Unit - 3

cDNA Libraries

We have already seen how to prepare cDNA beginning with RNA using reverse transcriptase. Thus we know that cDNAs are blunt ended. Therefore to clone cDNA, the first action is to fuse them with appropriate adaptors so that on subsequent cutting, a suitable sticky end will be generated which will help further during ligation and creation of the recombinant.

Any RNA can be converted into a cDNA. Also, when we have said that a cDNA library is a library of those regions of the genome that are expressed, we have conveyed an idea that this library is representative of all the genes, whether they be genes coding for proteins or for rRNAs or tRNAs. *In practice, it is just mRNAs that are normally used as templates for making cDNA and therefore a cDNA library most of the time represents the heterogeneous population of mRNAs present in the cell. This library then is mostly a representative of protein coding genes of the cell.*

Most mRNAs are not bigger than 10 kb. Thus, for cDNA libraries we cannot normally use λ replacement vectors. The vectors used here are either λ insertion vectors or the plasmids. The former are preferred.

We have already considered the shortcomings of a cDNA library as far as representation of all the expressed regions of the genome are concerned (high rate of transcription vs low rate of transcription, high vs short mRNA half life, gene not expressed all the time etc.). We will now discuss the advantages of a cDNA library.

⇒ Obviously, a cDNA library consists of far less number of clones than a genomic library. This makes screening procedures for detecting a particular gene easy.

CONSTRUCTION OF cDNA LIBRARY

A *cDNA library* is a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism or tissue is represented as its cDNA insertion in a plasmid or a phage vector. The frequency of a specific cDNA in such a library would ordinarily depend on the frequency of the concerned mRNA in the tissue/organism in question.

➢ Isolation of mRNA

For isolation of mRNA, total RNA is first extracted from a suitable organism/tissue. The amount of desired mRNA in this sample is then increased by using one of the several procedures, some of which are listed below.

1. Chromatography on poly-U sepharose or oligo-T cellulose, which retains mRNA molecules since they have 3' poly-A tails; *this enriched the preparation with mRNAs of all kinds.*
2. In some specific cases, density gradient centrifugation can be used to increase the frequency of desired mRNA molecules.
3. When the protein produced by a gene is known, it is purified and used to produce antibodies specific to it. These antibodies are used to precipitate the polysomes (mRNAs associated with ribosomes and newly synthesized polypeptide chains) engaged in the synthesis of the concerned polypeptide. Now mRNA is isolated from the precipitated polysomes and purified. Maize zein (seed storage protein) mRNA was isolated in this way.
4. Some genes are expressed only in specific tissues, *e.g.*, seed storage protein genes in developing seeds, chicken ovalbumin gene in oviduct, globin gene in erythrocytes, insulin gene in pV cells of pancreas, etc. Therefore, mRNA preparations from such tissues are exceptionally rich in the concerned mRNA or may even contain only this mRNA.

➢ Preparation of cDNA

cDNA is the copy or complementary DNA produced by using mRNA (usually) as a template. In fact, any RNA molecule can be used to produce cDNA. DNA copy of an RNA molecule is produced by the enzyme ***reverse transcriptase*** (RNA-dependent DNA polymerase; discovered by Temin and Baltimore in 1970) generally obtained from avian myeloblastosis virus (AMV). This enzyme performs similar reactions as DNA polymerase, and has an absolute requirement for a primer with a free 3'—OH.

When eukaryotic mRNA is used as a template, a poly-T oligonucleotide (more specifically,-oligodeoxynucleotide) is conveniently used as the primer since these mRNAs have a poly-A tail at their 3'-ends. But special tricks are

required to utilize primers for other RNAs, *e.g.*, prokaryotic mRNA, rRNA, RNA virus genomes, etc. In such cases, a poly-A tail may be added to the 3'-end of the RNA to make it analogous to eukaryotic mRNA (oligo-T is now used as primer); this reaction is catalyzed by the enzyme poly-A polymerase. Alternatively, a oligonucleotide complementary to a region, preferably at the 3'-end, of the RNA molecule may be used as a primer; for this approach, the sequence of a segment (preferably, the 3'-terminal segment) of the RNA must be known.

The appropriate oligonucleotide primer (oligo-T for eukaryotic mRNA) is annealed with the mRNA; this primer will base-pair to the 3'-end of mRNA. Reverse transcriptase extends the 3'-end of the primer using mRNA molecule as- a template. This produces a RNA-DNA hybrid molecule; the DNA strand of this hybrid is obviously the DNA copy (cDNA) of the mRNA strand. The RNA strand is digested either by RNase H or alkaline hydrolysis; this frees the single stranded cDNA. Curiously, the end of this cDNA serves as its own primer and provides the free 3'-OH required for the synthesis of its complementary strand; therefore a primer is not required for this step. The complementary strand of cDNA single strand is synthesized by either the reverse transcriptase itself or by *E. coli* DNA polymerase I. Since the 3'-end of the cDNA single strand is used as primer for this reaction; a short hairpin loop is generated at this end. The hairpin loop is cleaved by a single strand-specific nuclease to yield a regular DNA duplex.

Cloning of cDNA

cDNAs are usually cloned in phage insertion vectors because they afford high efficiency packaging *In vitro* and, as a result, give large numbers of cDNA clones. Typically, 10^5-10^6 cDNA

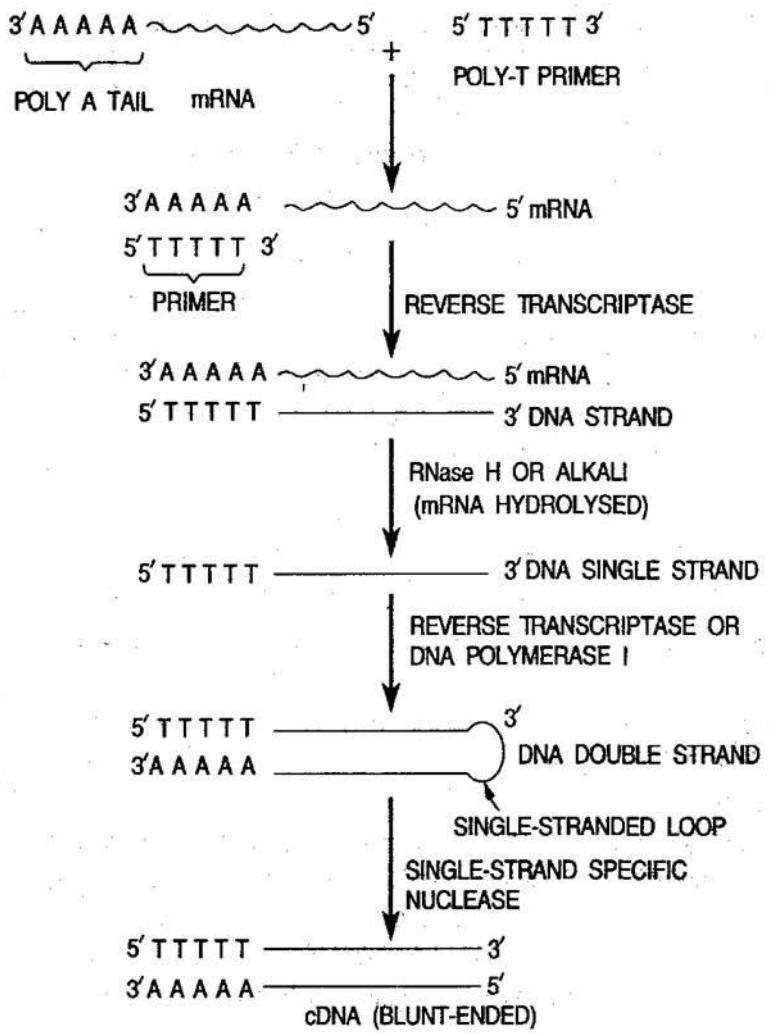


Fig. 2.9. Production of cDNA (complementary or copy DNA) from an mRNA molecule used as template by the enzyme reverse transcriptase.

clones are sufficient for the isolation of low-abundance mRNAs that are present at about 15 molecules per cell or above. In case the mRNA of interest is less than this, it is advisable to enrich the mRNA preparation before library construction. One way to do is by size fractionation coupled with testing the various fractions for the presence of desired mRNAs, *e.g.*, by injecting the mRNAs into ***Xenopus*** oocytes and assaying for the protein product of the concerned mRNA.

2.8.4. Problems in cDNA Preparation

The ideal situation of Fig. 2.9 is rare. In fact, starting with even a pure mRNA preparation, the end product consists of a mixture of cDNAs most of which are somewhat shorter than the complete RNA molecule. This is because of the following problems: (1) incomplete copying of the mRNA by reverse transcriptase (5′-end of mRNA missing from cDNA), (2) incomplete copying of the cDNA single strand so that the 3′-end of mRNA will be missing from cDNA, and (3) the nuclease used for cleaving the hairpin loop may also nibble away the ends of the duplex.

The first situation listed above arises when oligo-T primers are used to prime reverse transcription; this gives rise to a ***3′-end bias,*** *i.e.*, preferential recovery of clones representing 3′-ends of the mRNAs, in the library. This can be avoided by using random, usually, hexamer, primers to prime reverse transcription as well as for copying of the cDNA single strand. This yields smaller cDNA fragments, which must be assembled to recover the full length cDNAs. In addition, as the size

The amount of mRNA (=cDNA) increases, it becomes increasingly more difficult to isolate full length clones. These problems arise because of the poor processivity of reverse transcriptase purified from avian myeloblastosis virus (AMV) or produced in *E. coli* from the gene of Moloney murine leukaemia virus (MMLV). These enzymes also possess intrinsic RNase activity, which degrades the RNA template. But engineered MMLV reverse transcriptase is now available, *e.g.,* Superscript II marketed by Life Technologies, that lack RNAse H activity, and is more efficient in generating full-length cDNAs. Superscript II has another advantage: it functions at 50°C, which increase its efficiency. In contrast, native enzymes function optimally at 37°C; as a result, they tend to stall in regions rich in secondary structures, *e.g.,* in the 5'- and 3'-untranslated regions of mRNAs.

Elaborate strategies have been developed to overcome these problems. For example, incomplete copying of the RNA can be ameliorated by using a specially designed *E. coli* vector to which the RNA is attached before copying; all subsequent steps are performed in association with this vector. Similarly, the use of single strand-specific nuclease can be avoided by adding a poly-C tail, using the enzyme terminal deoxynucleotidyltransferase, to the 3'-end of the cDNA single-strand produced by copying of the mRNA; the complementary oligonucleotide (poly-G, in this case) is now used as a primer for the synthesis of complementary strand to yield double-stranded cDNA without a hairpin loop. This can be combined with PCR to obtain a large number of copies of the cDNA (Fig. 3.16).

