MOLECULAR BIOLOGY

- Albinism arises out of a defect in a single gene
- The consequences are a defect in the biosynthesis of melanin, a pigment
- Thus, geneticists define the albino gene as a heritable defect in melanin synthesis
- Melanin are aromatic compounds. Inability to synthesise melanin. For genetists. Recessive classic mendelian
- The final step in melanin synthesis is the conversion of tyrosine to melanin
- For molecular biologist--> what is the biochemical pathway that synthesise melanin. In albino there is no tyrosinase activity. Genetists have no view of the molecular basis. But biochemists cannot understand the hereditability. Must combine those views together.

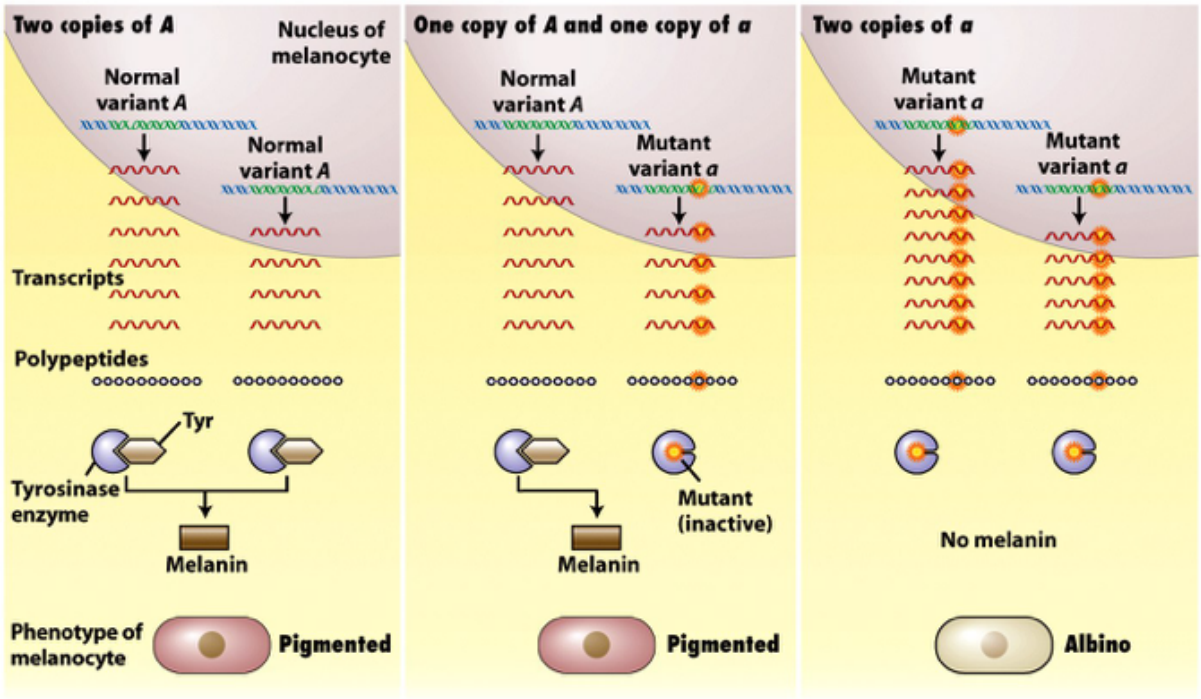


Figure 1-8 part 3
Introduction to Genetic Analysis, Ninth Edition

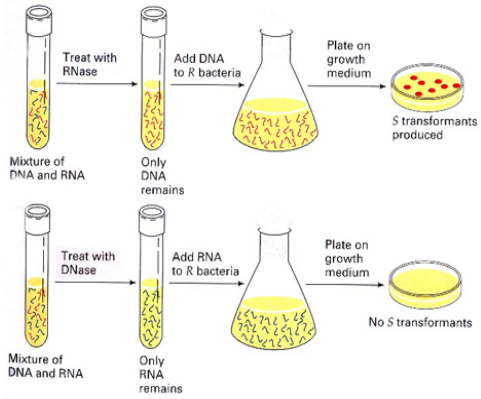
In modern times, genes are defined by neither geneticists, nor biochemists but by computers. The complete DNA sequence of human and other animal, plant and microorganism genomes vastly increases our knowledge of genes

HISTORY OF DNA

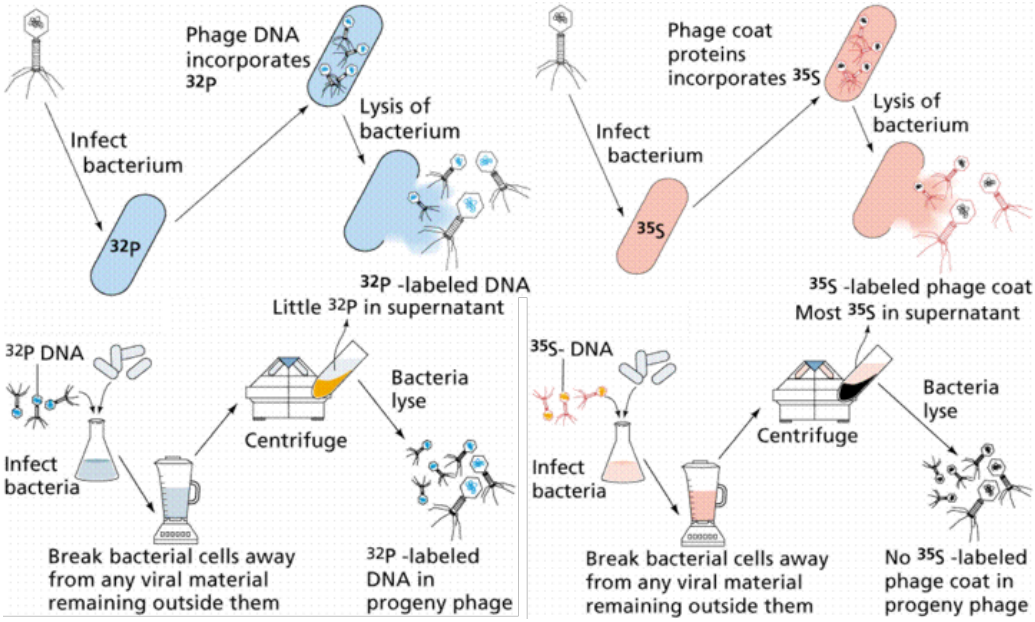
- In 1928, Frederick Griffith shows that *Streptococcus pneumoniae* can be transformed from the “rough” R, to the “smooth” S, phenotype

- In 1952 Alfred Hershey and Martha Chase used a kitchen blender to prove the genetic role of DNA in Bacteriophage
- Despite the convincing data of Avery, McCarty and Macleod, people still didn’t believe in DNA - it took the experiments of Hershey and Chase to convince the sceptics - they were

part of the “in crowd” at Cold Spring Harbor where Max Delbruck and his followers established the phage group which were to many the origin of molecular biologists.



- Hershey and Chase had difficulty getting the agitation conditions right to separate viral particles from infected bacteria. It wasn’t until they borrowed Margaret MacDonalds kitchen blender that the experiment was a success. All of the DNA of the phage has been transmitted to the offspring and no proteins.



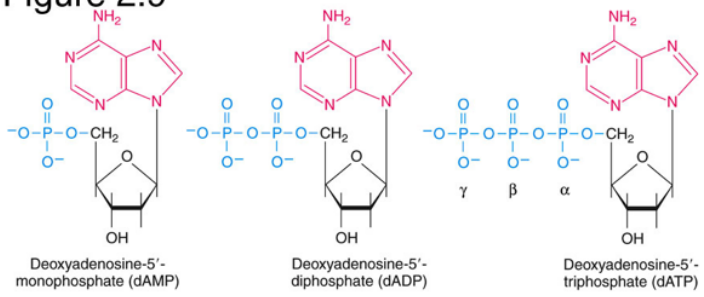
- Erwin Chargaff carefully purified DNA from human (various tissues), yeast, bacteria and cow and after chemical analysis, estimated the composition of the different nucleotide bases. This provided strong additional evidence that DNA was the hereditary material, since these species are clearly different from each other and they have different amounts of the bases.
- Chargaffs 1950 paper helped identify DNA as the hereditary material, but also showed a vital clue as to its structure.
- He showed that the GC content of DNA from different species varied, but that the A content was always roughly the same as the T content and that G = C
- In a race against Linus Pauling, Watson and Crick suggested that the X ray diffraction data of Wilkins and Franklind was best explained by an anti-parallel double helix
- Chargaffs rules were elegantly explained by specific base pairing of G with C and A with T. This was achieved by proposing specific hydrogen bond pairs between the bases - three for a GC pair and two for an AT pair.

PRIMARY STRUCTURE OF NUCLEIC ACIDS

- Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are chemically very similar.
- They are linear polymers composed of varying sequences of four different building blocks called nucleotides.
- A nucleotide is composed of

1. A **pentose sugar**
    - In RNA the pentose sugar is ribose, a five-carbon sugar
    - (Carbons are numbered 1', 2', etc to distinguish from carbons in the bases)
    - Arranged as five-membered ring with 1'C linked by oxygen to 4'
    - …with the fifth carbon, 5'C attached as a side chain to 4'C
    - In DNA the pentose sugar is 2'-deoxyribose with a -H group rather than an -OH group attached to 2'C
    - The planar ring is subject to steric strain, relieved by puckering such that the 2' or 3' carbon is out of the plane.
2. A **phosphate moiety**
    - Free nucleotides have phosphate groups attached via phosphoester bonds (P-O-C) to the side chain 5'C.
    - One phosphate - nucleotide monophosphate (eg AMP)
    - Two phosphates – nucleotide diphosphate (e.g. ADP)
    - Three phosphates – nucleotide triphosphate (e.g. ATP)

Figure 2.9



3. An **organic nitrogenous base**
    - The 1'C is joined by a ß glycosidic linkage to the base, which is in the plane above the sugar ring.
    - They form planar rings.
    - Bases are of two types: purines and pyrimidines
    - Purines have a double ring structure (five membered ring fused with six membered ring). Adenine (A) and guanine (G) are purines.
    - Pyrimidines have a single six-membered ring structure. Cytosine(C), thymine (T) and uracil (U) are pyrimidines.

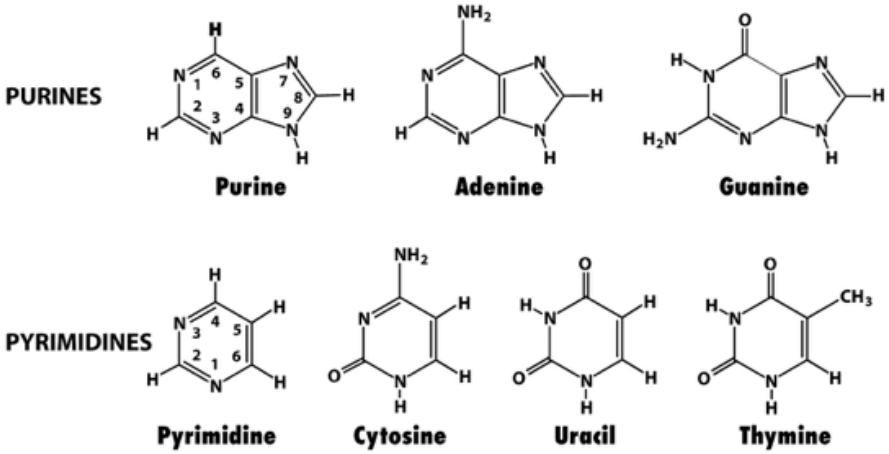
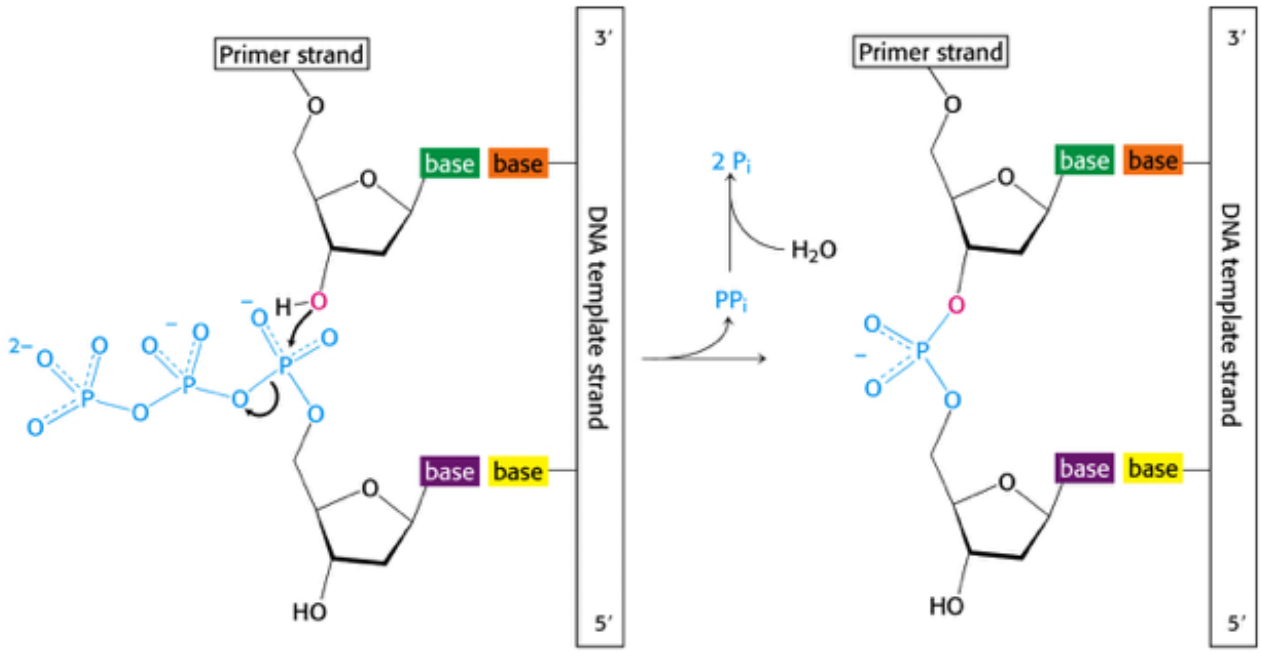


Figure 4-4
Biochemistry, Sixth Edition

- The groups at position 4 of pyrimidines and position 6 of purines determine whether the base is a:
- keto base (group is C=O), i.e. guanine G, thymine T and uracil U, or an
- amino base (group is C-NH2), i.e. cytosine C, adenine A.
- DNA contains adenine A, guanine G, cytosine C and thymine T, while
- RNA contains uracil U in place of thymine T.

POLYMERIZATION

- During polymerization, the a phosphoryl group of a free nucleotide triphosphate undergoes nucleophilic attack by the 3'C –OH group of the nucleotide at the 3' end of the growing chain.
- This results in addition of a nucleotide to the 3'C via a phosphodiester bond (C-O-P-O-C), with the elimination of pyrophosphate.
- Polymerization is catalysed by polymerase enzymes (see later), and is fuelled by the energy-rich phosphoanhydride bonds (P-O-P).

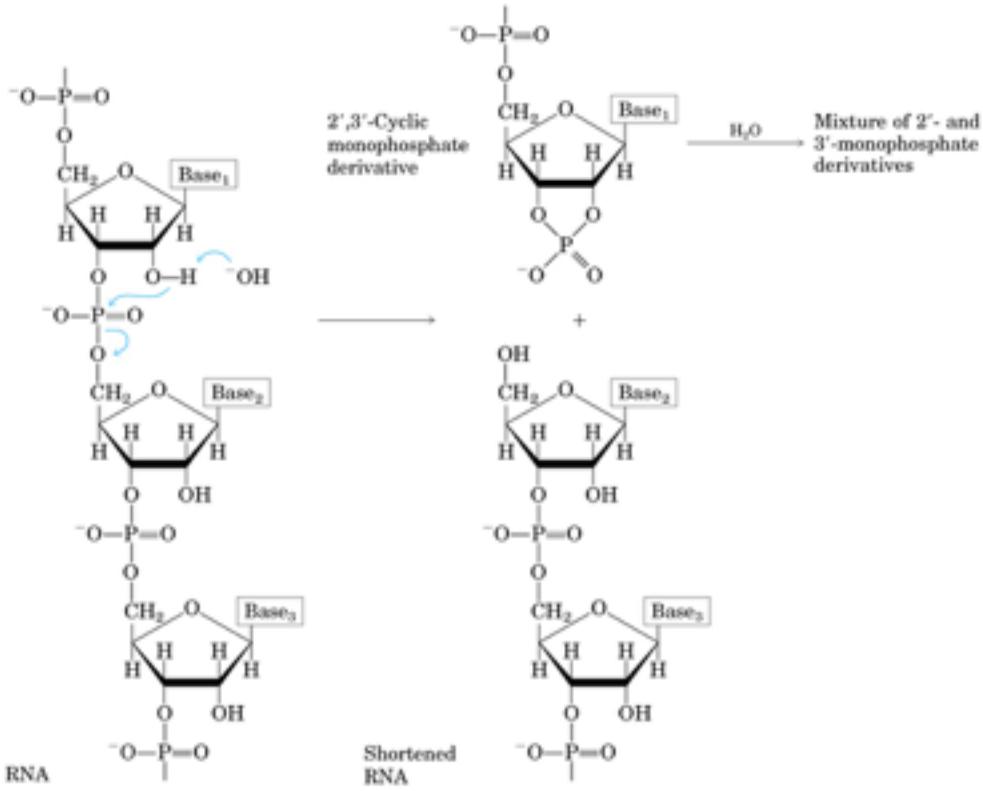


NUCLEOTIDES HAVE DIRECTIONS:

- Phosphodiester bonds link the 3' and 5' C of all the sugars within the chain.
- At one end there will be a free 5' phosphate group (we call this the 5' end)
- At the other end there will be a free 3' –OH group (the 3' end).
- Thus DNA & RNA polymers have directionality.
- Convention has it that we write nucleic acid sequences in the 5' to 3' direction.

WHY DNA IS THE HEREDITARY MATERIAL:

- The presence of the 2' –OH group in RNA makes it susceptible to base-catalysed hydrolysis.
- Thus DNA is much more stable than RNA.
- DNA is better suited to being the hereditary material

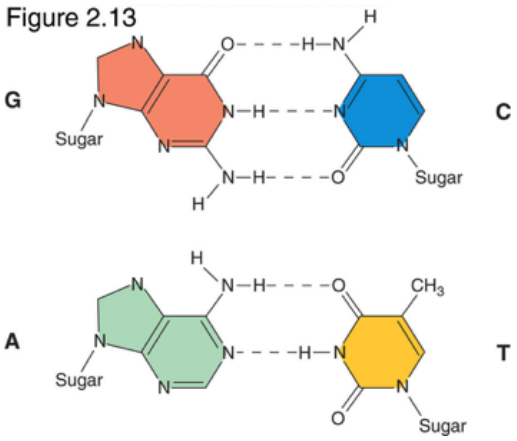


- The high free energy of the phosphoanhydride bond in ATP is used to drive very many biochemical reactions
- A lot of biochemistry is involved in making ATP
- Some nucleotides are important, intracellular signaling molecules (e.g. cAMP)
- Nucleotides can act as intercellular, neurotransmitters, thus regulating cell and organismal behaviour.

SECONDARY STRUCTURE OF NUCLEIC ACIDS

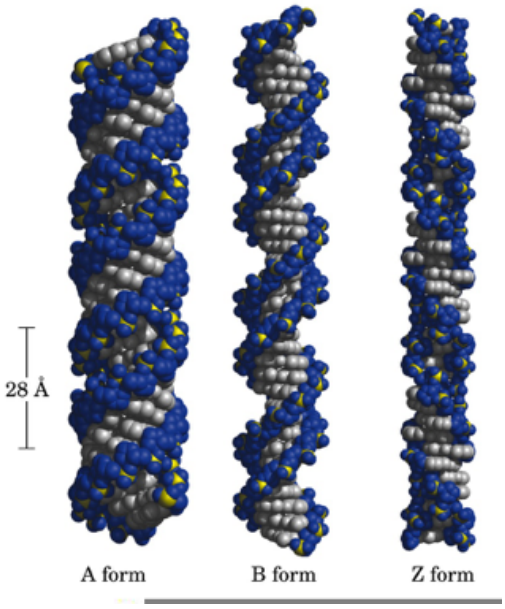
DNA

- The two strands run antiparallel to one another, and the complementary bases on each strand form hydrogen bonds with one another.
- Watson-Crick base-pairing:
    - A – T 2 H bonds
    - G – C 3 H bonds
- purines always face pyrimidines (constant width)



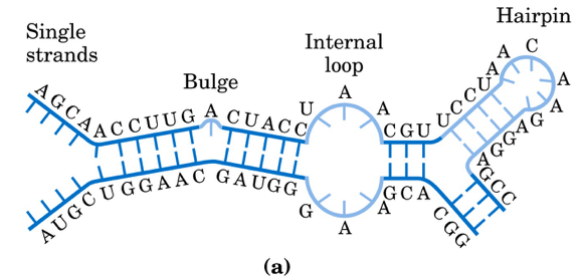
- Bases project perpendicular to the helical axis.
- If the backbone is a spiral staircase, the bases are the steps.
- Bases are parallel to one another and stack, partially overlapping.
- **Van-der-Waals & hydrophobic interactions between the planar base rings stabilize the DNA structure**.
- In 'average' B form DNA :
    - The **helices are right handed**
    - The rise between adjacent bases is **0.34 nm.**
    - The helical repeat is 3.4 nm.
    - There **are 10 base-pairs per turn**.
    - The helix is **2 nm wide.**
- Owing to the hydrogen bonding, base-paired nucleotides are not directly opposite each other
- This results in two grooves of different widths: the major groove and minor groove.
- The major groove is large enough to allow intimate protein binding to the double stranded DNA molecule.

