# GENETIC ENGINEERING - (E CONTENT)


## DR. DHIVYA L

**Assistant Professor Department of Biotechnology**

**M.M.E.S. WOMEN'S ARTS AND SCIENCE COLLEGE**

# Cloning and Genetic Engineering

## Learning Objectives

By the end of this section, you will be able to:

- Explain the basic techniques used to manipulate genetic material
- Explain molecular and reproductive cloning

Biotechnology is the use of artificial methods to **modify the genetic material** of living organisms or cells to produce novel compounds or to perform new functions. Biotechnology has been used for improving livestock and crops since the beginning of agriculture through selective breeding. Since the discovery of the structure of DNA in 1953, and particularly since the development of tools and methods to manipulate DNA in the 1970s, biotechnology has become synonymous with the manipulation of organisms' DNA at the molecular level. The primary applications of this technology are in medicine (for the production of vaccines and antibiotics) and in agriculture (for the genetic modification of crops). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels, as well as many household applications such as the use of enzymes in laundry detergent.

## Manipulating Genetic Material

To accomplish the applications described above, biotechnologists must be able to extract, manipulate, and analyze nucleic acids.

## Review of Nucleic Acid Structure

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base). The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus of eukaryotic organisms is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases.

Unlike DNA in eukaryotic cells, RNA molecules leave the nucleus. Messenger RNA (mRNA) is analyzed most frequently because it represents the protein-coding genes that are being expressed in the cell.

## Isolation of Nucleic Acids

To study or manipulate nucleic acids, the DNA must first be extracted from cells. Various techniques are used to extract different types of DNA (Figure 10.2). Most nucleic acid extraction techniques involve steps to break open the cell, and then the use of enzymatic reactions to destroy all undesired macromolecules. Cells are broken open using a detergent solution containing buffering compounds. To prevent degradation

and contamination, macromolecules such as proteins and RNA are inactivated using enzymes. The DNA is then brought out of solution using alcohol. The resulting DNA, because it is made up of long polymers, forms a gelatinous mass.

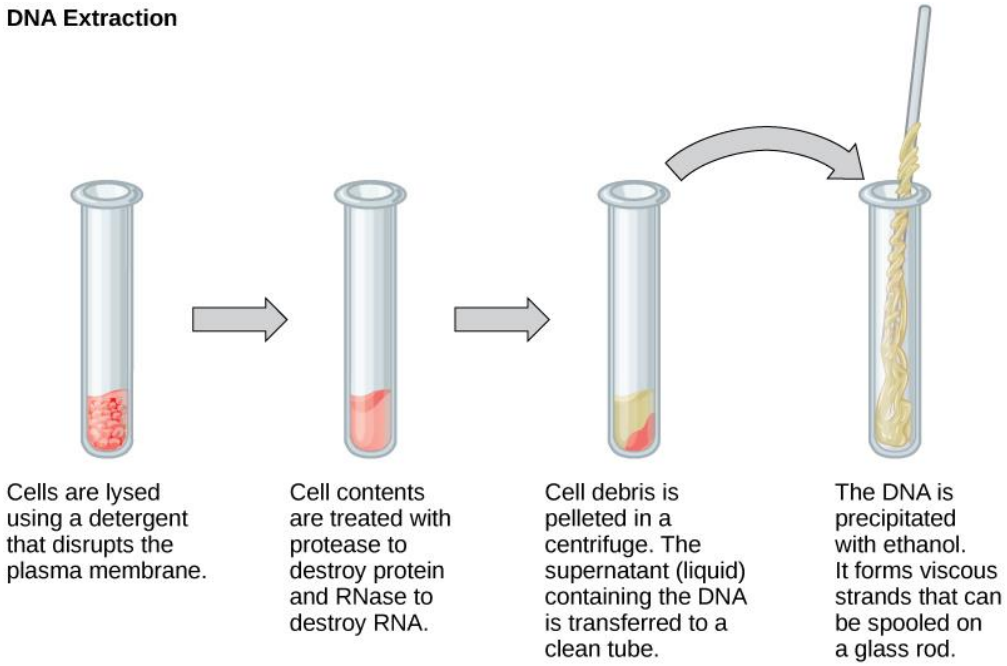



Figure 10.2 This diagram shows the basic method used for the extraction of DNA.

RNA is studied to understand gene expression patterns in cells. RNA is naturally very unstable because enzymes that break down RNA are commonly present in nature. Some are even secreted by our own skin and are very difficult to inactivate. Similar to DNA extraction, RNA extraction involves the use of various buffers and enzymes to inactivate other macromolecules and preserve only the RNA.

**Gel Electrophoresis**

Because nucleic acids are negatively charged ions at neutral or alkaline pH in an aqueous environment, they can be moved by an electric field. Gel electrophoresis is a technique used to separate charged molecules on the basis of size and charge. The nucleic acids can be separated as whole chromosomes or as fragments. The nucleic acids are loaded into a slot at one end of a gel matrix, an electric current is applied, and negatively charged molecules are pulled toward the opposite end of the gel (the end with the positive electrode). Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. The nucleic acids in a gel matrix are invisible until they are stained with a compound that allows them to be seen, such as a dye. Distinct fragments of nucleic acids appear as bands at specific distances from the top of the gel (the negative electrode end) that are based on their size (Figure 10.3). A mixture of many fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.

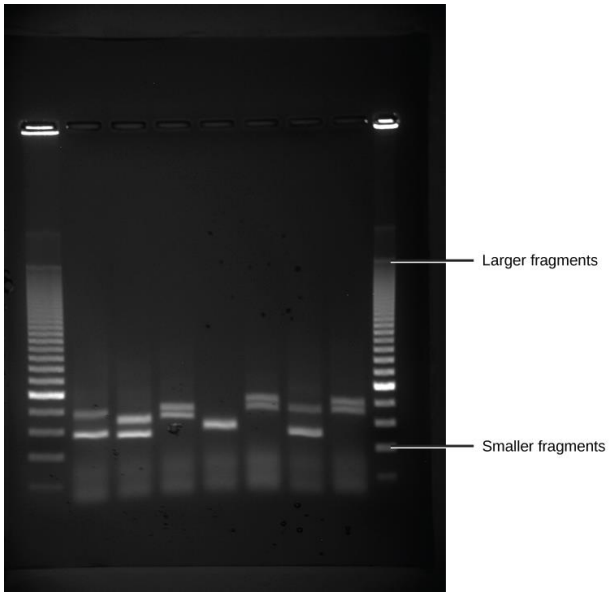



Figure 10.3 Shown are DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light. (credit: modification of work by James Jacob, Tompkins Cortland Community College)

**Polymerase Chain Reaction**

DNA analysis often requires focusing on one or more specific regions of the genome. It also frequently involves situations in which only one or a few copies of a DNA molecule are available for further analysis. These amounts are insufficient for most procedures, such as gel electrophoresis. **Polymerase chain reaction (PCR)** is a technique used to **rapidly increase the number of copies** of specific regions of DNA for further analyses (Figure 10.4). PCR uses a special form of DNA polymerase, the enzyme that replicates DNA, and other short nucleotide sequences called primers that base pair to a specific portion of the DNA being replicated. PCR is used for many purposes in laboratories. These include: 1) the identification of the owner of a DNA sample left at a crime scene; 2) paternity analysis; 3) the comparison of small amounts of ancient DNA with modern organisms; and 4) determining the sequence of nucleotides in a specific region.

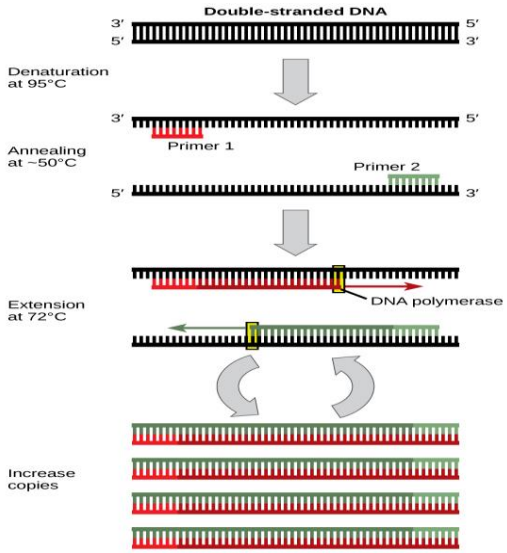


Figure 10.4 Polymerase chain reaction, or PCR, is used to produce many copies of a specific sequence of DNA using a special form of DNA polymerase. Figure showing PCR in 4 steps. First, the double strand of DNA is denatured at 95 degrees Celsius to separate the strands. The 2 strands are then annealed at approximately 50 degrees Celsius using primers. DNA polymerase then extends the new strands at 72 degrees Celsius. The fourth step shows that this procedure takes place many times, resulting in an increase in copies of the original DNA.

**Cloning**

In general, cloning means the creation of a perfect replica. Typically, the word is used to describe the creation of a genetically identical copy. In biology, the re-creation of a whole organism is referred to as "reproductive cloning." Long before attempts were made to clone an entire organism, researchers learned how to copy short stretches of DNA—a process that is referred to as molecular cloning.

**Molecular Cloning**

Cloning allows for the creation of **multiple copies of genes, expression of genes, and study of specific genes**. To get the DNA fragment into a bacterial cell in a form that will be copied or expressed, the fragment is first inserted into a plasmid. A **plasmid** (also called a vector in this context) is a small circular DNA molecule that replicates independently of the chromosomal DNA in bacteria. In cloning, the plasmid molecules can be used to provide a "vehicle" in which to insert a desired DNA fragment. Modified plasmids are usually reintroduced into a bacterial host for replication. As the bacteria divide, they copy their own DNA (including the plasmids). The inserted DNA fragment is copied along with the rest of the bacterial DNA. In a bacterial cell, the fragment of DNA from the human genome (or another organism that is being studied) is referred to as foreign DNA to differentiate it from the DNA of the bacterium (the host DNA).

Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as antibiotic resistance (the ability to be unaffected by antibiotics). Plasmids have been highly engineered as vectors for molecular cloning and for the subsequent large-scale production of important molecules, such as insulin. A valuable characteristic of plasmid vectors is the ease with which a foreign DNA fragment can be introduced. These plasmid vectors contain many short DNA sequences that can be cut with different commonly available restriction enzymes. **Restriction enzymes** (also called restriction endonucleases) recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction enzymes make **staggered cuts** in the two strands of DNA, such that the cut ends have a 2- to 4-nucleotide single-stranded overhang. The sequence that is recognized by the restriction enzyme is a four- to eight-nucleotide sequence that is a **palindrome**. Like with a word palindrome, this means the sequence reads the same forward and backward. In most cases, the sequence reads the same forward on one

strand and backward on the complementary strand. When a staggered cut is made in a sequence like this, the overhangs are complementary (Figure 10.5).

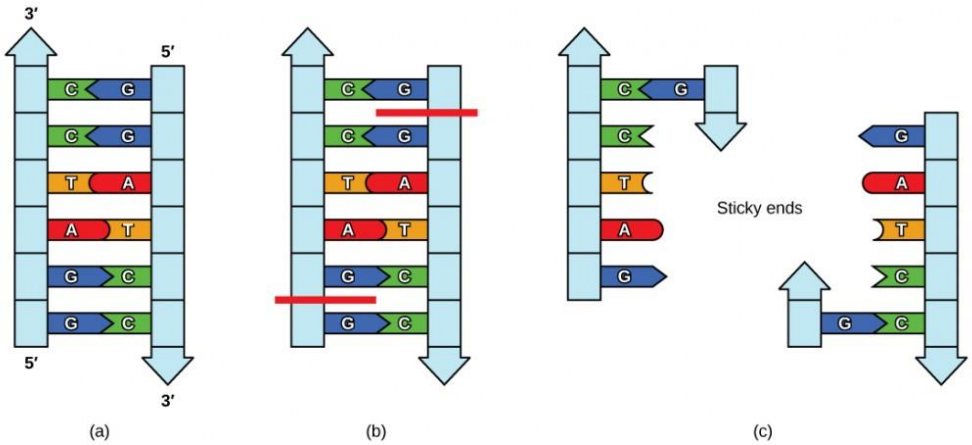



Figure 10.5 In this (a) six-nucleotide restriction enzyme recognition site, notice that the sequence of six nucleotides reads the same in the 5′ to 3′ direction on one strand as it does in the 5′ to 3′ direction on the complementary strand. This is known as a palindrome. (b) The restriction enzyme makes breaks in the DNA strands, and (c) the cut in the DNA results in "sticky ends". Another piece of DNA cut on either end by the same restriction enzyme could attach to these sticky ends and be inserted into the gap made by this cut.

Because these overhangs are capable of coming back together by hydrogen bonding with complementary overhangs on a piece of DNA cut with the same restriction enzyme, these are called "sticky ends." The process of forming hydrogen bonds between complementary sequences on single strands to form double-stranded DNA is called annealing. Addition of an enzyme called DNA ligase, which takes part in DNA replication in cells, permanently joins the DNA fragments when the sticky ends come together. In this way, any DNA fragment can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction enzyme (Figure 10.6).

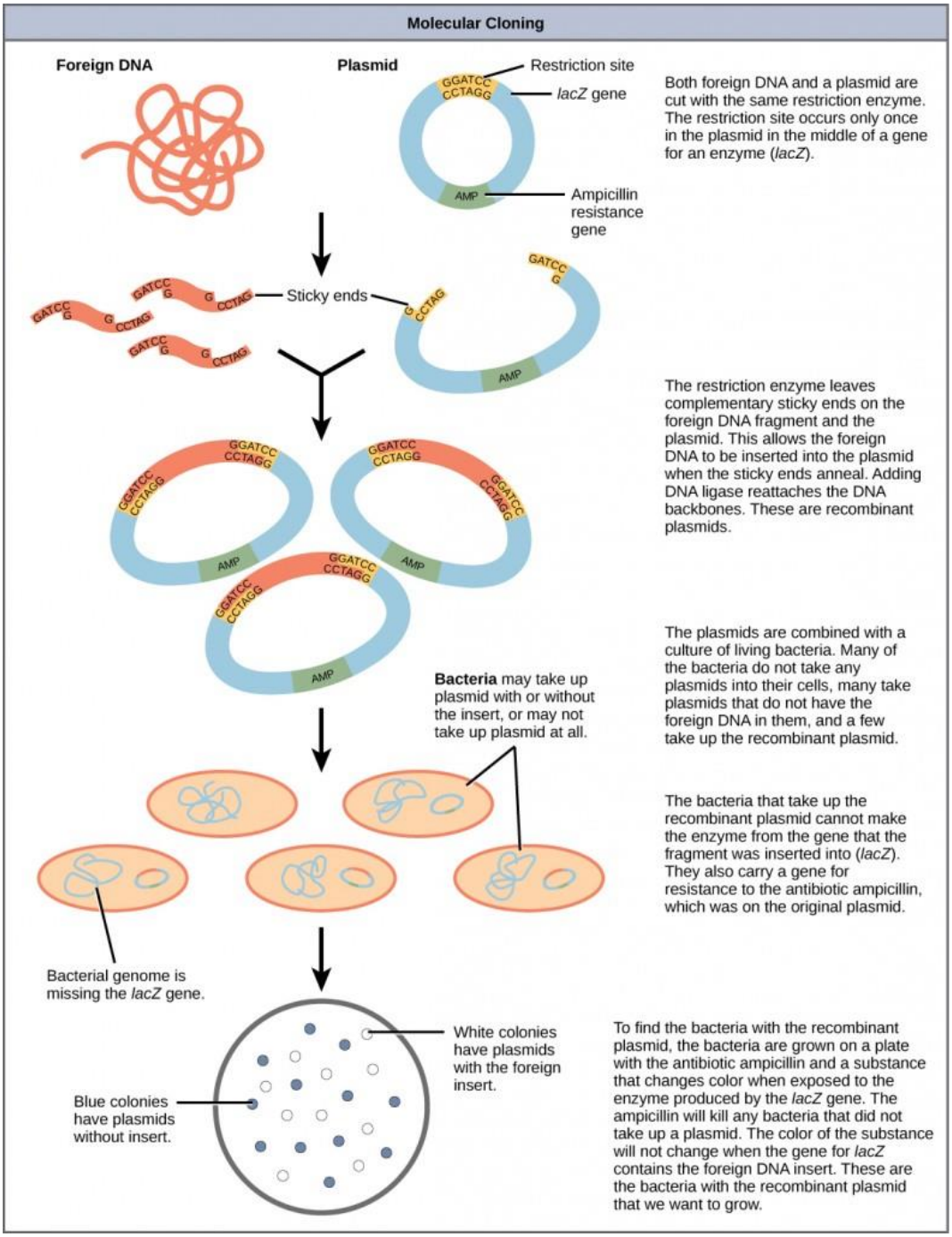



Figure 10.6 This diagram shows the steps involved in molecular cloning.

