Subject Name: Microbiology

Semester: IV

Name of the Teacher:

Dr. Kuntal Kanti Goswami

Topic:

DNA Amplification and DNA Sequencing (CC10; Unit 4)

B.Sc (HONOURS) MICROBIOLOGY (CBCS STRUCTURE)

CC-10: DNA Amplification and DNA Sequencing (THEORY)

SEMESTER -IV

Basics of Polymerase Chain Reaction (PCR):

The polymerase chain reaction (PCR) is used to amplify a sequence of DNA using a pair of oligonucleotide primers each complementary to one end of the DNA target sequence. These are extended towards each other by a thermostable DNA polymerase in a reaction cycle of three steps: denaturation, primer annealing and polymerization.

The PCR cycle: In the first cycle, the target DNA is separated into two strands by heating to 95°C typically for around 60 seconds. The temperature is reduced to around 55°C (for about 30 sec) to allow the primers to anneal to the template DNA. The actual temperature depends on the primer lengths and sequences. After annealing, the temperature is increased to 72°C (for 60–90 sec) for optimal polymerization which uses up dNTPs in the reaction mix and requires Mg2+. In the first polymerization step, the target is copied from the primer sites for various distances on each target molecule until the beginning of cycle 2, when the reaction is heated to 95°C again which denatures the newly synthesized molecules. In the second annealing step, the other primer can bind to the newly synthesized strand and during polymerization can only copy till it reaches the end of the first primer. Thus at the end of cycle 2, some newly synthesized molecules of the correct length exist, though these are base paired to variable length molecules. In subsequent cycles, these soon outnumber the variable length molecules and increase two-fold with each cycle. If PCR was 100% efficient, one target molecule would become 2n after n cycles. In practice, 20–40 cycles are commonly used.

Template: Because of the extreme amplification achievable, it has been demonstrated that PCR can sometimes amplify as little as one molecule of starting template. Therefore, any source of DNA that provides one or more target molecules can in principle be used as a template for PCR. This includes DNA prepared from blood, sperm or any other tissue, from older forensic specimens, from ancient biological samples or in the laboratory from bacterial colonies or phage plaques as well as purified DNA. Whatever the source of template DNA,

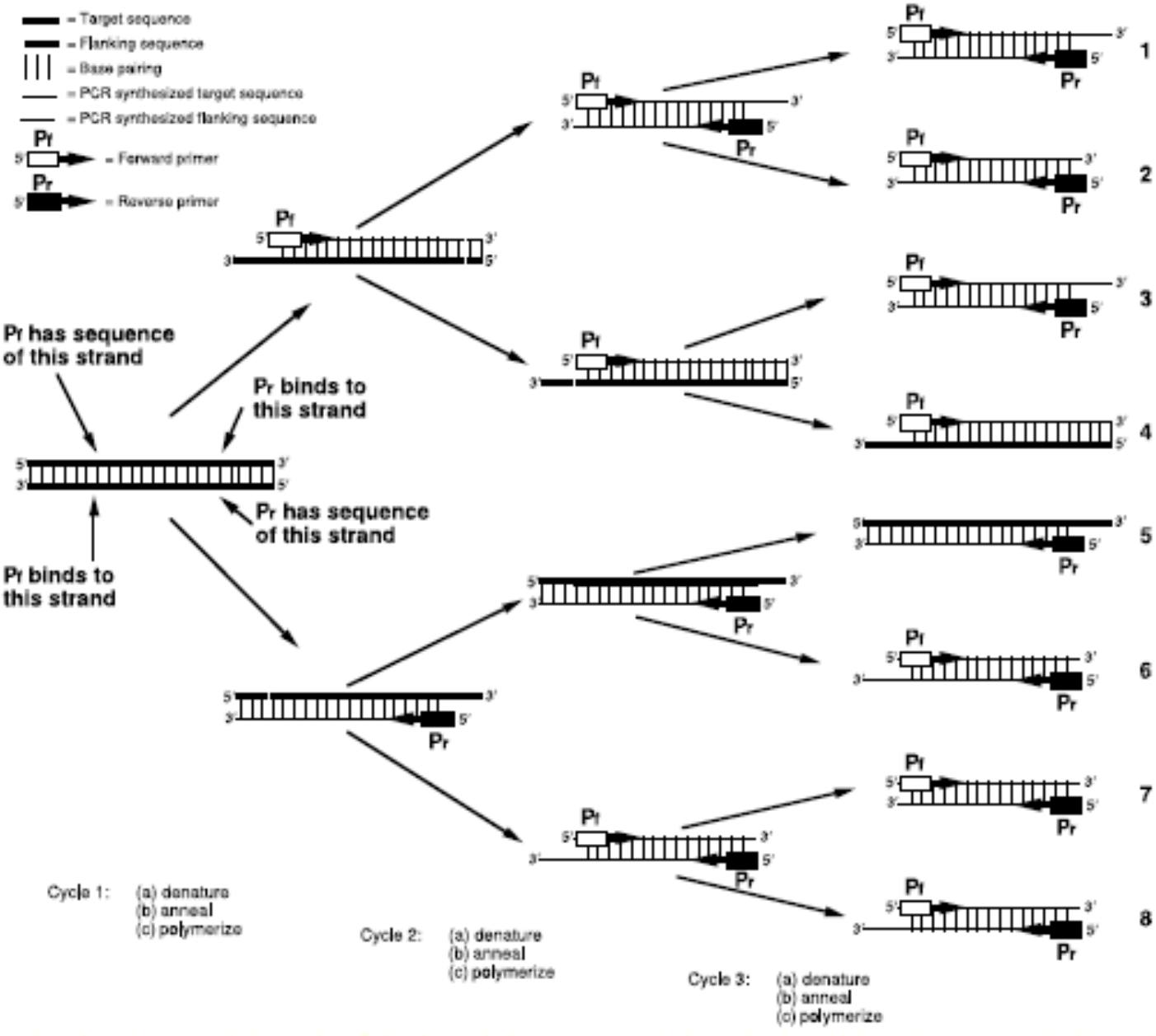
PCR can only be applied if some sequence information is known so that primers can be designed.