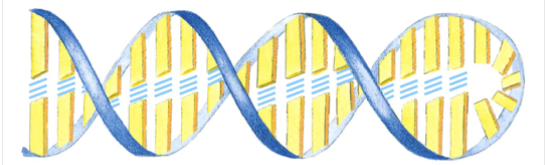
- N.B. Double-stranded DNA is also called duplex DNA
- **A form DNA** is a more compact right hand helix with 11 bases per turn, large tilt of base pairs and central hole.
- **Z form DNA** is a left-handed helix with 12 bases per turn and a zig-zag appearance.
- Biochemical conditions in the cell, and particularly protein-DNA interactions can locally alter DNA secondary structure.



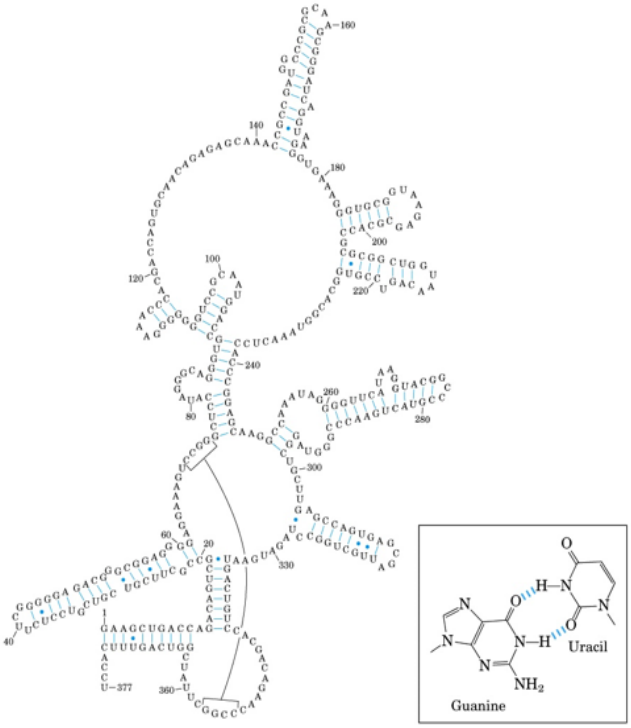
RNA:

- RNA is flexible and unstructured, but with pockets of structure where complementary base pairing forms hairpin or stem-loop structures.
- Hairpins form when complementary sequences of bases are close together.
- Stem-loop structures form where the complementary sequences are more distant.
- Note that single stranded RNA is a stable structure because of the rigidity provided by base stacking, Van der Waals forces.





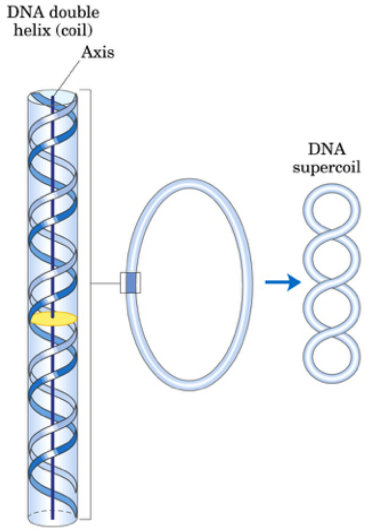
- Well-defined RNA secondary structures containing hairpin and stem-loop structures are important for ribosomal RNA (rRNA) and transfer RNA (tRNA). (See translation lectures).
- Duplex RNA in these regions resembles A DNA.
- Base-paired helical structures in an RNA. Shown here is the possible secondary structure of the M1 RNA component of the enzyme Rnase P, with many hairpins. Rnase P, which also contains a protein component (not shown), functions in the processing of transfer RNAs. The two brackets indicate additional complementary sequences that may be paired in the three dimensional structure. The blue dots indicate non Watson-Crick G=U base pairs (boxed inset). Note that G=U base pairs are allowed only when presynthesized strands of RNA fold up or anneal with each other. There are no RNA polymerases that insert a U opposite a template G or vice versa during RNA synthesis.



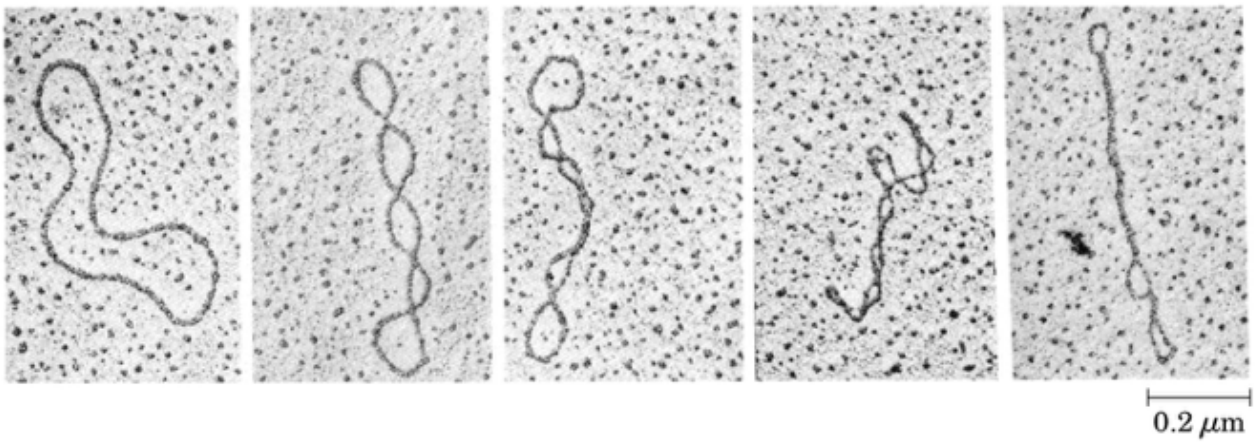
- RNA-DNA STRUCTURE:
- Hybrid DNA-RNA duplexes can form, and also have an A DNA-like conformation.
- Occur during:
    - Transcription, when mRNA is synthesised on a DNA template, and
    - Replication, when RNA primers are synthesised to initiate DNA synthesis.

DNA TERTIARY STRUCTURE

- The DNA double helix is packaged in different chromosomes
- In prokaryotes, these are often closed, circular stretches of DNA
- In eukaryotes, these are more often linear stretches of DNA
- The DNA is compacted via tertiary structure
- DNA supercoiling is the first stage of DNA compaction.
- DNA molecules tend to be very large and have to be packaged within a relatively small volume.



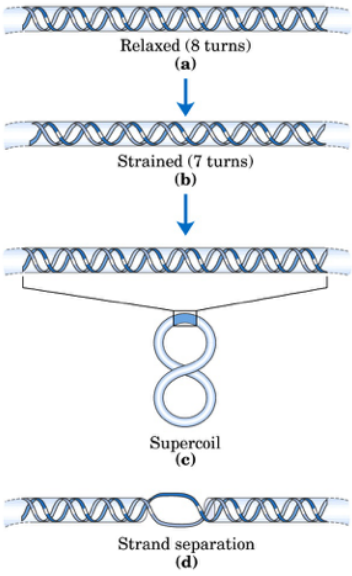
- Supercoiling of DNA. When the axis of the DNA double helix is coiled on itself, it forms a new helix (superhelix). The DNA superhelix is usually called a supercoil.
- A typical phone cord is coiled like a DNA helix and the coiled cord can itself coil in a supercoil. The illustration is especially appropriate because an examination of phone cords helped lead Jerome Vinograd and his colleagues to the insight that many properties of small, circular DNAs can be explained by supercoiling. They first detected supercoiling in small, circular viral DNAs in 1965.
- Supercoiling induced by separating the strands of a helical structure. Twist two linear strands of rubber band into a right handed double helix as shown. Fix one end by having a friend hold onto it, then pull apart the two strands at the other end. The resulting strain will produce supercoiling.



- DNA is supercoiled when it is twisted in space about its own axis.
    - Other terms for supercoiling are supertwisting or superhelicity.
- Supercoiling can only occur in closed structures.
- Bacterial chromosomes and plasmids are circular and hence closed. Even eukaryotic DNA is a closed structure, forming large loops with ends held together with proteins.
- The greater the number of supercoils, the greater the torsion.
- Only a break in either strand will allow untwisting to release the supercoils.
- An open or closed molecule that lacks supercoils is said to be relaxed.
- DNA can be negatively or positively supercoiled.
- Most naturally-occuring DNA is negatively supercoiled.
- Negatively supercoiled DNA is said to be underwound.
- Negative supercoils twist the DNA in the opposite direction from the turns of the right handed double helix; thus the torsional stress can be relieved by loosening the winding of the duplex and limited disruption of base pairing, this is important for DNA replication and DNA transcription.

What causes supercoiling ?

- Supercoils are introduced and removed by specific enzymes.
- Type I topoisomerases incrementally **relax** supercoiled DNA (prokaryotes & eukaryotes).
- Type II topoisomerases introduce **supercoils** (prokaryotes). **Requires ATP hydrolysis**. The energy stored in supercoiling fuels transcription and replication.
- Supercoiling in the same direction as the intrinsic duplex helical turns is termed positive supercoiling and results in overwound DNA.



- In the nucleus of eukaryotic cells DNA is wound around nucleosomes
- Nucleosomes consist of basic proteins called histones
- Nucleosomes assemble the DNA into 30nm threads

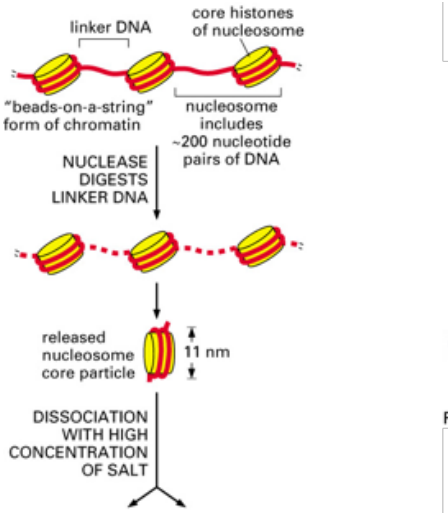


Figure 4–24 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

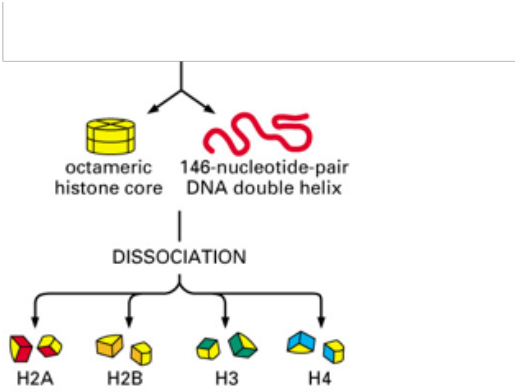
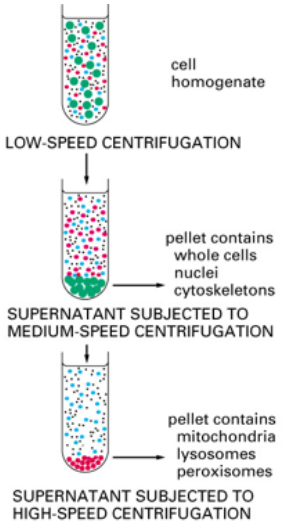


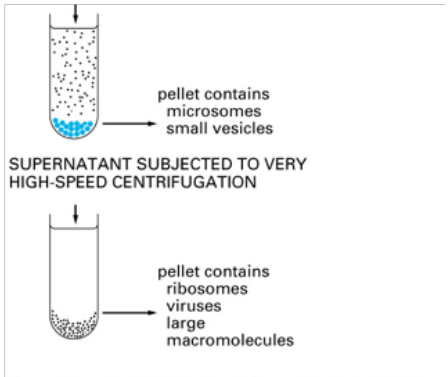
Figure 4–24 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

- Figure 4-24. Structural organization of the nucleosome. A nucleosome contains a protein core made of eight histone molecules. As indicated, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, an enzyme that breaks down DNA. (The nuclease can degrade the exposed linker DNA but cannot attack the DNA wound tightly around the nucleosome core.) After dissociation of the isolated nucleosome into its protein core and DNA, the length of the DNA that was wound around the core can be determined. This length of 146 nucleotide pairs is sufficient to wrap 1.65 times around the histone core.

NUCLEIC ACID PURIFICATION

- DNA is one of the largest molecules in the cell
- In prokaryotes it is found in a shared cytoplasm with all other cell components
- In eukaryotes it is tightly packaged within the nucleus
- Any purification procedure can take advantage of the subcellular location of nucleic acids
- Repeated centrifugataion at progressively higher speeds will fractionate homogenates of cells into their components. In general, the smaller the subcellular component, the greater is the centrifugal force required to sediment it. Typical values for the various centrifugation steps referred to in the figure are: low speed = 1000 times gravity for 10 minutes; medium speed = 20,000 times gravity for 20 minutes; high speed = 80,000 times gravity for 1 hour; very high speed = 150,000 times gravity for 3 hours.





- In velocity sedimentation, (A) subcellular components sediment at different speeds according to their size and shape when layered over a dilute solution containing sucrose. To stabilize the sedimenting bands against convective mixing caused by small differences in temperature or solute concentration, the tube contains a continuous shallow gradient of sucrose that increases in concentration towards the bottom of the tube (typically from 5% to 20% sucrose). Following centrifugation, the different components can be collected individually most simply by puncturing the tube and collecting drops from the bottom, as illustrated here.

  In equilibrium sedimentation (B) subcellular components move up or down when centrifuged in a gradient until they reach a position where their density matches their surroundings. Although a sucrose gradient is shown here, denser gradients, which are especially useful for protein and nucleic acid separation, can be formed from cesium chloride. The final bands can be collected as in (A).

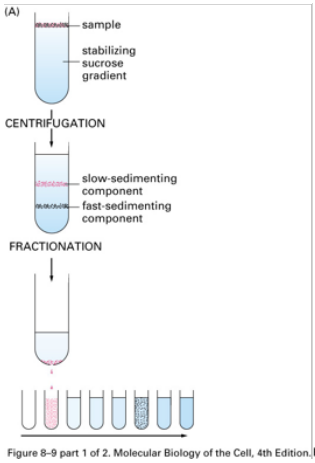


Figure 8–9 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

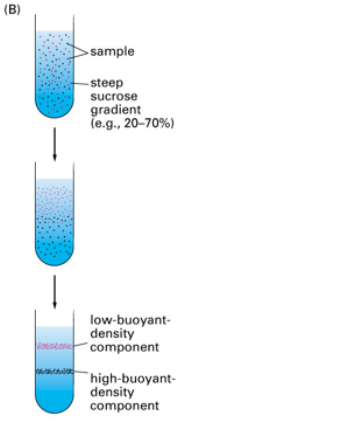
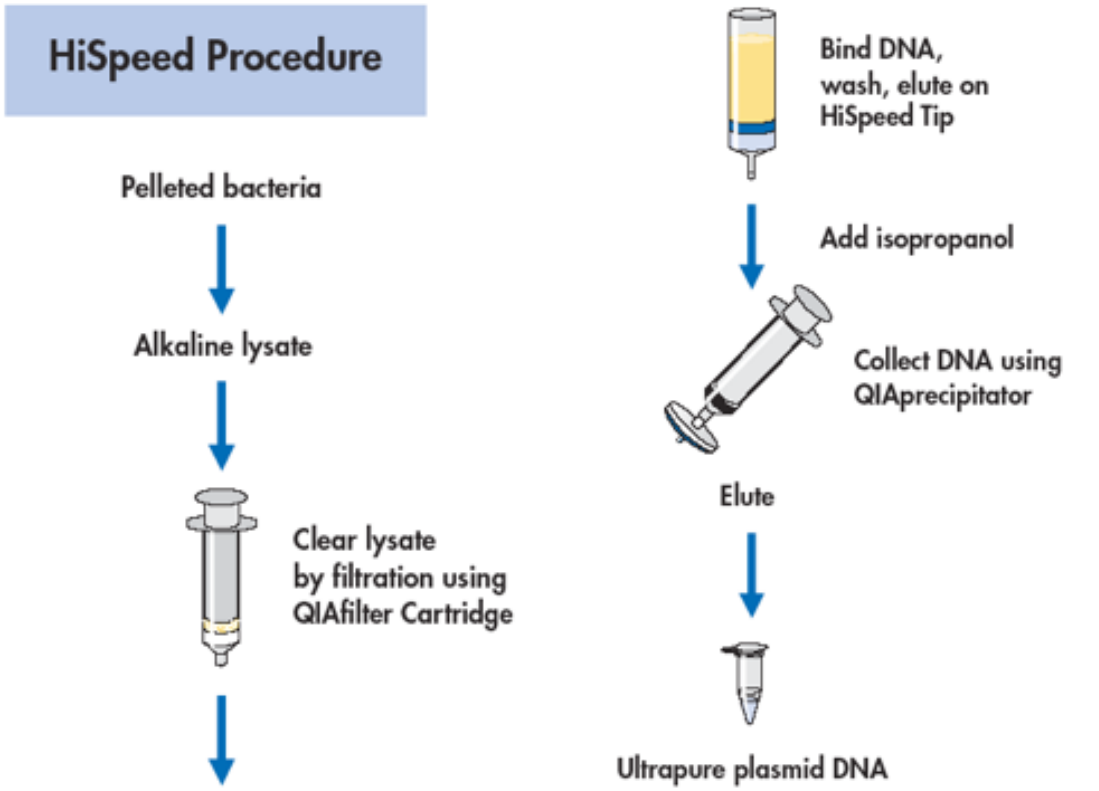


Figure 8–9 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

- **Nucleic Acids are polyanions**
  - due to phosphates in backbone
  - not soluble in organic solvents
  - not denatured by them
  - negative charges neutralized by association with cations - e.g. Na^+ or histones
- **Strand denaturation:**
  - The sugar - phosphate backbone of RNA is hydrolysed under alkaline conditions
  - Under the same conditions, the backbone of DNA is not affected.
  - Strands are denatured due to interference of H bonding between bases.
  - At elevated temperatures, H bonding of bases is insufficient to keep strands together, and they are denatured.
- **Nucleases:**
  - Naturally occurring enzymes in the cell attack and hydrolyse the phosphodiester bonds of both RNA and DNA.
  - Deoxyribonucleases specifically catalyse the degradation of DNA
  - Ribonucleases specifically catalyse the degradation of RNA

- **Purification:**
    - Treat cells under conditions which:
    - Inactivate (denature) cellular components e.g. proteins (Proteinase, chaotropic agents)
    - Do not denature DNA / RNA
    - Note that hydrodynamic forces lead to different parts of the DNA molecule moving in different directions
    - Causes shearing of the large DNA molecules of chromosomes into smaller ones
    - Use strong detergents but minimal agitation
    - e.g. Sodium Dodecyl Sulphate (SDS)
    - Use chelating agents to complex divalent cations (e.g. Mg^{2+}) that nucleases require for activity, e.g. EDTA (ethylene diamine tetra-acetic acid).
    - Remove contaminating proteins by organic extraction, e.g. phenol or phenol/chloroform
    -
- **Anion Exchange Chromatography:**



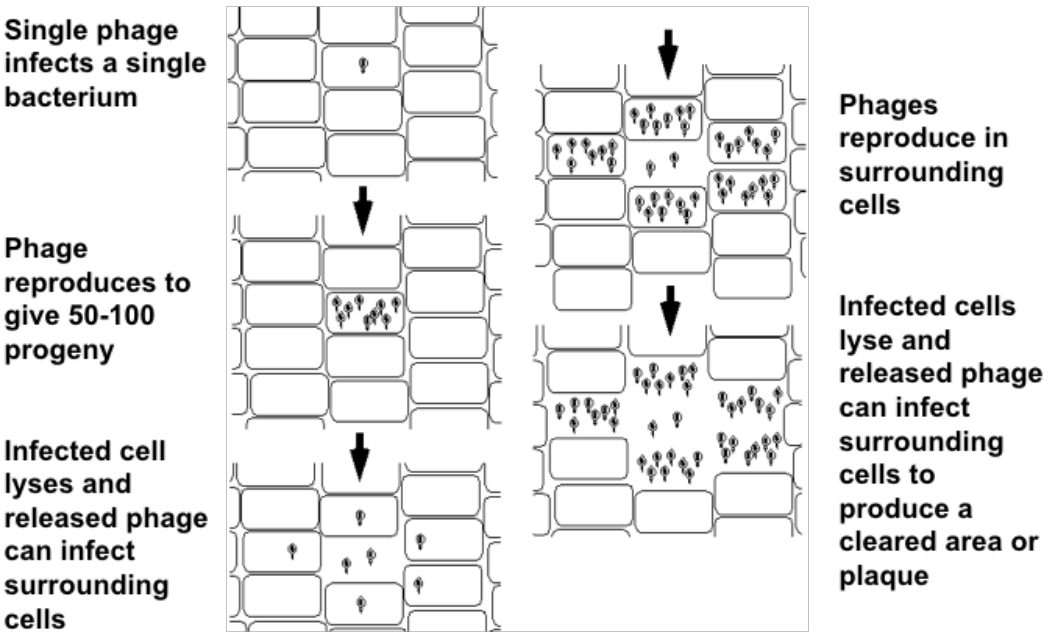
- The Qiagen hispeed plasmid isolation procedure. Bacteria are grown, harvested and resuspended in buffer. Then subject to high salt, strong detergent and strong alkali. This denatures and precipitates the bulk of the macromolecules such as proteins, cell wall, genomic DNA. Bind the lysate to anion exchange resin in the hispeed tip, elute with salt and precipitate with isopropanol, then catch the precipitate on a filter, dry and elute with a small volume of TE.