Plasmids with foreign DNA inserted into them are called recombinant DNA molecules because they contain new combinations of genetic material. Proteins that are produced from recombinant DNA molecules are called recombinant proteins. Not all recombinant plasmids are capable of expressing genes. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

**Reproductive Cloning**

Reproductive cloning is a method used to make a clone or an **identical copy of an entire multicellular organism**. Most multicellular organisms undergo reproduction by sexual means, which involves the contribution of DNA from two individuals (parents), making it impossible to generate an identical copy or a

clone of either parent. Recent advances in biotechnology have made it possible to reproductively clone mammals in the laboratory.

Natural sexual reproduction involves the union, during fertilization, of a sperm and an egg. Each of these gametes is haploid, meaning they contain one set of chromosomes in their nuclei. The resulting cell, or zygote, is then diploid and contains two sets of chromosomes. This cell divides mitotically to produce a multicellular organism. However, the union of just any two cells cannot produce a viable zygote; there are components in the cytoplasm of the egg cell that are essential for the early development of the embryo during its first few cell divisions. Without these provisions, there would be no subsequent development. Therefore, to produce a new individual, both a diploid genetic complement and an egg cytoplasm are required. The approach to producing an artificially cloned individual is to take the egg cell of one individual and to remove the haploid nucleus. Then a diploid nucleus from a body cell of a second individual, the donor, is put into the egg cell. The egg is then stimulated to divide so that development proceeds. This sounds simple, but in fact it takes many attempts before each of the steps is completed successfully.

The first cloned agricultural animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low. Dolly lived for six years and died of a lung tumor (Figure 10.7). There was speculation that because the cell DNA that gave rise to Dolly came from an older individual, the age of the DNA may have affected her life expectancy. Since Dolly, several species of animals (such as horses, bulls, and goats) have been successfully cloned.

There have been attempts at producing cloned human embryos as sources of embryonic stem cells. In the procedure, the DNA from an adult human is introduced into a human egg cell, which is then stimulated to divide. The technology is similar to the technology that was used to produce Dolly, but the embryo is never implanted into a surrogate mother. The cells produced are called embryonic stem cells because they have the capacity to develop into many different kinds of cells, such as muscle or nerve cells. The stem cells could be used to research and ultimately provide therapeutic applications, such as replacing damaged tissues. The benefit of cloning in this instance is that the cells used to regenerate new tissues would be a perfect match to the donor of the original DNA. For example, a leukemia patient would not require a sibling with a tissue match for a bone-marrow transplant.

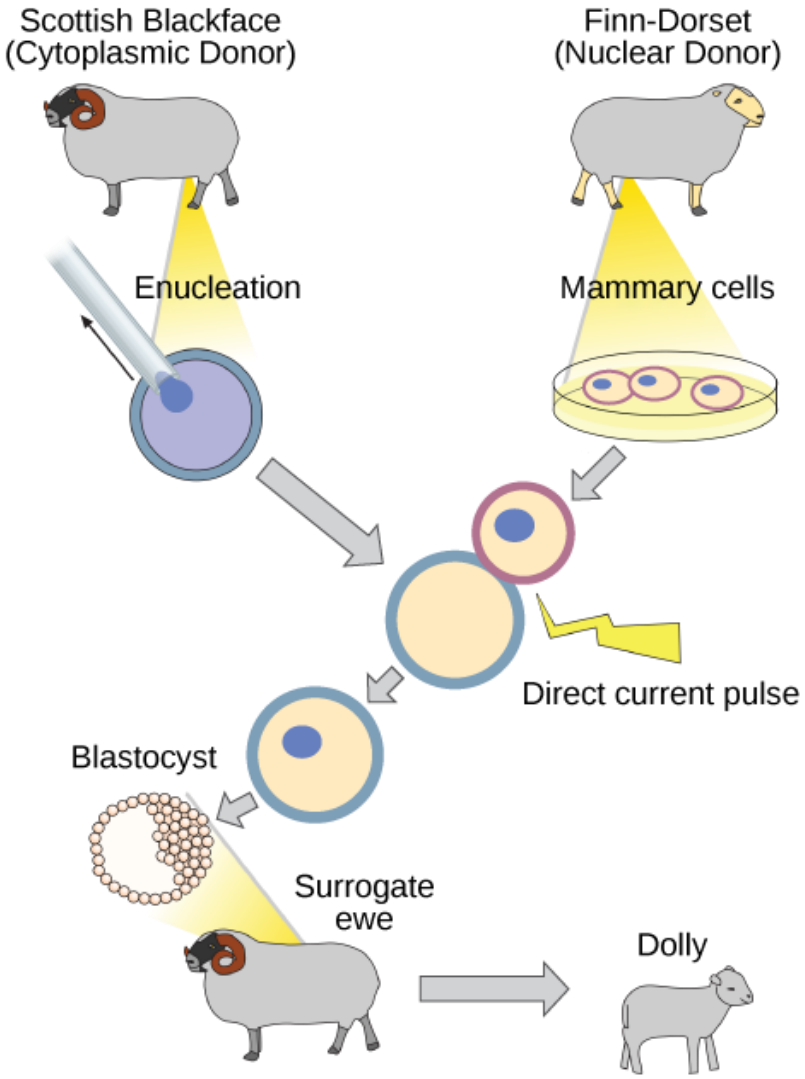



Figure 10.7 Dolly the sheep was the first agricultural animal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The enucleated egg was placed next to the other cell, then they were shocked to fuse. They were shocked again to start division. The cells were allowed to divide for several days until an early embryonic stage was reached, before being implanted in a surrogate mother.

Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

Because even though the original cell came from a Scottish Blackface sheep and the surrogate mother was a Scottish Blackface, the DNA came from a Finn-Dorset.

**Genetic Engineering**

Using recombinant DNA technology to modify an organism's DNA to achieve desirable traits is called genetic engineering. Addition of foreign DNA in the form of recombinant DNA vectors that are generated by molecular cloning is the most common method of genetic engineering. An organism that receives the recombinant DNA is called a genetically modified organism (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is called transgenic. Bacteria, plants, and animals have

been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. These applications will be examined in more detail in the next module.

Although the classic methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, called reverse genetics, has resulted in reversing the classical genetic methodology. One example of this method is analogous to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the wing's function is flight. The classic genetic method compares insects that cannot fly with insects that can fly, and observes that the non-flying insects have lost wings. Similarly in a reverse genetics approach, mutating or deleting genes provides researchers with clues about gene function. Alternately, reverse genetics can be used to cause a gene to overexpress itself to determine what phenotypic effects may occur.

**Section Summary**

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA can be amplified by PCR. DNA can be cut (and subsequently re-spliced together) using restriction enzymes. The molecular and cellular techniques of biotechnology allow researchers to genetically engineer organisms, modifying them to achieve desirable traits.

Cloning may involve cloning small DNA fragments (molecular cloning), or cloning entire organisms (reproductive cloning). In molecular cloning with bacteria, a desired DNA fragment is inserted into a bacterial plasmid using restriction enzymes and the plasmid is taken up by a bacterium, which will then express the foreign DNA. Using other techniques, foreign genes can be inserted into eukaryotic organisms. In each case, the organisms are called transgenic organisms. In reproductive cloning, a donor nucleus is put into an enucleated egg cell, which is then stimulated to divide and develop into an organism.

In reverse genetics methods, a gene is mutated or removed in some way to identify its effect on the phenotype of the whole organism as a way to determine its function.

**Glossary**

**anneal:** in molecular biology, the process by which two single strands of DNA hydrogen bond at complementary nucleotides to form a double-stranded molecule

**biotechnology:** the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions

**cloning:** the production of an exact copy—specifically, an exact genetic copy—of a gene, cell, or organism

**gel electrophoresis:** a technique used to separate molecules on the basis of their ability to migrate through a semisolid gel in response to an electric current

**genetic engineering:** alteration of the genetic makeup of an organism using the molecular methods of biotechnology
genetically modified organism (GMO): an organism whose genome has been artificially changed

**plasmid:** a small circular molecule of DNA found in bacteria that replicates independently of the main bacterial chromosome; plasmids code for some important traits for bacteria and can be used as vectors to transport DNA into bacteria in genetic engineering applications

**polymerase chain reaction (PCR):** a technique used to make multiple copies of DNA

**recombinant DNA:** a combination of DNA fragments generated by molecular cloning that does not exist in nature
strong>recombinant protein: a protein that is expressed from recombinant DNA molecules

**restriction enzyme:** an enzyme that recognizes a specific nucleotide sequence in DNA and cuts the DNA double strand at that recognition site, often with a staggered cut leaving short single strands or “sticky” ends

**reverse genetics:** a form of genetic analysis that manipulates DNA to disrupt or affect the product of a gene to analyze the gene’s function

**reproductive cloning:** cloning of entire organisms

**transgenic:** describing an organism that receives DNA from a different species

## DNA MODIFYING ENZYMES

Restriction enzymes and DNA ligases represent the cutting and joining functions in DNA manipulation. All other enzymes involved in genetic engineering fall under the broad category of enzymes known as DNA modifying enzymes. These enzymes are involved in the degradation, synthesis and alteration of the nucleic acids.

**DNA ligase** is an important cellular enzyme, as its function is to repair broken phosphodiester bonds that may occur at random or as a consequence of DNA replication or recombination. In genetic engineering it is used to seal discontinuities in the sugar—phosphate chains that arise when recombinant DNA is made by joining DNA molecules from different sources. It can therefore be thought of as molecular glue, which is used to stick pieces of DNA together.This function is crucial to the success of many experiments, and DNA ligase is therefore a key enzyme in genetic engineering.

The enzyme used most often in experiments is T4 DNA ligase, which is purified from E. coli cells infected with bacteriophage T4. Although the enzyme is most efficient when sealing gaps in fragments that are held together by cohesive ends, it will also join blunt-ended DNA molecules together under appropriate conditionsT4 DNA Ligase catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini in duplex DNA or RNA. The enzyme repairs single-strand nicks in duplex DNA, RNA, or DNA/RNA hybrids. It also joins DNA fragments with either cohesive or blunt termini, but has no activity on single-stranded nucleic acids. The T4 DNA Ligase requires ATP as a cofactor.

## RESTRICTION ENDONUCLEASES

- Also called restriction enzymes
- 1962: “molecular scissors” discovered in in bacteria
- E. coli bacteria have an enzymatic immune system that recognizes and destroys foreign DNA
- 3,000 enzymes have been identified, around 200 have unique properties, many are purified and available commercially

Named for bacterial genus, species, strain, and type

Example: EcoR1

Genus:Escherichia

Species:coli

Strain:R

Order discovered: 1

Enzymes recognize specific 4-8 bp sequences which are palindromic.

Some enzymes cut in a staggered fashion - "sticky ends"

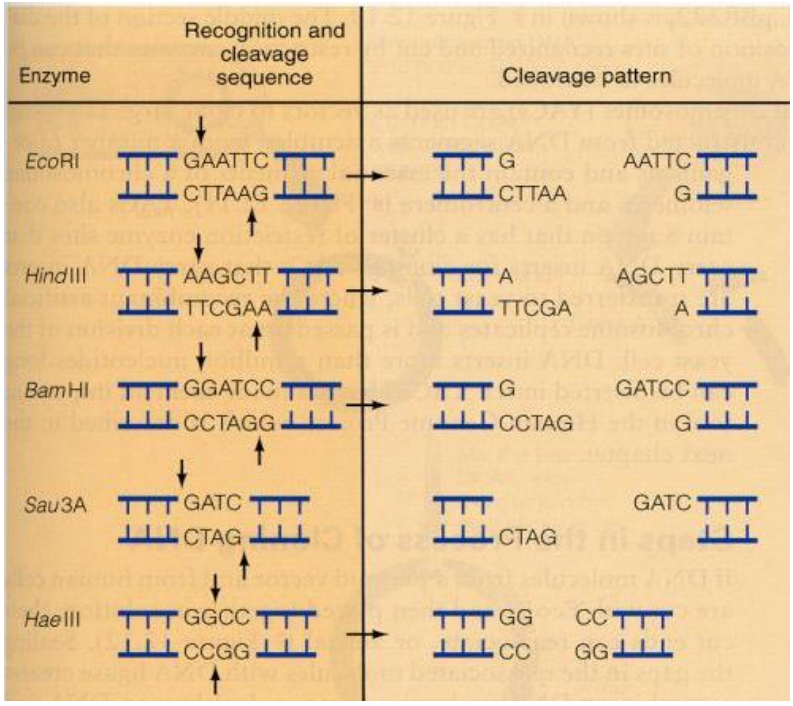
EcoRI 5'…GAATTC…3'

3'…CTTAAG…5'

Some enzymes cut in a direct fashion – "blunt ends"