The double-stranded cDNA preparations are always a mixture of different kinds of molecules due to the above problems in copying of the RNA and also because even highly purified mRNAs are never absolutely pure. Physical and chemical methods are incapable of resolving these mixtures. Therefore, the cDNA mixture itself is used for cloning and the desired cDNA is identified and isolated in pure form from the appropriate bacterial clone. An elaborate strategy to obtain full-length cDNAs uses two enzymes, alkaline phosphatase and acid pyrophosphatase to select only full-length mRNA molecules for cDNA synthesis. The mRNA preparation is first treated with alkaline phosphatase, which removes the 5'-phosphate groups from all such mRNA molecules that lack the 5-cap, but it does not affect the 5'-capped molecules. The mRNA is then treated with acid phosphatase, which removes the 5'-cap from the 5'-end of mRNA leaving the 5'-phosphate group intact. Now

a specific oligonucleotide is attached to the 5'-ends of the full length mRNA molecules by RNA ligase to obtain ***oligocapped*** mRNAs. The mRNA molecules that were originally without the 5'-cap cannot be oligocapped since they will lack 5''-phosphate residue needed for ligation. *This step ensures that only those mRNA molecules that have their 5'-cap intact are oligocapped.* The mRNA is now reverse transcribed using oligo-T as primer; *this ensures reverse transcription of only such mRNA molecules that have their 3'-ends intact.* The cDNA single-strand so obtained is replicated by PCR using the oligonucleotide (used for oligocapping of mRNA) sequence as primer; this generates full-length duplex cDNAs. Those mRNA molecules that were only partially reverse-transcribed will lack the oligo-cap; as a result, they will not be replicated and will remain single stranded. *This step eliminates 3'-end bias and will ensure generation of only full length cDNA duplexes-,* these are further amplified using the oligonucleotide and oligo-T as primers.

RT-PCR can be used to generate cDNA for library construction when the source material is in a very small amount or the tissue is fixed. But DNA polymerases used in PCR are more error-prone than those used for synthesis of the second strand of DNA; as a result, the cDNAs would contain a large number of mutations. In addition, there may be some distortion due to competition among templates, and a bias towards shorter DNAs. Contamination with genomic DNA may give false results since even trace amounts of the DNA will be amplified. This can be taken care of by treating the RNA preparation with DNase to digest the contaminating DNA.

Properties of cDNAs and cDNA Libraries

1. Eukaryotic cDNAs are free from intron sequences.
2. As a result of the above, they are smaller in size than the corresponding genes, *i.e.*, the genes that encoded them.
3. A comparison of the cDNA sequence with the corresponding genome sequence permits the delineation of intron/exon boundaries.
4. The contents of cDNA libraries from a single organism will vary widely depending on the developmental stage and the cell type used for preparation of the library. In contrast, the genomic libraries will remain essentially the same irrespective of the developmental stage and the cell type used.

5. A cDNA library will be enriched for abundant mRNAs, but may contain only a few or no clones representing rare mRNAs.

Applications of cDNA Library

Use of cDNA is absolutely essential when the expression of an eukaryotic gene is required in a prokaryote, e.g., *a bacterium.* The gene expression may be needed either

(1) for detection of the clone or

(2) the polypeptide product may be the primary objective of cloning.

This is because eukaryotic genes have introns, which must be removed from their transcripts to yield mature mRNAs. Bacteria do not possess the enzymes necessary for removal of introns; therefore, they do not have the ability to produce mature ready-to-translate mRNA molecules from the transcripts of complete native eukaryotic genes. In contrast, functional mRNA molecules do not have introns; hence their cDNA is also free from introns and can be cloned and expressed in bacteria. For example, cDNAfor interferon, blood clotting factor VIII C (both human) and several other mRNAs have been expressed in bacteria. In addition, cDNAs are used to generate expressed sequence tags (ESTs), which are very useful in high-throughput genome research.

==========

- **GENOMIC LIBRARY**

A *genomic library* is a collection of plasmid clones or phage lysates containing recombinant DNA molecules so that the sum total of DNA inserts in this collection, ideally, represents the entire genome of the concerned organism. However, inspite of all the care taken in the preparation of genomic libraries, certain DNA fragments should be expected to be under- or over-represented or even missing. The possible reasons for this may be that certain fragments code for a toxic product, or might replicate slowly or might have been altered by recombinational events during cloning. In addition, endonuclease cleavage sites are often not recognised equally well. Certain DNA fragments may, therefore, never ippear in partial digests by a restriction endonuclease used for the construction of genomic libraries.

Construction of a Genomic Library

For preparation of a genomic library, the total genomic DNA of an organism is extracted. The DNA is broken into fragments of appropriate size either by mechanical shearing (this generates blunt-ended fragments), or sonication, or by using a suitable restriction endonuclease for partial digestion of the DNA; complete digestion is avoided since it generates fragments that are too heterogeneous in size. It is desirable that the DNA is cut randomly so that the DNA fragments (= clones) overlap one another and no sequence of the genome is systematically excluded. For partial digestion, restriction enzymes having 4-base (tetrameric) recognition sequences are employed in preference to those having 6-base (hexameric) target sites. This is because a given 4-base recognition site is expected to occur every 4^4 (= 256) base pairs in a DNA molecule, while a 6 base target site would occur only after every 4^6 (= 4,096) base pairs. (It is assumed here that the arrangement of the 4 bases in DNA molecules is random). Therefore, the fragments produced in partial digests with enzymes having 4-base recognition sites are more likely to be of appropriate size for cloning than those generated by enzymes having 6-base recognition sites. Single or mixed digestions with the enzymes *Alu*I (AG/CT), *Hae*III (GG/CC) have been used for constructing genomic libraries. The use of restriction enzymes has the advantage that the same set of fragments are obtained from a DNA each time a specific enzyme is used, and many of the enzymes produce cohesive ends.

The partial digests of genomic DNA are subjected to agarose gel electrophoresis or sucrose gradient centrifugation for separation from the mixture of fragments of appropriate size. These fragments are then inserted into a suitable vector for cloning. This constitutes the ***shotgun approach*** to gene cloning. In principle, any vector can be used, but X vectors and cosmids have been the most commonly used since DNA inserts of upto 23-25 kb (kilobase pairs) can be cloned in these vectors, particularly in replacement vectors. In addition, high capacity vectors like BACs (bacterial artificial chromosomes), PACs (Pi-derived artificial chromosomes), and YACs (yeast artificial chromosomes) are also used for construction of genomic libraries. The vectors containing the DNA inserts are cloned in a suitable bacterial host.

The minimum size (number of clones or bacterial colonies) of a genomic library depends on the following two factors: (1) the complexity of the genome

(the more complex the genome the larger the size), (2) the size of the DNA insert or fragment length used for cloning (the smaller the fragment size, the larger the number of clones for the same genome), and (3) the probability of a gene being represented in the library. The minimum number of colonies in a genomic library constructed with 20 kb inserts for a 99% probability of a genomic sequence being represented would be 1,157 for *E. coli* (genome of 2.6 x 10^3 kb), 3,462 for yeast (genome, 1.8 X 10^4 kb), 38,000 for *Drosophila* (genome, 1.2 x 10^5 kb), and 6,90,819 for man (genome, 3.2 x 10^6 kb). The minimum number (N) of independent recombinant clones/plaques in a genomic library required for a given DNA insert being represented in the library with a given probability *(P)* is given by the formula, $N = \ln (1-P)/\ln [1-(1/n)]$

where, *n* equals the ratio of the genome size (in bp) to the DNA insert size (in bp).

Identification of the Desired Clone

Identification of the bacterial colony containing the desired DNA fragment from among those making up the library generally employs a suitable hybridization probe in colony hybridization. The probe may be mRNA of the gene, cDNA of its mRNA, homologous gene from another organism, or a synthetic oligonucleotide representing the sequence of a part of the desired gene/DNA fragment. In order to detect hybridization, the probe must be labelled, usually with a radioactive isotope.

The genomic DNA fragments may be subjected to electrophoresis or high performance reverse phase liquid chromatography in order to increase the frequency of desired DNA inserts before using them for cloning. Electrophoresis is usually done in agarose gel, which effectively separates fragments ranging from 0.3 to 23 kb (polyacrylamide gel has the effective range of 2 to 6 kb). These techniques separate DNA fragments mainly on the basis of their size and base composition. The desired DNA fragment is obtained by collecting the appropriate eluate from the chromatography column or by eluting the fragment from appropriate piece of the gel. The location of this piece in the gel is identified by hybridization with a labelled probe specific for the desired DNA fragment following Southern hybridization. Once the location of the fragment becomes known, the comparable region of an unhybridized gel is eluted to obtain the fragments for cloning. Similarly, the different fractions

collected from chromatography may be subjected to dot blot assay to identify the fraction containing the desired DNA fragment. However, the fragments isolated by either method from complex mixtures obtained from large genomes will be contaminated by other fragments of similar size or elution properties. Therefore, the clones obtained from these fragments are subjected to colony hybridization for picking out the clones containing the desired fragment. Obviously, these techniques greatly enhance the probability and, therefore, proportionately reduce the work of locating the desired colony or clone.

==========

- **PROBES**

Probes are small (10-30 bases long) nucleotide sequences used to detect the presence of complementary sequences in nucleic acid samples. This is achieved by permitting the probes to base-pair with the sample nucleic acids and then identifying the samples that show base pairing with the probes, *i.e.,* ***hybridization.*** The detection of hybridization is highly precise and extremely sensitive provided the probes are suitably labelled for an easy detection. Clearly, hybridization can occur only when the base sequence of a probe is present within the gene or DNA segment, which it is aimed to detect. Both DNA and RNA are used as probes. Single-stranded DNA probes are more convenient and preferable, but denatured double stranded DNA molecules can also be used. RNA probes are ordinarily single-stranded.

- **PREPARATION OF PROBES**

Probes can be obtained in several ways: some of the important ones are briefly described below.

1. Highly purified mRNA can be used as probe; mRNAs are naturally single-stranded.
2. Single-stranded RNA probes can be readily prepared by cloning the corresponding DNA sequence into a special vector like *pGEM, p*Bluescript, etc. Such a vector has a different and specific prokaryotic promoter beyond the two ends of the DNA insert. The recombinant vector is linearized and transcribed with the appropriate prokaryotic

RNA polymerase to obtain RNA molecules complementary to one or the other strand of the DNA insert.

