الجامعة المستنصرية

كلية العلوم

قسم علوم الحياة

المرحلة الثالثة

فرع التقانة

الهندسة الوراثية المنظري

يطلب من

مكتبة حسنين

السعر: ١٥٠٠ ا

Genetic engineering

The term genetic engineering is probably the label that most people would Several terms may be used to describe the technologies involved in manipulating genes. However, there are several other terms that can be used to describe the technology, including gene manipulation, gene cloning, recombinant DNA technology, genetic modification, and the new genetics. There are also legal definitions used in administering regulatory mechanisms in countries where genetic engineering is practised.

Genetic engineering is the direct manipulation of an organism's genome using biotechnology. It is a set of technologies used to change the genetic makeup of cells, including the transfer of genes within and across species boundaries to produce improved or novel organisms. New DNA may be inserted in the host genome by first isolating and copying the genetic material of interest using molecular cloning methods to generate a DNA sequence, or by synthesizing the DNA, and then inserting this construct into the host organism.

The science of genetic engineering originated in the late 1960s, the enzyme DNA ligase was isolated. This enzyme can join two strands of DNA together, a prerequisite for the construction of recombinant molecules, and can be regarded as a sort of molecular glue. and early 1970s with the discovery of restriction enzymes . The 1978 Nobel prize for physiology went to the discoverer of restriction enzymes, Hamilton O. Smith, and the first people to use these tools to analyze the genetics of a virus, Daniel Nathans and Werner Arber.

Restriction enzymes make it possible to remove a bit of DNA from one organism's chromosome and to insert it into another organism's chromosome . This allows for the production of new combinations of genes that may not exist in nature.

The first recombinant DNA molecules were generated at Stanford University in 1972, utilising the cleavage properties of restriction enzymes (scissors) and the ability of DNA ligase to join DNA strands together (glue). The importance of these first tentative experiments cannot be overestimated. Scientists could now join different DNA

molecules together and could link the DNA of one organism to that of a completely different organism. The methodology was tended in 1973 by joining DNA fragments to the plasmid pSC101, ensure that the target sequence is replicated in a suitable host cell. which is an extrachromosomal element isolated from the bacterium Escherichia coli. These recombinant molecules behaved as replicons; that is, they could replicate when introduced into E. coli cells. Thus, by creating recombinant molecules in vitro, and placing the construct in a bacterial cell where it could replicate in vivo, specific fragments of DNA could be isolated from bacterial colonies that formed clones (colonies formed from a single cell, in which all cells are identical) when grown on agar plates. This development marked the emergence of the technology that became known as gene cloning

An organism that is generated through genetic engineering is considered to be a genetically modified organism (GMO). The first GMOs were bacteria generated in 1973 and GM mice in 1974. Insulin-producing bacteria were commercialized and the first diabetic patient in the world was injected with human insulin made in bacteria in December 1980, making this the first genetically engineered product to enter medical practice , and genetically modified food has been sold since 1994. GloFish, the first GMO designed as a pet, was first sold in the United States in December 2003.

The interferons are another medically important group of peptides that became available in abundance only after the development of genetic engineering techniques. Interferon was useful for treating viral infections, and there were strong indications that it might be effective against some cancers. Before the advent of genetic engineering techniques, it took laborious processing of thousands of units of human blood to obtain enough interferon to treat a few patients. Other medically useful human peptides that have been made widely available because of genetic engineering are human growth hormone, which is used to treat persons with congenital dwarfism and tissue-type plasminogen activator (t-PA), which is a promising new treatment for persons who suffer a heart attack. With the development of retroviral vectors in the early 1980s, the possibility of efficient gene transfer into mammalian cells for the purpose of gene therapy became widely accepted.

1990 deemed America to became the first country to allow new genes be introduced into human beings . A gene drug was used to treat a 4 year-old girl with severe combined immune deficiency (SCID). Victims of SCID the lack of gene that controls the production commands vital to immune functioning. SCID patients prior to gene treatment had to live inside sanitized plastic bubbles. In early 1991, a 9 year-old girl with SCID deficiency was also treated with the same gene therapy. In 2000 it was announced that three French infants born with SCID had been cured using a more refined version of this technique

A **genetically modified organism (GMO)** is any organism whose genetic material has been altered using genetic engineering techniques (i.e. genetically *engineered* organism). GMOs are the source of medicines and genetically modified foods and are also widely used in scientific research and to produce other goods. The term A more specifically defined type of GMO is a "Transgenic Organism". This is an organism whose genetic makeup has been altered by the addition of genetic material from another, unrelated organism. This should not be confused with the more general way in which "GMO" is used to classify genetically altered organisms, as typically GMOs are organisms whose genetic makeup has been altered without the addition of genetic material from an unrelated organism.

The first genetically modified mouse was in 1973 and the first plant was produced in 1983.

if genetic material from another species is added to the host, the resulting organism is called transgenic. If genetic material from the same species or a species that can naturally breed with the host is used the resulting organism is called cisgenic. Genetic engineering can also be used to remove genetic material from the target organism, creating a gene knockout organism

Genetic Engineering Applications

Medicine

Genetic engineering has resulted in a series of medical products. The first two commercially prepared products from recombinant DNA technology were insulin and human growth hormone, both of which were cultured in the E. coli bacteria. Since then a plethora of products have appeared on the market, including the following abbreviated list, all made in E. coli:

Bionote

A **vaccine** is usually a harmless version of a bacterium or virus that is injected into an organism to activate the immune system to attack and destroy similar substances in the future.

- **Tumor necrosis factor**. Treatment for certain tumor cells
- **Interleukin-2 (IL-2)**. Cancer treatment, immune deficiency, and HIV infection treatment
- **Prourokinase**. Treatment for heart attacks
- **Taxol**. Treatment for ovarian cancer
- **Interferon**. Treatment for cancer and viral infections

In addition, a number of *vaccines* are now commercially prepared from recombinant hosts. At one time vaccines were made by denaturing the disease and then injecting it into humans with the hope that it would activate their immune system to fight future intrusions by that invader. Unfortunately, the patient sometimes still ended up with the disease.

With DNA technology, only the identifiable outside shell of the microorganism is needed, copied, and injected into a harmless host to create the vaccine. This method is likely to be much safer because the actual disease-causing microbe is not transferred to the host. The immune system is activated by specific proteins on the surface of the microorganism -e. DNA technology takes that into account and only utilizes identifying surface features for the vaccine. Currently vaccines

for the hepatitis B virus, herpes type 2 viruses, and malaria are in development for trial use in the near future.

Agriculture

Crop plants have been and continue to be the focus of biotechnology as efforts are made to improve yield and profitability by improving crop resistance to insects and certain herbicides and delaying ripening (for better transport and spoilage resistance). The creation of a transgenic plant, one that has received genes from another organism, proved more difficult than animals. Unlike animals, finding a vector for plants proved to be difficult until the isolation of the *Ti plasmid*, harvested from a tumor-inducing (Ti) bacteria found in the soil. The plasmid is "shot" into a cell, where the plasmid readily attaches to the plant's DNA. Although successful in fruits and vegetables, the Ti plasmid has generated limited success in grain crops.

Creating a crop that is resistant to a specific herbicide proved to be a success because the herbicide eliminated weed competition from the crop plant. Researchers discovered herbicide-resistant bacteria, isolated the genes responsible for the condition, and "shot" them into a crop plant, which then proved to be resistant to that herbicide. Similarly, insect-resistant plants are becoming available as researchers discover bacterial enzymes that destroy or immobilize unwanted herbivores, and others that increase nitrogen fixation in the soil for use by plants.

Geneticists are on the threshold of a major agricultural breakthrough. All plants need nitrogen to grow. In fact, nitrogen is one of the three most important nutrients a plant requires. Although the atmosphere is approximately 78 percent nitrogen, it is in a form that is unusable to plants. However, a naturally occurring *rhizobium* bacterium is found in the soil and converts atmospheric nitrogen into a form usable by plants. These nitrogen-fixing bacteria are also found naturally occurring in the legumes of certain plants such as soybeans and peanuts. Because they contain these unusual bacteria, they can grow in nitrogen-deficient soil that prohibits the growth of other crop plants. Researchers hope that by isolating these bacteria, they can identify the DNA segment that codes for nitrogen fixation, remove the segment, and insert it into the DNA of a profitable cash crop! In so doing, the new transgenic crop plants could

live in new fringe territories, which are areas normally not suitable for their growth, and grow in current locations without the addition of costly fertilizers!

Projects with genetically modified products

Bt cotton: Since we began developing cotton varieties in 1984, using GM technologies, enormous improvements have been made. Australia now has the highest cotton yields in the world, exporting cotton worth $2.5 billion each year. We've reduced Australian growers' reliance on insecticides and improved their water use efficiency. Currently, more than 95 per cent of the Australian cotton crop is grown from CSIRO bred varieties, which have reduced pesticide use by up to 85 per cent and herbicide use by about 52 per cent.

DHA canola: We have developed canola plants which produce high quality oils rich in omega-3 DHA (docosahexaenoic acid). This nutrient is currently only found in beneficial quantities in ocean-based algae, and the fish that eat it. This product could break the world's reliance on fish stocks while meeting the increasing demand for these healthy long-chain omega-3 oils.

SHO Safflower: By engineering safflower to contain 94 per cent oleic acid, we've produced the world's highest source of oleic acid. This chemical can be used in lubricants, oleochemicals and transformer oil.

Leaf oil: We have engineered tobacco plants to have oilseed-like levels of oil in their leaves (around 35 per cent). This product could provide an economically competitive renewable alternative to petroleum diesel.

BT cowpeas: We are part of a global project to improve cowpea production in Africa and are making progress towards incorporating 'built-in' insect pest protection that could help reduce food shortages in some African regions.

Animal Husbandry

Neither the use of animal vaccines nor adding bovine growth hormones to cows to dramatically increase milk production can match the real excitement in animal husbandry: transgenic animals and clones.

Transgenic animals model advancements in DNA technology in their development. The mechanism for creating one can be described in three steps:

1. Healthy egg cells are removed from a female of the host animal and fertilized in the laboratory.
2. The desired gene from another species is identified, isolated, and cloned.
3. The cloned genes are injected directly into the eggs, which are then surgically implanted in the host female, where the embryo undergoes a normal development process.

It is hoped that this process will provide a cheap and rapid means of generating desired enzymes, other proteins, and increased production of meat, wool, and other animal products through common, natural functions.

Ever since 1997 when Dolly was cloned, research and experimentation to clone useful livestock has continued unceasingly. The attractiveness of cloning is the knowledge that the offspring will be genetically identical to the parent as in asexual reproduction. Four steps describe the general process:

1. A differentiated cell, one that has become specialized during development, with its diploid nucleus is removed from an animal to provide the DNA source for the clone.
2. An egg cell from a similar animal is recovered and the nucleus is removed, leaving only the cytoplasm and cytoplasm organelles.
3. The two egg cells are fused with an electric current to form a single diploid cell, which then begins normal cell division.
4. The developing embryo is placed in a surrogate mother, who then undergoes a normal pregnancy.

Working with nucleic acids

Before examining some of the specific techniques used in gene manipulation, it is useful to consider the basic methods required for handling, quantifying, and analysing nucleic acid molecules. It is often difficult to make the link between theoretical and practical aspects of a subject, and an appreciation of the methods used in routine work with nucleic acids may be of help when the more detailed techniques of gene cloning and analysis are described.

Laboratory requirements

One of the striking aspects of gene manipulation technology is that many of the procedures can be carried out with a fairly basic laboratory setup. The requirements can be summarised under three headings:

1. General laboratory facilities
2. Cell culture and containment
3. Processing and analysis

Isolation of DNA and RNA

Every gene manipulation experiment requires a source of nucleic acid, in the form of either DNA or RNA. It is therefore important that reliable methods are available for isolating these components from cells. There are three basic requirements: (1) opening the cells in the sample to expose the nucleic acids for further processing, (2) separation of the nucleic acids from other cell components, and (3) recovery of the nucleic acid in purified form.

A variety of techniques may be used, ranging from simple procedures with few steps up to more complex purifications involving several different stages.

The first step in any isolation protocol is disruption of the starting material, which may be viral, bacterial, plant, or animal. The method used to open cells should be as gentle as possible, preferably utilising enzymatic degradation of cell wall material (if present) Cells have to be opened to enable nucleic acids to be isolated; opening cells should be done as gently as possible to avoid shearing large DNA molecules. and detergent lysis of cell membranes. If more vigorous methods of cell

disruption are required (as is the case with some types of plant cell material.

Following cell disruption, most methods involve **a deproteinisation** stage. This can be achieved by one or more extractions using phenol or phenol/chloroform mixtures. On the formation of an emulsion and subsequent centrifugation to separate the phases, protein molecules partition into the phenol phase and accumulate at the interface. The nucleic acids remain mostly in the upper aqueous phase and may be precipitated from solution using isopropanol or ethanol. Some techniques do not require the use of phenolic mixtures and are safer and more pleasant to use than phenol-based extraction media.

If a DNA preparation is required, the **enzyme ribonuclease (RNase)** can be used to digest the RNA in the preparation. If mRNA is needed Once broken open, cell preparations can be deproteinised and the nucleic acids purified by a range of techniques. Some applications require highly purified nucleic acid preparations; some may be able to use partially purified DNA or RNA. for cDNA synthesis, a further purification can be performed by **affinity chromatography** using oligo(dT)-cellulose to bind the poly(A) tails of eukaryotic mRNAs . This gives substantial enrichment for mRNA and enables most of the contaminating DNA, rRNA, and tRNA to be removed.

The technique of gradient centrifugation is often used to prepare DNA, particularly plasmid DNA (pDNA). In this technique a caesium chloride (CsCl) solution containing the DNA preparation is spun at high speed in an ultracentrifuge. Over a long period (up to 48 h in some cases) a density gradient is formed and the pDNA forms a band Solutions of nucleic acids are used to enable very small amounts to be handled easily, measured, and dispensed. At one position in the centrifuge tube. The band may be taken off and the CsCl removed by dialysis to give a pure preparation of pDNA. As an alternative to gradient centrifugation, size exclusion chromatography (gel filtration) or similar techniques may be used.

