DNA Amplification and DNA sequencing

Class No-1

SEMESTER-VI

What is PCR?

Polymerase chain reaction (**PCR**) is a common laboratory technique used to make many copies (millions or billions!) of a particular region of DNA. This DNA region can be anything the experimenter is interested in. For example, it might be a gene whose function a researcher wants to understand, or a genetic marker used by forensic scientists to match crime scene DNA with suspects.

Typically, the goal of PCR is to make enough of the target DNA region that it can be analyzed or used in some other way. For instance, DNA amplified by PCR may be sent for sequencing, visualized by gel electrophoresis, or clonedinto a plasmid for further experiments.

PCR is used in many areas of biology and medicine, including molecular biology research, medical diagnostics, and even some branches of ecology.

Taq polymerase

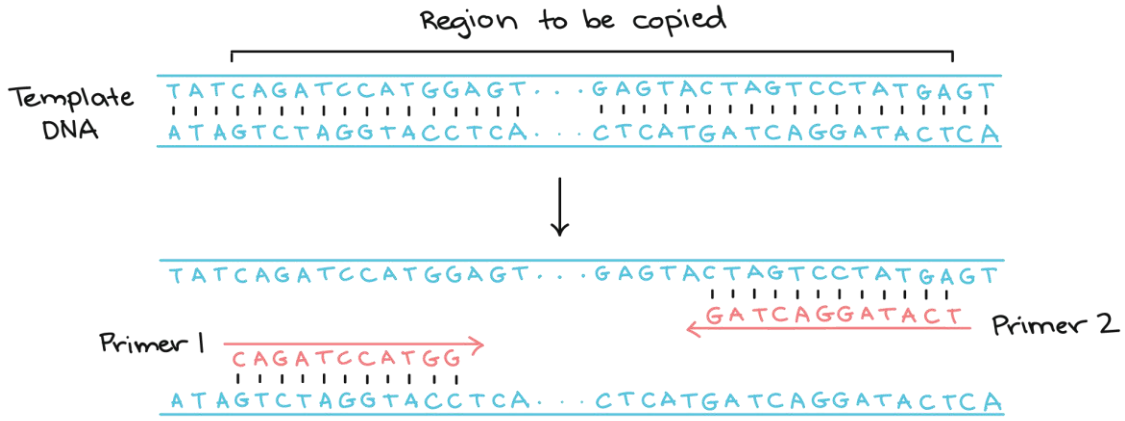
Like DNA replication in an organism, PCR requires a DNA polymerase enzyme that makes new strands of DNA, using existing strands as templates. The DNA polymerase typically used in PCR is called ***Taq* polymerase**, after the heat-tolerant bacterium from which it was isolated (***T****hermus* ***aq****uaticus*).

T. aquaticus lives in hot springs and hydrothermal vents. Its DNA polymerase is very heat-stable and is most active around 70°C (a temperature at which a human or *E. coli* DNA polymerase would be nonfunctional). This heat-stability makes Taq polymerase ideal for PCR. As we'll see, high temperature is used repeatedly in PCR to **denature** the template DNA, or separate its strands.

PCR primers

Like other DNA polymerases, *Taq* polymerase can only make DNA if it's given a **primer**, a short sequence of nucleotides that provides a starting point for DNA synthesis. In a PCR reaction, the experimenter determines the region of DNA that will be copied, or amplified, by the primers she or he chooses.

PCR primers are short pieces of single-stranded DNA, usually around 202020nucleotides in length. Two primers are used in each PCR reaction, and they are designed so that they flank the target region (region that should be copied). That is, they are given sequences that will make them bind to opposite strands of the template DNA, just at the edges of the region to be copied. The primers bind to the template by complementary base pairing.

