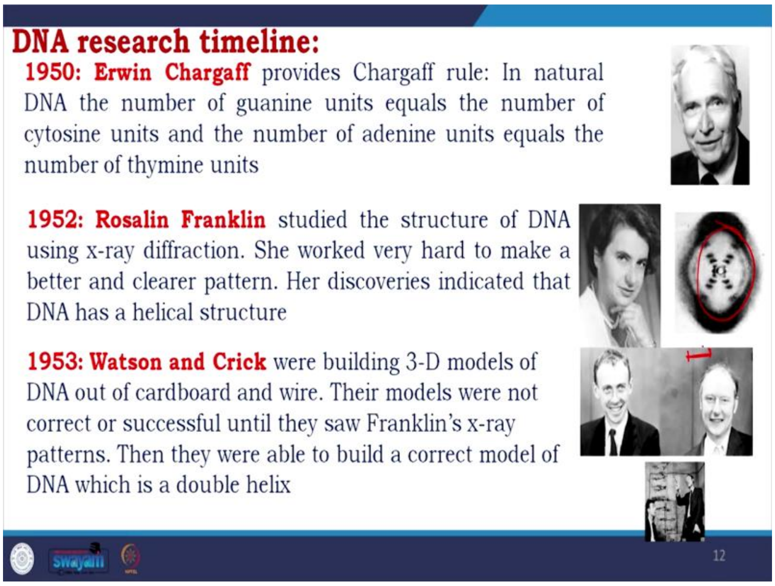
11

In 1928 Fred Griffith experimented on the pathogenicity of Streptococcus pneumoniae using a virulent and un-virulent strain. His experiment would not be very conclusive, but that opened the field, which was continued in 1944 by Oswal Avery, Collin Macleod and Maclyn

McCarty, who have taken the factor to add protease and DNAase and so that when there is DNAase, the factor which is virulence factor is not effective. So it is DNA.

Again in 1952, a more biochemistry-oriented confirmation came from the famous Harshey Chase experiment where the labelling of protein and DNA using Sulphur and phosphorus confirmed that this is the genetic material. This is DNA.

(Refer Slide Time: 23:48)



DNA research timeline:

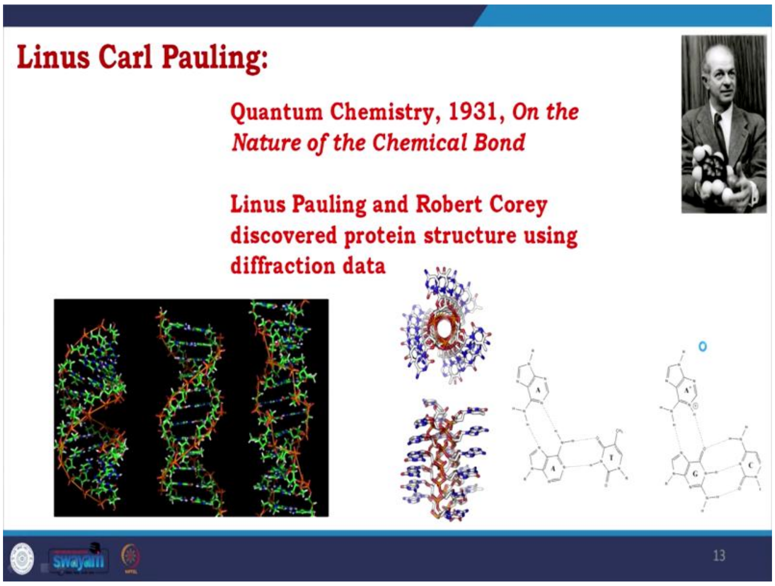
1950: Erwin Chargaff provides Chargaff rule: In natural DNA the number of guanine units equals the number of cytosine units and the number of adenine units equals the number of thymine units

1952: Rosalin Franklin studied the structure of DNA using x-ray diffraction. She worked very hard to make a better and clearer pattern. Her discoveries indicated that DNA has a helical structure

1953: Watson and Crick were building 3-D models of DNA out of cardboard and wire. Their models were not correct or successful until they saw Franklin's x-ray patterns. Then they were able to build a correct model of DNA which is a double helix

In 1950 Erwin Chargaff provided the Chargaff rule, which is in natural DNA. The number of guanine units equals the number of cytosine units, and the number of adenine units is equal to the number of thymine units. In 1952 the real breakthrough came through the work of Rosalin Franklin. She studied the structure of DNA using X-ray diffraction. The first time people have used DNA to determine the structure, they worked very hard to make a better and clearer pattern. Multiple time data collection and discoveries indicated that DNA has a helical structure. In 1953 Watson and Crick were working but were trying to fit a model using the Chargaff AGC rule, but they were unable. Still, looking at Rosalin Franklin work, it helped them build a 3D model of DNA out of cardboard and wire where their models were not correct for successful until they saw Franklin X-ray patterns and then were able to build the correct model of DNA is a double helix.