Handling and quantification of nucleic acids

It is often necessary to use very small amounts of nucleic acid (typically micro-, nano-, or picograms) during a cloning experiment. It is obviously

impossible to handle these amounts directly, so most of the measurements that are done involve the use of aqueous solutions of DNA and RNA. The concentration of a solution of nucleic acid can be determined by measuring the absorbance at 260 nm, using a spectrophotometer. An A260 of 1.0 is equivalent to a concentration of 50 $\mu g\ ml^{-1}$ for double-stranded DNA, or 40 $\mu g\ ml^{-1}$ for single stranded DNA or RNA. If the A280 is also determined, the A260/A280 ratio indicates if there are contaminants present, such as residual phenol or protein. The A260/A280 ratio should be around 1.8 for pure DNA and 2.0 for pure RNA preparations. In addition to spectrophotometric methods, the concentration of DNA may be estimated by monitoring the fluorescence of bound ethidium bromide. This dye binds between the DNA bases (intercalates) and fluoresces orange when illuminated with ultraviolet (uv) light. By comparing the fluorescence of the sample with that of a series of standards, an estimate of the concentration may be obtained. This method can detect as little as 1--5 ng of DNA and may be used when uv-absorbing contaminants make spectrophotometric measurements impossible. Having determined the concentration of a solution of nucleic acid, any amount (in theory) may be dispensed by taking the appropriate volume of solution. In this way nanogram or picogram amounts may be dispensed with reasonable accuracy. Precipitation of nucleic acids is an essential technique that is Nucleic acids can be concentrated by using alcohol to precipitate the DNA or RNA from solution; the precipitate is recovered by centrifugation and can then be processed as required. used in a variety of applications. The two most commonly used precipitants are isopropanol and ethanol, ethanol being the preferred choice for most applications. When added to a DNA solution in a ratio, by volume, of 2:1 in the presence of 0.2 M salt, ethanol causes the nucleic acids to come out of solution. Although it used to be thought that low temperatures (−20◦C or −70◦C) were necessary, this is not an absolute requirement, and 0◦C appears to be adequate. After precipitation the nucleic acid can be recovered by centrifugation, which causes a pellet of nucleic acid material to form at the bottom of the tube. The pellet can be dried and the nucleic acid resuspended in the buffer appropriate to the next stage of the experiment.

Labelling of nucleic acids

A major problem encountered in many cloning procedures is that of keeping track of the small amounts of nucleic acid involved. This problem is magnified at each stage of the process, because losses mean that the amount of material usually diminishes after each step. One way of tracing the material is to label the nucleic acid with a marker of some sort, so that the material can be identified at each stage of the procedure. So what can be used as the label?

Types of label – radioactive or not?

Radioactive tracers have been used extensively in biochemistry and molecular biology for a long time, and procedures are now well established. The most common isotopes used are tritium (3H), carbon-14 (14C), sulphur-35 (35S), and phosphorus-32 (32P). Tritium and 14C are low-energy emitters, with 35S being a 'medium'-energy emitter and 32P being a high-energy emitter. Thus, 32P is more hazardous than Radioactive isotopes are often used to label nucleic acids, although they are more hazardous than non-radioactive labelling methods. the other isotopes and requires particular care in use. There are also strict statutory requirements for the storage and disposal of radioactive waste materials. Partly because of the inherent dangers of working with high-energy isotopes, the use of alternative technologies such as fluorescent dyes, or enzyme-linked labels, has become popular. Although these methods do offer advantages for particular applications (such as DNA sequencing), for routine tracing experiments a radioactive label is still often the preferred choice. In this case the term radiolabelling is often used to describe the technique. One way of tracing DNA and RNA samples is to label the nucleic acid with a radioactive molecule (usually a deoxynucleoside triphosphate (dNTP), labelled with 3H or 32P), so that portions of each reaction may be counted in a scintillation counter to determine the amount of nucleic acid present. This is usually done by calculation, taking into account the amount of radioactivity present in the sample. A second application of radiolabelling is in the production of highly radioactive nucleic acid molecules for use in hybridisation experiments. Such molecules are known as radioactive probes and Radioactive probes are very useful for identifying specific DNA or RNA

sequences. have a variety of uses. The difference between labelling for tracing purposes and labelling for probes is largely one of specific activity, that is, the measure of how radioactive the molecule is. For tracing purposes, a low specific activity will suffice, but for probes a high specific activity is necessary. In probe preparation the radioactive label is usually the high-energy ß-emitter 32P.

End labelling

In the end labelling technique, the enzyme polynucleotide kinase is used to transfer the terminal phosphate group of ATP onto 5 -hydroxyl termini of nucleic acid molecules. If the ATP donor is radioactively labelled, this produces a labelled nucleic acid of relatively low specific activity, as only the termini of each molecule become radioactive

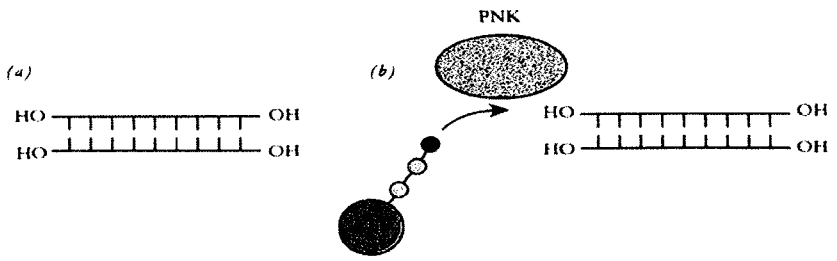


Fig. 3.2 End labelling DNA using polynucleotide kinase (PNK). (*a*) DNA is dephosphorylated using phosphatase to generate 5′-OH groups. (*b*) The terminal phosphate of [γ-^{32}P]ATP (solid circle) is then transferred to the 5′ terminus by PNK. The reaction can also occur as an exchange reaction with 5′-phosphate termini.

Nick translation

Nick translation relies on the ability of the enzyme DNA polymerase I (to translate (move along the DNA) a nick created in the phosphodiester backbone of the DNA double helix. Nicks may occur naturally or may be caused by a low concentration of the nuclease DNase I in the reaction mixture. DNA polymerase I catalyses a strand-replacement reaction that incorporates new dNTPs into the DNA chain. If one of the dNTPs supplied is radioactive, the result is a highly labelled DNA molecule

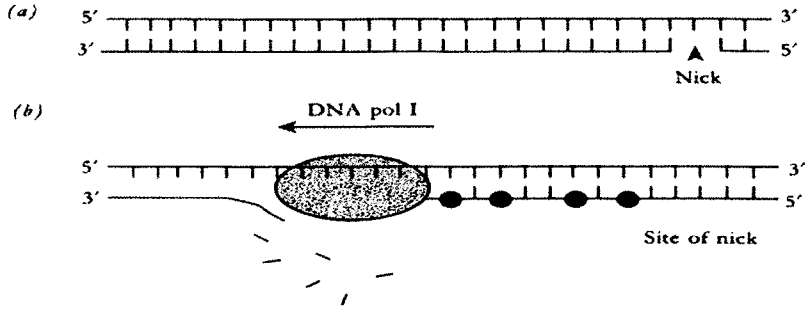


Fig. 3.3 Labelling DNA by nick translation. (*a*) A single-strand nick is introduced into the phosphodiester backbone of a DNA fragment using DNase I. (*b*) DNA polymerase I then synthesises a copy of the template strand, degrading the non-template strand with its 5′→3′ exonuclease activity. If [α-^{32}P]dNTP is supplied this will be incorporated into the newly synthesised strand (solid circles).

Labelling by primer extension

Labelling by primer extension refers to a technique that uses random oligonucleotides (usually hexadeoxyribonucleotide molecules -- sequences of six deoxynucleotides) to prime synthesis of a DNA strand In most labelling reactions not all the radioactive dNTP is incorporated into the target sequence, and non-incorporated isotope is usually removed before using the probe. by DNA polymerase. The DNA to be labelled is denatured by heating, and the oligonucleotide primers annealed to the single-stranded DNAs. The Klenow fragment of DNA polymerase .can then synthesise a copy of the template, primed from the 3 -hydroxyl group of the oligonucleotide. If a labelled dNTP is incorporated, DNA of very high specific activity is produced . In a radiolabelling reaction it is often desirable to separate the labelled DNA from the unincorporated nucleotides present in the reaction mixture. A simple way of doing this is to carry out a smallscale gel filtration step using a suitable medium. The whole process can be carried out in a Pasteur pipette, with the labelled DNA coming through the column first, followed by the free nucleotides. Fractions can be collected and monitored for radioactivity, and the data used to calculate total activity of the DNA, specific activity, and percentage incorporation of the isotope.

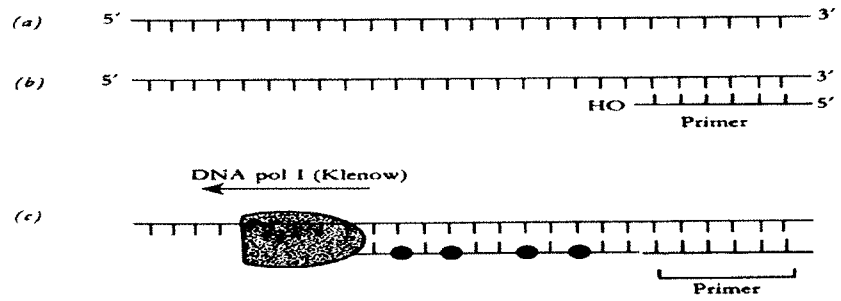
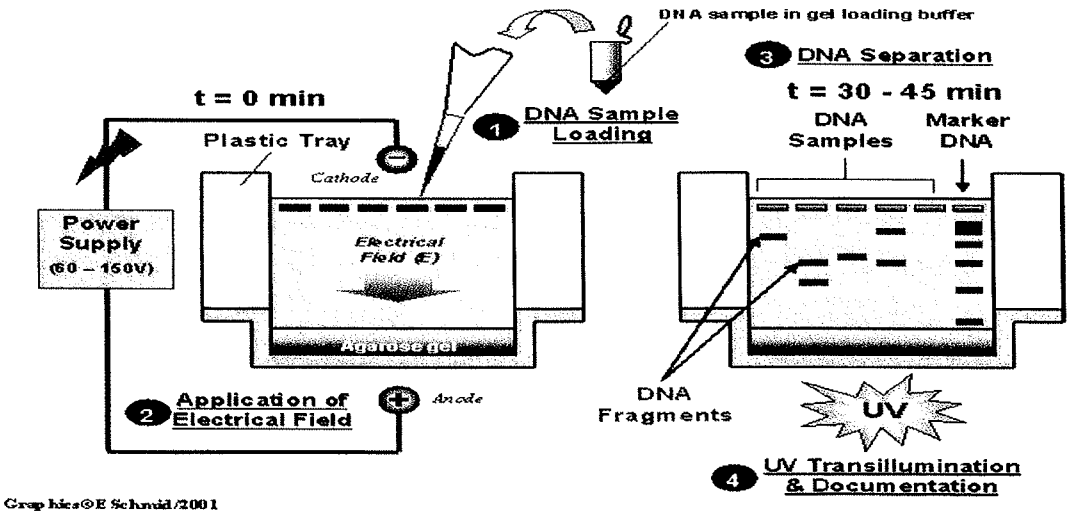


Fig. 3.4 Labelling DNA by primer extension (oligolabelling). (*a*) DNA is denatured to give single-stranded molecules. (*b*) An oligonucleotide primer is then added to give a short double-stranded region with a free 3′-OH group. (*c*) The Klenow fragment of DNA polymerase I can then synthesise a copy of the template strand from the primer, incorporating [α-^{32}P]dNTP (solid circles) to produce a labelled molecule with a very high specific activity.

Gel electrophoresis

The technique of gel electrophoresis is vital to the genetic engineer, as it represents the main way by which nucleic acid fragments may be visualised directly. The method relies on the fact that nucleic acids are polyanionic at neutral pH; that is, they carry multiple negative charges because of the phosphate groups on the phosphodiester backbone of the nucleic acid strands. This means that the molecules will migrate towards the positive electrode when placed in an electric field. As the negative charges are distributed evenly along the DNA molecule, the charge/mass ratio is constant; thus, mobility depends on fragment length. The technique is carried out using a gel matrix, which separates the nucleic acid molecules according to size. A typical nucleic acid electrophoresis setup is shown in Fig. 3.5. The type of matrix used for electrophoresis has important conseSeparation of biomolecules by quences for the degree of separation achieved, which is dependent gel electrophoresis is one of the most powerful techniques in molecular biology. on the porosity of the matrix. Two gel types are commonly used: agarose and polyacrylamide. Agarose is extracted from seaweed and can be purchased as a dry powder that is melted in buffer at an appropriate concentration, normally in the range 0.3--2.0% (w/v). On cooling, the agarose sets to form the gel. Agarose gels are usually run in the apparatus shown in Fig. 3.5, using the submerged agarose gel electrophoresis (SAGE) technique. Polyacrylamide-based gel electrophoresis (PAGE) is sometimes used to separate small nucleic acid molecules; in applications such as DNA sequencing (see Section 3.7), as the pore size is smaller than that achieved with agarose. Electrophoresis is carried out by placing the

nucleic acid samples in the gel and applying a potential difference across it. This potential difference is maintained until a marker dye (usually bromophenol blue, added to the sample prior to loading) reaches the end of the gel. The nucleic acids in the gel are usually visualised by staining with the intercalating dye ethidium bromide and examining under uv light. Nucleic acids show up as orange bands, which can be photographed to provide a record . The data can be used to estimate the sizes of unknown fragments by construction of a calibration curve using standards of known size, as migration is inversely proportional to the log10 of the number of base pairs. This is particularly useful in the technique of restriction mapping . In addition to its use in the analysis of nucleic acids, PAGE is used extensively for the analysis of proteins. The methodology is different from that used for nucleic acids, but the basic principles are similar. One common technique is SDS-PAGE, in which the detergent SDS (sodium dodecyl sulphate) is used to denature multisubunit proteins and cover the protein molecules with negative charges. In this way the inherent charge of the protein is masked, and the charge/mass ratio becomes constant. Thus, proteins can be separated according to their size in a similar way to DNA molecules.



Restriction enzymes – cutting DNA

The restriction enzymes, which cut DNA at defined sites, represent one of the most important groups of enzymes for the manipulation of DNA. These enzymes are found in bacterial cells, where they function as part of a protective mechanism called the restriction-modification system. In this system the restriction enzyme hydrolyses any exogenous DNA that appears in the cell. To prevent the enzyme acting on Restriction enzymes act as a 'protection' system for bacteria in that they hydrolyse exogenous DNA that is not methylated by the host modification enzyme. the host cell DNA, the modification enzyme of the system (a methylase) modifies the host DNA by methylation of particular bases in the restriction enzyme's recognition sequence, which prevents the enzyme from cutting the DNA.

There are two different kinds of restriction enzymes:

(1) Exonucleases catalyses hydrolysis of terminal nucleotides from the end of DNA or RNA molecule either 5'to 3' direction or 3' to 5' direction. Example: exonuclease I, exonuclease II etc.