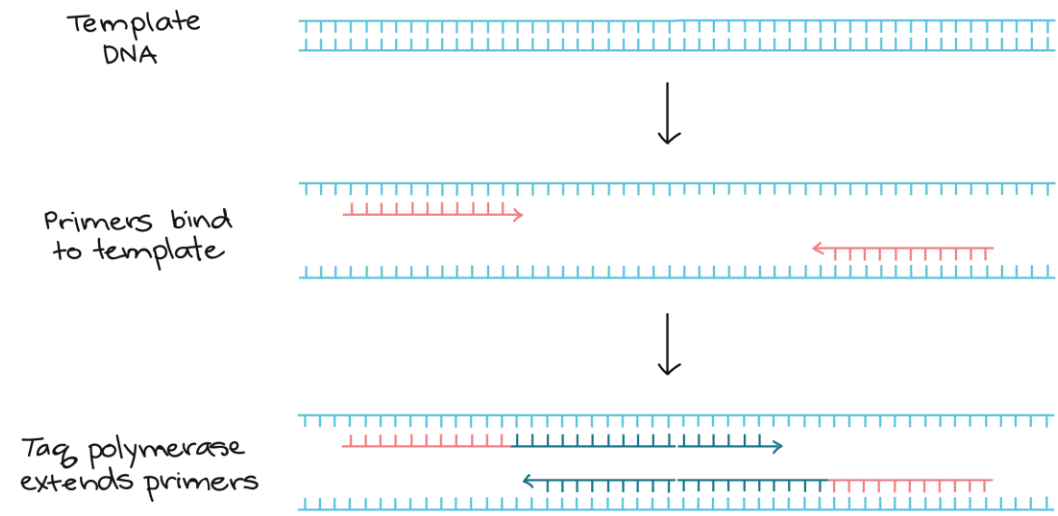


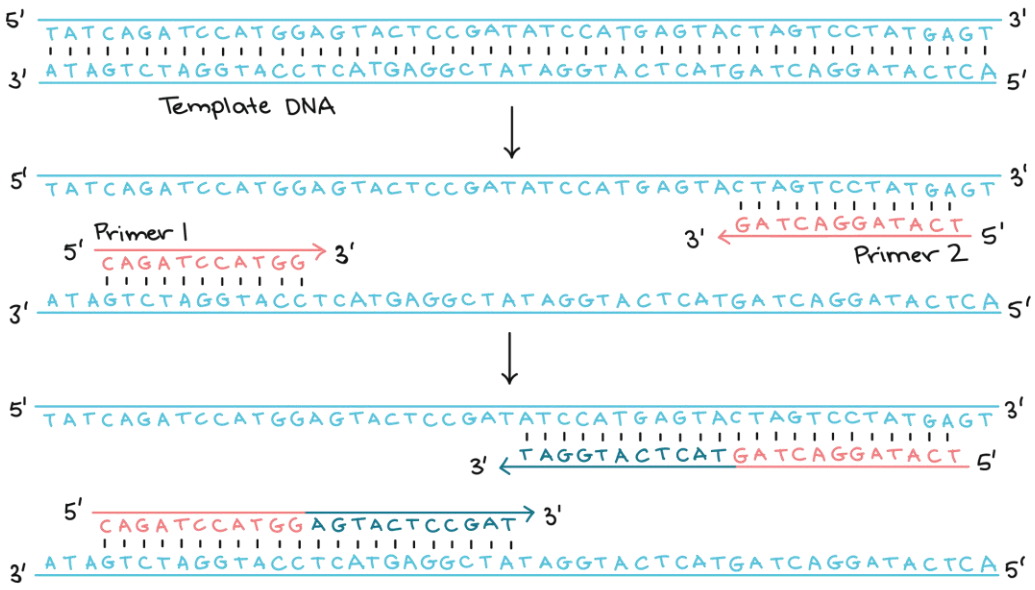
Template DNA:

5' TATCAGATCCATGGAGT...GAGTACTAGTCCTATGAGT 3' 3' ATAGTCTAGGTACCTCA...CTCATGATCAGGATACTCA 5'

Primer 1: 5' CAGATCCATGG 3' Primer 2:

When the primers are bound to the template, they can be extended by the polymerase, and the region that lies between them will get copied.