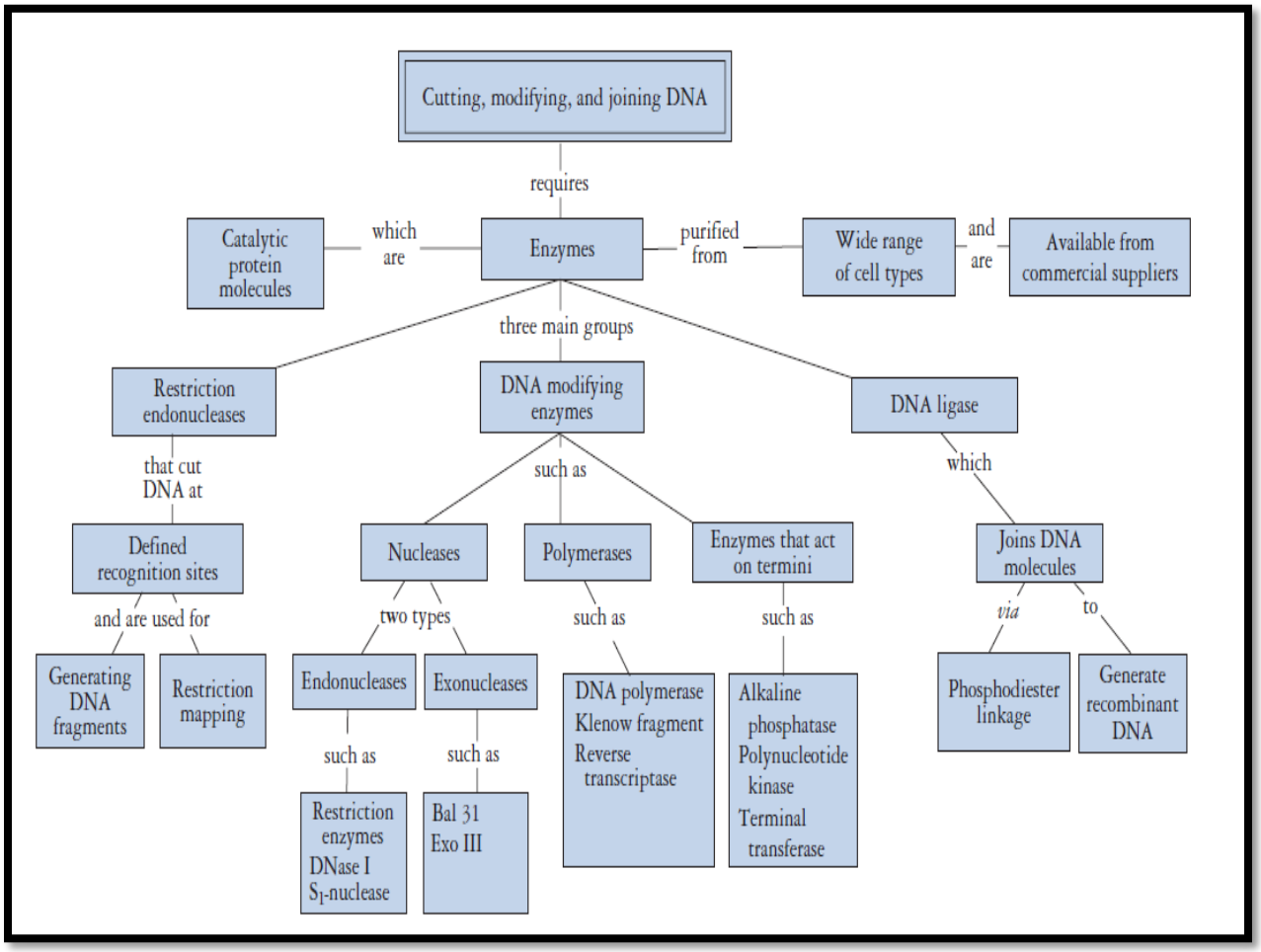
PvuII 5'…CAGCTG…3'

3'…GTCGAC…5'



Uses of restriction enzymes

- RFLP analysis (Restriction Fragment Length Polymorphism)
- DNA sequencing
- DNA storage – libraries
- Transformation
- Large scale analysis – gene chips




**TYPES OF DNA MODIFYING ENZYMES**

Nuclease enzymes degrade nucleic acids by breaking the phosphodiester bond that holds the nucleotides together. Restriction enzymes are good examples of endonucleases, which cut within a DNA strand. A second group of nucleases, which degrade DNA from the

termini of the molecule, are known as **exonucleases.** Apart from restriction enzymes, there are four useful nucleases that are often used in genetic engineering. These are

- **Bal 31**
- **exonuclease III** (exonucleases),
- **deoxyribonuclease I** (DNase I)
- **S1-nuclease** (endonucleases).

These enzymes differ in their precise mode of action and provide the genetic engineer with a variety of strategies for attacking DNA.

- Nuclease Bal 31 is a complex enzyme. Its primary activity is a fast-acting 3' exonuclease, which is coupled with a slow-acting endonuclease. When Bal 31 is present at a high concentration these activities effectively shorten DNA molecules from both termini.
- Exonuclease III is a 3' exonuclease that generates molecules with protruding 5' termini.
- DNase I cuts either single-stranded or double-stranded DNA at essentially random sites.
- Nuclease S1 is specific for single-stranded RNA or DNA.

**POLYMERASES**

Polymerase enzymes synthesise copies of nucleic acid molecules and are used in many genetic engineering procedures. When describing a polymerase enzyme, the terms 'DNA-dependent' or 'RNA-dependent' may be used to indicate the type of nucleic acid template that the enzyme uses.

DNA POLYMERASE I

The enzyme DNA polymerase I has, in addition to its polymerase function, 5'→3' and 3'→5' exonuclease activities. The enzyme catalyses a strand-replacement reaction, where the 5'→3' exonuclease function degrades the non-template strand as the polymerase synthesises the new copy. A major use of this enzyme is in the nick translation procedure for radiolabelling DNA.The 5'→3' exonuclease function of DNA polymerase I can be

removed by cleaving the enzyme to produce what is known as the **Klenow fragment.** This retains the polymerase and 3'→5' exonuclease activities. The Klenow fragment is used where a single-stranded DNA molecule needs to be copied; because the 5'→3' exonuclease function is missing, the enzyme cannot degrade the non-template strand of dsDNA during synthesis of the new DNA. The 3'→5' exonuclease activity is supressed under the conditions normally used for the reaction. Major uses for the Klenow fragment include radiolabelling by primed synthesis and DNA sequencing by the dideoxy method in addition to the copying of single-stranded DNAs during the production of recombinants.A modified form of DNA polymerase I called the Klenow fragment is a useful polymerase that is used widely in a number of applications

**Reverse transcriptase** (RTase) is an RNA-dependent DNA polymerase, and therefore produces a DNA strand from an RNA template.It has no associated exonuclease activity. The enzyme is used mainly for copying mRNA molecules in the preparation of cDNA (**complementary** or **copy DNA**) for cloning, although it will also act on DNA templates.Reverse transcriptase is a key enzyme in the generation of cDNA; the enzyme is an RNA-dependent DNA polymerase, which produces a DNA copy of an mRNA molecule.

ENZYMES THAT MODIFY THE ENDS OF DNA MOLECULES

The enzymes alkaline phosphatase, polynucleotide kinase, and terminal transferase act on the termini of DNA molecules and provide important functions that are used in a variety of ways. The phosphatase and kinase enzymes, as their names suggest, are involved in the removal or addition of phosphate groups.

- Bacterial alkaline phosphatase (there is also a similar enzyme, calf intestinal alkaline phosphatase) removes phosphate groups from the 5 ends of DNA, leaving a 5-OH group. The enzyme is used to prevent unwanted ligation of DNA molecules, which can be a problem in certain cloning procedures. It is also used prior to the addition of radioactive phosphate to the 5 ends of DNAs by polynucleotide kinase.

- Terminal transferase (terminal deoxynucleotidyl transferase) repeatedly adds nucleotides to any available 3 terminus. Although it works best on protruding 3 ends, conditions can be adjusted so that blunt-ended or 3-recessed molecules may be utilised. The enzyme is

mainly used to add homopolymer tails to DNA molecules prior to the construction of recombinants

**VECTORS**

A cloning **vector** is a DNA molecule used as a vehicle to transfer foreign genetic material

Few of the generally used cloning vectors include:

1. Plasmids
2. Bacterophages
3. Cosmids
4. Yeast artifical chromosomes (YAC's)
5. Bacterial arifical chromosomes (BAC's)

**Comparison of vectors available for cloning DNA fragments**

| Vector | Host cell | Vector structure | Insert range (kb) |
|---|---|---|---|
| M13 | *E. coli* | Circular virus | 1–4 |
| Plasmid | *E. coli* | Circular plasmid | 1–5 |
| Phage λ | *E. coli* | Linear virus | 2–25 |
| Cosmids | *E. coli* | Circular plasmid | 35–45 |
| BACs | *E. coli* | Circular plasmid | 50–300 |
| YACs | *S. cerevisiae* | Linear chromosome | 100–2000 |

**Vectors for yeast and other fungi**

The yeast *Saccharomyces cerevisiae* is one of the most important organisms in biotechnology. Development of cloning vectors for yeast was stimulated by the discovery of a plasmid that is present in most strains of *S. cerevisiae*. The 2 µm plasmid, as it is called, is one of only a very limited number of plasmids found in eukaryotic cells.

**Why is it popularly used?**

Yeast cells are much easier to grow and manipulate than plant and animal cells.

The cellular biochemistry and regulation of yeast are very like those of higher eukaryotes.

There are many yeast homologues of human genes, e.g. those involved in cell division. Thus yeast can be a very good surrogate host for studying the structure and function of eukaryotic gene products.

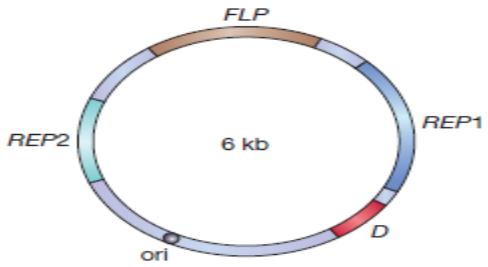
**The different kind of cloning vector derived from yeast are**:

1. 2 µm plasmid
2. yeast episomal plasmids (YEps)
3. Yeast integrative plasmids (YIps)
4. Yeast replicative plasmids (YRps)
5. yeast artificial chromosome (YAC)

| Vector | Transformation frequency | Copy no./cell | Loss in non-selective medium | Disadvantages | Advantages |
|---|---|---|---|---|---|
| YIp | $10^2$ transformants per µg DNA | 1 | Much less than 1% per generation | 1 Low transformation frequency<br>2 Can only be recovered from yeast by cutting chromosomal DNA with restriction endonuclease which does not cleave original vector containing cloned gene | 1 Of all vectors, this kind give most stable maintenance of cloned genes<br>2 An integrated YIp plasmid behaves as an ordinary genetic marker, e.g. a diploid heterozygous for an integrated plasmid segregates the plasmid in a Mendelian fashion<br>3 Most useful for surrogate genetics of yeast, e.g. can be used to introduce deletions, inversions and transpositions (see Botstein & Davis 1982) |
| YEp | $10^3$–$10^5$ transformants per µg DNA | 25–200 | 1% per generation | Novel recombinants generated *in vivo* by recombination with endogenous 2 µm plasmid | 1 Readily recovered from yeast<br>2 High copy number<br>3 High transformation frequency<br>4 Very useful for complementation studies |
| YRp | $10^4$ transformants per µg DNA | 1–20 | Much greater than 1% per generation but can get chromosomal integration | Instability of transformants | 1 Readily recovered from yeast<br>2 High copy number. Note that the copy number is usually less than that of YEp vectors but this may be useful if cloning gene whose product is deleterious to the cell if produced in excess<br>3 High transformation frequency<br>4 Very useful for complementation studies<br>5 Can integrate into the chromosome |
| YCp | $10^4$ transformants per µg DNA | 1–2 | Less than 1% per generation | Low copy number makes recovery from yeast more difficult than that with YEp or YRp vectors | 1 Low copy number is useful if product of cloned gene is deleterious to cell<br>2 High transformation frequency<br>3 Very useful for complementation studies<br>4 At meiosis generally shows Mendelian segregation |
| YAC | | 1–2 | Depends on length: the longer the YAC the more stable it is | Difficult to map by standard techniques | 1 High-capacity cloning system permitting DNA molecules greater than 40 kb to be cloned<br>2 Can amplify large DNA molecules in a simple genetic background |

## Vectors for yeast and other fungi

### 2 µm plasmid




2 µm plasmid.

The 2 µm plasmid is an excellent basis for a cloning vector. It is 6 kb in size and has a high copy number of between 70 and 200.

Replication makes use of a plasmid origin, several enzymes provided by the host cell, and the proteins coded by the REP1 and REP2 genes carried by the plasmid.

However, all is not perfectly straightforward in using the 2 µm plasmid as a cloning vector.

First, there is the question of a selectable marker. For this purpose a normal yeast gene is used, generally one that codes for an enzyme involved in amino acid biosynthesis. An example is the gene LEU2, which codes for β-isopropyl-malate dehydrogenase, one of the enzymes involved in the conversion of pyruvic acid to leucine.

In order to use LEU2 as a selectable marker, a special kind of host organism is needed. The host must be an auxotrophic mutant that has a non-functional LEU2 gene.

Such a leu2− yeast is unable to synthesize leucine and can survive only if this amino acid is supplied as a nutrient in the growth medium. Selection is possible because transformants contain a plasmid-borne copy of the LEU2 gene, and so are able to grow in the absence of the amino acid. In a cloning experiment, cells are plated out onto minimal medium, which contains no added amino acids. Only transformed cells are able to survive and form colonies and thus are easily differentiated.

Yeast artificial chromosome (YAC) is a human-engineered DNA molecule used to clone DNA sequences in yeast cells. YACs are often used in connection with the mapping and sequencing of genomes. Segments of an organism's DNA, up to one million base pairs in length, can be inserted into YACs. The YACs, with their inserted DNA, are then taken up by yeast cells. As the yeast cells grow and divide, they amplify the YAC DNA, which can be isolated and used for DNA mapping and sequencing. It is abbreviation is YAC.

They carry large amounts of DNA so a long distance map of the region can be obtained in several steps. The repetitive rAegions may be 10-20 kb in length they are rarely longer than 50 kb. Thus a YAC with 100kb will contain some region that is single copy which can be used for further steps in the walk.

**Features of YACs**

1. Large DNA (>100 kb) is ligated between two arms. Each arm ends with a yeast telomere so that the product can be stabilized in the yeast cell. Interestingly, larger YACs are more stable than shorter ones, which favors cloning of large stretches of DNA.
2. One arm contains an autonomous replication sequence (ARS), a centromere (CEN) and aselectable marker (*trp1*). The other arm contains a second selectable marker (*ura3*).
3. Insertion of DNA into the cloning site inactivates a mutant expressed in the vector DNA and red yeast colonies appear.
4. Transformants are identified as those red colonies which grow in a yeast cell that is mutant for *trp1* and *ura3*. This ensures that the cell has received an artificial chromosome with both telomeres (because of complementation of the two mutants) and the artificial chromosome contains insert DNA (because the cell is red).