(2) Endonucleases can recognize specific base sequence (restriction site) within DNA or RNA molecule and cleave internal phosphodiester bonds within a DNA molecule. Example: EcoRI, Hind III, BamHI etc.

Types of restriction endonucleases

Type I restriction endonucleases are complex endonucleases, and have recognition sequences of about 15 bp; they cleave the DNA about 1000 bp away from the 5'-end of the sequence "TCA" located within the recognition site, e.g., Eco K, Eco B, etc.

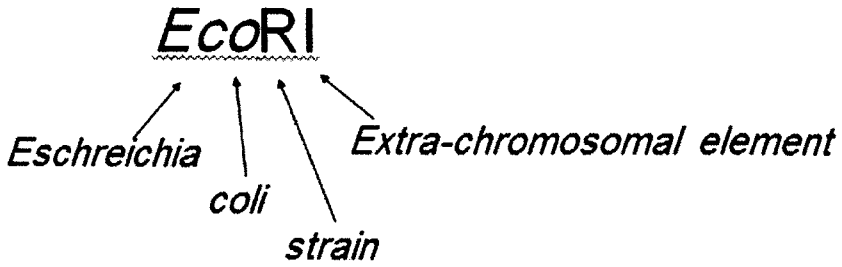
Type II restriction endonucleases are remarkably stable and induce cleavage either, in most cases, within or immediately outside their recognition sequences, which are symmetrical. More than 350 different type II endonucleases with over 100 different recognition sequences are known.

They require Mg^{2+} ions for cleavage. The first type II enzyme to be isolated was HindII in 1970. Only type II restriction endonucleases are used for restriction mapping and gene cloning in view of their cleavage only at specific sites.

Type III restriction endonucleases are intermediate between the type I and type II enzymes; they cleave DNA in the immediate vicinity of their recognition sites, e.g., EcoP1, EcoP15, HinfIII, etc. Type 1 and Type III restriction enzymes are not used in gene cloning. The Type III enzymes recognize asymmetric target sites, and cleave the DNA duplex on one side of the recognition sequence up to 20 bp away.

Restriction Endonuclease Nomenclature:

Restriction enzyme nomenclature is based on a number of conventions. The generic and specific names of the organism in which the enzyme is found are used to provide the first part of the designation which comprises the first letter of the generic name and the first two letters of the specific name. Thus, an enzyme from a strain of *Escherichia coli* is termed Eco, one from Bacillus amyloliquefaciens is Bam, and so on. Further descriptors may be added, depending on the bacterial strain involved and on the presence or absence of extrachromosomal elements.



Recognition Sequences:

In a palindrome, the base sequence in the second half of a DNA strand is the mirror image of the sequence in its first half; consequently, the

complementary DNA strand of a double helix also shows the same situation .

But in a palindrome with rotational symmetry, the base sequence in the first halt of one strand of a DNA double helix is the mirror image of the second half of its complementary strand . Thus in such palindromes, the base sequence in both the strands of a DNA duplex reads the same when read from the same end (either 5′ or 3′) of both the strand

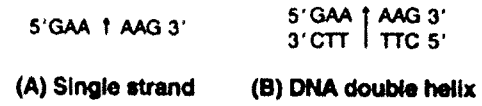


FIG. 2.1. A palindromic sequence. (A) Sequence in a single DNA strand. (B) Sequence in a DNA double helix. The arrow represents the axis of symmetry.

Most of the type II restriction endonucleases have recognition sites of 4, 5 or 6 bp (base pairs), which are predominantly GC-rich. Longer palindromic target sequences are also known, and so are nonpalindromic ones (specific for some enzymes). Some restriction enzymes have ambiguities in their recognition sites, e.g., Hi nil, EcoRII, etc., so that they may recognise upto 4 or even more, e.g., Sfil, different target sequences

Cleavage Patterns of Some Common Restriction Endonucleases:

Three types of fragment may be produced:

(1) blunt ends (sometimes known as flush-ended fragments),

(2) fragments with protruding 3 ends,

(3) fragments with protruding 5 ends.

→ cutting two different DNA samples with the same enzyme and mixing the fragments together, recombinant DNA can be produced. This is one of the most useful applications of restriction enzymes and is a vital part of many manipulations in genetic engineering.

Table 4.1. Recognition sequences and cutting sites for some restriction endonucleases

Enzyme	Recognition sequence	Cutting sites	Ends
BamHI	5′-GGATCC-3′	G↓GATCC CCTAG↑G	5′
EcoRI	5′-GAATTC-3′	G↓AATTC CTTAA↑G	5′
HaeIII	5′-GGCC-3′	GG↓CC CC↑GG	Blunt
HpaI	5′-GTTAAC-3′	GTT↓AAC CAA↑TTG	Blunt
PstI	5′-CTGCAG-3′	CTGCA↓G G↑ACGTC	3′

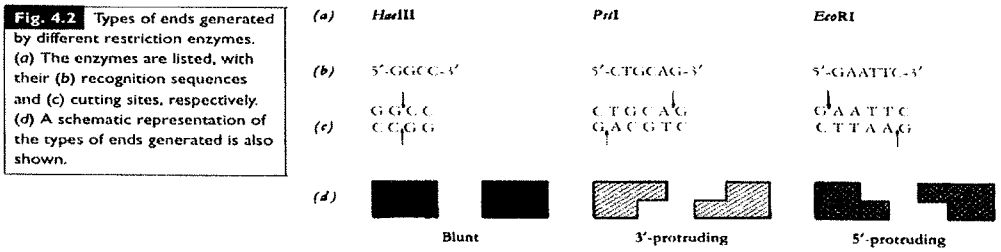
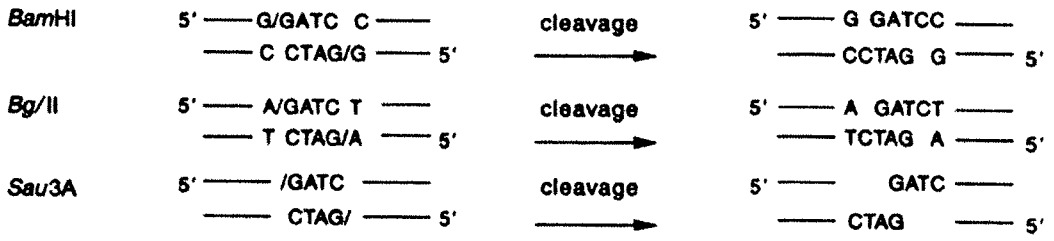


Fig. 4.2 Types of ends generated by different restriction enzymes. (*a*) The enzymes are listed, with their (*b*) recognition sequences and (*c*) cutting sites, respectively. (*d*) A schematic representation of the types of ends generated is also shown.

In addition, two or more restriction enzymes having different recognition sites generate the same sticky end for BamHI, BglII and Sau3A; they all produce GATC sticky ends.



→ **There are cases where two different restriction enzymes recognize the same target sequence, but one of them is able to recognize both methylated as well non-methylated target sequences, while the other enzyme can recognize only the non-methylated target sequence; such enzymes are known as isoschizomers.**

→ For example, restriction enzymes HpaII and MspI are isoschizomers; they both recognize the sequence 5'CCGG3' when it is unmethylated. But when the second C of the sequence is methylated, HpaII can no longer recognize it, while MspI recognizes it just as well as it does the unmethylated sequence.

Isoschizomers are very useful in determining the state of methylation of a DNA molecule.

Frequency of Recognition Sites:

The frequency of recognition sites for a restriction enzyme likely to be present in a DNA molecule can be estimated theoretically if we make the following two assumptions: (1) the frequencies of A, T. G and C in the molecule are equal, i.e., it has a G: C content of 50%, and (2) the bases are present in a random order.

In such a case, any given base, say, A, is expected to occur on, an overage, once every 4 bases (=41), a sequence of two bases, e.g., AT, would occur every 16 bases (=42) and so on. Therefore, a recognition sequence of 4 bases, e.g., 5'CCGG3' (for Hpall), is expected to occur, on an average, once every 44 (= 256) bp along a DNA molecule. In contrast, a 6 base recognition sequence, e.g., 5'GAATTC3' (for EcoRI) should be much less frequent and occur once every 46 bp (= 4,096 bp.(

The above estimates are, however, theoretical estimates, and the actual situation may differ remarkably from the expectation. For example, X phage genome is 49 kb long; therefore, it is expected to contain about 12 (= 49,000/4,096) recognition sites for a restriction enzyme having recognition sequence of 6 bp. But the actual number of sites present for BamHl, Bglll and Sail is only 5, 6 and 2, respectively, which is a reflection of the less than 50% G: C content of λ DNA.

Restriction mapping:

Most pieces of DNA will have recognition sites for various restriction enzymes, and it is often beneficial to know the relative locations of some of these sites. The technique used to obtain this information is known as restriction mapping. This involves cutting a DNA fragment with a selection of restriction enzymes, singly and in various combinations. The fragments produced are run on an agarose gel and their sizes determined. From the data obtained, the relative locations of the cutting sites can be worked out. A fairly simple example can be used to illustrate the technique, as outlined in the following.

Table 4.2. Digestion of a 15 kbp DNA fragment with three restriction enzymes

BamHI	*EcoRI*	*PstI*	*BamHI* + *EcoRI*	*BamHI* + *PstI*	*EcoRI* + *PstI*	*BamHI* + *EcoRI* + *PstI*
14	12	8	11	8	7	6
1	3	7	3	6	5	5
			1	1	3	3
						1

Note: Data shown are lengths (in kbp) of fragments that are produced on digestion of a 15 kbp DNA fragment with the enzymes *BamHI*, *EcoRI*, and *PstI*. Single, double, and triple digests were carried out as indicated. Fragments produced by each digest are listed in order of length.

Let us say that we wish to map the cutting sites for the restriction enzymes BamHI, EcoRI, and PstI, and that the DNA fragment of interest is 15 kb in length. Various digestions are carried out, and the fragments arising from these are analysed and their sizes determined. The results obtained are shown in Table above As each of the single enzyme reactions produces two DNA fragments, we can conclude that the DNA has a single cutting site for each enzyme. The double digests enable a map to be drawn up, and the triple digest confirms this.

Applications:

In various applications related to genetic engineering DNA is cleaved by using these restriction enzymes.

• They are used in the process of insertion of genes into plasmid vectors during gene cloning and protein expression experiments.

• Restriction enzymes can also be used to distinguish gene alleles by specifically recognizing single base changes in DNA known as single nucleotide polymorphisms (SNPs). This is only possible if a mutation alters the restriction site present in the allele.

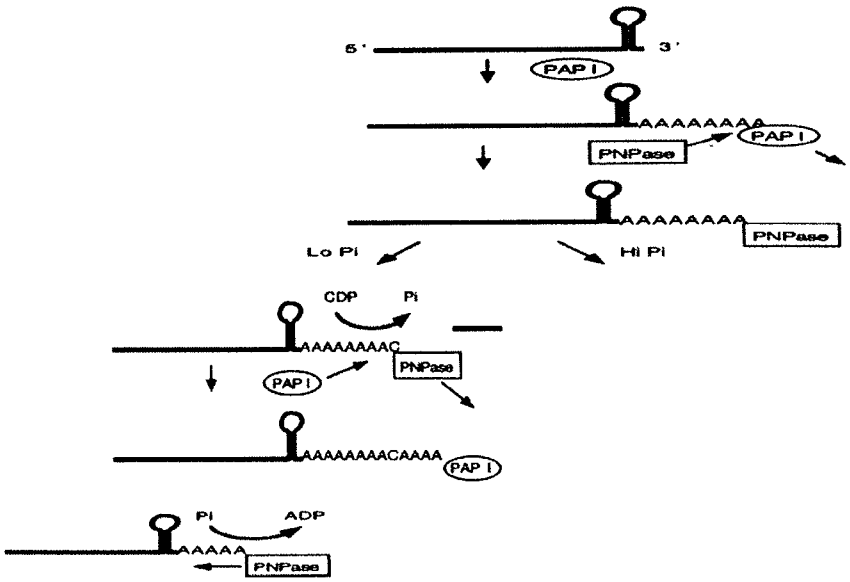
• Restriction enzymes are used for Restriction Fragment Length Polymorphism (RFLP) analysis for identifying individuals or strains of a particular species.

ENZYMES IN MODIFICATION

Polynucleotide phosphorylase: was first discovered from extracts of Azotobacter agile by Grunberg-Manago and Ochoa. catalyzes the synthesis of long chain polyribonucleotides (RNA) in 5' to 3' direction from nucleotide diphosphates as precursors and reversibly catalyzes phosphorolytic cleavage of polyribonucleotides in 3' to 5' direction with a release of orthophosphate in presence of inorganic phosphate.

PNPase is a bifunctional enzyme and functions in mRNA processing and degradation inside the cell.

These enzyme can catalyze not only the synthesis of RNA from the mixtures of naturally occurring ribonucleoside diphosphates, but also that of non-naturally occurring polyribonucleotides. In rDNA cloning technology, it has been used to synthesize radiolabelled polyribonucleotides from nucleoside diphosphate monomers.



In E.coli, polynucleotide phosphorylase regulates mRNA processing either by adding ribonucleotides to the 3' end or by cleaving bases in 3' to 5' direction. The function of PNPase depends upon inorganic phosphate (Pi) concentration inside the cell. The transcripts are polyadenylated using enzyme polyadenylate polymerase I (PAPI). After primary polyadenylylation of the transcript by PAP I, PNPase may bind

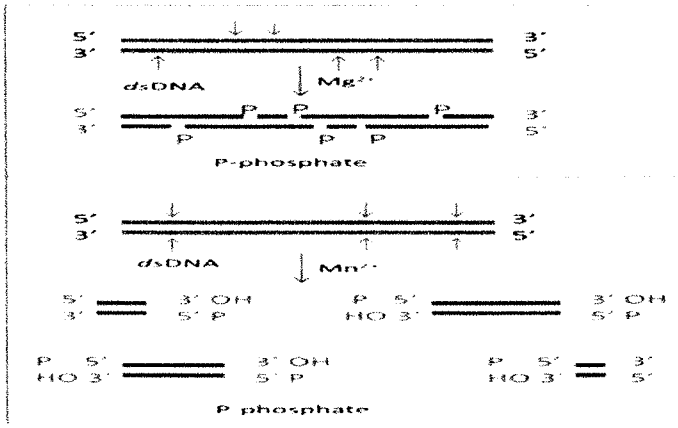
to the 3′ end of the poly(A) tail. PNPase works either degradatively or biosynthetically inside the cell depending on the Pi concentration. Under high Pi concentration, it degrades the poly(A) tail releasing adenine diphosphates. If the Pi concentration is low, PAP I initiates addition of one or more nucleotides to the existing poly (A) tail and in the process generates inorganic phosphate. On dissociation of PNPase, the 3′ end again is available to PAP I for further polymerization.

