Summary-of-MB1 - Summary Molecular Biology 1

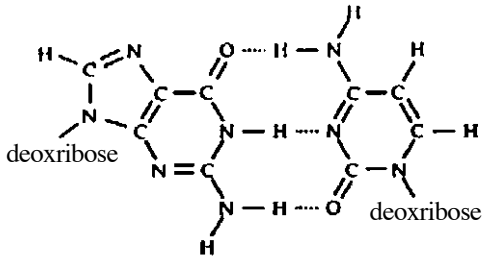
Molecular Biology 1 (University of Technology Sydney)

REVIEW OF NUCLEIC ACID STRUCTURE AND PROPERTIES

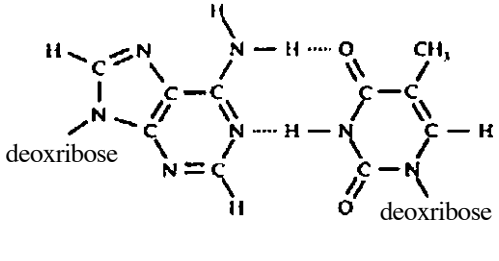
DOUBLE HELICAL STRUCTURE OF DNA

Complementarity of base-pairing (A - T and G - C)
Anti-parallel strands (5' - 3' on the top strand and 3' - 5' on the bottom strand)

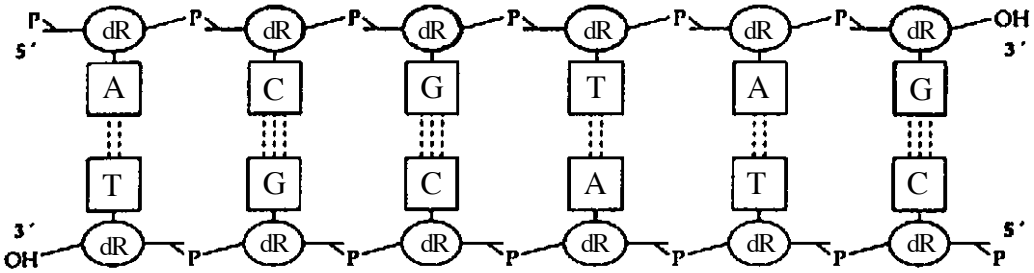
5'...AAGGCTTTTAGCTAGCTTTCGATCGATAGATG...3'
3'...TTCCGAAAATCGATCGAAAGCTAGCTATCTAC...5'

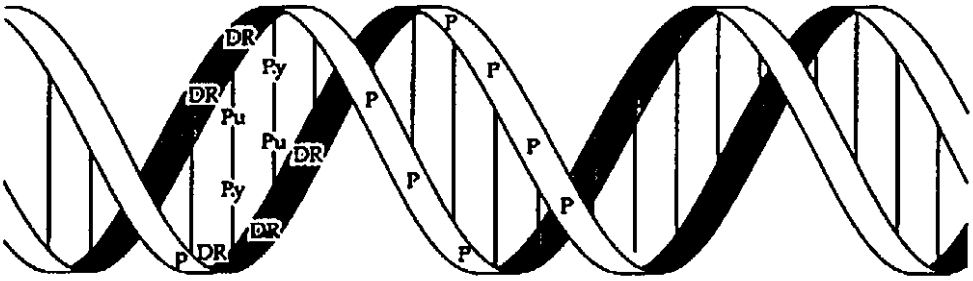


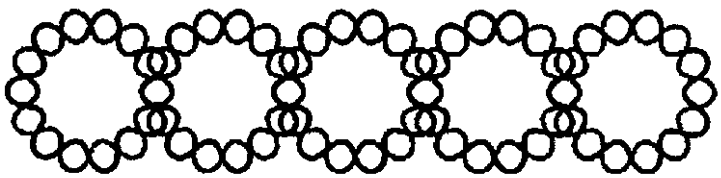
Guanine Cytosine



Adensoine Thymine





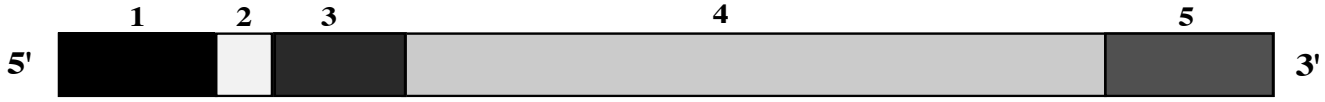


ELEMENTS OF GENE STRUCTURE

Important features:-

i. Collinear gene and polypeptide sequences.
ii. Intron / exon structure of eukaryotic genes; splicing of pre-mRNAs.
iii. Polycistronic prokaryotic genes (operons).
iv. Regulatory elements upstream from coding regions: promoters, SD sequences, transcriptional activators and terminators, repressors / inducers.

Elements of the Gene Unit



1. Upstream Promoter region. Contains -10 and -35 consensus regions for RNA polymerase binding.

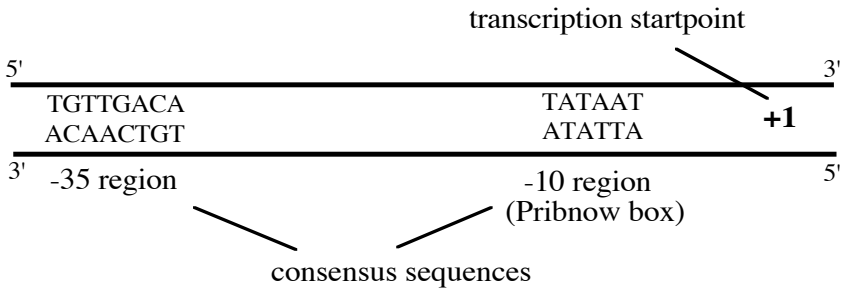
2. Ribosome Binding Site (RBS) or Shine-Dalgano (SD) sequence. Consensus is AGGA or some combination of several purines. The resulting mRNA transcript uses this sequence to bind to 3' end of the 16S large ribosomal subunit.

3. Leader or signal sequence. Not always present. Forms stem loop with promoter region to prevent RNA polymerase binding. Or, contains non-polar amino acid sequence for protein export through the membranes of cells.

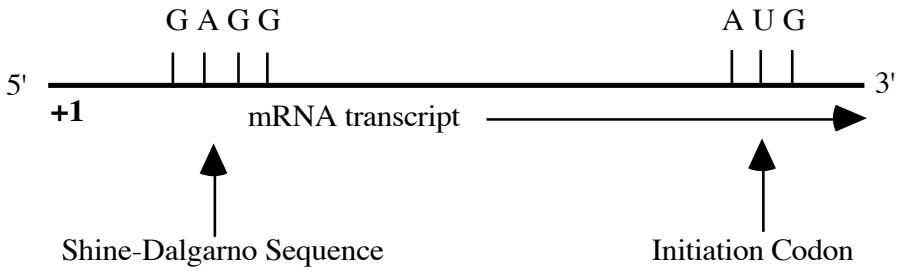
4. Open Reading Frame (ORF). Contains the triplet amino acid codons for the polypeptide sequence, starting with AUG (methionine) and finishing with one of the three stop codons (UAA, UGA, UAG).

5. Terminator region. Contains sequence that forms a stem loop structure for the detachment of RNA polymerase. Only present at the end of single genes or operons.

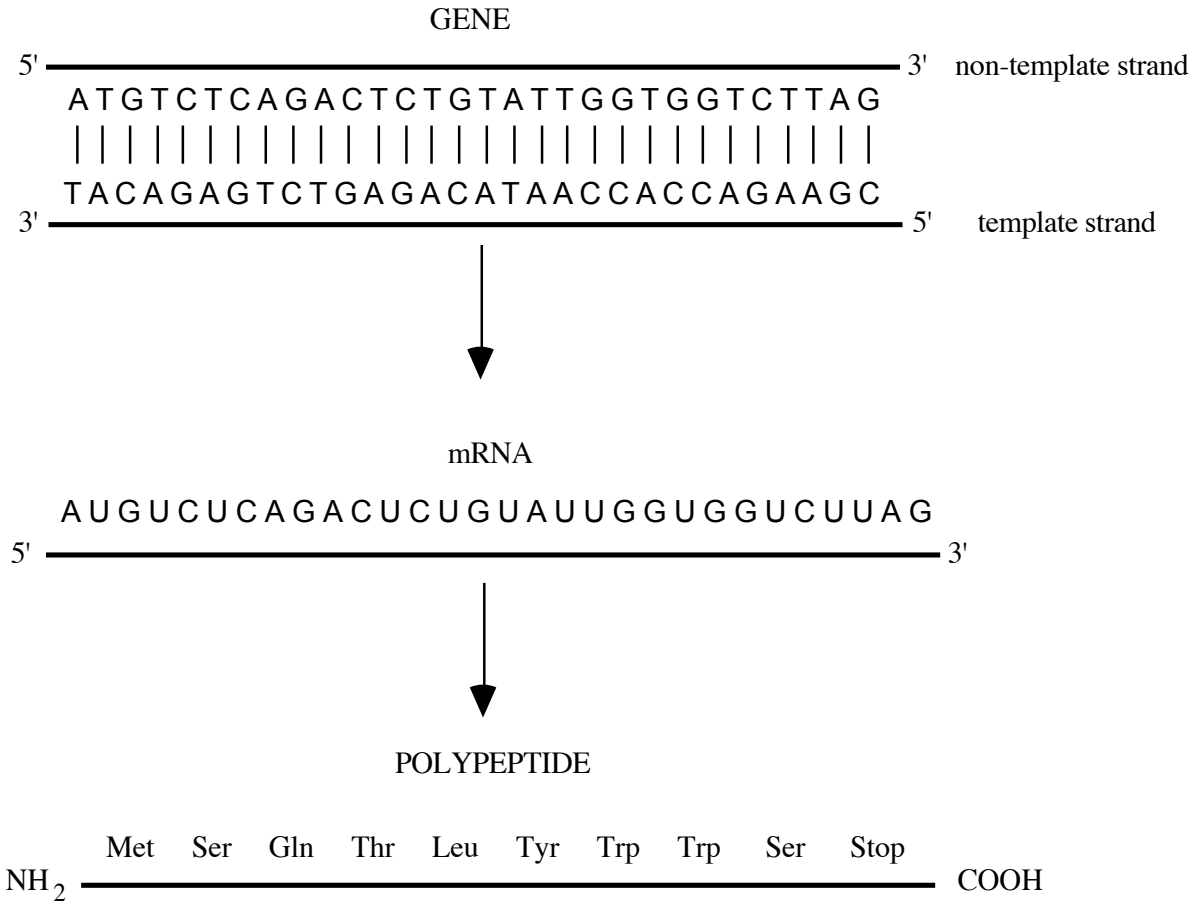
Promoter



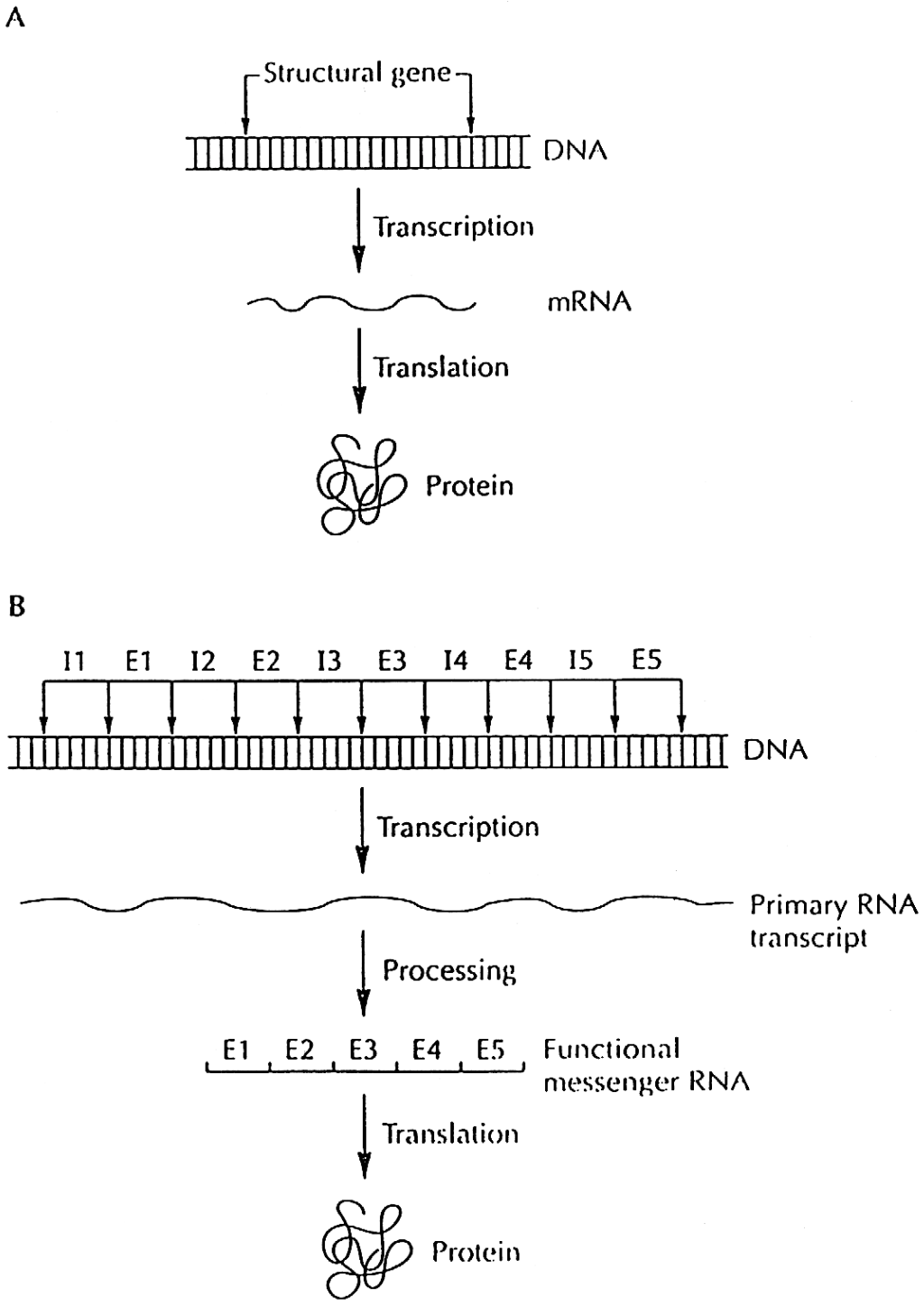
Ribosome Binding Site



Colinearity of Gene and Polypeptide



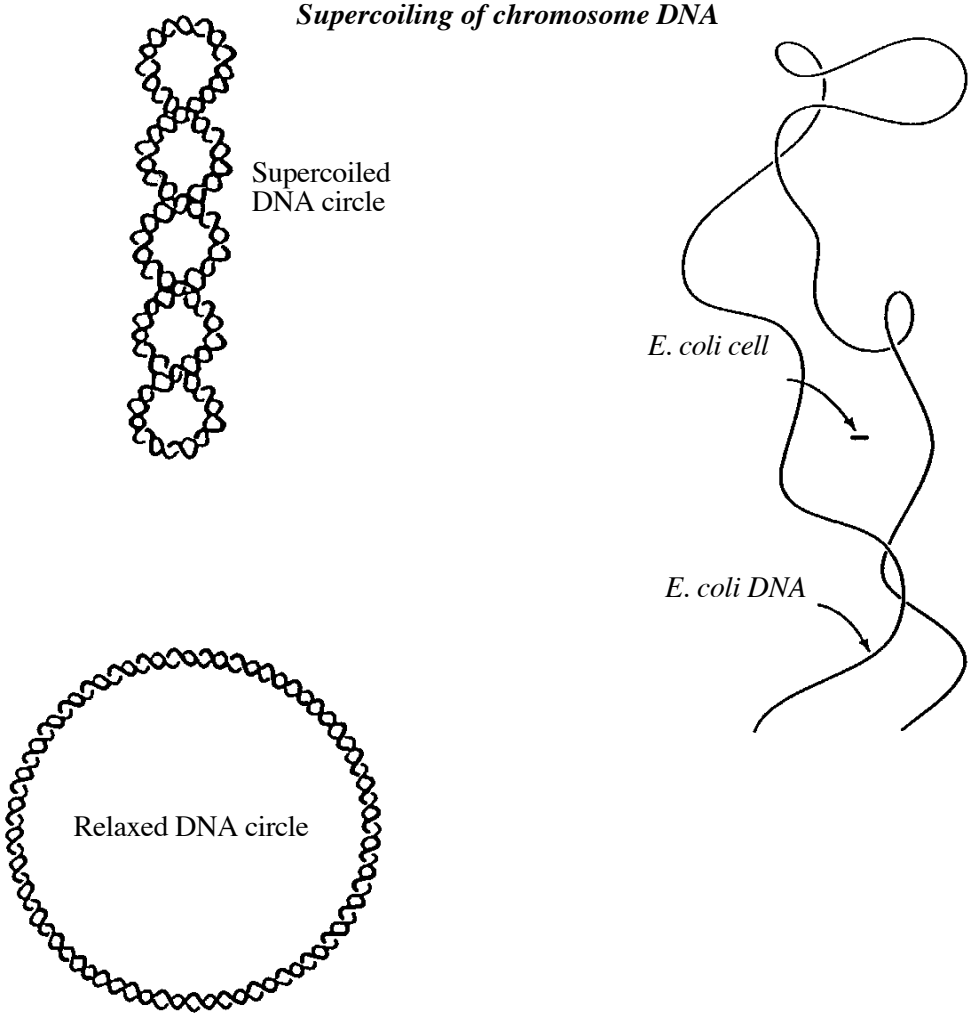
Structural genes, processing of mRNA transcripts and translation into proteins, in prokaryotes (A) and eukaryotes (B)



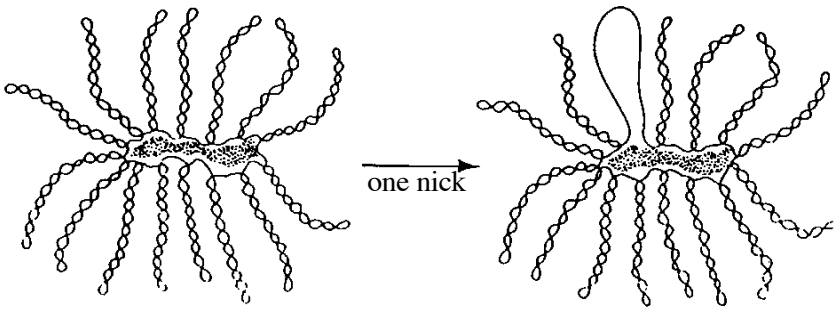
SUPERCOILED AND RELAXED DNA MOLECULES

- Plasmids and bacterial chromosomes are mostly supercoiled in their native, vegetative state.
- Relaxed, open circular structures occur transiently during transcription and replication.
- Supercoiling is a way for the cell to hold the amount of DNA required in its chromosome and also as resting stage between transcriptional or replication events. DNA topoisomerase (gyrases) carry out the reversible transition from supercoiled to relaxed states.

Supercoiling of chromosome DNA

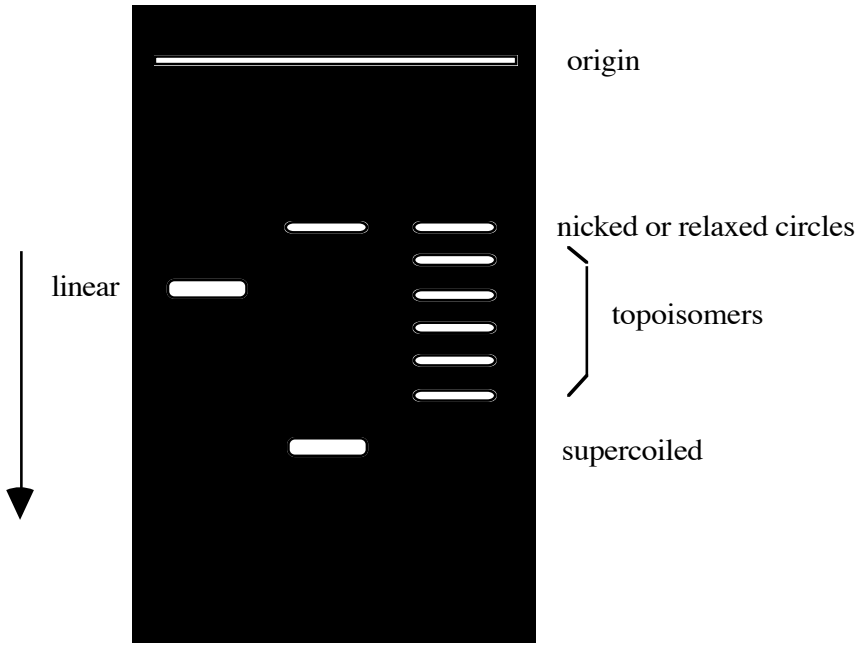


Schematic drawing of the highly folded and supercoiled E. coli chromosome, showing only 15 of the 46 loops (domains)



The gel profile below shows the relative migration positions of different DNA conformations including transitional structures formed by the activity of topoismerases.

DNA Conformations in Agarose Gels



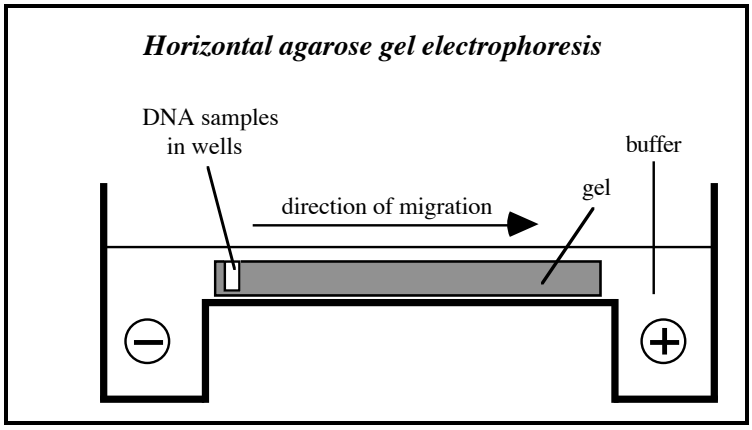
AGAROSE GEL ELECTROPHORESIS

Agarose is a purified component of agar but behaves like agar in melting at 100°C and gelling at about 40°C. Its suitability as an electrophoresis medium is due to its low electrolyte content (inertness) and its tendency to form gels with a matrix of homogeneously defined pore sizes depending on the concentration of agarose.

Agarose gels are cast into trays for horizontal submarine gel electrophoresis. A well-forming toothed comb is inserted into the poured, molten agarose at one end of the tray; and is removed form the hardened gel to expose the sample wells. The gel is placed into a Perspex box filled with buffer (Tris-acetate-EDTA or Tris-borate-EDTA) so that the gel surface and sample wells are submerged (this helps to dissipate heat during a run).

DNA or RNA samples added to the wells will migrate from the cathode (-) to the anode (+) end of the gel because all molecules are negatively charged at the buffer pH of about 7.6.

The DNA samples are combined with a small volume of 10 times sample buffer which contains bromophenol blue (BPB) as a dye front that runs ahead of most nucleic acid molecules, and a high density component (sucrose or glycerol) that enables the samples to drop into and remain within the wells during sample loading.



GEL CONCENTRATION vs RESOLUTION RANGE

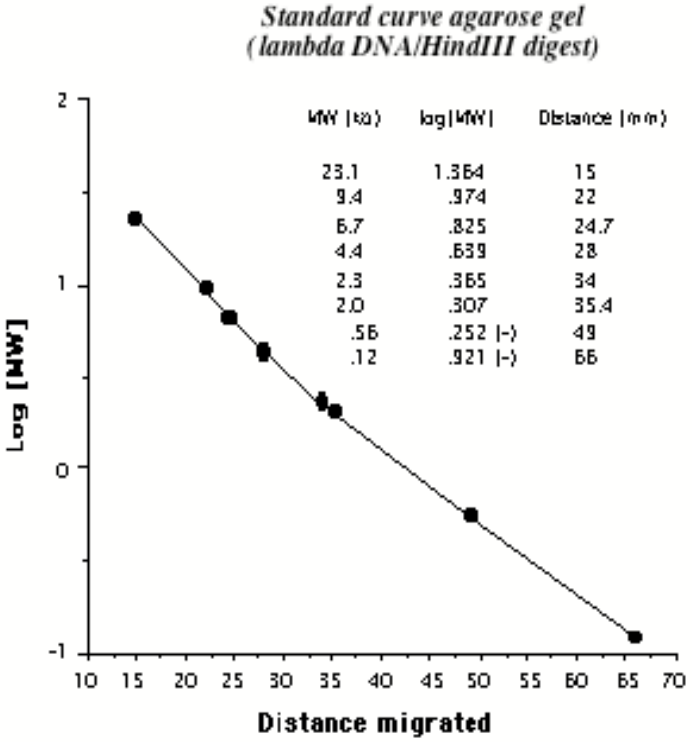
Separation Characteristics for Agarose and Polyacrylamide Gels

Gel type	Separation range (base pairs)
0.3% agarose	50,000 to 1000
0.7% agarose	20,000 to 300
1.4% agarose	6,000 to 300
4.0% acrylamide	1,000 to 100
10.0% acrylamide	500 to 25
20.0% acrylamide	50 to 1

Most DNA and RNA molecules are run and resolved in agarose gels of an appropriate concentration. Polyacrylamide gels are used for very small DNA fragments only or for DNA sequencing gels.

ESTIMATION OF FRAGMENT SIZE

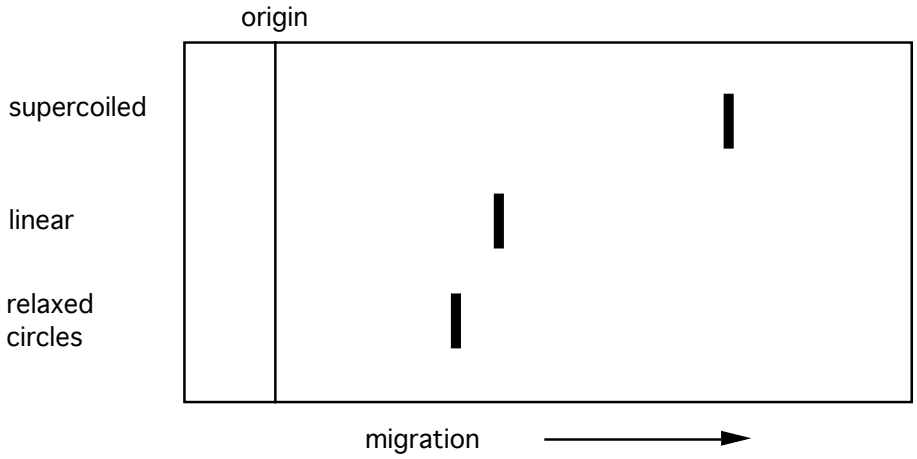
The migration rates of DNA molecules of the same conformation in agarose gels are proportional to the $\log_{10}$ of their molecular sizes.



CONFORMATION VERSUS MIGRATION RATE

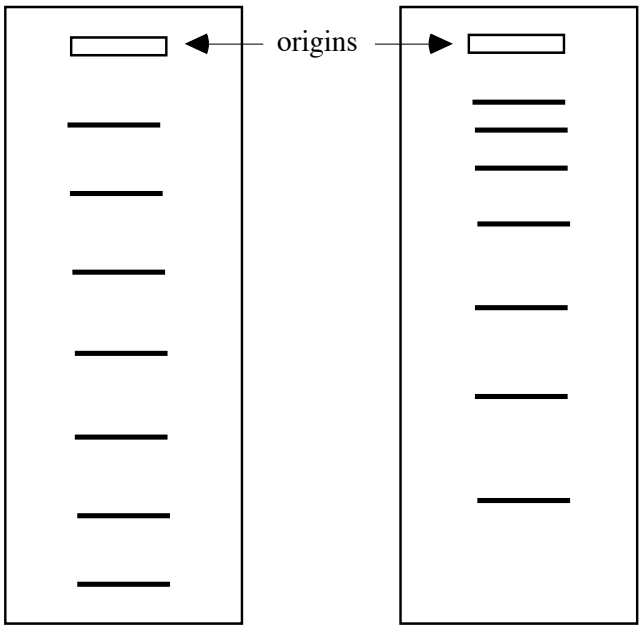
The migration rate of DNA molecules of the same conformation is governed by the expression $\log_{10}$MW versus distance migrated. However, if three different conformations of exactly the same DNA molecule are run in an agarose gel, their migration rates will differ according to the tightness or floppiness of their different conformations, even though their molecular compositions are identical. See profile below of linear, relaxed and supercoiled plasmid DNA.

Different migration rates of the same molecules as different conformations



In the gel profile diagram above, note that in practice, the linear and relaxed bands migrate close to one another. Also, but not shown in the profile, when plasmid DNA is isolated from cell lysates, only two bands of plasmid DNA are seen - supercoiled (approx. 80%) and relaxed circular (approx. 20%). Why?