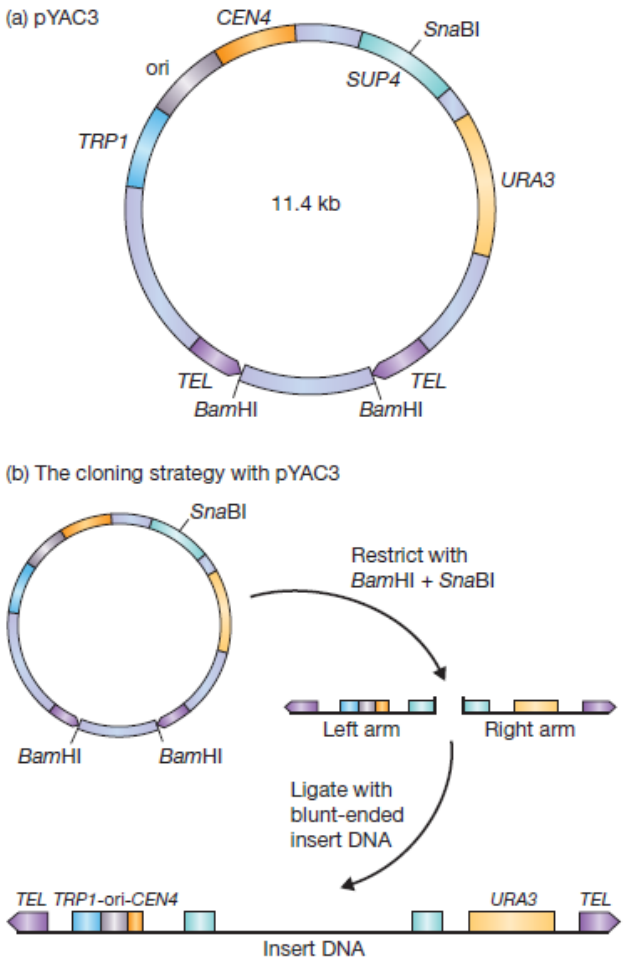
**Example : pYAC3**

pYAC3 is essentially a pBR322 plasmid into which a number of yeast genes have been inserted. Two of these genes, URA3 and TRP1, have been encountered already as the selectable markers for YIp5 and YRp7, respectively. As in YRp7, the DNA fragment that carries TRP1 also

contains an origin of replication, but in pYAC3 this fragment is extended even further to include the sequence called CEN4, which is the DNA from the centromere region of chromosome 4. The TRP1–origin–CEN4 fragment therefore contains two of the three components of the artificial chromosome.

The third component, the telomeres, is provided by the two sequences called TEL.

These are not themselves complete telomere sequences, but once inside the yeast nucleus they act as seeding sequences onto which telomeres will be built. This just leaves one other part of pYAC3 that has not been mentioned: SUP4, which is the selectable marker into which new DNA is inserted during the cloning experiment.



## Bacterial artificial chromosome

A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on naturally occurring F-factor plasmid found in the bacterium *E. coli.*

Bacterial Artificial Chromosomes (BAC) have been developed to hold much larger pieces of DNA than a plasmid can. BAC vectors were originally created F' plasmid present in few bacterial plamids. it was found through studies that F' plasmid were able hold up to a million basepairs of DNA from nonself bacteria. In 1992, Hiroaki Shizuya took the parts of F' that were important, cleaned it up, and turned it into a vector.

BAC vectors are able to hold up to 350 kb of DNA and have all of the tools that a vector needs to work properly, like replication origins, antibiotic resistance genes, and convenient places where clone DNA can insert itself.[8] By these vector that can hold larger pieces of DNA, the number of clones required to cover the human could be reduced drastically from 1.8 billion to about 50 million.

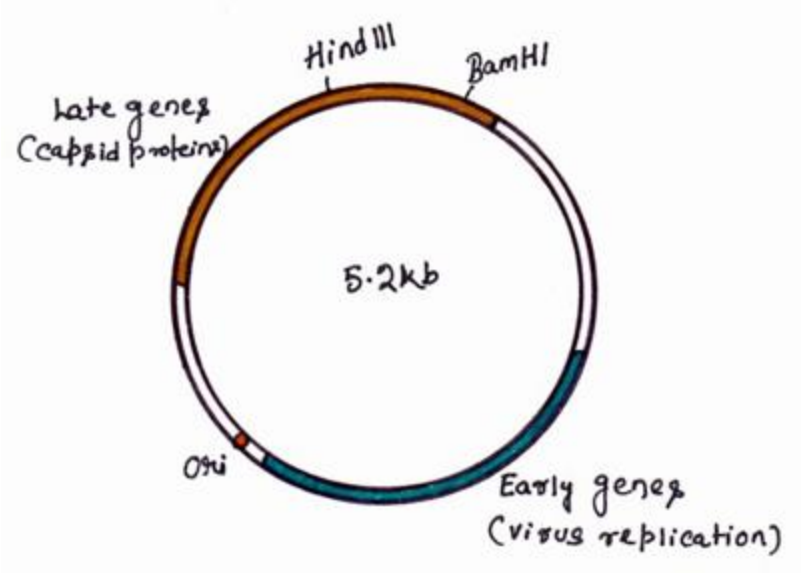
Researchers have modified BAC vectors to become more convenient to use and more useful in specialized situations. In addition to the antibiotic resistance gene that was added to identify transfected bacteria, a gene was added that enabled the bacteria to turn the colourless substance X-gal/IPTG blueSo it is possible to tell not only if a bacteria had been transfected (meaning incorporated into the cell), but also if the bacteria was transfected with the vector containing insert DNA or just the vector alone (remembering that if the vector has properly incorporated the clone DNA, it will have lost its ability to change X-gal/IPTG blue).- Modifications to BAC vectors make them more specialized.

**Few advantages**:

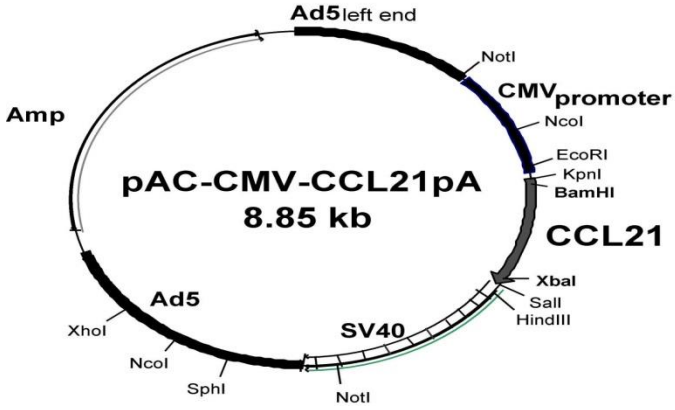
- The creation of BAC vectors has allowed researchers to do many things that they could not do before and do them more quickly and more easily. [8]
- BACs have allowed researchers to look at microbial DNA without having to actually grow the organisms[8]
- BAC vectors are also useful for studying *pathogens*, and are helpful in the development of vaccines. [8] BAC vectors are playing a tremendous role in discovering new and powerful antibiotics in the environment.
- Discovering enzymes that are able to help clean up oil spills or help farmers breed healthier farm animals or even process radioactive waste are just a few examples of what Bacterial Artificial Chromosomes can do.

## Animal viruses as cloning vectors

The first eukaryotic DNA virus was SV40, for which a complete nucleotide sequence and a detailed understanding of transcription were available. The genome of SV40 contains very little non-essential DNA so it is necessary to insert the foreign gene in place of essential viral genes and to propagate the recombinant genome in the presence of a helper virus. This virus is capable of infecting several mammalian species, following a lytic cycle in some host and a lysogenic cycle in others. The genome is 5.2kb in size and contains two sets of genes, the early genes, expressed early in the infection cycle and coding for proteins involved in viral DNA replication, and the late genes, coding for viral capsid proteins (Figure-18). However, all work using SV40 virions to propagate recombinant DNA molecules is severely constrained by the facts that the viral genome is small, 5.24 kb, and that the packaging limits are strict. Such systems cannot, therefore, be used for the analysis of most eukaryotic genes.

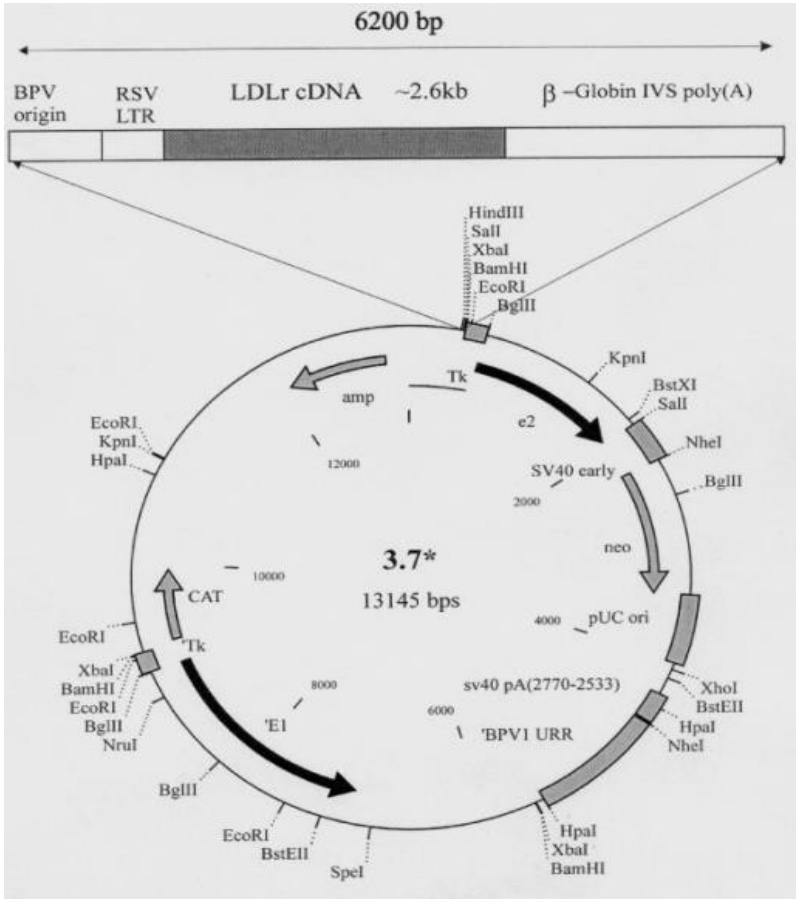



**Adenoviruses,** are a group of viruses which enable larger fragments of DNA to be cloned than it is possible with an SV40 vector. Adenoviruses are more difficult to handle because the genomes are bigger.




**Papillomaviruses**, which also have a relatively high capacity for inserted DNA, have the important advantage of enabling a stable transformed cell line to be obtained. Papillomavirus transformed cells don't contain integrated viral DNA rather they contain between 50 and 300 copies of unintegrated, circular viral DNA although some proportion of these viral genomes exists as concatamers and/or catenates. Bovine papillomavirus(BPV), which causes warts on cattle, has an unusal infection cycle in mouse cells, taking the form of a multi copy plasmid with about 100 molecules present per cell. It doesn't cause the death of the mouse cell, and BPV molecules are passed to daughter cells on cell division. Shuttle vectors consisting of BPV and

pBR322 sequences, and capable of replication in both mouse and bacterial cells, are therefore of great value in animal cell biotechnology.



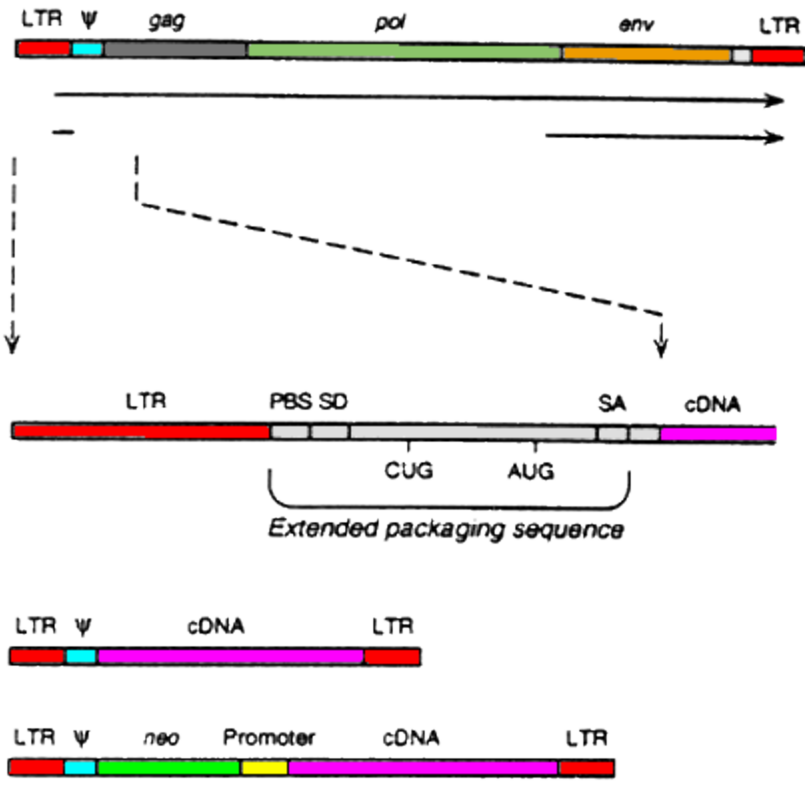

**Retroviruses,** though have single-stranded RNA genomes but provides perhaps the most promising vector system of all. During the process of reverse transcription, sequences from the termini of viral RNA are duplicated to generate long terminal repeats(LTRs). These long terminal repeats contain both the promoter and the polyadenylation signal for the transcription of viral mRNAs. The specificity of proviral DNA integration is also determined by the long terminal repeats. Although retroviruses can integrate at many sites within the cellular genome, integrative recombination always occurs at particular sites at the ends of the LTRs. The sequences appropriately inserted between the two LTRs will be integrated intact which contrasts sharply with the integration of papovavirus or adenovirus DNA, during which extensive rearrangements of the integrated viral sequences are commonplace. A further great advantage of retroviruses is that they are natural transducing viruses.

Retroviral vectors are most frequently based upon the Moloney murine leukaemia virus (Mo-MLV), which is an amphotrophic virus, capable of infecting both mouse cells, enabling vector development in mouse models, & human cells, enabling human treatment. The viral genes (gag, pol & env) are replaced with the transgene of interest & expressed on plasmids in the packaging cell line. Because the non-essential genes lack the packaging sequence (psi) they are not included in the virion particle. To prevent recombination resulting in replication competent retroviruses, all regions of homology with the vector backbone should be removed & the non-essential genes should be expressed by at least two transcriptional units.

The essential regions include the 5' & 3' LTRs & the packaging sequence lying downstream of the 5' LTR. To aid identification of transformed cells selectable markers, such as neomycin & beta galactosidase, can be included & transgenes expression can be improved with the addition of internal ribosome sites. The available carrying capacity for retroviral vectors is approximately 7.5 kb, which is too small for some genes even if the cDNA is used.