The steps of PCR

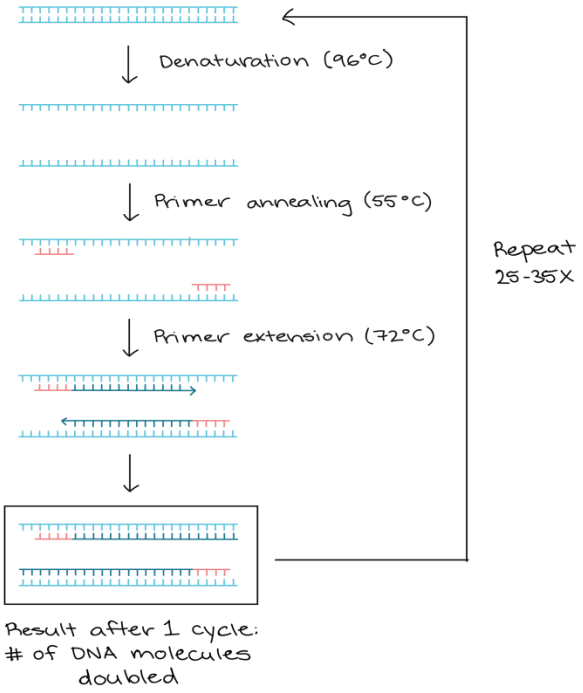
The key ingredients of a PCR reaction are *Taq* polymerase, primers, template DNA, and nucleotides (DNA building blocks). The ingredients are assembled in a tube, along with cofactors needed by the enzyme, and are put through repeated cycles of heating and cooling that allow DNA to be synthesized.

The basic steps are:

1. **Denaturation** (96°C): Heat the reaction strongly to separate, or denature, the DNA strands. This provides single-stranded template for the next step.

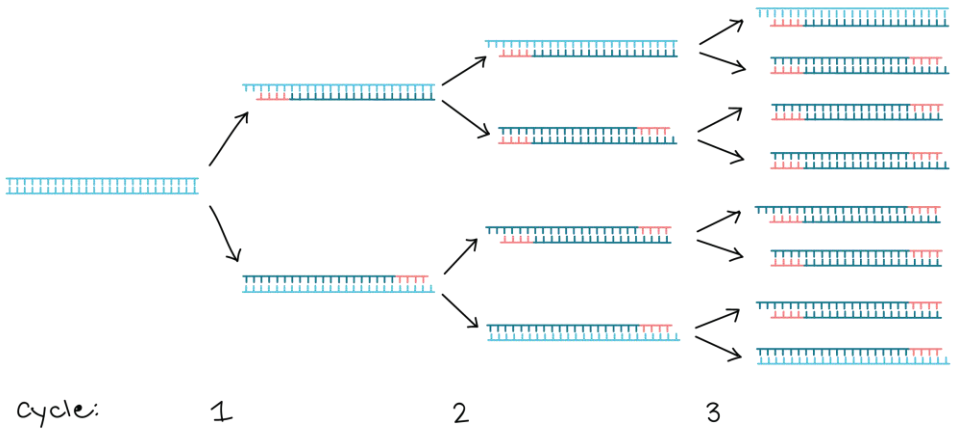
2. **Annealing** (555555 - 656565°C): Cool the reaction so the primers can bind to their complementary sequences on the single-stranded template DNA.

3. **Extension** (72°C): Raise the reaction temperatures so *Taq* polymerase extends the primers, synthesizing new strands of DNA.



This cycle repeats 252525 - 353535 times in a typical PCR reaction, which generally takes 222 - 444 hours, depending on the length of the DNA region being copied. If the reaction is efficient (works well), the target region can go from just one or a few copies to billions.

That's because it's not just the original DNA that's used as a template each time. Instead, the new DNA that's made in one round can serve as a template in the next round of DNA synthesis. There are many copies of the primers and many molecules of *Taq* polymerase floating around in the reaction, so the number of DNA molecules can roughly double in each round of cycling. This pattern of exponential growth is shown in the image below.



Advantages of Polymerase Chain Reaction:

PCR is so sensitive that DNA sequences present in an individual cell can be amplified. The isolation and amplification of a specific DNA sequence by PCR is faster and less technically difficult than traditional cloning methods using recombinant DNA techniques.