The two gel traces below are of the same set of linear standards. Can you explain the reason for the different distribution of positions of the bands in the two gels?



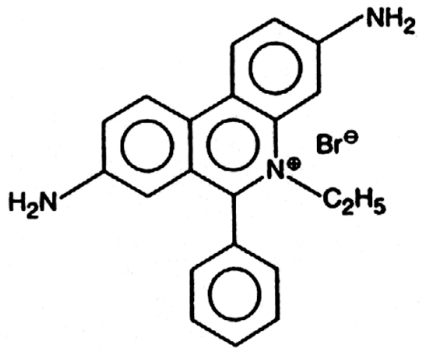
RECOVERY OF DNA FRAGMENTS FROM AGAROSE GELS

Many molecular biology cloning and sequencing strategies rely on the recovery of specific DNA fragments from agarose gels. There are a variety of recovery methods:

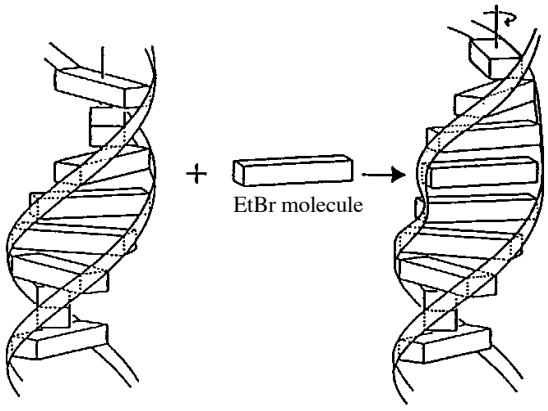
- Electroelution into wells cut out of the agarose ahead of a migrating band, or onto DEAE-cellulose paper strips inserted just ahead of a DNA band.
- Ethanol precipitation of DNA recovered from solubilised low melting temperature agarose slices or phenol extraction of normal agarose.
- Dissolution of agarose slices with a combination of heat (55^{o}C) and sodium iodide followed by binding of the DNA molecules to silica particles. After washing to remove loosely bound impurities, the DNA is eluted with a small volume of heated water or TE buffer.
- High-speed centrifugation of macerated gel slices.

ETHIDIUM BROMIDE UV FLUORESCENCE DETECTION OF NUCLEIC ACIDS

- Ethidium bromide (EtBr) is a planar molecule that intercalates DNA (that is, stacks between the base pairs). It absorbs ultraviolet light strongly with a maximum at 302 nm, and fluoresces by emitting visible orange light at 590 nm.
- Larger DNA fragment bands in agarose gels appear brighter than smaller fragments because they bind more of the EtBr dye.
- Supercoiled circular DNA molecules bind less EtBr than relaxed circular or linear DNA molecules because it is more difficult for the dye to invade supercoiled DNA. But, note well, isolated plasmids have much brighter supercoiled bands than relaxed bands because there is much more of the former, though it binds less EtBr per molecule.
- When EtBr binds to DNA of any conformation, but especially supercoiled molecules, the DNA become more relaxed, and therefore migrates at a slower rate in agarose gels, or bands at a lower buoyant density (higher in the gradient) in CsCl ultracentrifugation.
- Ethidium bromide staining of DNA can be useful in discriminating DNA doublet fragments that migrate together due to their nearly equivalent molecular sizes. The recognition of a doublet band is made on the increased brightness of EtBr staining when compared to other bands nearby.



Ethidium Bromide



Effect of EtBr binding on DNA conformation

METHODS FOR THE EXTRACTION AND PURIFICATION OF NUCLEIC ACIDS

The three main classes of nucleic acids that can be extracted and purified from cells or tissues of bacteria, other microorganisms, plants and animals are:

1. **Chromosomal (genomic) DNA**
2. **Plasmid DNA**
3. **RNA (total RNA = mRNA, tRNA, and rRNA)**

The choice of method depends upon the physical properties of the particular nucleic acid (size and conformation), and the behaviour of the nucleic acids under different extracting and fractionating conditions. Plasmid DNA is mostly isolated using an alkaline lysis procedure. Chromosomal DNA is extracted under mild lysis conditions. RNA is a special case in that mRNA can only be separated from total RNA from eukaryotic cells or tissues.

Following the precipitation of nucleic acids from cell lysates, further purification can be made by a second precipitation, by dialysis, or by CsCl/EtBr density gradient ultracentrifugation. If pure DNA is required, RNA can be removed by RibonucleaseA digestion. If RNA (or mRNA) is required, it can be isolated by ultracentrifugation which takes account of the greater density of RNA, but in practice the isolation of RNA is a special case that needs to take account of the susceptibility of RNA to degradation by ribonucleases (see further on). Calculation of the concentration of DNA or RNA is meaningless until the particular nucleic acid has been purified.

There are a number of commercial purification kits now available that facilitate the purification of nucleic acids from cell lysates. They owe their suitability to the use of a binding matrix (example, hydroxylated silica particles for DNA) that selectively binds nucleic acids from cell lysates. After washing to remove contaminants, the bound DNA is eluted in water or Tris-EDTA (TE) buffer.

ISOLATION OF CHROMOSOMAL DNA

In molecular biology, chromosomal DNA is mostly used for cloning or southern hybridisation which use restriction enzymes to firstly digest the DNA; or PCR (Polymerase Chain Reaction) amplification which only requires the DNA as a template for the attachment of oligonucleotide primers to the heat denatured strands. All of these applications are satisfactorily performed with semi-purified chromosomal DNA. A detailed isolation procedure for bacterial chromosomal DNA for such purposes will be performed in Experiment 2 of the Practical Course.

Standard extraction procedures use mild, neutral lysis (for example, Tris/EDTA/SDS or CTAB/salt based buffers) to release the high MW Chr DNA from cells. After removing the membrane, polysaccharide and protein debris by centrifugation, the Chr DNA is purified by one of the four methods below.

- Dialysis: Used only after phenol/chloroform extractions and RNase treatment. If the concentration of Chr DNA is too low, it may be concentrated by repeated butanol extractions (reduces the volume of aqueous phase).

- Precipitation in salt/ethanol: Precipitated high molecular weight genomic DNA is particularly difficult to solubilise, requiring at least one hour at 50°C.
- Column matrix binding: Lysates are passed through a column that binds the Chr DNA. Extraneous impurities are washed through the column before eluting the Chr DNA.
- Ultracentrifugation: EtBr is removed by extraction with isoamyl alcohol or butanol. CsCl is removed by dialysis or precipitation with salt/ethanol.

PROPERTIES OF PLASMIDS (VECTORS) USED IN MOLECULAR CLONING

Plasmids are circular self-replicating molecules that have been engineered to make them ideal vehicles for the cloning of small DNA fragments. The cloned DNA can be propagated, amplified and analysed in a bacterial host. Cloned genes can be expressed and protein products isolated.

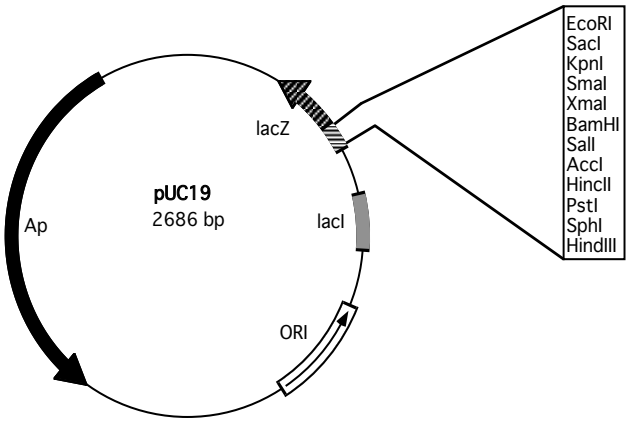
Large plasmids are naturally occurring in bacteria, yeasts, fungi and plants. Properties include:

- one or two copies per cell.
- stringent replication control.
- size range of 10 to 400 kbps.

Small plasmids (vectors) have been engineered from large plasmids with a view to providing ideal cloning properties, which are:

- 10 to 50 copies per cell.
- size range of 2 to about 8 kbps.
- relaxed replication leading to higher copy number (up to 1000 copies per cell after amplification).
- one or more selectable markers (antibiotic resistance) for the detection of transformants, and for maintenance in bacterial hosts.
- unique restriction sites in non-essential genes.
- detection methods for the insertion of cloned fragments (e.g. positive selection by an expressed property of the insert or indirect detection methods for recombinant clones).

A common cloning vector, pUC19, with the main features shown



This section is concerned only with the features of small vector plasmids that make them ideal for molecular cloning. The strategies of cloning will be dealt with in a later lecture following the basic instructional knowledge of the tools of molecular cloning, namely restriction endonuclease digestion, DNA ligation and DNA transformation into competent bacterial cells and the subsequent selection of recombinant clones.

ISOLATION OF PLASMID DNA

Many methods are available for the isolation of plasmid DNA. All involve three basic steps:

i. Growth of bacteria and amplification (optional) of plasmid DNA

ii. Harvesting and lysis of the bacterial culture

iii. Purification of the plasmid DNA

Isolation and purification methods for small plasmids exploit the two major differences between chromosomal and plasmid DNAs, namely,

i Chromosomes are much larger than plasmids and are susceptible to shear forces. Plasmid vectors are purposely constructed to be of small size.

ii Most of the bulk *E. coli* DNA extracted from lysed cells is obtained as broken, linear molecules, whereas plasmid DNA is largely recovered as supercoiled circles that resist breakage and denaturation during cell lysis.

A typical plasmid DNA isolation procedure involves the following steps:

1 Growth of the culture in medium plus antibiotic (to ensure that the cells that grow will all have the plasmid). Amplification with chloramphenicol or spectinomycin to reduce cell mass but increase plasmid yield.

Amplification is a process by which a great deal more plasmid DNA can be produced per unit cell. A protein synthesis inhibitor antibiotic is added at the mid-log phase of growth of the culture. The cells cease to divide, the chromosome is not replicated, the cell mass remains constant. However, the relaxed replication control of the plasmid that is not governed by the cell cycle allows the plasmid population to increase many-fold. In this way, an amplified culture (at about 4 x 10^8 cells per ml) can produce approximately 1000 copies of plasmid per cell versus about 50 to 100 copies per cell for an overnight, fully grown saturated culture (at about 4 x 10^9 cells per ml). Not a great deal more DNA, perhaps 1 to 3 times more, but a much reduced cell mass and therefore cleaner lysate.

2 Lysis of resuspended cell pellets in SDS/alkali. Viscosity increases; chromosomal DNA is irreversibly denatured at high pH. Plasmid DNA is transiently denatured but renatures when the lysate is neutralised because the two strands of each plasmid molecule are interlocked.

3 Addition of potassium acetate lowers the pH (denatured plasmid molecules regain their native conformation, but linearised, denatured high MW chromosomal DNA forms complexes with K/SDS/proteins, and is precipitated during centrifugation).

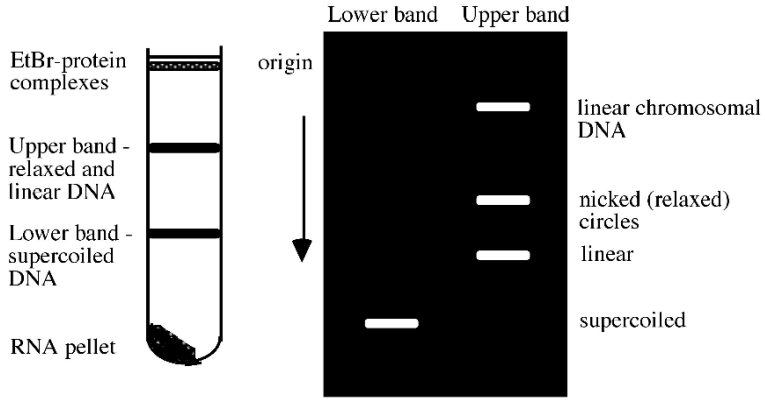
4 The supernatant (cleared cell lysate) is extracted with phenol and chloroform to remove residual proteins which form a flocculent precipitate at the aqueous/organic interface. The nucleic acids remain in solution in the aqueous layer.

5 Nucleic acids are precipitated with two volumes of ethanol in the presence of divalent cations. The centrifuged plasmid DNA pellet is resuspended in a small volume of buffer.

6 Alternatively, after step 4, caesium chloride (CsCl) and ethidium bromide (EtBr) are added and the sample is ultracentrifuged to equilibrium buoyant density. The plasmid band is removed, EtBr is extracted, and the plasmid DNA is dialysed or precipitated to remove CsCl.

A detailed plasmid isolation procedure will be performed in two of the practical experiments.

The distribution of DNAs in a CsCl gradient and of the same DNAs in an agarose gel



ISOLATION OF RNA

Single- and double-stranded structure of RNA.

- Messenger RNAs (mRNAs) are single-stranded.
- Ribosomal and transfer RNAs are partially single- and double-stranded.
- Viral RNA chromosomes are single-stranded (e.g. retroviruses) or double-stranded (e.g. tobacco mosaic viruses).

Prokaryotic.....total RNA is made up of,

- 23S (rRNA)
- 16S (rRNA)
- 5S (rRNA and tRNAs)
- mRNA (cannot be purified away from the total RNA)

Eukaryotic.....total RNA is made up of,

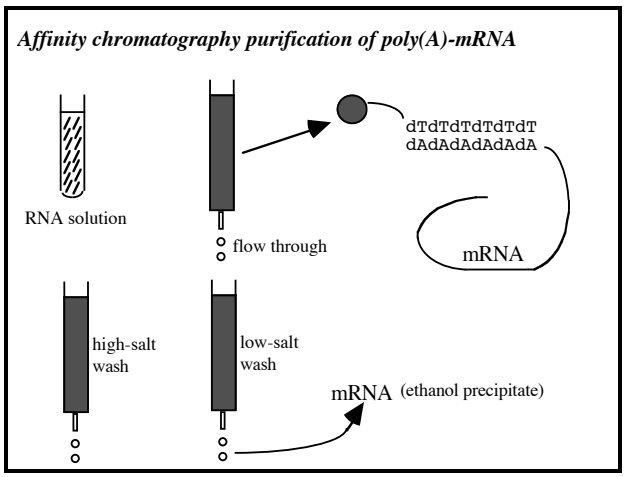
- 28S (rRNA)
- 18S (rRNA)
- 5S (rRNA and tRNAs)
- mRNA (poly A) (can be purified by dT-cellulose affinity chromatography)

The major problem in isolating RNA is degradation by ribonucleases. This is chiefly because RNA is essentially single-stranded and every ribonuclease nick in the phosphodiester backbone will result in the breakage of molecules (unlike double-stranded DNA in which a nick by deoxyribonuclease in one strand will not shorten or break the molecule.

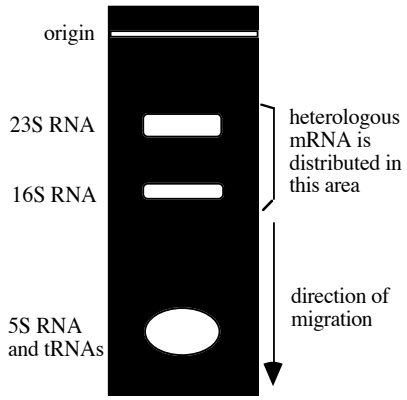
The sources of these ribonucleases and the methods of control (in parentheses) are as follows:

i. Released from cells at the same time as the total RNA following lysis with phenol-based buffers (guanidine thiocyanate in the lysis buffer is a powerful inhibitor)
ii. Airborne dust and bacteria (work in a laminar flow hood)
iii. The hands of workers (use gloves)
iv. Glassware and solutions (bake glassware at 200°C overnight; autoclave solutions in 0.1% diethylpyrocarbonate [DEPC], an inhibitor of ribonucleases).

Contaminating DNA can be digested with deoxyribonuclease I; or be extracted (along with proteins) in a selective phenol / chloroform / salt / guanidinium buffer; or can be eliminated by differential ultracentrifugation during which the RNA forms a pellet at the bottom of the tube and the DNA forms a band at a higher position (lower buoyant density) in the tube. The final RNA preparation is precipitated in 75% ethanol and can be held indefinitely at -70°C.



Analysis of RNA is performed in denaturing agarose gels containing formaldehyde or guanidinium, which assist and maintain the unravelling of the secondary structures in RNA, in order that the single-stranded molecules migrate according to size. An RNA isolation and gel isolation and electrophoresis procedure is detailed in Experiment 3 in the Practical Manual. A typical denaturing gel profile of total RNA is depicted below.



INTRODUCTION TO MICROBIAL GENETICS: REGULATION AND CONTROL : THE LAC OPERON

ESCHERICHIA COLI AND MOLECULAR BIOLOGY

Many types of hosts are used in molecular biology, including viruses (retroviruses, baculoviruses and bacteriophages) and bacteria (the commonest that are used are *Escherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium*). For eukaryotic work, yeasts (*Saccharomyces cerevisiae*), animals (mice, hamsters, rabbits and humans) and a diverse variety of cultured tissue cells are employed. Most experimentation, though, relies on *Escherichia coli* at some stage or other.

Bacteria divide by binary fission and can reach very large numbers in a short time. The doubling time for *E. coli* is 20 minutes in an aerated broth culture at 37^{o}C. When the culture is just visibly turbid - early log phase - there will be millions of cells per ml. At the end of the late log phase - several hours - the cell number per ml will be about 2 to 5 x 10^9. After late log phase, the culture enters stationary phase, where there is no further increase in cell numbers, due to the exhaustion of nutrients and build up of wastes. On semi-solid agar, a single cell produces a colony of 1 to 2 mm diameter after an overnight incubation. Each colony will contain many millions of cells.

Escherichia coli (*E. coli*) laboratory strains are derived from the parental K-12 strain created in the 1930s so as to be non-pathogenic to humans. *E. coli* has a circular, haploid chromosome of approximately 4000 kilobase pairs (kb or kbp). In comparison, the *S. cerevisiae* (yeast) chromosome is 20,000 kb, *D. melanogaster* (fruit fly) is 165,000 kb, *H. sapiens* (human) is 2,800,000 kb, amphibians are in the range of 950,000 to 10,000,000 kb, and plants from 95,000 to 120,000,000 kb. At the time of writing (July 1999), the chromosomes of nine bacterial species have been sequenced completely, including *E. coli.*

E. coli is also an ideal host for extrachromosomal elements (plasmids) that are supercoiled, circular, self-replicating DNA molecules. These can be large (10 to 200 kb) and transferable (conjugative) to other bacteria, or small and non-conjugative, the so-called genetically engineered plasmid vectors used in molecular cloning. These cloning vectors are about 2 to 6 kb in size, are isolated in large numbers from cell lysates, and may be transferred to other bacteria by the process of transformation.

The complete genome of an organism is all of the DNA with the chromosome(s) including the coding (genes) and non-coding sequences. The percentage of DNA that makes up the complement of genes can be high (in *E. coli*, > 60%) or low (in humans only about 10%).

GENOTYPES AND PHENOTYPES

The genotype comprises the wild-type gene complement of the strain; the phenotype is the visual manifestation of the genotype (lactose utilisation, antibiotic resistance etc.). Mutations are changes that give rise to defective (dysfunctional) genes or to newly acquired properties of the modified genes. The ways in which mutations are generated will be discussed later.

A summary of the nomenclature of genotypes and phenotypes is as follows:

- A mutant locus is given a three-letter, lower case, italicised abbreviation, irrespective of the species type (e.g. *lac* = lactose utilisation).
- A capital letter following the gene locus name refers to the individual mutant gene (e.g. *lacZ* = beta-galactosidase). The phenotype depends upon the particular status of the gene. In this case, the enzyme designation is LacZ (note, three letters, but not italicised and the first letter is in upper case). Beta-galactosidase hydrolyses lactose to glucose and galactose and only the wild-type locus is capable of doing so, that is, $lacZ^+$. Wild-type genes are seldom indicated.
- Specific mutated alleles are indicated by numbers (e.g. *lacZ*165).
- Antibiotic resistance loci are sometimes emphasised with a superscript 'r' or 'R' (e.g. kan^r = genotype; Kan^R or Km^R = phenotype; also, *bla* = gene for beta-lactamase, that is, ampicillin resistance; Amp^R or Ap^R = phenotype).
- Deletions are denoted by 'Δ', e.g. Δ(*gal-uvrB*) = deletion of the region from *gal* to *uvrB*.
- The presence of an extrachromosomal element such as a plasmid is denoted by a bracketed identification following the strain name, e.g. *E. coli* JM109(pUC19).

The genotypes of some laboratory K-12 strains are given in the table below. Only the relevant (mutant) properties are listed. All of the other genes are assumed to be wild-type (functional).

Genotypes of common E. coli strains used in molecular biology

Strain	Genotype
C-1A	wild-type
594	*rpsL*
DH5a	*supE44*, *lacZΔM15*, *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*
JM109	*recA1*, *supE44*, *endA1*, *hsdR17*, *gyrA96*, *relA1*, *thi*, *Δ*(*lac-proAB*), F'[*traD36*, $proAB^+$, $lacI^q$, *lacZΔM15*]
HB101	*supE44*, *hsdS20*, *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20*, *xyl-5*

MUTATION

Much of what we now know about microbial genetics comes, historically, from the study of mutations. With the present day advantages of molecular biology and its technological advances, we have been able to 'speed up' the types and numbers of DNA mutations.

In nature, mutations are rare events and this is necessarily so or else an organism would quickly lose its integrity and identity. Also, since most mutations are deleterious (harmful), their effects would soon compromise the functionality and even the viability of cells. Mutations are changes in the base composition of a cell's DNA. Most of these changes are single base alterations and are corrected by the 3'-5' exonuclease activity of DNA polymerases.

Mutations can arise spontaneously giving rise to *E. coli* variants at a frequency of about 1 in 10 million (10^{-7}). These mutations arise in the absence of any selective agent. That mutations are truly spontaneous can be proved by doing a Fluctuation Test that shows that mutations can arise at any time in the growth cycle of otherwise identical cell cultures (see table below).

Culture number (inoculum: 10^2 cells per ml of overnight culture)	**Number of cells resistant to streptomycin per 10^9 cells**
1	0
2	8
3	0
4	766
5	2

These results show that mutations may be absent in some cultures, or occur either early or late in the growth cycle. The very high number in culture number '4' is referred to as a 'jackpot'. Streptomycin inhibits protein synthesis by binding to the S12 protein on the 30S ribosome subunit and interfering in the mechanism of tRNA-mRNA binding. Streptomycin resistance, therefore, involves a change in the S12 protein that prevents the binding of the drug so that protein synthesis is unaffected.

Wild-type cells are known as prototrophs. Mutants that lack the ability to synthesis an essential metabolite are known as auxotrophs (e.g. inability to grow on glucose or lactose or to synthesise certain amino acids). Auxotrophs require the addition of the 'missing' metabolite to the growth medium. Conditional mutants are those that only express a mutant property under certain conditions, the best example being temperature-sensitive mutants. Spontaneous "back" mutations - known as reversions - also occur in cells at all times.

MUTAGENESIS

Mutagenesis is the process of generating and isolating mutants. The mutation rate can be increased by the use of mutagens. These agents include ionising radiation, ultraviolet light, growth at high temperatures and a host of chemical mutagens (e.g. ethyl methane sulphonate, nitrosoguanidine, acridine dyes). In fact, virtually any situation that induces stress in bacterial cells will give rise to mutations representing induced errors in the nucleotide composition, that are inherited by the succeeding generations, if repair processes do not regenerate the wild-type and if the mutation is not deleterious to the survival of the cell.

Molecular biology has now enabled a new, targeted method of generating mutants called site-directed mutagenesis (see later lecture). Whether induced spontaneously or generated by site-directed mutagenesis, DNA mutations are expressed mostly as single amino acid changes. Many of the amino acid changes are benign in that they will not alter the structure or function of a protein or enzyme, or might cause only very minor alterations. Others have more pronounced effects, with some single-base substitutions completely abrogating function. It is also possible to generate deletions or insertions of two to several codons.

Single-base substitutions can be classified as silent, missense, nonsense or frameshift.

- Transitions are substitutions of one purine for another or pyrimidine for another pyrimidine: AT <-----> GC or GC <-----> AT

- Transversions are changes from purine to pyrimidine or from pyrimidine to purine:
  AT <-----> TA or AT <-----> CG or GC <-----> TA or GC <-----> CG

If, for example, the third position of the codon GTT for the amino acid valine is changed to any of the other possible codons, GTA, GTC or GTG, the amino acid specified will still be valine and the amino acid sequence within the protein is unchanged. This is due to the degeneracy of the Genetic Code whereby the 20 amino acids are specified by 61 codons so that most of the amino acids have at least two codons and some up to six codons.

If, in our example, the valine codon GTT had the first base changed to C, becoming CTT, then leucine would be specified. But the structural similarity of valine and leucine (small, non-polar side chains) might cause little or no alteration in the protein structure or function. These are called silent mutations. On the other hand, if the second base of the valine codon was changed to A so that GTT became GAT, then the new amino acid will be aspartic acid, a very polar amino acid, and the change to the protein is likely to be more severe. This last example is called a missense substitution mutation.

Some changes can be extremely severe. This is because the Genetic Code has three stop codons, one or other of which is required at the normal terminus of a gene's coding frame for a protein. So, for example, if the codon for tryptophan, TGG, was changed in the third position to TGA, then this stop or nonsense codon would cause the premature termination of the protein sequence at wherever in the chain the tryptophan had been present. This is a called a nonsense mutation.

As well as base change substitutions, another form of sequence modification involves base insertions or deletions. These cause frameshifts. A frameshift is an alteration in the triplet codon sequence beyond the insertion or deletion point - unless the insertion/deletion is a multiple of three. This frameshift invariably produces a dysfunction polypeptide even if the frameshift is in the last amino acid codon of the protein, remember that the next codon - the stop codon - will also be changed and a stop codon will not be present in the right place. Moreover, frameshifts often introduce stop codons earlier in the reading frame, which results in truncated polypeptides.

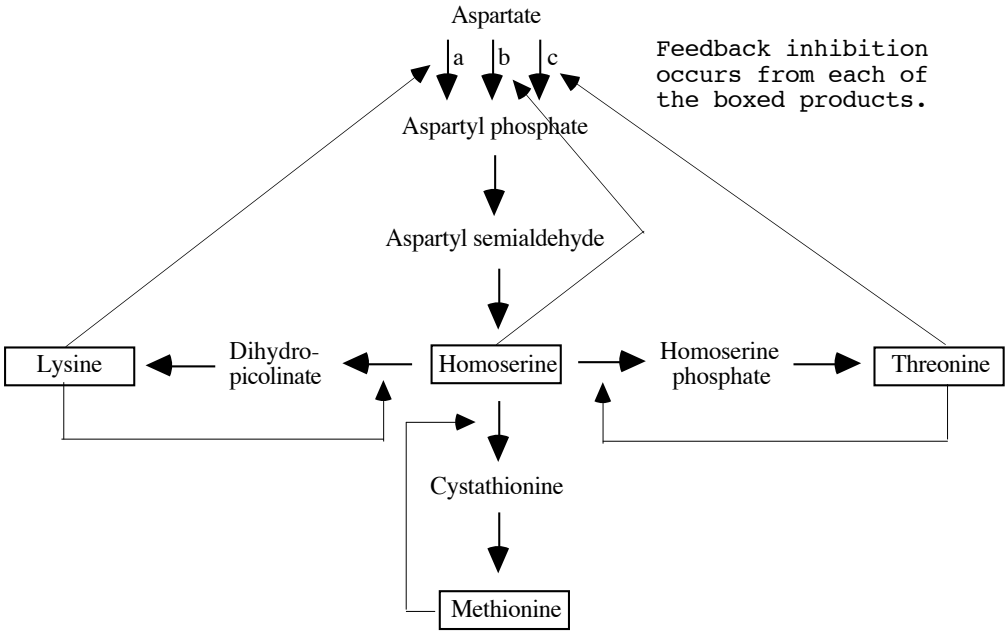
REGULATION AND CONTROL

A cluster of genes controlled from a single promoter and whose expression is regulated in some way is called an operon. This type of gene arrangement is typical of metabolic pathways in which several enzymes act in concert in the pathway.

Either all or none of the enzymes of an operon system are produced and this feature is termed coordinate regulation. The enzymes are either all produced or all not produced according to the conditions in the cell, and in particular, to the availability of certain metabolites.

When an operon has been working at high capacity to produce the enzymes to complete a metabolic process, the final product of the pathway often acts to "switch off" the process when enough of the final product is produced. This is termed feedback inhibition. Sensibly, feedback inhibition is exerted mostly on the first enzyme in the metabolic pathway.