(Refer Slide Time: 26:01)



Linus Carl Pauling:

Quantum Chemistry, 1931, *On the Nature of the Chemical Bond*

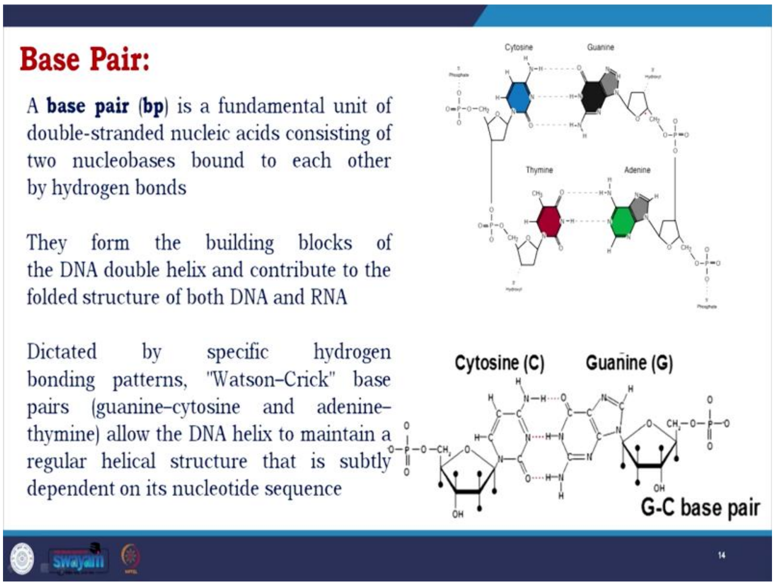
Linus Pauling and Robert Corey discovered protein structure using diffraction data

Linus Pauling was a chemist. In 1931 Linus Pauling contributed to Quantum Chemistry with the work on the nature of the chemical bond. He talked about the improvement of valence bond theory and the concept of resonance.

Linus Pauling and Robert Corey discovered the protein structure using diffraction data. He was awarded Nobel Prize for this work. He was not stopping even.

What is not mentioned regularly, the first model came from Pauling; this is the first model where he talked about the helical nature of DNA. So the first breakthrough the first concept was given by him the only thing he talked about was a triple helix. But it was proved that DNA is a double helix proving him wrong. But today, if you look at you will see that there are hooked in base pairs where there is evidence of three nucleosides forming interaction together and forming a triple helix.

(Refer Slide Time: 28:53)



Base Pair:

A **base pair** (**bp**) is a fundamental unit of double-stranded nucleic acids consisting of two nucleobases bound to each other by hydrogen bonds

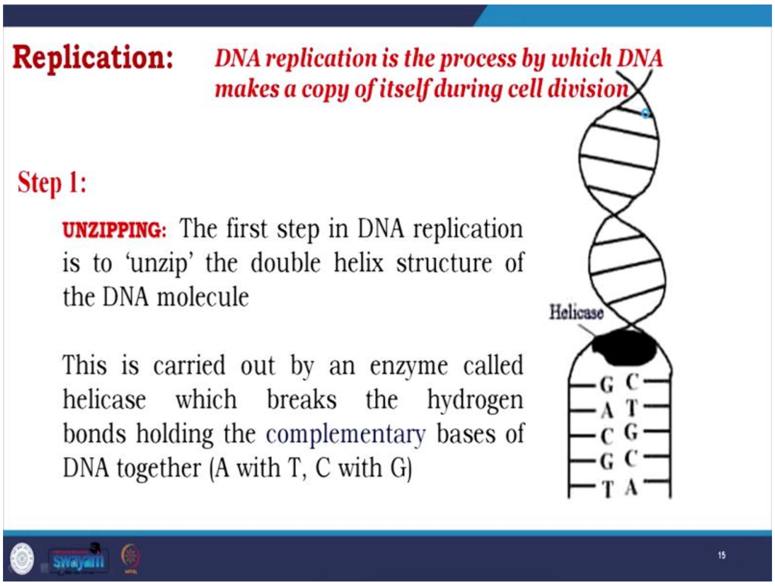
They form the building blocks of the DNA double helix and contribute to the folded structure of both DNA and RNA

Dictated by specific hydrogen bonding patterns, "Watson–Crick" base pairs (guanine–cytosine and adenine–thymine) allow the DNA helix to maintain a regular helical structure that is subtly dependent on its nucleotide sequence

Now we will talk about the basic properties of DNA. The base pair is the fundamental unit of double standard nucleic acids consisting of two nucleobases bound by hydrogen bonds. They form building blocks of the DNA double helix and contribute to the folded structure of both DNA and RNA. Dictated by a specific hydrogen bonding pattern; the Watson-Crick base pair allowed the DNA helix to maintain a regular helical structure that is subtly dependent on nucleotide sequence.

There are three hydrogen bonds between cytosine and guanine and two hydrogen bonds between adenine and thymine.