3. DNA segments isolated from the genome of an organism or cDNA molecules prepared by using mRNAs can be cloned in *E. coli* and used as probes. These probes, of necessity, will be double-stranded.
4. Single strand DNA probes can be prepared by cloning the concerned DNA segment in a vector like phage M13 vectors of *E. coli.* Different recombinant clones of this vector will yield probes complementary to one or the other strand of the cloned DNA segment.
5. Single-stranded cDNA probes can be prepared by limiting the copying of mRNA by reverse transcriptase to only one strand.
6. PCR can be used to generate single-stranded copies of a DNA sequence; this is called ***asymmetric PCR.***
7. Synthetic oligonucleotides can be prepared for use as probes. This is rather easy when the base sequence of DNA to be detected or of RNA produced by it is known. But the base sequence of a part of the gene can be deducted from the amino acid sequence of a small, say, 5 amino acid long segment of the protein encoded by it. But it is impossible to determine the exact base sequence of the gene due to the degeneracy of genetic code, *i.e.,* one amino acid being coded by two or more codons. This problem is tackled by synthesizing a mixture of oligonucleotides containing all the possible base sequences predicted by the degenerate code. For example, there are four different base sequences possible on the basis of the amino acid sequence. The oligonucleotide mixture to be used as probe should contain all the four sequences. This is easily achieved by using the appropriate mixture of nucleotides (in place of a single nucleotide) at the specific step in oligonucleotide synthesis where alternative bases can be present, *e.g.,* at step 6, a mixture of T and C mononucleotides will be used.

➢ LABELLING OF PROBES

The probes can either be labelled with (1) radioactivity, *e.g.,* ^{32}P, or (2) with nonradioactive labels, *e.g.,* biotin, digoxigenin, etc.

1. Radioactive Labelling.

The various techniques for labelling of nucleic acids are as follows:

a. nick translation,
b. primer extension,
c. methods based on RNA polymerase,
d. end-labelling
e. direct labelling methods.

(i) In *direct labeling* a nucleotide labelled with ^{32}P is provided during production of the probe so that this radioactive nucleotide is included in the probe. (ii) The procedure of *nick translation* is widely used to label (radioactive or non-radioactive labelling) double-stranded DNA molecules

2. Non-radioactive Labelling.

There are several strategies for nonradioactive labelling of nucleic acids, *e.g.*, labelling with biotin, digoxigenin, fluorescent molecules, etc. Radioactive labelling poses problems in handling and, especially, disposal. Further, due to the relatively short half-life of the radioisotopes, the probes have to be used within a short period of time after they are prepared. In| contrast, some non-radioactively labelled probes can be stored at-20°C for long periods of time.

Biotin labelled probes are prepared by nick translation in which biotin-conjugated *(biotinylated)* nucleotides are used. Long tails of biotinylated nucleotides may be added to the probes to increase the number of biotin (vit. H) molecules available for the colour development necessary for detection; these tails have a nonspecific base sequence so that the same tail can be added to any probe. These probes are hybridized with the test nucleic acid fixed to a solid support and the nonhybridized probes are washed off. The hybridized samples are detected by the development of blue colour following a series of cytochemical reactions, which basically utilize the affinity of egg white glycoprotein avidin for biotin. This approach has the advantage of being much faster than that with radioactive probes, which requires autoradiography. But its chief disadvantage relates to the inability to reuse the filter (solid support) and the nucleic acids fixed to it for hybridization with other probes since the reactions leading to the colour development product insoluble precipitates. In contrast, the radioactive probes are easily removed by washing under conditions

favouring denaturation, *e.g.,* high pH, and the same filter can be reused for hybridization with a series of probes, used one at a time, for the identification of desired samples.

Nucleotides conjugated with digoxigenin, a plant-derived chemical, may be used in nick-translation to produce digoxigeninlabelled probes. The probes are used in hybridization; after washing away the free probes, the filter is incubated in a detection buffer containing a digoxigenin specific antibody (anti-digoxigenin) coupled with an enzyme (usually alkaline phosphatase). After appropriate washing, the alkaline phosphatase activity is detected by using a suitable substrate that yields colour due to the enzyme action.

Nucleotides have been conjugated with other ligands that produce some detectable signal, *e.g., fluores*cence, enzyme activity, etc. The signals from probes can be amplified by any one of several approaches, *e.g., (i)* attaching multiple enzyme molecules to each probe molecule, *(ii)* adding nonspecific labelled tails to the probes. *(iii)* using multiple secondary probes that hybridize with multiple target- specific primary probes (christmas tree or forest approach), etc.

- **APPLICATIONS OF PROBES**

:

Probes have a variety of applications since they permit a highly reliable and extremely efficient detection of nucleic acid sequences complementary to the employed probes. Some of their applications are briefly summarised below.

1. Identification of the recombinant clone carrying the desired DNA insert; this is the most critical step in DNA cloning (techniques used: Southern, northern, colony, and dot blot hybridizations).
2. Confirmation of the integration (by Southern hybridization) of a DNA insert into the host genome, and its expression in transformed cells (northern hybridization).
3. Development of RFLP (restriction fragment length polymorphism) maps.
4. DNA finger-printing for the unequivocal identification of plant varieties, criminals, parental relationships, etc.
5. *In situ* hybridization for determining the locations of specific sequences in specific, chromosomes.
6. Accurate diagnosis of diseases caused by parasites, pathogens or defective viruses.
7. Preparation of genome maps of eukaryotes, including man.

==========

- **PCR**

The polymerase chain reaction (PCR) technique, developed by Kary Mullis in 1985, is extremely powerful. It generates microgram (µg) quantities of DNA copies (up to billion copies) of the desired DNA (or RNA) segment, present even as a single copy in the initial preparation, in a matter of few hours. The PCR process has been completely automated and compact thermal cyclers are available in the market.

The PCR is carried out *in vitro*. It utilizes the following:

1. a DNA preparation containing the desired segment to be amplified *(target sequence)*,
2. two nucleotide primers (about 20 bases long) specific, *i.e.*, complementary, to the two S'-borders (the sequences present at or beyond the 3'-ends of the two strands) of the desired segment,
3. the four deoxynucleoside triphosphates, *viz.*, TTP (thymidine triphosphate), dCTP (deoxycyctidine triphosphate), dATP (deoxyadenosine triphosphate) and dGTP (deoxyguanosine triphosphate),
4. a heat stable DNA polymerase, *e.g.*, *Taq* (isolated from the bacterium *Thermus acquaticus)*, *Pfu* (from *Pyrococcus furiosus)* and *Vent* (from *Thermococcus litoralis)* polymerases. *Pfu* and *Vent* polymerases are more efficient than the *Taq* polymerase.

➢ PROCEDURE OF PCR

At the start of PCR, the DNA from which a segment is to be amplified, an excess of the two primer molecules, the four deoxyriboside triphosphates and the DNA polymerase are mixed together in the reaction mixture that has appropriate quantities of Mg^{2+}. The following operations are now performed sequentially (Fig. 3.9).

i. Denaturation

The reaction mixture is first heated to a temperature between 90-98°C (commonly 94°C) that ensures DNA denaturation. This is the *denaturation* step. The duration of this step in the first cycle of PCR is usually 2 min at 94°C.

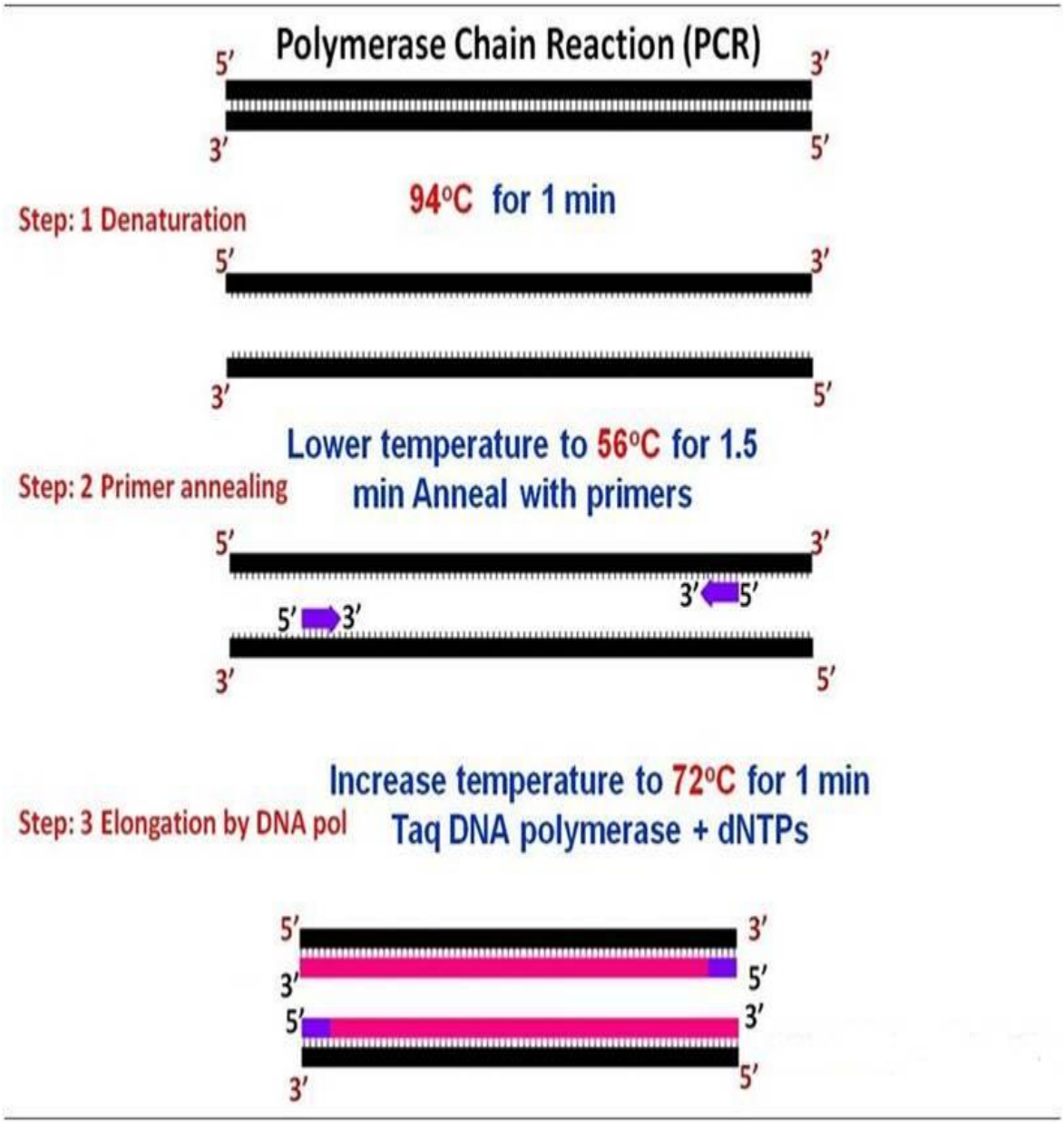
ii. Annealing

The mixture is now cooled to a temperature (generally 40-60°C) that permits annealing of the primer to the complementary sequences in the DNA; these sequences are located at the 3'-ends of the two strands of the desired segment. This step is called *annealing*. The duration of annealing step is usually 1 min during the first as well as the subsequent cycles of PCR. Since the primer concentration is kept very high relative to that of the template DNA, primer-template hybrid formation is greatly favoured over reannealing of the template strands.