Deoxyribonuclease (DNase):

A nuclease enzyme that can catalyze the hydrolytic cleavage of phosphodiester bonds in the DNA backbone are known as deoxyribonuclease (DNase).

Based on the position of action, these enzymes are broadly classified as endodeoxyribonuclease (cleave DNA sequence internally) and exodeoxyribonuclease (cleave the terminal nucleotides).

Unlike restriction enzymes, DNase does not have any specific recognition/restriction site and cleave DNA sequence at random locations.



Phosphatase:

catalyses the cleavage of a phosphate (PO4 -2) group from substrate by using a water molecule (hydrolytic cleavage). This reaction is not reversible.

On the basis of their activity there are two types of phosphatase i.e acid phosphatase and alkaline phosphatase. In both forms the alkaline

phosphatase are most common. Special class of phosphatase that remove a phosphate group from protein, called "Phosphoprotein phosphatase"

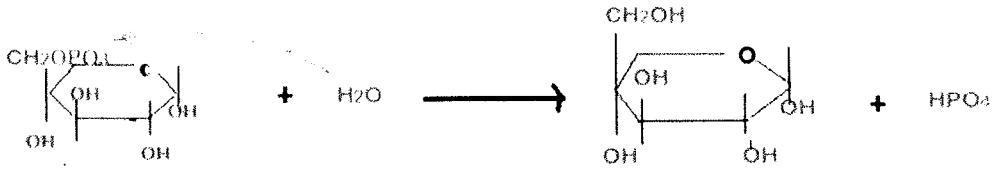


Fig 2-3.1: Schematic representation of hydrolytic cleavage of phosphate group (-PO_4^{-2}).

Methylase:

Methyltransferase or methylase catalyzes the transfer of methyl group (-CH3) to its substrate. The process of transfer of methyl group to its substrate is called methylation.

Methylation is a common phenomenon in DNA and protein structure.

Methyltransferase uses a reactive methyl group that is bound to sulfur in Sadenosyl methionine (SAM) which acts as the methyl donor.

Methylation normally occurs on cytosine (C) residue in DNA sequence. In protein, methylation occurs on nitrogen atom either on N-terminus or on the side chain of protein.

DNA methylation regulates gene or silence gene without changing DNA sequences, as a part of epigenetic regulation.

In bacterial system, methylation plays a major role in preventing their genome from degradation by restriction enzymes. It is a part of restriction – modification system in bacteria. Methyltransferase can be classified in three groups:

a) m6A-generates N6 methyladenosine,

b) m4C-generates N4 methylcytosine,

c) m5C-generatesN5 methylcytosine.

m6A and m4C methyltransferase are primarily found in prokaryotes. These enzymes are responsible for methylation of DNA sequences in order to prevent the host from digesting its own genome via its restriction enzyme.

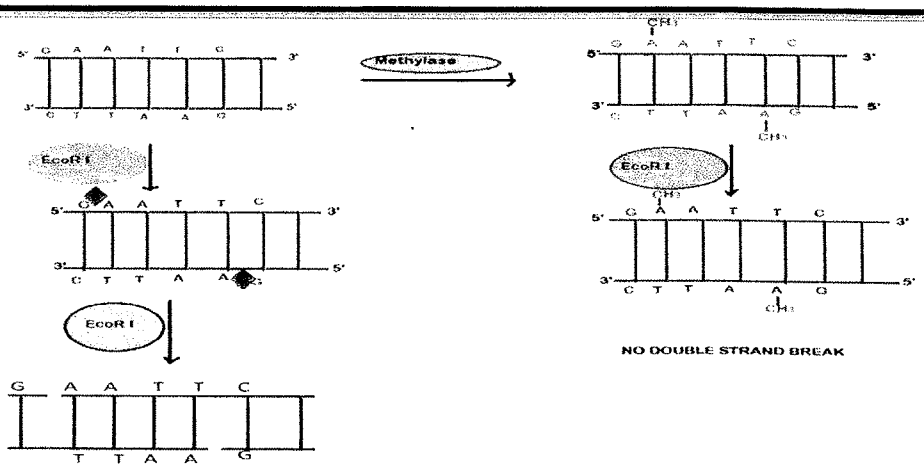


Fig 2-3.2: Activity of restriction and methylase enzymes

Ligases:

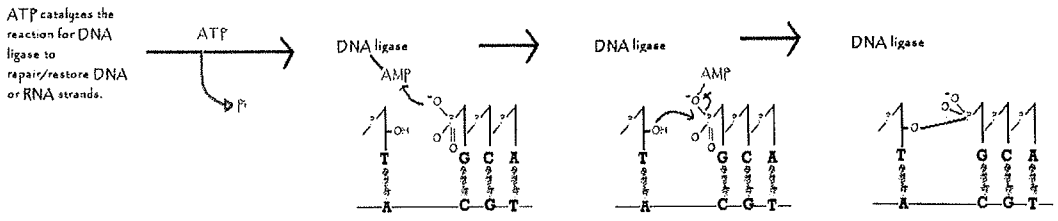
DNA ligase catalyses the formation of phosphodiester bond between two deoxynucleotide residues of two DNA strands.

DNA ligase enzyme requires a free hydroxyl group at the 3′ -end of one DNA chain and a phosphate group at the 5′-end of the other and requires energy in the process.

The role of DNA ligase is to seal nicks in the backbone of double-stranded DNA after lagging strand formation to join the okazaki fragments. This joining process is essential for the normal synthesis of DNA and for repairing damaged DNA. It has been exploited by genetic engineers to join DNA chains to form recombinant DNA molecules. Usually single stranded break are repaired using the complimentary strand as the template but sometimes double stranded breaks can also be repaired with the help of DNA ligase IV.

Application:

1. DNA ligase enzyme is used by cells to join the "okazaki fragments" during DNA replication process. In molecular cloning, ligase enzyme has been routinely used to construct a recombinant DNA. Followings are some of the examples of application of ligase enzyme in molecular cloning.Joining of adapters and linkers to blunt end DNA molecule.
2. Cloning of restricted DNA to vector to construct recombinant vector.



Polynucleotide Kinase:

PNK is a homotetramer with phosphatase activity at 3' end and kinase activity at 5' end with a tunnel like active site. The active site has side chains which interact with NTP donor's beta-phosphate and 3' phosphate of acceptor with an acid which activated 5' –OH. Lys-15 and Ser-16 are important for the kinase activity of the enzyme.

The basic residues of active site of PNK interact with the negatively charged phosphates of the DNA.

Polynucleotide kinase (PNK) catalyzes the transfer of a phosphate group (PO4 -2) from γ position of ATP to the 5' end of either DNA or RNA and nucleoside monophosphate.

PNK can convert 3' PO4/5' OH ends into 3' PO4/5' PO4 ends which blocks further ligation by ligase enzyme.

PNK is used to label the ends of DNA or RNA with radioactive phosphate group.

T4 polynucleotide kinase is the most widely used PNK in molecular cloning experiments, which was isolated from T4 bacteriophage infected E.coli.

There are two major uses of PNK: The linkers and adopters are phosphorylated along with the fragments of DNA before ligation, which requires a 5' phosphate. This includes products of polymerase chain reaction, which are generated by using non-phosphorylated primers. PNK is also used for radio labelling oligonucleotides, generally with ^{32}P for preparing hybridization probes.

Ribonuclease (RNase):

Nuclease that can catalyze hydrolysis of ribonucleotides from either single stranded or double stranded RNA sequence are called ribonucleotides (RNase).

RNase are classified into two types depending on position of cleavage, i.e. endoribonuclease (cleave internal bond) and exoribonuclease (cleave terminal bond).

RNase is important for RNA maturation and processing.

RNaseA and RNaseH play important role in initial defence mechanism against RNA viral infection

Application:

1. RNaseA is used to remove RNA contamination from DNA sample.
2. During cDNA library preparation from RNA sample, RNaseH enzyme is used to cleave RNA strand of DNA-RNA duplex

Cloning

DNA cloning is the starting point for many genetic engineering approaches to biotechnology research. Large amounts of DNA are needed for genetic engineering.

In biology, **cloning** is the process of producing similar populations of genetically identical individuals that occurs in nature when organisms such as bacteria, insects or plants reproduce asexually.

Cloning in biotechnology refers to processes used to create copies of DNA fragments (molecular cloning), cells (cell cloning), or organisms. The term also refers to the production of multiple copies of a product such as digital media or software.

How is DNA cloned in cells?

To get multiple copies of a gene or other piece of DNA you must isolate, or 'cut', the DNA from its source and then 'paste' it into a DNA vector that can replicate (or copy) itself.

The four main steps in DNA cloning are:

Step 1. The chosen piece of DNA is 'cut' from the source organism using restriction enzymes.

Once DNA has been purified, it must be cut into pieces before the chromosomal DNA and the plasmid DNA can be joined. The problem is to cut the DNA so that it will be easy to join the cut ends of the chromosomal DNA to the cut ends of the plasmid DNA. A group of enzymes, called restriction enzymes, are used for this purpose.

Step 2. The piece of DNA is 'pasted' into a vector and the ends of the DNA are joined with the vector DNA by ligation.

The double-stranded ends must be covalently attached. A version of this reaction is normally carried out in the cell by an enzyme known as DNA ligase. At the end of this process, there is a nick in the DNA that is sealed by DNA ligase. The double-stranded break formed by the restriction enzyme can be thought of as two nicks, each of which is a substrate for ligase.

Step 3. The vector is introduced into a host cell, often a bacterium or yeast, by a process called transformation. The host cells copy the vector DNA along with their own DNA, creating multiple copies of the inserted DNA.

Step 4. The vector DNA is isolated (or separated) from the host cells' DNA and purified.

DNA that has been 'cut' and 'pasted' from an organism into a vector is called recombinant DNA. Because of this, DNA cloning is also called recombinant DNA technology.

What is cloned DNA used for?

DNA cloning is used to create a large number of copies of a gene or other piece of DNA. The cloned DNA can be used to:

- Work out the function of the gene
- Investigate a gene's characteristics (size, expression, tissue distribution)
- Look at how mutations may affect a gene's function
- Make large concentrations of the protein coded for by the gene

Cell cloning

The process of producing a group of cells that are genetically identical (clones) to a single ancestral cell.

Reproductive Cloning:

Reproductive cloning is the term for type of cloning that most people associate human cloning in general with, and has been performed on many animals, including sheep. The purpose of reproductive cloning is to recreate an exact copy of an existing organism by copying one's genetic code. This is done by taking an egg from a female, removing the existing DNA, and replacing it with DNA from the person being cloned. The ovum with the new DNA is then implanted in the womb and allowed to develop. However, there is an extreme risk of serious genetic defects

associated with this process. This form of cloning is universally considered immoral, and has not yet successfully been done.

Research Cloning (Therapeutic cloning, Biomedical cloning): This type of human cloning involves the creation of a cloned embryo for the purpose study. The process of creating the cloned embryo is exactly the same as in Reproductive cloning. The difference is that in research cloning, the embryo is intentionally destroyed, rather that being implanted in the womb and allowed to develop into a infant. When the zygote (fertilized egg) splits repeatedly, eventually forming a blastocyst (100-200 cells), stem cells are removed and the embryo dies. The ultimate goal of research cloning is to "trick" the stem cells into forming desired kinds of cell, and eventually create entire organs.

Embryo Cloning:

Embryo Cloning is quite different from the other two types of human cloning. It is a technique in which an embryo is encouraged to develop into two or more identical embryos. Embryo cloning follows the same process as nature does, in producing identical (monozygotic) twins. It has been performed many times on animals, but hardly at all on humans.

Ligation – joining DNA molecules together

The final step in construction of a recombinant DNA molecule is the joining together of the vector molecule and the DNA to be cloned. This process is referred to as ligation, and the enzyme that catalyzes the reaction is called DNA ligase.

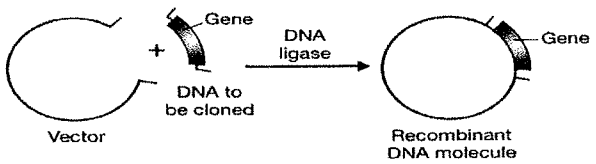


Figure 4.19
Ligation: the final step in construction of a recombinant DNA molecule.

Sticky ends increase the efficiency of ligation:

The ligation reaction of two blunt-ended fragments being joined together. Although this reaction can be carried out in the test tube, it is not very efficient. This is because the ligase is unable to "catch hold" of

the molecule to be ligated, and has to wait for chance associations to bring the ends together. If possible, blunt end ligation should be performed at high DNA concentrations, to increase the chances of the ends of the molecules coming together in the correct way. In contrast, ligation of complementary sticky ends is much more efficient. This is because compatible sticky ends can base pair with one another by hydrogen bonding , forming a relatively stable structure for the enzyme to work on. If the phosphodiester bonds are not synthesized fairly quickly then the sticky ends fall apart again. These transient, base-paired structures do, however, increase the efficiency of ligation by increasing the length of time the ends are in contact with one another.

Putting sticky ends onto a blunt-ended molecule:

For the reasons detailed in the preceding section, compatible sticky ends are desirable on the DNA molecules to be ligated together in a gene cloning experiment. Often these sticky ends can be provided by digesting both the vector and the DNA to be cloned with the same restriction endonuclease, or with different enzymes that produce the same sticky end, but it is not always possible to do this. A common situation is where the vector molecule has sticky ends, but the DNA fragments to be cloned are blunt-ended. Under these circumstances one of three methods can be used to put the correct sticky ends onto the DNA fragments.

Linkers:

The first of these methods involves the use of linkers. These are short pieces of doublestranded DNA, of known nucleotide sequence, that are synthesized in the test tube.

A typical linker is is blunt-ended, but contains a restriction site, BamHI in the example shown. DNA ligase can attach linkers to the ends of larger bluntended DNA molecules. Although a blunt end ligation, this particular reaction can be performed very efficiently because synthetic oligonucleotides, such as linkers, can be made in very large amounts and added into the ligation mixture at a high concentration. More than one linker will attach to each end of the DNA molecule, producing the chain structure shown in Figure. However, digestion with BamHI cleaves the chains at the recognition sequences, producing a large number of cleaved

linkers and the original DNA fragment, now carrying BamHI sticky ends. This modified fragment is ready for ligation into a cloning vector restricted with BamHI.