A requirement for retroviral integration & expression of viral genes is that the target cells should be dividing. This limits gene therapy to proliferating cells in vivo or ex vivo, whereby cells are removed from the body, treated to stimulate replication & then transduced with the retroviral vector, before being returned to the patient. When treating cancers in vivo, tumour cells are preferentially targeted However, ex vivo cells can be more efficiently transduced, due to exposure to higher virus titres & growth factors. Furthermore ex vivo treated tumour cells will associate with the tumour mass & can direct tumouricidal effects.

Though transgene expression is usually adequate in vitro & initially in vivo, prolonged expression is difficult to attain. Retroviruses are inactivated by c1 complement protein & an anti-alpha galactosyl epitope antibody, both present in human sera. Transgene expression is also reduced by inflammatory interferons, specifically IFN-alpha & IFN-gamma acting on viral LTRs. As the retroviral genome integrates into the host genome it is most likely that the viral LTR promoters are being inactivated, therefore one approach has been to use promoters for host cell genes, such as tyrosine. Clearly this is an area where continued research is needed.




**Baculoviruses,** enable large amounts of proteins to be obtained from genes cloned in insect cells. One of the major proteins encoded by the virus genome is polyhedrin, which accumulates in very large quantities in the nuclei of infected cells, since the gene has an extremely active promoter. The same promoter can be used to drive the over expression of a foreign gene engineered into the baculovirus genome, and large quantities of protein can be produced in infected insect cells in culture. This method is being used increasingly for large-scale culture of proteins of animal origin, since the insect cells can produce many of the post-translational modifications of animal proteins which a bacterial expression system.

**Expression vector**

An **expression vector**, otherwise known as an **expression construct**, is usually a plasmid or virus designed for protein expression in cells. The vector is used to introduce a specific gene into a target cell, and can commandeer the cell's mechanism for protein synthesis to produce the protein encoded by the gene. The plasmid is engineered to contain regulatory sequences that act as enhancer and promoter regions and lead to efficient transcription of the gene carried on the

expression vector.[1] The goal of a well-designed expression vector is the production of significant amount of stable messenger RNA, and therefore proteins. Expression vectors are basic tools for biotechnology and the production of proteins. An example is insulin which is used for medical treatments of diabetes.

**Elements for expression**

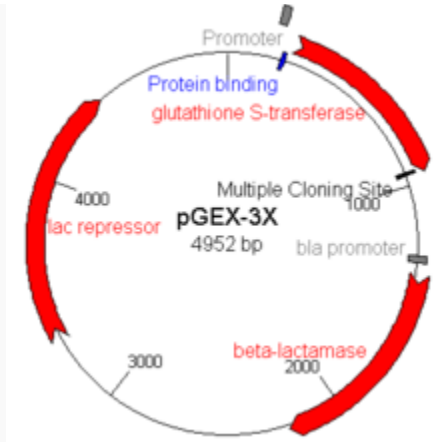
An expression vector must have elements necessary for protein expression. These may include a strong promoter, the correct translation initiation sequence such as a ribosomal binding site and start codon, a strong termination codon, and a transcription termination sequence.[2] There are differences in the machinery for protein synthesis between prokaryotes and eukaryotes, therefore the expression vectors must have the elements for expression that is appropriate for the chosen host. For example, prokaryotes expression vectors would have a Shine-Dalgarno sequence at its translation initiation site for the binding of ribosomes, while eukaryotes expression vectors would contain the Kozak consensus sequence.

The promoter initiates the transcription and is therefore the point of control for the expression of the cloned gene. The promoters used in expression vector are normally inducible, meaning that protein synthesis is only initiated when required by the introduction of an inducer such as IPTG. Protein expression however may also be constitutive (i.e. protein is constantly expressed) in some expression vectors. Low level of constitutive protein synthesis may occur even in expression vectors with tightly controlled promoters.

**Expression systems**

Different organisms may be used to express a target protein, the expression vector used therefore will have elements specific for use in the particular organism. The most commonly used organism for protein expression is the bacterium *Escherichia coli*. However not all proteins can be successfully expressed in *E. coli*, and other systems may therefore be used.

**Bacterial**




An example of a bacterial expression vector is the pGEX-3x plasmid

The expression host of choice for the expression of many proteins is *Escherichia coli* as the production of heterologous protein in *E. coli* is relatively simple and convenient, as well as being rapid and cheap. A large number of *E. coli* expression plasmids are also available suitable for a wide variety of needs. Other bacteria used for protein expression include *Bacillus subtilis*.

Most heterologous proteins are expressed in the cytoplasm of *E. coli*. However, not all proteins formed may be soluble in the cytoplasm, and incorrectly folded proteins formed in cytoplasm can form insoluble aggregates called inclusion body. Such insoluble proteins will require refolding which can be an involved process and may not produce high yield. Where necessarily, for example when the protein can only fold correctly in an oxidizing environment due to the presence of disulphide bonds, the protein may be targeted to the periplasmic space by the use of an N-terminalsignal sequence. Other more sophisticated systems are being developed; such systems may allow for the expression of proteins previously thought impossible in *E. coli*, such as glycosylated proteins.

The promoters used for these vector are usually based on the promoter of the *lac* operon or the T7 promoter, and they are normally regulated by the *lac* operator. These promoters may also be hybrids of different promoters, for example, the tac promoter is a hybrid of *trp* and *lac* promoters.Note that most commonly used *lac* or *lac*-derived promoters are based on the *lac*UV5 mutant which is insensitive to catabolite repression. This mutant allows for expression of protein under the control of the *lac* promoter when the growth medium contains

glucose since glucose would inhibit protein expression if wild-type *lac* promoter is used.Presence of glucose nevertheless may still be used to reduce background expression through residual inhibition in some systems.

Examples of *E. coli* expression vectors are the pGEX series of vectors where glutathione-S-transferase is used as a fusion partner and protein expression is under the control of the tac promoter, and the pET series of vectors which uses a T7 promoter.

It is possible to simultaneously express two or more different proteins in *E. coli* using different plasmids. However, when 2 or more plasmids are used, each plasmid needs to use a different antibiotic selection as well as a different origin of replication, otherwise the plasmids may not be stably maintained. Many commonly-used plasmids are based on the ColE1replicon and are therefore incompatible with each other; in order for a ColE1-based plasmid to coexist with another in the same cell, the other would need to be of a different replicon, e.g. a p15A replicon-based plasmid such as the pACYC series of plasmids. Another approach would be to use a single two-cistron vector or design the coding sequences in tandem as a bi- or poly-cistronic construct.

**Mammalian**

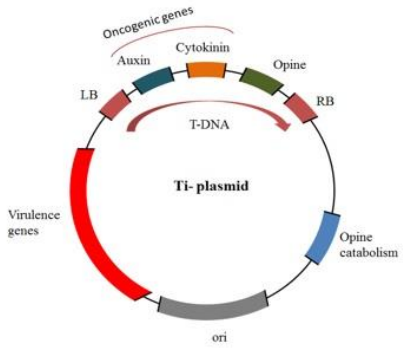
Mammalian expression vectors offer considerable advantages for the expression of mammalian proteins over bacterial expression systems - proper folding, post-translational modifications, and relevant enzymatic activity. It may also be more desirable than other eukaryotic non-mammalian systems whereby the proteins expressed may not contain the correct glycosylations. It is of particular use in producing membrane-associating proteins that require chaperones for proper folding and stability as well as containing numerous post-translational modifications. The downside, however, is the low yield of product in comparison to prokaryotic vectors as well as the costly nature of the techniques involved. Its complicated technology, and potential contamination with animal viruses of mammalian cell expression have also placed a constraint its use in large-scale industrial production.

Cultured mammalian cell lines such as the Chinese hamster ovary (CHO), HEK, HeLa, and COS cell lines may be used to produce protein. Vectors are transfected into the cells and the DNA may be integrated into the genome by homologous recombination in the case of stable transfection, or the cells may be transiently transfected. Examples of mammalian expression vectors include the adenoviral vectors, the pSV and the pCMV series of plasmid

vectors, vaccinia andretroviral vectors, as well as baculovirus. The promoters for cytomegalovirus (CMV) and SV40 are commonly used in mammalian expression vectors to drive protein expression. Non-viral promoter, such as the elongation factor (EF)-1 promoter, is also known.

**Ti-plasmid**

The virulent strains of *A. tumefaciens* harbor large plasmids (140–235 kbp) known as tumor-inducing (Ti) plasmid involving elements like T-DNA, vir region, origin of replication, region enabling conjugative transfer and o-cat region (required for catabolism of opines).




Ti-plasmid of *Agrobacterium*

**T-DNA**

It is a small, specific segment of the plasmid, about 24kb in size and found integrated in the plant nuclear DNA at random site. This DNA segment is flanked by right and left borders.

The functions of T-DNA genes are listed

| **Gene** | **Product** | **Function** |
|---|---|---|
| ***ocs*** | Octopine synthase | Opine synthesis |
| ***nos*** | Nopaline synthase | Opine synthesis |
| ***trns1 (iaaH, auxA)*** | Tryptophan-2-mono-oxygenase | Auxin synthesis |
| ***trns2 (iaaM, auxB)*** | Indoleacetamide hydrolase | Auxin synthesis |
| ***trnr (ipt, cyt)*** | Isopentyltransferase | Cytokinin synthesis |

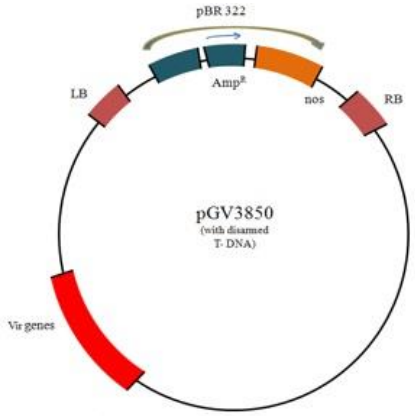
| | | |
|---|---|---|
| ***frs*** | Fructopine synthase | Opine synthesis |
| ***mas*** | Mannopine synthase | Opine synthesis |
| ***ags*** | Agropine synthase | Opine synthesis |

**T- DNA:Border Sequences**

- T-regions are defined by direct repeats known as T-DNA border sequences (Right and Left Border i.e. RB and LB of 25 bp each).
- These are not transferred intact to the plant genome, but are involved in the transfer process.
- The RB is rather precise, but the LB can vary by about 100 nucleotides.
- Deletion of the RB repeat abolishes T-DNA transfer, but the LB seems to be non-essential. The LB repeat has little transfer activity alone.

**Disarmed Ti-plasmid derivatives as plant vectors**

Ti plasmid is a natural vector for genetically engineering plant cells due to its ability to transfer T-DNA from the bacterium to the plant genome. But wild-type Ti plasmids are not suitable as vectors due to the presence of oncogenes in T-DNA that cause tumor growth in the recipient plant cells. For efficient plant regeneration, vectors with disarmed T-DNA are used by making it non-oncogenic by deleting all of its oncogenes. The foreign DNA is inserted between the RB and LB and then integrated into the plant genome without causing tumors.



Structure of the Ti-plasmid pGV3850 with disarmed T-DNA.

- The creation of disarmed T-DNA is an important step forward, but the absence of tumor formation makes it necessary to use an alternative method for the identification of transformed plant cells. Opine production using *pGV3850* was exploited as a screenable phenotype, and the *ocs* and *nos*genes are now widely used as screenable markers.

**Drawbacks**

Several drawbacks are associated with disarmed Ti- vector systems as discussed below;

- Necessity to carry out enzymatic assays on all potential transformants.
- Not convenient as experimental gene vectors due to large size.
- Difficulty in *in vitro* manipulation and
- Absence of unique restriction sites in the T-DNA.

**Co- integrate vectors**

Co-integrate vectors are the deletion derivatives of Ti-plasmids. The DNA to be introduced into the plant transformation vector is sub cloned in a conventional *Escherichia coli* plasmid vector for easy manipulation, producing a so-called *intermediate vector*. These vectors are incapable of replication in*A. tumefaciens* and also lack conjugation functions. Transfer is achieved using a 'triparental mating' in which three bacterial strains are mixed together:

(i) An *E. coli* strain carrying a helper plasmid able to mobilize the intermediate vector in *trans*;
(ii) *The E. coli* strain carrying the recombinant intermediate vector;
(iii) *A. tumefaciens* carrying the Ti plasmid.
Conjugation between the two *E. coli* strains transfers the helper plasmid to the carrier of the intermediate vector, which in turn is mobilized and transferred to the recipient *Agrobacterium*.Homologous recombination between the T-DNA sequences of the Ti plasmid and intermediate vector forms a large co- integrate plasmid resulting in the transfer of recombinant T-DNA to the plant genome.

**Binary vector**

- Binary vector was developed by Hoekma *et al* (1983) and Bevan in (1984).
- It utilizes the trans- acting functions of the vir genes of the Ti-plasmid and can act on any T-DNA sequence present in the same cell.
- Binary vector contains transfer apparatus(the vir genes) and the disarmed T-DNA containing the transgene on separate plasmids.

**Advantages of Binary vector**

- Small size due to the absence of border sequences needed to define T-DNA region and vir region.
- Ease of manipulation

## Marker genes

Marker systems are tools for studying the transfer of genes into an experimental organism. In gene transfer studies, a foreign gene, called a transgene, is introduced into an organism, in a process called transformation. A common problem for researchers is to determine quickly and easily if the target cells of the organism have actually taken up the transgene. A marker allows the researcher to determine whether the transgene has been transferred, where it is located, and when it is expressed. Marker genes exist in two broad categories:

## I. Selectable marker genes and II. Reporter genes.

## Selectable Marker Genes:

The selectable marker genes are usually an integral part of plant transformation system. They are present in the vector along with the target gene. In a majority of cases, the selection is based on the survival of the transformed cells when grown on a medium containing a toxic substance (antibiotic, herbicide, antimetabolite). This is due to the fact that the selectable marker gene confers resistance to toxicity in the transformed cells, while the non- transformed cells get killed.

A large number of selectable marker genes are available and they are grouped into three categories—antibiotic resistance genes, antimetabolite marker genes and herbicide resistance genes.

**(a) Antibiotic Resistance Genes**

In many plant transformation systems, antibiotic resistance genes (particularly of *E. coli*) are used as selectable markers. Despite the plants being eukaryotic in nature, antibiotics can effectively inhibit the protein biosynthesis in the cellular organelles, particularly in chloroplasts.

**Eg: Neomycin phosphotransferase II (npt II gene):**

The most widely used selectable marker is npt II gene encoding the enzyme neomycin phospho-transferase II (NPT II). This marker gene confers resistance to the antibiotic kanamycin. The trans-formants and the plants derived from them can be checked by applying kanamycin solution and the resistant progeny can be selected.

**(b) Antimetabolite Marker Genes**

**Eg: Dihydrofolate reductase (dhfr gene)**

The enzyme dihydrofolate reductase, produced by dhfr gene is inhibited by the antimetabolite methotrexate. A mutant dhfr gene in mouse that codes for this enzyme which has a low affinity to

methotrexate has been identified. This dhfr gene fused with CaMV promoter results in a methotrexate resistant marker which can be used for the selection of transformed plants.