iii. Primer Extension

The temperature is now so adjusted that the DNA polymerase synthesizes the complementary strands by utilizing 3'-OH of the primers; this reaction is the same as that occurs *in vivo* during replication of the leading strand of a DNA duplex. The primers are extended towards each other so that the DNA segment lying between the two primers is copied; *this is ensured by employing primers complementary to the 3'-ends of the segment to be amplified.* The duration of primer extension is usually 2 min at 72°C. It has been shown that in case of longer target sequences, best results are obtained when the period of extension is kept at the rate of 1 min per kb of the target sequence and the extension is carried out at 68°C in the place of usual 72°C. *Taq* polymerase usually amplifies

DNA segments of up to 2 kb; special reaction conditions are necessary for amplification of longer DNA segments.



In case of *Taq* polymerase, the optimum temperature for synthesis is between 70° and 75°C; the temperature of reaction mixture is, therefore, adjusted to this temperature. This situation has the following additional advantages. *Between 70-75°C, the base pairing between the primers and the template DNA is much more specific than that at 37°C, the optimal temperature for* E. coli *DNA polymerase. This minimises the chances of annealing of primers to imperfectly matched sequences*

and, thereby, amplification of unwanted DNA. The specificity of annealing is further increased by selecting appropriate conditions like ionic strength, primer length, etc.

The completion of the extension step completes the first cycle of amplification; each cycle may take few (ordinarily 4-5) minutes. It should be noted that extension of the primer continues till the strands are separated during the denaturation step of the next PCR cycle. Therefore, the products of

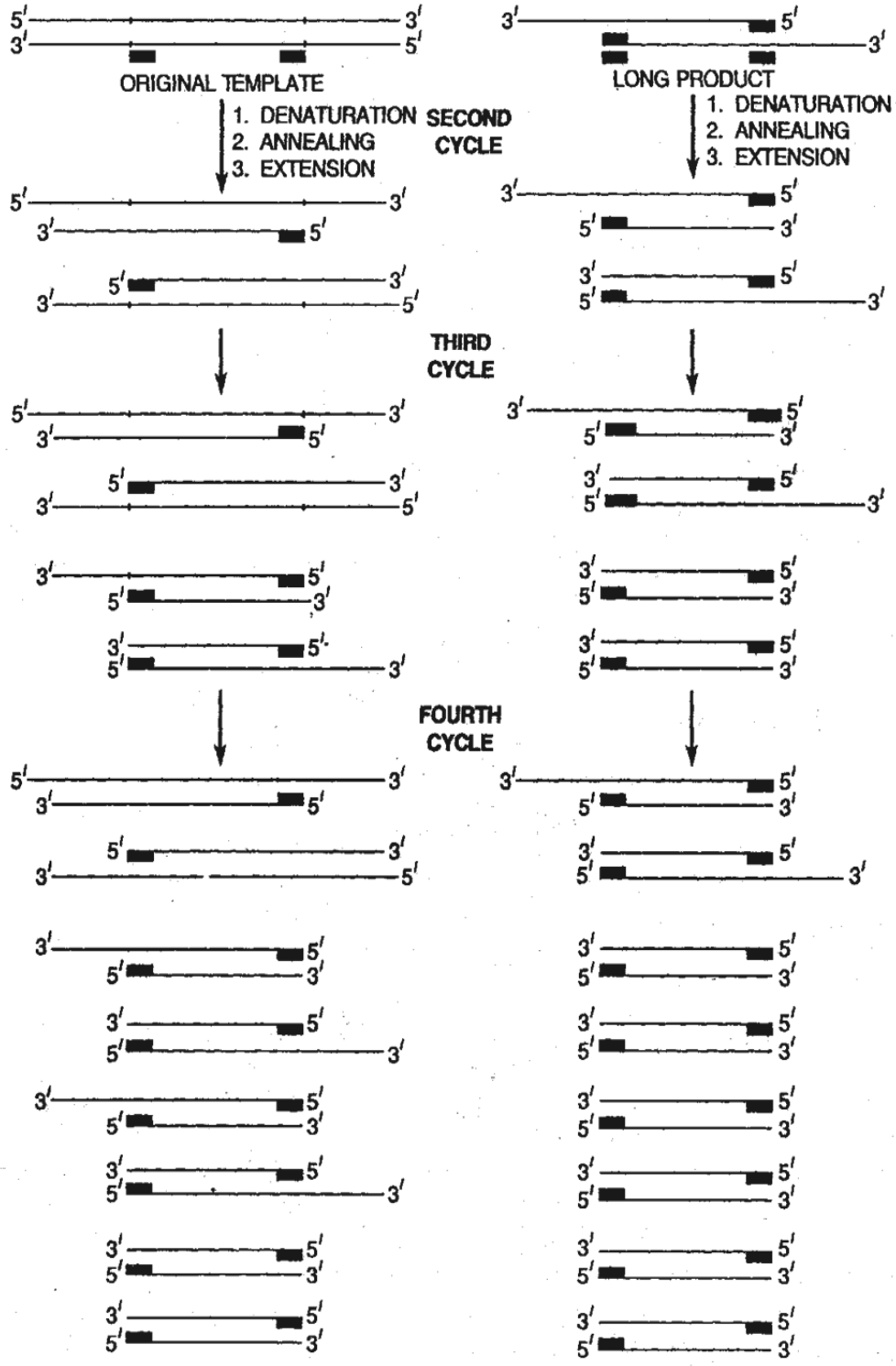


Fig. 3.10. During second and subsequent cycles, the 'long product' accumulates lineraly, while the 'correct' copy of the target sequence increases exponentially. At the end of 4th cycle, there are only 8 long product strands, as compared to 22 strands of the correct product.

this extension step, and indeed of every cycle based on the original template DNA, is of indefinite length; this PCR product is usually called the *long product* (Fig. 3.9). At the second cycle of PCR, primers will anneal to the *'long product'* at sites where sequences complementary to them are located, which will be much before the actual 3'-end of the long product. Extension step in this cycle will produce a product that will be much shorter than the 'long product'; this is the *'correct' PCR product* and represents the target sequence (Fig. 3.10). The original template sequence will also be copied during the second and all subsequent cycles to generate the 'long product'. Therefore, the long product continues to increase linearly, while the 'correct PCR' product will multiply exponentially.

The next cycle of amplification is initiated by denaturation (Step 1), which separates the newly synthesized DNA strands from the old DNA strands. This step is usually of one minute as against 2 min in the first cycle. But in the case of long target sequences, a denaturation step of only 10s at 94°C gives better results than 30s or 60s at 94°C. Annealing allows the primers to base pair with both the new and old strands, the total number of strands being twice their original number. Synthesis of new strands takes place, which doubles the number of copies of the desired DNA segment present at the end of first amplification cycle. This completes the second cycle.

Thus at each cycle, both new and old strands anneal to the primers and serve as templates for DNA synthesis. As a result, at the end of each cycle, the number of copies of the desired segment becomes twice the number present at the end of the previous cycle. Thus at the end of *n* cycles *2"* copies of the segment are expected; the real values are quite close to this expectation. Usually, 20-30 cycles are carried out in most PGR experiments. In case of automated PCR machines, called *thermal cyclers*, the researcher has to only specify the number and duration of cycles, etc. after placing the complete reaction mixture for incubation, and the machine performs the entire programme of operations precisely. After PCR cycles, the amplified DNA segment is purified by gel electrophoresis and can be used for the desired purpose.

==========

- **NUCLEIC ACID AND PROTEIN BLOTTING TECHNIQUES**

- **SOUTHERN BLOTTING / HYBRIDIZATION**
- NORTHERN **BLOTTING / HYBRIDIZATION**
- WESTERN **BLOTTING / HYBRIDIZATION**

- **SOUTHERN BLOTTING / HYBRIDIZATION**

The name of this technique is derived from the following: (1) the name of its inventor, *EM.* Southern and (2) the DNA-DNA hybridization that forms its basis. It is also called *Southern blotting* since the procedure for transfer of DNA from the gel to the nitrocellulose filter resembles blotting. This technique has since been extended to the analysis of RNA *(northern blotting)* and proteins *(western blotting);* these names are only jargon terms, *i.e.,* reverse of Southern being northern and so on, and do not reflect any functional or historical significance.

In Southern hybridization, a sample of DNA containing fragments of different sizes is subjected to electrophoresis using either polyacrylamide or agarose gel. The DNA sample may either be subjected to mechanical shearing or to restriction endonuclease digestion in order to generate the fragments. Agarose gel is useful in separating DNA fragments of few hundred to 20 kb in size, while polyacrylamide is preferred for smaller fragments. Very large DNA fragments of up to 1000-2000 kb are separated in agarose gel with pulsed electrical fields, or field inversion.

The gel provides a complex network of polymeric molecules through which DNA fragments migrate, depending on their sizes, under an electric field since DNA molecules are negatively charged. Smaller molecules of DNA migrate relatively faster than the larger ones. Marker DNA fragments of known sizes are run in a separate lane; this permits an accurate determination of the size of an unknown DNA molecule by interpolation. The gels are stained with the intercalating dye ethidium bromide, which gives visible fluorescence on illumination of the gel with UV light; as little as 0.05 µg of DNA in one band can be detected by using this dye. This approach is useful when few DNA fragments with considerable lengths are to be separated and studied. This

approach also separates the closed circular (supercoiled), nicked (relaxed) and linear configurations of a single DNA molecule.

In many situations, it is critical to detect and identify DNA fragments in a sample that are complementary to a given DNA sequence, *e.g.*, to demonstrate the presence of the gene in question in transgenics, to detect and study RFLP (restriction fragment length polymorphism), etc. This is achieved by Southern hybridization in which the following steps are performed (Fig).