- From relatively pure solutions, concentrate nucleic acids by selective precipitation
    - exploits large size of nucleic acids
    - use addition of e.g. ethanol
    - recover by centrifugation and resuspend in aqueous solution
    - Avoid extremes of pH and heat which would otherwise:
    - Cause DNA strands to denature
    - Cause sugar phosphate backbone to be hydrolysed

RNA PURIFICATION:

- Harder to purify than DNA
- Partly due to the action of contaminating ribonucleases (e.g. in sweat)
- Partly due to difficulties in dissociating RNA from protein complexes
- Use very strong denaturing conditions and the presence of proteases
- Use density gradient centrifugation to separate from DNA

A virus that infects only bacteria is called a **bacteriophage.** The nucleic acid of a virion is enclosed within a protein coat, or **capsid**, composed of multiple copies of one protein or a few different proteins, oeach of which is encoded by a single viral gene. In many DNA bacetiophages, the viral DNA is located within an icosahedral head that is attached to a rodlike tail. During infection, some icosahedral viruses interact with host-cell receptors via clefts in between the capsid subunits. The number of infectious viral particles in a sample can be quantified by a **plaque assay**. This assay is performed by culturing diluted sample of viral particles on a plate covered with host cells and then counting the number of local lesion, called plaqies, that develop. A plaque develops on the plate wherever a single virion initially infects a single cell.
A virus overtakes a cell and uses the host cellular machinery to reproduce via the **lytic cycle**.



Different bacterial host strains supported growth of phage to different extent. Phage grown on *E.coli* strain B would grow well (high plating efficiency) if regrown on same strain, but poorly (very few plaques) if grown on other strains such as *E.coli* strain K.

Resuspend phage from a single plaque grown on *E.coli* strain B in 1 ml of buffer and plate out 1 ml on either strain B or K:

E.coli strain B → 1000 plaques

E.coli strain K → 1 plaque

The phage that did escape "restriction" by strain K would now grow well on strain K. Resuspend phage from the single plaque that grew on *E.coli* strain K in 1 ml of buffer and plate out 1 ml on either strain B or K:

E.coli strain B → 1 plaque

E.coli strain K → 1000 plaques

They were said to have been "modified". Much later discovered that "restriction" was due to degradation of the DNA and that the "modification" was due to methylation of the DNA.

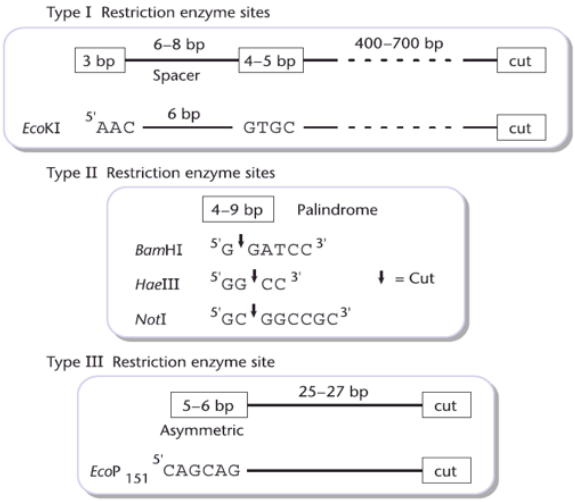
- Restriction is due to restriction enzymes, nucleases that cut nucleic acids:
    - May show specificity for sugar so there are DNases, RNases and non-specific nucleases.
    - May be specific for single-strands or double-strands or be non-specific
    - Differ by where in the phosphodiester bond they cut, leaving either 5' phosphate and 3' hydroxyl or 5' hydroxyl and 3' phosphate
    - May show specificity for where they attack, so there are endonucleases (cleave site within molecule) and exonucleases (degrade from end of molecule - 3' to 5' exonuclease attacks from the 3' end and 5' to 3' exonuclease attacks from the 5' end)

Nomenclature for restriction enzyme endonucleases

First three letters	Bacterial species of origin (genus + strain)
Fourth letter	Bacterial strain
Number (Roman numerals)	Differentiates between different enzymes in strain

EcoRI *Escherichia coli* strain RY13 1st (I) enzyme discovered
EcoRII *Escherichia coli* strain RY13 2nd (II) enzyme discovered
BamHI Bacillus amyloliquefaciens strain H 1st (I) enzyme discovered

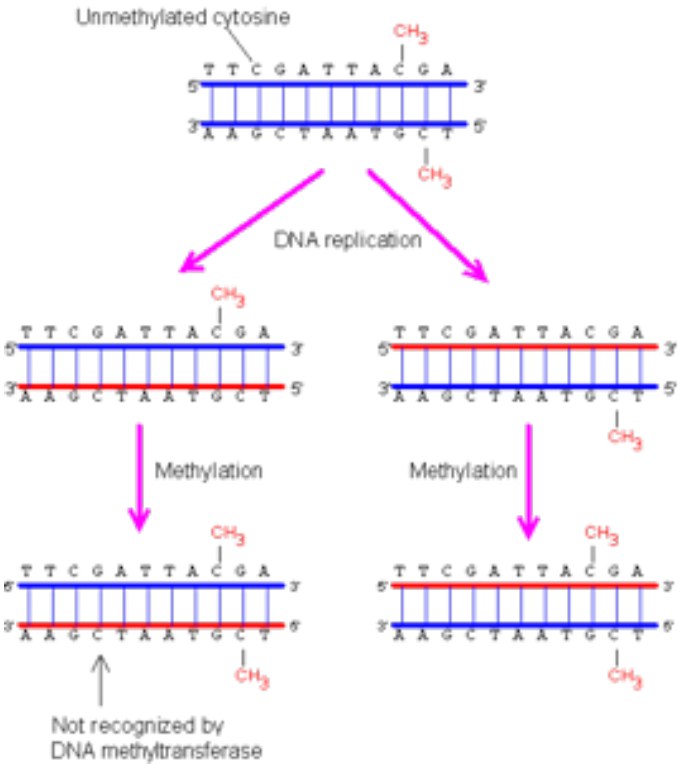
- **RESTRICTION ENZYME ENDONUCLEASES:**
    - **Type I** enzymes are complex, multisubunit, combination restriction-and-modification enzymes that cut DNA at random far from their recognition sequences. Originally thought to be rare, we now know from the analysis of sequenced genomes that they are common. Type I enzymes are of considerable biochemical interest, but they have little practical value since they do not produce discrete restriction fragments or distinct gel-banding patterns.
    - **Type II** enzymes cut DNA at defined positions close to or within their recognition sequences. They produce discrete restriction fragments and distinct gel banding patterns, and they are the only class used in the laboratory for routine DNA analysis and gene cloning. Rather than forming a single family of related proteins, Type II enzymes are a collection of unrelated proteins of many different sorts. Type II enzymes frequently differ so completely in amino acid sequence from one another, and indeed from every other known protein, that they exemplify the class of rapidly evolving proteins that are often indicative of involvement in host-parasite interactions.
    - **Type III** enzymes are also large combination restriction-and-modification enzymes. They cleave outside of their recognition sequences and require two such sequences in opposite orientations within the same DNA molecule to accomplish cleavage; they rarely give complete digests.



- **Inverted repeat palindrome**: the sequence of top strand read left to right (5' to 3') is same as the sequence of the bottom strand read right to left (5' to 3') - more common in restriction enzymes

- **Methylation**: Modification of DNA by methylation within the restriction enzyme recognition site protects the DNA from degradation

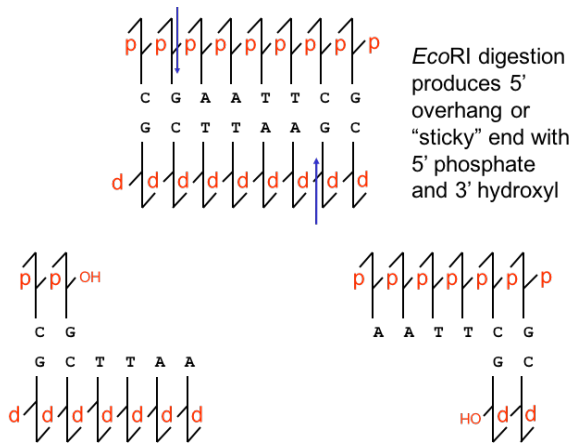
C5 or N4 positions of cytosine / N6 position of adenine base



- During DNA replication only normal, non-methylated bases are used by the DNA polymerase
- After DNA replication of a duplex where both strands are methylated, the DNA will be methylated on one strand only - hemi-methylated
- Hemi-methylated DNA is not cleaved by the restriction enzyme
- Hemi-methylated DNA is the best substrate for the methylases so it quickly becomes fully methylated before the next round of DNA replication

- **RESTRICTION ENZYMES:**

  HaeIII and AluI cut straight across the double helix producing "blunt" ends. However, many restriction enzymes cut in an offset fashion. The ends of the cut have an overhanging piece of single-stranded DNA. These are called **"sticky ends"** because they are able to form base pairs with any DNA molecule that contains the complementary sticky end. Any other source of DNA treated with the same enzyme will produce such molecules. Mixed together, these molecules can join with each other by the base pairing between their sticky ends. The union can be made permanent by another enzyme, a **DNA ligase**, that forms covalent bonds along the backbone of each strand. The result is a molecule of **recombinant DNA** (**rDNA**).



- **ISOSCHIZOMERS**: enzymes that have the same recognition sequence but not necessarily the same cleavage site or the same sensitivity to methylation.

DpnI	**G^{m}ATC**	**but not**	**GATC**
DpnII	**GATC**	**but not**	**G^{m}ATC**
MboI	**GATC**	**but not**	**G^{m}ATC**
Sau3AI	**GATC**	**and**	**G^{m}ATC**

GEL ELECTROPHORESIS

"the migration of charged molecules (in solution or through a solid support) when under the influence of an electric field".
At equilibrium, there is no net force on molecule:
Frictional force = Force exerted by electric field on particle

$$f \times v = q \times E$$

Electrophoretic Mobility:

$$\mu = v/E = q/f = eZ/f$$

MOBILITY is dependent on NET CHARGE and MOLECULAR DIMENSIONS (size and shape)

Materials needed:

- Something to act as supporting media for our sample that allows separation to occur using the parameter(s) we want
- Some apparatus to facilitate the electric field and hold our support media / sample
- Some way to visualize our electrophoresed molecules

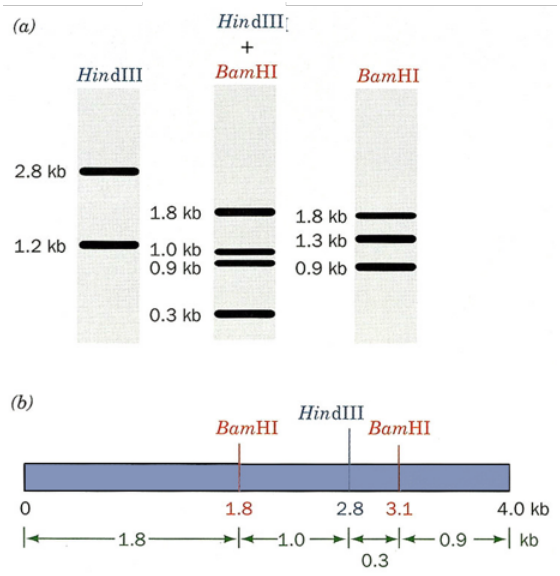
Medium molecule	Molecules separated	Size of
Paper	amino acids peptides	Size and Charge
Polyacrylamide	Proteins DNA/RNA	10-200 kDa 5-500 bp
Agarose	DNA/RNA	100-50,000 bp

- Fragments of linear DNA migrate through agarose gels with a mobility that is inversely proportional to the $\log_{10}$ of their molecular weight. In other words, if you plot the distance from the well that DNA fragments have migrated against the $\log_{10}$ of either their molecular weights or number of base pairs, a roughly straight line will appear.
- Circular forms of DNA migrate in agarose distinctly differently from linear DNAs of the same mass. Typically, uncut plasmids will appear to migrate more rapidly than the same plasmid when linearized. Additionally, most preparations of uncut plasmid contain at least two topologically-different forms of DNA, corresponding to supercoiled forms and nicked circles

- **Visualizing DNA/RNA on Agarose gel:**

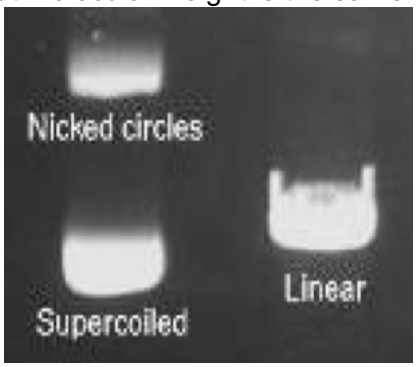
Ethidium bromide (SyBr SAFE)
Properties:
1. Intercalating agent
binds to DNA by slipping between stacked bases
2. Strongly fluorescent when bound to duplex DNA
(also binds to ssDNA/RNA but less well)
UV illuminated DNA bands detected as orange
3. Carcinogen/ mutagen
alternatives (cost) eg: SyBr SAFE

- **Construction of Restriction Maps:**

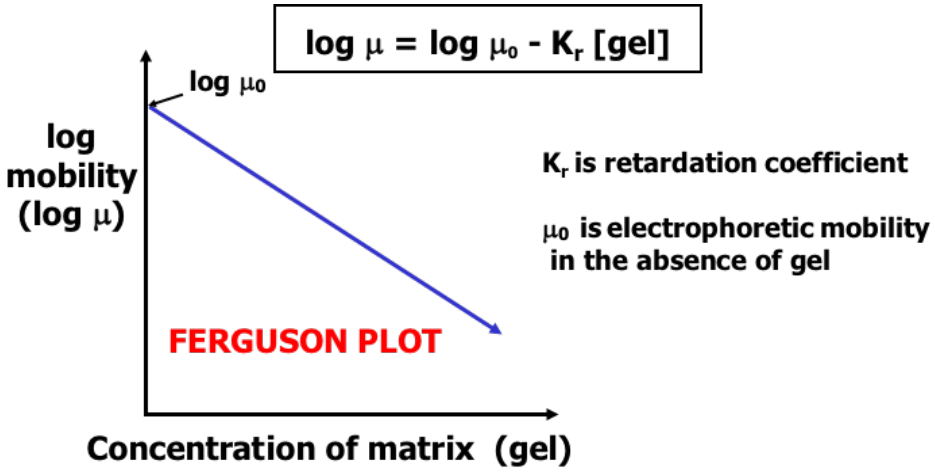


Factors affecting the migration on agarose gel:

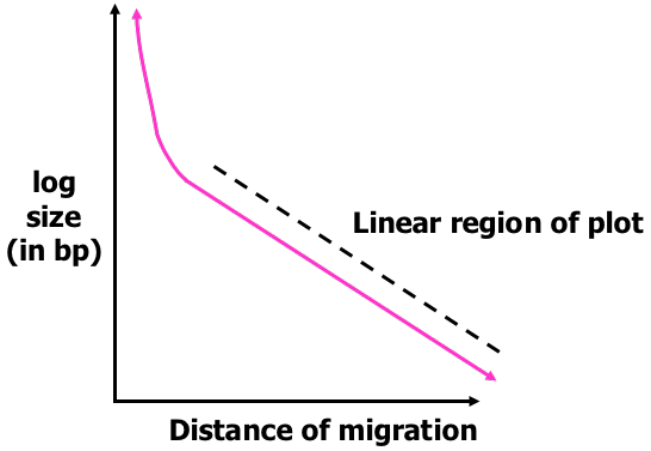
1. conformation of the DNA molecule: plasmids are circular, supercoiled DNA travels faster than relaxed one (it's more compact). Looks smaller but molecular weight is the same.



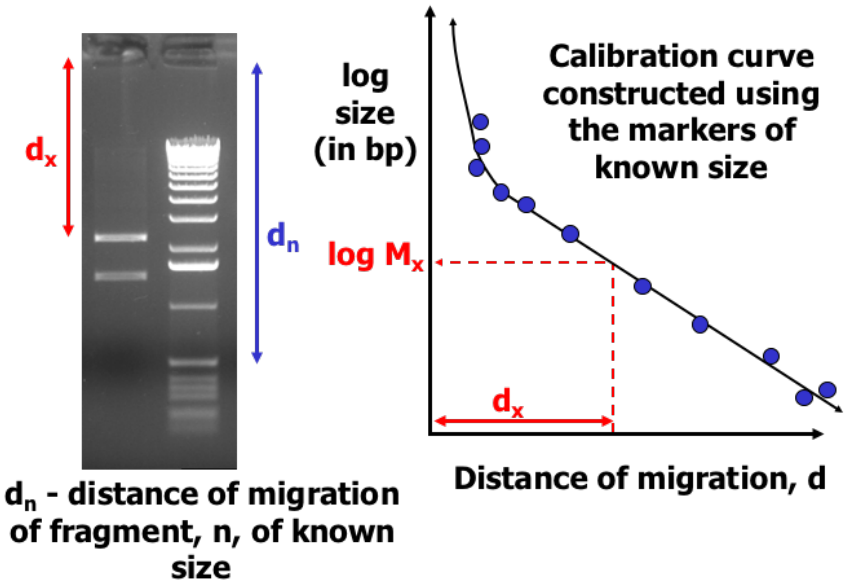
2. agarose gel concentration: electrophoreticmobility decreases with increasing of gel concentration.



3. size of DNA molecule: the molecular sieve effect. For DNA, charge/unit length is constant, for linear DNA separation occurs by molecular sieving effect and is dependent on size alone.

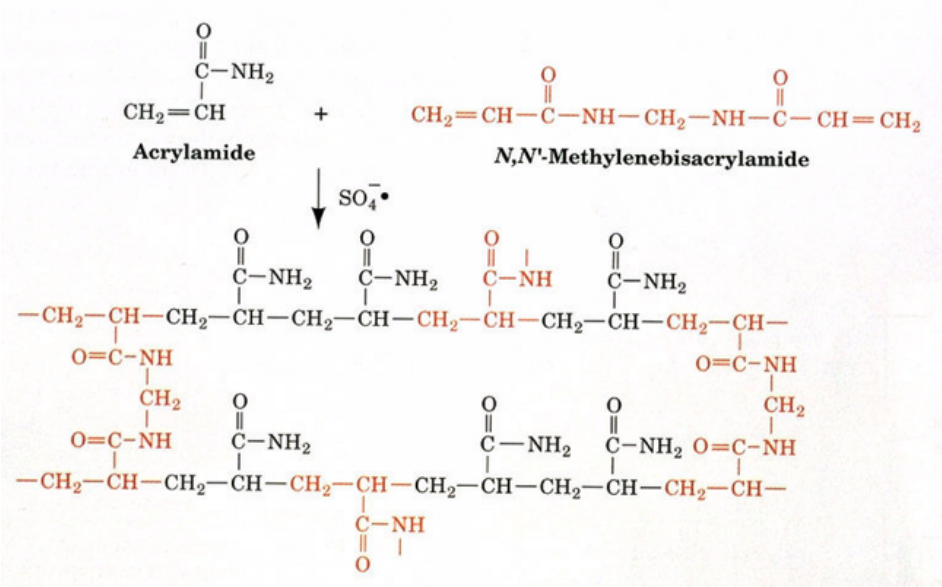


Determination of the size of linear DNA fragments--> calibration curve: constructed using the markers of known size.



4. electrophoresis conditions: voltage applied, time, presence of intercalating dyes, electrophoresis buffer (TAE -->Tris-acetate (acetic acid)-EDTA; TBE--> Tris-borate (boric acid); EDTA pH 7.5-7.8 -->Ethylenediaminetetraacetate (EDTA) is a chelating agent for metal cations)

POLYACRYLAMIDE GEL ELECTROPHORESIS



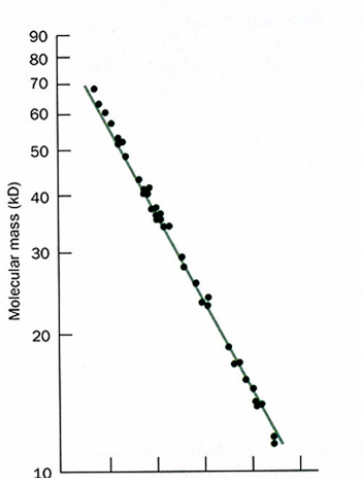
Polyacrylamide is ideal for protein separations because it is chemically inert, electrically neutral, hydrophilic, and transparent for optical detection at wavelengths greater than 250 nm. Additionally, the matrix does not interact with the solutes and has a low affinity for common protein stains.