**(c) Herbicide Resistance Markers**

**Eg: Enolpyruvylshikimate phosphate synthase (epsps/aroA genes):**

The herbicide glyphosate inhibits photosynthesis. It blocks the activity of enolpyruvylshikimate phosphate (EPSP) synthase, a key enzyme involved in the biosynthesis of phenylalanine, tyrosine and tryptophan. Mutant strains of *Agrobacterium* and *Petunia hybrida* that are resistant to glyphosate have been identified. The genes epsps/aroA confer resistance to transgenic plants which can be selected.

**Lists of selectable markers genes:**

**TABLE 49.3 A selected list of selectable marker genes used for gene transfer in plants, their source and substrates used for their selection**

| *Selectable marker gene (encoded enzyme)* | *Abbreviation* | *Source of gene* | *Substrate(s) used for selection* |
|---|---|---|---|
| **Antibiotic resistance** | | | |
| Neomycin phosphotransferase II | *nptII* | *E. coli* | Kanamycin, geneticin (G418) |
| Neomycin phosphotransferase III | *nptIII* | *Streptococcus faecalis* | Kanamycin, geneticin (G418) |
| Hygromycin phosphotransferase | *hpt/hyg* | *E. coli* | Hygromycin |
| Bleomycin resistance | *ble* | *E. coli* | Bleomycin |
| Aminoglycoside adenyltransferase | *aadA* | *Shigella flexneri* | Streptomycin, spectinomycin |
| **Antimetabolite markers** | | | |
| Dihydrofolate reductase | *dhfr* | Mouse | Methotrexate |
| Dihydropteroate synthase | *dhps/sul* | *E. coli* | Sulfonamides |
| **Herbicide resistance** | | | |
| Phosphinothricin acetyltransferase | *bar/pat* | *Streptomyces hygroscopicus/ S. viridochromogenes* | Glufosinate, L-phosphinothricin, Bialophos |
| Enolpyruvyl shikimate phosphate synthase | *epsps/aroA* | *Agrobacterium* sp/ *Petunia hybrida* | Glyphosate |
| Acetolactase synthase | *als* | *Arabidopsis* sp/maize/tobacco | Sulfonylureas |
| Glyphosate oxidoreductase | *gox* | *Achromobacter LBAA* | Glyphosate |
| Bromoxynil nitrilase | *bxn* | *Klebsiella pneumoniae* | Bromoxynil |
| **Others** | | | |
| β-Glucuronidase | *gus/uidA* | *E. coli* | Cytokinin glucuronide |
| Xylose isomerase | *xylA* | *Thermoanaerobcterium thermosulfurogenes* | Xylose |
| Mannose 6-phosphate isomerase | *pmi/manA* | *E. coli* | Mannose |
| Betaine aldehyde dehydrogenase | *badh* | Spinach | Betaine aldehyde |

## Reporter Genes:

A reporter gene may be regarded as the test gene whose expression can be quantified. The plant transformation can be assessed by the expression of reporter genes. In general, an assay for the reporter gene is carried out by estimating the quantity of the protein it produces or the final products formed. A selected list of the reporter genes along with the detection assays is given in Table and some of the important ones are discussed below.

**Table 49.4 A selected list of reporter genes used for gene transfer in plants, their sources, and detection assays**

| *Reporter gene (enzyme/protein encoded)* | *Abbreviation* | *Source of gene* | *Detection assay* |
|---|---|---|---|
| Octopine synthase | *ocs* | *Agrobacterium tumefaciens* | Electrophoresis, chromatography |
| Nopaline synthase | *nos* | *Agrobacterium tumefaciens* | Electrophoresis, chromatography |
| β-Glucuronidase | *gus/uidA* | *E. coli* | Fluorometric or histochemical or colorimetric |
| Green fluorescent protein | *gfp* | *Aequorea victoria* (jelly fish) | Fluorescence |
| Luciferase (bacterial) | *luxA/luxB* | *Vibrio harveyi* | Bioluminescence |
| Luciferase (firefly) | *luc* | *Photonus pyralis* | Bioluminescence |
| Chloramphenicol acetyltransferase | *cat* | *E. coli* | Autoradiography |

### β-Glucuronidase (GUS gene)

β-Glucuronidase producing gene (gus/uidA) is the most commonly used reporter gene in assessing plant transformation for the following reasons:

i. β-Glucuronidase assays are very sensitive.

ii. Quantitative estimation of the enzyme can be done by fluorometric method (using substrate 4-methylumbelliferryl P-D-glucuronide which is hydrolysed to 4-methylumbelliferone).

iii. Qualitative data on the enzyme can be obtained by histochemical means (enzyme localization can be detected by chromogenic substance such as substrate X-gluc).

iv. No need to extract and identify DNA.

### Green fluorescent protein (GFP gene)

Green fluorescent protein (GFP), coded by gfp gene, is being widely used in recent years. In fact, in many instances, GFP has replaced GUS since assays of GFP are easier and non-destructive. Thus, screening of even the primary transplants can be done by GFP which is not possible with other reporter genes.

Gene for GFP has been isolated from jelly fish *Aequorea victoria* which is a luminescent organism. The original gfp gene has been significantly modified to make it more useful as a reporter gene. GFP emits fluorescence which can be detected under a fluorescent microscope.

## Luciferase

### 1) Bacterial luciferase (luxA/luxB genes)

The bacterial luciferase genes (luxA and luxB) have originated from *Vibrio harveyi*. They can be detected in some plant transformation vectors. The detection assay of the enzyme is based on the principle of bioluminescence. Bacterial luciferase catalyses the oxidation of long-chain fatty aldehydes that results in the emission of light which can be measured.

### Firefly luciferase (luc gene)

The enzyme firefly luciferase, encoded by the gene luc, catalyses the oxidation of D-luciferin (ATP dependent) which results in the emission of light that can be detected by sensitive luminometers. The firefly luciferase gene, however, is not widely used as a marker gene since the assay of the enzyme is rather cumbersome.

# Restriction enzymes & DNA ligase

## Key points:

- **Restriction enzymes** are DNA-cutting enzymes. Each enzyme recognizes one or a few target sequences and cuts DNA at or near those sequences.
- Many restriction enzymes make staggered cuts, producing ends with single-stranded DNA overhangs. However, some produce blunt ends.
- **DNA ligase** is a DNA-joining enzyme. If two pieces of DNA have matching ends, ligase can link them to form a single, unbroken molecule of DNA.
- In DNA cloning, restriction enzymes and DNA ligase are used to insert genes and other pieces of DNA into plasmids.

## How do you cut and paste DNA?

In DNA cloning, researchers make many copies of a piece of DNA, such as a gene. In many cases, cloning involves inserting the gene into a piece of circular DNA called a **plasmid**, which can be copied in bacteria.

How can pieces of DNA from different sources (such as a human gene and a bacterial plasmid) be joined together to make a single DNA molecule? One common method is based on restriction enzymes and DNA ligase.

- A **restriction enzyme** is a DNA-cutting enzyme that recognizes specific sites in DNA. Many restriction enzymes make staggered cuts at or near their recognition sites, producing ends with a single-stranded overhang.
- If two DNA molecules have matching ends, they can be joined by the enzyme **DNA ligase**. DNA ligase seals the gap between the molecules, forming a single piece of DNA.

  Restriction enzymes and DNA ligase are often used to insert genes and other pieces of DNA into plasmids during DNA cloning.

## Restriction enzymes

Restriction enzymes are found in bacteria (and other prokaryotes). They recognize and bind to specific sequences of DNA, called **restriction sites**. Each restriction enzyme recognizes just one or a few restriction sites. When it finds its target sequence, a restriction enzyme will make a

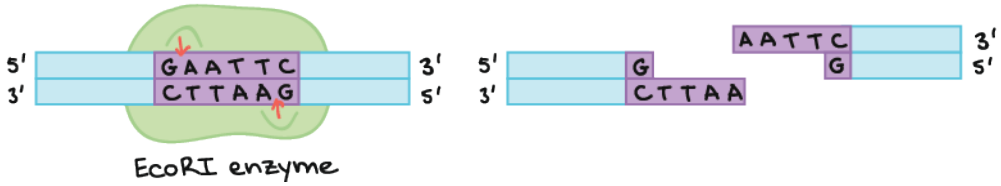
double-stranded cut in the DNA molecule. Typically, the cut is at or near the restriction site and occurs in a tidy, predictable pattern.
[Why do bacteria have restriction enzymes?]

As an example of how a restriction enzyme recognizes and cuts at a DNA sequence, let's consider *Eco*RI, a common restriction enzyme used in labs. *Eco*RI cuts at the following site:


5'-...GAATTC...-3' 3'-...CTTAAG...-5'

**EcoRI**site

When EcoRI recognizes and cuts this site, it always does so in a very specific pattern that produces ends with single-stranded DNA “overhangs”:



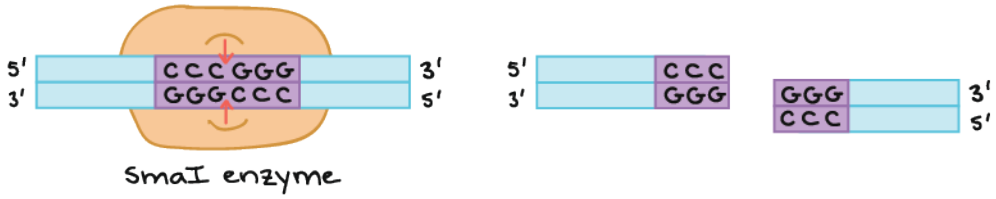

An *Eco*RI enzyme binds to an *Eco*RI site in a piece of DNA and makes a cut on both strands of the DNA. The pattern of the cut is:

5'-...G|AATTC...-3' 3'-...CTTAA|G...-5'

Thus, it produces an overhang of 5'-AATT-3' on each end of the cut DNA.

If another piece of DNA has matching overhangs (for instance, because it has also been cut by EcoRI), the overhangs can stick together by complementary base pairing. For this reason, enzymes that leave single-stranded overhangs are said to produce **sticky ends**. Sticky ends are helpful in cloning because they hold two pieces of DNA together so they can be linked by DNA ligase.

Not all restriction enzymes produce sticky ends. Some are “blunt cutters,” which cut straight down the middle of a target sequence and leave no overhang. The restriction enzyme *Sma*I is an example of a blunt cutter:




A *Sma*I enzyme binds to the *Sma*I restriction site, which is:

5'-...CCCGGG...-3' 3'-...GGGCCC...5'

It makes a cut right in the middle of this sequence on both strands, producing blunt ends. The cut sites are:

5'-...CCC|GGG...-3' 3'-...GGG|CCC...5'

Blunt-ended fragments can be joined to each other by DNA ligase. However, blunt-ended fragments are harder to ligate together (the ligation reaction is less efficient and more likely to fail) because there are no single-stranded overhangs to hold the DNA molecules in position.

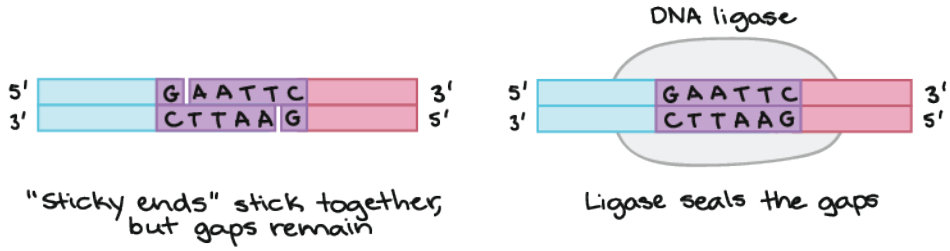
[Where do restriction enzymes get these weird names?]

3,3, comma000000

| | |
|---|---|
| 5′–GAATTC–3′3′–CTTAAG–3′ | 5′–G↓AATTC–3′3′–CTTAA↑G–5′ |
| 5′–GGATCC–3′3′–CCTAGG–3′ | 5′–G↓GATCC–3′3′–CCTAG↑G–5′ |
| 5′–AAGCTT–33′–TTCGAA–5′ | 5′–A↓AGCTT–33′–TTCGA↑A–5′ |
| 5′–CCCGGG–33′–GGGCCC–5′ | 5′–CCC↓GGG–33′–GGG↑CCC–5′ |
| 5′–GAGCTC–33′–CTCGAG–5′ | 5′–GAGCT↓C–33′–C↑TCGAG–5′ |

# DNA ligase

If you've learned about DNA replication, you may already have met DNA ligase. In DNA replication, ligase's job is to join together fragments of newly synthesized DNA to form a seamless strand. The ligases used in DNA cloning do basically the same thing. If two pieces of DNA have matching ends, DNA ligase can join them together to make an unbroken molecule.




Fragment 1 of DNA:

5'-...G 3'-...CTTAA

Fragment 2 of DNA:

AATTC...-3' G...-5'

The single-stranded regions of the two molecules can stick together by hydrogen bonding, but there are still gaps in the backbone:

5'-...G|AATTC...-3' 3'-...CTTAA|G...-5'

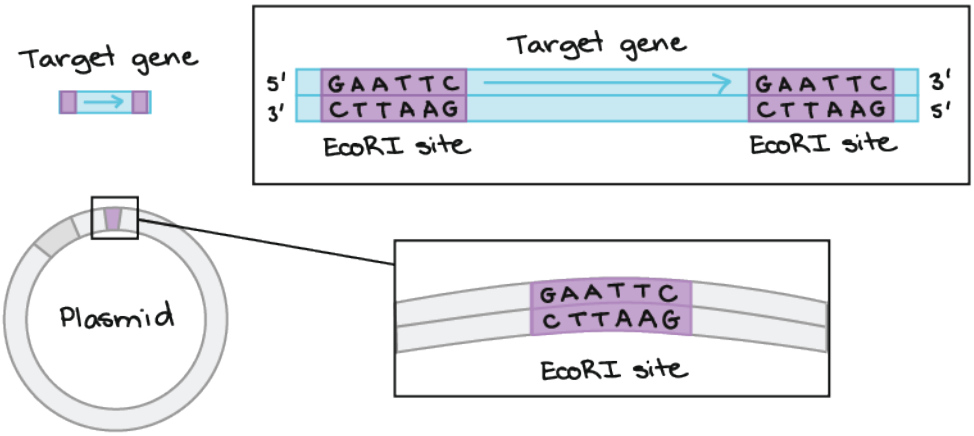
DNA ligase seals the gaps to make an unbroken molecule of DNA:

5'-...GAATTC...-3' 3'-...CTTAAG...-5'

How does DNA ligase do this? Using ATP as an energy source, ligase catalyzes a reaction in which the phosphate group sticking off the 5’ end of one DNA strand is linked to the hydroxyl group sticking off the 3’ end of the other. This reaction produces an intact sugar-phosphate backbone.