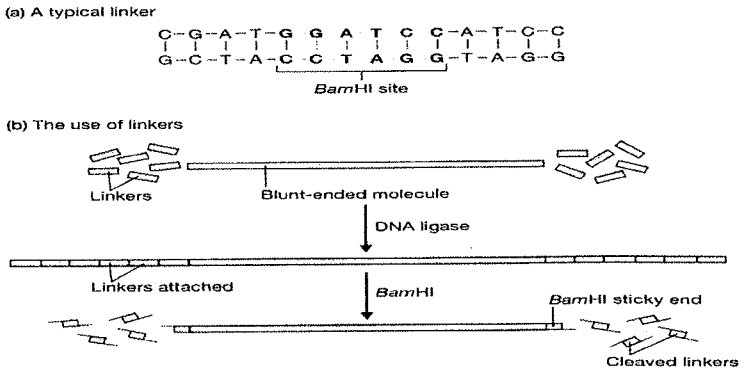


Figure 4.21
Linkers and their use: (a) the structure of a typical linker; (b) the attachment of linkers to a blunt-ended molecule.

Adaptors:

There is one potential drawback with the use of linkers. Consider what would happen if the blunt-ended molecule contained one or more BamHI recognition sequences. If this was the case, the restriction step needed to cleave the linkers and produce the sticky ends would also cleave the blunt-ended molecule . The resulting fragments will have the correct sticky ends, but that is no consolation if the gene contained in the blunt-ended fragment has now been broken into pieces.

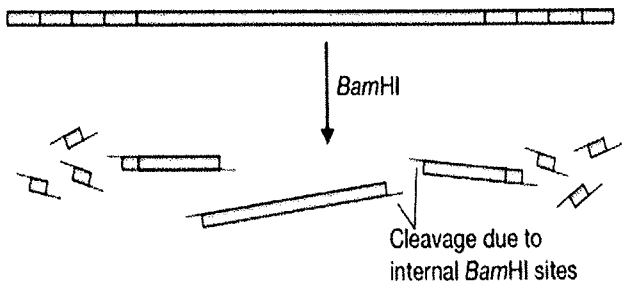
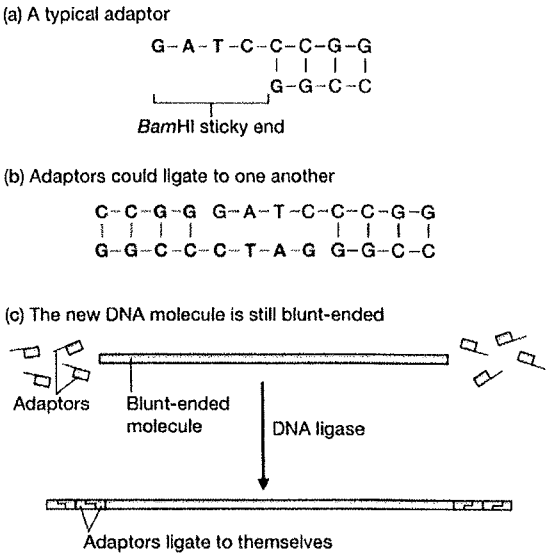


Figure 4.22
A possible problem with the use of linkers. Compare this situation with the desired result *BamHI* restriction, as shown in Figure 4.21(b)

Figure 4.23

Adaptors and the potential problem with their use. (a) A typical adaptor. (b) Two adaptors could ligate to one another to produce a molecule similar to a linker, so that (c) after ligation of adaptors a blunt-ended molecule is still blunt-ended and the restriction step is still needed.



The second method of attaching sticky ends to a blunt-ended molecule is designed to avoid this problem. Adaptors, like linkers, are short synthetic oligonucleotides. But unlike linkers, an adaptor is synthesized so that it already has one sticky end . The idea is of course to ligate the blunt end of the adaptor to the blunt ends of the DNA fragment, to produce a new molecule with sticky ends. This may appear to be a simple method but in practice a new problem arises. The sticky ends of individual adaptor molecules could base pair with each other to form dimers , so that the new DNA molecule is still blunt-ended . The sticky ends could be recreated by digestion with a restriction endonuclease, but that would defeat the purpose of using adaptors in the first place. The answer to the problem lies in the precise chemical structure of the ends of the adaptor molecule. Normally the two ends of a polynucleotide strand are chemically distinct, a fact that is clear from a careful examination of the polymeric structure of DNA . One end, referred to as the 5′ terminus, carries a phosphate group (5′-P); the other, the 3′ terminus, has a hydroxyl group (3′-OH). In the double helix the two strands are antiparallel (Figure 4.24b), so each end of a double-stranded molecule consists of one 5′-P terminus and one 3′-OH terminus. Ligation takes place between the 5′-P and 3′-OH ends . Adaptor molecules are synthesized so that the blunt end is the same as "natural" DNA, but the sticky end is different. The 3′-OH terminus of the sticky end is the same as usual, but the 5′-P terminus is

modified: it lacks the phosphate group, and is in fact a 5′-OH terminus . DNA ligase is unable to form a phosphodiester bridge between 5′-OH and 3′-OH ends. The result is that, although base pairing is always occurring between the sticky ends of adaptor molecules, the association is never stabilized by ligation . Adaptors can therefore be ligated to a blunt-ended DNA molecule but not to themselves. After the adaptors have been attached, the abnormal 5′-OH terminus is converted to the natural 5′-P form by treatment with the enzyme polynucleotide kinase , producing a sticky-ended fragment that can be inserted into an appropriate vector.

Producing sticky ends by homopolymer tailing

The technique of homopolymer tailing offers a radically different approach to the production of sticky ends on a blunt-ended DNA molecule. A homopolymer is simply a polymer in which all the subunits are the same. A DNA strand made up entirely of, say, deoxyguanosine is an example of a homopolymer, and is referred to as polydeoxyguanosine or poly(dG). Tailing involves using the enzyme terminal deoxynucleotidyl transferase to add a series of nucleotides onto the 3′-OH termini of a double-stranded DNA molecule. If this reaction is carried out in the presence of just one deoxyribonucleotide, a homopolymer tail is produced.

Figure 4.23
The use of adaptors: (a) the actual structure of an adaptor, showing the modified 5′-OH terminus; (b) conversion of blunt ends to sticky ends through the attachment of adaptors.

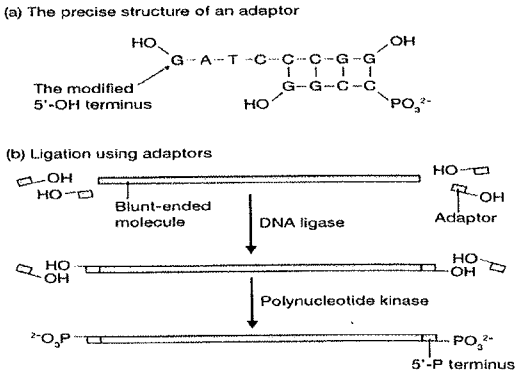
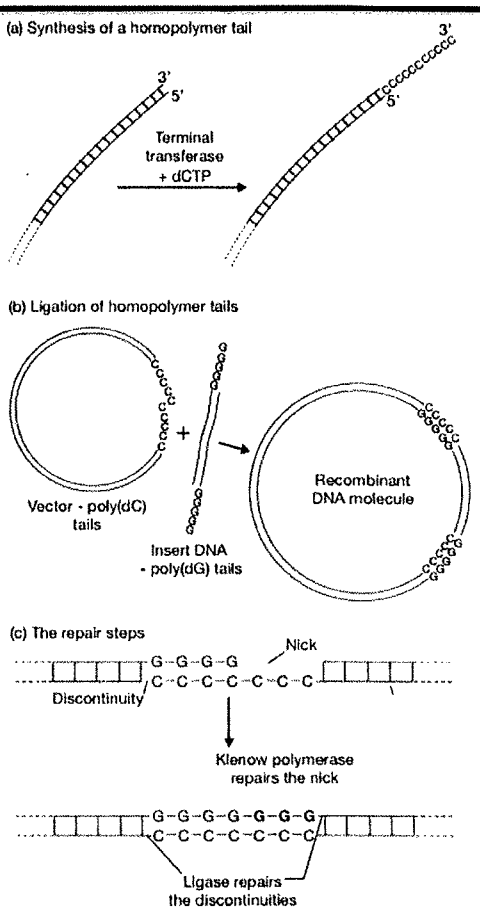


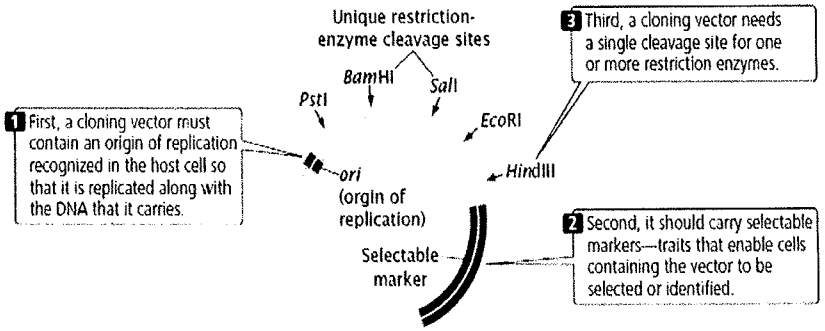
Figure 4.26
Homopolymer tailing: (a) synthesis of a homopolymer tail; (b) construction of a recombinant DNA molecule from a tailed vector plus tailed insert DNA; (c) repair of the recombinant DNA molecule.



two tailed molecules, the homopolymers must be complementary. Frequently polydeoxycytosine (poly(dC)) tails are attached to the vector and poly(dG) to the DNA to be cloned. Base pairing between the two occurs when the DNA molecules are mixed . In practice, the poly(dG) and poly(dC) tails are not usually exactly the same length, and the base-paired recombinant molecules that result have nicks as well as discontinuities . Repair is therefore a two-step process, using Klenow polymerase to fill in the nicks followed by DNA ligase to synthesize the final phosphodiester bonds. This repair reaction does not always have to be performed in the test tube. If the complementary homopolymer tails are longer than about 20 nucleotides, then quite stable base-paired associations are formed. A recombinant DNA molecule, held together by base pairing although not completely ligated, is often stable enough to be introduced into the host cell in the next stage of the cloning experiment . Once inside the host, the cell's own DNA polymerase and DNA ligase repair the recombinant DNA molecule, completing the construction begun in the test tube.

Cloning vectors

A **cloning vector** is a stable, replicating DNA molecule to which a foreign DNA fragment can be attached for introduction into a cell. An effective cloning vector has three important characteristics:
(1) an origin of replication, which ensures that the vector is replicated within the cell;
(2) selectable markers, which enable any cells containing the vector to be selected or identified;
(3) one or more unique restriction sites into which a DNA fragment can be inserted.
The restriction sites used for cloning must be unique; if a vector is cut at multiple recognition sites, generating several pieces of DNA, there will be no way to get the pieces back together in the correct order.
Three types of cloning vectors are commonly used for cloning genes in bacteria: plasmids, bacteriophages, and cosmids.



18.6 Three characteristics of an idealized cloning vector. An origin of replication, one or more selectable markers, and one or more unique restriction sites.

1: Plasmid vectors

Plasmids are circular DNA molecules that exist naturally in bacteria. They contain origins of replication and are therefore able to replicate independently of the bacterial chromosome. Plasmids are the most used cloning tools because they are easy to purify, to manipulate *in vitro* and to introduce into cells in culture, whether they multiply apart from the host genome or they integrate totally or partially (the sequence of interest carried by the plasmid) into the genome.
One of the first vectors to be developed was pBR322. pBR322 lacks the more sophisticated features of the newest cloning vectors, and so is no

longer used extensively in research, it still illustrates the important, fundamental properties of any plasmid cloning vector.

The nomenclature of plasmid cloning vectors

The name "pBR322" conforms with the standard rules for vector nomenclature: 1 "p" indicates that this is indeed a plasmid. 1 "BR" identifies the laboratory in which the vector was originally constructed (BR stands for Bolivar and Rodriguez, the two researchers who developed pBR322). 1 "322" distinguishes this plasmid from others developed in the same laboratory (there are also plasmids called pBR325, pBR327, pBR328, etc.)

The useful properties of pBR322

The first useful feature of pBR322 is its size. it was stated that a cloning vector ought to be less than 10 kb in size, to avoid problems such as DNA breakdown during purification. pBR322 is 4363 bp, which means that not only can the vector itself be purified with ease, but so can recombinant DNA molecules constructed with it.

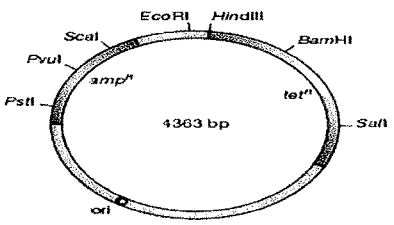


Figure 6.1
A map of pBR322 showing the positions of the ampicillin resistance (*amp*^{R}) and tetracycline resistance (*tet*^{R}) genes, the origin of replication (ori) and some of the most important restriction sites.

The second feature of pBR322 it carries two sets of antibiotic resistance genes. Either ampicillin or tetracycline resistance can be used as a selectable marker for cells containing the plasmid, and each marker gene includes unique restriction sites that can be used in cloning experiments.

A third advantage of pBR322 is that it has a reasonably high copy number. Generally there are about 15 molecules present in a transformed E. coli cell, but this number can be increased, up to 1000–3000, by plasmid amplification in the presence of a protein synthesis inhibitor such as chloramphenicol

pUC8—a Lac selection plasmid

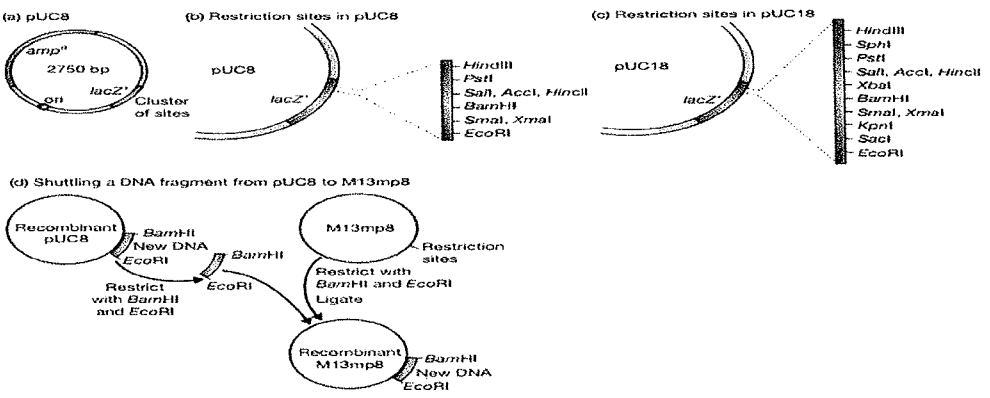
is descended from pBR322, although only the replication origin and the amp^R gene remain. The nucleotide sequence of the amp^R gene has been changed so that it no longer contains the unique restriction sites: all these cloning sites are now clustered into a short segment of the lacZ′ gene carried by pUC8.

pUC8 has three important advantages that have led to it becoming one of the most popular E. coli cloning vectors.