- Visualizing Proteins/DNA/RNA on polyacrylamide gel:

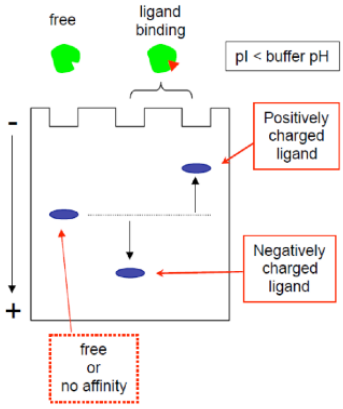
Proteins	**-**	**Coomassie Blue** **Silver staining** **SyPro dyes**
DNA/RNA	**-**	**Silver staining** **Radioactivity** **SyPro dyes** **Fluorescence/luminescence**

- **Factors affecting the migration on acrylamide gel:**

  - **Denaturing conditions**: native protein is denatured uning SDS (sodium dodecyl sulphate) and Mercaptoethanol. SDS is a detergent that gives the protein a constant mass/charge ratio and even the charges of the protein. Mercaptoethanol is a reducing agent which disrupts disulphide bridges



- **Non-denaturing conditions:** Native-Page: Proteins are prepared in a non-reducing non-denaturing sample buffer, which maintains the proteins' secondary structure and native charge density. Therefore you can easily see multiple bands from the camshot of your native PAGE gel if your target protein has polymerized forms in your sample. In native PAGE electrophoresis most proteins have an acidic or slightly basic pl (isoelectric point) (~3–8) and migrate towards the negative polar



Electrophoresis can be used for:

- Estimating the size of unknown DNA molecules
- Checking the purity and amount of DNA
- Isolating a particular DNA molecule
- Restriction Mapping
- DNA sequencing

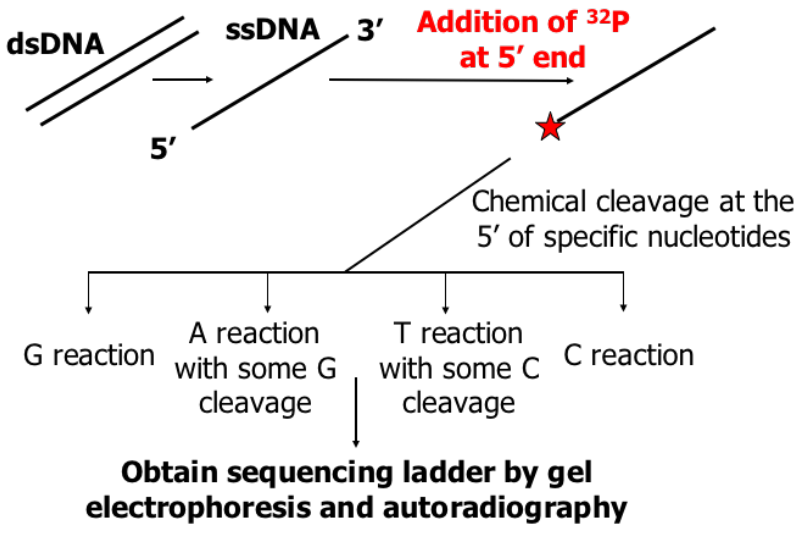
DNA Sequencing:

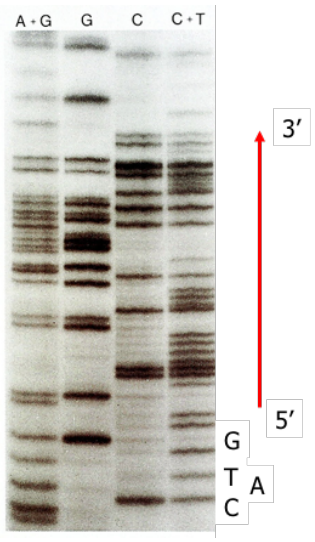
- Degradation: chemical cleavage method (Maxam and Gilbert)
- Synthesis : Dideoxy chain-terminator method (Sanger)

Common Features:

- Production of a series of single-stranded DNA molecules with common 5' end
- The identity pf the base at the 3' end is known
- The molecules are separated by polyacrylamide gel electrophoresis in the presence of urea.

- **Maxam and Gilber Method**:

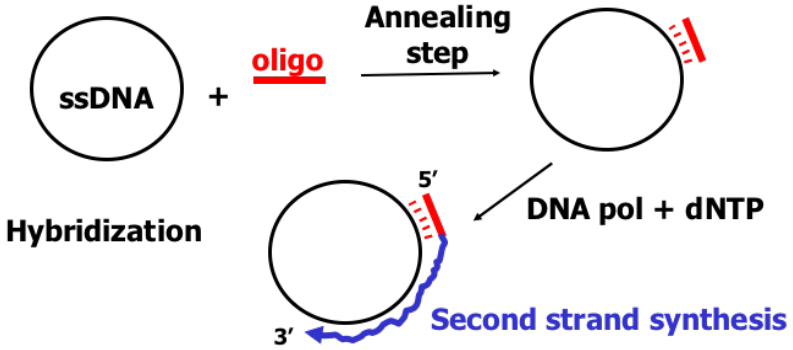


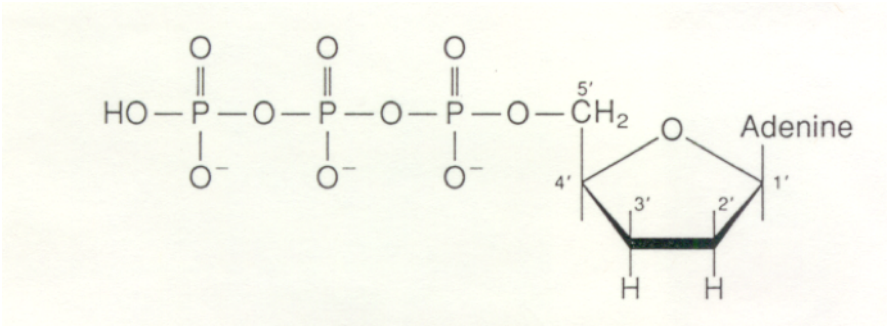


Can label with P32 or S35 or non-radioactive label

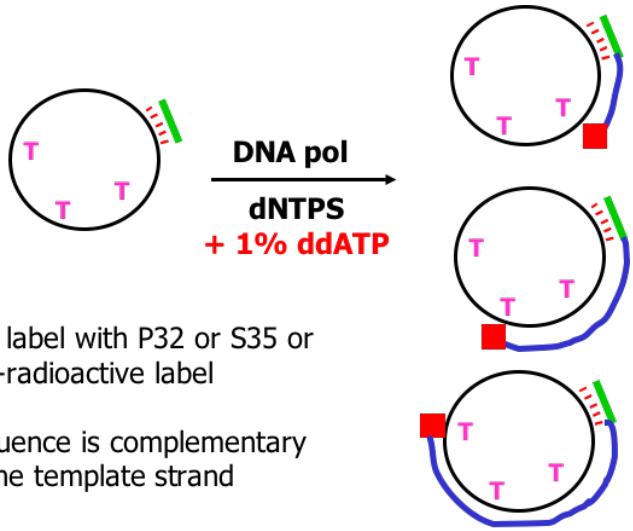
Sequence is complementary to the template strand

- **Sanger Method:**
  - DNA polymerases synthesise DNA using dNTPs, a single-stranded DNA template and an oligonucleotide primer
  - Basis of sequencing and the polymerase chain reaction (PCR)





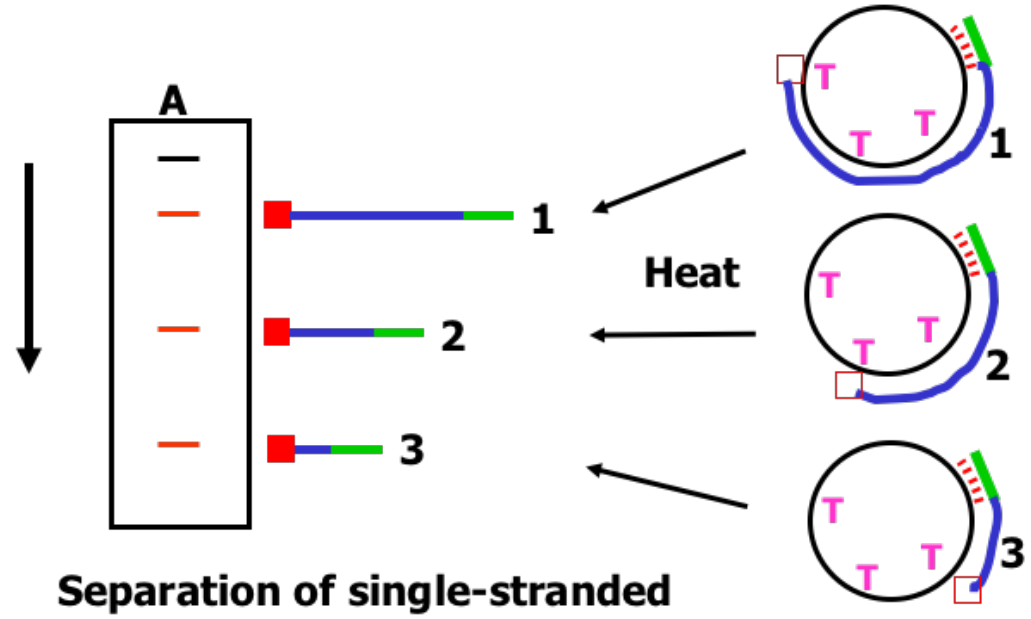
- Dideoxy nucleosides lack the 3’ -OH group and so act as terminators of DNA synthesis



Can label with P32 or S35 or non-radioactive label

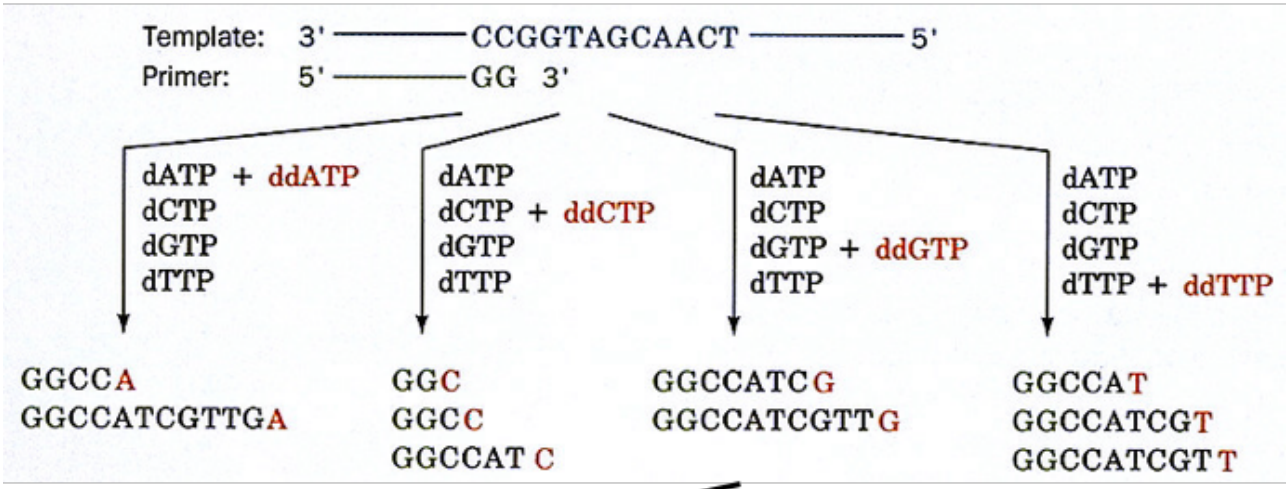
Sequence is complementary to the template strand

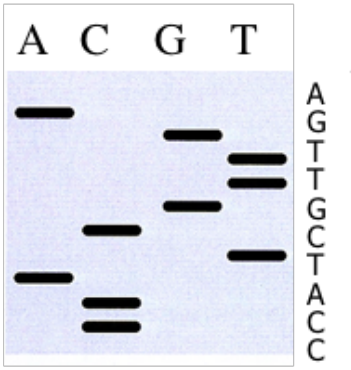
Single-stranded products of the sequencing reaction are separated on a polyacrylamide gel

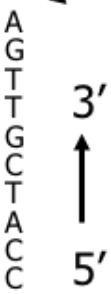


Separation of single-stranded DNA on urea-containing denaturing polyacrylamide gels at 70°C

Four dideoxy reactions are used to sequence DNA



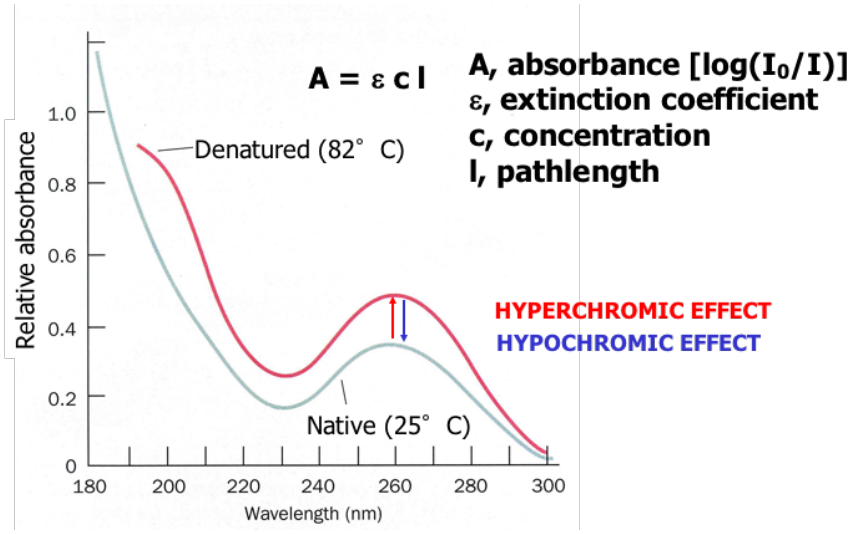




- Fragments differing in size by one base can be resolved
- Sequence complementary to template strand

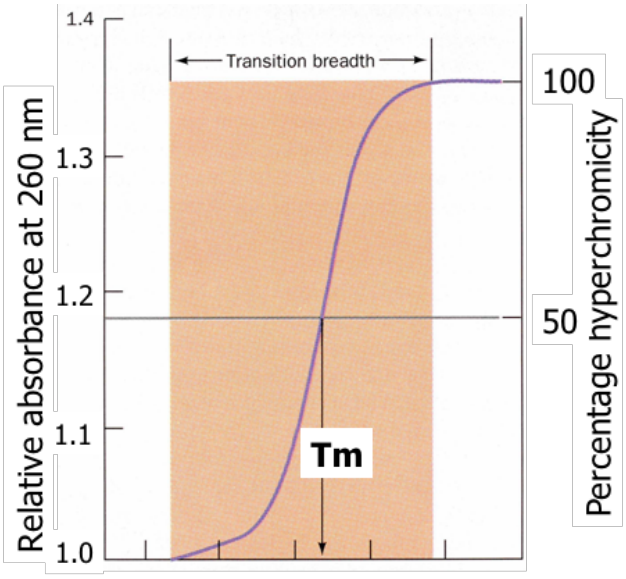
DNA DENATURATION

Disruption of the hydrogen bonds between paired bases and of base stacking causes unwinding of the double helix to form two single strands, completely separate from each other along the entire length (partial denaturation) of the molecule. No covalent bonds in the DNA are broken. Renaturation of a DNA molecule is a rapid one-step process, as long as a double-helical segment of a dozen or more residues still unites the two strands. When the temperature or pH is returned to the range in which most organisms live, the unwound segments of the two strands spontaneously rewind, or **anneal,** to yield the intact duplex.



- DNA has an absorbance peak of 260nm.
- The close interactions between stacked bases in a nucleic acid has the effect of decreasing its absorption of UV light relative to that of a solution with the same concentration of free nucleotides, and the aborption is decreased further when two complementary nucleic acid strands are paired. This is called the **hypochromic effect.**
- Denaturation of a double-stranded nucleic acid prodices the opposite result: an increase in absorption called the **hyperchromic effect**. The transition from double-stranded DNA to the single-stranded can thus be detected by monitoring absorption at 260nm.

DNA Melting Curve:



- Each species of DNA has a characteristic denaturation temperature (melting point): the higher its content og G≡C base pairs, the higher the melting point of DNA (because G≡C have three hydrogen bonds and require more heat energy to dissociate)
- The melting point can yeld an estimate of base composition

Factors affecting T_m --> T_m increases with:

- increasing G-C content (increased H-bond pairs)
- increasing Na^+ concentration
- decreasing formamide concentration
- increasing length of hybrid

T_m (ºC) = 81.5 + 0.41 x (% G-C) + 16.6 log [Na^+] - 0.63 x (% formamide) - (600/length)

HYBRIDISATION

Formation of double-stranded helix or hybrid usually between a probe and its complementary target

Stability of hybrids:

DNA/DNA < DNA/RNA < RNA/RNA

Stringency of hybridisation:

Low: temperature of hybridisation much lower then T_m for perfect match

High: temperature of hybridisation closer to T_m for perfect match

Types of hybridisation:

- **Homologous hybridisation**: Probe and target contain identical sequences
  HIGHER STRINGENCY
- **Heterologous hybridisation**: Probe and target contain similar but non-identical sequences
  e.g. Finding gene in humans using a drosophila probe
  LOWER STRINGENCY

Complexity of DNA: refers to number of base pairs per repeating unit

DNA sequence	Complexity (x)
polyA	1
poly(AGCT)	4
E. coli genome	4.7 million

DNA RENATURATION

For a given amount of DNA (g/L), the initial concentration of target DNA molecules is dependent on:

- The size of genome (determines total number of DNA molecules)
- The number of times the target is repeated in the DNA

Total number of target molecules =
number of DNA molecules x number of repeats

- Rate of renaturation gives information on the size of DNA molecules and the sequence complexity of the sample

NUCLEIC ACID LABELLING

Labelled DNA can be generated by

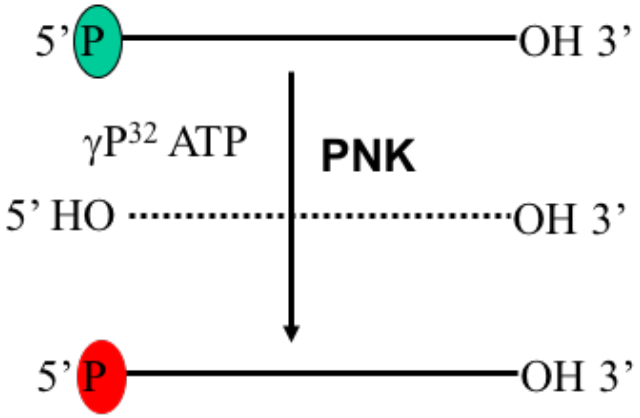
- **End labelling (two types)**
- **Random priming**

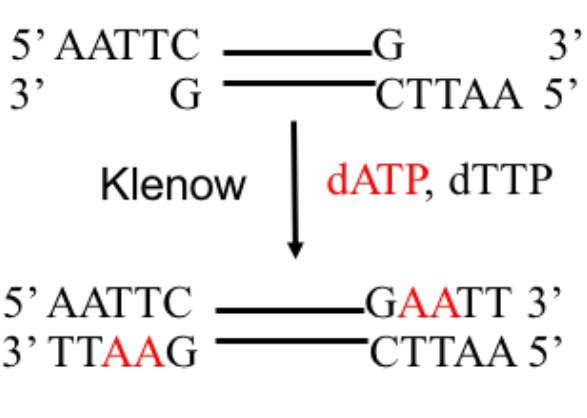
Labelling can be isotopic or nonisotopic

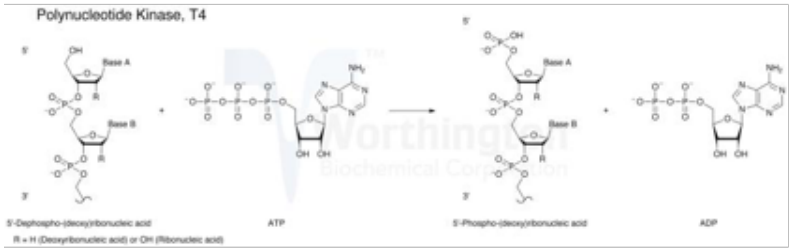
- Non isotopic labelling can be direct or indirect
- **Traditionally** nucleic acids were labeled with radioisotopes for detection by autoradiography. Isotopes including: ^{32}P, ^{33}P, ^{35}S, and ^{3}H each had advantages. ^{3}H was used for chromosome *in situ* hybridization. ^{32}P, ^{33}P, ^{35}S, were used for DNA sequencing.
- **More contemporarily** nucleic acids are labelled either directly with a fluorophore (eg: Flourescein) or indirectly with a reporter molecule (eg: DIG or biotin).
  These are visualised by laser detectors, fluorescence, chemilumenescence or colorimetric assay.