(Refer Slide Time: 29:44)



Replication:

DNA replication is the process by which DNA makes a copy of itself during cell division

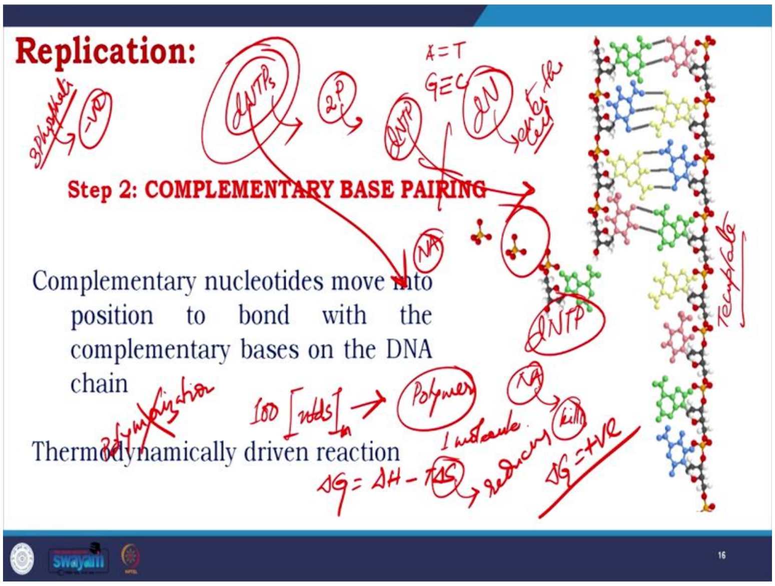
Step 1:

UNZIPPING: The first step in DNA replication is to 'unzip' the double helix structure of the DNA molecule

This is carried out by an enzyme called helicase which breaks the hydrogen bonds holding the complementary bases of DNA together (A with T, C with G)

DNA replication is how DNA makes a copy of itself during cell division. The first step is unzipping. It is to unzip the double helix because it is bound with non-covalent interactions and base stacking and hydrogen bond. So unzip the double helix structure of the DNA molecule. The helicase enzyme carries out this, which breaks the hydrogen bond, holding the complementary bases of DNA together. There is a complex protein specially called single-strand binding protein, which binds to single strands after unzip, and in that way, many proteins come and play their role to develop complexes.

(Refer Slide Time: 31:24)



Replication:

Step 2: COMPLEMENTARY BASE PAIRING

Complementary nucleotides move into position to bond with the complementary bases on the DNA chain

Thermodynamically driven reaction

The second is complementary base pairing. So, in complementary base pairing, the complementary nucleotides move to bond with the complementary bases on the DNA chain. So you see that this is the templates strand, and the bases are coming accordingly. If there is A is T, G is C, and vice versa. But if you look at a very interesting thing here, you see that the nucleotides are coming as triphosphate. So they are coming as DNTP's, and two phosphates are released. So, if you think thermodynamics

$\Delta G = \Delta A - T\,\Delta S$

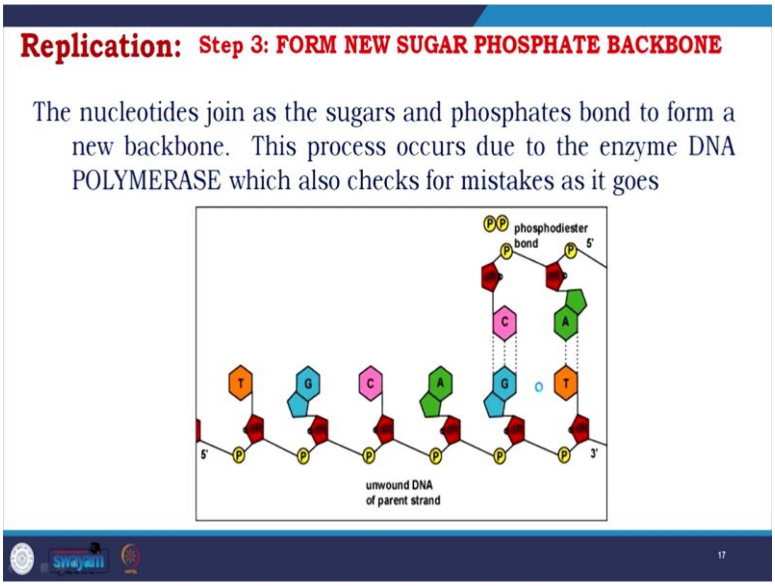
entropy is significantly reducing, next to the ΔG positive, which means the polymerization reaction is not spontaneous.

So you need more energy to make ΔG more negative, that complementation is coming through the inorganic phosphates. When they break and release a negative amount of energy- 7.3 kj/mol, energies compensate the ΔG positive value that is one thing. So they are coming as the DNTP's because they compliment the thermodynamic energy.

But there is something more to think about nucleoside could have come to the cell from food or other things. Now when we say deoxynucleoside, it could enter into the cell. Enter the cell, but the DNTP molecule could not enter the cell. A DNTP molecule could not enter the cell because, in the presence of 3 phosphates, they are very negative in charge, and membranes outside the cell membrane are negative, so they repelled.

So, if they have to enter the cell, any nucleoside or, more importantly, nucleoside analogs cannot enter in the form of the DNTP. If that would happen, then there would be many. Foods that might give nucleoside analogs could be killing us. We are saved because DNA cannot receive monophosphate; it takes it only in the state of triphosphate, hence selectivity of proper nucleosides and not nucleoside analogs entering the DNA. So, as I told is thermodynamically driven reaction and specificity matter.