Applications of PCR

- Identification and characterization of infectious agents
  - Direct detection of microorganisms in patient specimens
  - Identification of microorganisms grown in culture
  - Detection of antimicrobial resistance
  - Investigation of strain relatedness of pathogen of interest
- Genetic fingerprinting (forensic application/paternity testing)
- Detection of mutation (investigation of genetic diseases)
- Cloning genes
- PCR sequencing

Limitations of PCR

- **Need for target DNA sequence information**
  Primer Designing for unexplored ones.
  Boundary regions of DNA to be amplified must be known.
- **Infidelity of DNA replication.**
  Taq Pol – no Proof reading mech – Error 40% after 20 cycles
- **Short size and limiting amounts of PCR product**
  Up to 5kb can be easily amplified .
  Up to 40kb can be amplified with some modifications.
  Cannot amplify gene >100kb
  Cannot be used in genome sequencing projects.

RT-PCR:

RT-PCR (Reverse Transcription PCR) is a method used to amplify, isolate or identify a known sequence from a cellular or tissue RNA. The PCR reaction is preceded by a reaction using reverse transcriptase to convert RNA to cDNA.

RT-PCR is widely used in expression profiling, to determine the expression of a gene or to identify the sequence of an RNA transcript, including transcription start and termination sites and, if the genomic DNA sequence of a gene is known, to map the location of exons and introns in the gene. The 5′ end of a gene (corresponding to the transcription start site) is typically identified by a RT-PCR method, named RACE-PCR, short for Rapid Amplification of cDNA Ends.

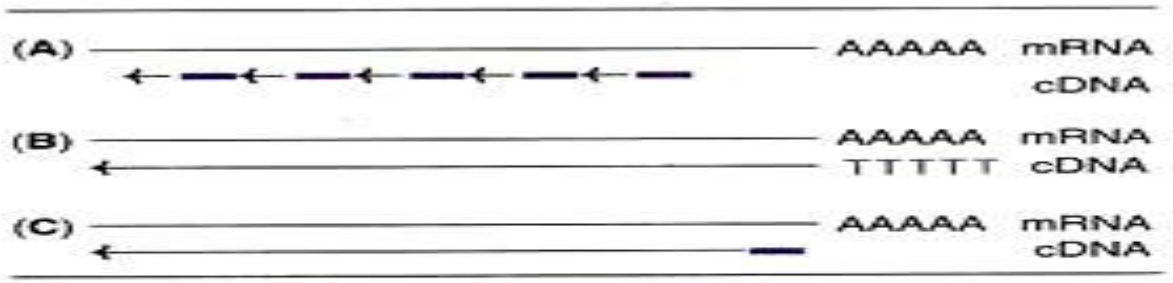
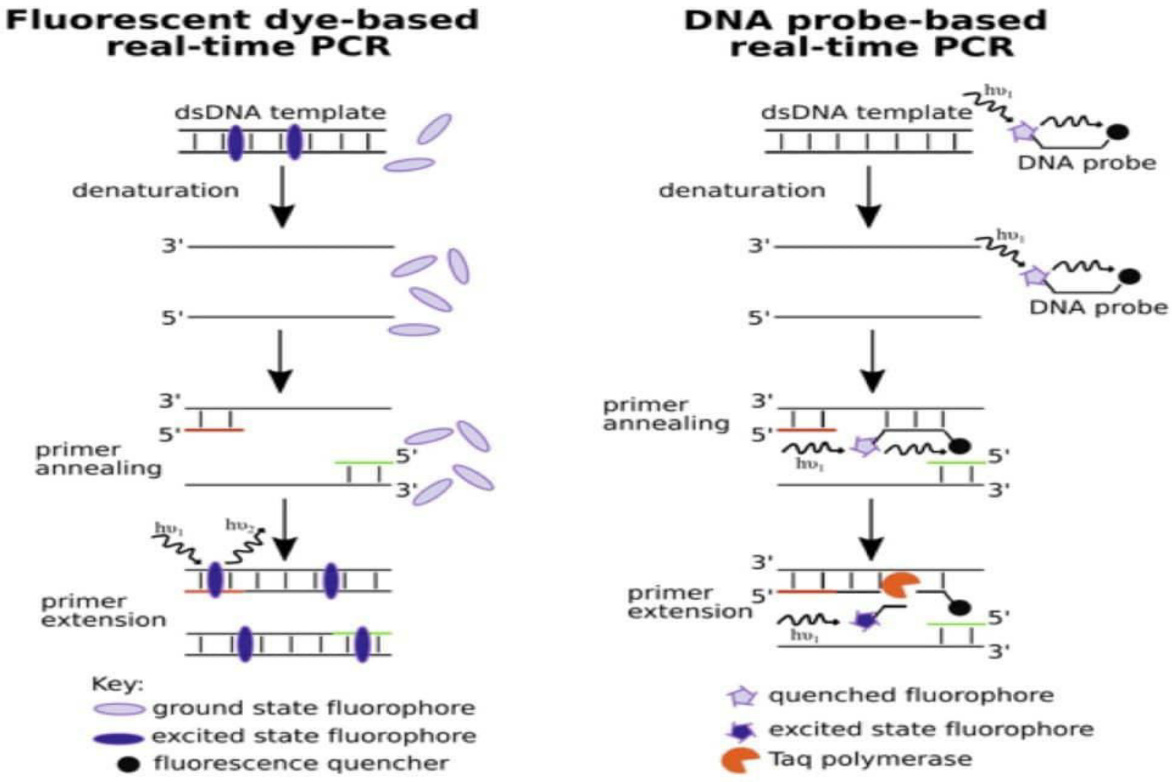