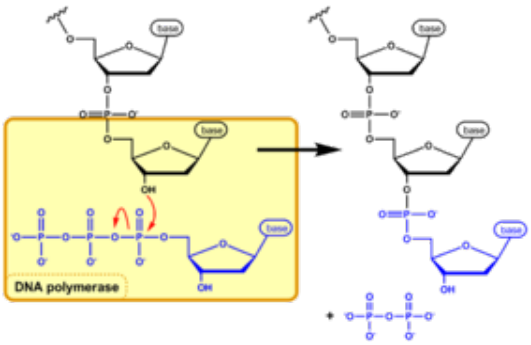
- **END labelling :** with polynucleotide kinase (PNK) , with Klenow (DNA pol)

  Forward or exchange reactions (fill-in reaction)

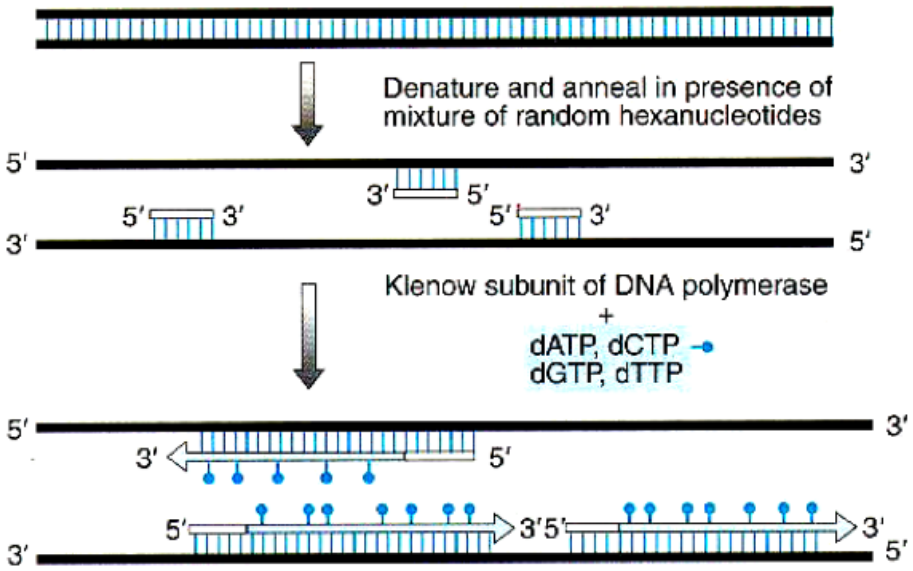






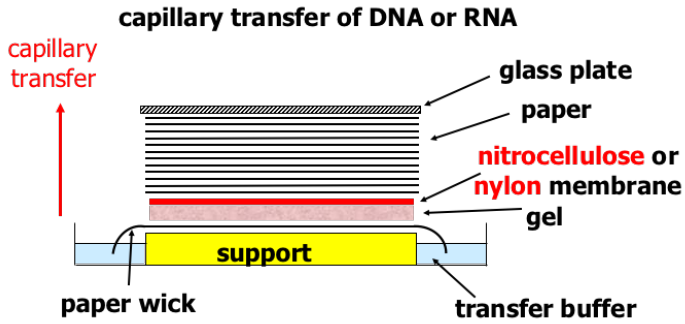


- **Random Priming :**

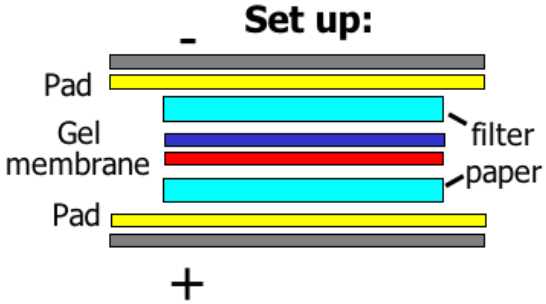


- **Hybridisation techniques:**
  - **Filter or membrane hybridisation-** a techniques that prevents parental strands re-annealing.
    e.g. Southern, Northern, colony hybridisation, (Western)
  - **Microarrays-** hybridisation analysis of thousands of DNAs (genes) simultaneously
  - ***In situ* hybridisation-** allows localisation of target DNA/RNA in tissue (e.g. fluorescence *in situ* hybridisation (FISH), chromosome painting and "SKY")
  - **DNA sequencing**
  - **PCR**

BLOTTING: transfer of nucleic acids (or protein) to a solid membrane support

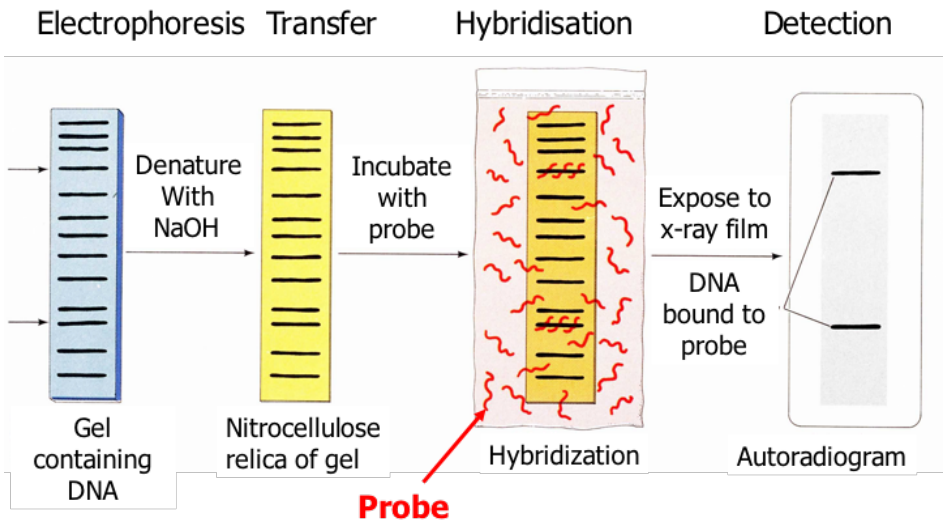


- **Electroblotting:** transfer proteins or nucleic acids onto a membrane by using PVDF or nitrocellulose, after gel electrophoresis



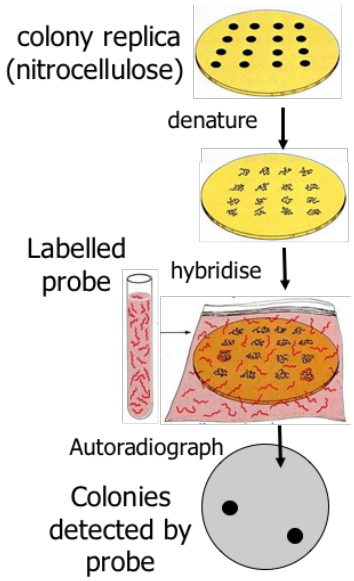
Type	Target	Probe
Southern	DNA	ssDNA or RNA
Northern	RNA	ssDNA or RNA
Western	Protein	Antibody

- **Southern Blotting:**
  - • DNA is separated by size on an agarose gel and transferred as onto nitrocellulose, nylon or PVDF membrane.
  - • A radio-rabelled DNA fragment (probe) is incubated with the membrane, and detect the hybridised bands by exopsing to a film.
  - **• This can be used to detect a specific DNA sequence**
  - **in DNA samples.**



- **Northern Blotting:**
  - • Used to analyse RNA.
  - • RNA is separated by size on an agarose gel and transferred as in Southern blotting, onto nitrocellulose, nylon or PVDF.
  - • A radio-labelled RNA or DNA fragment (probe) is incubated with the membrane, and detect the hybridised bands by exposing to a film.
  - • This can be used to identify what is transcribed (size and type) and the level of transcription.

- **Colony Hybridisation:** If a bacterial colony or plaque is potentially carrying a fragment of interest, it can be screened by hybridisation



- **Microarrays:**
    - Each spot on a microarray contains multiple identical strands of DNA.
    - The DNA sequence on each spot is unique.
    - Each spot represents one gene.
    - Thousands of spots are arrayed in orderly rows and columns on a solid surface (usually glass).
    - The precise location and sequence of each spot is recorded in a computer database.
    - Microarrays can be the size of a microscope slide, or even smaller.

    - RNA isolation
    - cDNA generation
    - Labelling of Probe
    - Hybridise to array
    - Imaging and Analysis

- ***In Situ* Hybridisation:**
    - RNA or DNA is fixed and probed *in situ* i.e. in the position it occupied in the living tissue
    - Cut into sections, hybridise with probe, examine with microscope.

  Chromosome Painting --> FISH (fluorescent in situ hybridisation)

DNA REPLICATION

The purpose is to copy ALL of the genetic material ACCURATELY before cell division so that both the daughter cells receive a full complement of genetic material.

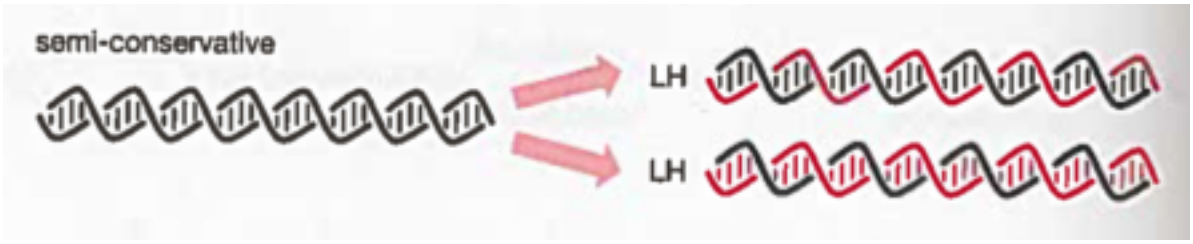
Replication errors cause mutations in the genetic code, which may be heritable if in germ-line cells or somatic if in other cells (e.g. cancer)

In the mammalian cell cycle, DNA replication occurs during 'S' phase (synthesis phase), preceding mitosis (M), and is tightly regulated

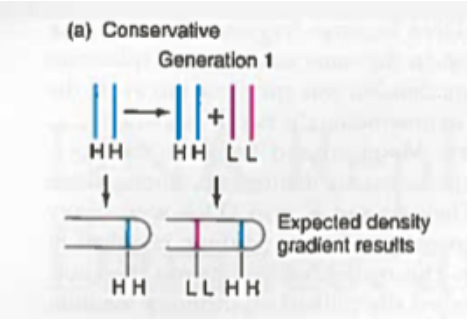
DNA replication must occur only once per cell division cycle. Daughter cells must receive only one genome copy.

Replication "starts" at many points: needs co-ordination. Otherwise you get chromosome instability & mis-expressed genes

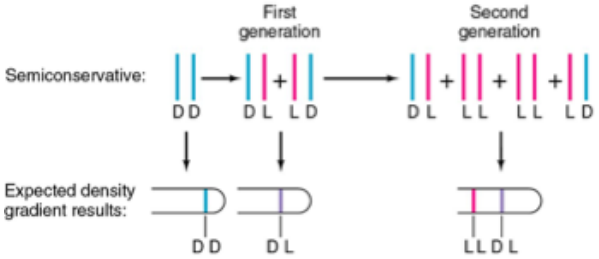
"**Semi-conservative replication**": parental DNA strand in each daughter helices



- Key experiment:
    - In 1958, Meselson & Stahl performed a CsCL equilibrium **density-gradient centrifugation** to distinguish between *E. coli* duplex DNA containing only ^{14}N only ^{15}N or a mixture of both
    - Parental DNA labelled by growing bacteria in ^{15}N for several generations – H/H duplex only; Medium abruptly changed to contain only ^{14}N and samples tested periodically.
- Expected results:
    - **Conservative**: the 2 "heavy" parental strands remain together. Daughter duplex has two "light" strands. Both will separate from each other on gradient



- **Semi-Conservative**: after 1 generation, the 2 “heavy” parental strands separate and will bind to a “light” partner creating a hybrid with a density halfway between H/H parental & L/L ordinary



- Actual results:

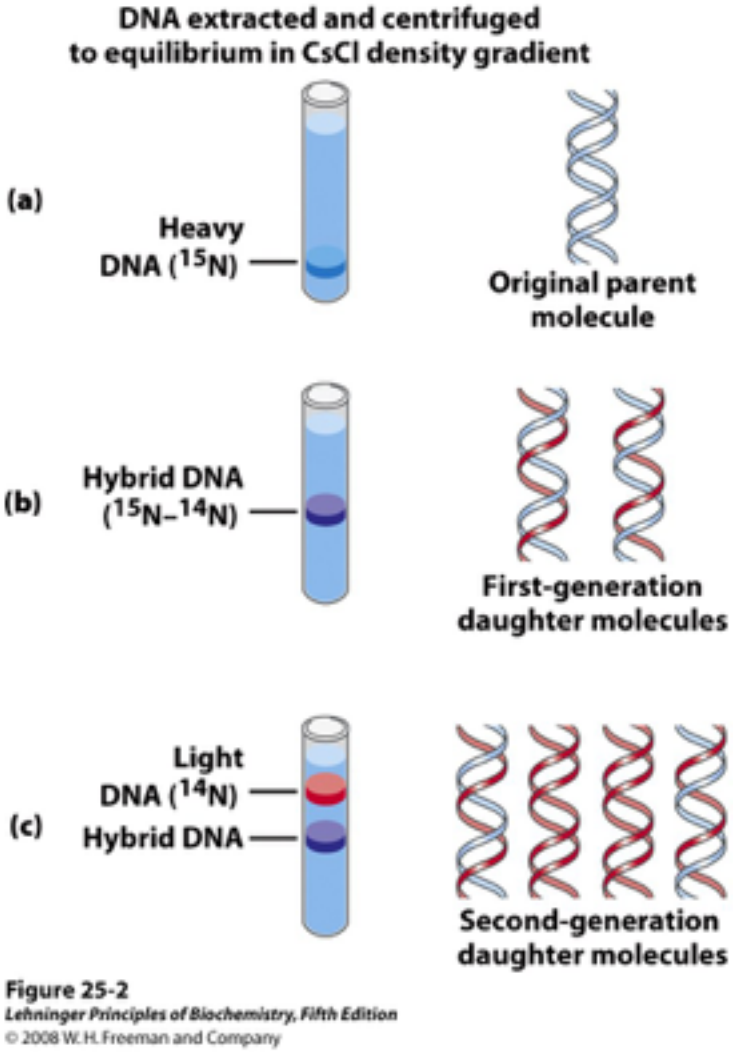


Figure 25-2
Lehninger Principles of Biochemistry, Fifth Edition

Replicating circular (bacterial) chromosomes appear as θ structures - known as replication **eyes** or **bubbles**.
Electron microscopy studies show two growing **replication forks** potentially moving away from a central **origin…**
Replication forks are bidirectional

- **DNA POLYMERASES**:
    - Use single-stranded DNA as a template (are template-directed enzymes)
    - Add dNTPs to free 3'-OH of a base-paired nucleotide to synthesise a complementary strand
    - Catalyse the nucleophilic attack by the 3' on the α phosphoryl group of the nucleoside triphosphate to be added
    - Incoming nucleotides are selected by ability to form Watson-Crick base pairs with template; The binding of the dNTP containing the proper base is favoured by the formation of a base pair with its partner on the template strand
    - New DNA strand forms duplex with template strand.
    - Synthesis is rapid (up to 1000 nucleotides per second)
    - Have proof-reading activity to ensure accuracy
      Dna polymerase goes only 5' to 3'. Proof reading mechanism is able to correct the errors and repairs them.

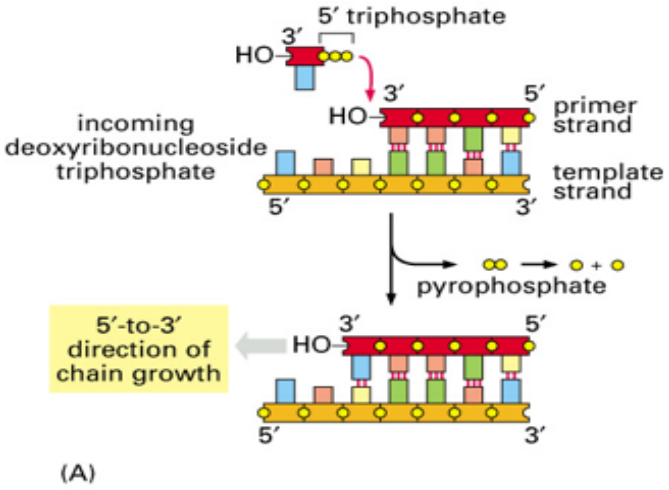


Figure 4–5 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

There are two major DNA polymerases

- **DNA polymerase III**
    - 5' – 3' polymerase activity
    - 3' – 5' exonuclease activity (to remove the error on the other side while polymerase is adding towards 3')

- **DNA polymerase I**
    - 5' – 3' polymerase activity
    - 3' – 5' exonuclease activity
    - 5' – 3' exonuclease activity → involved in different stage of DNA replication

DNA replication requires template strands. However duplex DNA is very stable – very high temperatures are required to separate the strands *in vitro*. --> **DNA HELICASE**

- A diverse group of enzymes
- Harness the hydrolysis of ATP
- unwind short sections of AT-rich parental DNA duplex DNA
- Specifically at recognised **origins of replication**.

DNA helicase binds ssDNA and continues moving along the strand when it encounters dsDNA, thus prising the helix apart at a rate of 1000 bp per second (same rate as DNA polymerase) is coordinated

- How are the separated DNA strands prevented from re-annealing? They are after all, spatially close & aligned:

- **Single-stranded DNA binding protein (SSB)**

    - Keep strands apart
    - Stop the formation of secondary structures (e.g. hairpins)
    - Help align strands
    - Interact with other replication proteins at the replication forks
    - Stimulate polymerases
    - Each SSB protein prefers to bind next to a previously-bound protein. This **cooperative binding** straightens out the DNA template.

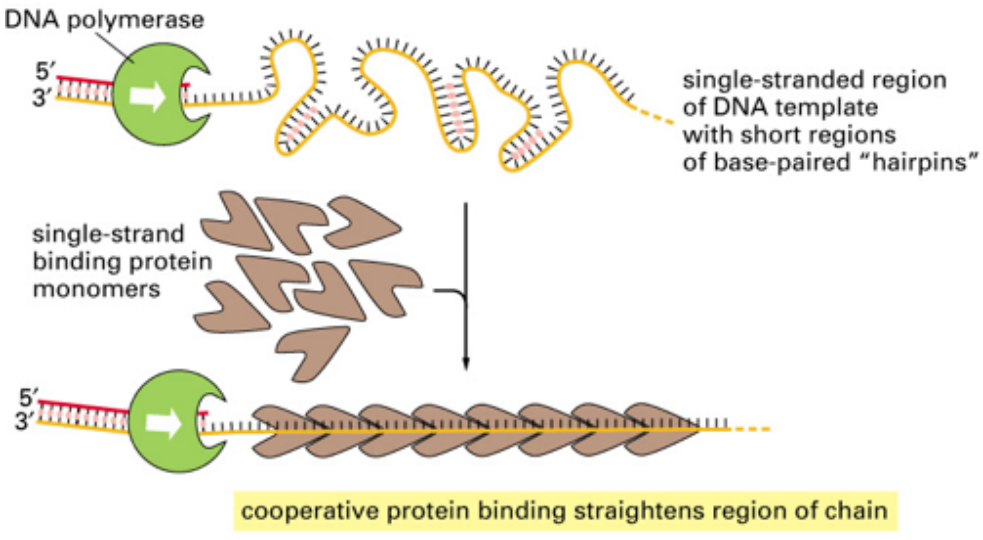
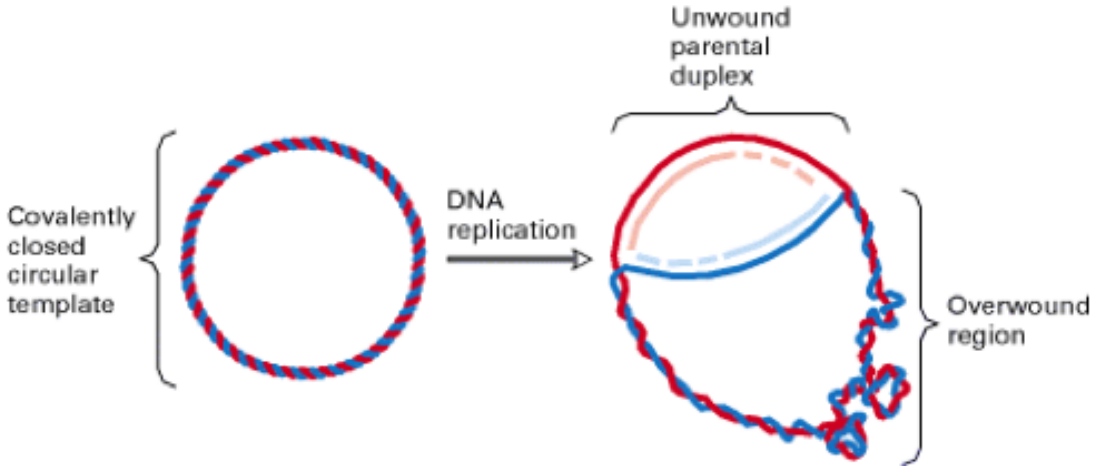


Figure 5–17. Molecular Biology of the Cell, 4th Edition.

DNA is a double helix whose strands wrap around each other once every 10.5 bp. In order for the replication fork to advance, the helix ahead would have to **rotate rapidly**. If 4.6 Mbp *E. coli* genome were to be unwrapped, it would need 11,000 revolutions min^{-1}. Impractical & damaging

Topological issue: Circular DNA or Eukaryotic chromosomes, ends aren't free to unwind…as the DNA unwinds, **positive supercoils** form ahead of the helicase.