## Example: Building a recombinant plasmid

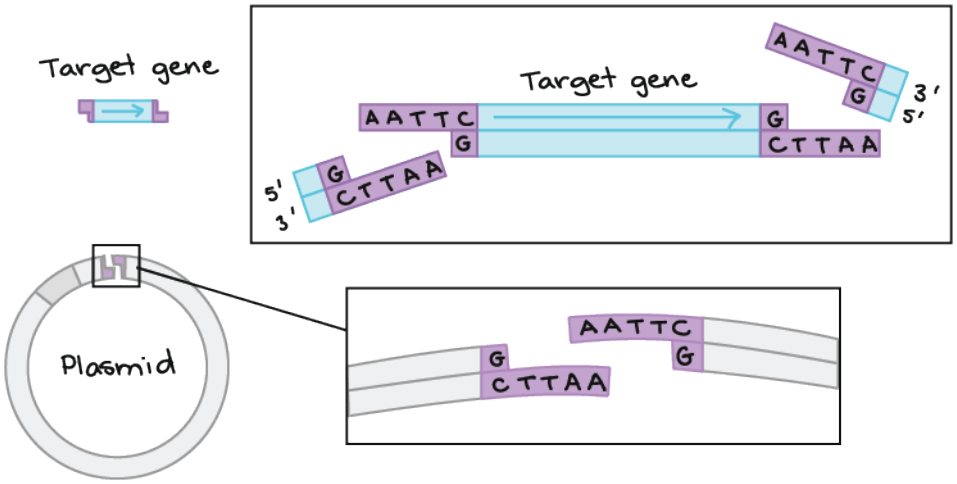
Let's see how restriction digestion and ligation can be used to insert a gene into a plasmid. Suppose we have a target gene, flanked with *Eco*RI recognition sites, and a plasmid, containing a single *Eco*RI site:




We start off with a target gene and a circular plasmid. The target gene has two *Eco*RI restriction sites near its ends. The plasmid has one *Eco*RI site in it, lying just after a promoter that drives expression in bacteria. The sequence of the *Eco*RI sites is:

5'-GAATTC-3' 3'-CTTAAG-5'

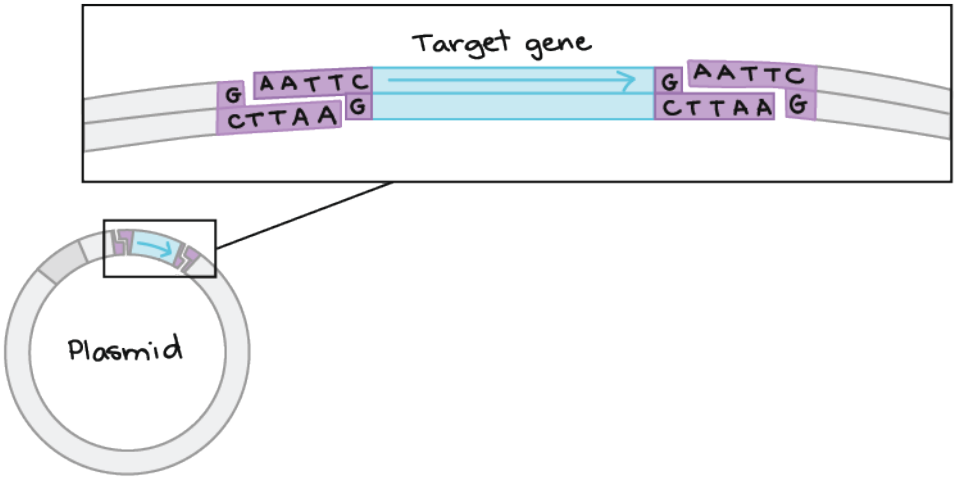
Our goal is to use the enzyme *Eco*RI to insert the gene into the plasmid. First, we separately digest (cut) the gene fragment and the plasmid with *Eco*RI. This step produces fragments with sticky ends:




We separately digest (cut) the gene fragment and the plasmid with *Eco*RI. This step produces fragments with sticky ends. All of the ends have an overhang of four nucleotides, with the sequence 5'-AATT-3'. That's because *Eco*RI's cut pattern is:

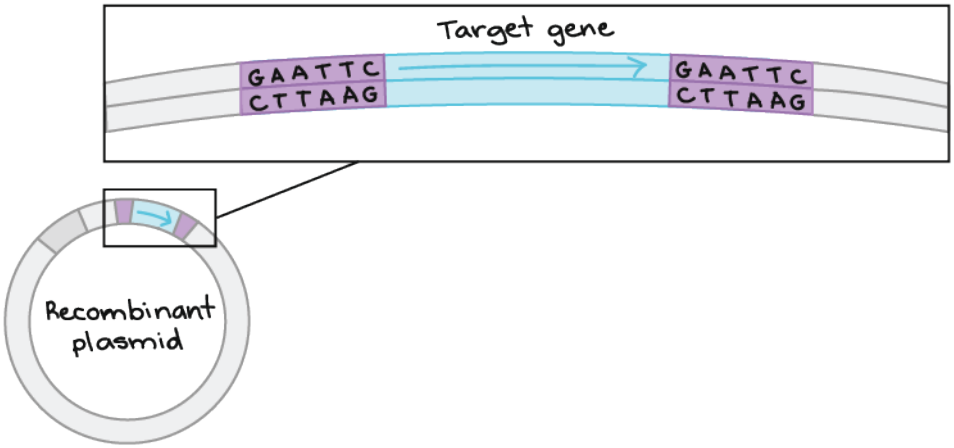
5'-G|AATTC-3' 3'-CTTAA|G-5'

Next, we take the gene fragment and the linearized (opened-up) plasmid and combine them along with DNA ligase. The sticky ends of the two fragments stick together by complementary base pairing:




Next, we take the gene fragment and the linearized (opened-up) plasmid and combine them along with DNA ligase. The sticky ends of the two fragments stick together by complementary base pairing. However, there are still gaps in the sugar-phosphate backbones of the DNA double helix at the junction sites where the gene and plasmid DNA meet.

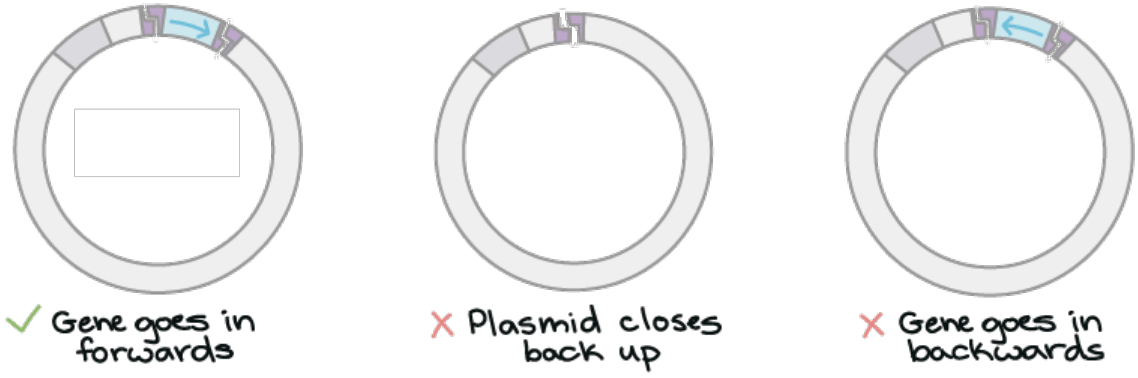
Once they are joined by ligase, the fragments become a single piece of unbroken DNA. The target gene has now been inserted into the plasmid, making a recombinant plasmid.




Once they are joined by ligase, the fragments become a single piece of unbroken DNA. The target gene has now been inserted into the plasmid, making a recombinant plasmid. In the plasmid, the gene is now flanked by two *Eco*RI sites that were generated when the cut ends were ligated together.

## Restriction digests and ligations involve many molecules of DNA

In the example above, we saw one outcome of a ligation between a gene and plasmid cut with *Eco*RI. However, other outcomes could happen in this exact same ligation. For instance, the cut plasmid could recircularize (close back up) without taking in the gene. Similarly, the gene could go into the plasmid, but flipped backwards (since its two *Eco*RI sticky ends are identical).




Left: recombinant plasmid produced when gene goes in forwards ("pointing" away from the promoter that is already in the plasmid).

Middle: non-recombinant plasmid produced when the cut plasmid simply closes back up (its ends ligate with each other).

Right: recombinant plasmid produced when gene goes in backwards ("pointing" back towards the promoter that is already in the plasmid).

Restriction digests and ligations like this one are performed using many copies of plasmid and gene DNA. In fact, billions of molecules of DNA are used in a single ligation! These molecules are all bumping into one another, and into DNA ligase, at random in different ways. So, if multiple products can be made, all of them *will* be made at some frequency – including ones we don't want.

How can we avoid the "bad" plasmids? When we transform bacteria with DNA from a ligation, each one takes up a different piece of DNA. We can check the bacteria after transformation and use only the ones with the correct plasmid. In many cases, plasmid from transformed bacteria is analyzed using another restriction digest to see if it contains the right insert in the right orientation.

## Marker genes

Marker systems are tools for studying the transfer of genes into an experimental organism. In gene transfer studies, a foreign gene, called a transgene, is introduced into an organism, in a process called transformation. A common problem for researchers is to determine quickly and easily if the target cells of the organism have actually taken up the transgene. A marker allows the researcher to determine whether the transgene has been transferred, where it is located, and when it is expressed. Marker genes exist in two broad categories:

## I. Selectable marker genes and II. Reporter genes.

## Selectable Marker Genes:

The selectable marker genes are usually an integral part of plant transformation system. They are present in the vector along with the target gene. In a majority of cases, the selection is based on the survival of the transformed cells when grown on a medium containing a toxic substance (antibiotic, herbicide, antimetabolite). This is due to the fact that the selectable marker gene confers resistance to toxicity in the transformed cells, while the non- transformed cells get killed.

A large number of selectable marker genes are available and they are grouped into three categories—antibiotic resistance genes, antimetabolite marker genes and herbicide resistance genes.

**(a) Antibiotic Resistance Genes**

In many plant transformation systems, antibiotic resistance genes (particularly of *E. coli*) are used as selectable markers. Despite the plants being eukaryotic in nature, antibiotics can effectively inhibit the protein biosynthesis in the cellular organelles, particularly in chloroplasts.

**Eg: Neomycin phosphotransferase II (npt II gene):**

The most widely used selectable marker is npt II gene encoding the enzyme neomycin phospho-transferase II (NPT II). This marker gene confers resistance to the antibiotic kanamycin. The trans-formants and the plants derived from them can be checked by applying kanamycin solution and the resistant progeny can be selected.

**(b) Antimetabolite Marker Genes**

**Eg: Dihydrofolate reductase (dhfr gene)**

The enzyme dihydrofolate reductase, produced by dhfr gene is inhibited by the antimetabolite methotrexate. A mutant dhfr gene in mouse that codes for this enzyme which has a low affinity to

methotrexate has been identified. This dhfr gene fused with CaMV promoter results in a methotrexate resistant marker which can be used for the selection of transformed plants.

**(c) Herbicide Resistance Markers**

**Eg: Enolpyruvylshikimate phosphate synthase (epsps/aroA genes):**

The herbicide glyphosate inhibits photosynthesis. It blocks the activity of enolpyruvylshikimate phosphate (EPSP) synthase, a key enzyme involved in the biosynthesis of phenylalanine, tyrosine and tryptophan. Mutant strains of *Agrobacterium* and *Petunia hybrida* that are resistant to glyphosate have been identified. The genes epsps/aroA confer resistance to transgenic plants which can be selected.

**Lists of selectable markers genes:**

**TABLE 49.3 A selected list of selectable marker genes used for gene transfer in plants, their source and substrates used for their selection**

| *Selectable marker gene (encoded enzyme)* | *Abbreviation* | *Source of gene* | *Substrate(s) used for selection* |
|---|---|---|---|
| **Antibiotic resistance** | | | |
| Neomycin phosphotransferase II | *nptII* | *E. coli* | Kanamycin, geneticin (G418) |
| Neomycin phosphotransferase III | *nptIII* | *Streptococcus faecalis* | Kanamycin, geneticin (G418) |
| Hygromycin phosphotransferase | *hpt/hyg* | *E. coli* | Hygromycin |
| Bleomycin resistance | *ble* | *E. coli* | Bleomycin |
| Aminoglycoside adenyltransferase | *aadA* | *Shigella flexneri* | Streptomycin, spectinomycin |
| **Antimetabolite markers** | | | |
| Dihydrofolate reductase | *dhfr* | Mouse | Methotrexate |
| Dihydropteroate synthase | *dhps/sul* | *E. coli* | Sulfonamides |
| **Herbicide resistance** | | | |
| Phosphinothricin acetyltransferase | *bar/pat* | *Streptomyces hygroscopicus/ S. viridochromogenes* | Glufosinate, L-phosphinothricin, Bialophos |
| Enolpyruvyl shikimate phosphate synthase | *epsps/aroA* | *Agrobacterium* sp/ *Petunia hybrida* | Glyphosate |
| Acetolactase synthase | *als* | *Arabidopsis* sp/maize/tobacco | Sulfonylureas |
| Glyphosate oxidoreductase | *gox* | *Achromobacter LBAA* | Glyphosate |
| Bromoxynil nitrilase | *bxn* | *Klebsiella pneumoniae* | Bromoxynil |
| **Others** | | | |
| β-Glucuronidase | *gus/uidA* | *E. coli* | Cytokinin glucuronide |
| Xylose isomerase | *xylA* | *Thermoanaerobcterium thermosulfurogenes* | Xylose |
| Mannose 6-phosphate isomerase | *pmi/manA* | *E. coli* | Mannose |
| Betaine aldehyde dehydrogenase | *badh* | Spinach | Betaine aldehyde |

## Reporter Genes:

A reporter gene may be regarded as the test gene whose expression can be quantified. The plant transformation can be assessed by the expression of reporter genes. In general, an assay for the reporter gene is carried out by estimating the quantity of the protein it produces or the final products formed. A selected list of the reporter genes along with the detection assays is given in Table and some of the important ones are discussed below.

**Table 49.4 A selected list of reporter genes used for gene transfer in plants, their sources, and detection assays**

| *Reporter gene (enzyme/protein encoded)* | *Abbreviation* | *Source of gene* | *Detection assay* |
|---|---|---|---|
| Octopine synthase | *ocs* | *Agrobacterium tumefaciens* | Electrophoresis, chromatography |
| Nopaline synthase | *nos* | *Agrobacterium tumefaciens* | Electrophoresis, chromatography |
| β-Glucuronidase | *gus/uidA* | *E. coli* | Fluorometric or histochemical or colorimetric |
| Green fluorescent protein | *gfp* | *Aequorea victoria* (jelly fish) | Fluorescence |
| Luciferase (bacterial) | *luxA/luxB* | *Vibrio harveyi* | Bioluminescence |
| Luciferase (firefly) | *luc* | *Photonus pyralis* | Bioluminescence |
| Chloramphenicol acetyltransferase | *cat* | *E. coli* | Autoradiography |

### β-Glucuronidase (GUS gene)

β-Glucuronidase producing gene (gus/uidA) is the most commonly used reporter gene in assessing plant transformation for the following reasons:

i. β-Glucuronidase assays are very sensitive.

ii. Quantitative estimation of the enzyme can be done by fluorometric method (using substrate 4-methylumbelliferryl P-D-glucuronide which is hydrolysed to 4-methylumbelliferone).

iii. Qualitative data on the enzyme can be obtained by histochemical means (enzyme localization can be detected by chromogenic substance such as substrate X-gluc).

iv. No need to extract and identify DNA.