(Refer Slide Time: 36:43)

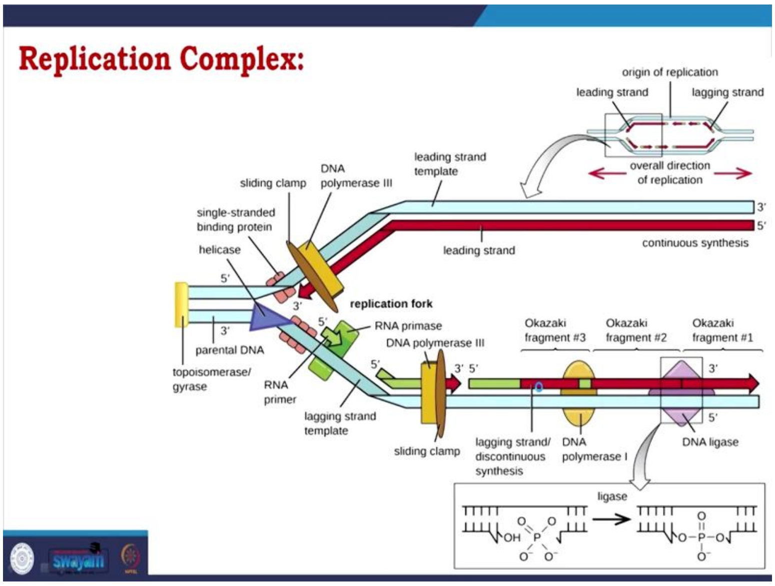


Replication: **Step 3: FORM NEW SUGAR PHOSPHATE BACKBONE**

The nucleotides join as the sugars and phosphates bond to form a new backbone. This process occurs due to the enzyme DNA POLYMERASE which also checks for mistakes as it goes

The third one is forming the new sugar-phosphate backbone. The nucleotides join as the sugar and phosphates bond to form a new backbone. These processes occur due to the enzyme DNA polymerase, which also checks for mistakes.

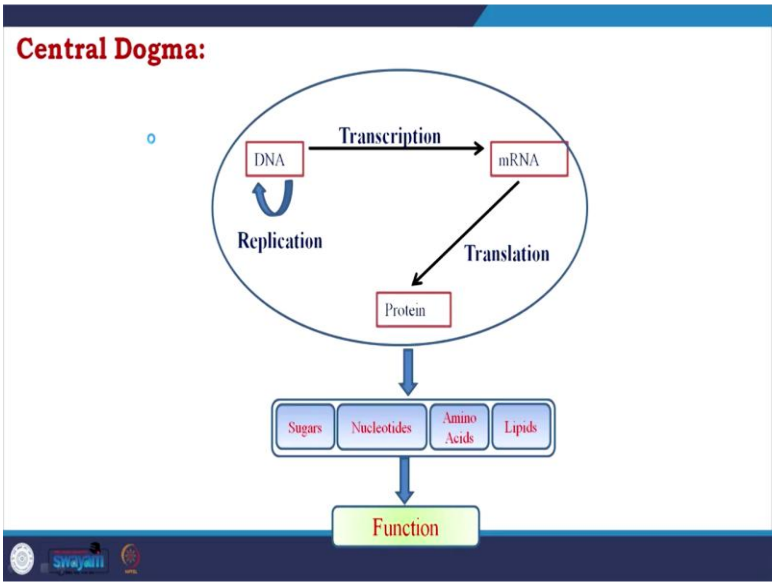
(Refer Slide Time: 37:04)



Replication Complex:

So DNA polymerase is important. But if you see that, I am trying to say. Many other enzymes are taking the role of processing the whole thing. It is complex, but DNA polymerase is the main enzyme with a processivity (proof-reading activity).

(Refer Slide Time: 38:14)



Central Dogma:

DNA

Transcription

mRNA

Replication

Translation

Protein

Sugars

Nucleotides

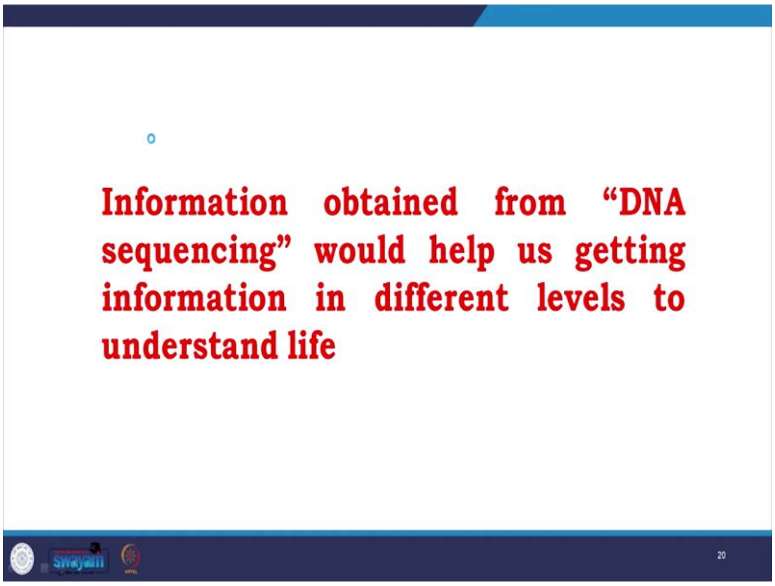
Amino Acids

Lipids

Function

In the central dogma, DNA could develop another DNA through replication. DNA could develop through mRNA through transcription, and mRNA make protein to translation.