Fig. 8.5 : Synthesis of first strand of cDNA in reverse transcription-PCR with different primers (A) Random primers (B) oligo dT primer (C) Sequence specific primer (Note : The primers are shown in colour).

Real-Time Quantitative PCR:

The quantification of PCR products in different cycles is not as simple as projected by theoretical considerations (Table 8.1). In practice, large variations occur. The most commonly used technique for measuring the quantity of PCR is by employing a fluorescence compound like eithidium bromide.

The principle is that the double-stranded DNA molecules bind to ethidium bromide which emit fluorescence that can be detected, and DNA quantified. The synthesis of genes by PCR and the role of PCR in site-directed mutagenesis are described elsewhere.



Sanger's method of gene sequencing

- Sanger's method of gene sequencing is also known as dideoxy chain termination method. It generates nested set of labelled fragments from a template strand of DNA to be sequenced by replicating that template strand and interrupting the replication process at one of the four bases.
- Four different reaction mixtures are produced that terminates in A. T. G or C respectively.

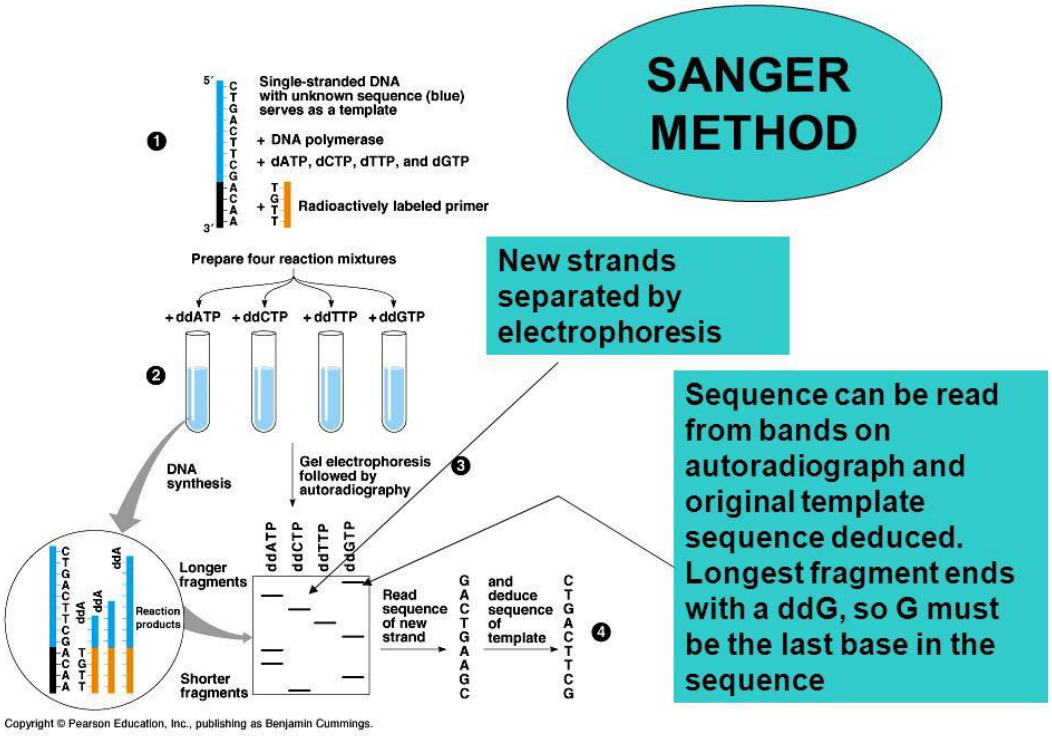
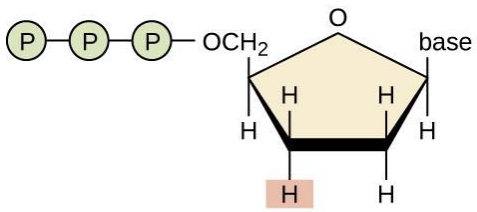