### Green fluorescent protein (GFP gene)

Green fluorescent protein (GFP), coded by gfp gene, is being widely used in recent years. In fact, in many instances, GFP has replaced GUS since assays of GFP are easier and non-destructive. Thus, screening of even the primary transplants can be done by GFP which is not possible with other reporter genes.

Gene for GFP has been isolated from jelly fish *Aequorea victoria* which is a luminescent organism. The original gfp gene has been significantly modified to make it more useful as a reporter gene. GFP emits fluorescence which can be detected under a fluorescent microscope.

## Luciferase

### 1) Bacterial luciferase (luxA/luxB genes)

The bacterial luciferase genes (luxA and luxB) have originated from *Vibrio harveyi*. They can be detected in some plant transformation vectors. The detection assay of the enzyme is based on the principle of bioluminescence. Bacterial luciferase catalyses the oxidation of long-chain fatty aldehydes that results in the emission of light which can be measured.

### Firefly luciferase (luc gene)

The enzyme firefly luciferase, encoded by the gene luc, catalyses the oxidation of D-luciferin (ATP dependent) which results in the emission of light that can be detected by sensitive luminometers. The firefly luciferase gene, however, is not widely used as a marker gene since the assay of the enzyme is rather cumbersome.

# Difference between Selectable Marker and Reporter Gene

Genetic engineering is a budding phenomenon in biotechnology. It primarily involves the transformation of cells with desired genes to gain better traits and quality in the offspring. Selectable markers and reporter genes are two such sequences that aid in the process of transformation. Let us discuss the two genes before differentiating between them.

## What is a Selectable Marker?

Selectable markers are gene sequences that are introduced into plasmids for differentiation between transformed and non-transformed cells.

For example, a plasmid has a neomycin resistant gene as the selectable marker. When the cell is transformed, the selectable marker will be incorporated into the cell's genome. Now when these cells are grown on a media containing neomycin antibiotic, only those cells that have been transformed will be able to grow, and the non-transformed cells will not.

## What is a Reporter Gene?

A reporter gene is a test gene whose expression can be quantified. It is found in the T-DNA of plasmids that are incorporated into a cell's genome. When a cell has been transformed, it is necessary to assess the expression of those genes.

Reporter genes are those sequences that can be assessed to determine the expression of transformed genes. A reporter gene assay is carried out by calculating the total amount of protein formed. They are usually fluorescent in nature and give visual cues for final estimation. E.g., luciferase, octopine synthase, green fluorescent proteins, etc.

## Selectable Marker vs Reporter Gene

| Selectable Marker | Reporter Gene |
|---|---|
| **Definition** | |
| Selectable markers are sequences that can be used to distinguish between transformed and non- | Reporter genes are test sequences whose expression can be quantified. |

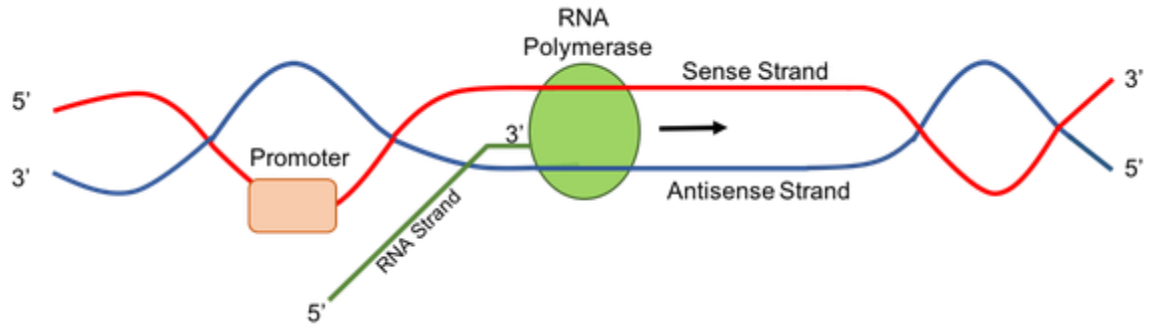
| | |
|---|---|
| transformed genes. | |
| **Function** | |
| They act as markers for transformed genes by providing resistance against toxic substances. | The amount of protein a reporter gene produces can be quantified to assess the transformation of a gene. |
| **Examples** | |
| Antibiotic resistance genes, herbicide resistance genes and antimetabolite marker genes. | Green fluorescent protein, luciferase, octopine synthase. |

# Promoters

## Definition

A promoter is a region of DNA where transcription of a gene is initiated. Promoters are a vital component of expression vectors because they control the binding of RNA polymerase to DNA. RNA polymerase transcribes DNA to mRNA which is ultimately translated into a functional protein. Thus the promoter region controls when and where in the organism your gene of interest is expressed.

## Summary



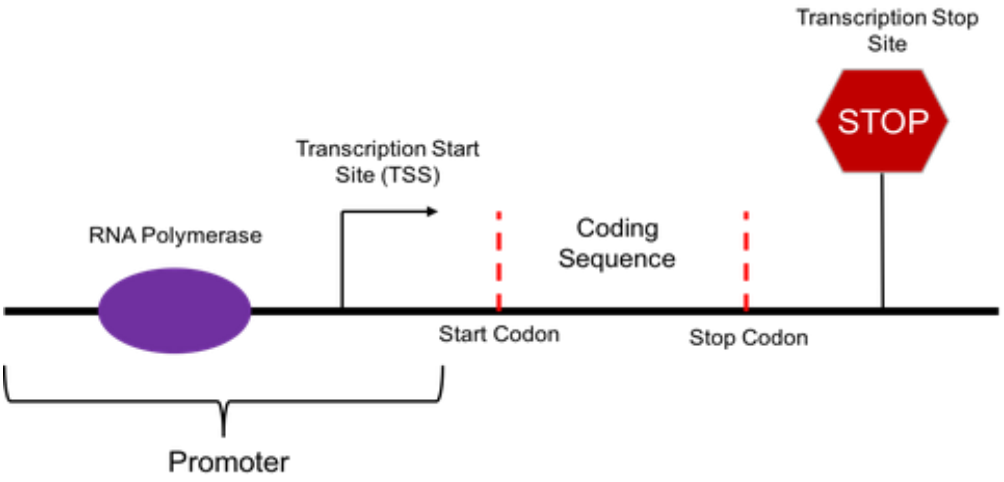

Promoters are about 100-1000 base pairs long and are adjacent and typically upstream (5') of the sense or coding strand of the transcribed gene. The coding strand is the DNA strand that encodes codons and whose sequence corresponds to the mRNA transcript produced. The antisense strand is referred to as the template strand or non-coding strand as this is the strand that is transcribed by the RNA polymerase.

DNA sequences called response elements are located within promoter regions, and they provide a stable binding site for RNA polymerase and transcription factors. Transcription factors are proteins which recruit RNA polymerase and control and regulate the transcription of DNA into mRNA.

Promoter binding is very different in bacteria compared to eukaryotes. In bacteria, the core RNA polymerase requires an associated sigma factor for promoter recognition and binding. On the other hand, the process in eukaryotes is much more complex. Eukaryotes require a minimum of seven transcription factors in order for RNA polymerase II (a eukaryote-specific RNA polymerase) to bind to a promoter. Transcription is tightly controlled in both bacteria and

eukaryotes. Promoters are controlled by various DNA regulatory sequences including enhancers, boundary elements, insulators, and silencers.

## Promoter Regions




There are three main portions that make up a promoter: core promoter, proximal promoter, and distal promoter. Below describes the specifics of these regions in eukaryotic cells.

### Core Promoter

The core promoter region is located most proximal to the start codon and contains the RNA polymerase binding site, TATA box, and transcription start site (TSS). RNA polymerase will bind to this core promoter region stably and transcription of the template strand can initiate. The TATA box is a DNA sequence (5'-TATAAA-3') within the core promoter region where general transcription factor proteins and histones can bind. Histones are proteins found in eukaryotic cells that package DNA into nucleosomes. Histone binding prevents the initiation of transcription whereas transcription factors promote the initiation of transcription. The most 3' portion (closest to the gene's start codon) of the core promoter is the TSS which is where transcription actually begins. Only eukaryotes and archaea, however, contain this TATA box. Most prokaryotes contain a sequence thought to be functionally equivalent called the Pribnow box which usually consists of the six nucleotides, TATAAT.

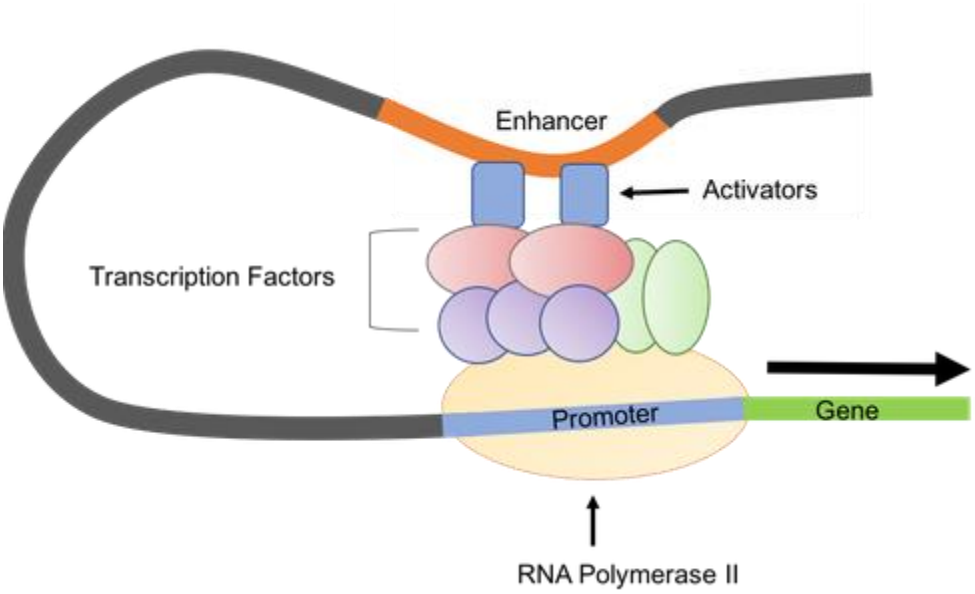
### Proximal Promoter

Further upstream from the core promoter you will find the proximal promoter which contains many primary regulatory elements. The proximal promoter is found approximately 250 base pairs upstream from the TSS and it is the site where general transcription factors bind.

### Distal Promoter

The final portion of the promoter region is called the distal promoter which is upstream of the proximal promoter. The distal promoter also contains transcription factor binding sites, but mostly contains regulatory elements.

# Eukaryotic Promoters




Eukaryotic Transcription

Eukaryotic promoters are much more complex and diverse than prokaryotic promoters. Eukaryotic promoters span a wide range of DNA sequences. It is not unusual to have several regulatory elements such as enhancers several kilobases away from the TSS. Eukaryotic promoters are so complex in structure that the DNA tends to fold back on itself which helps to explain how many physically distant DNA sequences can affect transcription of a given gene. The TATA-binding protein binds the TATA box and helps in the subsequent binding of the RNA polymerase. A transcription complex is constructed from the RNA polymerase and several transcription factor proteins.

### Common Eukaryotic Promoters Used in Research

| Promoter | Expression | Description |
|---|---|---|
| CMV | Constitutive | Strong mammalian promoter from human cytomegalovirus |
| EF1a | Constituitve | Strong mammalian promoter from human elongation factor 1 alpha |
| CAG | Constitutive | Strong hybrid mammalian promoter |

| Promoter | Expression | Description |
|---|---|---|
| PGK | Constitutive | Mammalian promoter from phospholycerate kinase gene |
| TRE | Inducible | Tetracycline response element promoter |
| U6 | Constitutive | Human U6 nuclear promoter for small RNA expression |
| UAS | Specific | Drosophila promoter containing Gal4 binding sites |

## Bacterial Promoters

Promoters in bacteria contain two short DNA sequences located at the -10 (10 bp 5' or upstream) and -35 positions from the transcription start site (TSS). Their equivalent to the eukaryotic TATA box, the Pribnow box (TATAAT) is located at the -10 position and is essential for transcription initiation. The -35 position, simply titled the -35 element, typically consists of the sequence TTGACA and this element controls the rate of transcription. Bacterial cells contain sigma factors which assist the RNA polymerase in binding to the promoter region. Each sigma factor recognizes different core promoter sequences.

### Operons

Although bacterial transcription is simpler than eukaryotic transcription bacteria still have complex systems of gene regulation, like operons. Operons are a cluster of different genes that are controlled by a single promoter and operator. Operons are common in prokayotes, specifically bacteria, but have also been discovered in eukaryotes. Operons consist of a promoter, which is recognized by the RNA polymerase, an operator, a segment of DNA in which a repressor or activator can bind, and the structural genes that are transcribed together.

Operon regulation can be either negative or positive. Negative repressible operons, are normally bound by a repressor protein that prevents transcription. When an inducer molecule binds to the repressor, it changes its conformation, preventing its binding to the operator and thus allowing for transcription. The Lac operon in bacteria is an example of a negatively controlled operon.

A positive repressible operon works in the opposite way. The operon is normally transcribed until a repressor/corepressor binds to the operator preventing transcription. The trp operon involved in the production of tryptophan is an example of a positively controlled operon.

### Common Bacterial Promoters used in Research

| Promoter | Expression | Description |
|---|---|---|
| T7 | Constitutive but requires T7 RNA polymerase | Promoter from T7 bacteriophage |
| Sp6 | Constitutive but requires Sp6 RNA polymerase | Promoter from Sp6 bacteriophage |
| lac | Constitutive in the absense of lac repressor (lacI or lacIq). Can be induced by IPTG or lactose | Promoter from Lac operon |
| araBad | Inducible by arabinose | Promoter of the arabinose metabolic operon |
| trp | Repressible by tryptophan | Promoter from *E. coli* tryptophan operon |
| Ptac | Regulated like the lac promoter | Hybrid promoter of lac and trp |

## Types of RNA Polymerases

Promoters control the binding of RNA polymerase to DNA to initiate the transcription of genes. There are three types of RNA polymerases that all transcribe different genes.

**RNA polymerase I** transcribes genes encoding ribosomal RNA (rRNA) which is a main component of a cell's ribosome structure. Ribosomes are the site of protein syntehsis where mRNA is translated into a protein.

**RNA polymerase II** transcribes messenger RNA (mRNA) which is the RNA responsible for providing a stable template for the translation of a protein.

**RNA polymerase III** transcribes genes encoding transfer RNAs (tRNA), the adaptor molecules that are responsible for bringing amino acids to the ribosome when proteins are being synthesized. RNA Polymerase III also transcribes small RNAs, such as shRNAs and gRNAs.