Primers: Each one of a pair of PCR primers needs to be about 18–30 nt long and to have similar G+C content so that they anneal to their complementary sequences at similar temperatures. For short oligonucleotides (<25 nt), the annealing temperature (in °C) can be calculated using the formula: $Tm = (4 \times [G + C]) + (2 \times [A + T])°C$ where Tm is the melting temperature and the annealing temperature is approximately 3–5°C lower. The primers are designed to anneal on opposite strands of the target sequence so that they will be extended towards each other by addition of nucleotides to their 3′-ends. Short target sequences amplify more easily, so often this distance is less than 500 bp, but, with optimization, PCR can amplify fragments over 10 kb in length. If the DNA sequence being amplified is known, then primer design is relatively easy. The region to be amplified should be inspected for two suitable sequences of about 20 nt with a similar G+C content, either side of the region to be amplified. If the PCR product is to be cloned, it is sensible to include the sequence of unique restriction enzyme sites within the 5′-ends of the primers. If the DNA sequence of the target is not known, for example when trying to clone a cDNA for a protein for which there is only some limited amino acid sequence available, then primer design is more difficult. For this, degenerate primers are designed using the genetic code to work out what DNA sequences would encode the known amino acid sequence. PCR using degenerate oligonucleotide primers is sometimes called DOP-PCR.

Enzymes: Thermostable DNA polymerases which have been isolated and cloned from a number of thermophilic bacteria are used for PCR. The most common is Taq polymerase from *Thermus aquaticus*. It survives the denaturation step of 95°C for 1–2min, having a half-life of more than 2 h at this temperature. Because it has no associated 3′ to 5′ proofreading exonuclease activity, Taq polymerase is known to introduce errors when it copies DNA – roughly one per 250 nt polymerized. For this reason, other thermostable DNA polymerases with greater accuracy are used for certain applications.

DNA polymerase	Source
Tma	*Thermotoga maritima*
Deep Vent™	*Pyrococcus* sp.
Tli	*Thermococcus litoralis*
Pfu	*Pyrococcus furiosus*
Pwo	*Pyrococcus woesi*

Thermostable DNA Polymerase with Proofreading activity



Factors Affecting PCR

The specificity of the PCR depends crucially upon the primers. The following factors are important in choosing effective primers.

• Primers should be 17 to 30 nucleotides in length.

• A GC content of about 50% is ideal. For primers with a low GC content, it is desirable to choose a long primer so as to avoid a low melting temperature.

• Sequences with long runs (i.e. more than three or four) of a single nucleotide should be avoided.

• Primers with significant secondary structure are undesirable.