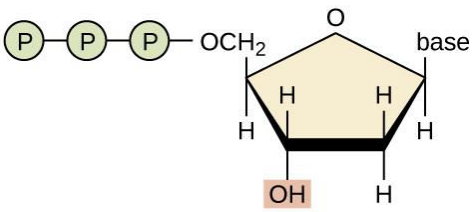
Figure: Diagrammatic representation of Sanger sequencing

Principle

- A DNA primer is attached by hybridization to the template strand and deoxynucleosides triphosphates (dNTPPs) are sequentially added to the primer strand by DNA polymerase.
- The primer is designed for the known sequences at 3' end of the template strand.
- M13 sequences is generally attached to 3' end and the primer of this M13 is made.
- The reaction mixture also contains dideoxynucleoside triphosphate (ddNTPs) along with usual dNTPs.
- If during replication ddNTPs is incorporated instead of usual dNTPs in thegrowing DNA strand then the replication stops at that nucleotide.



dideoxynucleotide (ddNTP)

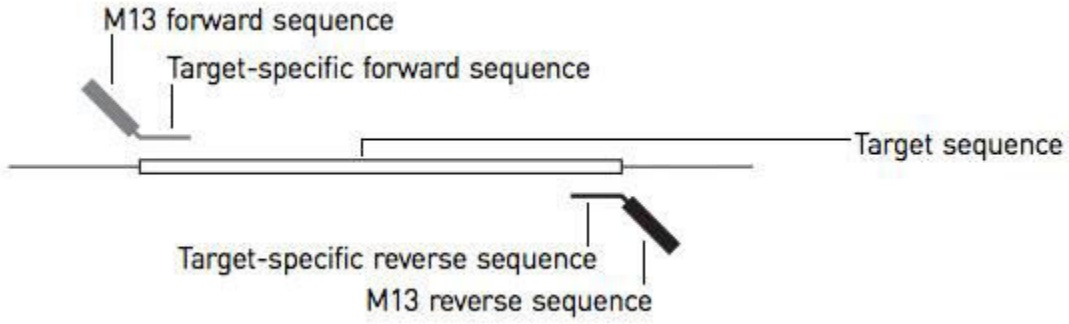


deoxynucleotide (dNTP)

- The ddNTPs are analogue of dNTPs
- ddNTPs lacks hydroxyl group (-OH) at c3 of ribose sugar, so it cannot make phosphodiester bond with nest nucleotide, thus terminates the nucleotide chain
- Respective ddNTPs of dNTPs terminates chain at their respective site. For example ddATP terminates at A site. Similarly ddCTP, ddGTP and ddTTP terminates at C, G and T site respectively.