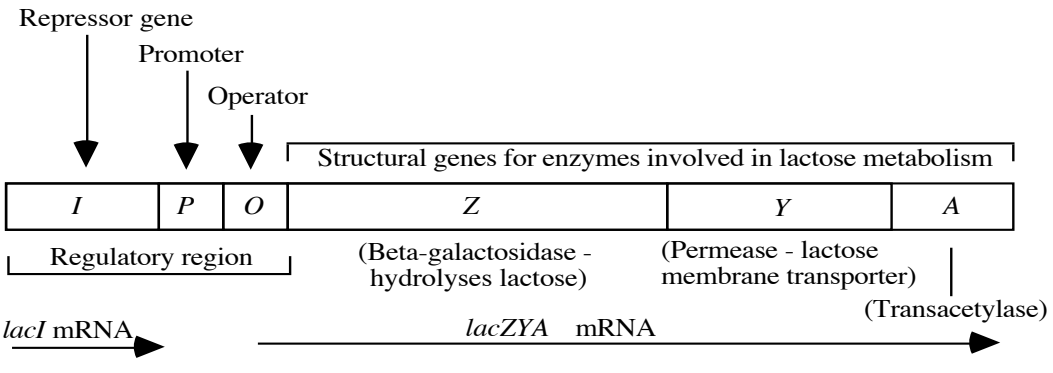
Feedback Inhibition



The single messenger RNA transcribed from an operon is called a polycistronic mRNA. All of the genes of an operon are regulated from a single promoter upstream from the first gene's Open Reading Frame (ORF). Operons are mostly regulated by repressors that prevent RNA polymerase from gaining access to the promoter region. This is termed negative regulation.

The most fully described example of an operon is the *lac* operon in *E. coli*. The structure and regulation of this operon is illustrated in the diagrams below. Note, the promoter and operator sequences are not drawn to scale and they are much shorter than the genes in the operon.

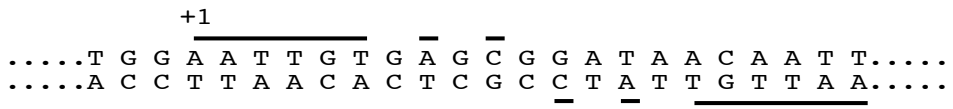
The lac Operon in E. coli



Repressor proteins are specified by repressor genes that may be transcribed divergently from, or convergently with, the structural genes of the operon. Repressor genes are constitutively expressed from their own promoters; that is, they are incapable of being repressed themselves.

A repressor protein binds to a short sequence adjacent to, and downstream from, a promoter. This sequence is called an operator. It is this operator sequence that gives an "operon" its name. If an operator is defective or deleted, all of the genes of the operon will be expressed constitutively, since the repressor cannot bind to a modified or deleted operator sequence.

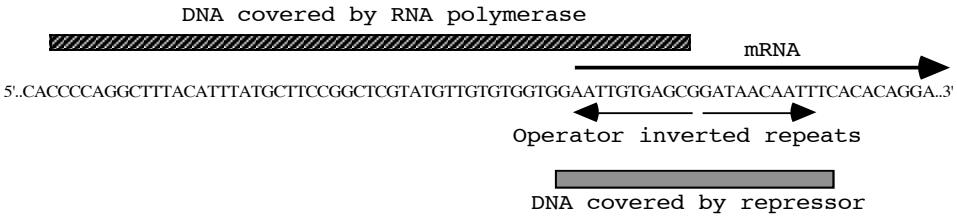
The lac Operator Inverted Repeat Sequence



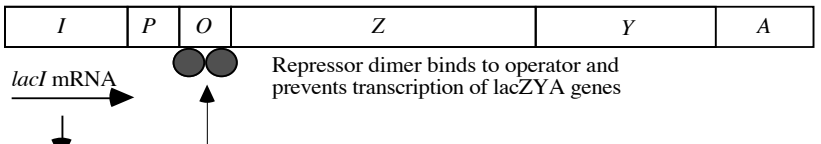
The bars on the top and bottom strands represent the inverted repeat (palindrome) sequences. The +1 site is the first nucleotide of the polycistronic mRNA. The Lac repressor is a dimer with each monomer bound to one of the repeat regions.

Repressors are titrated out as they are synthesised, or detached from operators, by inducer molecules that are antagonists of the repressors. The inducer forms an inactive complex with the repressor, which cannot bind to the operator. The inducer of the *lac* operon is lactose or the isomeric allolactose.

Lactose Operon Promoter and Operator Sites



Repressed lac Operon

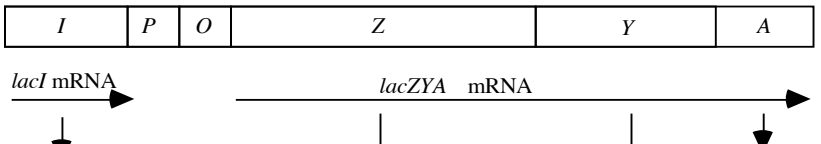


lacI mRNA

Repressor dimer binds to operator and prevents transcription of lacZYA genes

Repressor

Induced lac Operon



lacI mRNA

lacZYA mRNA

Repressor protein

Inducer (lactose or allolactose)

Beta-galactosidase

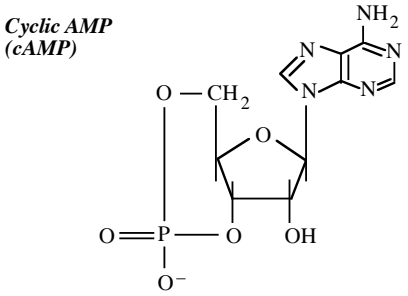
Permease

Transacetylase

Operons can also be affected by activator molecules that increase the rate of transcription (see below). This is termed positive regulation. Activators bind to short sequences - activator sequences - just upstream, that is 5', to promoters. Many operons can be both negatively and positively regulated, which is possible by having promoters straddled on either side by activator and operator regions.

If glucose and lactose are present in the growth medium simultaneously, then the cell will use glucose only. In such cases, the *lac* operon is repressed since it is not needed. The initiation of *lac* mRNA synthesis is regulated indirectly by the concentration of glucose in the cell. The glucose effect is linked to the *lac* operon by another molecule, cyclic AMP (cAMP) that is synthesised by adenylate cyclase.

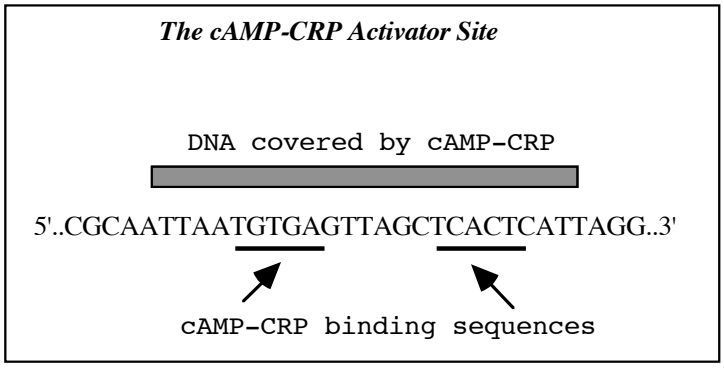
Cyclic AMP is ubiquitously distributed in prokaryotes and eukaryotes. In the latter, it is important in regulating the action of many hormones. In bacteria such as *E. coli*, the cAMP concentration is high if the carbon source is anything other than glucose, or if the cells are under starvation conditions.



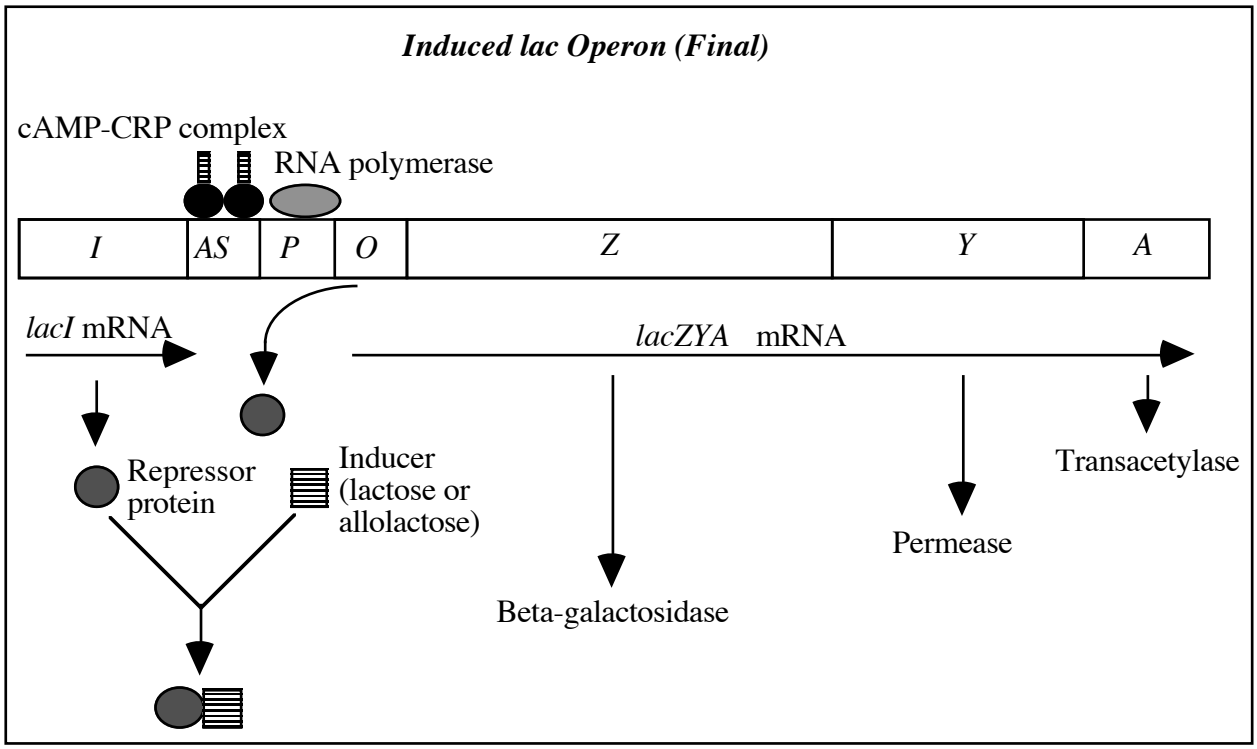
When the cellular glucose level is high, the cAMP level is low. Conversely, when the glucose level is low, the cAMP level is high.

In *E. coli*, cAMP binds to the product of the *crp* gene, the **c**AMP **r**eceptor **p**rotein (CRP). A cAMP-CRP dimer complex is the positive activator/regulator of the *lac* operon. It binds to the activator site just adjacent to the promoter and enhances the binding of RNA polymerase to the promoter site, provided that repressor molecules have been removed by the action of inducer molecules.

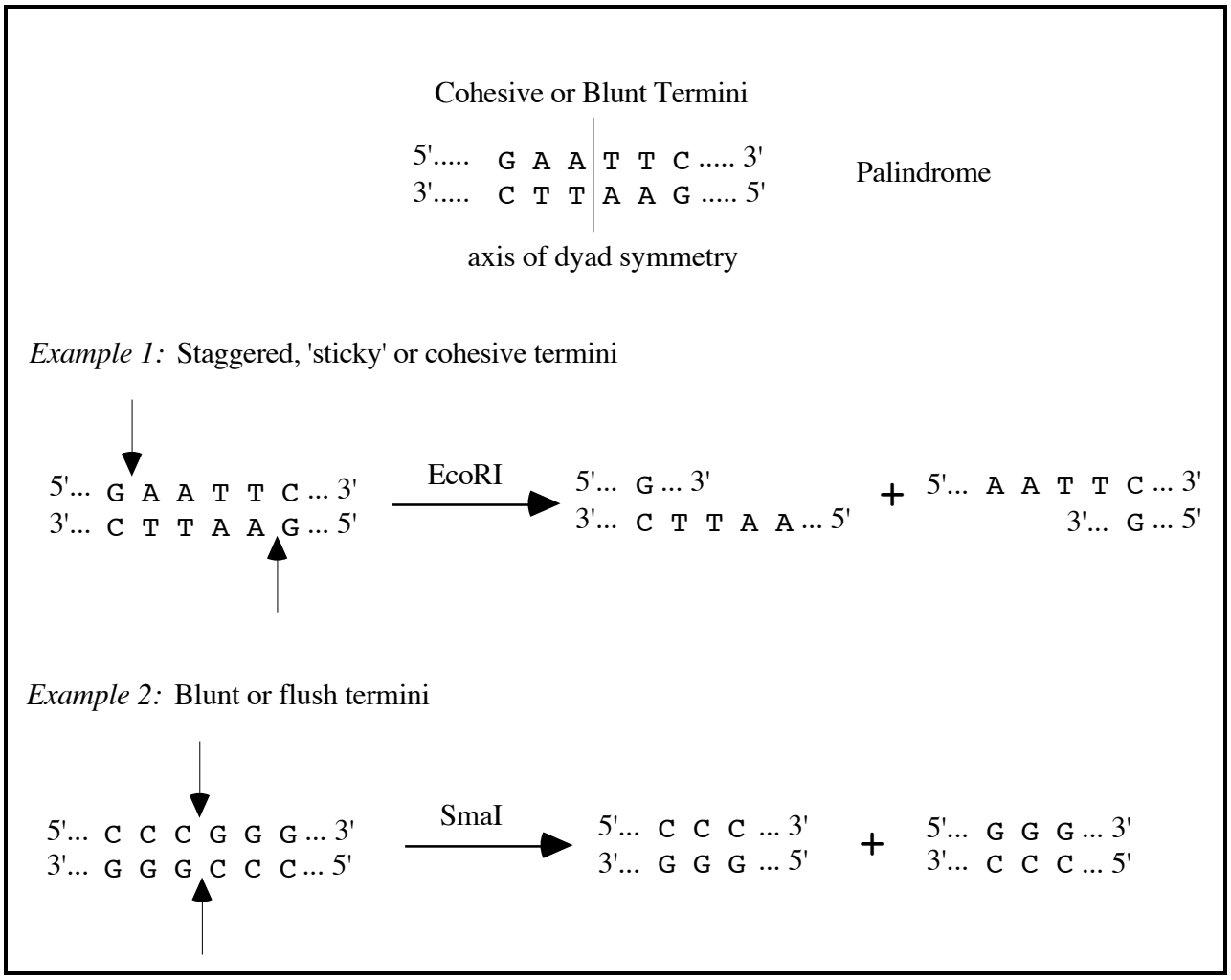
The activator site is - like the operator - a binding site for a regulatory complex, in this case cAMP-CRP. Two protruding alpha-helices (one form each dimer) fit into the major grooves of the DNA double helical activation site, making contact with the two consensus sequences in the grooves.



The final diagram for regulation and control of the *lac* Operon is:



RESTRICTION ENDONUCLEASES



Selected range of restriction endonucleases, their target sites and products of digestion

enzyme	target sequence	products	
HindIII	AAGCTT TTCGAA	A TTCGA	AGCTT A
SmaI	CCCGGG GGGCCC	CCC GGG	GGG CCC
HaeIII	GGCC CCGG	GG CC	CC GG
PstI	CTGCAG GACGTC	CTGCA G	G ACGTC

RECOGNITION SITES, FRAGMENT SIZES AND CUTTING FREQUENCY

1. For recognition sites of four (4) base pair palindromes, cut sites occur on average every $4^4 = 256$ base pairs.

2. For recognition sites of six (6) base pair palindromes, cut sites occur on average every $4^6 = 4096$ base pairs.

Most recognition sites are strict palindromes, but there are some sites that have unusual recognition sequences that are variable palindromes or not palindromes at all, for example:

(A/G)GCGC(T/C)	$GAAGA(N)_8$
(T/C)CGCG(A/G)	$CTTCT(N)_7$

Cutting site can be very near - or even at - the ends of a DNA molecule.

Examples of cutting near or at the ends of DNA molecules

enzyme	sequence	oligo length	% cleaved 2 hr	% cleaved 20 hr
AccI	GGTCGACC	8	0	0
	CCGGTCGACCGG	12	0	0
AscI	GGCGCGCC	8	>90	>90
AvaI	CCCCGGGG	8	50	>90
	CCCCCGGGGG	12	>90	>90
EcoRI	GGAATTCC	8	>90	>90
NcoI	CCCATGGG	8	0	0
	CATGCCATGGCATG	14	50	75
PstI	GCTGCAGC	8	0	0
	TGCACTGCAGTGCA	14	10	10

Cutting frequency.

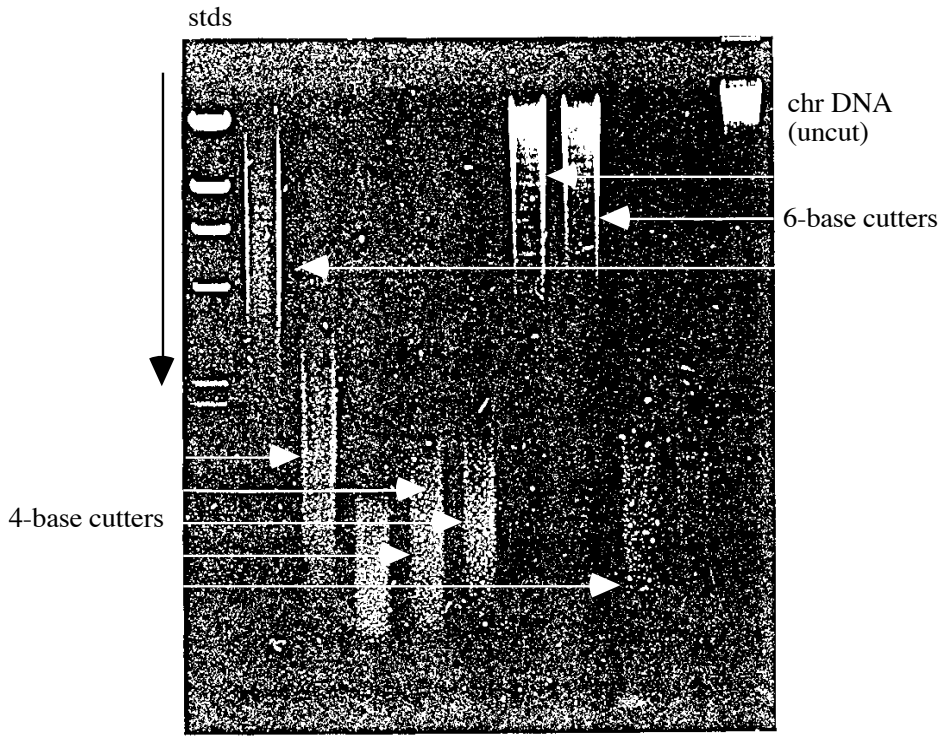
The *Escherichia coli* single chromosome is 4000 kilobase pairs (4×10^6 base pairs). Using the figures above, there should be approximately 15,625 sites for a four base pair enzyme, and 976 sites for a six base pair enzyme, if the sites are distributed randomly - which they are not, nor are they at regular intervals.

In practice, the number of sites is dependent on the nucleotide content of the DNA sequence though, in general, there will be more cutting sites for an enzyme that recognises a four-base site versus a six-base site (see the figure below).

The average size of a gene is 1 kb (1000 base pairs). Around four sites (variations are common) would occur for a 4-base cutting enzyme, and no sites for a 6-base cutting enzyme, within the 1000 base pairs.

In addition to the above examples of 4- and 6-base pair sites - which represent the majority of recognition sequences - there are a small number of restriction endonucleases that have recognition sequences of between 5 to 16 base pairs.

Chromosomal restriction digests showing the distribution of fragments



In the gel above, typical profiles for the electrophoresis of uncut and restricted chromosomal DNA are depicted. The lanes with profiles at higher positions in the gel contain chromosomal DNA that was digested with enzymes that recognise six-base palindromes, while the lanes with profiles at lower positions in the gel contain chromosomal DNA that was digested with enzymes that recognise four-base palindromes. A profile just like the one above will be produced in one of the practical class experiments.

Different banding geometry for chromosomal and plasmid DNA bands.

A simple way of discriminating chromosomal DNA bands from plasmid DNA bands is by visual observation of the bands in gels. Uncut chromosomal DNA migrates as a broad band with extensive striation and tailing (due to the drag of the large molecules through the pores of the gel). Plasmid bands are even in appearance with occasional tailing edges ('smiling').

Doublets.

Missing fragments are almost always 'present' as doublets or as small fragments that have migrated ahead of the bromophenol blue dye and off the gel (next section). Doublets are two restriction fragments of DNA migrating in the same position because they are almost identical in size. The tell tale sign is that the band is brighter than those immediately below and above. The presence of a doublet is best demonstrated by:

- excising the gel fragment containing the band

- extracting the DNA from the gel slice
- restricting it with a different enzyme (a four base cutter - why?)
- running the new digest in a gel
- estimating the sizes of the new fragments using linear standards.

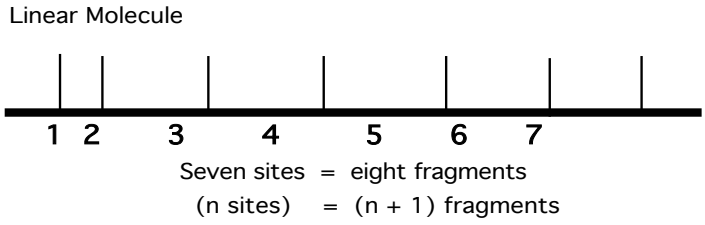
Small DNA fragments are resolved in acrylamide or highly cross-linked agarose gels.

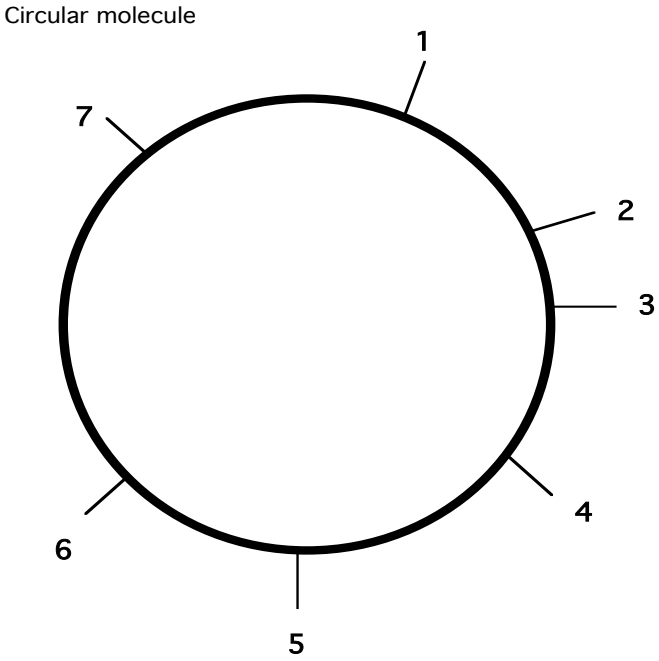
Very small fragments of DNA (less than 300 bps) either run as diffuse bands near the end of an agarose gel, or run off the end of the gel altogether since they migrate ahead of the BPB dye front. The best ways around this problem are to:

- use an agarose gel of higher percentage (1.5 to 2.5%) or higher cross-linked pore matrix
- use an acrylamide gel (10 to 15%) which will resolve small fragments (10 to 200 bps)

There are (n + 1) linear and (n) circular fragments for the same molecule.

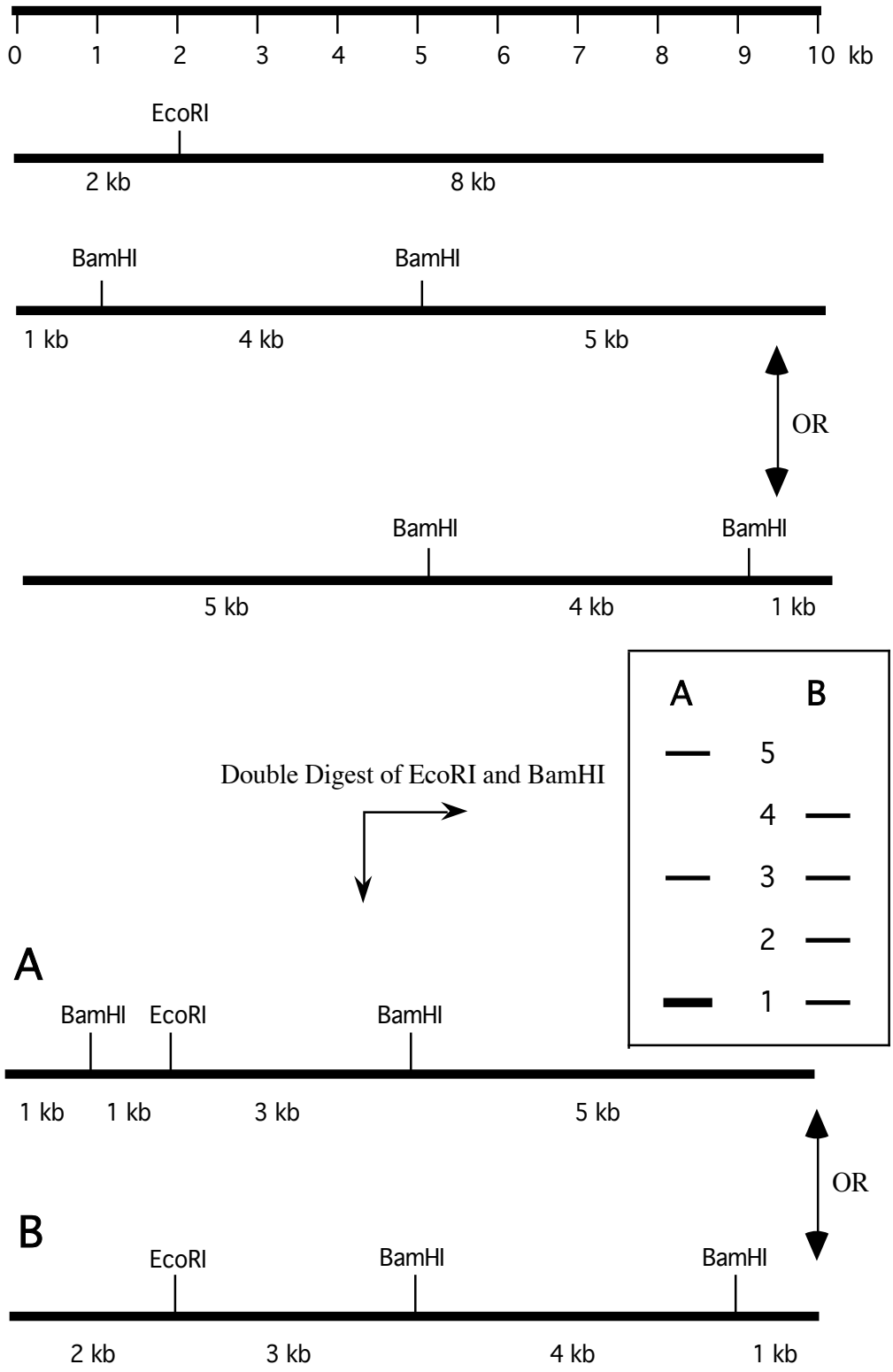
DNA Sites in linear and circular versions of the same molecule





Restriction mapping.

Restriction maps can be difficult. They require sequential single, double, and sometimes triple digests. In some complex cases, it may be necessary to cut one or more fragments out of an initial restriction gel, then do secondary or further digests on the isolated fragments. These can then be put back on the map of the whole original molecule.



DNA LIGATION

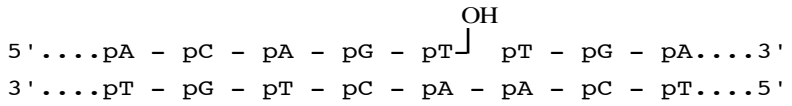
T4 DNA LIGASE

T4 DNA ligase catalyses the formation of a phosphodiester bond between adjacent 5'-PO_4 and 3'-OH ends of the same or different DNA molecule(s). The enzyme has an absolute requirement for ATP and Mg^{++}. Reactions can be carried out from 4°C to 37°C depending on the constraints of time and the required efficiency of ligation. T4 ligase is the ligation enzyme of choice.

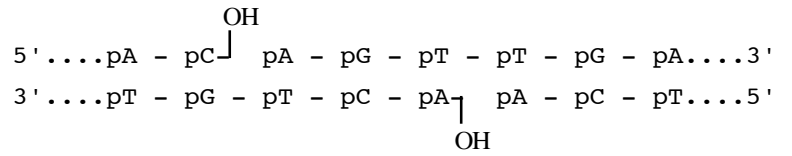
T4 DNA ligase is a single polypeptide enzyme (MW = 68,000) that catalyses the formation of a phosphodiester bond between adjacent 3'-OH and 5'-PO_4 ends.

Its substrates are:

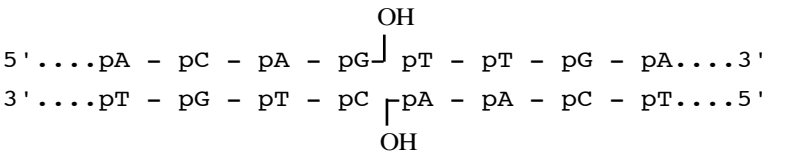
- Double-stranded DNA molecules with 'nicked' bonds,



- Double-stranded DNA (the same molecule or different molecules) with compatible cohesive (staggered) ends (generated from restriction digests),



- Double-stranded DNA (the same or different molecules) with blunt ends (generated from restriction digests or from PCR or from sheared DNA whose ends have been 'polished' with filling-in enzymes such as Klenow polymerase or T4 polynucleotide kinase),



T4 DNA ligase is also used routinely for attaching linkers and adaptors (short, double-stranded oligonucleotides) to DNA molecules in order to add or change the restriction sites on the ends of the DNA (see later lecture). In this way, virtually any DNA molecule can be adapted to have 'designer' restriction ends for specific cloning purposes.