• There should be no complementarity between the two primers. The great majority of primers which conform with these guidelines can be made to work, although not all comparable primer sets are equally effective even under optimized conditions

Reverse Transcriptase PCR (RT-PCR):

The thermostable polymerase used in the basic PCR requires a DNA template and hence is limited to the amplification of DNA samples. There are numerous instances in which the amplification of RNA would be preferred. For example, in analyses involving the diffierential expression of genes in tissues during development or the cloning of DNA derived from an mRNA (complementary DNA or cDNA), particularly a rare mRNA. In order to apply PCR methodology to the study of RNA, the RNA sample must first be reverse-transcribed to cDNA to provide the necessary DNA template for the thermostable polymerase. This process is called reverse transcription (RT), hence the name RT-PCR.

Avian myeloblastosis virus (AMV) or Moloney murine leukaemia virus (MuLV) reverse transcriptases are generally used to produce a DNA copy of the RNA template. Various strategies can be adopted for first-strand cDNA synthesis.

Real Time PCR: The amount of product that is synthesized during a set number of cycles of a PCR depends on the number of DNA molecules that are present in the starting mixture. If there are only a few DNA molecules at the beginning of the PCR then relatively little product will be made, but if there are many starting molecules then the product yield will be higher. This relationship enables PCR to be used to quantify the number of DNA molecules present in an extract.

Number of short products synthesized after 25 cycles of PCR with different numbers of starting molecule.

NUMBER OF STARTING MOLECULES	NUMBER OF SHORT PRODUCTS
1	4,194,304
2	8,388,608
5	20,971,520
10	41,943,040
25	104,857,600
50	209,715,200
100	419,430,400

In quantitative PCR (qPCR) the amount of product synthesized during a test PCR is compared with the amounts synthesized during PCRs with known quantities of starting DNA. In the early procedures, agarose gel electrophoresis was used to make these comparisons. After staining the gel, the band intensities were examined to identify the control PCR whose product was most similar to that of the test. Although easy to perform, this type of qPCR is imprecise, because large differences in the amount of starting DNA give relatively small differences in the band intensities of the resulting PCR products.

Today, quantification is carried out by real-time PCR, a modification of the standard PCR technique in which synthesis of the product is measured over time, as the PCR proceeds through its series of cycles. There are two ways of following product synthesis in real time:

- A dye (eg. SYBR green) that gives a fluorescent signal when it binds to double-stranded DNA can be included in the PCR mixture. This method measures the total amount of double-stranded DNA in the PCR at any particular time, which may over-estimate the actual amount of the product because sometimes the primers anneal to one another in various non-specific ways, increasing the amount of double-stranded DNA that is present. **(Figure 1)**
- A short oligonucleotide called a reporter probe, which gives a fluorescent signal when it hybridizes to the PCR product, can be used. Because the probe only hybridizes to the PCR product, this method is less prone to inaccuracies caused by primer-primer

annealing. Each probe molecule has pair of labels. A fluorescent dye is attached to one end of the oligonucleotide, and a quenching compound, which inhibits the fluorescent signal, is attached to the other end. Normally there is no fluorescence because the oligonucleotide is designed in such a way that its two ends base pair to one another, placing the quencher next to the dye. Hybridization between the oligonucleotide and the PCR product disrupts this base pairing, moving the quencher away from the dye and enabling the fluorescent signal to be generated.**(Figure 2)**

Both systems enable synthesis of the PCR product to be followed by measuring the fluorescent signal. Quantification again requires comparison between test and control PCRs, usually by identifying the stage in the PCR at which the amount of fluorescent signal reaches a pre-set threshold. The more rapidly the threshold is reached, the greater the amount of DNA in the starting mixture.