Procedure

1. Template preparation:



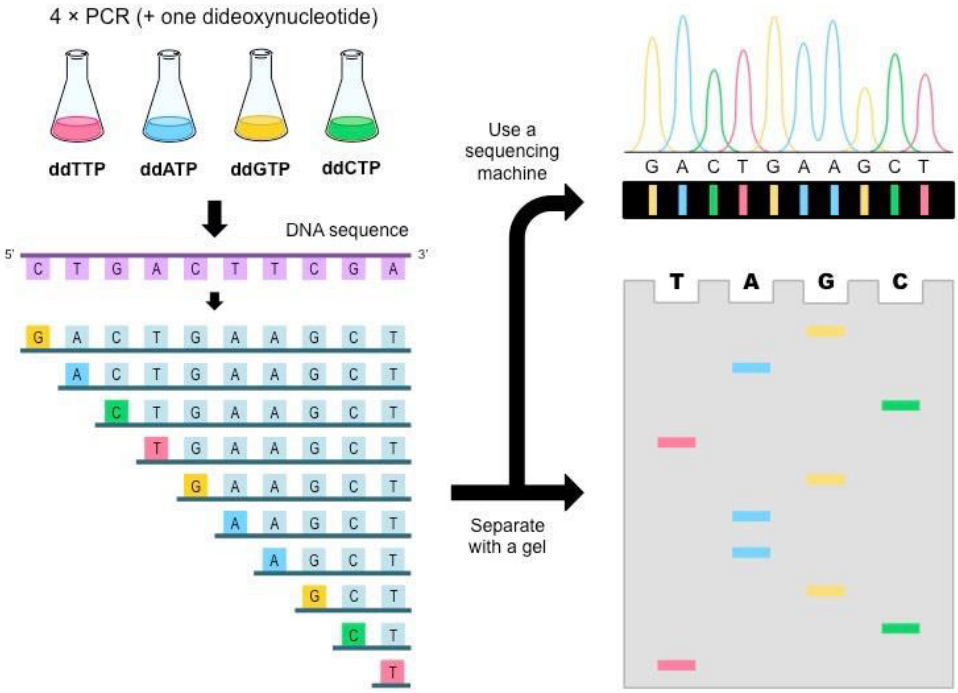
m13-forward-sequence

- Copies of template strand to be sequenced must be prepared with short known sequences at 3' end of the template strand.
- A DNA primere is essential to initiate replication of template , so primer preparation of known sequences at 3'end is always required.
- For this purpose a single stranded cloning vector M13 is flanked with template strand at 3'end which serves as binding site for primer.

2. Generation of nested set of labelled fragments:

- Copies of each template is divided into four batches and each batch is used for different replication reaction.
- Copies of standard primer and DNA polymerase I are used in all four batches.

- To synthesize fragments that terminates at A, ddATP is added to the reaction mixture on batch I along with dATP, dTTP,dCTP and dGTP, standard primer and DNA polymerase I.
- Similarly, to generate, all fragments that terminates at C, G and T, the respective ddNTPs ie ddCTP,ddGTP and ddTTP are added respectively to different reaction mixture on different batch along with usual dNTPs.



3. Electrophoresis and gel reading:

- The reaction mixture from four batches are loaded into four different well on polyacrylamide gel and electrophoresed.
- The autoradiogramof the gel is read to determine the order of bases of complementary strand to that of template strand.
- The band of shortest fragments are at the bottom of autoradiogram so that the sequences of complementary strand is read from bottom to top.

Automatic DNA Sequencer:

A variant of the above dideoxy-method was developed, which allowed the production of automatic sequencers. In this new approach, different fluorescent dyes are tagged either to the oligonucleotide primer (dye primers) in each of the four reaction tubes (blue for A, red for C, etc), or to each of the four ddNTPs (dye terminators) used in a single reaction tube: when four tubes are used, they are pooled.

After the PCR reaction is over, the reaction mixture is subjected to separation of synthesized fragments through electrophoresis (Fig. 23.9). Depending upon the electrophoretic system used, whether slab gel electrophoresis or capillary electrophoresis, following two types of automatic sequencing systems have been designed.