1. The restriction fragments of DNA present in agarose gel (after electrophoresis) are denatured into single-stranded form by alkali treatment.
2. They are then transferred onto a nitrocellulose filter membrane; this is done by placing the gel on top of a buffer saturated filter paper, then laying the nitrocellulose filter membrane on the top of gel, and finally placing some dry filter papers on top of this membrane. The buffer moves, due to capillary action, from the bottom filter paper through the gel carrying with it the denatured DNA present in the gel; the DNA becomes trapped in the nitrocellulose membrane as the buffer passes through it. This process is known as *blotting*, and takes several hours to complete. The relative positions of the bands on the membrane remain the same as those in the gel and there is a minimal loss in their resolution (sharpness).
3. The nitrocellulose membrane is now removed from the blotting stack, and the DNA is permanently immobilized on the membrane by baking it at 80°C *in vacuo*.
4. Single-stranded DNA has a high affinity for nitrocellulose filter membrane. (Note that RNA lacks this affinity). Therefore, the baked membrane is treated with a solution containing 0.2% each of Ficoll (an artificial polymer of sucrose), polyvinylpyrrolidone and bovine serum albumin; this mixture is often supplemented with an irrelevant nucleic acid, *e.g.*, tRNA (***pretreatment***). This treatment prevents nonspecific binding of the radioactive probe (to be used in the next step) probably by attaching macromolecules to all the free binding sites on the membrane. Often the above mixture is included in hybridization reaction itself.
5. The pretreated membrane is placed in a solution of radioactive, single-stranded DNA or an oligodeoxynucleotide (a DNA segment having few to several nucleotides) called ***probe.*** The name probe signifies the fact that this DNA molecule is used to detect and identify the DNA fragment in the gel/membrane that has a sequence complementary to the probe. The conditions during this step are chosen so that the probe hybridizes with the complementary

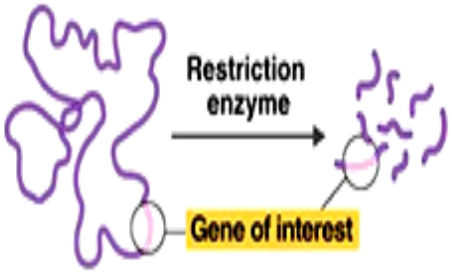
DNA on the membrane to the greatest extent with a low nonspecific binding on the membrane; this step is known as ***hybridization reaction.*** Usually the initial hybridization reaction is carried out under conditions of relatively low stringency of hybridization to pennit a high rate of hybridization; this is followed by a series of post-hybridization washes of increasing stringency, *i.e.*, higher temperature or, more commonly, lower ionic strength, with a view to eliminate the pairing of radioactive probe to related sequences and to allow only perfectly complementary pairing.

6. After the hybridization reaction, the membrane is washed to remove the unbound probes.

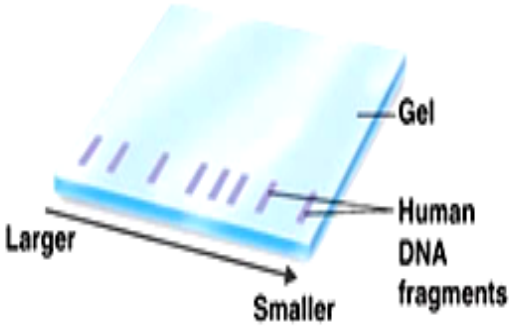
7. The membrane is now placed in close contact with an X-ray film and incubated for a desired period to allow images due to the radioactive probes to be formed on the film. The film is then developed to reveal distinct band(s) indicating positions in the gel of the DNA fragments that are complementary to the radioactive probe used in the study.

It should be kept in mind that electrophoresis of sheared or restricted DNA produces a smear in which the fragments are distributed in a continuum according to their size, and there are no distinct bands. The distinct bands are produced by the hybridization reaction of the selected probe with one or few fragment sequences present in the gel. The Southern blotting technique is extremely sensitive. It can be used to map the restriction sites around a single copy gene sequence in any genome (even of man). It is used for DNA fingerprinting; preparation of RFLP maps, detection and identification of the transferred genesin transgenic individuals, etc.

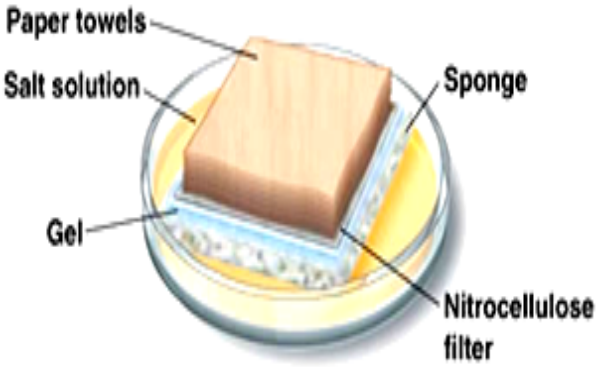
Recently, some new membrane materials, *e.g.*, nylon membranes, have been developed, which have the following advantageous features: (1) they are physically robust in comparison to nitrocellulose filter membranes, (2) both DNA and RNA become cross-linked to them by a brief exposure to UV light, which (3) saves the time needed for baking *in vacuo* in the case of nitrocellulose membranes, and (4) the same membrane blot, *e.g.*, a membrane onto which DNA/RNA has been transferred from a gel and cross-linked by UV exposure, can be used for search with more than one probe after removing the earlier probe by high temperature washing or some other denaturing procedure; in other words, *the nylon membranes are reusable.*



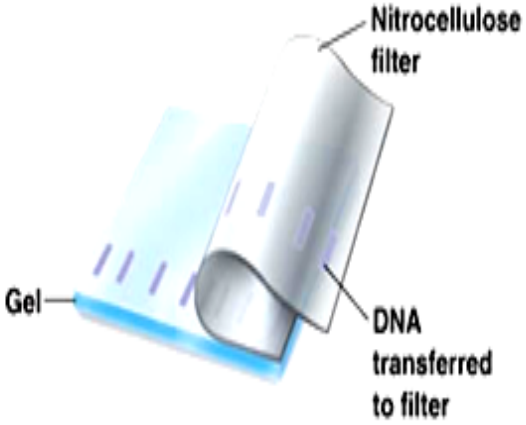
1 DNA containing the gene of interest is extracted from human cells and cut into fragments by restriction enzymes.



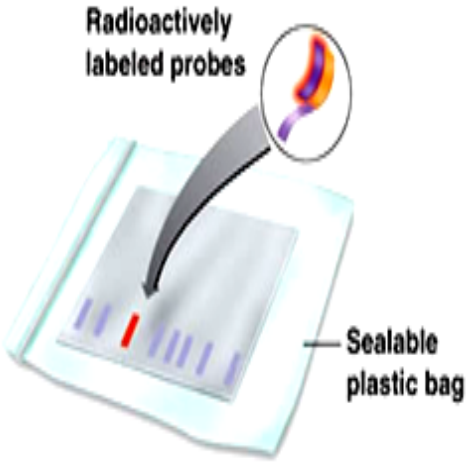
2 The fragments are separated according to size by gel electrophoresis. Each band consists of many copies of a particular DNA fragment. The bands are invisible but can be made visible by staining.



3 The DNA bands are transferred to a nitrocellulose filter by blotting. The solution passes through the gel and filter to the paper towels.



4 This produces a nitrocellulose filter with DNA fragments positioned exactly as on the gel.



5 The filter is exposed to a radioactively labeled probe for a specific gene. The probe will base-pair (hybridize) with a short sequence present on the gene.

6 The filter is then exposed to X-ray film. The fragment containing the gene of interest is identified by a band on the developed film.

- **NORTHERN HYBRIDIZATION**

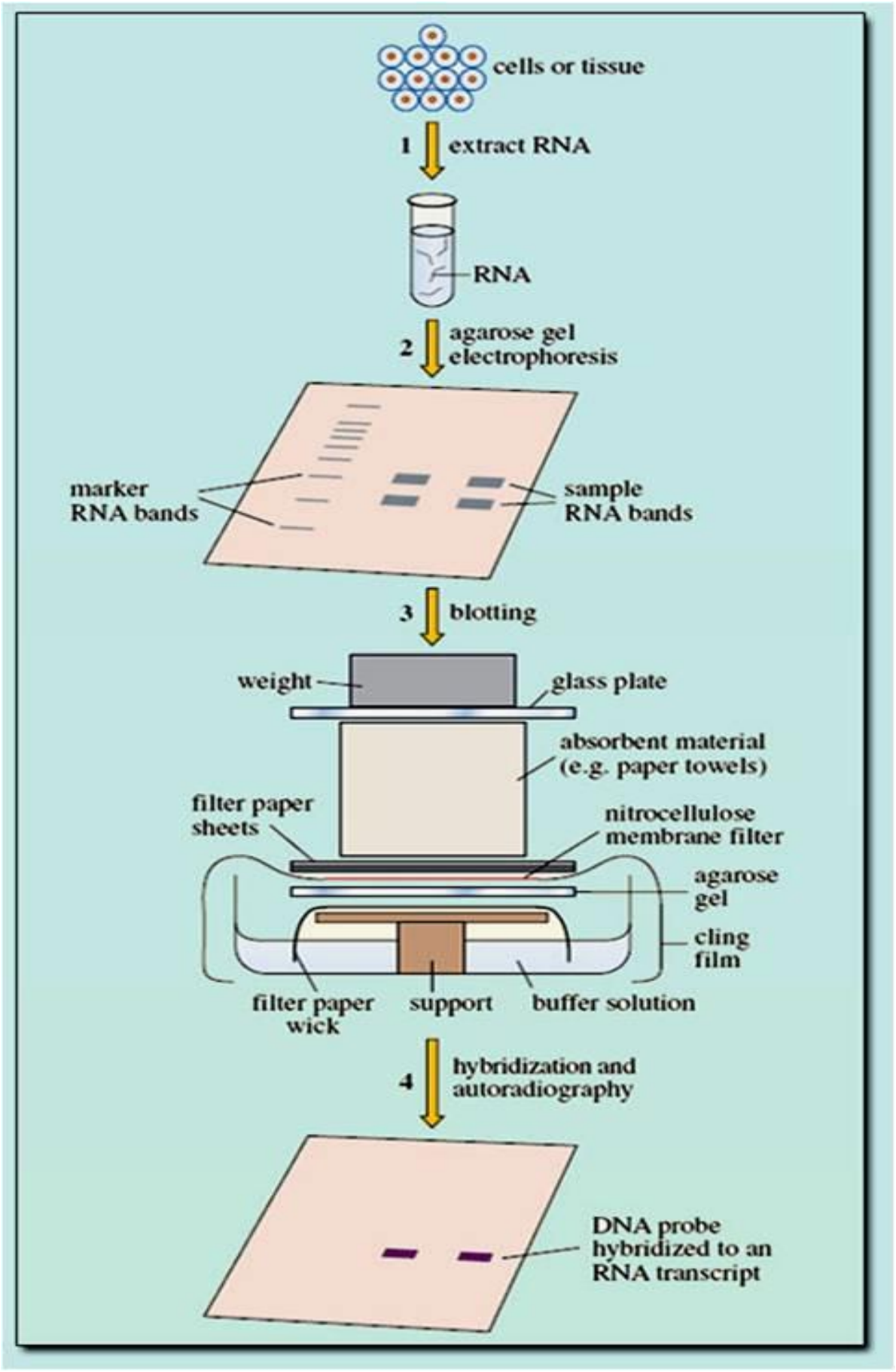
In this technique, RNAs are separated by gel electrophoresis, the RNA bands are transferred onto a suitable membrane, *e.g.*, diazobenzyloxymethyl (DBM) paper or nylon membranes, and immobilized; the bands are hybridized with radioactive, single-stranded DNA probes, and the bands showing hybridization are detected by autoradiography.