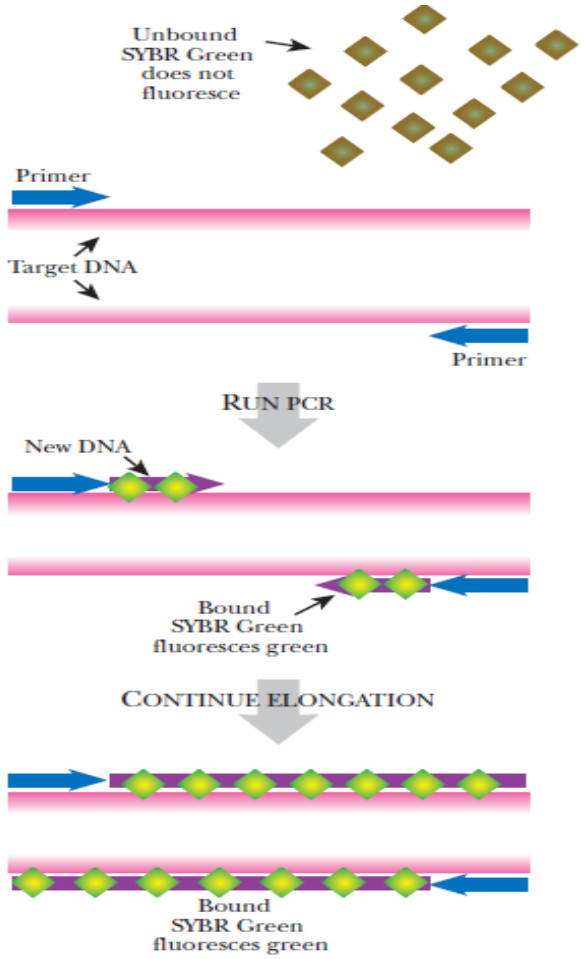


Figure 1: Real Time Fluorescent PCR with SYBR green

When the fluorescent probe SYBR® green is present during a PCR reaction, it binds to the double-stranded PCR product and emits light at 520 nm. The SYBR® Green dye only fluoresces when bound to DNA. Hence, the amount of fluorescence correlates with the amount of PCR product produced. This allows the accumulation of PCR product to be followed through many cycles.

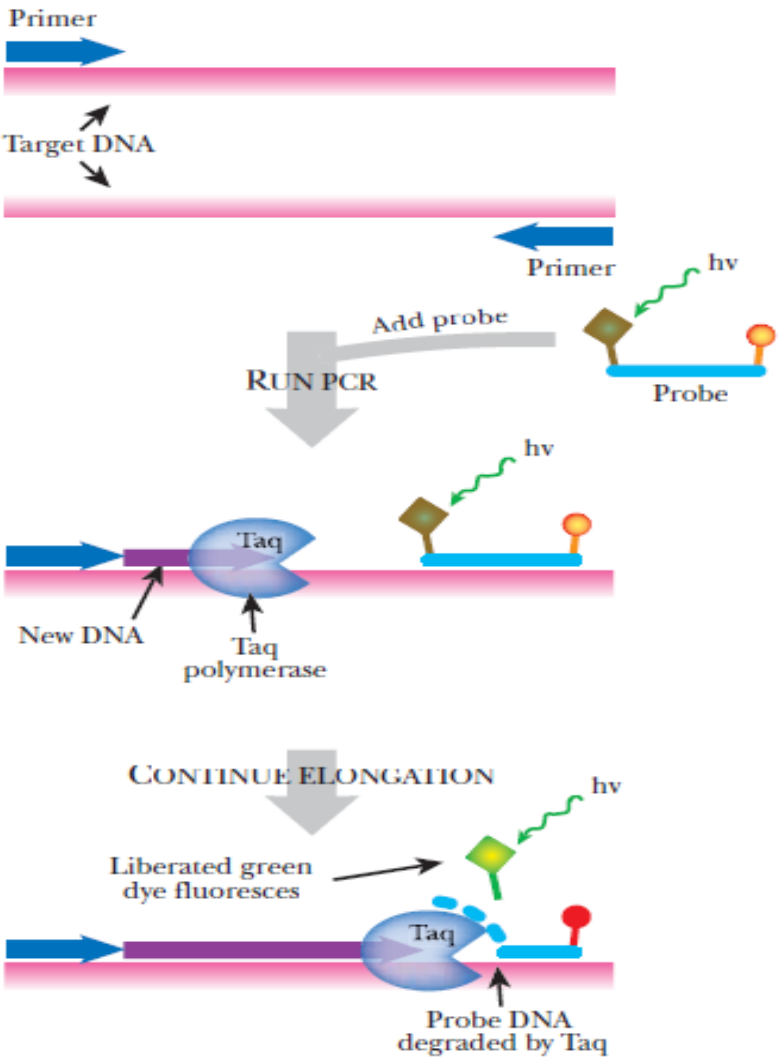


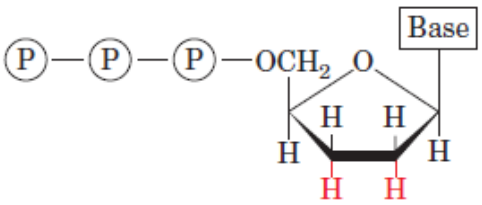
Figure 2: Real Time Fluorescent PCR with Reporter Probe

The TaqMan® probe has three elements: a short-wavelength fluorophore on one end (diamond), a sequence that is specific for the target DNA (blue), and a long-wavelength fluorophore at the other end (circle). The two fluorophores are so close that fluorescence is quenched and no green light is emitted. This probe is designed to anneal to the center of the target DNA. When Taq polymerase elongates the second strand during PCR, its nuclease activity cuts the probe into single nucleotides. This releases the two fluorophores from contact and abolishes quenching. The short-wavelength fluorophore can now fluoresce and a signal will be detected that is proportional to the number of new strands synthesized.