In practice, DNA ligations are notoriously difficult, being captive to a range of controlling parameters that can result in undesirable by-products (see next two pages).

LIGATION PRODUCTS

Many types of products can be formed in any ligation reaction. The bias for any product type can be controlled reasonably well by the choice of conditions. Some products, such as re-ligated vectors, can be a problem unless steps are taken to reduce the background.

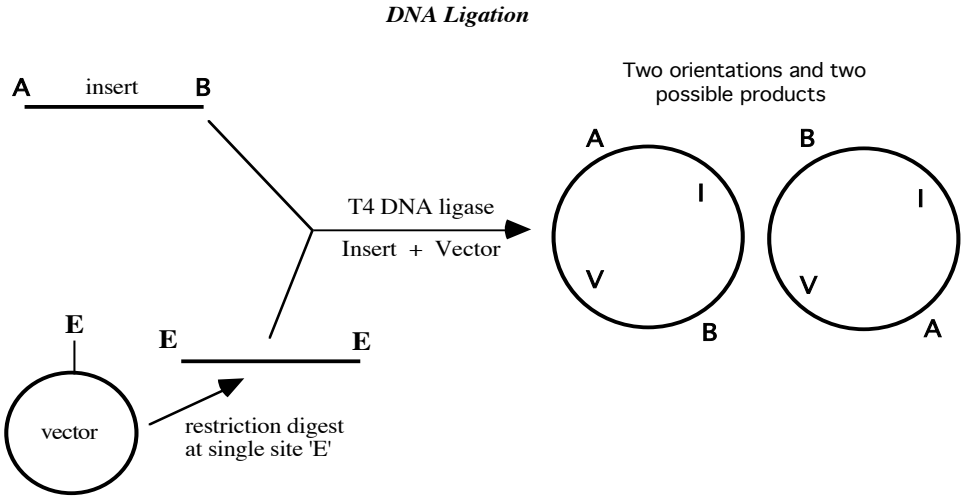
Optimising the formation of circular, recombinant plasmids depends upon trading off the two major competing reactions - the formation of *intermolecular* and *intramolecular* ligation products. If the ratio of vector:insert is too high, then intermolecular ligation is favoured. That is, long linear molecules are produced. On the other hand, if the vector concentration is too low, then intramolecular ligation is favoured with recircularised vectors being the major product. Ideally, the vector insert ratio should be about 1:1 irrespective of the size of either molecule. It is the equimolar concentration of ends of the two reactants that is critical for producing hybrid circles and <u>not</u> the respective masses of the molecules. See the outcome profiles on the next page.

COHESIVE VERSUS BLUNT-ENDED LIGATIONS

The ligation of DNA molecules with compatible cohesive ends is relatively simple since there is a degree of annealing of ends before DNA ligase forms the new covalent bond. However, blunt-ended ligations are more difficult since no annealing of ends can occur. In these cases, the efficiency can be improved markedly by macromolecular crowding (addition of 10% polyethylene glycol). Or, T4 RNA ligase can be used in concert with T4 DNA ligase. Since RNA ligase joins single-stranded molecules, it has no problem in attaching to blunt ends.

END-TO-END LIGATIONS (ORIENTATION)

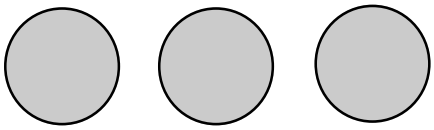
The orientation of DNA molecules to be ligated cannot be entirely controlled. The joining of two fragments to form a circle - the common example of a linearised vector and linear insert - will produce two recombinant plasmid products having the insert (A - B) in the example below) in two possible orientations in relation to the ends of the plasmid vector.



DNA Ligation of Vectors and Inserts

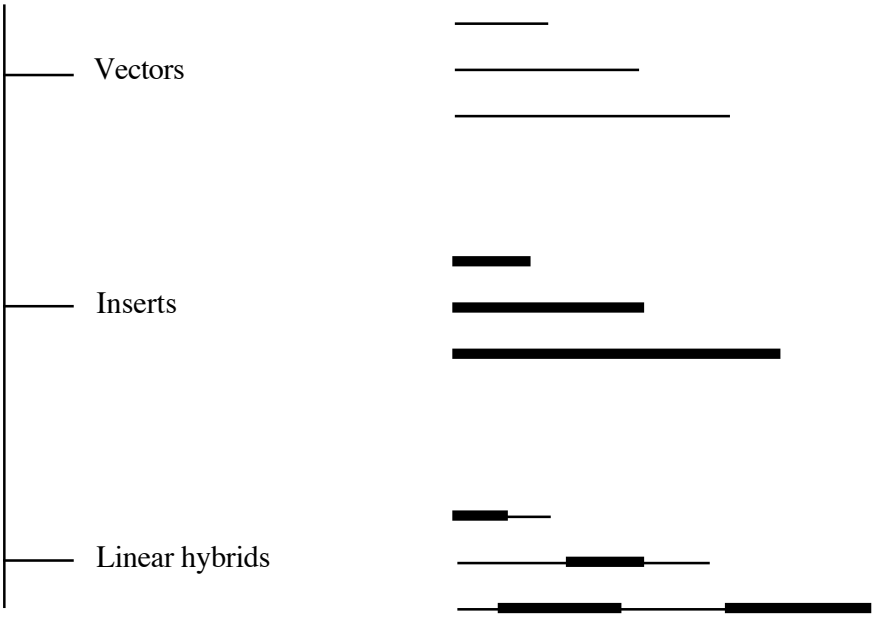
1. Recircularised vectors

2. Recombinants
(circular hybrids)



3. Circularised inserts

4. Linear concatemers (end-to-end)



TRANSFORMATION AND ELECTROPORATION OF DNA

Large naturally occurring plasmids can transfer between bacterial cells by conjugation. This process relies only on functions provided by genes on the incoming plasmid. The three limiting conditions for conjugation are -

- Transfer is often limited to a narrow host range (the same or very closely related species).
- Conjugative plasmids are greater than 20 kb since the number of genes required to encode the proteins needed for the transfer process comprises an operon of about 15 kbp, and the remaining DNA encodes replication genes and other traits (for example, antibiotic resistance).
- The new host (recipient) receives only one copy of the plasmid.

In molecular biology, in which cloning experiments depend upon small multicopy plasmids, an alternative to conjugation needed to be found for delivering plasmids into appropriate hosts. The first discovered of these - in 1972 - was **Transformation** of plasmid DNA. DNA could be transformed into bacterial cells that were made competent by prior treatment with calcium chloride at 4^{o}C. In 1988, an improved process of DNA transfer into cells was discovered. **Electroporation** is a process in which cells take up DNA when exposed to a brief but intense high voltage pulse. A point summary of each of these processes follows:

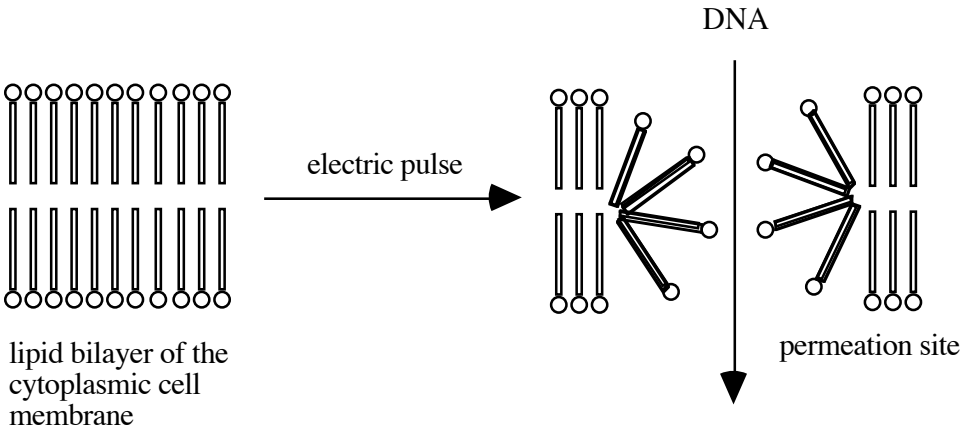
TRANSFORMATION

- Only circular DNA is transformed with any real efficiency
- Size limitation has inverse functionality (2 kb plasmid is transformed at 10 times the efficiency of a 4 kb plasmid, 100 times more efficiently than 8 kb, etc up to about 30 kb)
- Only a small percentage of “competent” cells are transformed. Competence is achieved after treatment with - and transformation in the presence of - calcium chloride
- Transformation is only achieved if DNA and cells are held at 4^{o}C for about 30 minutes, followed by a 2 minute heat shock at 42^{o}C
- Optimum results are achieved with competent cells prepared from early or late log phase cells, but not mid-phase cells (reason not known)
- There is a DNA concentration dependence on the frequency of transformation with an increase in the low range of 1 to 100 ng of plasmid DNA per 100 μl of cells, to a plateau of about 500 ng, then a decrease in which excess DNA is actually inhibitory to further uptake
- Maximum number of transformants is 10^8 to 10^9 per μg of DNA

ELECTROPORATION

- Cell will take up circular or linear DNA during an electric pulse but linear DNA is taken up at about 4 logs less than circular DNA

- Electrocompetent cells are prepared by repeatedly washing mid-log phase cells in cold ion-free water. DNA should also be in a low ionic strength buffer or in water, and in a small volume so as not to exceed the cuvette capacity
- There is very little size dependence over the range of 2 to 10 kbp of plasmid DNA, and plasmids up to 100 kbp can be introduced into cells
- Frequency is dependent on [DNA] over at least six orders of magnitude
- Method is based on transient permeabilisation of cell membranes (formation of pores) in a short pulse of a high intensity electric field
- A short recovery period in growth medium is required for the survivors (about 20 to 40%), but up to 80% of these can take up DNA)
- Maximum number of transformants is about 10^{10} per μg of DNA



The principle of electroporation. An electric pulse creates transient permeation sites in the lipid bilayer

DNA SEQUENCING

INTRODUCTION: GENERAL INFORMATION

Fragments of DNA including whole genes can be subjected to one of several methods now available for DNA sequencing. The sequence generated is that of each consecutive nucleotide in the 5'-PO_4 to 3'-OH direction of a DNA single strand. The opposite strand can be deduced by A-T and G-C complementarity, or can itself be sequenced in another set of reactions.

DNA sequencing involves several stages of experimentation and analysis. Once a sequence is produced, it can then be analysed for a number of features, namely open reading frames, promoters, termination signals, restriction sites, etc. Database searching and matching will then identify other sequences of variable relatedness. Further analysis can identify protein sequences or motifs and information about structure and function.

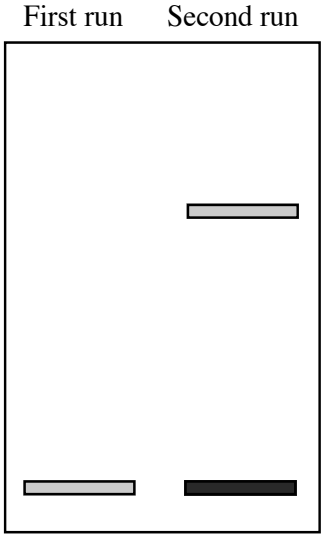
Nucleotide sequences can be obtained by the traditional manual method, or by one of several automated methods now available (see below). When the sequence is obtained, it is usually entered as a file into a computer. Subsequent related, overlapping sequences (Contigs) are added as separate files as they are generated. A computer analysis program (for example, 'MacVector AssemblyLIGN') will then line up the overlapping contigs and produce a completed consecutive order of nucleotide bases.

That nucleotide sequences can even be obtained depends upon several important technical protocols. For instance, sequencing reactions may use vanishingly small aliquots of templates and reagents. This is possible because the DNA is labelled with a radioactive isotope that can be detected easily in an autoradiogram of a developed gel. The small volumes used mean that the gel thickness must be at a minimum (usually 0.2 or 0.4 mm). This thinness is helpful for the autoradiographic detection in X-ray film. For instance, if the nucleotide isotope label is ^{32}P, the gel need not be processed any further after it is run - it is covered in gladwrap and the X-ray film is placed against it. If ^{33}P or ^{35}S are used, their weaker β-emissions require the gel to be dried down either on the glass plate or on thick blotting paper. The X-ray exposure times are also longer. Newer automated sequencers make use of nucleotide reagents with attached fluorescent dyes that can be detected by lasers in the sequencing apparatus. More template is required.

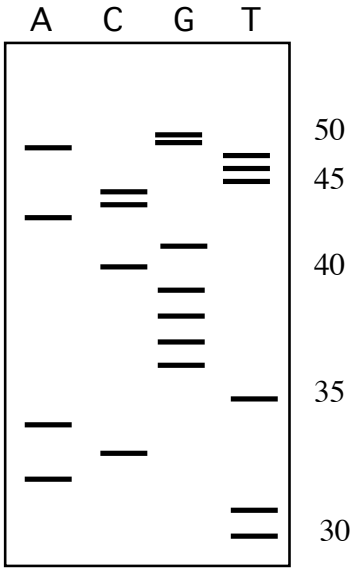
The choice of isotope may depend on the experimental requirements but more often is only determined by the personal preference of the investigator. For example, α-^{32}P-ATP will give quick results (several hours to overnight) due to the extra energy of the β-emissions. However, the bands are not very sharp. Also, the half-life of ^{32}P is only 14 days meaning that the isotope must be used over a short time. Indeed, once the sequencing reactions have been done, they must be run in a gel the same day. On the other hand, α-^{35}S-ATP will give very sharp bands (which helps to discriminate bands that are close together but in separate lanes); however, against this is the need to leave the gel exposed to the X-ray film longer (say, 2 to 4 days). An advantage of the longer half-life of ^{35}S - at 87 days - is that the isotope can be used for reactions over several weeks at least, and once the sequencing reactions are done, they can be stored frozen for up to a week before running the gel. The ^{33}P isotope has properties that give results about midway between ^{32}P and ^{35}S.

A second technical feature is the type of gel used. The length of sequence that can be read will depend mainly on the resolution of individual fragments in the sequencing gel, and thus the longer the gel, the longer the sequence that can be read. There are of course physical limits to

this notion, and in practice most sequencing gels are about 40 to 60 cm long. One way in which the length of a sequence can be extended to perhaps 2 to 3 times the length of the gel is to use second and even third loadings of the same sequencing reaction mixtures onto the same gel at different times. This is illustrated in the diagram over the page. Note the positions of the overlapping dye fronts. In practice, the size of sequencing fragments can be estimated from the positions of the dyes. For example, in a 6% gel, the bromophenol blue marker dye will co-migrate with a 26-base single strand and the xylene cyanol dye with a 106-base single strand. Note that the bands in a sequencing gel are single-stranded.



Most gels are cast as 4, 6 or 8% polyacrylamide plus 8M urea (as a denaturant that enables fragments to migrate according to linear proportionality). In this way, fragments that differ in length by a single nucleotide can be resolved over a range of approximately 30 to 300 nucleotides. This means that the first nucleotide that can be read from the gel from the bottom will be perhaps 30 bases in length and the next one that is in a slightly higher position will be 31 bases long, the next 32 bases and so on, as in the diagram below.



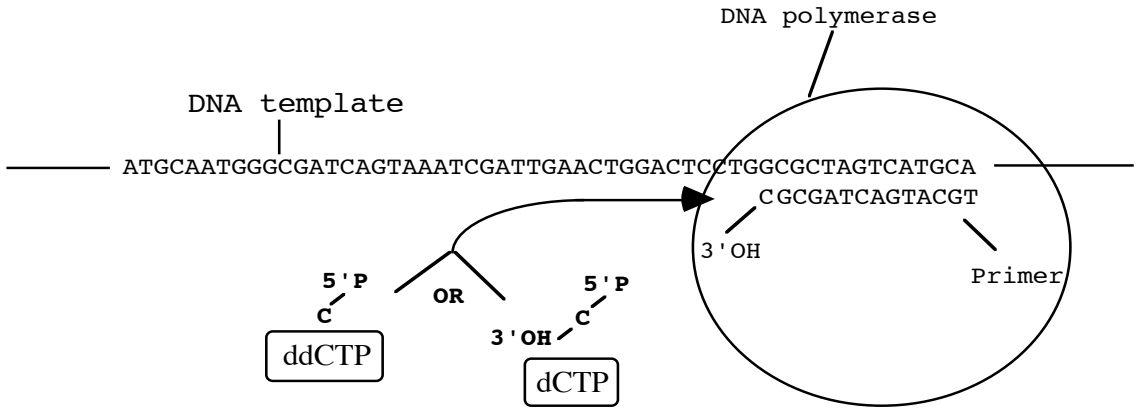
The sequence is read from the bottom up beginning at the 5' end and with the shortest fragments. Thus, the sequence above would be: 5' - TTACATGGGGCGACCTTTAGG - 3'.

The bands are closer together higher up in the gel because the resolution of longer fragments in this percentage of acrylamide gel is being reached. So, only about 200 bases can be read in a set of lanes. One way around this problem is to load the same samples at different times.

Before the availability of improved sequencing protocols and computer analysis packages, DNA sequencing was a long and arduous process. To give some indication of the current improvements, perhaps as few as just 10 to 15 years ago, a PhD student might have cloned and sequenced a single gene over a 3-year candidature. Today, this task can be achieved in a matter of days, and is very often carried out by laboratory technicians using automated methods.

SANGER DIDEOXY CHAIN TERMINATION METHOD

The principle of the dideoxy method involves the use of chain terminating dideoxynucleotides (ddNTPs). These lack the 3'-OH group on the deoxyribose moiety, which means that they cannot contribute to the formation of the next phosphodiester bond in the growing DNA chain, which is therefore terminated (see the diagram below).



In a typical set of 4 chain termination sequencing reactions for G, A, T and C, all 4 reactions will contain all 4 dNTPs, but each separate reaction will contain one only of the 4 ddNTPs.

Typical reaction mixes:

- DNA template
- an oligonucleotide primer
- DNA polymerase
- dATP, dGTP, dCTP and dTTP
- α-^{35}S-ATP (or α-^{32}P-ATP or α-^{33}P-ATP)

After mixing and incubating for several minutes to prime the template, the mixture is divided into four equal aliquots and a different termination cocktail is added to each aliquot. For example, in tube 1, ddGTP is added so that its ratio to dGTP is 1 : 20.

- The tubes with the ddNTPs are given the nucleotide name. So, if tube 1 contains ddGTP, it will generate the 'G' lane in the sequencing gel, and so on.
- In practice, the ratio of ddNTPs and dNTPs are pre-determined in the reagent kits so that the strands can be extended and terminated simultaneously. This means that in any one reaction, a

template will be extended to its entire length and to every point that is shorter by the positions of the nominal terminating base. This results in a family of strands with a common 5' end, but of different lengths corresponding to where the chain was terminated by the ddNTP incorporation.

- In automated fluorescent dye cycle sequencing, the four ddNTPs each come with a different coloured fluorescent dye tag. There is no need for a separate radiolabelled nucleotide. All four reactions are run in a single lane and the migrating bands are read near the end of the gel by a laser detector, which converts the fluorescent light to an electrical signal and then to a base in a computer file.

The Four Chain Terminating Reactions for the same DNA Template

ddCTP terminating reactions

5'-GATCGATCGATCGATCGATCGATCGATCGATCGATCGAT-3'
3'CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTAGCTA-5'
3'CTAGCTAGCTA-5'
3'CTAGCTA-5'
3'CTA-5'

ddTTP terminating reactions

5'-GATCGATCGATCGATCGATCGATCGATCGATCGATCGAT-3'
3'TAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTAGCTA-5'
3'TAGCTAGCTAGCTA-5'
3'TAGCTAGCTA-5'
3'TAGCTA-5'
3'TA-5'

ddATP terminating reactions

5'-GATCGATCGATCGATCGATCGATCGATCGATCGATCGAT-3'
3'AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTAGCTA-5'
3'AGCTAGCTAGCTA-5'
3'AGCTAGCTA-5'
3'AGCTA-5'
3'A-5'

ddGTP terminating reactions

5'-GATCGATCGATCGATCGATCGATCGATCGATCGATCGAT-3'
3'GCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTAGCTA-5'
3'GCTAGCTAGCTA-5'
3'GCTAGCTA-5'
3'GCTA-5'

- Following denaturation (boiling for 3 min) to separate the labelled synthesised strands from the template, the four reactions are loaded into separate wells of a denaturing polyacrylamide gel to separate the strands by size. The pattern of labelled bands indicates each base at which chain termination occurred. The sequence is read from the bottom in order of increasing size of synthesised strands in the 5'----> 3' direction.
- If the complementary strand of the DNA template is to be sequenced as well, then a separate primer is needed and 4 new reactions are set up as above for the complementary template.
- In automated sequencing, all reactions are done in the one tube with all 4 ddNTPs labelled with a different colour dye. The reaction is loaded into a single gel lane and the various terminated fragments are discriminated by their different colours as they pass by the laser detector in the instrument. A computer is interfaced to the gel apparatus and detector.

DNA POLYMERASES FOR SEQUENCING

DNA polymerase synthesises a DNA strand complementary to a single-stranded DNA template, adding dNTPs to the 3'-OH end of an oligonucleotide primer annealed to the DNA template.

- Modified T7 DNA polymerase (Sequenase). Genetically engineered to remove the 3'-5' exonuclease activity of bacteriophage T7 DNA polymerase. Sequenase is highly processive, producing long lengths of synthesised DNA.
- Thermostable DNA polymerases (egs. Taq or Pfu). Active at higher temperatures than Sequenase. Useful for sequences with strong secondary structure (regions of mostly G-C pairs that are more difficult to denature). Used for cycle sequencing and PCR.

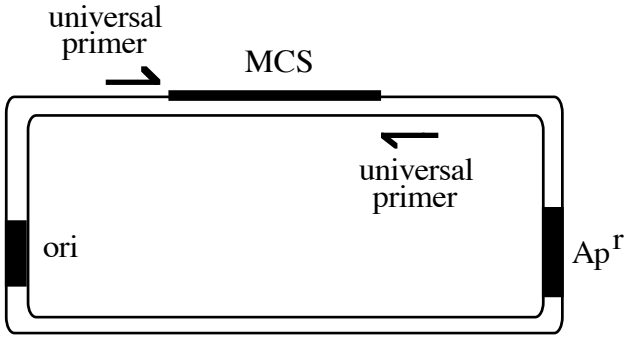
SEQUENCING OF PCR PRODUCTS

PCR products) can be sequenced without subcloning into vectors,. using the same PCR primers for DNA sequencing.

PLASMIDS AS SEQUENCING VECTORS

Common examples are the pGEM, pBluescript and pUC series of vectors. These also use the X-gal / IPTG selection system. Plasmid DNA must be denatured by heat or alkali to generate single strands prior to setting up sequencing reactions. Primers are annealed during the slow cooling or neutralisation processes. Universal primers complementary to sequences upstream and downstream of the MCS (and therefore either side of cloned inserts) are commercially available.

Plasmid vector



CYCLE SEQUENCING

Cycle sequencing is the best method for sequencing double-stranded plasmid DNA. This is because it avoids the main problem of template annealing that reduces the efficiency of ds plasmid sequencing. The technique employs a thermal cycler (the same as for PCR) and a thermostable DNA polymerase (egs. Taq, Vent, Pfu). Once reagents (template, primer, dNTPs, ddNTPs and buffer) are mixed together and enzyme added, repeated cycles of heat denaturation (95^{o}C), annealing (54^{o}C) and extension (72^{o}C) will generate a family of linear fragments all with the same 5' end and endpoints at each nucleotide (as tagged by a dideoxynucleotide). The newly synthesised DNA can be labelled with an isotope or with fluorescent dyes (the dyes can be attached to the primers, dNTPs or ddNTPs).

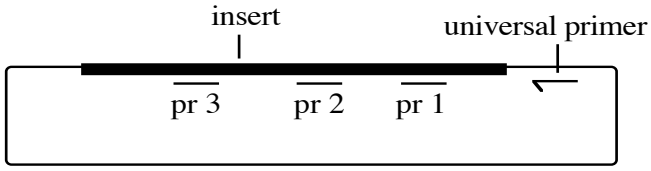
PRIMER DESIGN

If universal primers cannot be used, oligonucleotide primers can be designed using standard software provided that some sequence is already known (for example, if sequencing into a large insert having begun with vector universal primers). The major constraints are that the GC content

should be in the range of 40 to 60% and not be self-complementary. The longer the better but anywhere from 16 to 24 bases is the usual range (depending on the fidelity required and cost).

SEQUENCING ACROSS A LARGE INSERT

One of the simplest ways to do this that doesn't involve any sub-cloning of smaller fragments, is to sequence 300-600 bps into the insert from a vector primer, then have a primer synthesised equivalent to about 20 bps of the last clear sequence obtained and use this to prime the next sequencing reactions. Repeat until the insert has been sequenced completely (see diagram below).



SHOTGUN SEQUENCING

DNA is firstly fragmented by ultrasonication, DNase I treatment, or partial restriction digestion with Sau3A. In the first two examples, repair the ends with T4 DNA polymerase and clone fragments into a blunt-ended restriction site of a sequencing vector (in either orientation). In the case of Sau3A, clone into the compatible BamHI site of a plasmid vector. Since the order of sequenced fragments will be initially unknown there is a need to sequence more than the total template DNA because sequences of overlapping fragments (contigs) can only be aligned correctly using a sequence alignment computer program.

TROUBLESHOOTING IN DNA SEQUENCING

Stops:

- These are places in the gel where no further sequence occurs, and they are due to regions of secondary structure in the template caused by homopolymeric tracts (runs of the same base). With sequenase (a highly processive polymerase) or in cycle sequencing at a higher temperature with a thermostable polymerase, this problem is overcome.

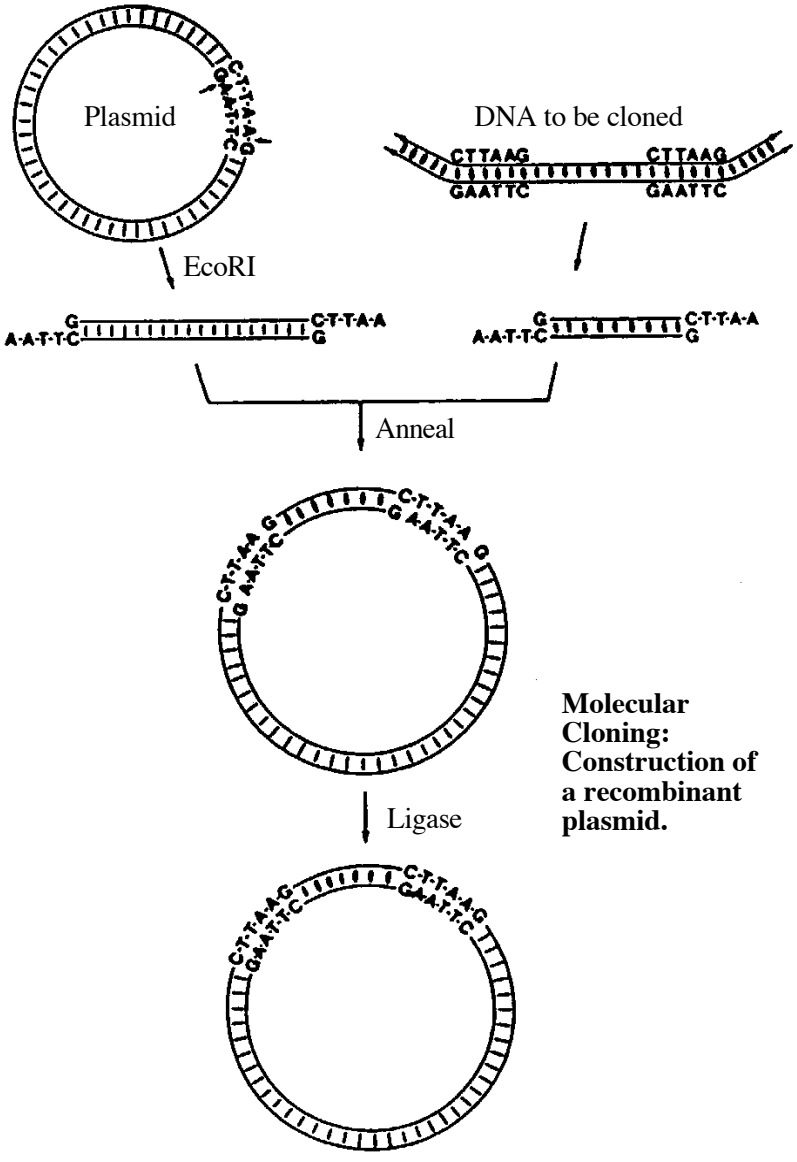
Compressions:

- These occur when bands in a particular position in the gel are crowded together (compressed) making the sequence unreadable. Compressions are gel artefacts caused by short stretches of dyad symmetry (hairpin loops or inverted repeats), especially regions of high G-C content. The problem may be overcome in one of several ways. Simply sequence the opposite strand; or add 40% formamide to the gel mix; or add 10% dimethylsulfoxide to the sequencing reactions; or use dITP in place of dGTP. dITP is a dGTP analog that forms I-C pairs but with 2 hydrogen bonds only (compressions are therefore more difficult to form and maintain).