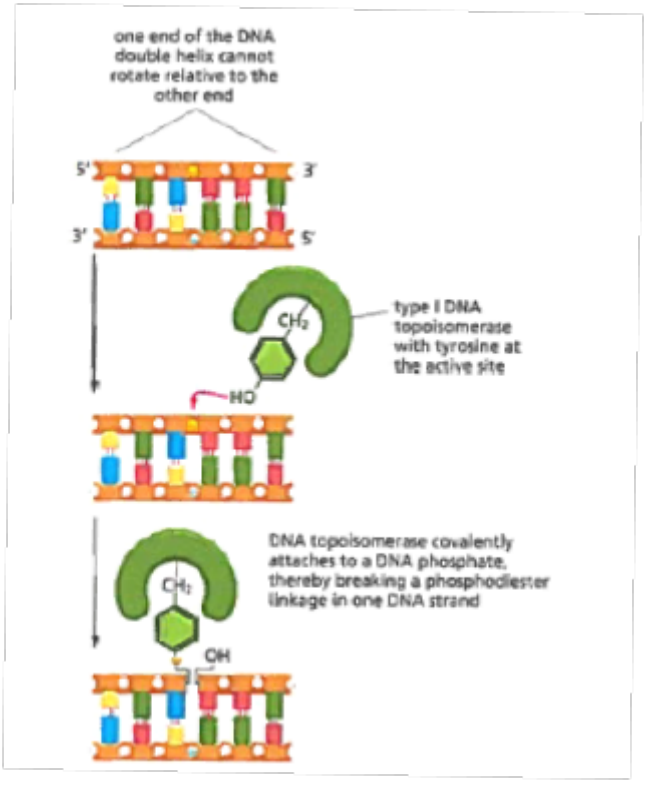


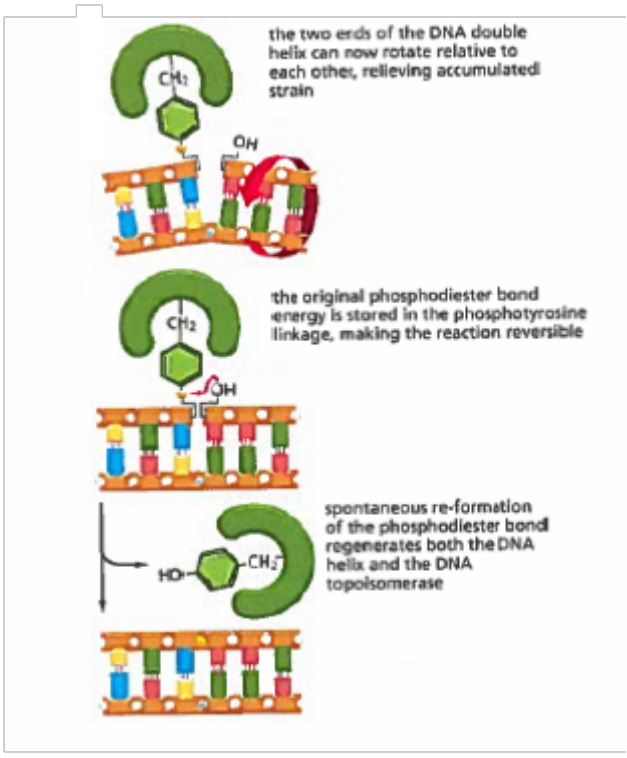
As DNA unwinds superhelical turns will be introduced meaning the linking number (the no. of times one strand crosses another) will remain the same.

This tightening of the helix will create intolerable strain and the energy required for unwinding the DNA will become too great unless it is relaxed:

Bacterial, archaeal & eukarayla genomes all encode **Topoisomerase** enzymes:

- 2 Classes (I & II): I makes ss breaks, II makes staggered ds breaks
- Cause Transient interruptions of DNA backbone
- Then reseal them





- DNA polymerase cannot initiate chain synthesis. It can only add a nucleotide to the 3’ end of a base-paired nucleotide on the primer strand.

  The Solution: **DNA primase**

  catalyses synthesis of a short RNA primer at the origin of replication, then stops. (Unlike DNA polymerase, it can start a new polynucleotide chain by joining two ribonucleotides together).

  RNA primes the synthesis of DNA (DNA primase is an RNA polymerase); after DNA synthesis has been initiated, the short stretch of RNA is hydrolized and replaced with DNA.

  DNA polymerase can then catalyse the addition of deoxynucleotides to the 3’ end.

  DNA helicase and DNA primase together comprise the primosome.

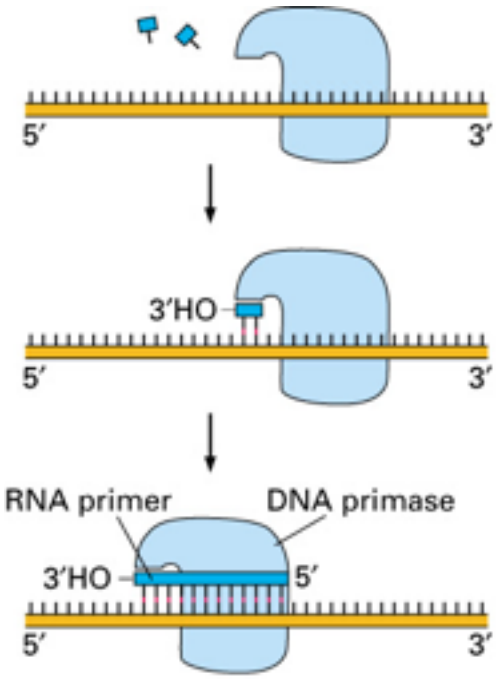


Figure 5–12. Molecular Biology of the Cell, 4th Edition.

DNA REPLICATION: THE MECHANISM

The leading strand is continuously synthesised 5'-3', in the direction of the replication fork movement

The lagging strand is also synthesised 5'-3' but discontinuously, as a series of short DNA pieces called **Okazaki fragments**. These are then joined to make long nascent DNA chains

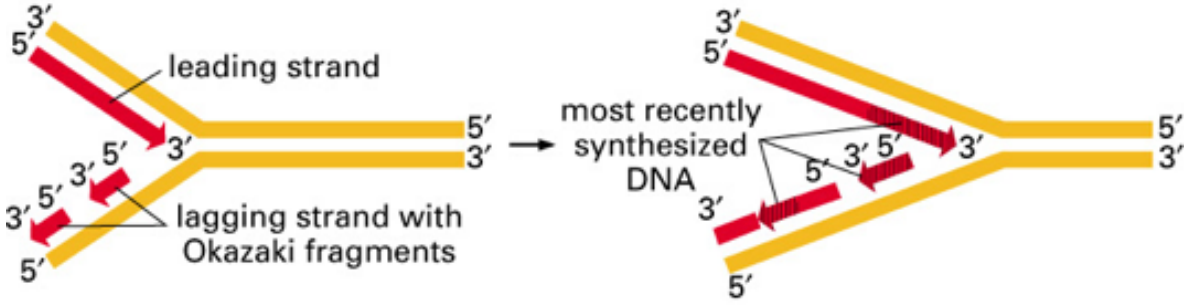


Figure 5–8. Molecular Biology of the Cell, 4th Edition.

While leading strand synthesis requires a single RNA primer, lagging strand synthesis requires multiple RNA primers, one for each segment – synthesised by DNA primase

Elongation of the lagging strand is in the opposite direction to the direction of replication fork advance
DNA replication is therefore semi-discontinuous

Since DNA replication is bi-directional, each strand in the duplex will serve as the template

Each replication fork will have a leading strand and a lagging strand

From a single origin (*OriC*, AT-rich in *E.coli*) these 2 forks advance in opposite directions at constant speed & meet up ~ ½ way round chromosome.

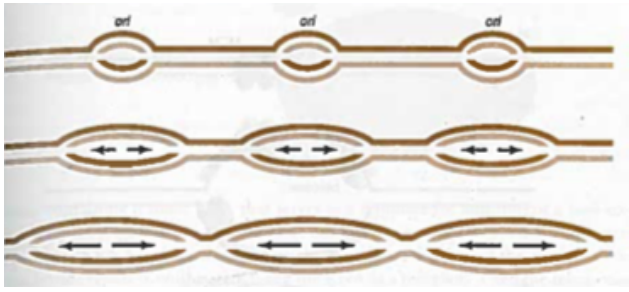
OriC is bound by an initiator protein DnaA. Opens up a ~45 bp segment into single strands. DnaC binds and permits helicase, DnaB, binding.

Replication terminates at specific *Ter* sites

At the end of replication, the two double-stranded daughter chromosomes are interlocked and need to be separated by a type II DNA topoisomerases

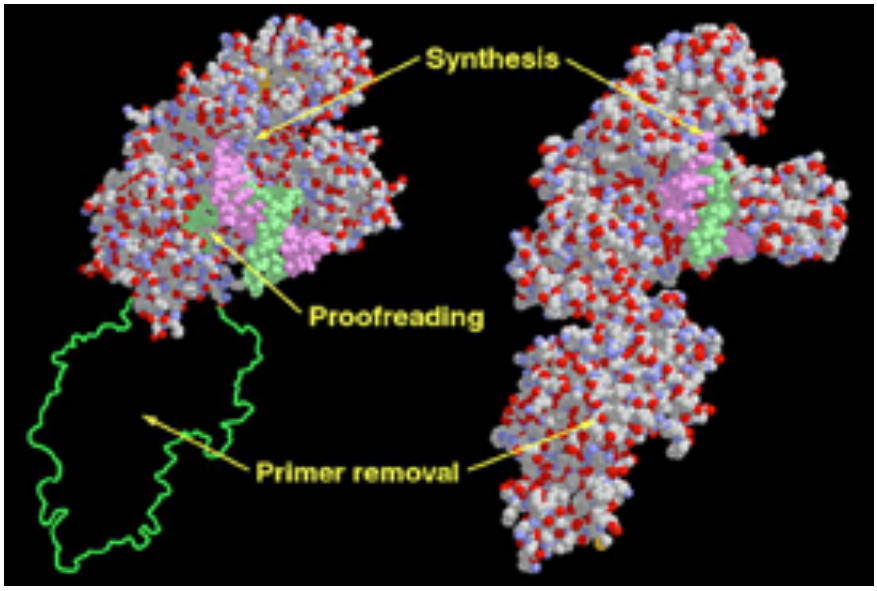
There are thousands of origins and replication forks in a eukaryotic cell….one every 3 to 300 kb….

A eukaryotic chromosome typically contains 60 times more DNA than a prokaryotic chromosome. If operating from a single origin replication could take months



- **DNA Polymerase I:**
    - 5’ – 3’ polymerase activity
    - 3’ – 5’ exonuclease activity
    - very important for proof-reading

- 5’-- 3’ exonuclease activity
  - allows degradation of a strand ahead of the advancing polymerase.
  Useful in DNA repair and **removal of RNA primers** (‘Nick Translation’ on the lagging strand)



103 kDa protein
proteolysis (trypsin/subtilisin)
68kDa = **Klenow** fragment (**3’-5’ exonuclease + polymerase**)
35kDa = **5’-3’ exonuclease**
The **Klenow Fragment** is often used in labs when you require DNA synthesis but no destruction of strand. It is often used for end “filling”…
DNA polymerase I (‘Nick translation’) can be used for the radioactive labelling of single-stranded DNA probes. Use DNase to create ss nicks in dsDNA. Add Pol I with ^{32}P-labelled dNTPs. It replaces unlabelled nucleotides with labelled nucleotides

Comparison of DNA Polymerases of *E. coli*

	DNA polymerase		
	I	II	III
Structural gene*	*pol*A	*pol*B	*pol*C (*dna*E)
Subunits (number of different types)	1	≥4	≥10
M_r	103,000	88,000†	830,000
3′→5′ Exonuclease (proofreading)	Yes	Yes	Yes
5′→3′ Exonuclease	Yes	No	No
Polymerization rate (nucleotides/sec)	16–20	40	250–1,000
Processivity (nucleotides added before polymerase dissociates)	3–200	1,500	≥500,000

- ***E. coli* DNA polymerase III**
    - Absence of Pol III is lethal. The enzyme is actually a 'holoenzyme' (multisubunit)
    - Will perform 3'-5' exonuclease (proof reading) but not 5'-3' (so no nick translation..)

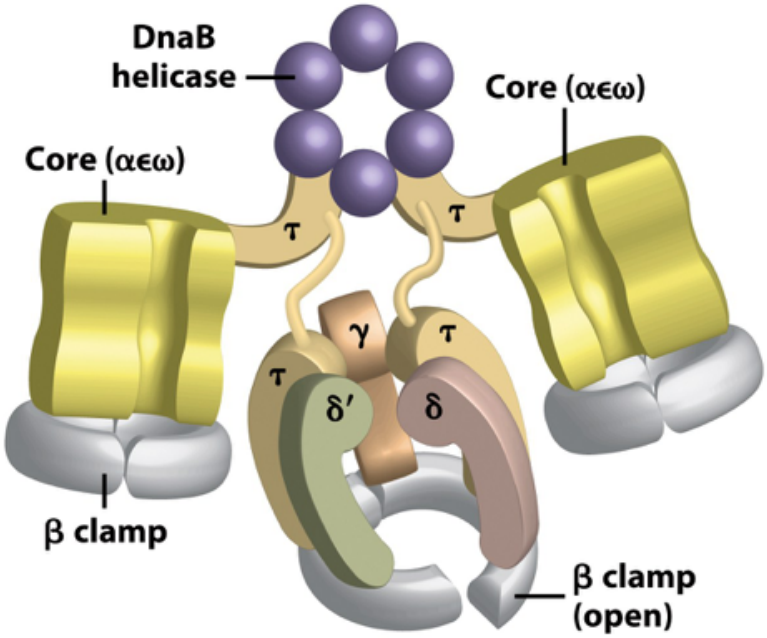


Figure 25-10a
Lehninger Principles of Biochemistry, Fifth Edition

    - Two core domains, composed of subunits ⟨, ∑, and ⟨, are linked by a five-subunit clamp-loading complex (also known as the © complex) with the composition τ_2©™™′. The γ and τ subunits are encoded by the same gene. The γ subunit is a shortened version of the ⁄ subunit; τ thus contains a domain identical to © along with an additional segment that interacts with the core polymerase. The other two subunits of DNA polymerase III*, ⁄ and ⁄ (not shown), also bind to the clamp-loading complex. Two ® clamps interact with the two-core subassembly; each clamp a dimer of the ® subunit. The complex interacts with the DnaB helicase through the ⁄ subunits.

If an enzyme carries out a single reaction and then dissociates from substrates= **distributive**
If an enzyme performs multiple actions before dissociating = **processive**

DNA polymerase can dissociate easily from the template, allowing it to recycle and begin synthesis of the next Okazaki fragment

The sliding clamp keeps the DNA polymerase on the DNA while it is moving – and increases the enzyme's processivity

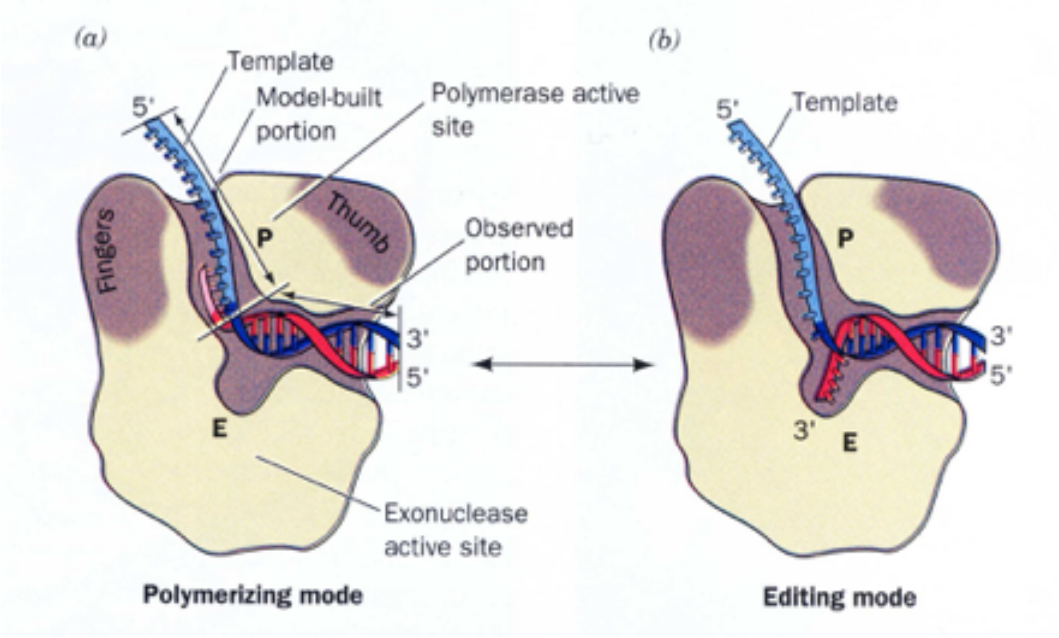
When the polymerase encounters a ds region, the clamp releases it

Reliance on complementary base pairing is not sufficient to ensure accuracy (non-Watson-Crick base-pairing is possible and does occur)

Proof-reading mechanisms ensure that fidelity is very high – only 1 error for every 10^9 nucleotides copied

- Correct nucleotide has higher affinity for polymerase as it correctly base-pair with template. Phosphodiester bond formation involves conformational change in DNA polymerase so incorrectly-bound nucleotides do not fit active site.

- **3'-5' exonuclease activity of DNA polymerases**
  - takes place immediately after incorrect addition to growing chain.
  - polymerase can't extend such a strand- requires a base-paired 3'-OH end of the primer strand



-The newly synthesised DNA transiently unpairs and the polymerase undergoes a **conformational change**, moving the editing catalytic site into place for removal.

- Replication involves a lot of proteins & in a confined space. The parental DNA has strands of opposite polarity. If replication on either strand was independent:
  - Pol III's could end up 3 kb away from one another…
  - DNA helicase could generate excess lagging-strand template
  - DnaG could be lost from the replication fork…

So, the lagging strand template DNA loops round, bringing the two DNA polymerases into a complex. This brings the 3' end of a completed Okazaki fragment close to the start site for the next fragment. Known as the "trombone model"
The lagging strand is looped so that DNA synthesis proceeds steadily on both the leading and lagging strand templates at the same time. Red arrows indicate the 3' end of the two new strands and the direction of DNA synthesis. The heavy black arrows show the direction of movement of

the parent DNA through the complex. An Okazaki fragment is being synthesized on the lagging strand.

The leading strands are synthesised simultaneously, although lagging strand synthesis begins a little later than the other. Unwinding is rate-limiting step

- As DNA unwound 3'-OH is continuously replicated ('**Leading Strand Synthesis**')
- DNA unwinding and synthesis is concomitant
- After sufficient ssDNA produced can '**lagging strand synthesis**' occur
- Discontinuous due to Okazaki fragment creation & RNA primer synthesis
- As lagging strand template is exposed it binds ssDNA-binding protein

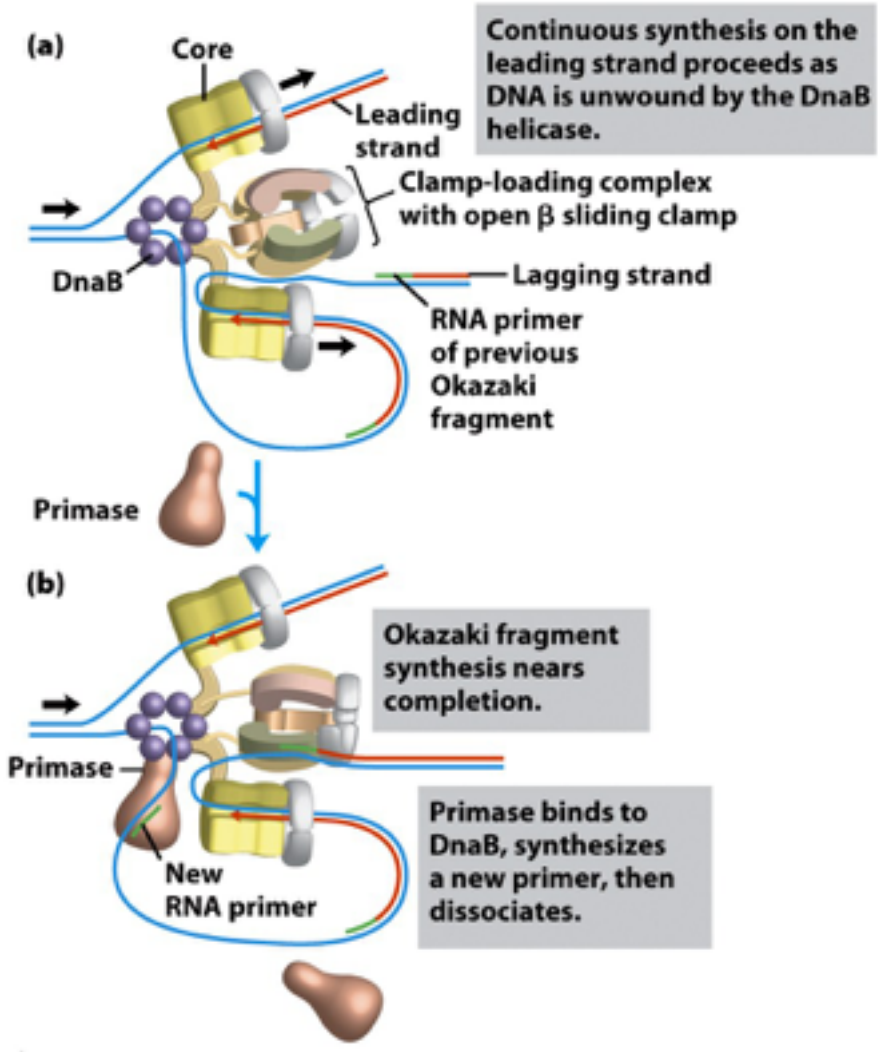


Figure 25-14 part 1
Lehninger Principles of Biochemistry, Fifth Edition

- **Lagging strand synthesis:**
- After completing a fragment the lagging strand holoenzyme relocates to a new primer.
- Then locked in with β-clamp (about 1 a second)

- *E. coli* replicates its DNA at a rate of ~1000 nt/s. Therefore, in lagging strand synthesis the DNA polymerase must be reloaded into the template strand every second or so.