The first the manipulations involved in construction of pUC8 were accompanied by a chance mutation, within the origin of replication, which results in the plasmid having a copy number of 500–700 even before amplification. This has a significant effect on the yield of cloned DNA obtainable from E. coli cells transformed with recombinant pUC8 plasmids.

The second advantage is that identification of recombinant cells can be achieved by a single step process, by plating onto agar medium containing ampicillin plus X-gal . (With both pBR322 and pBR327, selection of recombinants is a two-step procedure, requiring replica plating from one antibiotic medium to another) .

The third advantage of pUC8 lies with the clustering of the restriction sites, which allows a DNA fragment with two different sticky ends (say EcoRI at one end and BamHI at the other) to be cloned without resorting to additional manipulations such as linker attachment.



The pUC plasmids. (a) The structure of pUC8. (b) The restriction site cluster in the *lacZ′* gene of pUC8. (c) The restriction site cluster in pUC18. (d) Shuttling a DNA fragment from pUC8 to M13mp8.

2: Bacteriophage vector

The bacteriophages used for cloning are the phage λ and M13 phage. There is an upper limit on the amount of DNA that can be packed into a phage (a maximum of 53 kb), therefore to allow foreign DNA to be inserted into phage DNA, phage cloning vectors need to have some non-essential genes deleted, for example the genes for lysogeny in phage λ.[15] There are two kinds of λ phage vectors - insertion vector and replacement vector. Insertion vectors contain a unique cleavage site whereby foreign DNA with size of 5–11 kb may be inserted. In replacement vectors, the cleavage sites flank a region containing genes not essential for the lytic cycle, and this region may be deleted and replaced by the DNA insert in the cloning process, and a larger sized DNA of 8–24 kb may be inserted.

There is also a lower size limit for DNA that can be packed into a phage, and vector DNA that is too small cannot be properly packaged into the phage. This property can be used for selection - vector without insert may be too small, therefore only vectors with insert may be selected for propagation

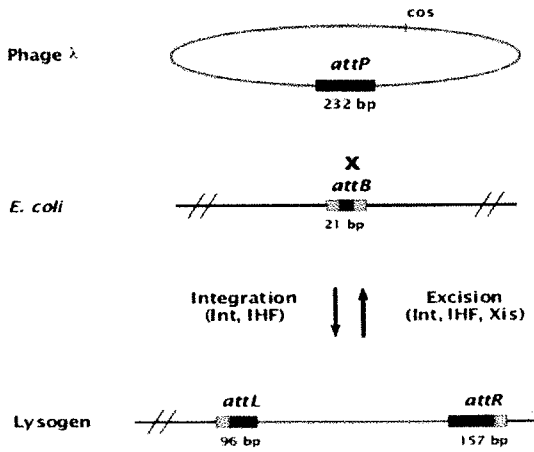
Phage lambda is a bacteriophage or phage, i.e. bacterial virus, that uses E. coli as host.

• Its structure is that of a typical phage: head, tail, tail fibres.

• Lambda viral genome: 48.5 kb linear DNA with a 12 base ssDNA "sticky end" at both ends; these ends are complementary in sequence and can hybridize to each other (this is the cos site: cohesive ends).

• Infection: lambda tail fibres adsorb to a cell surface receptor, the tail contracts, and the DNA is injected.

• The DNA circularizes at the cos site, and lambda begins its life cycle in the E. coli host.

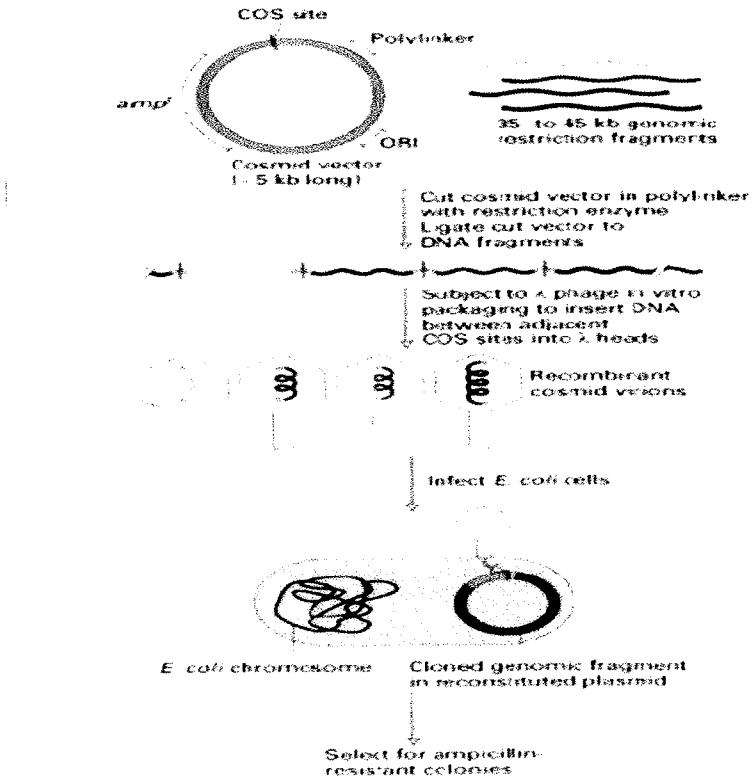


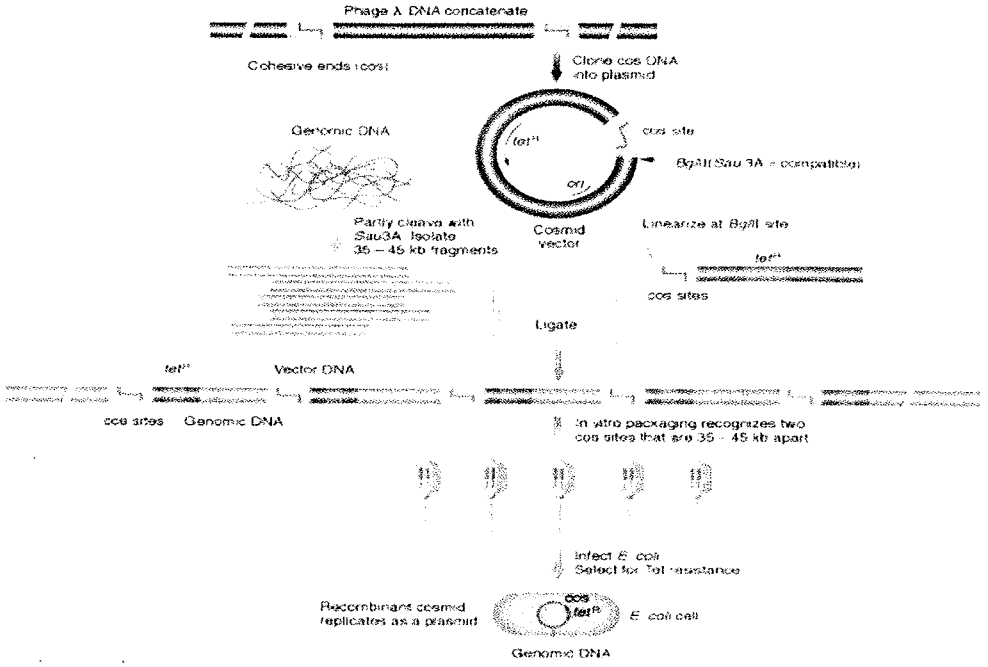
1. The Gateway® System relies on five sets of specific and non cross-reacting *att* sequences.
2. The specificity is given by the 7 nucleotides of the core region
3. The reaction is specific and directional:

   *att*B x *att*P ⇔*att*L x *att*R.
4. Each reaction is mediated by a different combination of enzymes.

3: Cosmid vector

is a combination of the plasmid vector and the COS site which allows the target DNA to be inserted into the λ head. It has the following advantages: – High transformation efficiency. – The cosmid vector can carry up to 45 kb whereas plasmid and λ phage vectors are limited to 25 kb.




Phage λ DNA concatemer
Cohesive ends (cos)
Clone cos DNA into plasmid
Genomic DNA
cos site
BglII (Sau 3A = compatible)
ori
Partly cleave with Sau3A Isolate 35 – 45 kb fragments
Cosmid vector
Linearize at BglII site
cos sites
Ligate
Vector DNA
cos sites
Genomic DNA
In vitro packaging recognizes two cos sites that are 35 – 45 kb apart
Infect E. coli
Select for Tet resistance
Recombinant cosmid replicates as a plasmid
E. coli cell
Genomic DNA

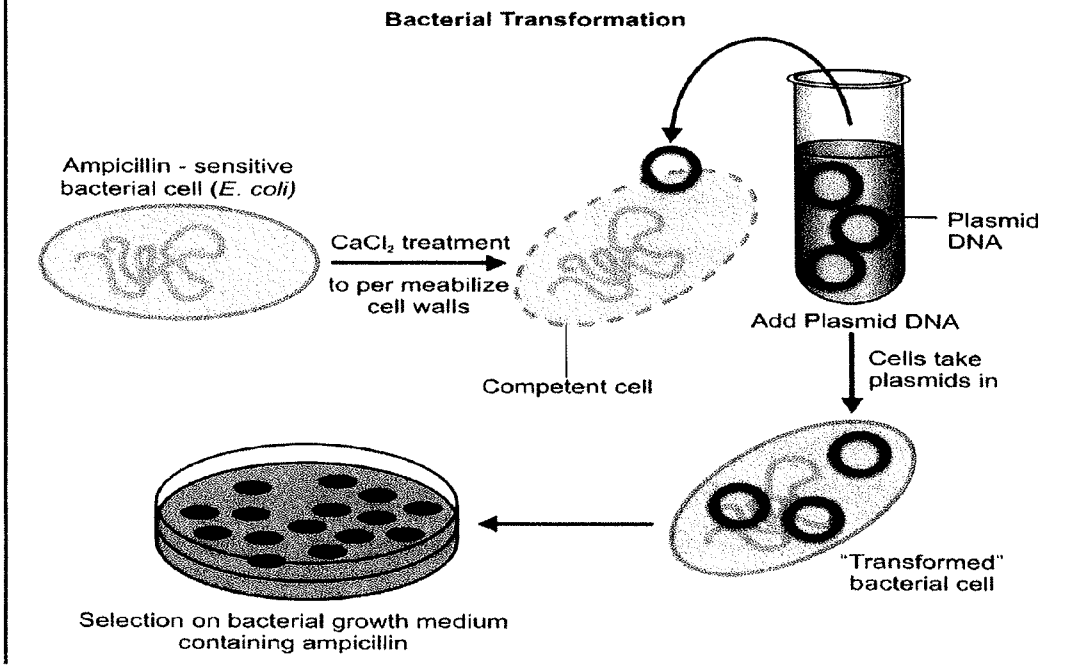
INTRODUCTION OF DNA INTO HOST CELLS

The next step in a gene cloning experiment is to introduce the recombinant molecule into living cells (usually bacteria), which will then grow and divide to produce clones.

Transformation: The uptake of DNA by bacterial cells

Transformation is the process of uptake of foreign DNA (normally plasmids) by bacteria. In nature, transformation is probably not a major process by which bacteria obtain genetic material. This is reflected by the fact that only a few species (notably members of the genera *Bacillus* and *Streptococcus*) can be transformed with ease in the laboratory, and studies have revealed that these species possess sophisticated mechanisms for DNA binding and uptake. Under normal circumstances, most bacterial species (including *E. coli*) will only take up limited amounts of DNA. In order to transform these species efficiently, the bacteria have to undergo some form of physical and/or chemical treatment that will enhance their ability to take up DNA. Cells that have undergone treatment are referred to as competent cells .

It was discovered in the early 1970s (cited by Old & Primrose, 1994) that *E. coli* cells treated with solutions containing Ca2+ ions were rendered susceptible to take up exogenous DNA. The precise mechanism of this is not understood. Ca2+ ions possibly cause the DNA to precipitate on the outside of the cells, or perhaps they are responsible for some kind of change in the cell wall that improves DNA binding. Thus, treatment with Ca2+ ions affects only DNA binding, and not the actual uptake of DNA . Furthermore, when DNA is added to treated cells, it remains attached to the exterior, and is not at this stage transported into the cytoplasm. The actual movement of the DNA into competent cells is stimulated by briefly rising the temperature to 42°C (referred to as heat shock) . This process induces enzymes involved in the repair of DNA and other cellular components, which allow the cell to recover from the unusual conditions of the transformation process, and increases the efficiency. After this, cells are incubated in a growth medium and finally spread out on an agar plate and incubated until single colonies of bacteria grow.



The quality of a given preparation of competent cells may be measured by determining the transformation efficiency. This can be defined as the number of colonies formed (on a selective plate) per microgram of input DNA, where that DNA is a pure plasmid, most commonly the vector to be used in a cloning experiment. Transformation efficiencies can range from 10^3 per μg for crude transformation protocols, to more than 10^8 per μg for very carefully prepared competent cells to be used in the construction of libraries. A transformation efficiency of 10^5 per μg would be adequate for a simple cloning experiment.

Introduction of phage DNA into bacterial cells:

A recombinant DNA molecule, constructed with a phage vector, can be introduced into a bacterial cell by transfection or *in vitro* packaging.

Transfection

The process of transfection is equivalent to transformation, the only difference being that phage DNA rather than a plasmid is involved. As with transformation, the purified recombinant phage molecule is added to competent *E. coli* cells and DNA uptake is induced by heat shock .

In vitro packaging

Another method of introducing a recombinant DNA molecule, constructed with a phage vector, is *in vitro* packaging. To introduce such recombinants into the host cell, they are first "packaged" into phage particles *in vitro*, and then the bacteria are infected (transfected) with the packaged recombinant.