MOLECULAR CLONING STRATEGIES

The molecular cloning of genes may be a relatively simple or difficult task. It is beyond the scope of this course to look at unusual and rare cloning strategies. The simplest cloning strategy requires the ligation of a restriction fragment of DNA to a plasmid vector cut with the same - or compatible - restriction endonuclease (shown in the diagram below).

Transformation or electroporation of the ligated DNA into competent bacterial cells is followed by an appropriate selection system for the recombinant plasmids in the host cells. In practice, transformed cells may also contain religated vector plasmid DNA, as it will be selected, like the recombinant plasmids, according to the antibiotic resistance property. The transformants (clones) are screened for the presence of the desired DNA fragment or gene in the recombinant plasmids. The systems described in the following sections illustrate both simple and complex cloning strategies; and methods for minimising vector-only background.



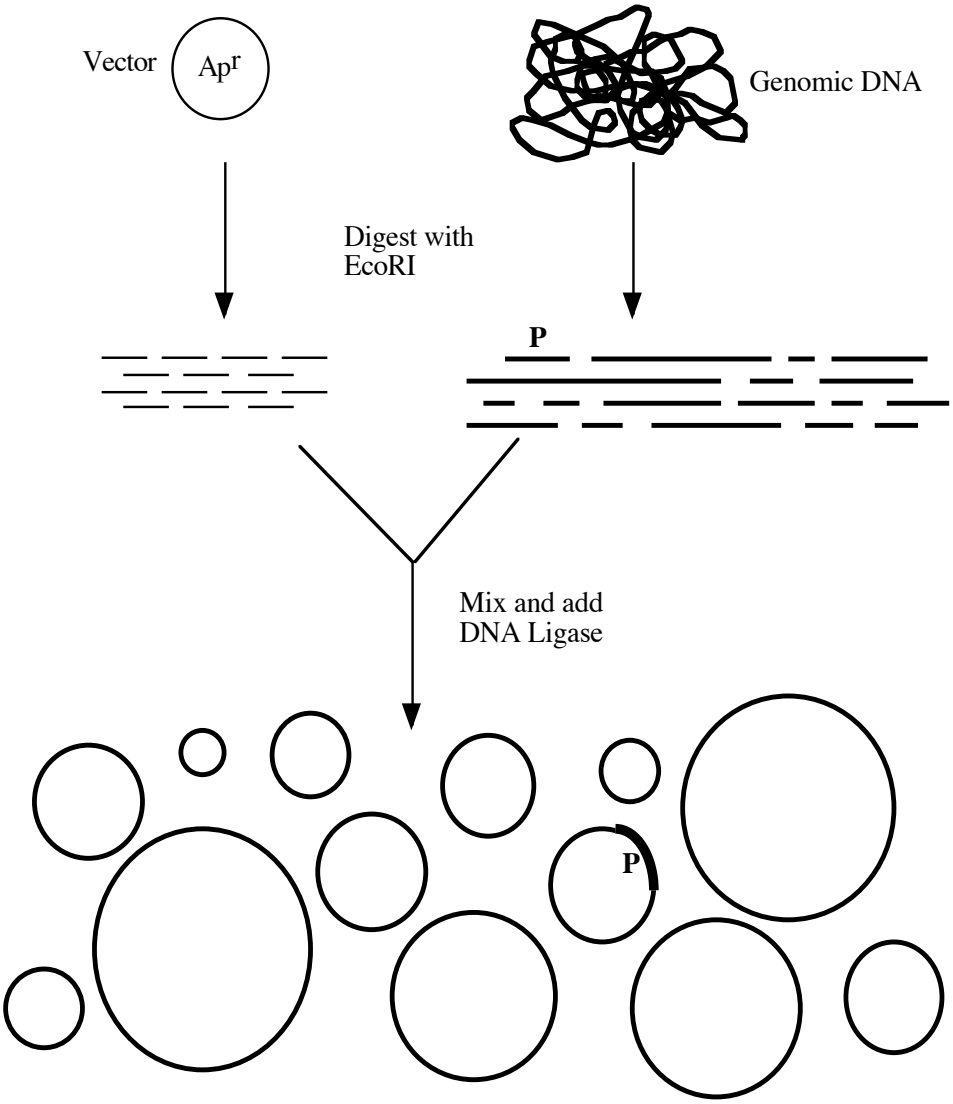
Molecular Cloning: Construction of a recombinant plasmid.

SELECTION FOR RECOMBINANTS: POSITIVE SELECTION

Positive selection for a product phenotype (detectable enzyme or protein)

The simplest detection scheme is one in which the cloned gene has a readily detectable phenotype. In some cases, the clone of interest will be the only type to grow, whilst in others, all transformants will grow but the desired clones can be discriminated by a phenotype identifier (antibiotic resistance such as: ampicillin, chloramphenicol, or tetracycline) or test (enzyme catalysed colour change, perhaps directly on the agar plate).

Molecular Cloning: Positive Selection



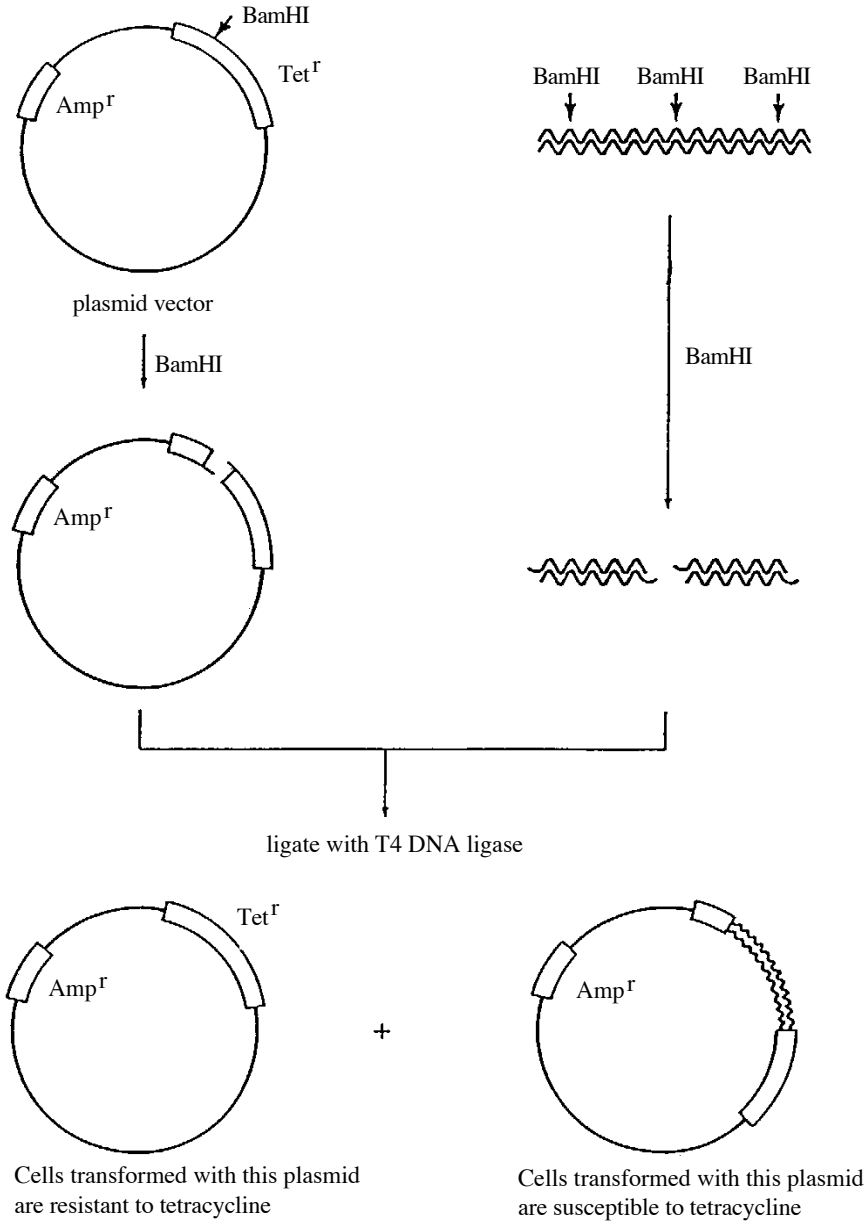
Transform competent *E. coli* cells and select Ap^r and Z^+

If a phenotype is not available, then 'silent' proteins can be detected by *Western Blotting*, using monoclonal antibodies (explained in a later lecture). Another method for the selection of positive clones is DNA probe detection in which the target clone is identified by hybridisation. This is called *Southern Hybridisation* and it will also be explained in a later lecture.

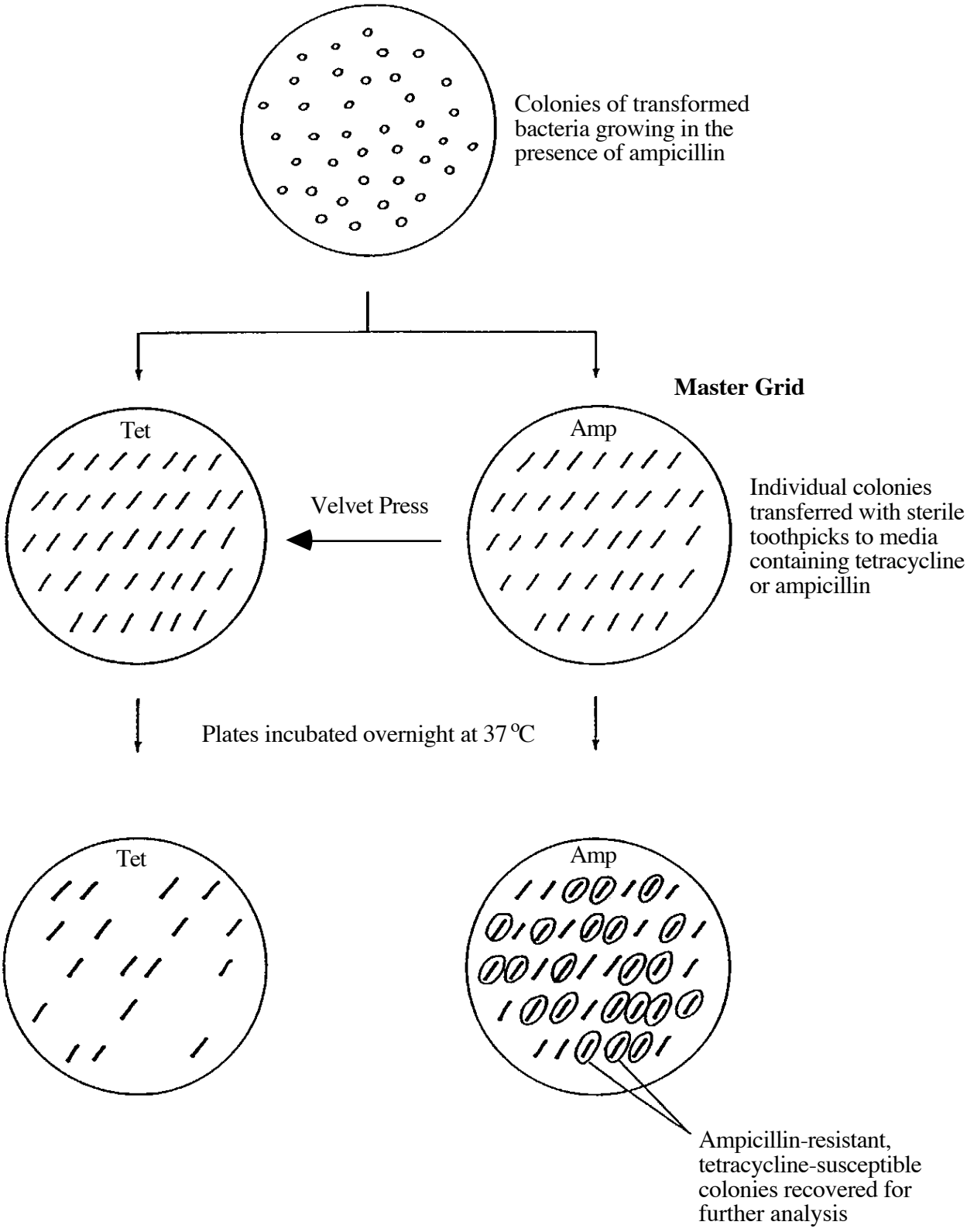
SELECTION FOR RECOMBINANTS: INSERTIONAL INACTIVATION

The simplest method of detecting recombinants is by insertional inactivation of an antibiotic resistance phenotype. This requires the vector to have two antibiotic resistance genes; in the example below, these are Amp^R and Tet^R. If inserts are ligated into the BamHI site within the tetracycline resistance gene sequence, then the Tet resistance gene will be interrupted by the insertion, and the phenotype of cells containing such a recombinant plasmid will be Tet^S, that is, these cells will now be susceptible to the antibiotic tetracycline.

Screening for recombinants is made on antibiotic containing selection plates. Transformants are first plated onto agar + ampicillin. These clones are then replica-plated onto agar + tetracycline. Recombinants (Tet^S) will not grow on the tetracycline plates whereas vector-only transformants will because these are still tetracycline resistant. The recombinants identified in this way may then be recovered from the master ampicillin plates.



Screening for insertions of foreign DNA by inactivation of plasmid-borne antibiotic resistance genes

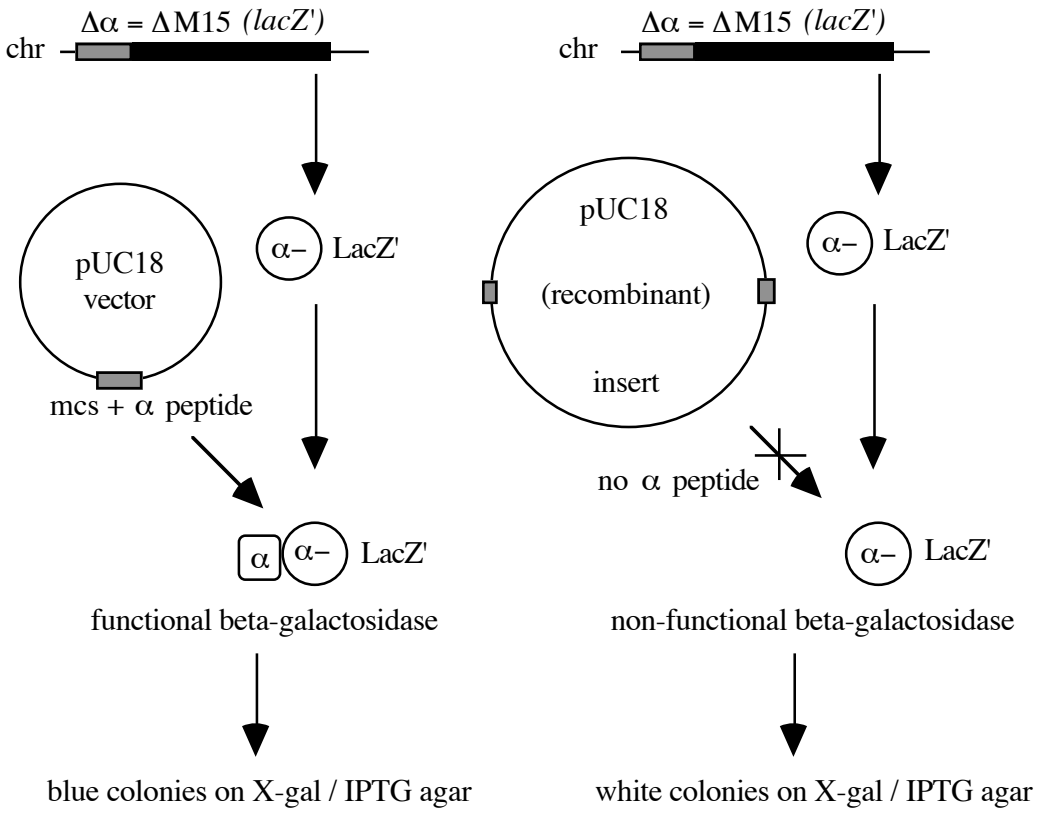


LacZα / β-galactosidase complementation system (blue / white colonies)

Complementation is a phenomenon in which two separated parts of a normally contiguous protein can associate into a functional composite complex. The *E. coli lacZ* gene encodes β-galactosidase, an enzyme that allows cells to utilise lactose. Many cloning vectors contain a small 5' fragment of the *lacZ* gene, which encodes the N-terminal portion of the LacZ protein (the α-peptide or M15 segment containing amino acids 11 to 41 of LacZ). The remaining portion of the gene is present in the *E. coli* host cell chromosome. Transformants containing such vectors will supply both portions of a now functional β-galactosidase. This effect is called *complementation.*

If the selection plates are seeded with the gratuitous inducer IPTG [isopropyl-β-D-thiogalactopyranoside] - to allow expression of the LacZ α-peptide from the vector plasmid - and the chromogenic substrate X-Gal [5-bromo-4-chloro-3-indolyl-β-D-galactoside], then cells containing vector plasmids will be able to hydrolyse the X-Gal substrate, releasing the intense blue coloured indolyl moiety which makes the colonies blue.

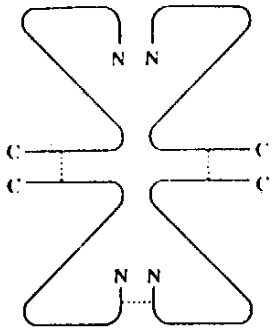
Vectors utilising this system also contain a MCS within the *lacZ*α sequence. True recombinants will be unable to express the alpha-peptide since there is now an inserted fragment in the *lacZ*α sequence. Complementation is prevented in these transformants and the resulting colonies will be white since the X-Gal substrate in the vicinity on the clone cannot be hydrolysed.



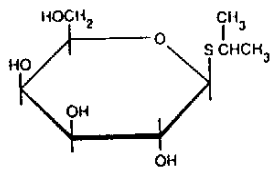
In the α-complementation scheme above, the chromosome (Chr) has a defective *lacZ* gene (*lacZ*') that cannot encode the N-terminal α-peptide of β-galactosidase. If a vector (pUC or M13) is present in the same cell (left panel), an α-peptide can be produced and can combine

with the remainder of the C-terminal portion of LacZ to generate a functional β-galactosidase (that can hydrolyse the X-gal substrate - see O/L - to produce a blue colour). If a DNA insert is cloned into the vector (right panel), the α-peptide cannot be encoded since its sequence is now interrupted by the insertion. In this case, complementation is impossible and the X-gal substrate cannot be hydrolysed, resulting in white colonies or plaques.

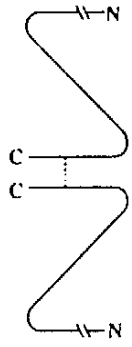
Biologically active tetrameric β-galactosidase; non-covalent forces maintain quaternary structure



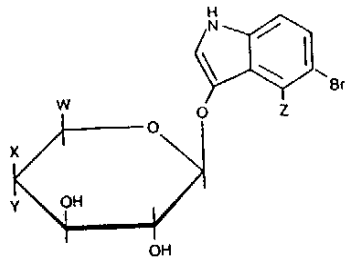
IPTG



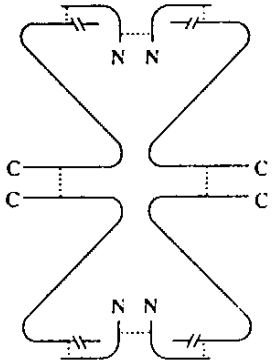
Inactive dimer produced in the deletion (lacZΔM15) which removes amino acids 11-41.



X-GAL



W = CH_2OH
X = OH
Y = H
Z = Cl



Dimer-dimer interaction restored by the α-fragment of β-galactosidase supplied by the vector.

REDUCING VECTOR BACKGROUND IN CLONING EXPERIMENTS

Directional (Forced) Cloning

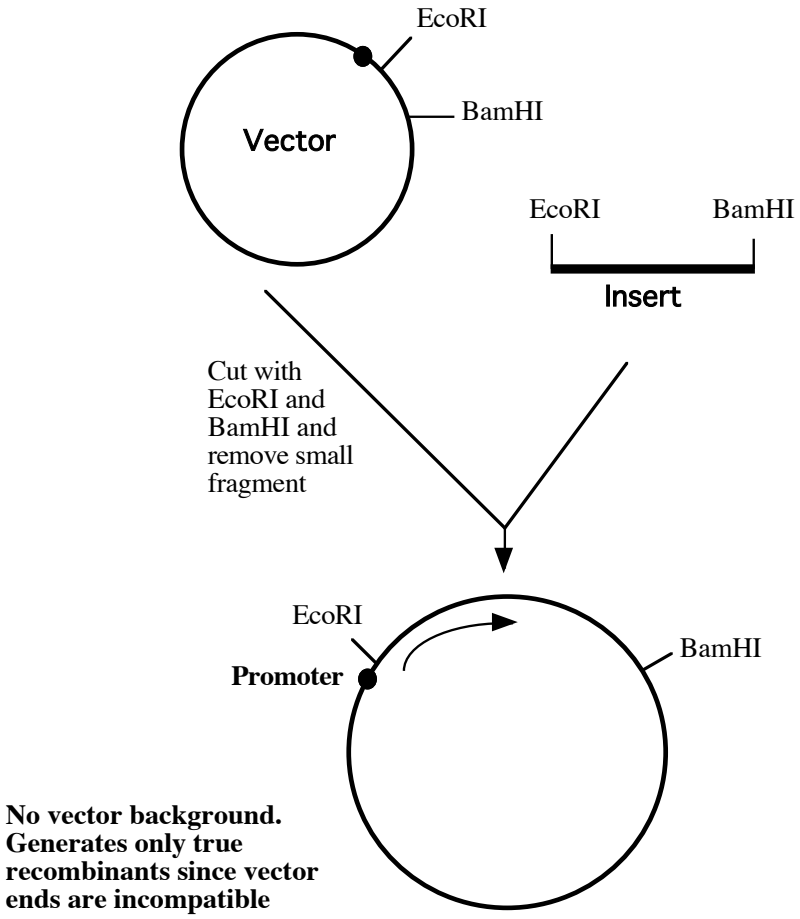
A Multi-Cloning Site (MCS) in a vector is a short sequence containing tandem restriction enzyme sites, giving different enzyme options for the insertion - and subsequent removal - of a fragment. The MCS is located within insertional inactivation detection regions (e.g. lacZα).

Directional cloning enables the insertion of a fragment in one orientation only, by using two different enzyme sites in the MCS and the same two sites on the fragment. Note, this method will not work if the two different enzymes generate blunt ends (Why not?).

There are two main reasons for using directional cloning:

- It is a simple means of reducing vector background - since the ends of the vector have been generated with two incompatible restriction endonucleases.
- Expression vectors have promoters and ribosomal binding sites just upstream from MCSs. These are used to generate mRNA transcripts from genes cloned without promoters (for example, cDNAs); or to increase transcription of a gene whose promoter is much weaker than the vector's promoter. The correct orientation of the inserted fragment is crucial.

Directional (Forced) Cloning



Alkaline Phosphatase to Reduce Vector Background

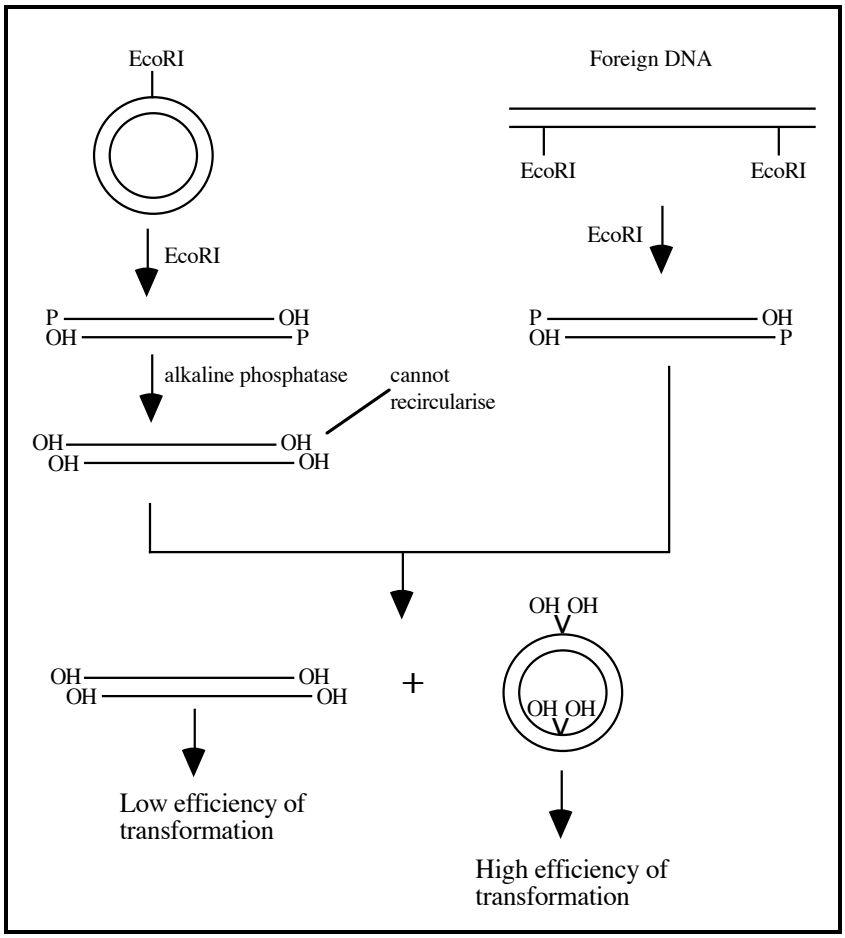
Double-stranded DNA molecules are anti-parallel. That is, they have 5'-PO_4 and 3'-OH groups opposite to one another on the ends of each strand. When these ends are ligated to other double-stranded DNA molecules, new phosphodiester bonds are formed.

One major problem in cloning DNA fragments into plasmids is that the ends of the restricted plasmid are more likely to ligate together than to be ligated to the ends of a fragment. This means that a high percentage of the clones consist of re-ligated vector only. This problem is made worse by the transformation bias of the vector over the larger recombinant plasmids.

A simple solution to this problem is to pre-treat the restricted vector molecules with alkaline phosphatase, an enzyme that removes 5'-PO_4 groups from the ends of DNA strands. This means that all four ends of the linearised vector now contain only hydroxyl (OH) groups and these cannot be ligated together, since there are no phosphate groups to form phosphodiester bonds with the hydroxyls (see figure below). The vector background is largely eliminated.

The dephosphorylated vector ends can still form two of the four phosphodiester bonds when ligated to other fragments. These partially joined circles are transformed at high efficiencies whereupon the annealed unjoined ends are repaired (ligated) in the host bacterial cells.

The elimination of vector background can also be achieved by directional cloning but there are limitations to its applicability in the available choice of restriction sites (see previous section).

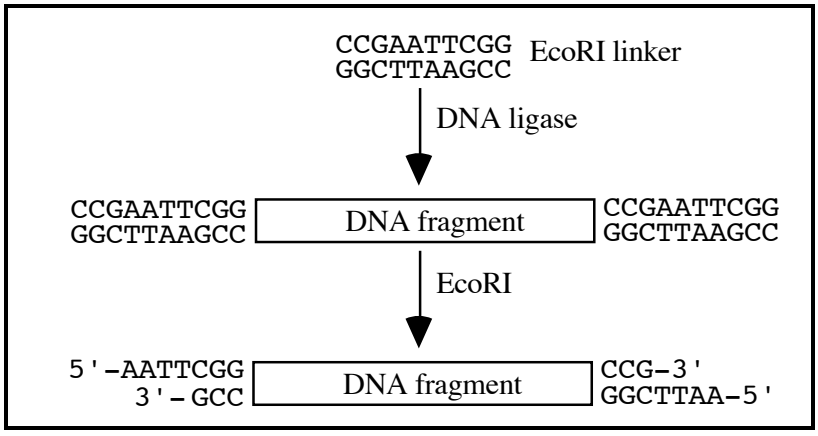


MODIFYING RESTRICTION SITES FOR CLONING

Linkers and Adaptors

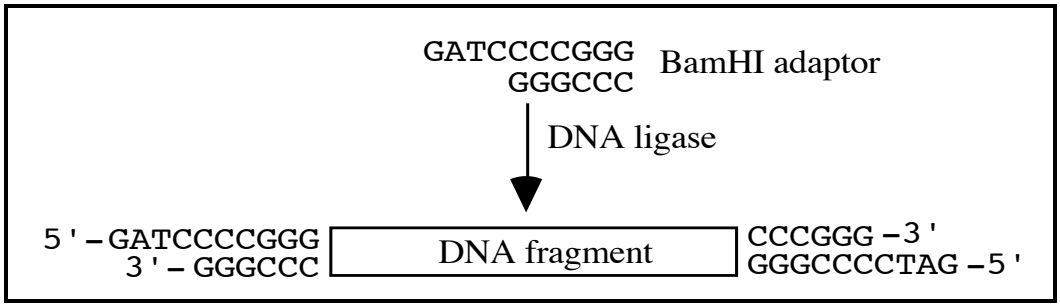
These are short DNA segments that are ligated to blunt-ended DNA in order to create a staggered end, or more commonly, to create a restriction site that is compatible with another site to which it is intended to be ligated. Synthetic linkers are short (8 to 12 bp) and can contain any one of the major restriction sites. They are ligated end-to-end very efficiently since they are present in such a high concentration in relation to the large fragment to which they are to be attached. If the linkers are dephosphorylated, only a single linker will be added to each end of the fragment. Following the ligation reaction and the heat inactivation of the T4 ligase, excess linkers are removed by the precipitation of the large fragment or separation in a column of Sephadex G-25 or similar resin. The ligated linkers are then pared back to the new site by digestion with the new site enzyme.