- When the polymerase encounters the previously synthesised fragment, the Pol III core releases the DNA and loses its affinity for the β-clamp. However, it is held in place by connections to the Pol III core involved in leading strand synthesis.

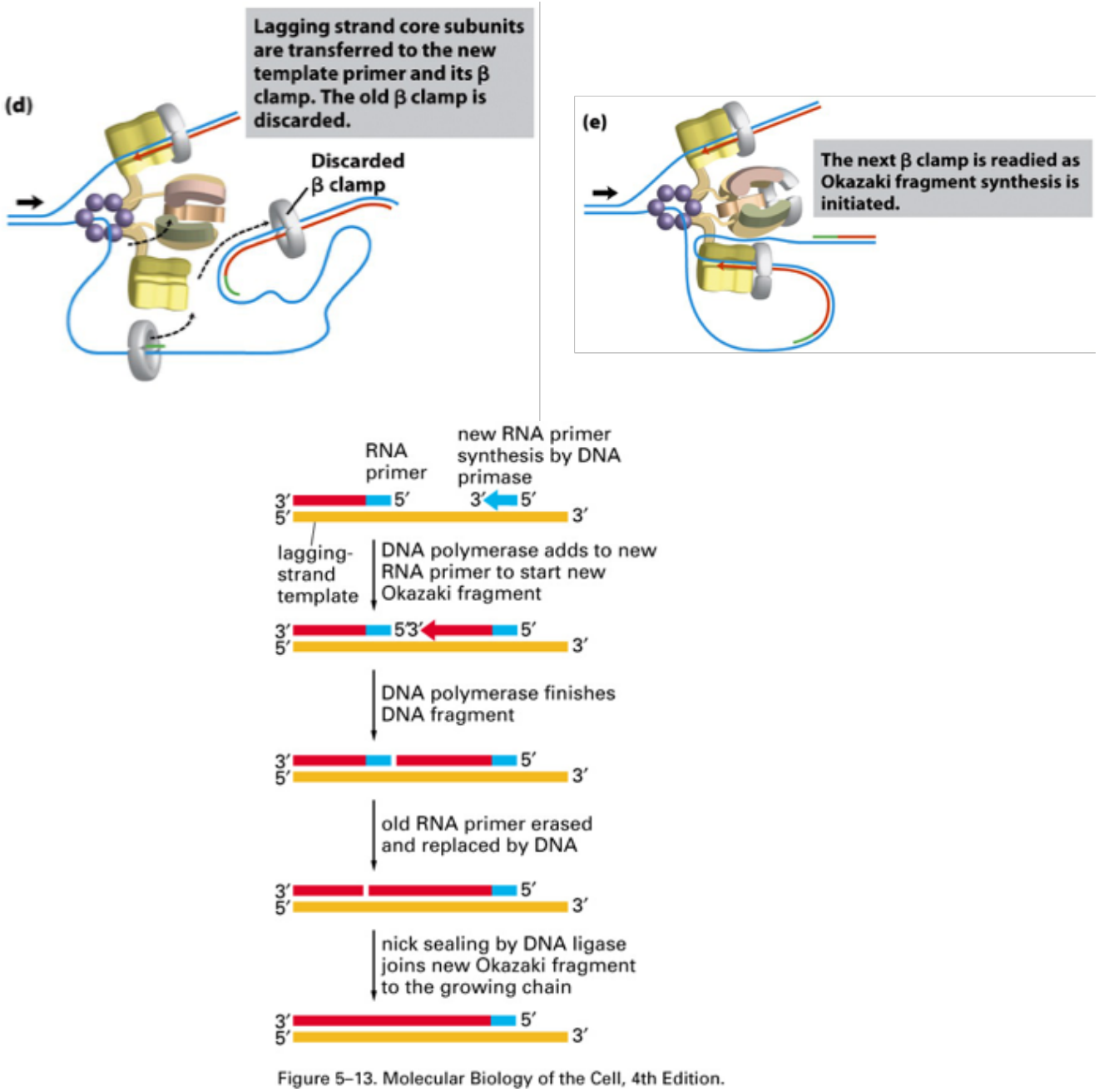


Figure 5–13. Molecular Biology of the Cell, 4th Edition.

DNA ligase : seals the nick, requires energy input
The 3'-OH and the 5' phosphate groups are joined together, catalysed by DNA ligase, in a reaction requiring ATP (in euakryotes, above), or NAD (nicotinamide adenine dinucleotide) in *E. coli*

RNA Transcription in Prokaryotes - Proteins and Mechanism

Similarities between transcription and replication:

- Fundamental chemical mechanism: Addition of nucleotides to 3' end of growing chain
- Polarity of polynucleotide growth (5' to 3')
- Use of a DNA template
- Three phases: Initiation, Elongation and Termination

Differences:

- Does not require a primer
- Not all of the DNA is transcribed
- Only one strand of a DNA template is transcribed by RNA polymerase

RNA POLYMERASE:

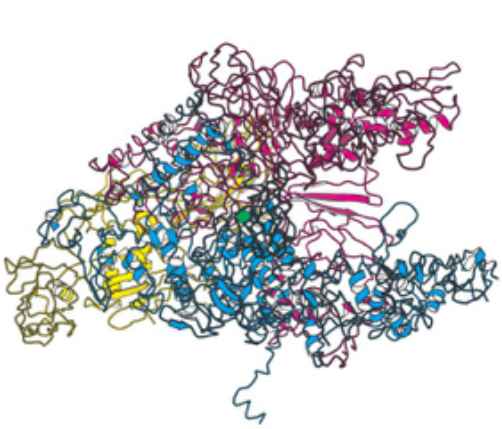
- Prokaryotic cells such as *E.coli* have one RNA polymerase
- Eukaryotic cells have three kinds
- 3D structure highly conserved
- Implies that catalytic basis identical
- $(NMP)_n + NTP \rightarrow (NMP)_{n+1} + PP_i$
- Requires DNA for activity, ribonucleotides (ATP, CTP, GTP, UTP) and Mg^{2+}

- ***Escherichia coli* RNA polymerase:**
  - Very big: ~400kD
  - Five kinds of subunits: α, β, β', ω, σ
  - Subunit composition of enzyme is $\alpha_2\beta\beta'\ \omega\ \sigma$

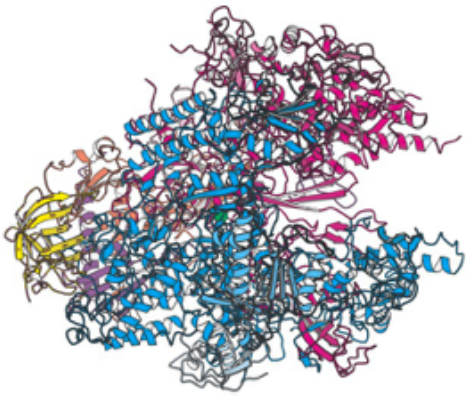
    Called the holoenzyme

    σ helps enzyme to recognize specific DNA sequences called promoter, initiate transcription, then dissociates

    Leaves the core enzyme $\alpha_2\beta\beta'\ \omega$ which carries out catalysis / chain elongation



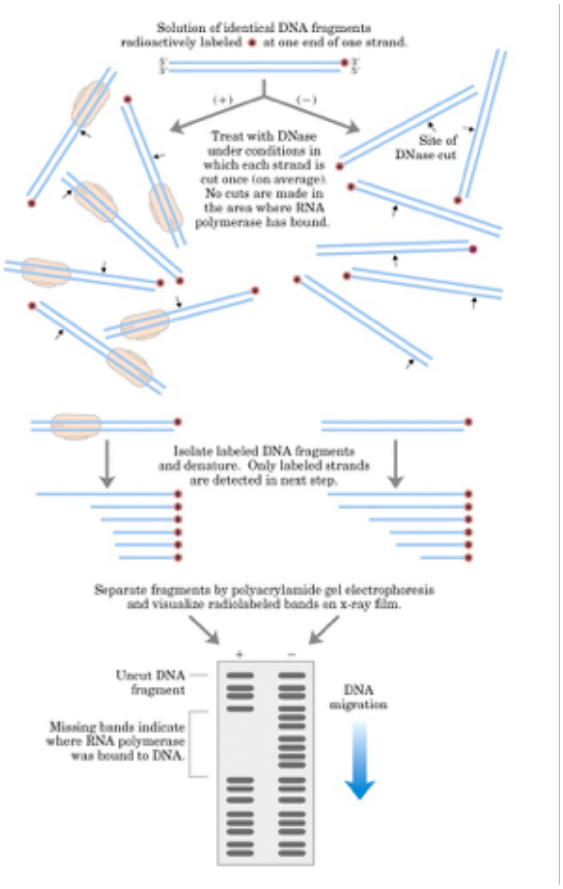
Prokaryotic RNA polymerase

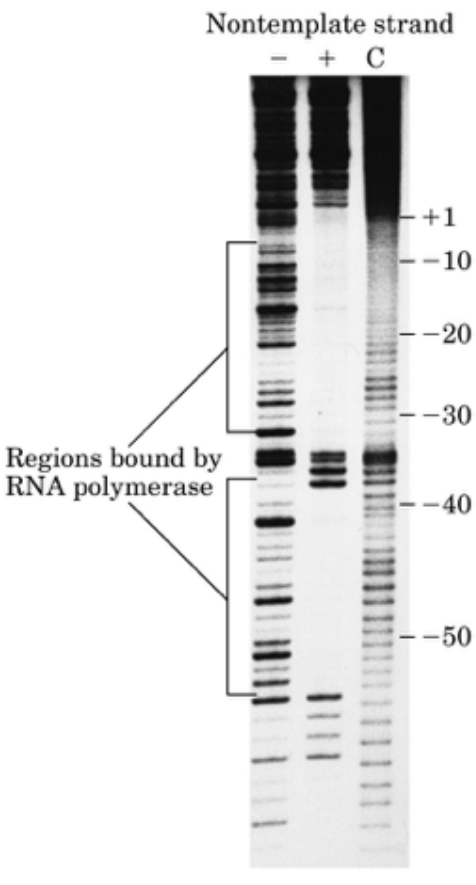


Eukaryotic RNA polymerase

- **TRANSCRIPTION INITIATION**:
    - The key step in transcription - it's the decision to express a gene
    - Holoenzyme binds to about 70bp before transcription start site
    - DNA protein binding can be determined by DNA footprinting experiments
    - Critical, conserved sequences occur at -35 and -10 (Pribnow box) regions

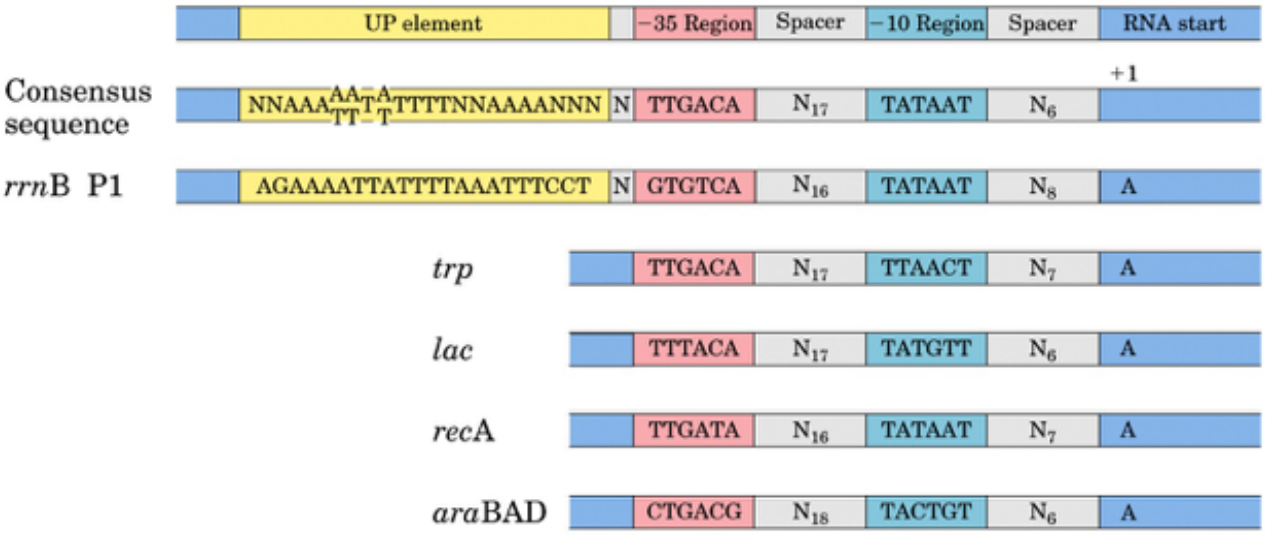
- DNA FOOTPRINTING: a technique derived from principles used in DNA sequencing, identifies the DNA sequences bound by a particular protein. Researchers isolate a DNA fragment thought to contain sequences recognized by a DNA-binding protein and radiolabel one end of one strand (FiS. 1). They then use chemical or enzymatic reagents to introduce random breaks in the DNA fragment (averaging about one per molecule). Separation of the labeled cleavage products (broken fragments of various lengths) by high-resolution electrophoresis produces a Ladder of radioactive bands. In a separate tube, the cleavage procedure is repeated on copies of the same DNA fragment in the presence of the DNA-binding protein. The researchers then subject the two sets of cleavage products to electrophoresis and compare them side by side. A gap ("footprint") in the series of radioactive bands derived from the DNA-protein sample, attributable to protection of the DNA by the bound protein, identifies the sequences that the protein binds. The precise location of the protein binding site can be determined by directly sequencing (see Fig. 8-34) copies of the same DNA fragment and including the sequencing lanes (not shown here) on the same gel with the footprint. Figure 2 shows footprinting results for the binding of RNA polymerase to a DNA fragment containing a promoter. The polymerase covers 60 to 80 bp; protection by the bound enzyme includes the - 10 and -35 regions.





- ***Escherichia coli* promoters:**

  Typical *E. coli* promoters recognized by an RNA polymerase holoenzyme containing sigma70. Sequences of the non-template strand are shown, read in the 5'-3' direction, as is the convention for representations of this kind. The sequences vary from one promoter to the next, but comparisons of many promoters reveal similarities, particularly in the -10 and -35 regions. The sequence UP, not present in all *E.coli* promoters, is shown in the P1 promoter for the highly expressed gene *rrnB*. UP elements, generally occurring in the region between -40 and -60, strongly stimulate transcription at the promoters that contain them. The UP element in the *rrnB* P1 promoter encompasses the region between -38 and -59. The consensus sequence for *E.coli* promoters recognized by sigma70 is shown second from the top. Spacer regions contain slightly variable numbers of nucleotides (N). Only the first nucleotide coding the RNA transcript (at position +1) is shown.



  - Clear similarities, particularly in the -10 and -35 regions
  - UP elements strongly stimulate transcription at the promoters that contain them and gene *rrnB is highly expressed*

- **Role of σ initiation:**
  - Decreases ability of core enzyme to bind DNA non-specifically (by about 10^4)
  - Allows holoenzyme to bind promoters (Binds -10 sequence)
  - Allows holoenzyme to migrate along the DNA until a promoter is encountered (random walk)
  - Different σ factors permit binding to different promoters (Allows for specific, regulated gene expression)

- β' subunit

  Largest subunit coded by the rpoC gene

  Responsible for RNA synthesis

- β subunit

  Second-largest subunit coded by the rpoB gene

  Responsible for RNA synthesis

- αI and αII:

  Third-largest subunit present in two copies, each with N-Terminal domain (αNTD) and C-terminal domain (αCTD)

  αNTD needed for assembly of RNAP

  αCTD interacts with promoter

- ω subunit

  Smallest subunit

  Required for assembly of RNAP and stabilizes it

- **TRANSCRIPTION INITIATION:**

  A series of changes occur in both the DNA and RNA

  Polymerase upon binding promoters

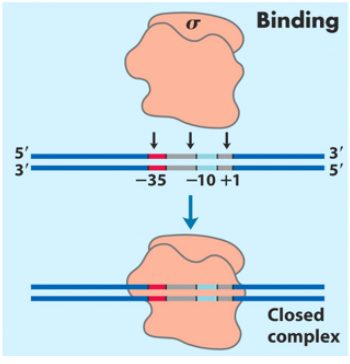
  Closed complex formation (DNA intact)

  Open complex formation (DNA partially unwound)

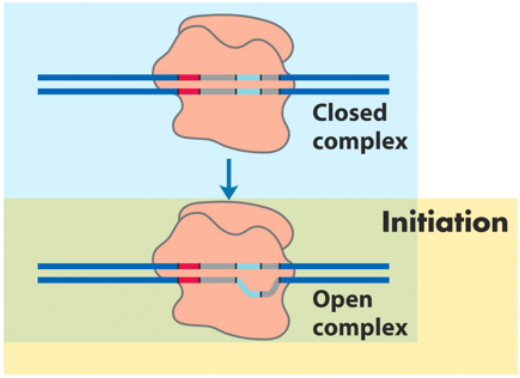
  Initiation (add the first ribonucleotide)

  Promoter clearance (loss of σ and change in holoenzyme to core enzyme - the elongation form of RNA polymerase)

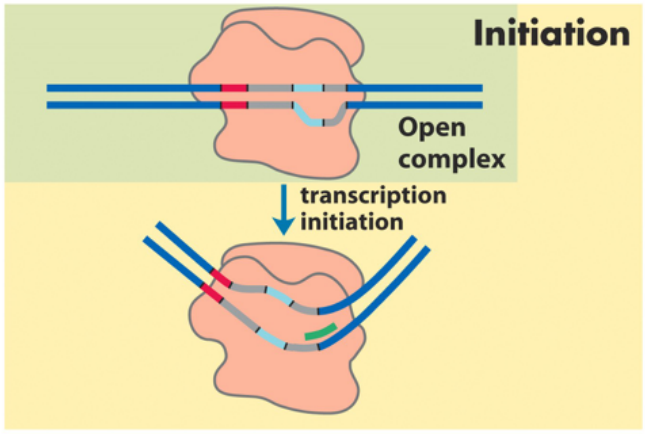
Transcription initiation and elongation by *E.coli* RNA polymerase. Initiation of transcription requires several steps generally divided into two phases, binding and initiation. In the binding phase, the initial interaction of the RNA polymerase with the promoter leads to formation of a closed complex, in which the promoter DNA is stably bound but not unwound.



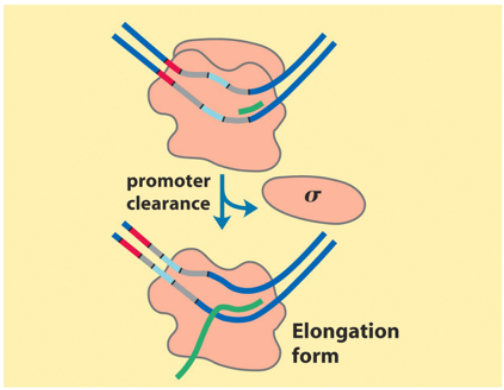
A 12 to 15 bp region of DNA from within the -10 region to position +2 or +3 is then unwound to form an open complex. Additional intermediates (not shown) have been detected in the pathways leading to the closed and open complexes, along with several changes in protein conformation.



The RNA polymerase transcribes the DNA, which is initiated by the beta subunit
Initially produces about 10 abortive transcripts of 2-8 nucleotides as exit channel blocked by the σ-factor



The initiation phase encompasses transcription initiation and promoter clearance. Once elongation commences, the sigma subunit is released and the polymerase leaves the promoter and becomes committed to elongation of the RNA.



- Only one strand is transcribed

  The strand that serves as a template for RNA polymerase is called the template strand

  The strand complementary to the template is called the non-template or coding strand (Note it is identical in base sequence to the RNA)

- **PRODUCTS OF RNA POLYMERASE:**
    - Messenger RNA (mRNA)
    - Non-coding RNA
    - Transfer RNA tRNA
    - Ribosomal RNA (rRNA)
    - lncRNA
    - Micro RNA—regulates gene activity
    - Catalytic RNA (Ribozyme)—enzymatically active RNA molecules

RNA Transcription in Prokaryotes - 2

Transcription is tightly regulated. Measuring gene expression (microarrays).