Clearly, northern blotting hybridization is simply an extension of the Southern blotting technique. The two techniques, however, show the following differences:

1. in Southern hybridization, DNA is separated by gel electrophoresis, while in northern blotting RNAs are separated;
2. as a result, in Southern hybridization DNA has to be denatured before blotting, while this step is not needed in northern hybridization;
3. nitrocellulose membrane is generally not used for northern, while it is often used for Southern hybridization; and finally,
4. hybridization with the probe produces DNA : DNA hybrid molecules in Southern, but it generates RNA : DNA hybrid molecules in northern hybridization.

Initially, specially prepared paper (diazobenzyloxymethyl, DBM, paper prepared by diazotization of aminobenzyloxymethyl paper) was used for northern blotting since RNA did not bind to nitrocellulose membrane. RNA becomes covalently bound to DBM paper due to which these blot-transfers are reusable. DBM is also equally effective in binding to denatured DNA, and is more efficient than nitrocellulose in binding to small DNA fragments. Recently developed nylon membranes have super ceded the use of DBM paper as they are robust, reusable and bind (by cross-linking) to RNA on a brief exposure to UV light.

Northern hybridization is useful in the identification and separation of the RNA that is complementary to a specific DNA probe; this is a sensitive test for the detection of transcription of a DNA sequence used as the probe.


cells or tissue
1
extract RNA
RNA
2
agarose gel
electrophoresis
marker
RNA bands
sample
RNA bands
3
blotting
weight
glass plate
absorbent material
(e.g. paper towels)
filter paper
sheets
nitrocellulose
membrane filter
agarose
gel
cling
film
filter paper
wick
support
buffer solution
4
hybridization and
autoradiography
DNA probe
hybridized to an
RNA transcript

- **WESTERN BLOTTING**

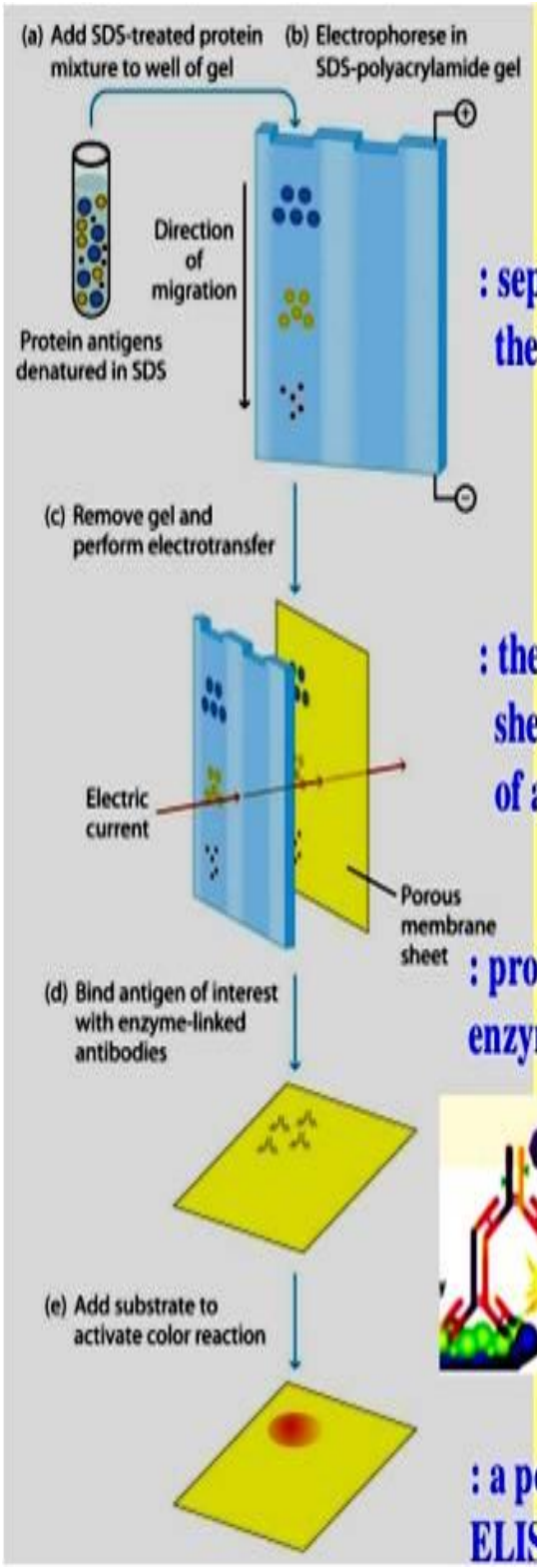
In western blotting, proteins are electrophoresed in polyacrylamide gel, transferred onto a nitrocellulose or nylon membrane (to which they bind strongly), and the protein bands are detected by their specific interaction with antibodies, lectins or some other compounds. A polyvinyl flouride hydrophobic membrane (0.45 µm pore size) is reported to have a high binding capacity for biomolecules, superior handling and staining characteristics and to be more sensitive than nitrocellulose. The various steps of this technique are briefly described below.

1. Protein bands are separated by polyacrylamide gel electrophoresis.

2. The protein bands are transferred onto a nitrocellulose or nylon membrane; initially, this was achieved by a capillary movement of buffer similar to Southern blotting *(capillary blotting)* but nowadays it is usually done by electrophoresis *(electrophoretic blotting).* Electrophoresis has been applied for the blotting step in Southern and northern hybridizations as well; in such cases, a buffer of low ionic strength (to avoid overheating during electrophoresis) and nylon membranes (since nucleic acids bind to nitrocellulose membrane only under conditions of high ionic strength) are used. The electrophoretic blotting, both of proteins and nucleic acids is much faster and more efficient than capillary blotting.

3. The specific protein bands are identified in a variety of ways. (/) Antibodies are the most commonly used as probes for detecting specific antigens, *(ii)* Lectins are used as probes for the identification of glycoproteins. These probes may themselves be radioactive or a radioactive molecule may be tagged to them. Often the identification process is based on a *'sandwich' reaction.* In such an approach, a species-specific second antibody or protein A of *Staphylococcus aureus* (protein A binds to certain subclasses of IgG antibodies) or streptavidin (it binds to biotinylated antibodies) is used to bind to the antibodies bound to the protein bands. These second molecules may be labelled with radioactive, enzyme or fluorescent tags; a single preparation of these labelled molecules can be employed as a general detector for various probes.

Western blotting



: separates the components according to their molecular weight.

: the proteins in the gel are transferred to the sheet of nitrocellulose or nylon by the passage of an electric current.

: probed with Ab & then radiolabeled or enzyme-linked 2nd Ab.

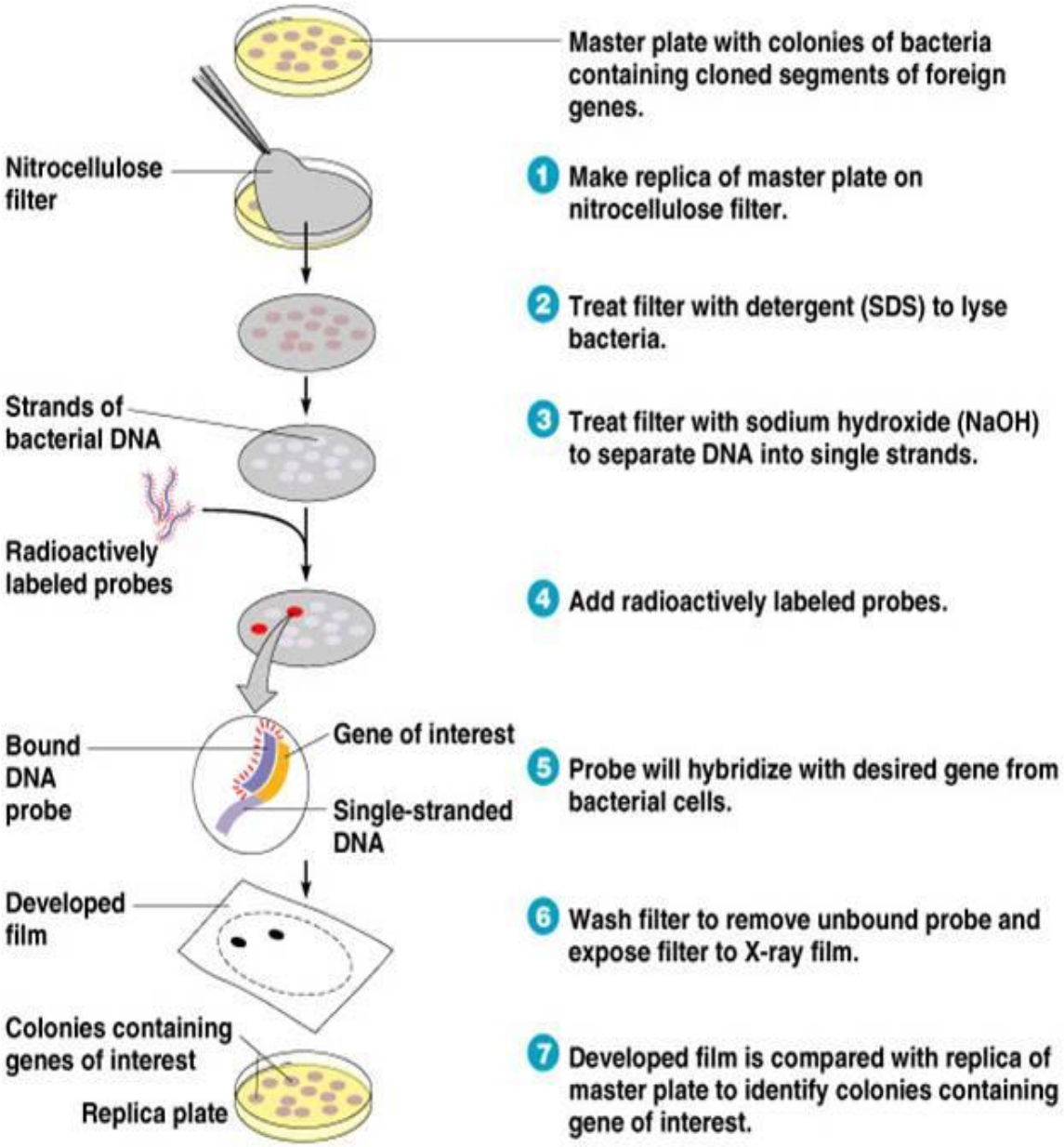
Expose x-ray film and develop

: a position is visualized by means of an ELISA reaction.