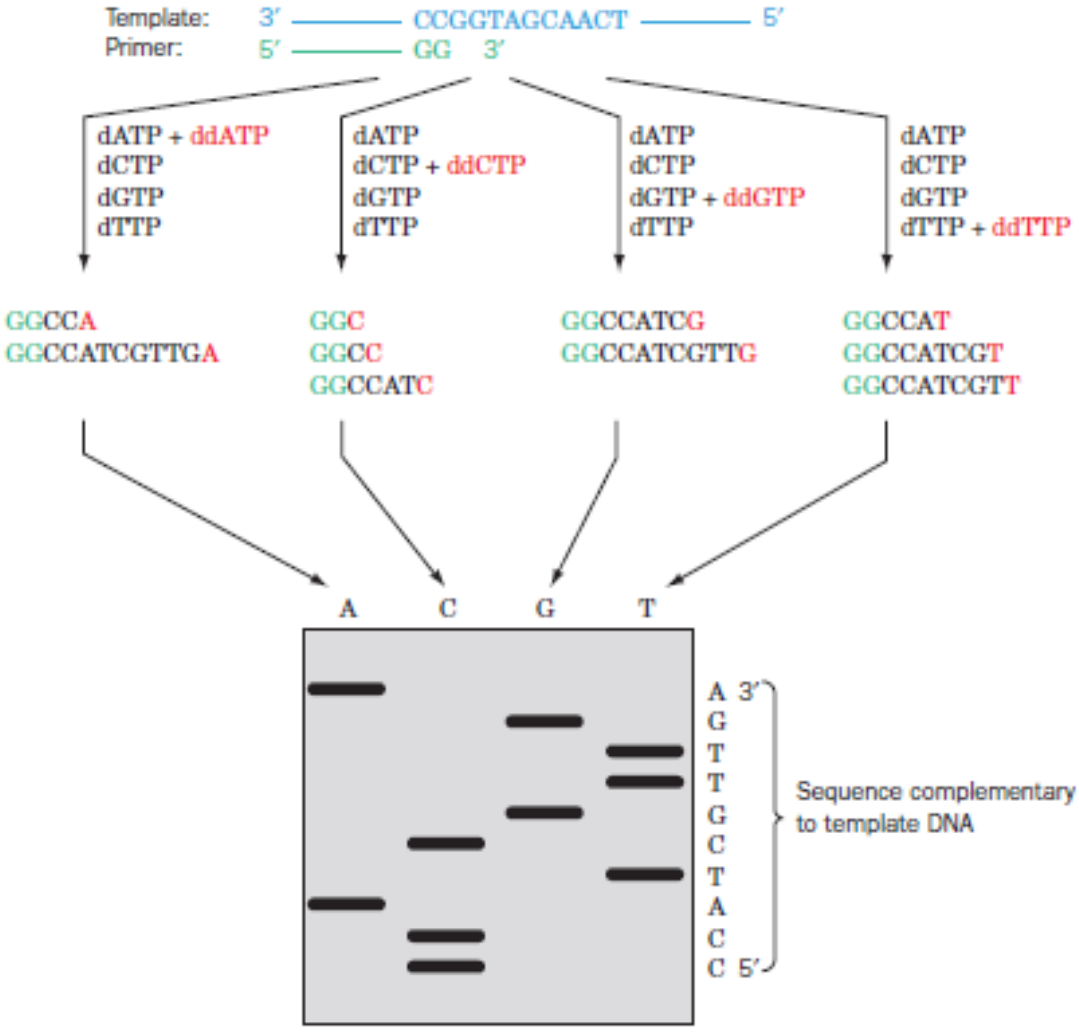
Sanger's method of DNA sequencing:

In the Sanger method (alternatively called the chain terminator method and the dideoxy method) , the DNA to be sequenced is incubated with the Klenow fragment of DNA polymerase I, a suitable primer, and the four deoxynucleoside triphosphates (dNTPs). Either at least one dNTP (usually dATP) or the primer is α-^{32}P-labeled. In addition, a small amount of the 2',3'-dideoxynucleoside triphosphate (ddNTP) of one of the bases is added to the reaction mixture. When the dideoxy analog is incorporated in the growing polynucleotide in place of the corresponding normal nucleotide, chain growth is terminated because of the absence of a 3'-OH group. By using only a small amount of the ddNTP, a series of truncated chains is generated, each of which is terminated by the dideoxy analog at one of the positions occupied

by the corresponding base. Each of the four ddNTPs is reacted in a separate vessel. The four reaction mixtures are simultaneously electrophoresed in parallel lanes on a sequencing gel. The sequence of the DNA that is complementary to the template DNA can then be directly read off an autoradiogram of the sequencing gel, from bottom to top.

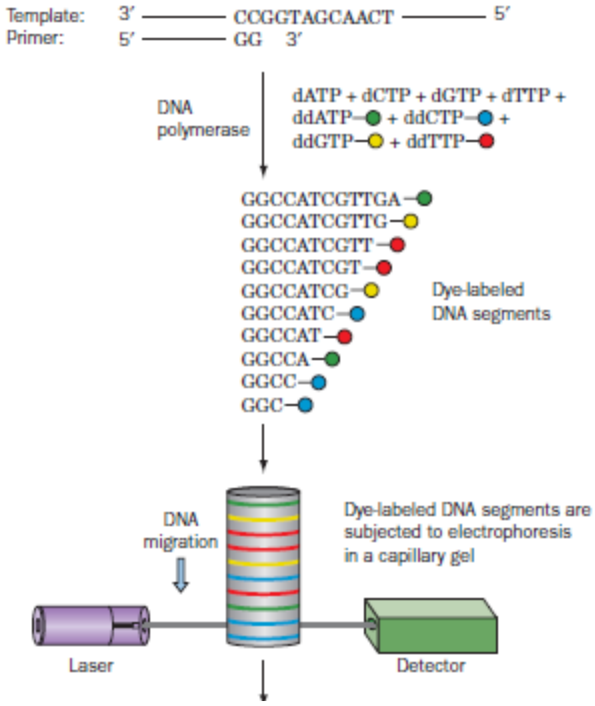