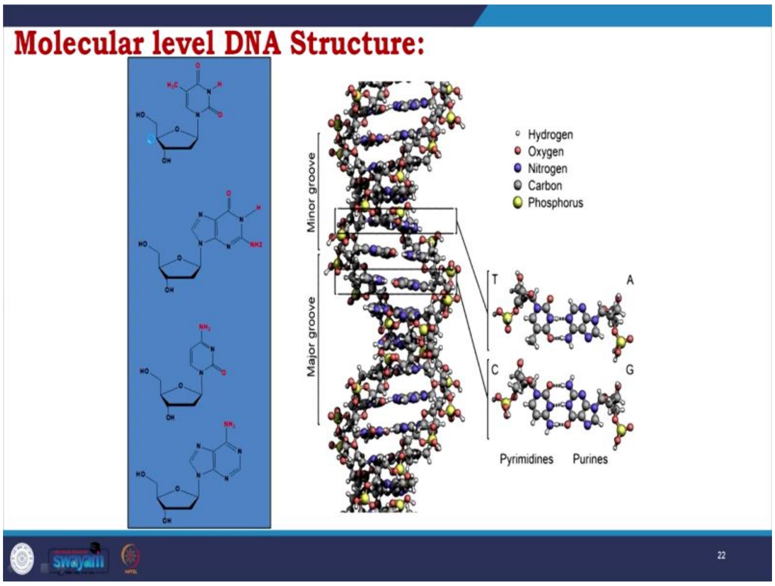
(Refer Slide Time: 39:31)



Information obtained from "DNA sequencing" would help us getting information in different levels to understand life

Information obtained from DNA sequencing would help us get information in different labels to understand life will know about RNA next level, and you also know about the protein.

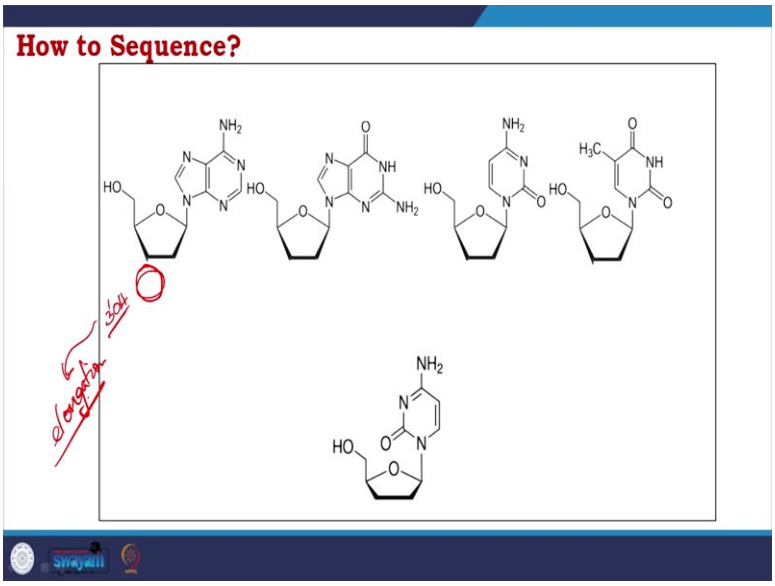
(Refer Slide Time: 39:52)



Molecular level DNA Structure:

But if you want to sequence, what would be the molecular mechanism to choose the selection. We already talked about the molecular level of DNA structure as we know that they differ only in nitrogenous bases.

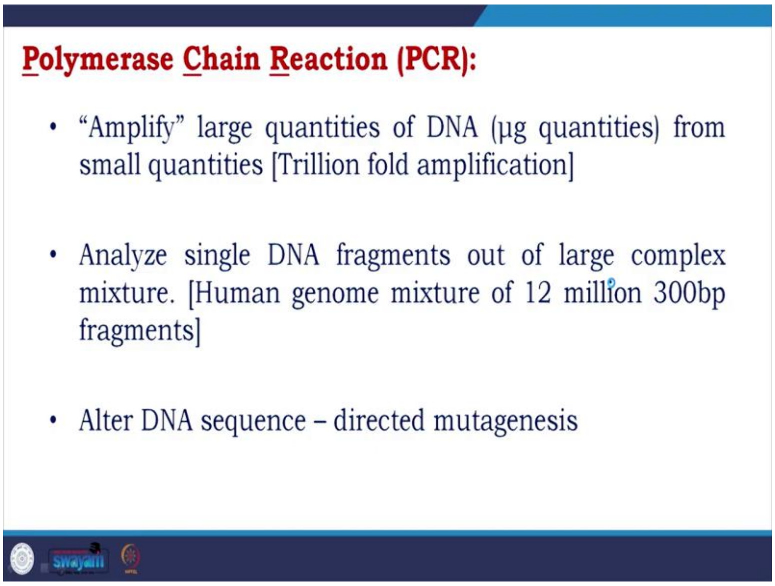
(Refer Slide Time: 40:09)



How to Sequence?