Primer walking (or Directed Sequencing) is a sequencing method of choice for sequencing DNA fragments between 1.3 and 7 kilobases. Such fragments are too long to be sequenced in a single sequence read using the chain termination method. This method works by dividing the long sequence into several consecutive short ones. The DNA of interest may be a plasmid insert, a PCR product or a fragment representing a gap when sequencing a genome. The term "primer walking" is used where the main aim is to sequence the genome. The term "**chromosome walking**" is used instead when the sequence is known but there is no clone of a gene. For example, the gene for a disease may be located near a specific marker such as an RFLP on the sequence.

The fragment is first sequenced as if it were a shorter fragment. Sequencing is performed from each end using either universal primers or specifically designed ones. This should identify the first 1000 or so bases. In order to completely sequence the region of interest, design and synthesis of new primers (complementary to the final 20 bases of the known sequence) is necessary to obtain contiguous sequence information.

Process

The overall process is as follows:

1. A primer that matches the beginning of the DNA to sequence is used to synthesize a short DNA strand adjacent to the unknown sequence, starting with the primer (see PCR).
2. The new short DNA strand is sequenced using the chain termination method.
3. The end of the sequenced strand is used as a primer for the next part of the long DNA sequence, hence the term "walking".

The method can be used to sequence entire chromosomes (hence "chromosome walking").[1] Primer walking was also the basis for the development of shotgun sequencing, which uses random primers instead of specifically chosen ones.

Shotgun sequencing

In genetics, **shotgun sequencing** is a method used for sequencing random DNA strands. It is named by analogy with the rapidly expanding, quasi-random firing pattern of a shotgun.

The chain termination method of DNA sequencing ("Sanger sequencing") can only be used for short DNA strands of 100 to 1000 base pairs. Due to this size limit, longer sequences are subdivided into smaller fragments that can be sequenced separately, and these sequences are assembled to give the overall sequence.

There are two principal methods for this fragmentation and sequencing process. Primer walking (or "chromosome walking") progresses through the entire strand piece by piece, whereas shotgun sequencing is a faster but more complex process that uses random fragments.

In shotgun sequencing, DNA is broken up randomly into numerous small segments, which are sequenced using the chain termination method to obtain *reads*. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

Shotgun sequencing was one of the precursor technologies that was responsible for enabling full genome sequencing.

Example

For example, consider the following two rounds of shotgun reads:

Strand	**Sequence**
Original	AGCATGCTGCAGTCATGCTTAGGCTA
First shotgun sequence	AGCATGCTGCAGTCATGCT------- ----------TAGGCTA
Second shotgun sequence	AGCATG---------- ------CTGCAGTCATGCTTAGGCTA
Reconstruction	AGCATGCTGCAGTCATGCTTAGGCTA

In this extremely simplified example, none of the reads cover the full length of the original sequence, but the four reads can be assembled into the original sequence using the overlap of their ends to align and order them. In reality, this process uses enormous amounts of information that are rife with ambiguities and sequencing errors. Assembly of complex genomes is additionally complicated by the great abundance of repetitive sequences, meaning similar short reads could come from completely different parts of the sequence.

Many overlapping reads for each segment of the original DNA are necessary to overcome these difficulties and accurately assemble the sequence. For example, to complete the Human Genome Project, most of the human genome was sequenced at 12X or greater *coverage*; that is, each base in the final sequence was present on average in 12 different reads. Even so, current methods have failed to isolate or assemble reliable sequence for approximately 1% of the (euchromatic) human genome, as of 2004

Next-generation sequencing

The classical shotgun sequencing was based on the Sanger sequencing method: this was the most advanced technique for sequencing genomes from about 1995–2005. The shotgun strategy is still applied today, however using other sequencing technologies, called next-generation sequencing. These technologies produce shorter reads (anywhere from 25–500bp) but many hundreds of thousands or millions of reads in a relatively short time (on the order of a day). This results in high coverage, but the assembly process is much more computationally intensive. These technologies are vastly superior to Sanger sequencing due to the high volume of data and the relatively short time it takes to sequence a whole genome.