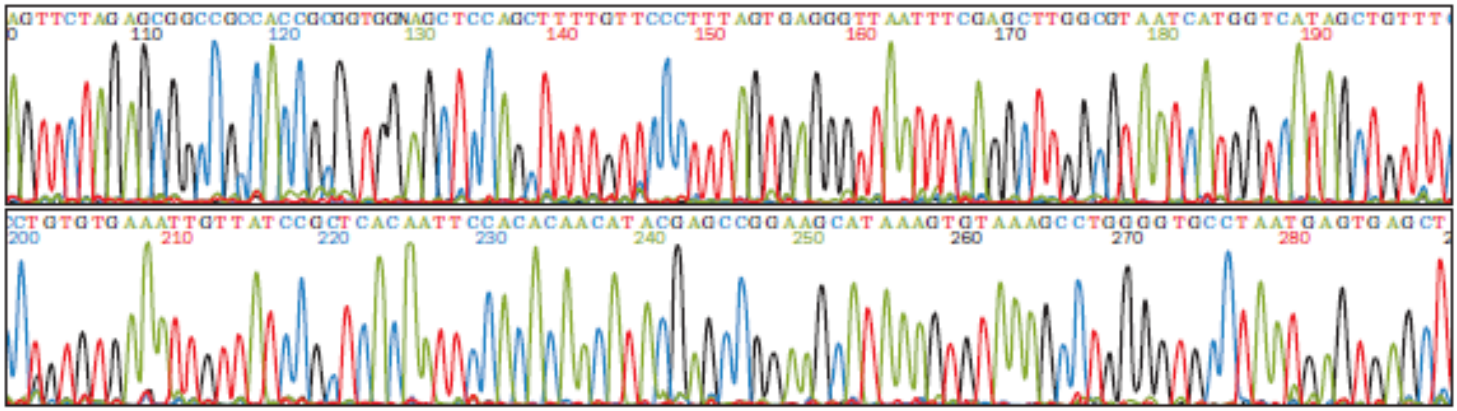
2′,3′-Dideoxynucleoside triphosphate



The Sanger Method Has Been Automated

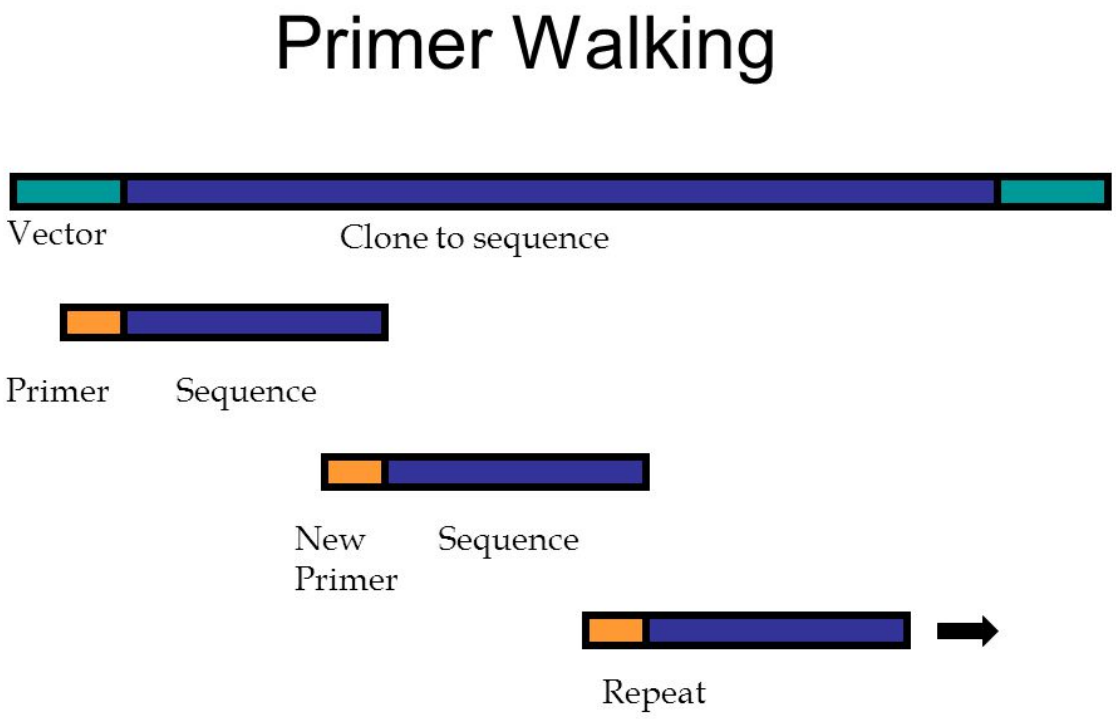
In order to sequence large tracts of DNA such as entire chromosomes, the Sanger method has been greatly accelerated through automation. This required that the above described radiolabeling techniques, which are not readily automated, be replaced by fluorescence labeling techniques (with the added benefit of eliminating the health hazards and storage problems of using radiolabeled nucleotides). In the most widely used such technique, each of the four ddNTPs used to terminate chain extension is covalently linked to a differently fluorescing dye, the chain-extension reactions are carried out in a single vessel containing all four of these labeled ddNTPs, and the resulting fragment mixture is subjected to sequencing gel electrophoresis in a single lane (Fig. 7-15a). As each fragment exits the gel, its terminal base is identified according to its characteristic fluorescence spectrum by a laser-activated fluorescence detection system.





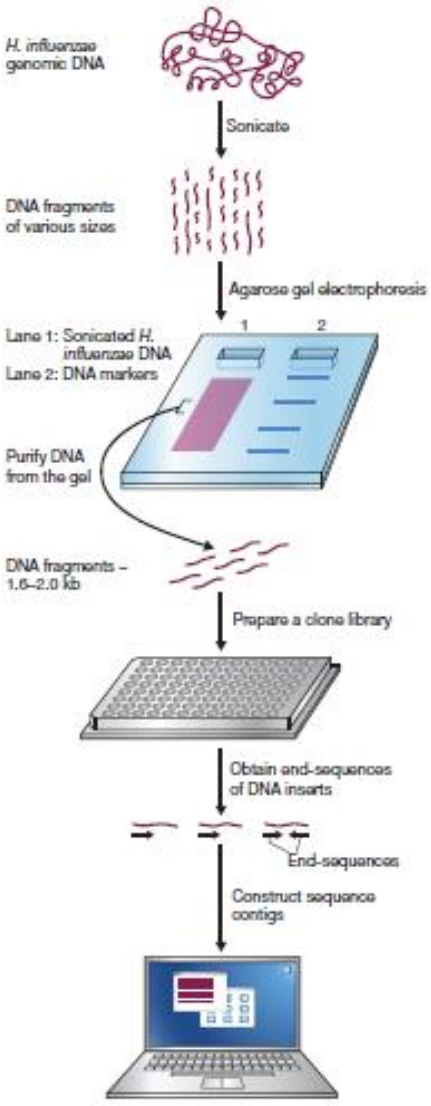
Primer walking:

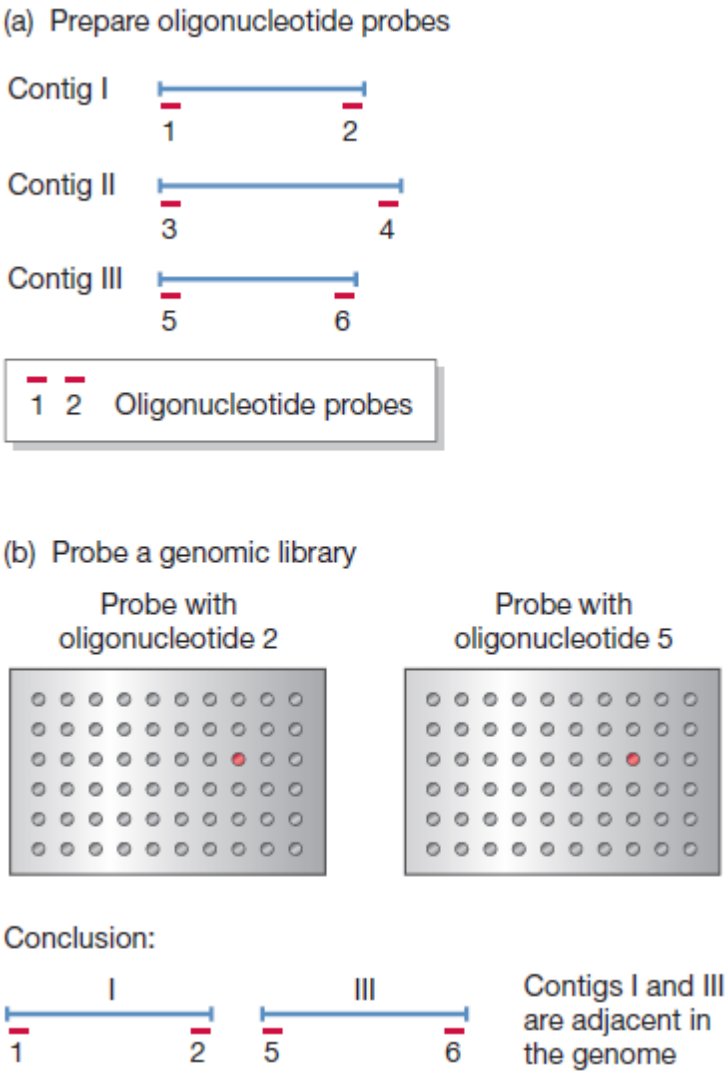
The Sanger method is fast, reliable and accurate, but is limited to its short reads of around 500 nucleotides per run. In order to extend the amount of reads, primer walking technique can be used. In Sanger sequencing, a primer about 10-20 base pairs is attached below the start of the target sequence. Since the strand terminates at around 500 nucleotides, any sequences longer cannot be read. To get around this, a second primer is added that is around 10-20 base pairs upstream of the termination of our first sequence. We can then sequence the next ~500 base pairs, and repeat this process until the entire cloned DNA is sequenced.



Shotgun Sequencing: In the shotgun approach, the genome is randomly broken into short fragments. The resulting sequences are examined for overlaps and these are used to build up the contiguous genome sequence.

The key requirement of the shotgun approach is that it must be possible to identify overlaps between all the individual sequences that are generated, and this identification process must be accurate and unambiguous so that the correct genome sequence is obtained. An error in identifying a pair of overlapping sequences could lead to the genome sequence becoming scrambled, or parts being missed out entirely. The probability of making mistakes increases with larger genome sizes, so the shotgun approach has been used mainly with the smaller bacterial genomes. The shotgun approach was first used successfully with the bacterium *Haemophilus influenzae.*





Using oligonucleotide hybridization to close gaps in the *H. influenzae* genome sequence. Oligonucleotides 2 and 5 both hybridize to the same λ clone, indicating that contigs I and III are adjacent. The gap between them can be closed by sequencing the appropriate part of the λ clone.