Transformation of non-bacterial cells:

Ways of introducing DNA into yeast, fungi, animals and plants are also needed if these organisms are to be used as hosts for gene cloning . The uptake of DNA into eukaryotic cells is more challenging than bacterial transformation, and the efficiency of the process is much lower. For example, in yeast and plant cells the cell wall must be digested with degradative enzymes to yield fragile protoplasts, which may then take up DNA quite readily. The cell walls are re-synthesized once the degrading enzymes are removed. In contrast, animal cells in culture, which have no cell wall, will take up DNA at low efficiency if it is precipitated on their surface with calcium phosphate. Treating the cells with a high voltage, which is believed to open transient pores in the cell membrane, may increase the efficiency of the process. This process is referred to as electroporation and can also be used to introduce cloned genes into bacterial cells. In addition, direct physical methods have been used. DNA may be microinjected directly into the cytoplasm or even the nucleus of individual animal or plant cells in culture. Alternatively, DNA may be introduced by the bombardment of the target cells with metallic microprojectiles coated with DNA. This technique is referred to as biolistics .

GENE LIBRARIES

A gene library is a collection of different DNA sequences from an organism each of which has been cloned into a vector for ease of purification, storage and analysis. Gene libraries are used for two main purposes: (i) the cloning of individual genes (e.g. genes associated with certain genetic disorders), and (ii) the construction of physical maps of genomes . There are essentially two types of gene libraries that can be made depending on the source of DNA used. If the DNA is genomic

DNA, the library is called a **genomic library**. If the DNA is a copy of an mRNA population, that is cDNA, then the library is called a **cDNA library**.

Basic steps in the construction of a gene library:

1. Extraction of genomic DNA of the organism of interest.

2. The extracted DNA is cut into gene-size pieces with restriction enzymes.

3. The appropriate vector (for this purpose a plasmid) is cut with the same restriction enzyme.

4. The gene-sized DNA and cut plasmids are combined into one test tube. Some of the enzyme cut DNA pieces will combine with the cut vectors and form recombinant DNA.

5. The recombinant plasmids are then transferred into bacteria using either electroporation or heat shock (transformation).

6. The bacteria are grown on a culture dish and allowed to grow into colonies. All the colonies on all the plates (cultures) are called a gene library.

7. The gene library is then screened in order to discover which bacterial colony is making copies of the one gene you are interested in. Library screening identifies colonies which have that particular gene. Screening can be based on detecting the DNA sequence of the cloned gene, detecting a protein that the gene encodes, or the use of linked DNA markers.

8. When the bacterial colony containing the desired gene is located, the bacteria can be propagated to make millions of copies of the recombinant plasmid that contains the gene. Furthermore, the plasmids can be extracted for the next steps of genetic engineering, gene modification, and transformation.

SCREENING AND SELECTION FOR RECOMBINANTS

The art of cloning is to find the one particular transformed cell that contains the cloning vector with the gene of interest (referred to as a

recombinant cell). It is thus most important to be able to select recombinant cells from transformed cells (containing the vector without insert DNA) . If this goal has been achieved the gene in question is said to have been cloned.

Although there are many different procedures by which the desired clone can be obtained, all are variations of two basic concepts :

• **Direct selection**, which means that the cloning experiment is designed in such a way that the clones obtained are the clones containing the required gene. Almost invariably, selection occurs at the plating-out stage. Direct selection is the method of choice, as it is quick and usually unambiguous. However, it is not applicable to all genes.

• **Identification of the clone from a gene library**, which entails an initial "shotgun" cloning experiment, to produce a clone library representing all or most of the genes present in the cell, followed by analysis of the individual clones to identify the correct one.

Direct selection:

Most cloning vectors are designed so that the insertion of a DNA fragment into the vector destroys the integrity of one of the genes present on the molecule. This is referred to as insertional inactivation . Two examples will be discussed: (i) direct antibiotic resistance screening, and (ii) blue-white colour screening.

Direct antibiotic resistance screening

Most cloning vectors are designed so that the insertion of a DNA fragment into the vector destroys the integrity of one of the genes present on the molecule (usually an antibiotic resistance gene). For instance, if the vector carries an ampicillin resistance gene (*ampr*), it will confer ampicillin resistance to the *E. coli* cells if the transformation process was successful. This would mean that when the clones are plated out on ampicillin-containing medium, only cells containing a plasmid would be resistant to the antibiotic and, therefore, be able to grow. However, this does not show whether the cell contains the recombinant plasmid (with the inserted gene of

interest) or only the re-ligated vector (without the inserted gene of interest).

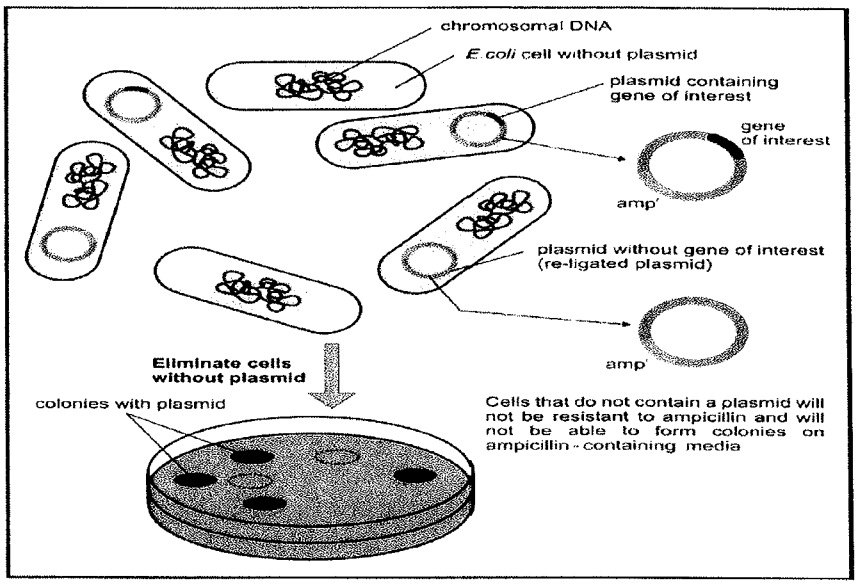


Figure 7: Screening for recombinants using direct antibiotic resistance screening.

Blue-white colour screening

This method also involves the insertional inactivation of a gene and uses the production of a blue compound as an indicator . The gene is *lacZ*, which encodes the enzyme ß-galactosidase, and is under the control of the *lac* promoter. If the host *E. coli* strain is expressing the lac repressor, expression of a *lacZ* gene on the vector may be induced using IPTG (isopropyl- ß-D-thiogalactopyranoside), and the expressed enzyme can utilize the synthetic substrate X-gal (5-bromo-4-chlore-3-indolyl- ß-D-galactopyranoside) to yield a blue product. Insertion of a DNA fragment into the *lacZ* gene (insertional inactivation of *lacZ*) in the production of a recombinant plasmid would prevent the development of the blue colour. In this method, the transformed cells are spread onto a plate containing an antibiotic (to select for transformants in the usual way), IPTG and X-gal, to yield a mixture of blue and white colonies. The white colonies have no expressed ß-galactosidase and are hence likely to contain the inserted target fragment. The blue colonies probably contain religated vector.

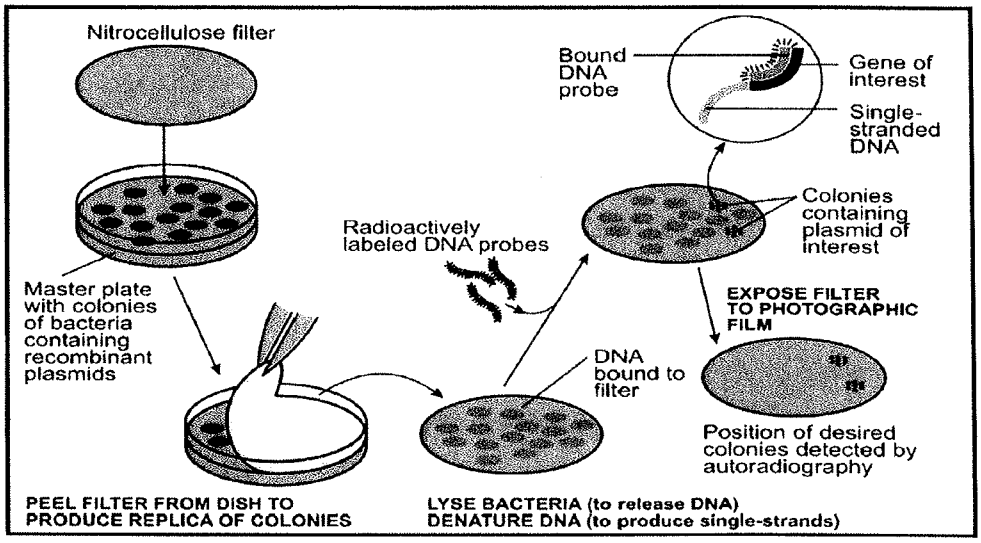
Identification of the clone from a gene library:

Many different techniques are available for screening a library. The most important approaches involve screening by nucleic acid hybridization and screening by functional analysis. Nucleic acid hybridization requires some prior knowledge of the DNA sequence either of the gene to be cloned or of stretches of DNA in the vicinity of the gene to be cloned. Functional screening involves the use of expression vectors that allow cells containing the vector with the desired gene to express the corresponding protein. Under these circumstances, cells containing a vector with the desired gene can be identified by means of antibodies directed against the protein.

Nucleic acid hybridization

Detection of an individual clone in a library can be achieved by employing strategies of nucleic acid hybridization in which short chemically synthesized labeled oligonucleotides (probes) are used to detect complementary sequences in individual cells or phages containing an insert. The success of colony or plaque hybridization will depend on the availability of a DNA molecule that can be used as probe. This probe must share at least a part of the sequence of the cloned gene. If the gene itself is not available, it can be derived, for example, from known protein sequences, or so-called degenerate oligonucleotides (mixtures of oligonucleotides that differ from each other by base substitutions at identical and/or different positions) can be used.

The first step of a hybridization screening experiment involves the transfer of the DNA in the plaque or colony to a nylon or nitrocellulose membrane. The bacteria on the membrane are lysed to release their DNA, and the DNA is denatured with alkali to produce single strands that are bounded to the membrane by heat treatment or UV irradiation. The membrane is then immersed in a solution containing a nucleic acid probe (usually radioactively labeled) and incubated to allow the probe to hybridize to its complementary sequence. After hybridization, the membrane is washed to remove unhybridized probe, and regions where the probe has hybridized are visualized with autoradiography .



Functional screening

The desired gene can also be identified by the activities of the encoded gene product. In this case, one uses a so-called expression library that has been established by cloning DNA (cDNA in the case of eukaryotes) fragments into special cloning vectors allowing the functional expression of cloned DNA fragments. Functional gene products and hence the desired clones can then be detected either by antibodies (immunological screening) or other ligands that specifically recognize the encoded proteins or by exploiting a bioactivity of the gene product, if known .

The procedure of immunological screening has similarities to hybridization screening, discussed before, though in this instance it is the protein encoded by the DNA (or cDNA) rather than the DNA itself that is detected on the membrane. The membrane is treated to covalently attach the protein, and immersed in a solution containing the antibodies. When the antibody has bound to its epitope, it is detected by other antibodies and/or chemicals that recognize it.

Chromosome walking

Often, a genomic clone may not include all of the sequence for a particular gene so it is necessary to isolate overlapping clones that cover the genomic region of interest. This process is known as chromosome walking.

DNA fragment sequencing

In order to establish exactly what changes have occurred in the gene under study, it will be necessary determine the DNA sequence of that gene. This is now a routine procedure in molecular biology laboratories, extending not just to individual genes or fragments but to the complete sequence of the DNA (the genome) of the organism.

The two main *in vitro* DNA sequencing methods were published in 1977, a few months apart, by Maxam and Gilbert and by Sanger *et al.* and will revolutionize in the same way as for PCR molecular biology and its applications in fundamental research as well as in diagnosis. Because it is simple and adaptable to the technological evolutions, the Sanger method is now universally used.

Principle of the Sanger method: the sequencing reaction

The Sanger sequencing method is applicable to a target fragment purified or specifically accessible. It is based on the possibility of performing *in vitro* replication of one of the two strands of the target, which supposes that a primer capable of hybridizingto the other strand thatwill be used as a template can be designed and added together with the trinucleotides and the DNA pol.

If the target fragment was obtained by PCR, its extremities are known and the problem of the sequencing primers is solved. If the target fragment is unknown, it is necessary to clone it into a vector whose sequences at the border of the cloning site are known and can be used to design primers.

The principle of the Sanger method consists of starting the synthesis of one of the two strands of the target fragment *in vitro* in the presence of small quantities of dideoxynucleotides (deoxy in 2_ but also in 3_) that will stopthe elongation in the event that they are incorporated because in the absence of a 3_ OH, it is impossible to add another nucleotide . The analysis of the size of the different randomly interrupted synthesized fragments with respect to dideoxynucleotide incorporation will allow the deduction of the sequence of the studied strand (the one synthesized starting from the primer). It is possible, and even recommended, to realize

the sequence of the other strand in four other tubes with a primer b/– that can hybridize with the (+) strand that will become the template.

Reading of the sequencing reaction products

After the sequencing reaction, the goal is to measure, with single-base precision, the length of the newly synthesized strands, which can only be done by separating the fragments by high-resolution gel electrophoresis (precise separation of strands differing by one single nucleotide) and considering that the newly synthesized fragments can be distinguished from the other DNA fragments that are not relevant (primers, fragment of interest, etc.). Therefore, it is necessary for the newly synthesized strands to be labelled in a way so that they can be specifically identified.

There are two ways of labelling synthesized strands: using a labelled primer (which is done with universal primers corresponding to the sequences in the cloning vector), or using labelled dideoxynucleotides (see below for the labelling techniques).

Even if radioactive labelling is now obsolete, the associated reading technique will be described in order to understand better the improvements made by fluorescent labelling and the use of automated sequencers.

Reading of a sequence with the labelled fragment emitting only one type of signal (radioactivity or cold labelling)

Using high-resolution polyacrylamide gel electrophoresis capable of separating fragments differing only by one nucleotide, it is possible to load on the gel the products of the four sequencing reactions and migrate them in four parallel lanes.

The DNA fragments will migrate according to their size and the revelation of the signal of the labelled molecules, meaning the newly synthesized strands whose size is the direct consequence of the sequence of the (+) strand, will allow the direct deduction of this sequence . The labelling signal never allows direct reading of the gel; in the case of a radioactive labelling (low intensity) the gel should be applied on a negative that will be revealed a few days later; in the case of a cold labelling, the gel should be revealed and exposed with an autoradiogram.

5′ Strand (+) TCGATTGGCAGCATGGACGC 3′

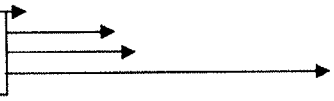
3′ Strand (–) AGCTAACCGTCGTACCTGCG 5′

Amplification primer a (+) /30 b

TUBE 1: in the presence of ddT

5′ + strand potential TCGATTGGCAGCATGGACGC 3′

Primer a (+) hybridized to the (–) strand
amplified: 30 nucleotides +?