Linkers



Synthetic adaptors are preformed restriction sites that contain both blunt and staggered ends. When ligated to a DNA fragment, the new site is available for ligating to a similar site. One difference with the use of adaptors is that they are supplied in the dephosphorylated form so that tandem ligations are avoided. If not dephosphorylated, the ligated adaptors need to be digested with the appropriate restriction endonuclease.

Adaptors



Methylase protection of restriction sites

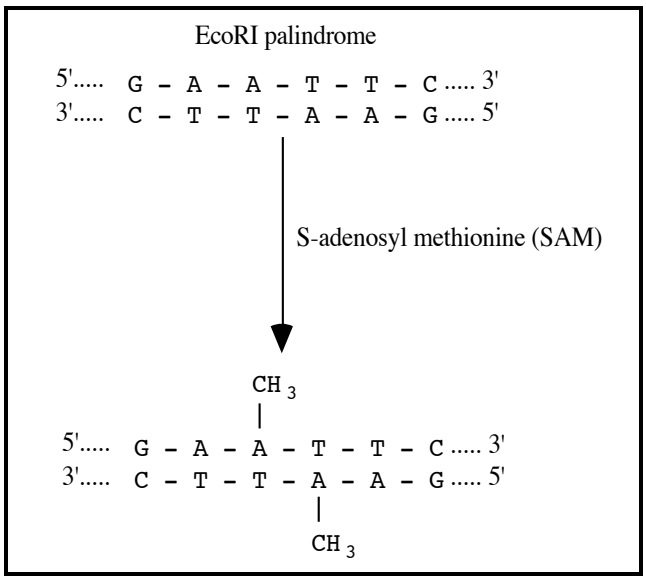
Methylases are enzymes that catalyse the transfer of methyl groups from a cofactor, S-adenosyl-methionine (SAM), to certain adenine or cytosine nucleotide bases in specific DNA sites. These sites are invariably restriction sites.

In bacteria, the methylases are used to add methyl groups to either adenines or cytosines within a specific restriction site. The bacterial cell also produces the restriction endonuclease that recognises this site. When the site is methylated, the enzyme cannot cut it and so the organism is protected from its own restriction enzyme. So why does the cell make the enzyme in the first place if it needs such an elaborate protection system against it? Read on....

This mechanism protects the cell against foreign, invasive DNA that might corrupt the fidelity of its genome. Since foreign DNA will not be methylated in the same pattern, the host restriction enzyme cuts the foreign DNA. Exonucleases then attack the fragmented DNA and digest it to completion. If the invasive DNA is from the same species, however, then the DNA is methylase protected in the same way as the host and recombination will be possible, since the exchange of DNA between the same organisms can be beneficial (selection and evolution).

Methylases have been found in eukaryotes. Since these are multicellular organisms and not exposed to invasive DNA, the protection mechanism is not required. However, eukaryotes have cleverly adapted the methylases for the regulation of gene expression. Control regions of genes are often methylated when transcription is undesirable and RNA polymerase is prevented from binding to the promoter DNA. At the correct time in the cell cycle, the DNA is de-methylated (activated), via a signal transduction event involving transcription factors, and transcription will take place normally.

Methylase protection is used to modify internal restriction sites. For example, EcoRI methylase catalyses the transfer of methyl groups from S-adenosyl-methionine to the two innermost adenine nucleotide bases in EcoRI sites (see diagram below). One application is the addition of EcoRI linkers to a DNA molecule that has one or more internal EcoRI sites. The strategy is firstly to methylate the internal sites, then ligate EcoRI linkers, and then digest with EcoRI to expose the new terminal EcoRI sites. The internal EcoRI site is methylase protected from this digestion.

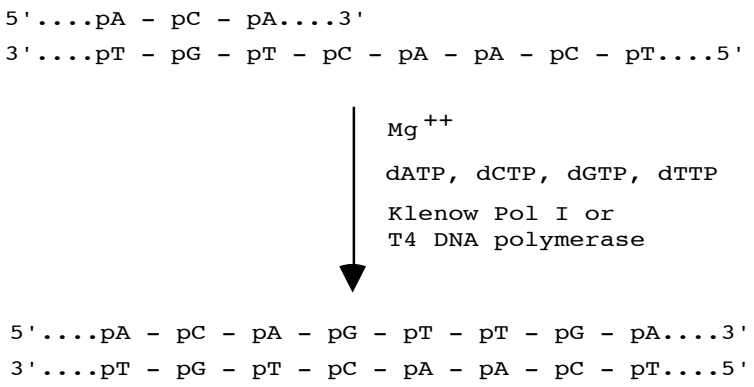


End-filling with DNA Polymerase I or T4 DNA polymerase

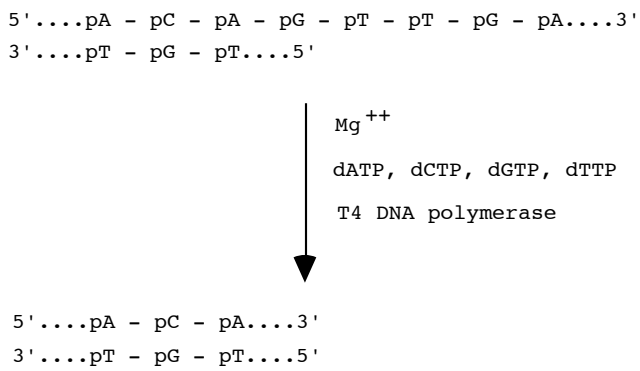
As in the last section, the reactions described here are designed to extend the range of options available for using restriction enzymes in molecular cloning. Even though there is a large number of restriction enzymes available, there may be limitations caused by particular cloning strategies. Many restriction endonucleases generate cohesive or staggered termini. If the need is to ligate fragments with staggered termini to vectors cut with enzymes producing blunt ends, or vice versa, then the staggered ends can be filled in using either Klenow DNA polymerase for filling in 3' ends; or T4 DNA polymerase for flushing 5' protruding ends or blunting 3' overhangs.

The Klenow fragment of DNA polymerase I (DNA Pol I) from *E. coli* consists of a single polypeptide chain (MW = 76,000). This fragment is produced by the cleavage of the DNA Pol I holoenzyme with subtilisin. This releases the 5' ----> 3' exonuclease moiety while the remainder - the Klenow fragment - retains the 5' ----> 3' polymerase and 3' ----> 5' exonuclease activities.

The major uses of the Klenow enzyme are for filling in the 3' recessed termini created by some restriction endonucleases (see below) or for labelling the termini of DNA molecules with ^{32}P-dNTPs in end-filling reactions for the creation of tagged, probe sequences.



Like the Klenow enzyme, T4 DNA polymerase possesses 5' ----> 3' polymerase and 3' ----> 5' exonuclease activities. However, the exonuclease activity is 200 times more active than that of the Klenow enzyme. For this reason, in addition to its main function of filling in recessed 3' termini as in the diagram above, it is also ideal for flush-ending recessed 5' ends (see below).



SITE-DIRECTED MUTAGENESIS

The method is called site-directed mutagenesis because the base to be modified is targeted by using a synthesised oligonucleotide that is identical to a short region of the gene sequence.

The major problem is to eliminate the background of plasmid molecules that would otherwise decrease the efficiency - by requiring extensive screening of transformants. There are several variations of site-directed mutagenesis that achieve this objective.

One of the most powerful tools in molecular biology is Site-Directed Mutagenesis in which base changes can be incorporated into DNA. These changes may encompass single or multiple nucleotide base substitutions, or deletions or insertions of one to many bases. The changes most often are designed to create or remove restriction sites or to change an amino acid at a particular position in a protein. The main use of the technique is to allow the study of structure-function relationships in proteins, chiefly by observing the effect of amino acid substitutions on structural elements or activity (specificity of substrate binding and catalysis).

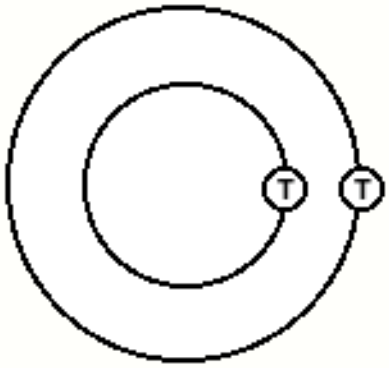
There are several variations of the method but one of the best ones - Stratagene's QuikChange method - is depicted on the next page. A common objective of the method is to change a single nucleotide within the gene sequence, usually in order to create or remove a restriction site or to bring about the substitution of one amino acid for another. The method is effective enough to tolerate several internal mismatches (see example below) provided that the primers are long enough and the ends are matched to the templates.

Beginning with a plasmid containing an inserted gene fragment, mutagenic primers are designed to straddle the site requiring the change. Ideally, the mismatched base should be near the middle of the primers so that the annealing is enhanced by the primers being perfectly matched to the template strands at both ends. The primers need not be precisely opposite one another, but should overlap significantly so that the same positional base change can be made in each strand.

Most of the factors affecting primer design discussed earlier for the PCR also apply here but there are some differences. In summary, these are:-

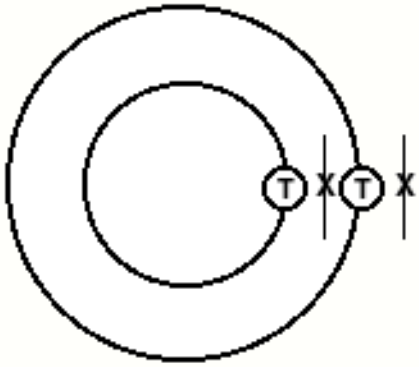
- primers should overlap either precisely or very nearly so
- primers should be between 25 to 45 bases in length
- melting temperatures (T_{ms}) of the primers should be equal to or greater than 78^{o}C
- desired mutation(s) should be near the middle of each primer with 10 to 15 bases of correct (perfectly matched to the template) sequence either side
- G or C at the 3'-end improves the yield
- primers should not be self-complementary or contain internal hairpin structure, or be complementary to one another near their 3'-ends

Step 1
Plasmid Preparation

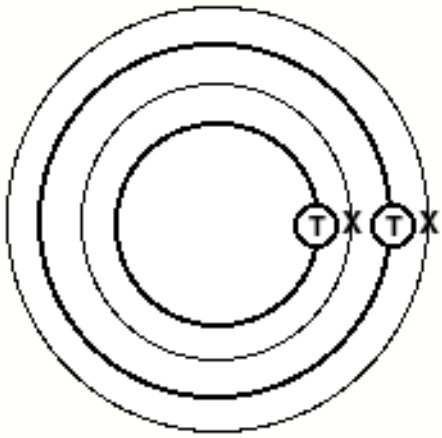


Gene in plasmid with target site (T) for mutation

Step 2
Temperature Cycling

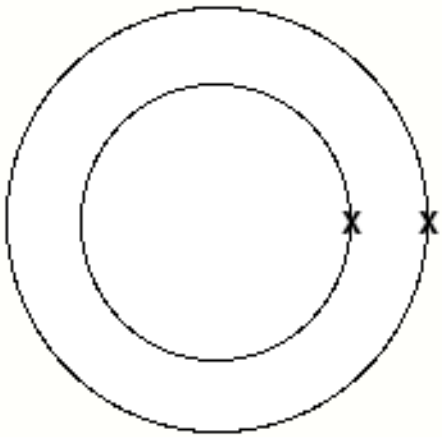


Denature the plasmid and anneal the oligonucleotide primers containing the desired mutation (X)



Using the nonstrand-displacing action of PfuTurbo DNA polymerase, extend the primers resulting in nicked circular strands

Step 3
Digestion and Transformation



Digest the methylated, parental DNA template plasmid with DpnI

Transform the circular, nicked ds DNA into competent cells

After transformation, the cells repair the nicks in the mutated plasmids

The claimed efficiency is greater than 80% of generated clones to contain the required mutation. The cycling protocol is typically as follows:

Cycles	Temperature	Time
1	95°C	30 seconds
12 - 18	95°C	30 seconds
	55°C	1 minute
	68°C	2 minutes / kb of plasmid length

The cycling protocol means that, for a plasmid of 4.5 kb for example, the 12 - 18 extension steps will need to be at least 9 minutes. Since the entire molecule must be replicated many times during the extension cycling - even though only a single base pair is being changed in a total of say 4500 base pairs - an enzyme with very high fidelity is used, namely PfuTurbo. As usual, the reaction mix is overlaid with mineral oil to prevent evaporation.

Following the PCR cycling, the reaction is cooled to 37°C and the restriction enzyme DpnI is added. This enzyme allows the method to enjoy its claimed effectiveness of greater than 80% of recovered clones to contain the desired mutation(s). The target sequence of DpnI is: 5'-G^{me}ATC-3'. DpnI, whose recognition sequence is commonplace in most DNA molecules, will digest all of the original (parental) DNA but none of the newly synthesised mutated DNA. This is because DpnI only digests DNA that is methylated on the adenine base of its recognition site which is the case for almost all laboratory *E. coli* strains which carry the *dam* methylase gene. The PCR synthesised DNA will not be methylated and is therefore not a substrate for DpnI.

As usual, the circular PCR-generated molecules will be transformed efficiently and the linearised parental template DNA will not. Miniprep plasmid DNA from several transformants is checked for the incorporation of the mutagenic base change. This can be done in one of several ways depending on the nature of the base changes. If a restriction site is added or removed, the plasmid DNA can be subjected to a digestion reaction with the enzyme. If the mutation brings about a single amino acid substitution, DNA sequencing is mandatory.

EXPRESSING FOREIGN (HETEROLOGOUS) PROTEINS IN *E. COLI*: EXPRESSION VECTORS

The major requirements for the expression of a cloned foreign gene in *E. coli* are that the:-

- **Gene must be placed under the control of an efficient *E. coli* promoter.**
- **mRNA must be stable and efficiently translated.**
- **Translated protein must not be rapidly degraded by proteases.**

These conditions are met by using specialised expression vector plasmids and *E. coli* strains.

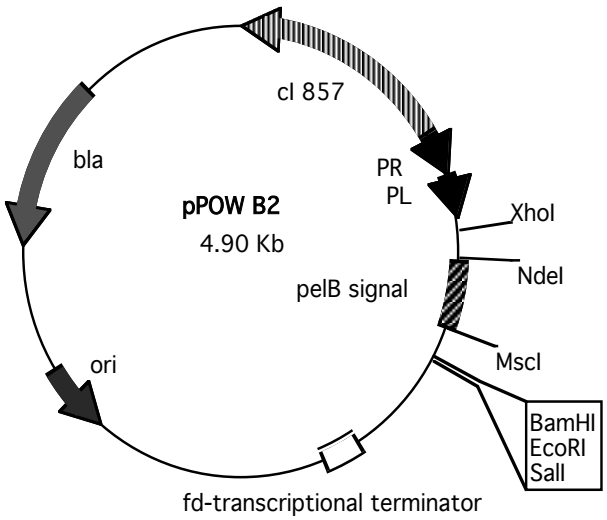
EXPRESSION VECTORS

An *expression vector* is one that has a promoter just upstream from a MCS, into which a foreign gene is ligated. Very often the foreign gene is from a eukaryotic source, in which case it will be a cDNA (complimentary) copy of the gene that has been produced from mRNA (Why must it be cDNA and not a chromosomal copy of the gene?). cDNAs do not have their own promoters and thus must be ligated in the correct orientation just downstream from the vector's promoter region.

pPOW B2 – an example of an expression vector

The expression vector below is one of many now available. It has ideal features of an expression plasmid. The chief advantages of pPOW B2 are its strong but tightly regulated lambda-derived promoters, that are under the control of the temperature-sensitive cI_{857} repressor gene.

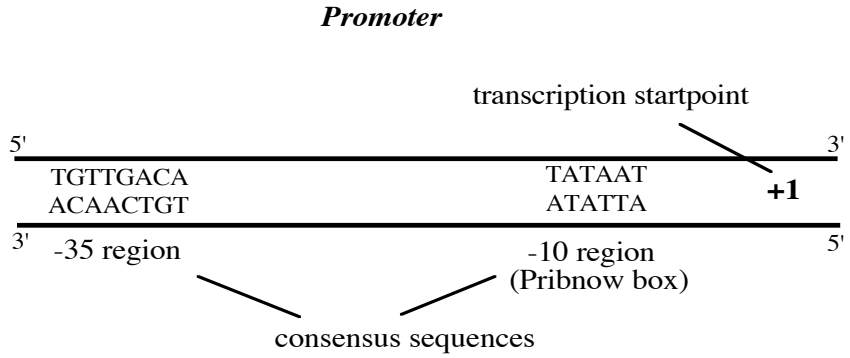
The transcriptional terminator is needed because the transcription is so efficient that it will continue into the origin of replication region (ori) and destabilise or destroy the plasmid. Also, the longer the transcript, the more likely it will be degraded by ribonucleases in the cell.



In most expression vectors, a *repressor* gene is located just 5'-prime (upstream) to the promoter site. Such repressor-promoter systems are inducible. Three of the best examples of these are the:-

- *lacI* (repressor) - *lacZ* (promoter), induced by IPTG (isopropyl-b-D-thiogalactoside)
- Stronger *trp-lac* promoter (*tac*), also induced by IPTG
- Tightly controlled bacteriophage lambda cI_{ts857} repressor - p_L promoter, induced at 42°C

A *promoter* is a short DNA sequence that binds RNA polymerase and therefore initiates (promotes) mRNA synthesis. Promoters in prokaryotic genes have two conserved regions located at about 10 bp (minus10 region; Pribnow or TATAAT box) and 35 bp (minus 35 region; consensus TTGACA) upstream from the transcription start point (+1).



The homology of the promoter regions to the consensus sequences and the distances between the regions determine the strength of the promoter. Many bacterial genes are controlled by weak promoters; and eukaryotic genes do not function in *E. coli*. Therefore, it is often necessary to insert genes downstream from strong, regulated promoters, in expression vectors.

A disadvantage of a strong promoter can be that the high level of transcription will cause instability of the plasmid or of the mRNA itself. This can be overcome by having a transcriptional terminator downstream from the MCS.

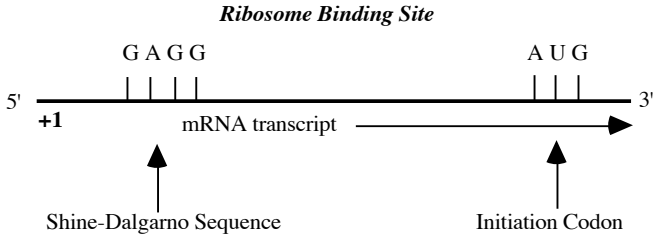
In some cases, a strong promoter will produce so much foreign protein as to be affect the survival of the host cell. However, the promoter strength can be moderated by only using partial induction, that is, shorter induction times or lesser amounts of the inducer molecule.

Ribosomal Binding Site

Efficient translation from mRNA requires a Ribosome Binding Site (RBS), also known as a Shine-Dalgarno or SD Sequence. This consists of an AG-rich sequence of 3-9 nucleotides located 3-11 nucleotides upstream from the initiation codon (ATG in DNA = AUG in mRNA).

The consensus RBS or SD sequence is part or all of 5'-UAAGGAGGU-3'. This sequence is partially complementary to the 3'-prime end of the *E. coli* 16S rRNA in the ribosomes. Overall, the effect of the SD sequence on the efficiency of translation can be affected by the:-

- Degree of complementarity between SD sequence and 16S rRNA.
- Spacing between the SD sequence and AUG codon.
- Identity of the triplet preceding, or the nucleotide following, the AUG start codon.



PROTEIN FOLDING AND STABILITY

In *E. coli*, many foreign proteins can be misfolded or rapidly degraded. For example,

- Newly synthesised protein cannot fold quickly enough into its native conformation. The polypeptide chains are degraded by proteases, or they become entangled into large clumps known as inclusion bodies. Sometimes, these inclusion bodies can be recovered, dissolved, and assembled correctly *in vitro*; or can be used for producing antibodies.
- The protein needs one or more disulfide bonds to complete its conformational folding. In *E. coli*, disulfide bonds are formed in the periplasm. Thus, many of these proteins will be misfolded and degraded by proteases before they can be exported to the periplasm; or could be degraded in the periplasm by proteases that are present in that compartment.

There are a variety of solutions to these problems:-

- Use *E. coli* strains that are deficient in proteases, reducing the degree of degradation.
- Use protease inhibitors in lysis and purification buffers.
- Use *E. coli* strains or plasmids that co-express an accessory foldase or chaperone protein.

In the bacterial cell, misfolding of polypeptides is prevented by two classes of accessory proteins known as foldases and chaperones. Foldases act to accelerate rate limiting covalent steps in folding. Chaperones mainly provide an environment for nascent proteins to fold without the competing process of self-association. These proteins bind to and usher nascent proteins during folding and delivery to sites inside or within the cell envelope.

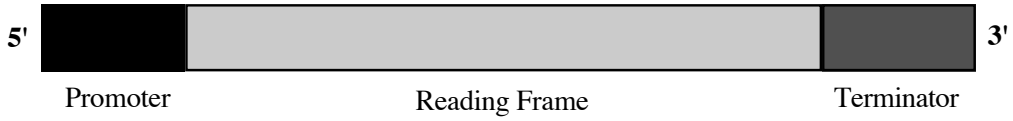
THE GENETIC CODE [CODON BIAS]

First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (Ochre)	Stop (Umber)	A
	Leu	Ser	Stop (Amber)	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Codon bias

Many eukaryotic genes use certain codons in preference to others. Some of these codons are rarely used in *E. coli*. The rarest codons are two out of the six arginine codons (AGA and AGG), and codons for leucine (CUA), isoleucine (AUA) and proline (CCC). The low abundance tRNAs for these codons cause backing up and slowing of the translation of nascent polypeptides at ribosome sites. This enables ribonucleases and proteases to attack and degrade mRNA and the polypeptides. The solution is to use a modified *E. coli* strain that has additional copies of the genes for the rare tRNAs, on a host plasmid.

TRANSLATIONAL READING FRAME



The **reading frame** consists of a number of **triplet codons**. The first is the **initiation** or **start** codon (usually methionine, ATG) and the last is one of three **stop** or **nonsense** codons (TAA,TGA,TAG). In between lies the number of codons that specify the amino acids of the polypeptide encoded by the gene.

```
  M   G   D   F   S   G   Q   L   K   T
ATG GGG GAC TTT TCA GGA CAG CTG AAA ACC TGA
Start                                 stop
```

This series of triplet codons must be **in-frame**, when a **hybrid** or **fusion** protein is constructed. If not, then the frameshift changes the triplet sequence.

```
1:        THE DOG AND THE CAT
2:         HED OGA NDT HEC AT
3:          EDO GAN DTH ECA T
            DOG AND THE CAT


 M   G   D   F   S   G   Q   L   K   T  stop
ATG GGG GAC TTT TCA GGA CAG CTG AAA ACC TGA

  W   G   T   F   Q   D   S  stop
 TGG GGG ACT TTT CAG GAC AGC TGA AAA CCT GA

   G   G   L   F   R   T   A   E   N   L
  GGG GGA CTT TTC AGG ACA GCT GAA AAC CTG A

    G   D   F   S   G   Q   L   K   T  stop
   GGG GAC TTT TCA GGA CAG CTG AAA ACC TGA
```

Any change in the reading frame may cause the polypeptide to be truncated (shortened) by introducing a premature stop codon; or, will otherwise change the identity of the "protein" so completely as to make it useless. These types of products are dysfunctional and are degraded by the cell since they cannot fold correctly or completely.

FUSION (HYBRID) PROTEINS

Reasons for generating fusion proteins are:-

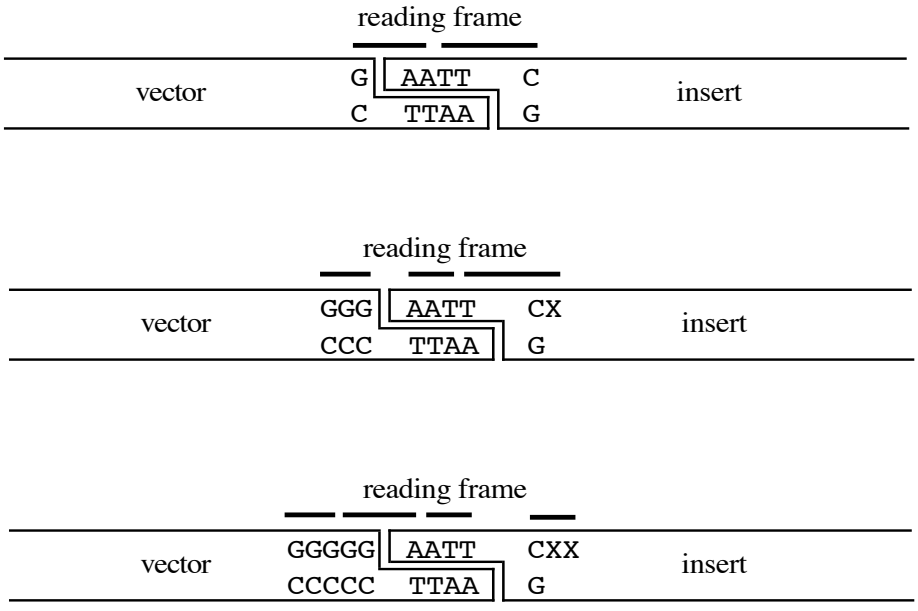
- Many fusion proteins are more stable in *E. coli* than the native protein
- Fusion proteins can be easily purified by affinity chromatography or immunoprecipitation
- Simple chemical treatments will cleave the fusion site and release the native protein

The most commonly used fusions are:

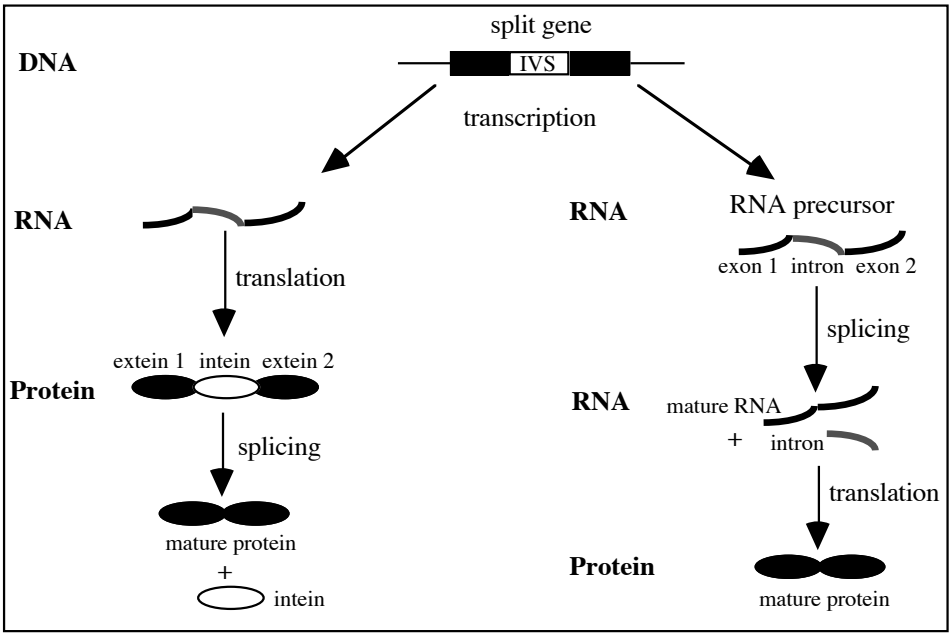
- Reporter enzymes (Beta-galactosidase; beta-lactamase; alkaline phosphatase; chloramphenicol acetyltransferase; glutathione-S-transferase). These not only provide one component of the hybrid construct but also serve as purification aids (affinity chromatography) and as indicators of expression by enzymatic assay.
- IMPACT-intein C-terminus fusion. Chitin binding domain for purification and cleavage from the intein C-terminal moiety.
- Short peptide epitopes, e.g. i-Spy QYPALT system (antibody affinity purification) or Histidine N- or C-octapeptide tag (affinity chromatography with Ni^{3+}).