So we took the nitrogenous bases along with the sugar. In DNA, there is only one hydroxyl group. And if you take 3 Prime hydroxyl groups, this molecule would not continue the DNA synthesis process because this residue is essential for elongation. So somehow, if you could introduce a molecule like this instead of a molecule with a hydroxyl group, you could stop the synthesis of DNA.

(Refer Slide Time: 41:16)



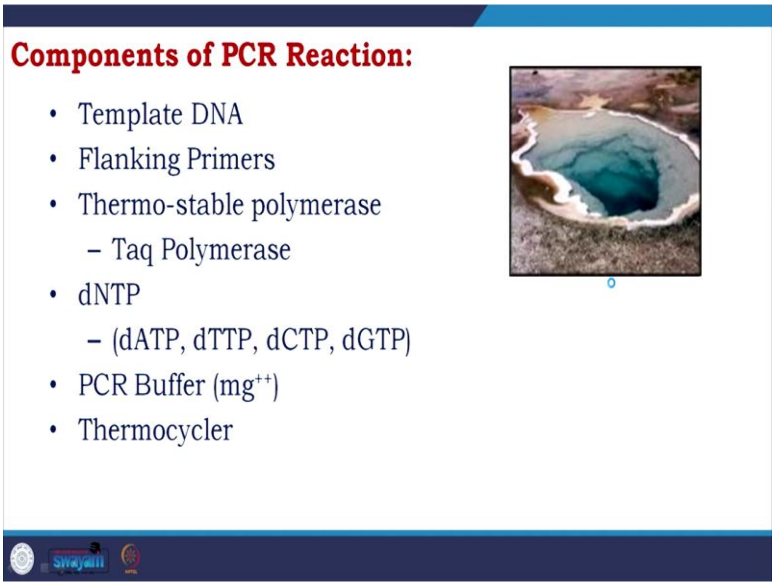
Polymerase Chain Reaction (PCR):

- "Amplify" large quantities of DNA (μg quantities) from small quantities [Trillion fold amplification]
- Analyze single DNA fragments out of large complex mixture. [Human genome mixture of 12 million 300bp fragments]
- Alter DNA sequence – directed mutagenesis

The DNA replication process was observed very keenly and based on that development. There is a process development called PCR (polymerase Chain Reaction). It amplifies large quantities of DNA in the microgram level to milligram from small quantity generally trillionfold of amplification.

To get a very tiny amount of DNA, and with the help of PCR now, you could get a handful to do your further experiments. Alter DNA sequence. If there is mutagenesis, you could have identified them. You want to do mutagenesis, you could perform that.

(Refer Slide Time: 42:26)



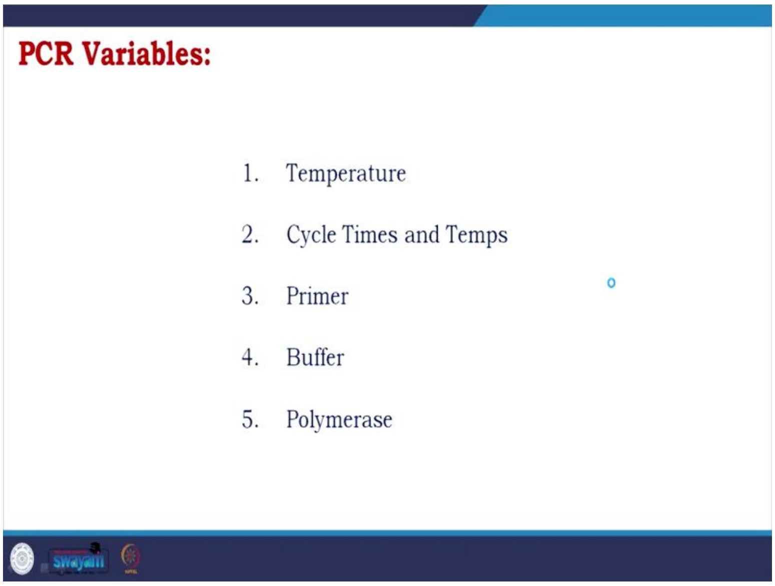
Components of PCR Reaction:

- Template DNA
- Flanking Primers
- Thermo-stable polymerase
  - Taq Polymerase
- dNTP
  - (dATP, dTTP, dCTP, dGTP)
- PCR Buffer (mg^{++})
- Thermocycler

What are the components of PCR reactions? You have to give the original DNA you want to elongate one template DNA. Primers, as we know now by our previous findings that DNA polymerase cannot work without the presence of Primers. Thermo stable polymerase: I already talked about DNA polymerase. In PCR, you have to provide the temperature changes and develop a temperature cycle, and protein being stable through non-covalent bonds could not be stable in that condition. So you need a Thermo Stable DNA polymerase, Tag polymerase from the organism Thermus Aquaticus. The finding of Taq polymerase is to recognize the Field of Biology by getting the instrumentation which is PCR.

So you need dNTPs there adenosine, thymine, cytosine, and guanine. You need a proper PCR buffer with magnesium. And ultimately, you need the instrumentation, which is called thermocycler.