COLONY HYBRIDIZATION

This technique is used to identify those bacterial colonies in a plate, which contain a specific DNA sequence. These bacterial colonies are obtained from bacterial cells into which this sequence was introduced through genetic engineering, and the given sequence is represented by the probe used in the hybridization experiment. The procedure for colony hybridization was developed by Grunstein and Hogness in 1975, and is briefly described below (Fig. 2.I.1).

1. The bacterial cells subjected to transformation are plated onto a suitable agar plate; this is the ***master plate.***
2. The colonies of master plate are replica-plated onto a nitrocellulose filter membrane placed on agar medium. For ***replica plating,*** a block of wood or cork, of suitable diameter for the master plate, is covered with velvet cloth. This block is sterilized and then lowered into the master plate till the velvet touches all the colonies; the block is withdrawn and gently lowered onto the nitrocellulose filter so that the bacterial cells sticking onto the velvet are transferred onto the filter. The master plate is retained intact for later use. A reference point is marked both on the master plate and on the replica plate to facilitate later comparisons.
3. After the colonies appear, the filter is removed from the agar plate and treated with alkali (0.5 N NaOH) to lyse the bacterial cells. This also denatures the DNA released from these cells.
4. The filter is neutralized and then treated with proteinase K to digest and remove the proteins; the denatured DNA remains bound to the filter.
5. The filter is now baked at 80°C to fix the DNA; this yields the DNA-print of bacterial colonies in the same relative positions as those of the colonies themselves in the master plate.
6. The filter is now hybridized with the radioactive probe; the probe represents the sequence of DNA segment used for transformation. The unhybridized probe is removed by repeated washing. The probe used originally was RNA, but DNA probes are frequently used. The probe may also be labeled nonradioactively. The probes are generally 16 nucleotides long, although 14-bases long probes have been successfully used.
7. The colonies whose DNA hybridizes with the probe are detected by autoradiography; only these colonies show up in the autoradiograph.



The positions of colonies showing up in the autoradiograph are compared with the master plate to identify these colonies; these colonies contain the DNA segment in question. The colonies are then picked up for further studies. A variation of this procedure can be applied to phage plaques as well; this method uses the procedure *of plaque lift,* in which a nitrocellulose filter is applied to the upper surface of agar plates so that filter comes in direct contact with the plaques. Up to 5 x 10^4 recombinant phage particles are plated onto the appropriate bacterial lawn in a 9 cm square Petri dish, and incubated for 6-8 hr (small plaques) or overnight (large plaques). The plate is then cooled at 4°C for 1 hr to stiffen the top agar/agarose; a nitrocellulose sheet is now placed on the

agar/agarose surface and left for 30 sec to 2 min, and reference marks are made both on the filter as well as the plate. Both phage particles and unpackaged recombinant phage DNA from the plaques will stick to the filter. The nitrocellulose filter is now lifted carefully and placed on to a filter paper soaked in alkali to denature the DNA. (The plate itself is stored at 4°C for later isolation of the desired recombinant phage.) The filter is then neutralized by transferring it onto a filter paper soaked in neutral buffer, and then baked at 80°C *in vacuo.* Hybridization with labeled nucleic acid probe allows detection of the plaques having the desired recombinant DNA. Agar plugs representing the identified plaques are picked from the plate and recombinant phage are isolated. Plaque lift procedure allows several identical DNA prints to be made from a single plate, and each print can be screened with a different probe or they can be used for replicated screening. This procedure is probably the most widely used method of library screening. More recently, automated procedures for library screening have been developed.

Merits and Application of Colony Hybridization (Nucleic Acid Hybridization)

The nucleic acid hybridization provides a very versatile tool for library screening, and *potentially any sequence from any library can be isolated provided an appropriate probe is available.*

1. Stringent hybridization conditions can be used to ensure that only such sequences that are identical to the probe are identified, *e.g.*, identification of genomic clones corresponding to a specific cDNA, identification of overlapping clones in a chromosome walk, etc.
2. Less stringent hybridization conditions can be used to identify both identical and relate sequences, *e.g.*, use of a probe from one species to isolate a homologous clone from another species.
3. Probes corresponding to a conserved functional domain of a gene can be used to identify members of a gene family. For example, a DNA sequence conserved between *Drosophila* developmental genes *fushi terazu and Antenapedia* was used to isolate several clones from vertebrate cDNA libraries representing the large family of *Hox* genes that are central to animal development. In such applications the hybridization condition are less stringent.

==========

- **DNA SEQUENCING (SANGER'S METHOD OR DIDEOXY METHOD)**

Twice Nobel Prize winner Frederick Sanger (1977) developed a powerfull method for DNA sequencing that utilizes single stranded DNA as a template. This method is also called

dideoxynucleotide cain termination method. The requirements are : a primer with free 3' -OH ends to start DNA synthesis, DNA polymerase and dNTPs.

Fig. 6.22 shows the presence of free 3'-OH group at 3' end in dATP and no 2',3' -OH in 2',3'ddATP. In 2',3', ddATP a hydrogen atom is attached at 2' and 3' carbons instead of –OH hydroxyl group. If any of four ddNTPs binds, the chain elongation is terminated. Because ddNTPs do not have free 3'-OH end which is required for chain elongation. Therefore, no phosphodiester bond will be formed.

Four reaction tubes are labelled with A,T, G and C each containing single stranded DNA termplate (obtained by

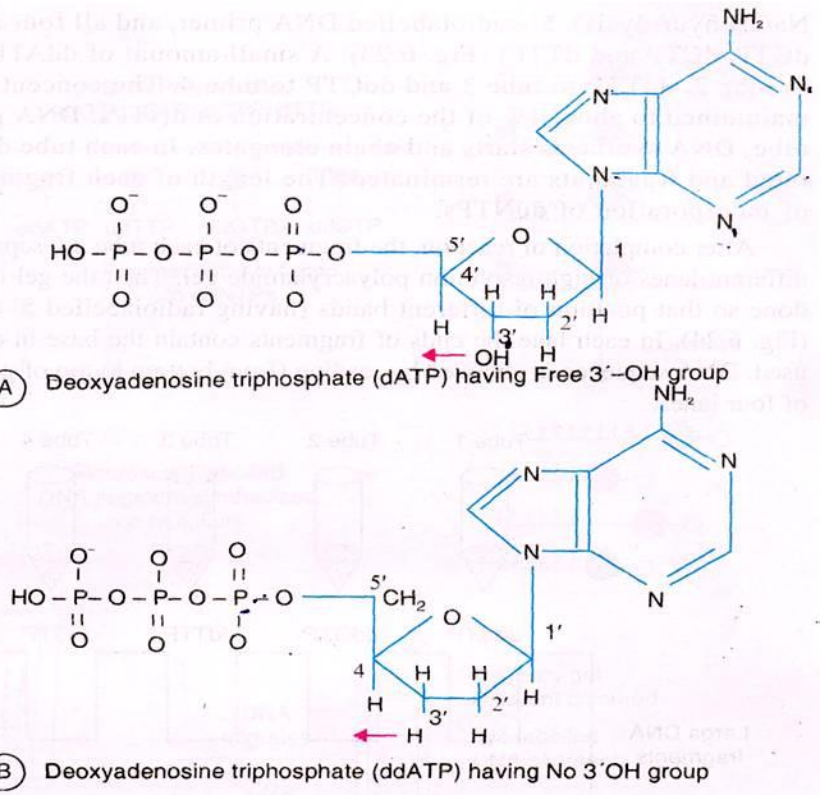


Fig. 6.22. Structure of deoxyadenosine triphosphate (dATP) (A) and dideoxyadenosine triphosphate (ddATP) (B)

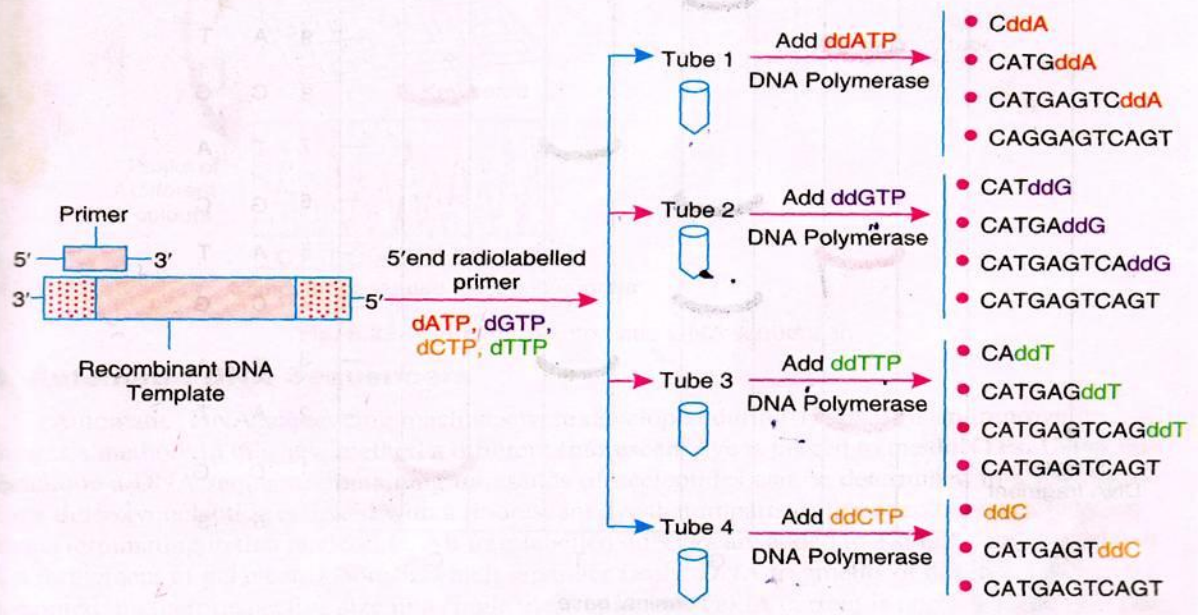


Fig. 6.23. Steps of chain termination by using ddNTPs base analogos.