When constructing a fusion hybrid in an expression vector, the most important consideration is that the fused ends of the hybrid must be in-frame in order to preserve the codon identity of the two segments of the fused polypeptide. This is achieved by using one of the three expression vectors provided by the commercial supplier. The three vectors will have the specific cloning site in the three possible reading frames. If the nucleotide sequence of the gene for the protein of interest is known, the correct vector can be chosen. If the sequence is unknown, then all three vectors must be used.

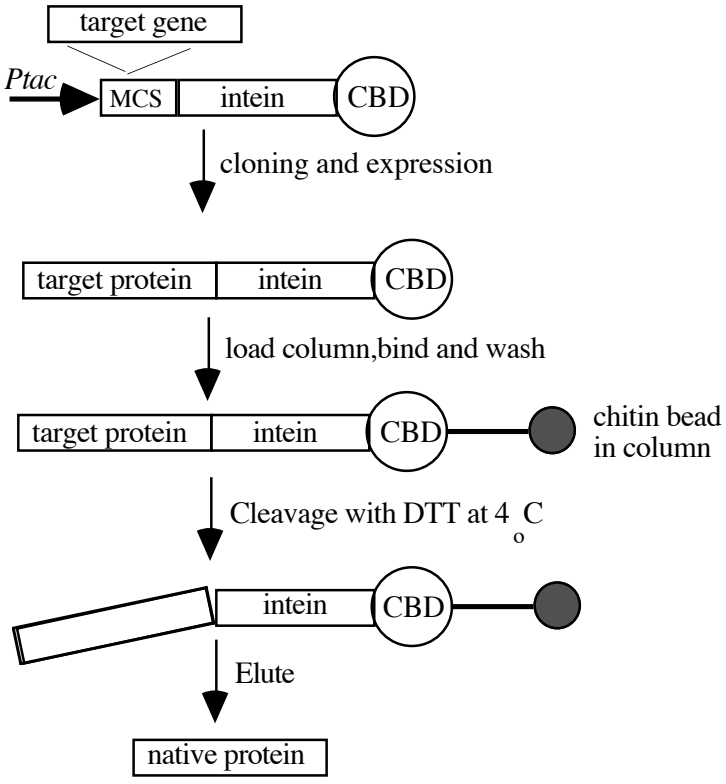
The different reading frames with respect to the translational initiation site of the lacZ gene presented by three different versions of the same vector



Splicing mechanisms in Prokaryotes and Eukaryotes



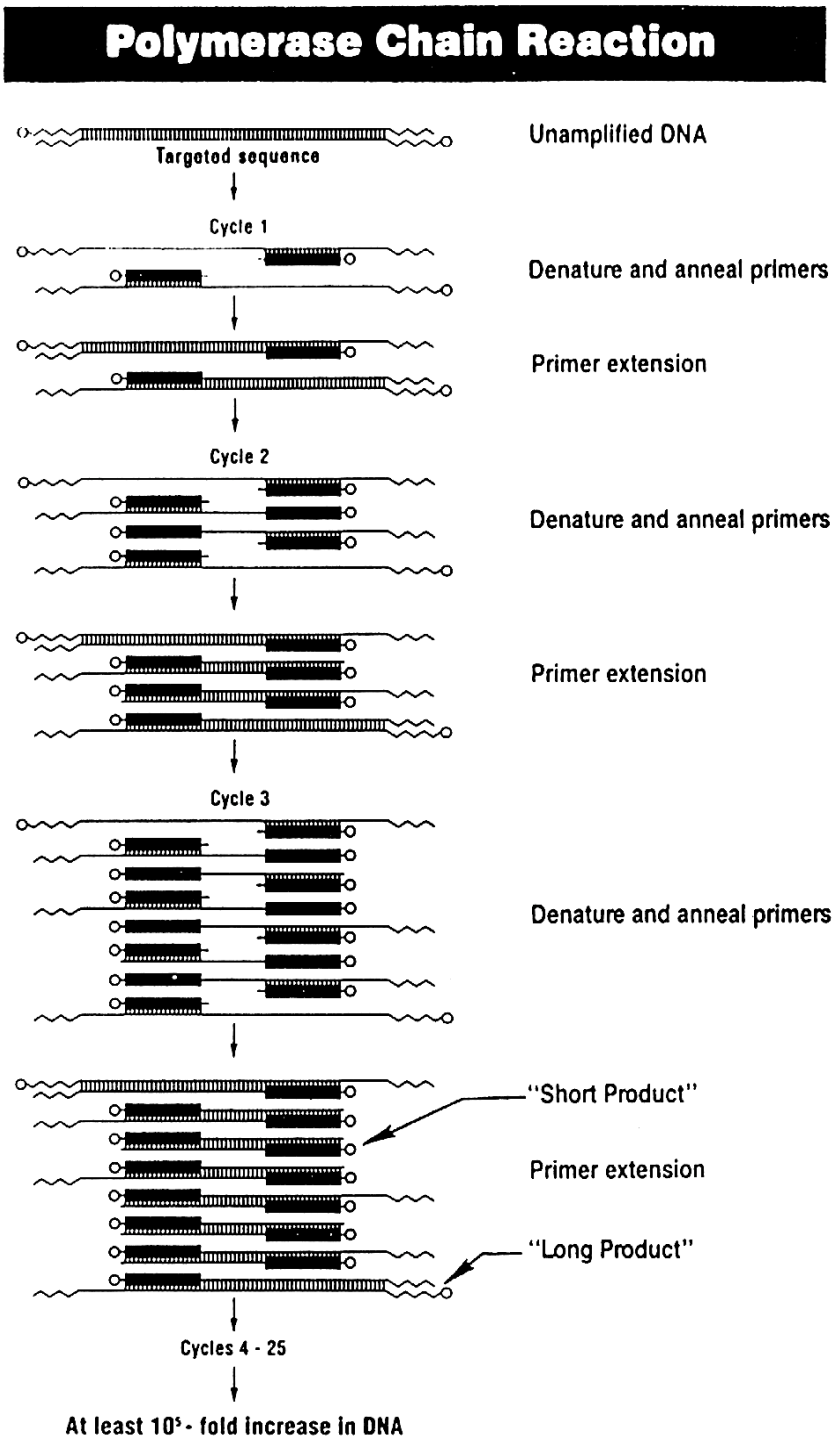
IMPACT Affinity Purification System



A schematic illustration of the IMPACT System. The target protein precursor is purified on a chitin column from a crude cell extract. Cleavage is induced by the addition of DTT and incubation at 4°C overnight. The target protein is eluted while the intein-chitin binding domain (CBD) fusion remains bound to the column.

POLYMERASE CHAIN REACTION

The Polymerase Chain Reaction (PCR) was devised in 1987 and has since become a major molecular biology method. It works on the principle that a given piece of double-stranded DNA can be replicated and amplified repetitively in a two-fold manner. The means by which this can be achieved depends upon flanking primers and an efficient DNA polymerase.



BASIS OF THE METHOD

The PCR derives its utility from the discovery of thermostable DNA polymerases in bacteria whose habitats include deep ocean volcanic trenches and hot springs. Commercial examples are Taq, Pfu and Vent polymerases. Pfu DNA polymerase has about 10-fold more fidelity (accuracy) than Taq polymerase (error rate of about 10 misincorporations per 10^6 bases). These enzymes have reaction optima of 72° to 74°C and as such are relatively resistant to heat denaturation. This led to the design of thermal cycler machines and the PCR protocol for the amplification of DNA.

Thermal cyclers have microprocessors that control heating and cooling coils that are capable of ramping the temperature of the tube holder block 10s of degrees Centigrade up or down in seconds. The rapid temperature changes are required because the optima of the three reaction steps are at different temperatures. These are approximately: 95°C for the denaturation of the double-stranded template DNA; 55°C for the primer annealing step; and 72°C for the polymerase extending step. There is an absolute requirement for a thermostable DNA polymerase due to the long periods of heating at high temperatures.

The PCR reaction uses a set of single-stranded oligonucleotide primers (18 to 30 bases) that flank a sequence of interest. A single reaction mixture can be set up to contain all the components needed for an amplification reaction (buffer, 4 x dNTPs, Mg^{2+}, template DNA, primer pair, thermostable polymerase, water). It relies on the use of a thermostable DNA polymerase (commonest examples are: Taq, Vent, Pfu) and a series of heating and cooling cycles that amplify the template DNA - between the primer pair - from a low number of copies (tens, hundreds or thousands) to many millions of copies in 25 to 35 cycles (refer to the table below for a guide to the approximate number of cycles versus the starting number of target molecules).

Number of target molecules	Number of cycles required
3 x 10^5	25 to 30
1 x 10^4	30 to 35
1 x 10^3	35 to 40
50	40 to 45

The entire process takes about 3 to 4 hours and is performed in a multi-slotted temperature cycler. Many different reactions can be done simultaneously in separate, thin-walled 0.5 ml plastic tubes. Because of the high temperatures used in the cycling, the reaction mixes need to be covered with a layer of oil to prevent evaporation.

How much DNA versus the source for a typical PCR? From the table above, about 10^5 to 10^6 target molecules is best. This is approximately 1 μg of human genomic DNA, 10 ng of yeast DNA or 1 ng of *E. coli* DNA. In order to appreciate the sizes of genome and numbers of DNA molecules, refer to the table below for a brief overview of quantification.

DNA quantity and type	Number of molecules
1 μg of 1 kb of ds DNA	9.12 x 10^{11}
1 μg of pUC18 (= 2.686 kilobase pairs)	2.5 x 10^{12}
1 μg of *E. coli* chr DNA (= 4000 kb)	2 x 10^8
1 μg of human chr DNA (= 2.8 x 10^6 kb)	3 x 10^5
5 mm blood spot (= ~ 5 μg DNA)	1.5 x 10^6
hair root (~ 10 - 200 ng)	3 x 10^3 - 6 x 10^4

Thermal cycling for the amplification reaction comprises three steps. In the first, the double-stranded DNA template is denatured at 95°C, then the reaction is cooled to 55°C to allow the primers to hybridise to their respective complementary sequences in the template strands, and finally the temperature is raised to 74°C to enable the DNA polymerase to synthesise new DNA along the template strands. This extension begins from the 3'- end of each primer.

Note that since the strands of a DNA molecule are antiparallel (5' ----> 3' and 3' <---- 5'), the annealed primers 'point' toward one another and only the DNA between them is amplified (see diagram overleaf). The conditions of PCR reactions can be adjusted to amplify short or long sequences. Originally, the length of amplifiable sequence was restricted to a range of about 200 to 800 base pairs. Currently, more than 20,000 bps (> 20 kb) may be reliably amplified.

A typical PCR reaction consists of the following steps in each cycle:

Step 1: Denaturation of template at 95°C.
Step 2: Annealing of primers at 56°C.
Step 3: Polymerisation or extension of templates at 72°C.

Each step is performed for 30-60 seconds (steps 1 and 2) and between 1-10 minutes for step 3 (depending on the length of template). The thermal cycler ramps the temperature between steps by rapid heating or cooling (usually taking up to 30 seconds). Reaction conditions (temperatures, length of primers, concentrations of dNTPs and Mg^{2+}) can be modified to improve the specificity and fidelity of the amplification process, but the one fixed parameter is the Taq or Pfu extension reaction. This needs to be performed at an optimum of 70 to 74°C for the enzyme. PCR reaction products are checked by running 10% of the mix in an agarose gel to detect a product of the required size. The reaction mix can then be used to clone the product into a vector, or can be sequenced, using the same amplification primers as sequencing primers.

FACTORS AFFECTING PCR

Hot-start procedure to avoid non-specific priming

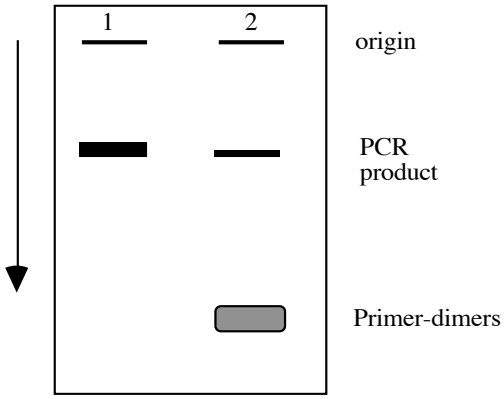
A 'hot-start' procedure is designed to increase to fidelity and efficiency of the reaction by adding the enzyme after the reactants have been heated beyond 70°C. This prevents the primers from annealing at room temperature to sites other than their precisely complementary sequences. Several base-pair mismatches can be tolerated at room temperature and the primers may thus bind to many sites in the very long chromosomal molecules. The reaction mix is overlaid with mineral oil then heated in the cycler to 95°C for one minute. Following addition of enzyme to the 'hot' reaction mix, a thermal cycling protocol such as the one indicated below is commenced.

Cycle 1	step 1	95°C	2.00 min
	step 2	55°C	0.30 min
	step 3	72°C	4.00 min
Cycle 2 (X 29)	step 1	95°C	1.00 min
	step 2	55°C	0.30 min
	step 3	72°C	4.00 min
Cycle 31	step 1	72°C	10.00 min
	step 2	4°C	30.00 min.

It takes approximately 3 to 4 hrs to reach step 2 in cycle 31. The reaction tubes are cooled to stop the reaction. The mineral oil is removed by adding an equal volume of chloroform which forms an immiscible bottom layer. Dye mix is added to the top layer and 10% of it is run in an agarose gel to check the product against linear DNA standards.

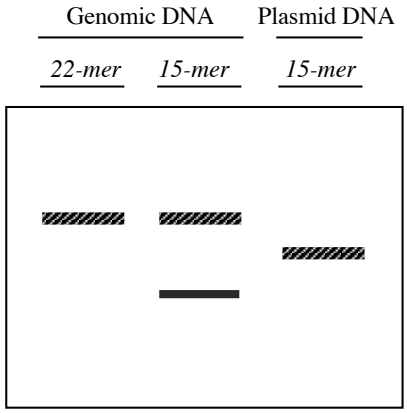
Primer design

It is an important condition of PCR reactions that the primer pair chosen must not be self-complementary, particularly at the 3'-end. If there is any complementarity, then primer-dimers will form and will be amplified, thus reducing the effective concentration of the primers and ultimately the product yield. Primer concentration can be too low (too little product) or too high (which leads to mispriming and a number of secondary products of various sizes.)



Primers must be designed to contain between 40-60% GC content, not have runs of Cs or Gs at the 3'-end, and have melting points (T_{ms}) that are close together so that priming is evenly achieved at the annealing temperature. Primers should not have internal hairpin structure. Exact matching of primer and template is more important at the 3'-end than at the 5'-end. Why?

Primer length can be an important consideration. Usually, a primer length of 18-30 bases is optimal. Theoretically, a primer of 18 bases represents a unique sequence amongst $4^{18} = 7 \times 10^{10}$ nucleotides, and should hybridise at only one position in most eukaryotic genomes (with complexities of 10^9-10^{10} base pairs). A shorter primer such as a 15-mer would hybridise at more than one site in a complex genome, but not in say a smaller bacterial genome or plasmid molecule (see gel profile below). Remember especially, that priming at multiple sites can occur if the annealing temperature is too low because a degree of mismatching can be tolerated.



APPLICATIONS OF PCR

Amplification of specific genomic DNA sequences (100 bps to > 20 kbps). This general use of the PCR has been discussed above.

Thermocycled DNA sequencing of double-stranded plasmid DNA. This method makes use of both the amplifying and extending properties of a thermostable DNA polymerase. The sequence bounded by the facing primers is continuously amplified and extended during repeated cycling (denaturation, annealing and extending). Since the dNTP mix is supplemented with chain terminating ddNTPs, single-stranded extensions from each primer will be terminated and amplified at each nucleotide position beyond the primers. In this way, the product bands - tagged with a fluorescent dye - will be detected in the DNA sequencing gel loaded with the reaction mixes. The real advantage of cycle DNA sequencing is that it allows direct sequencing of DNA cloned into a plasmid. The heating cycles enable denaturation of the double-stranded plasmids thereby overcoming the tendency of the plasmid strands to renature. No subcloning is required.

RAPD (Random Amplified Polymorphic DNA) PCR for the detection of new species-specific sequences. An assay for the detection of DNA sequence polymorphisms using random short primers in a PCR amplification reaction. Thus, the term 'RAPD' stands for Random Amplified Polymorphic DNA. The great utility of this technique is that it is not reliant on having any DNA sequence information about the target DNA. The method can be used to construct genetic maps within or between species. An assortment of primers with random sequences of a standard length - usually 8 - to 12-mers - is added to the template DNA along with the other components of a PCR reaction. Because of the desire to generate a series of products from different regions of very long target DNA molecules (usually full-length chromosomes), and because the primers are short, the annealing step is carried out at a lower temperature, in the range of 35^{o} to 40^{o}C instead of 55^{o}C. A typical reaction will generate a number of products of various lengths. These are identified in an agarose gel, excised and recovered from the agarose and cycle sequenced. In this way, genetic maps or specific probes can be constructed from template DNA whose DNA sequence was previously unknown.

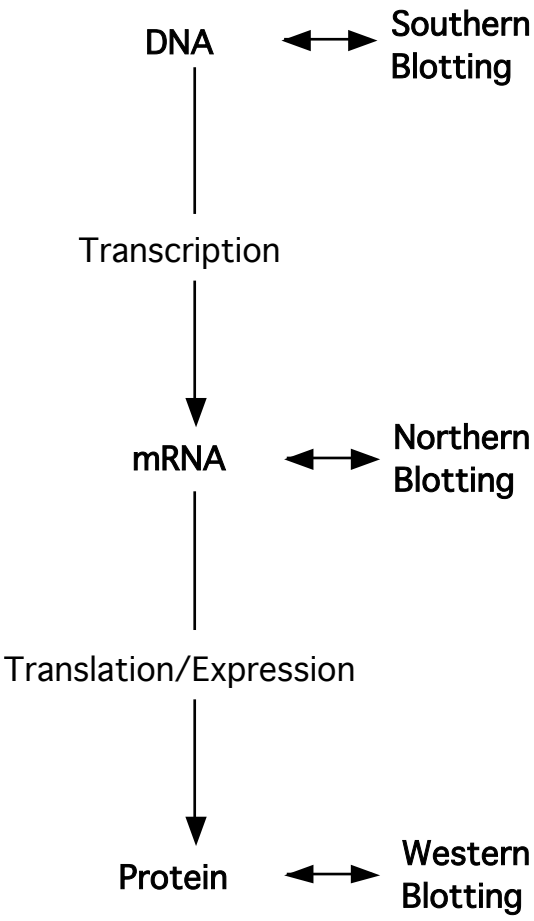
DNA fingerprinting to identify individual-specific sequences. This was dealt with in the earlier lecture on Southern Hybridisation. Briefly, hypervariable, tandemly repeated DNA sequences occur throughout the genomes of humans. This means that individuals will have an assortment of different restriction digestion site patterns (Restriction Fragment Length Polymorphisms or RFLPs). This pattern is identifiably closer in related individuals but more distinctive in distantly-related or unrelated persons. The way in which DNA is assorted when gametes are fused gives rise to the RFLPs. VNTRs (Variable Number of Tandem Repeats) are also used in DNA fingerprinting or profiling.

PCR in genetic diagnosis. Previously, the diagnosis of many genetic diseases meant taking a large blood sample (~20 ml) and a 4-5 day protocol involving DNA extraction, restriction, and southern hybridisation with specific radiolabelled probes. With PCR, it is possible to perform diagnoses in under 10 hours using such innocuous samples as a mouthwash or a drop of blood from a pinprick. The method relies only on a suitable set of primers that will amplify a defective gene or sequence. Some inherited disorders diagnosed by PCR diagnosis includes:

β-thalassaemia	Haemophilia A and B
Cystic fibrosis	Huntington's chorea
Duchenne muscular dystrophy	Phenylketonuria
Aplha-1-antitrypsin	Sickle cell anaemia
Myotonic dystrophy	Familial adenomatous polyposis

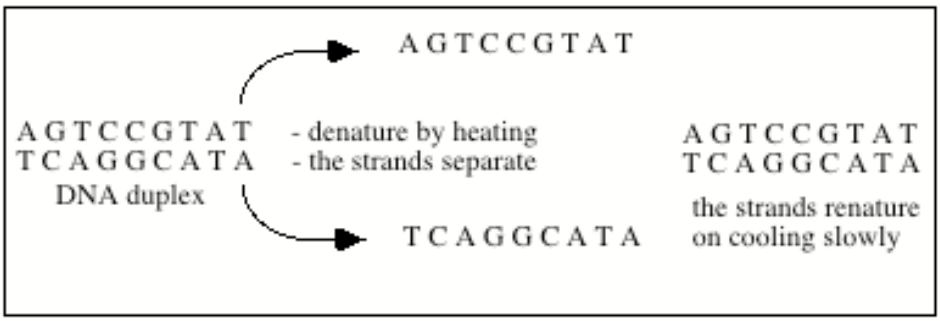
SOUTHERN HYBRIDISATION: DNA PROBES

Summary of the three blotting/hybridisation schemes



The method of Southern Hybridisation is based upon the nucleotide homology between a DNA probe and DNA target sequences. Hybridisation occurs under defined conditions between a denatured DNA probe in solution and denatured DNA target sequences immobilised on a membrane filter. Detection is by radiolabelled (^{32}P) probes, or by enhanced chemiluminescence (ECL) using Horse Radish Peroxide (HRP) labelled probes. Probes can be single- or double-stranded. Synthetic oligonucleotide probes are always single-stranded.

The Principle Of Hybridisation

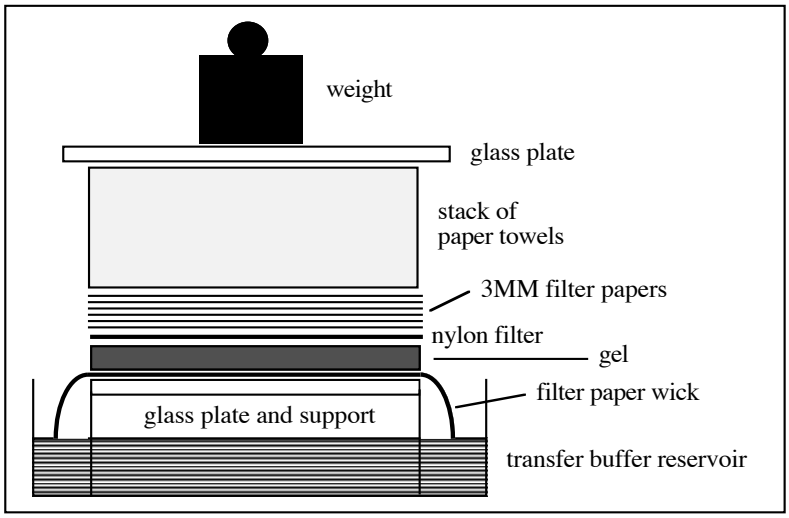


SOUTHERN BLOTTING

Target DNA can be DNA fragments in agarose gels, or clones or plaques seeded on agar plates, or DNA samples from any source.

Southern transfer / blotting to nitrocellulose or nylon membranes. After the tranfer in a low salt or high salt buffer, the DNA is fixed covalently to the membrane by alkali treatment and UV cross-linking. The blotting step can can done in alkali.

Setting up a Southern Blot



Notes:

- Nylon membrane does not require pre-wetting, has greater physical strength than nitrocellulose, and is more resistant to alkali. Nylon filters retain DNA better for re-probing. Disadvantages are high background and less efficient binding of small fragments (< 200 bps).
- Transfer buffer can be the same buffer used for the gel electrophoresis, or 20 x SSC (especially for RNA transfers), or alkali / salt (0.4N NaOH / 1.5M NaCl).
- The transfer of large fragments (> 8 kb) can be assisted by prior treatment of the gel with 0.5M HCl for 15 min, or by exposure of the gel to UV light.
- This type of capillary transfer set-up is the original version and is still very reliable if used for overnight blotting. There are now many alternatives including bi-directional blotting of the gel onto two membrane filters, or positive pressure apparatuses that use forced air or a vacuum to complete the transfer in 30 to 60 min.
- The transferred DNA is bound more strongly to a nylon filter if the transfer buffer is alkali / salt. Or, the DNA can be cross-linked to the nylon if it is placed in a microwave or UV oven for a short exposure. This preserves the DNA for subsequent probings of the same filter.

PREPARATION OF DNA PROBES

The length of a probe can be from 15 bases to several kilobases, but ideally should be no longer than several hundred bases for high specific activity - and specificity - of the labelled probe. Probes are labelled by one of three methods:-

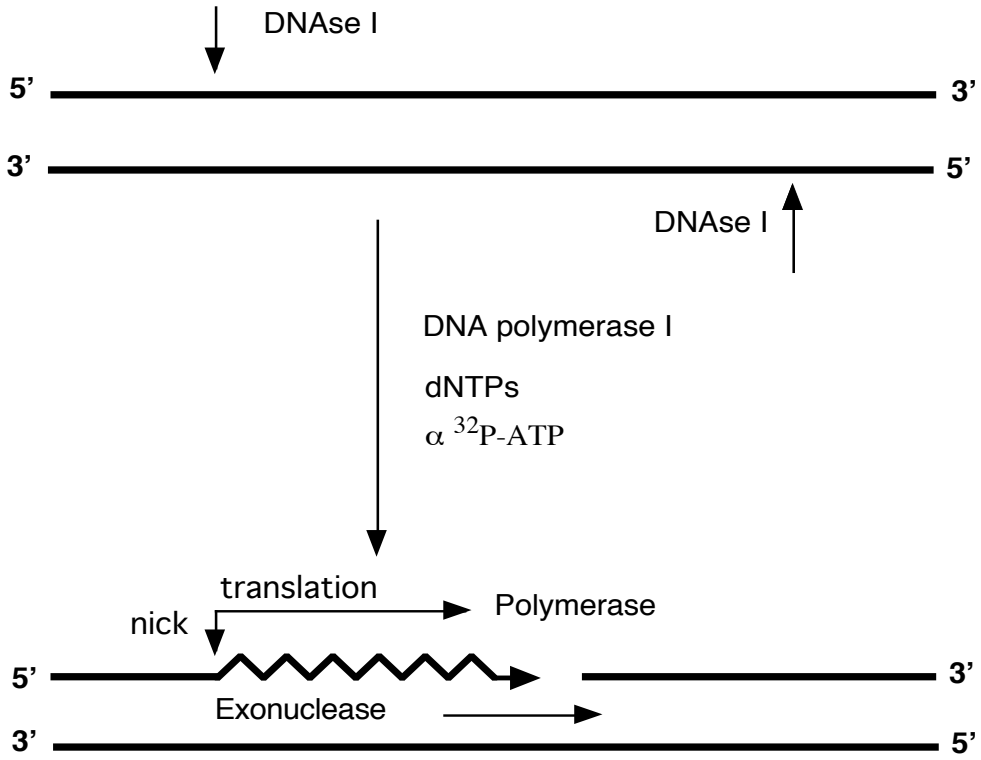
- **End-labelling** (for oligo probes < 50 bases)
- **Nick translation**
- **Random primer extension**

The probe synthesis methods incorporate radiolabelled ^{32}P-dATP or ^{32}P-dCTP (the latter gives slighter 'hotter' probes). The probe-target hybrids are detected eventually by autoradiography with X-ray film.