(Refer Slide Time: 43:50)

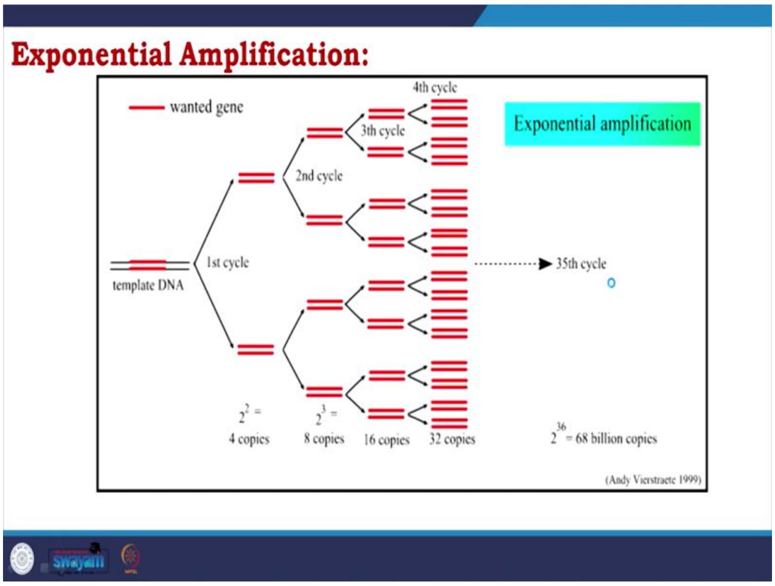


PCR Variables:

1. Temperature
2. Cycle Times and Temps
3. Primer
4. Buffer
5. Polymerase

There are variables in the PCR to optimize, which are temperature, cycle times and relative temperatures, primer, buffer, and the DNA polymerase. When it was started, we looked for a thermo stable enzyme. Later, we realized that the enzyme would be more or better proof-reading activity.

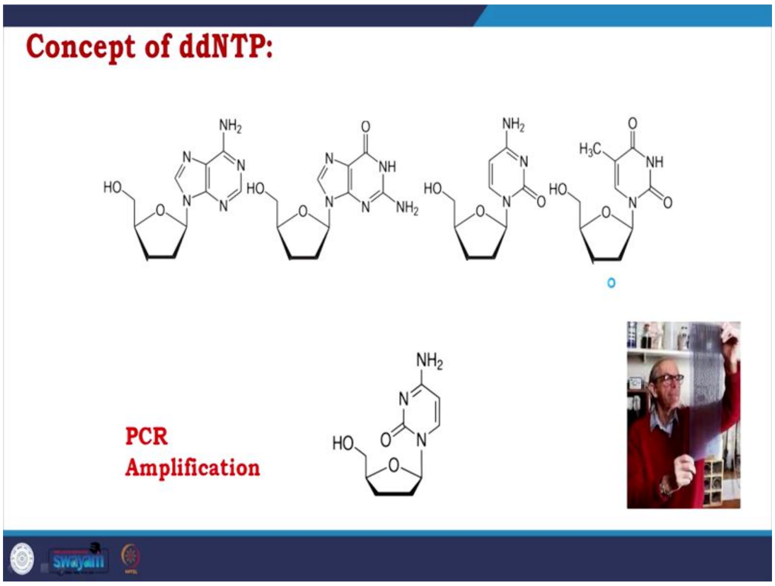
(Refer Slide Time: 44:43)



Exponential Amplification:

So ultimately, when you do that, there would be exponential amplification. If you have 30 cycles, you will get 2 trillion copies in theory, which is the revolution we have. We had DNA around, but now we could make a handful with a tiny amount of DNA.

(Refer Slide Time: 45:09)

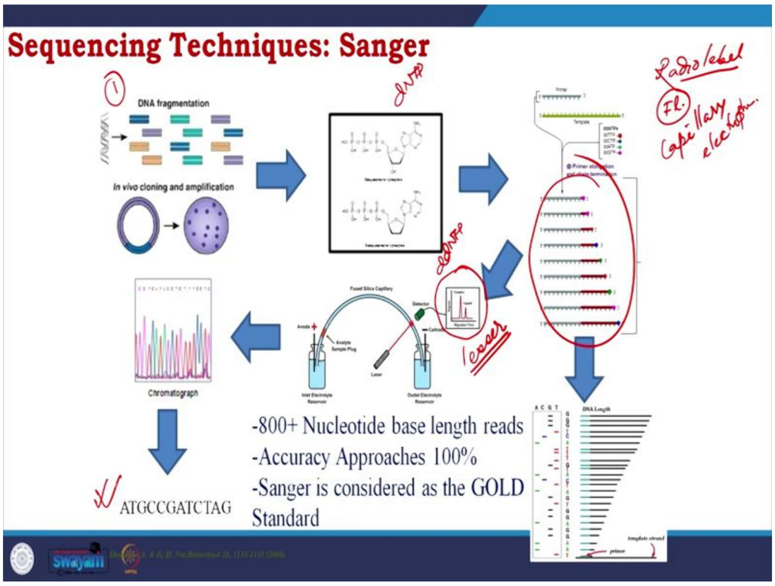


Concept of ddNTP:

PCR Amplification

So I already talked about the concept of this dihydroxy nucleotide triphosphate, which stops the reaction that, along with PCR amplification, leads us to something that we were dreaming of to read, came from Frederick Sanger known as Sanger Sequencing.