NaOH hydrolysis), 5'-radiolabelled DNA primer, and all four radiolabelled dNTPs (dATP, dGTP, dCTP and dTTP) (Fig. 6.23). A small amount of ddATP is added to tube 1, ddGTP to tube 2, ddTTP to tube 3 and ddCTP to tube 4. The concentration of ddNTPs should be maintained to about 1% of the concentration of dNTPs. DNA polymerase is added to each tube, DNA synthesis starts and chain elongates. In each tube ddNTP is randomly incorporated and fragments are terminated. The length of each fragment depends on the position of incorporation of ddNTPs.

After completion of reaction, the fragments of each tube are seprated by electrophoresis in four different lanes of high resolution polyacrylamide gel. Then the gel is dried and autoradiography is done so that position of different bands (having radiolabelled 5' end) in each lane is observed (Fig. 6.24). In each lane the ends of fragments contain the base in correspondence to the ddNTPs used. DNA sequence is obtained by reading (from bottom to top of gel) the bands on autoradiogram of four lanes.

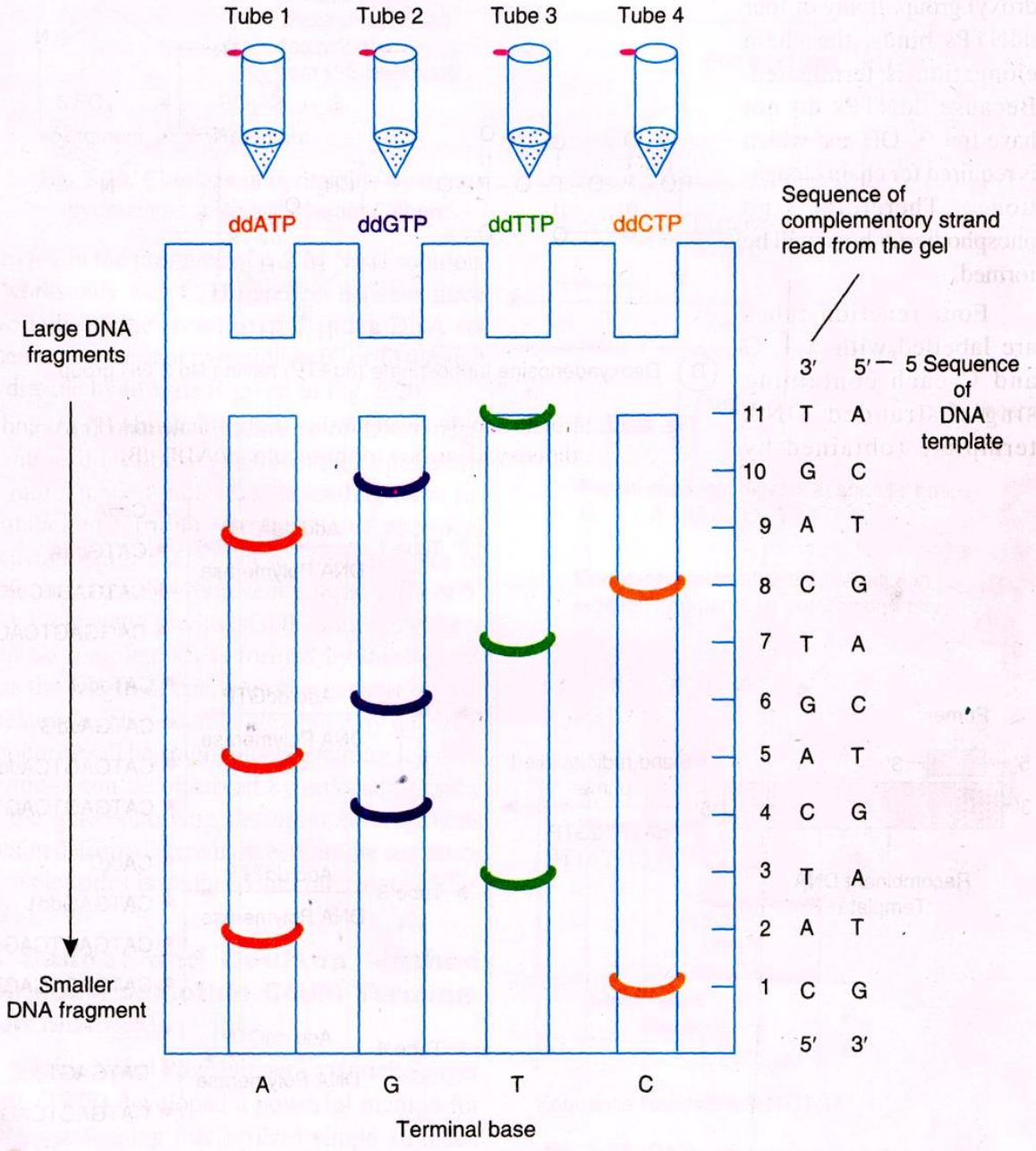


Fig. 6.24. Autoradiogram of electrophoresed DNA sequencing gel.

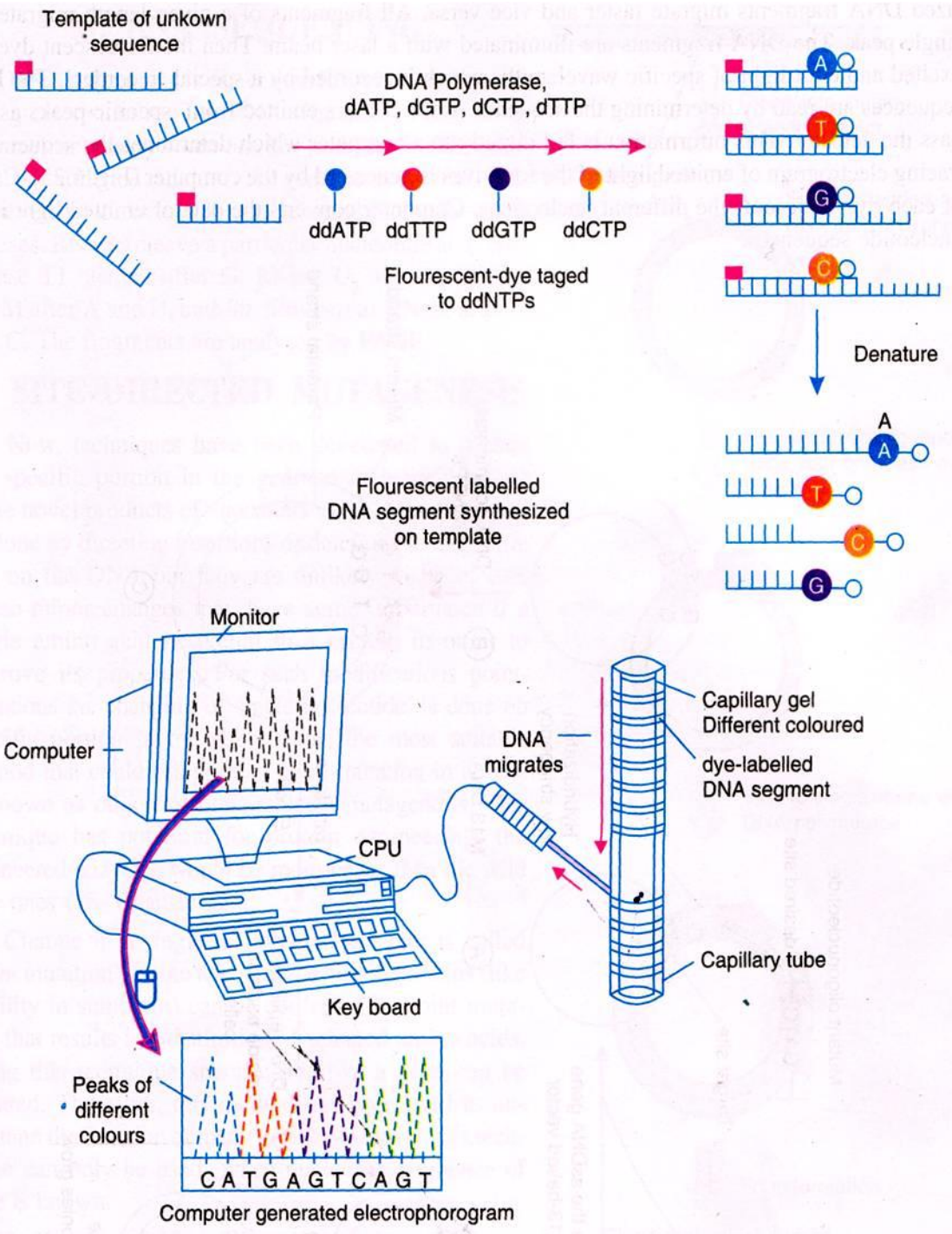


Fig. 6.25. Working of automatic DNA sequencer.

3. Automatic DNA Sequencers

Automatic DNA sequencing machines were developed during 1990s. It is an improvement of Sanger's method. In this new method a different fluorescent dye is tagged to the ddNTPs. Using this technique a DNA sequence containing thousands of necleotides can be determined in a few hours. Each dideoxynucleotide is linked with a fluorescent dye that imparts different colours to all the fragments terminating in that nucleotide. All four labelled ddNTPs are added to a single capillary tube. It is a refinement of gel electrophoresis which separates fastly. DNA fragments of different colours are separated by their respective size in a single electrophoretic gel. A current is applied to the gel. The negatively charged DNA strands migrate through the pores of gel towards the positive end. The small

sized DNA fragments migrate faster and vice versa. All fragments of a given length migrate in a single peak. The DNA fragments are illuminated with a laser beam. Then the fluorescent dyes are excited and emit light of specific wavelengths which is recorded by a special 'recorder'. The DNA sequences are read by determining the sequence of the colours emitted from specific peaks as they pass the detector. This information is fed directly to a computer which determines the sequence. A tracing electrogram of emitted light of the four dyes is generated by the computer (Fig. 6.25). Colour of each dye represents the different nucleotides. Computer converts the data of emitted light in the nucleotide sequences.