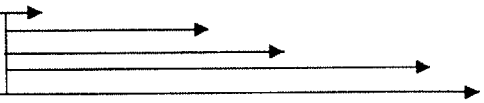


Result: size of the neo-strands depending on the possible incorporation of a ddT:
30 + 1 or 5 or 6 or 14

TUBE 2: in the presence of ddC

5′ + strand potential TCGATTGGCAGCATGGACGC 3′

Primer a (+) hybridized to the (–) strand
amplified: 30 nucleotides +?

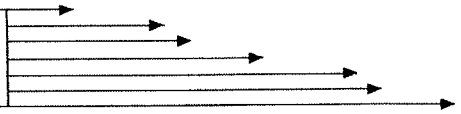


Result: size of the neo-strands depending on the possible incorporation of a ddC:
30 + 2 or 9 or 12 or 18 or 20

TUBE 3: in the presence of ddG

5′ + strand potential TCGATTGGCAGCATGGACGC 3′

Primer a (+) hybridized to the (–) strand
amplified: 30 nucleotides +?

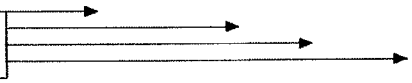


Result: size of the neo-strands depending on the possible incorporation of a ddG:
30 + 3 or 7 or 8 or 11 or 15 or 16 or 19

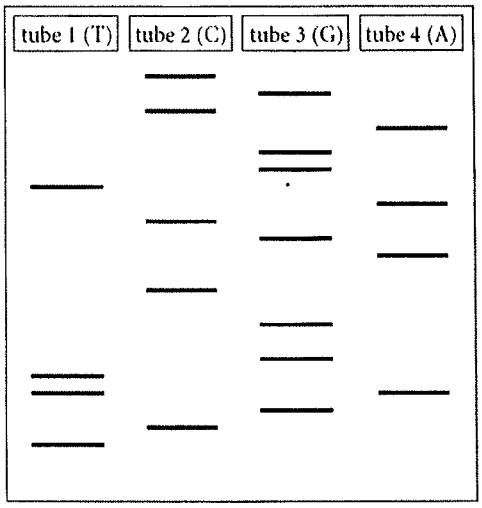
TUBE 4: in the presence of ddA

5′ + strand potential TCGATTGGCAGCATGGACGC 3′

Primer a (+) hybridized to the (–) strand
amplified: 30 nucleotides +?



Result: size of the neo-strands depending on the possible incorporation of a ddA:
30 + 4 or 10 or 13 or 17



Reading of the gel autoradiogram, with the sequence of the (+) strand after electrophoresis of the products of the Sanger reaction, labelled with the same type of signal.

- The smallest fragment corresponds to the non-elongated primer (dotted line) that is only visible if labelled primers are used instead of labelled ddNTPs.
- A fragment one base longer appears in the lane corresponding to the first tube and indicates that the first base following the primer is a T.
- Then, moving up on the gel, it appears that the second base after the primer is a C (tube 2), then a G (tube 3), etc., giving TCGATTGGCAGCATGGACGC.

Figure 1.13 Reading of the autoradiogram

Using the Polymerase Chain Reaction to Amplify DNA

A major problem in working at the molecular level is that each gene is a tiny fraction of the total cellular DNA. Because each gene is rare, it must be isolated and amplified before it can be studied. Before mid-1980, the only procedure available for amplifying DNA was gene cloning—placing the gene in a bacterial cell and multiplying the bacteria. Cloning is labor intensive and requires at least several days to grow the bacteria. In 1983, **Kary Mullis** discovered a new technique for amplifying DNA in a test tube. The **polymerase chain reaction** allows DNA fragments to be amplified a billion fold within just a few hours. It can be used with extremely small amounts of original DNA, even a single molecule. The polymerase chain reaction has revolutionized molecular biology and is now one of the most widely used of all molecular techniques.

The polymerase chain reaction in outline

PCR is based on ability of DNA polymerase to synthesize complementary strand to the template strand. As DNA polymerase can add a nucleotide only onto a 3'-OH group, it needs an artificial DNA strand (called DNA primer) of about 18 to 25 nucleotides complementary to 3' end of the DNA template. As shown below, each polynucleotide has a free 3' –OH group and 5' phosphate group. Moreover, a DNA strand has complimentary sequence, already paired by hydrogen bonding. Thus, primer can bind only when DNA strands are separated. This is generally done by heating. The primers anneal to the single-stranded DNA template at specific temperature (depends on primer sequence) and then DNA-Polymerase binds to this double stranded DNA produced. The again reaction mixture is heated to 72°C (extension); a temperature optimum for DNApolymerase functions. This starts synthesis of the new DNA strand. Than reaction mixture is cooled to lower temperature for short term storage, if required. This completes one cycle. After first cycle, one DNA molecule has become two. After multiple cycle of the PCR reaction, the specific sequence will be accumulated in billions of copies.

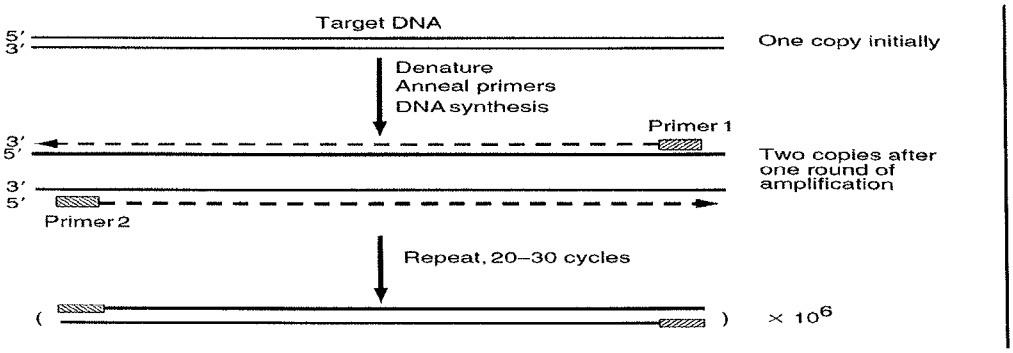
The PCR reaction requires the following components:

DNA template: DNA template is DNA target sequence. As explained earlier, at the beginning of the reaction, high temperature is applied to

separate both the DNA strands from each other so that primers can bind during annealing.

DNA polymerase: DNA polymerase sequentially adds nucleotides complimentary to template strand at 3'-OH of the bound primers and synthesizes new strands of DNA complementary to the target sequence. The most commonly used DNA polymerase is Taq DNA polymerase (from Thermus aquaticus, a thermophillic bacterium) because of high temperature stability. Pfu DNA polymerase (from Pyrococcus furiosus) is also used widely because of its higher fidelity (accuracy of adding complimentary nucleotide). Mg2+ ions in the buffer act as co-factor for DNA polymerase enzyme and hence are required for the reaction.

Primers: Primers are synthetic DNA strands of about 18 to 25 nucleotides complementary to 3' end of the template strand. DNA polymerase starts synthesizing new DNA from the 3' end of the primer

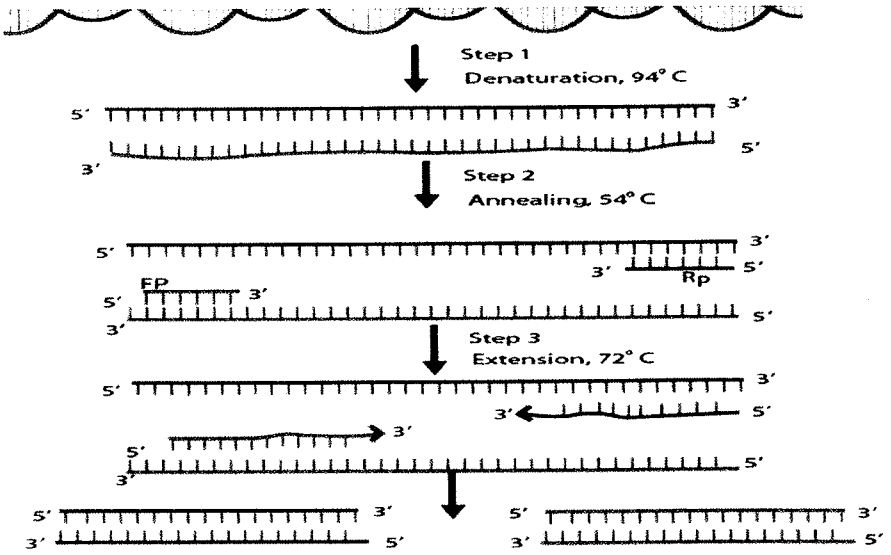


Two primers must be designed for PCR; the forward primer and the reverse primer. The forward primer is complimentary to the 3' end of antisense strand (3'-5') and the reverse primer is complimentary to the 3' end of sense strand (5'-3'). If we consider the sense strand (5'-3') of a gene, for designing primers, then forward primer is the beginning of the gene and the reverse primer is the reverse-compliment of the 3' end of the gene.

Nucleotides (dNTPs or deoxynucleotide triphosphates): All types of nucleotides are "building blocks" for new DNA strands and essential for reaction. It includes Adenine(A), Guanine(G), Cytosine(C), Thymine(T) or Uracil(U).

Procedure

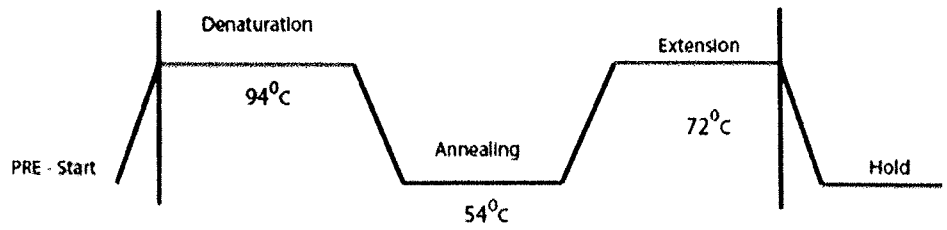
There are three major steps in a PCR, which are repeated for 30 or 40 cycles. This is done on an automated cycler, which can heat and cool the tubes with the reaction mixture in a very short time.



1. Denaturation at 94°C : During the heating step (denaturation), the reaction mixture is heated to 94°C for 1 min, which causes separation of DNA double stranded. Now, each strand acts as template for synthesis of complimentary strand.

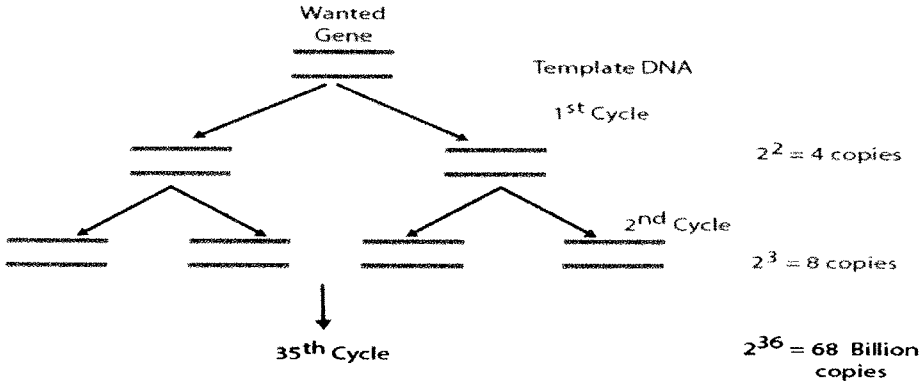
2. Annealing at 54°C : This step consist of cooling of reaction mixture after denaturation step to 54°C, which causes hybridization (annealing) of primers to separated strand of DNA (template). The length and GC-content (guanine-cytosine content) of the primer should be sufficient for stable binding with template. Please recall our discussion about DNA structure during earlier lectures. Guanine pairs with cytosine with three hydrogen bonding adenine binds with thymine with two hydrogen bonds. Thus, higher GC content results in stronger binding. In case GC content is less, length may be increased to have stronger binding (more number of H bonding between primer and template).

3. Extension at 72°C : The reaction mixture is heated to 72°C which is the ideal working temperature for the Taq polymerase. The polymerase adds nucleotide (dNTP's) complimentary to template on 3' –OH of primers thereby extending the new strand.



4. Final hold: First three steps are repeated 35-40 times to produce millions of exact copies of the target DNA. Once several cycles are completed, during the hold step, 4–15 °C temperature is maintained for short-term storage of the amplified DNA sample.

PCR-an exponential cycle: As both strands are copied during PCR, there is an exponential increase of the number of copies of the gene as shown in the figure. Suppose there is only one copy of the desired gene before the PCR starts, after one cycle of PCR, there will be 2 copies, after two cycles of PCR, there will be 4 copies. After three cycles there will be 8 copies and so on.



Various PCR methods

Nested PCR - use to synthesize more reliable product - PCR using a outer set of primers and the product of this PCR is used for further PCR reaction using an inner set of primers.

Inverse PCR - for amplification of regions flanking a known sequence. DNA is digested, the desired fragment is circularise by ligation, then PCR using primer complementary to the known sequence extending outwards.

AP-PCR (arbitrary primed)/RAPD (random amplified polymorphic DNA) - methods for creating genomic fingerprints from species with little-known target sequences by amplifying using arbitrary oligonucleotides. It is normally done at low and then high stringency to determine the relatedness of species or for analysis of Restriction Fragment Length Polymorphisms (RFLP).

RT-PCR (reverse transcriptase) - using RNA-directed DNA polymerase to synthesize cDNAs which is then used for PCR and is extremely sensitive for detecting the expression of a specific sequence in a tissue or cells. It may also be use to quantify mRNA transcripts. See also Quantiative RT-PCR, Competitive Quantitative RT-PCR, RT in situ PCR, Nested RT-PCR.

RACE (rapid amplificaton of cDNA ends) - used where information about DNA/protein sequence is limited. Amplify 3' or 5' ends of cDNAs generating fragments of cDNA with only one specific primer each (+ one adaptor primer). Overlapping RACE products can then be combined to produce full cDNA. See also Gibco manual.

DD-PCR (differential display) - used to identify differentially expressed genes in different tissues. First step involves RT-PCR, then amplification using short, intentionally nonspecific primers. Get series of band in a high-resolution gel and compare to that from other tissues, any bands unique to single samples are considered to be differentially expressed.

Multiplex-PCR - 2 or more unique targets of DNA sequences in the same specimen are amplified simultaneously. One can be use as control to verify the integrity of PCR. Can be used for mutational analysis and identification of pathogens.