(Refer Slide Time: 45:36)



Sequencing Techniques: Sanger

-800+ Nucleotide base length reads

-Accuracy Approaches 100%

-Sanger is considered as the GOLD Standard

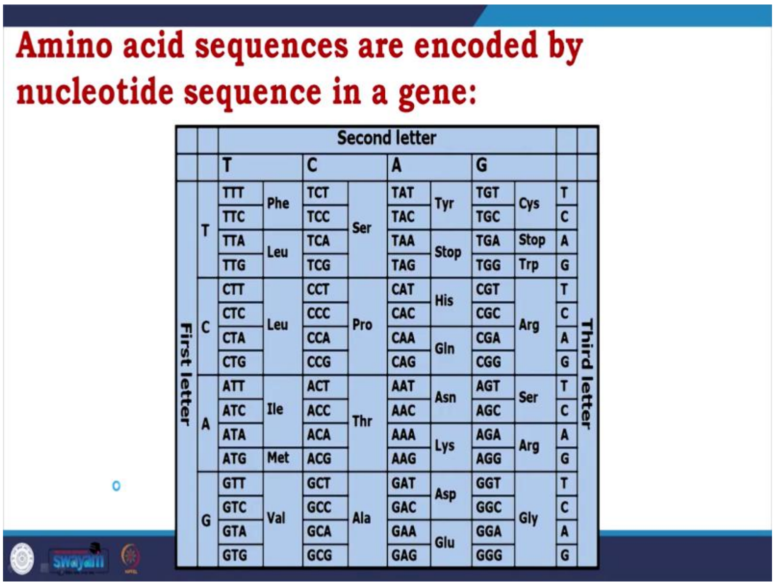
ATGCCGATCTAG

What is Sanger Sequencing? You have the first step where you have the DNA you make the DNA fragment then you add DNTP's as in PCR, but in addition to that you will add DDNTP's, then you continue the PCR so there would be a competition whenever the DDNTPS would enter in the DNA, it stops the reaction. So you get a series of different fragments. At the initial stage, the fragments were measured by radiolabel.

But with time, the fluorescent label came, followed by this one, which is Capillary electrophoresis. Capillary electrophoresis enhances the resolution much, much better, but more importantly, when it goes through the capillary, you start reading it by applying a laser.

So you hit with the laser the capability to get fluorescent signals different for different nucleotides, which would be recorded in a chromatogram. And from the chromatogram, you can read the sequence.

(Refer Slide Time: 50:30)

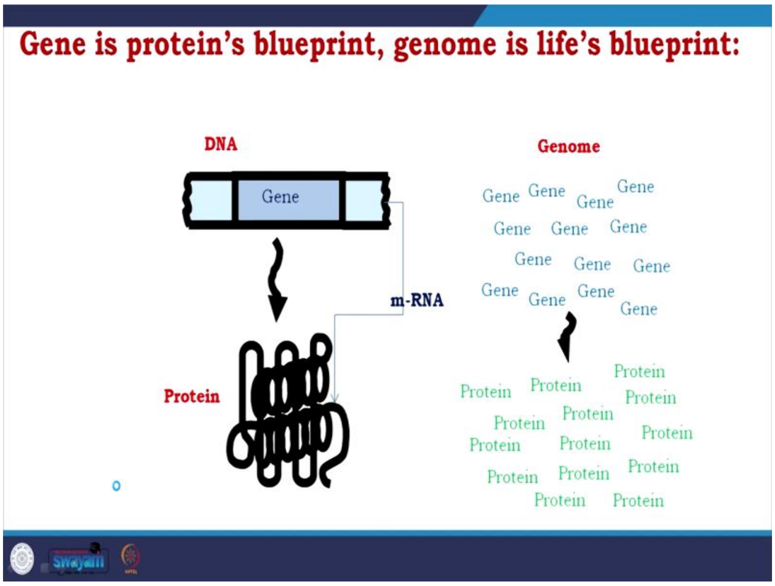


Amino acid sequences are encoded by nucleotide sequence in a gene:

		Second letter									
		T		C		A		G			
First letter	T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T	Third letter
		TTC		TCC		TAC		TGC		C	
		TTA	Leu	TCA		TAA	Stop	TGA	Stop	A	
		TTG		TCG		TAG		TGG	Trp	G	
	C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T	
		CTC		CCC		CAC		CGC		C	
		CTA		CCA		CAA	Gln	CGA		A	
		CTG		CCG		CAG		CGG		G	
	A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T	
		ATC		ACC		AAC		AGC		C	
		ATA		ACA		AAA	Lys	AGA	Arg	A	
		ATG	Met	ACG		AAG		AGG		G	
	G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T	
		GTC		GCC		GAC		GGC		C	
		GTA		GCA		GAA	Glu	GGA		A	
		GTG		GCG		GAG		GGG		G	

And if you see now and I talked about when you get the gene sequence you have the genetic code, you will get the protein sequence.

(Refer Slide Time: 50:41)



Gene is protein's blueprint, genome is life's blueprint:

So you get the DNA, the gene, you get to read the protein sequence. So the unraveling of the biology unraveling of life started with the finding and establishing PCR and Sanger Sequencing.

(Refer Slide Time: 51:06)