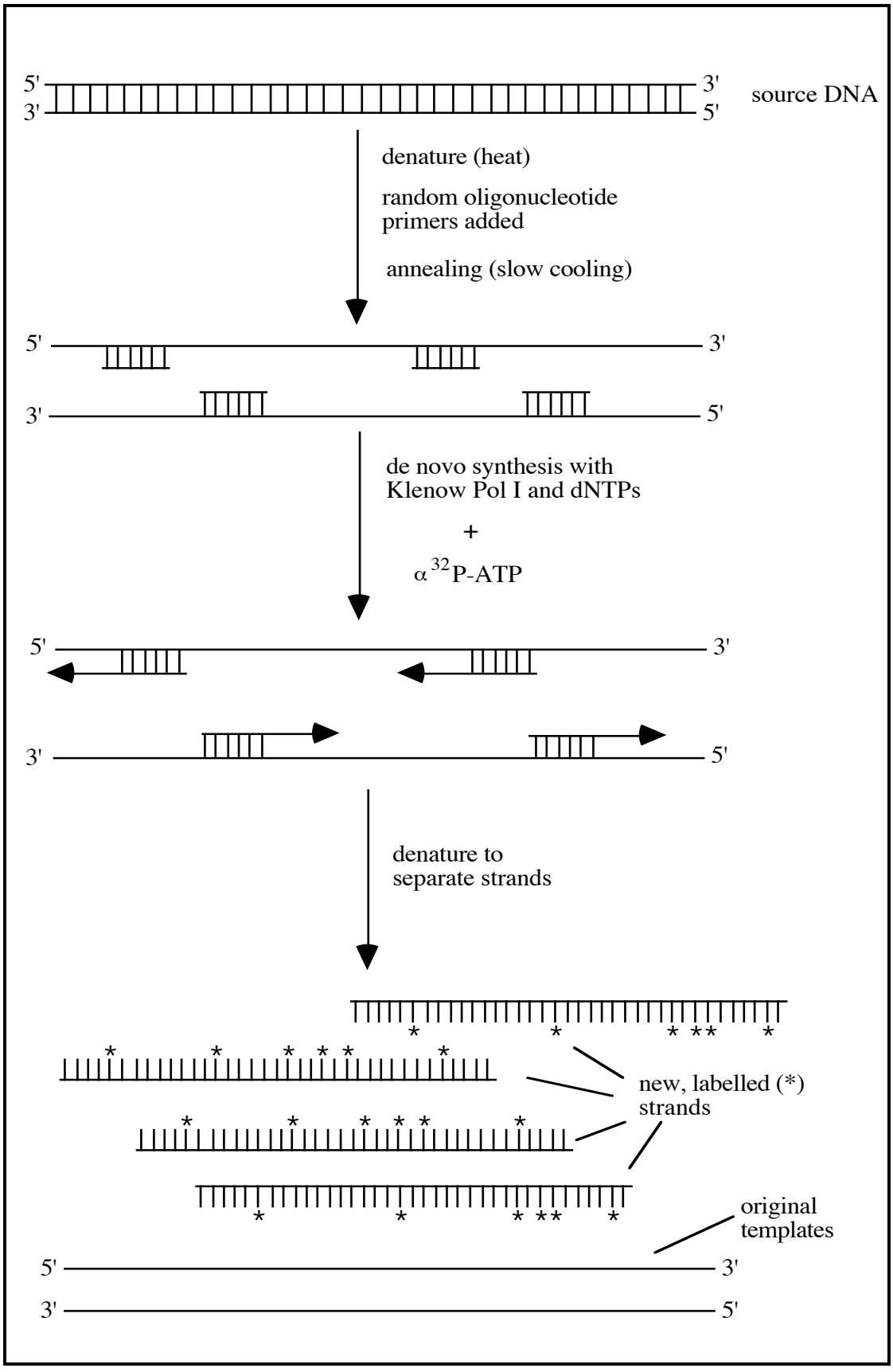
End-labelling

The end-labelling method uses polynucleotide kinase and γ^{32}P-ATP to carry out an exchange reaction between the labelled phosphate and the 5'-end phosphates of the short oligonucleotide strands. This results in every strand being radiolabelled with one γ^{32}P at each 5'-end.

Nick Translation



Random Primer Extension



PARAMETERS AFFECTING HYBRIDISATION REACTIONS

- Temperature
- Volume
- Time
- Ionic strength
- Stringency
- Accelerants
- Background

Temperature

The rate of reannealing of any DNA hybrid is maximal at 20-25^{o}C below the melting temperature (T_m) of the DNA. Under these conditions, hybrids formed between completely homologous nucleic acids are thermally stable. The hybrid formation reaction competes with reassociation of the probe's complementary strands. Every 1% base mismatch between two DNAs lowers the T_m of the hybrid molecule by 1.4^{o}C. Hybridisations are carried out at 65^{o}C with the probe in a formulated buffer solution containing accelerants and blocking agents.

Volume

The volume of a reaction should be as small as possible to improve the kinetics of hybrid formation. The membranes must be completely wetted at all times. Continual movement of the probe solution across the membrane filters is unnecessary.

Time

Most reactions require at least 6 hours but for convenience are run overnight (16 to 20 hours).

Ionic Strength

At low ionic strength (< 0.1M), nucleic acids hybridise very slowly, but as the ionic strength increases, the reaction rate increases (Why?). A good average is 0.5M NaCl.

Stringency

The discrimination of exact or closely-related sequences is made by controlling the stringency of the hybridisation and washing conditions. The lower the stringency, the more is the likelihood of detecting distantly-related sequences. Stringency is controlled by temperature and ionic strength in the hybridisation reaction and washes. Hybridisation depends more upon the nucleation frequency (hybrid formation), whilst washing depends on the thermal stability of the hybrids. Washing is also carried out to remove unhybridised probe and to dissociate unstable hybrids.

Accelerants

Accelerants of the hybridisation reaction include dextran sulphate or polyethylene glycol at 10%. These polymers work by excluding probe from that portion of the solution occupied by the polymers, thus effectively increasing the probe concentration, whilst minimising the volume.

Background

Background "noise" on the membrane filters needs to be suppressed, especially when using nylon membranes to which DNA binds quite tightly. Agents that reduce the background include any one of: Blotto (non-fat dried milk powder); Denhardt's reagent (containing SDS, BSA, and Ficoll); or denatured, fragmented salmon sperm DNA. These are used in pre-hybridisation buffers to block the non-specific attachment of the probe to the membranes, and also in hybridisation buffers to prevent non-specific DNA associations.

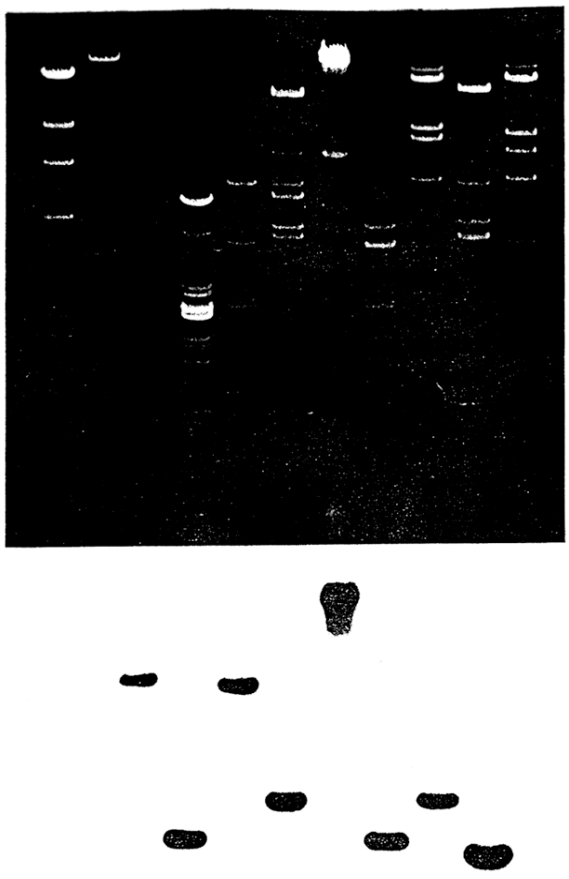
APPLICATIONS OF SOUTHERN HYBRIDISATION

Southern Hybridisation of DNA in Agarose Gel Blots

Southern hybridisation can be employed to analysis or detect sequence diversity in closely- or distantly-related species. The only differences required for detecting more distantly-related DNA are the use of lower stringency or degenerate probes (with one or more base changes). In this method, the DNA that is run in the agarose gel (before blotting and hybridisation) is a genomic DNA restriction digest that will give a 'smear' of DNA fragment bands right down the gel lane.

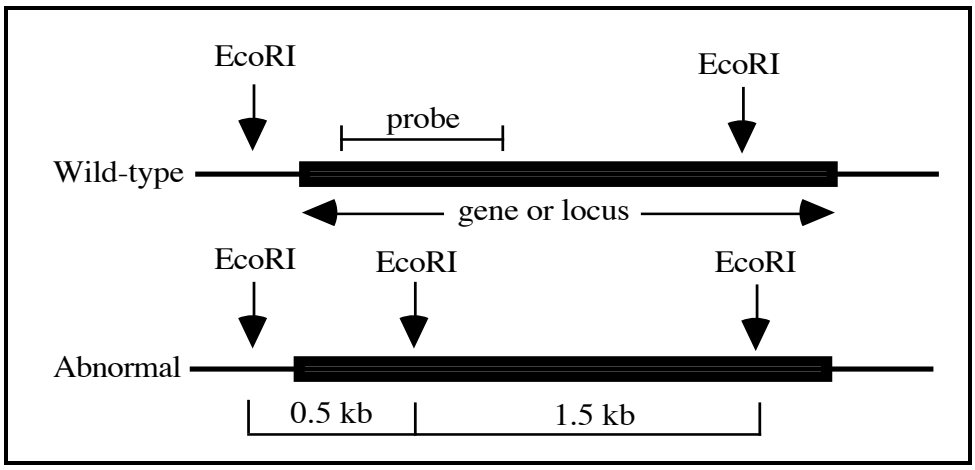
In the example below, the upper panel is a facsimile of an EtBr-stained gel depicting restriction fragment patterns for the same plasmid molecule cut with different restriction endonucleases. The bottom panel is a facsimile of the autoradiogram of the DNA blotted from the gel onto nylon and hybridised to a sequence-specific DNA probe. Note that the position of the probe-fragment hybrid bands are mostly in different positions in each of the lanes. This is because although the probe is the same, the restriction fragments that contain the probe-specific sequences are of different lengths.

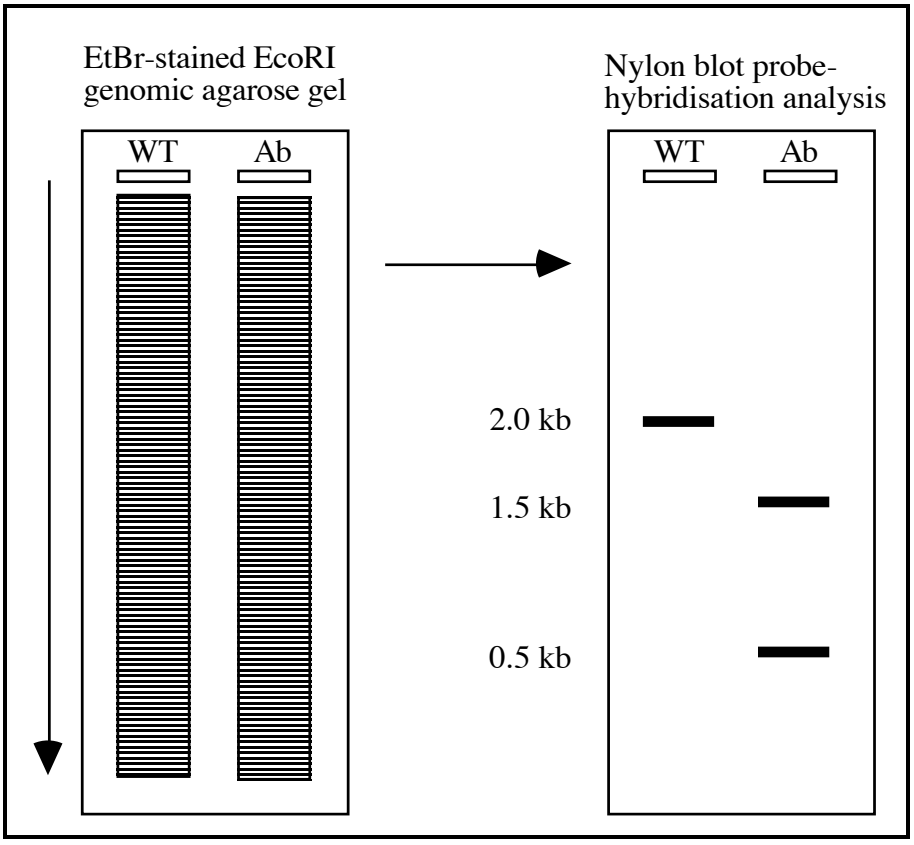
Southern Hybridisation of Agarose Gel Blots



Restriction Fragment Length Polymorphisms (RFLPs)

This technique is especially useful for the detection of genetic errors or abnormalities, or for the fine mapping analysis of alleles. It is a variation of southern hybridisation that relies on differences in a restriction endonuclease pattern at a particular locus that is a candidate for genetic or other errors that give rise to defective genes. A probe that straddles the polymorphic site is required. In the example below, the abnormal region has an additional EcoRI site that can be identified in a genomic digest-blot-hybridisation (bottom figure) using the probe indicated in the top figure. Most genetic abnormalities map to specific sites that can be characterised by restriction mapping to identify typical RFLPs, as in the example below.





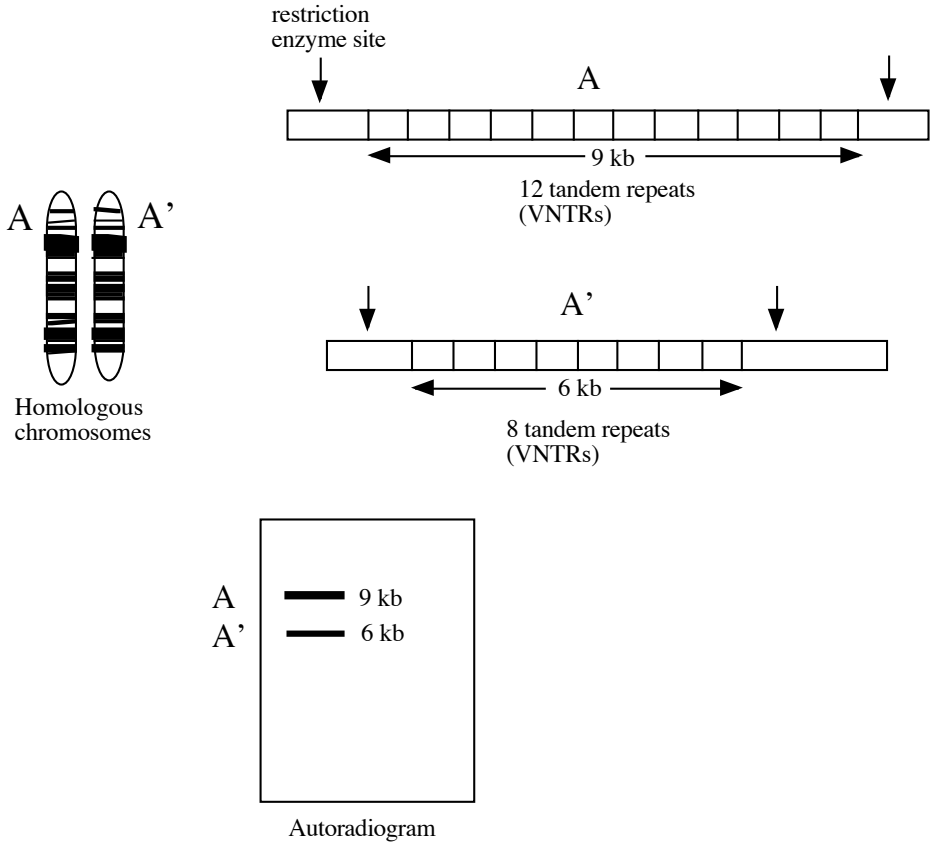
DNA Fingerprinting

This is a variation of southern blotting in which a probe (many commercial types are available) is used to identify homologous sequences in genomic DNA from different individuals. The homologous sequences are scattered throughout the genomes in DNA regions known as hypervariable satellite DNA. It is organised into highly repetitive clusters of nearly identical tandem sequences of as few as two to as many as several hundred base pairs depending on the species, is repeated many times, and is concentrated about the centromeres of chromosomes.

The hypervariable regions arise from unequal exchanges during replication, resulting in an alteration of the number of tandem repeats in any given 'minisatellite', which are thus highly polymorphic and individual specific. The tandem repeats have been shown to be inherited in a Mendelian manner. These regions are called **VNTRs** (Variable Number of Tandem Repeats). By definition, VNTRs are copies of identical-sequence DNA fragments arranged in direct succession (contiguity) within a chromosome. The number of copies varies in a random fashion.

In the example below, two alleles, A and A', occur at a specific locus on a pair of homologous chromosomes. The number of tandem repeats at these loci are different for each chromosome. If the same restriction sites occur at either end of the repeats, the restriction fragments will be of different lengths. This can be detected in the subsequent autoradiogram of the DNA profile.

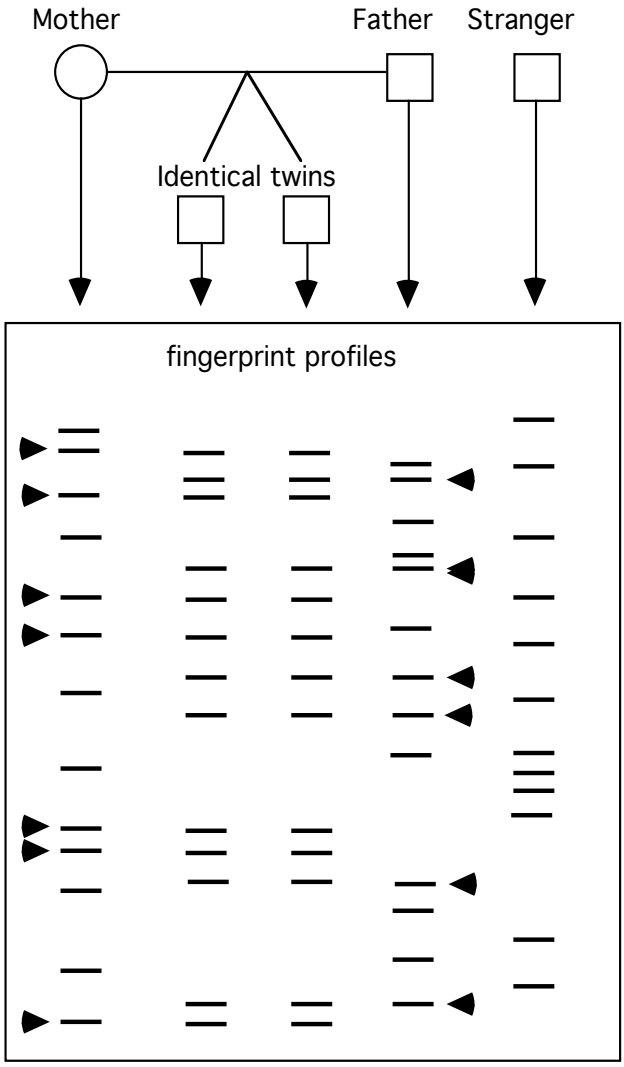
VNTR Analysis



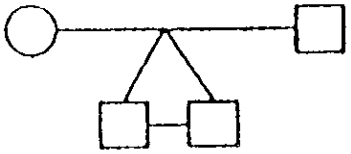
Any suitable restriction endonuclease will produce a family of fragments (RFLPs or VNTRs), each containing a number of tandem repeats of the core sequence. A short region of the highly conserved core sequence serves as the DNA probe to 'fingerprint' the polymorphic, genomic DNA. The two types of probes used in DNA fingerprinting are single-locus and multi-locus. The major difference between single-locus and multi-locus probing are:

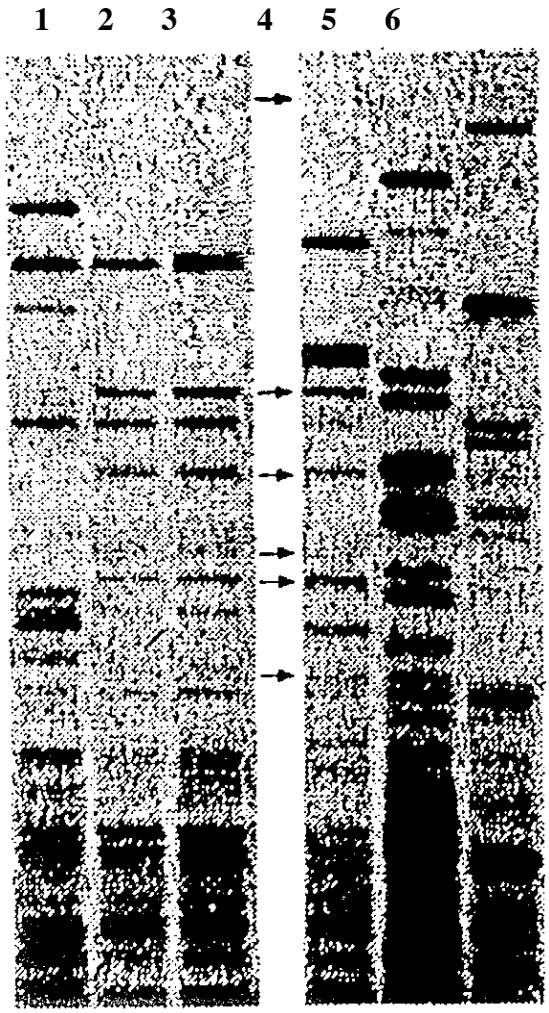
- Single-locus profiling requires far less target DNA but PCR has largely reduced this difference.
- Single-locus profiling produces only one or two bands, versus many for multi-locus profiling.
- Single-locus probes are species-specific. Multi-locus probes may cross-hybridise to other species.
- Single-locus profiling requires four or more different probes / profiles for confirmation. Multi-locus profiling only requires one or two probes.

DNA Fingerprinting (Schematic)



DNA Fingerprinting (Actual)





NORTHERN HYBRIDISATION

EXPRESSION AND REPRESSION OF GENES AND OPERONS

In terms of the normal operations of cells, many genes are expressed constitutively - these are the so-called 'housekeeping genes', for example, genes that are involved in DNA repair or the utilisation of glucose and generation of ATP.

Most other genes are only switched on when required, for example, at different stages of embryonic development, or a family of genes such as those within a regulated operon in a bacterium. The regulation is effected by the presence of a repressor gene (that is itself in the operon) which is the only working (expressed) gene in the absence of the substrate used by the operon. The repressor protein binds to the single promoter-operator for all of the structural genes of the operon. If the substrate is suddenly made available to the cell - for example lactose for the lactose operon - the repressor protein's conformation is changed when lactose binds to it. This causes the repressor to detach from the promoter site and allows access to RNA polymerase which then binds to the promoter and begins synthesis of an mRNA transcript.

RNA AND NORTHERN BLOTTING

The principle of northern blotting is essentially the same as southern blotting. Before describing the details, it should be stressed that the rationale behind the carrying out of northern blot experiments is to establish that a gene is being expressed by identifying its mRNA transcript.

The expression of a gene is detected as a unique mRNA transcript in a northern blot.

Northern blotting differs from southern blotting mainly in the preparation and electrophoresis of the target molecules. The target molecules are mRNAs that are isolated along with the bulk of ribosomal and transfer RNAs. The method is sensitive enough not to require any further fractionation of the total RNA except for the subsequent separation achieved in a denaturing agarose gel. However, in eukaryotic cells, mRNAs contain poly(A) tails that enable them to be separated from other RNA by oligo-dT affinity chromatography, if desired.

The DNA probe preparation, transfer (blotting) of RNA, hybridisation, washing and detection are carried out just as for southern blot experiments; except that alkaline transfer cannot be used.

Since RNA does not stain as intensely as DNA and since mRNAs from any cell or tissue type vary in size (according to the length of the genes from which they were transcribed), it is not possible to observe a complete profile of mRNA bands in an agarose gel.

However, the abundance of ribosomal RNAs (eg. 23S, 16S and 5S in prokaryotes) enables these major species to be detected in ethidium bromide stained gels as three distinct bands (see practical experiment 2). In general, the successful isolation of total RNA, as indicated by the intensity and compactness of the three major RNA bands in a denaturing agarose gel, is reasonably good evidence that the mRNA preparation is also satisfactory.

As discussed in an earlier lecture, the need for care in the isolation of total RNA is due to the following factors:

- Many mRNAs may be under-represented in the preparation since only a small number are transcribed in the particular cell type or time in the cell cycle.
- RNA is single-stranded and is more likely to be degraded by phosphodiester bond breakages.
- RNAs are easily and rapidly degraded by ribonucleases - from the same cells or the laboratory.

A northern blot experiment would involve the following sequential steps:

- Extraction of total RNA (see earlier lecture)
- Purification of mRNA (optional; for eukaryotic cells only)
- Electrophoresis in a denaturing agarose gel (see earlier lecture and practical expt. no. 2)
- Blotting or transfer to a nylon membrane
- Preparation of a DNA probe (to be denatured by boiling if double-stranded)
- Pre-hybridisation blocking of the membrane
- Hybridisation after addition of the probe
- Stringency washing
- Detection

In most of the above steps, the procedure is the same as for southern blotting and hybridisation.

RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction)

Northern blot experiments can achieve two outcomes. Namely, the detection of expressed mRNA transcripts and the size determination of the transcripts.

A method of quantifying gene expression that has become increasing popular is RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction), in which short primers are used to amplify mRNAs in a PCR reaction. The product is then run in an agarose gel and detected by ethidium bromide staining.

Note, this method does not involve northern blotting or hybridisation but since its intention is to detect expressed mRNAs, it is appropriate to make mention of it in this section. Also, RT-PCR is only used for the detection of mRNAs and is not used for estimating the size of mRNAs.

WESTERN BLOTTING

Western blotting represents the third method of following the sequence,

GENE -------> mRNA -------> PROTEIN

Although we have just seen how an expressed gene can be identified by the detection of the mRNA transcript in a northern blot, it is also possible to determine the expression of a gene by the detection of the translated protein by the method of 'Western Blotting'.

However, western blotting is only possible if an antibody probe is available (see below). This means that the protein must have been isolated previously and used to prepare an antibody, or that an antibody is available to a fused moiety on the (hybrid) protein of interest. This could be any one of a number of gene fusions, for example, peptides (8 x His), GST or LacZ tags.

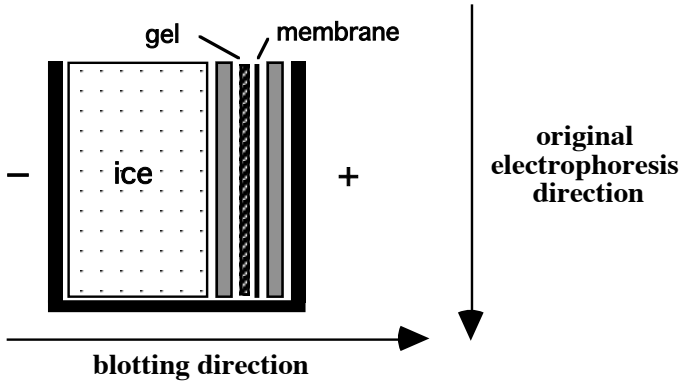
In western blotting, the target set is a profile of cell proteins that are subjected to electrophoresis in a polyacrylamide gel, and the probe is a specific antibody - ideally a monoclonal antibody - to an epitope of the protein or fusion tag.

Just as for the other blotting methods, this technique also involves the blotting of the target proteins from the gel onto a nylon or nitrocellulose membrane. However, one difference in this case is that the proteins must be electroblotted onto the membrane. This is because their small size would cause extensive diffusion if the capillary blotting method was used.

In a typical western blotting experiment, proteins are prepared from cells or cell compartments (eg. mitochondria, chloroplasts, nuclei, membranes etc). If many cases, genes cloned into expression vectors can be induced so that the protein of interest is produced in large quantities. In some instances, the protein may have been engineered as a fusion protein, that is, a hybrid of the protein plus a large (eg. LacZ; GST) or small ligand (eg. a histidine tag). This will allow affinity purification prior to electrophoresis.

Proteins are subjected to electrophoresis in polyacrylamide gels with size standards and positive and negative controls. A second, identical gel is usually run simultaneously and stained with Coomassie Blue to depict the complete protein profile and lane-to-lane loadings bias, if any. This objective may also be accomplished - albeit with less sensitivity - by staining the master gel with Ponceau S, which doesn't interfere with the subsequent western blot detection.

The target gel is electroblotted in the same electrophoresis apparatus using a sandwich assembly which transfers proteins to the membrane filter horizontally (see diagram on the next page), whereas the original PAGE gel electrophoresis was conducted vertically.



The electrotransfer takes 1 to 2 hours. Subsequent steps require about 1 to 2 hrs each. These are:

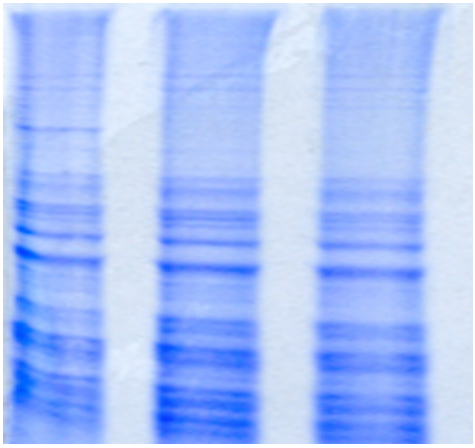
i. Membrane blocking of protein attachment sites with milk powder or BSA, to reduce background.

ii. Incubation in a solution containing the specific antibody.

iii. Washing cycles (three of these using large volumes) to remove non-specifically adsorbed antibody. A second antibody may be applied in those instances in which the detection method needs to rely on the second antibody reacting with colorimetric or chemiluminecent reagents. In cases in which a fusion protein is being detected, the primary antibody to the fusion tag (LacZ etc) will usually also carry the enzyme ligand for the detection reaction.

iv. Detection reactions include fluorescent antibody tags or the use of alkaline phosphatase or horse radish peroxidase and colorimetric or chemiluminescent reagents.

SDS-PAGE gel stained with Commassie Blue



Western blot / hybridisation of the proteins in the gel above