- **ELONGATION**

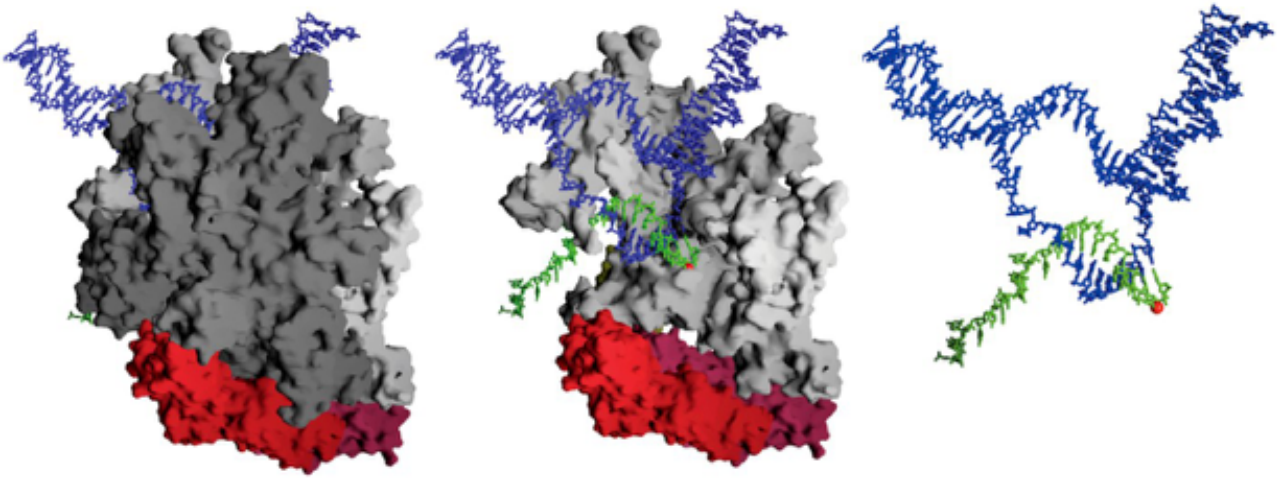
When the Open Complex between DNA and RNA polymerase forms, around 17 bp of DNA at the styart site are unwounded

After promoter clearance, RNA chain is extended and the polymerase moves along the DNA

Unwinds in front and rewinds behind

Keeps 17 bp as a bubble of unwound DNA as it goes.

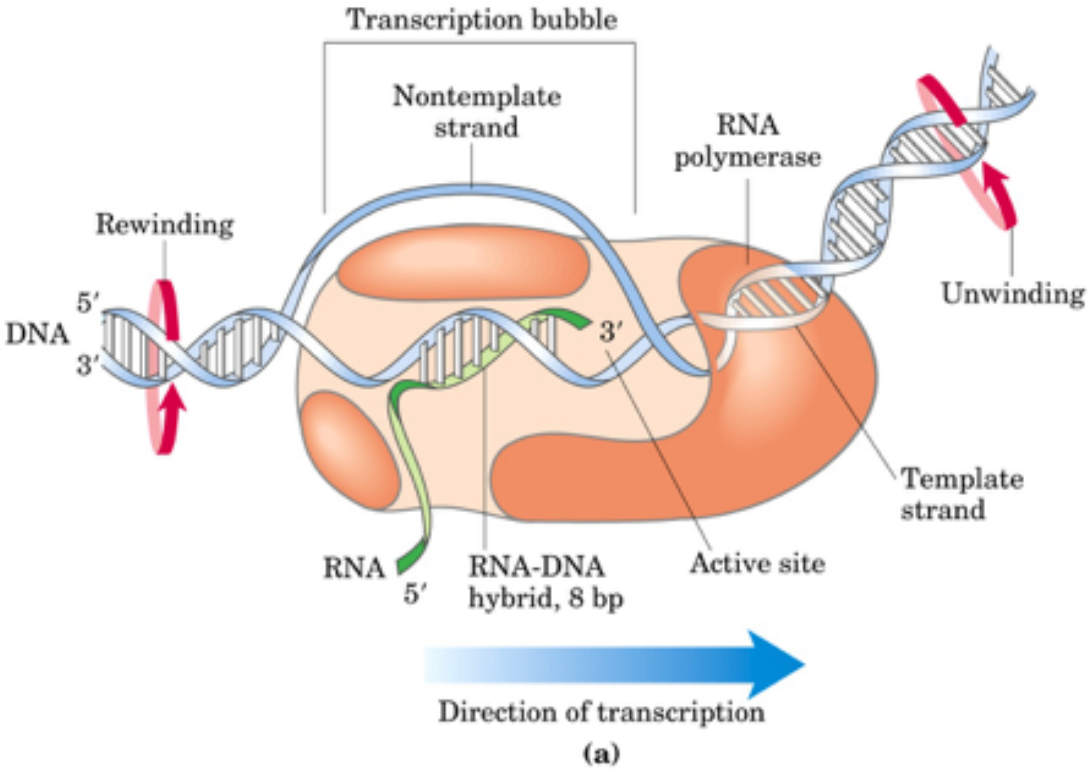
Magnesium ion is necessary for the transcription



Structure of the RNA core polymerase from E.coli. RNA and DNA are included here to illustrate a polymerase in the elongation phase. Subunit coloring matches: the b and b’ subunits are light gray and white; the a subunits, shades of red. The w subunit is on the opposite side of the complex and is not visible in this view. The s subunit is not present in this complex, having dissociated after the initiation steps. The left panel shows the entire complex. The active site for transcription is in

a cleft between the b and b' subunits . In the middle panel, the b subunit has been removed, exposing the active site and the DNA-RNA hybrid region. The active site is marked in part by a Mg^{2+} ion (red). In the right panel, all the protein has been removed to reveal the circuitous path taken by the DNA and RNA through the complex.

Negative supercoiling of circular DNA favours the transcription of genes because it facilitates unwinding. The introduction of negative supercoils into dnA by topoisomerase II can increase the efficiency of the promoters located at distant sites



Not all promoters sites are stimulated by negative supercoiling
The promoter site for topoisomerase II itself is an exception
Negative supercoiling decreases the rate of transcription of this gene. Produces a feedback control ensuring that DNA does not become excessively supercoiled.
Negative supercoiling could decrease the efficiency of this promoter by changing the structural relation of the -10 and -35 regions.

Template DNA strand selects appropriate ribonucleotides by Watson-Crick base pairing
In the transcription bubble, about 8bp of DNA and the newly synthesized RNA chain are bais paired
Transcription bubble moves at ~170Å sec^{-1} or about 50 nucleotides sec^{-1}
Not as fast as DNA polymerase (800 nucleotides sec^{-1}), it highly controlled.
Structure of RNA polymerase forces RNA to **exit** from the helix

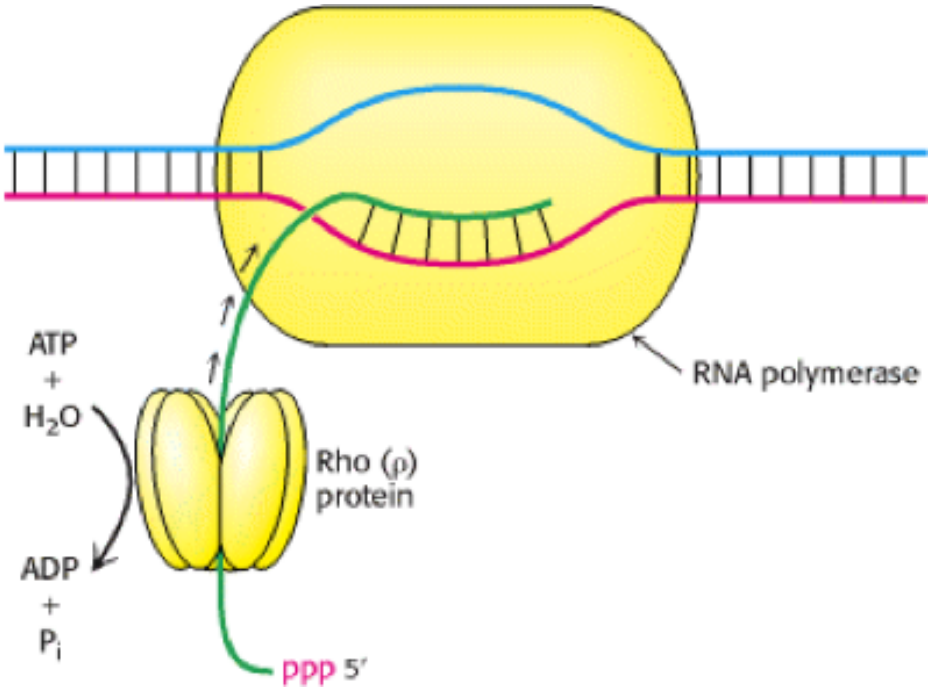
- **TERMINATION**

There are 2 mechanisms in *E. coli*

- Requirement for an additional protein factor, (rho)

Binds CA rich sequences in the RNA

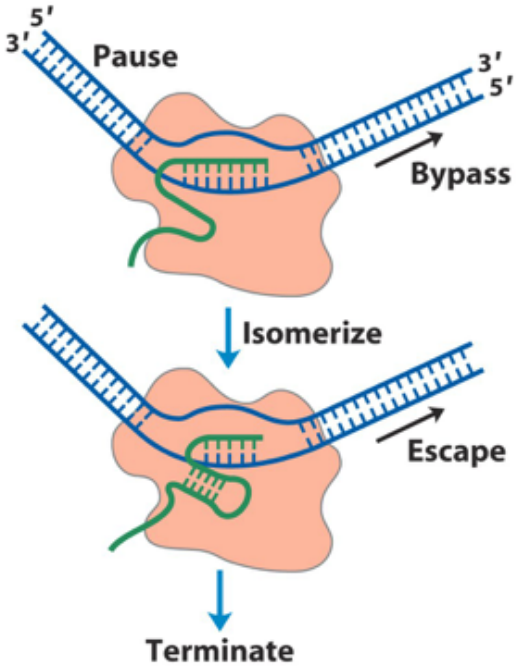
Uses ATP driven helicase activity to unwind the DNA/RNA helix



The ρ protein is an ATP-dependent helicase

Binds the nascent RNA chain and pulls it away from RNA polymerase and the DNA template

- Presence of **hairpin structures** in the RNA followed by UUU which causes polymerase to pause and release transcript due to **weaker A-U base pairs**



- **EUKARYOTIC TRANSCRIPTION**

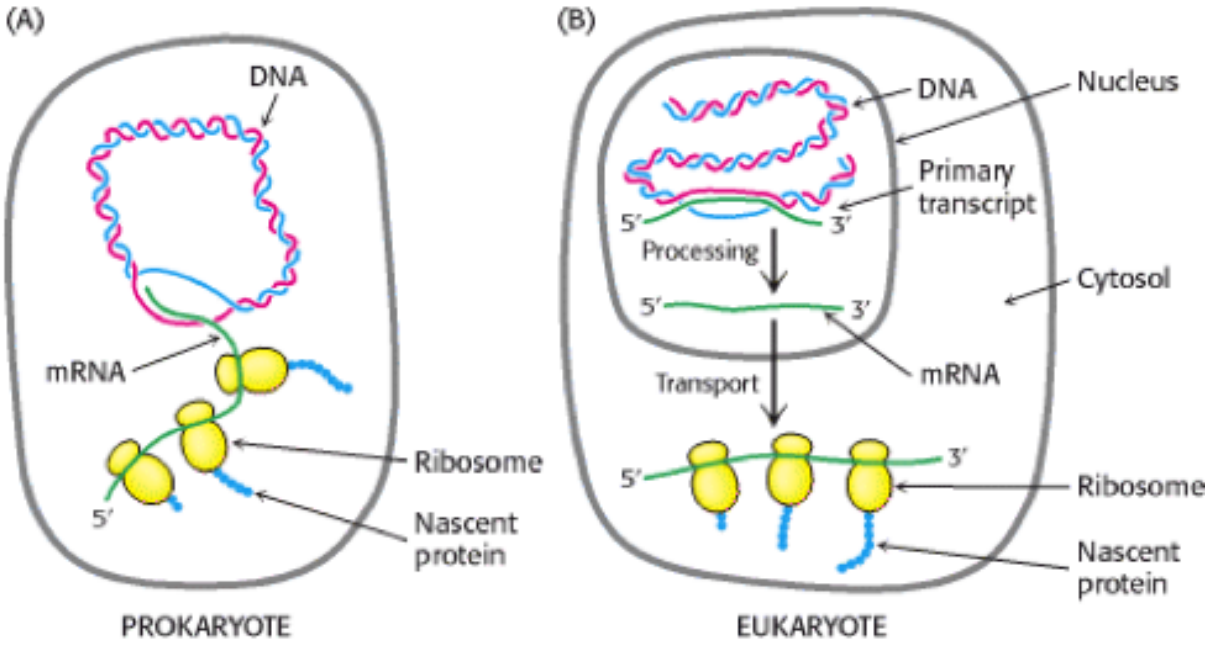
  It is much more complicated than prokaryotes (gene structure is more complicated: there are exons and introns)

- Occurs in the **nucleus** - transcripts are **exported**
  Three types of polymerase which share some subunits
  All are **large** (>500kDa) and **multi-subunit** (>8)

  - **Polymerase I**
    In the nucleolus makes pre-ribosomal RNA (processed to 18S, 5.8S, 28S rRNA)
  - **Polymerase II**
    Makes mRNAs + some small RNAs
    Most susceptible to α-aminitin
    α-Amanitin is produced by the poisonous mushroom *Amanita phalloides*, which is also called the *death cup* or the *destroying angel*. More than a hundred deaths result worldwide each year from the ingestion of poisonous mushrooms. α-Amanitin binds very tightly (*K*d = 10 nM) to RNA polymerase II and thereby blocks the elongation phase of RNA synthesis. Higher concentrations of α-amanitin (1 μM) inhibit polymerase III, whereas polymerase I is insensitive to this toxin. This pattern of sensitivity is highly conserved throughout the animal and plant kingdoms.

  - **Polymerase III**
    Makes tRNAs, 5S rRNA + other small RNAs

- Transcription and translation are closely coupled in prokaryotes, whereas they are spatially and temporally separate in eukaryotes
  In prokaryotes, the primary transcript serves as mRNA and is used immediately as the template for protein synthesis
  In eukaryotes, mRNA precursors are processed and spliced in the nucleus before being transported to the cytosol for translation into protein



Human genome project revealed the human genome contains just 25000 coding genes, which are not enough for the level of complexity of our cells. This is because some genes can encode for completely different proteins (splicing and other processes)

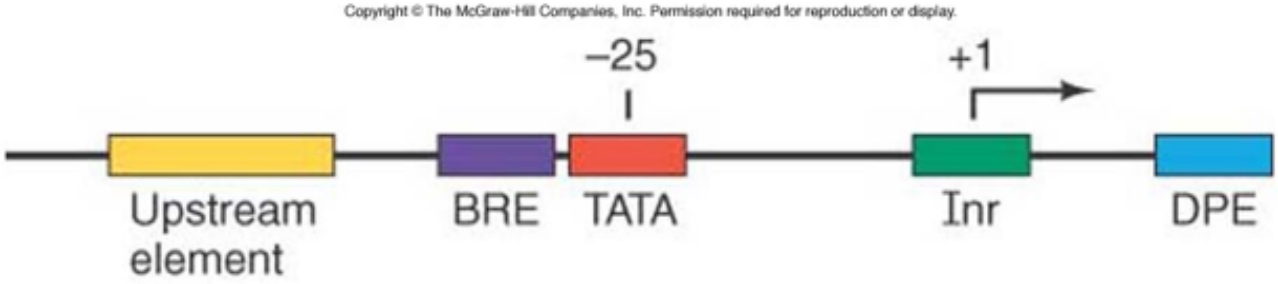
- Eukaryotes promoters are much larger and more complex than prokaryotes

  Bind more proteins (e.g. PolIII promoter sequences downstream, inside regions to be transcribed)

  Pol II promoters similar but more complex than prokaryotic promoters
  Core promoter may have
  "**TATA**" **box** and an initiator element (Inr) sequence near +1
  Upstream element(s) and Downstream element
  TATA box similar to -10 sequence in bacteria

- **TATA box**
  28–34 bp upstream of the TSS
  Associated with strong tissue-specific promoters
- **Initiator element**
  YYANWYY consensus where the A is at position +1
  Independent of the TATA box but can occur together
- **Downstream promoter element (DPE)**
  28–32 bp downstream of TSS in TATA-less promoters of Drosophila melanogaster
  Generally occurs together with Inr elements
  Thought to have similar function as TATA box
- **TFIIB recognition element (BRE)**
  The BRE element, with an SSRCGCC consensus, lies upstream of the TATA box in some TATA-dependant promoters

The TATA and Inr elements are the only known core promoter elements that alone can initiate transcription

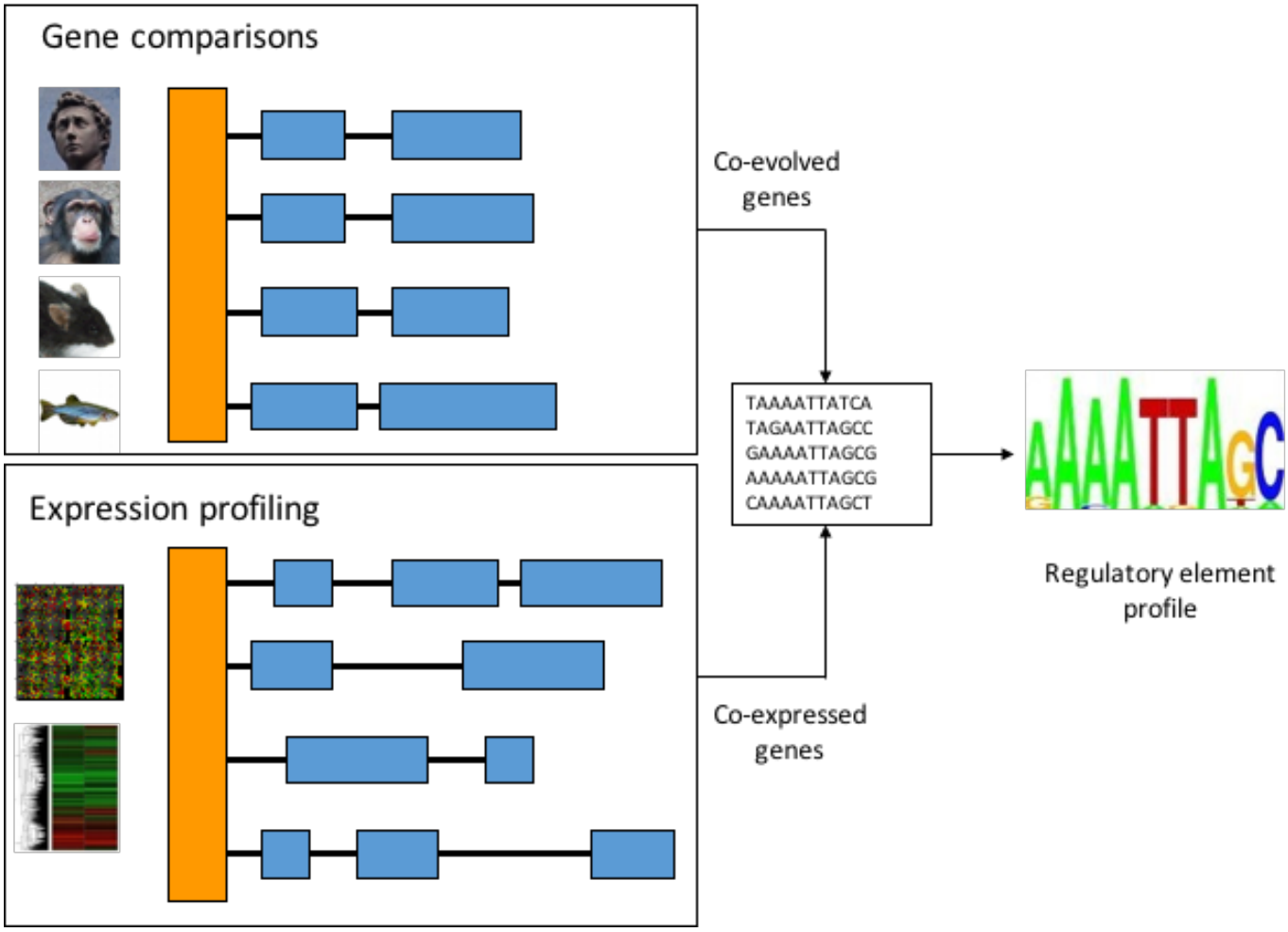


- Polymerase II does not contact these sequences directly and relies upon many other **Basal Transcription Factors** to bind to promoters
  In addition to **Core Promoter** sequences, **Enhancer** sequences my lie several hundred or thousand base pairs away

- Unlike prokaryotic RNA polymerase, Pol II **cannot initiate transcription** independently
  Transcription initiation requires assembly of a **very large transcription complex**
  RNA Pol II + basal transcription factors
  TFIID via TATA Binding Protein (**TBP**) especially important
  Also TFIIA, TFIIB, TFIIF, TFIIE, TFIIH
  All multi-subunit proteins
  Binding of TBP to the TATA box enforces the complex to select a TSS in a limited genomic space
  TBP is always involved, even if no TATA box and binds without sequence specificity

- Identifying where they bind is very difficult--> **Phylogenetic Footprinting**
  Take either a set of gene that are expressed at the same time, they are thus under control of the same transcription factor--> take the upstream regions and identifying any similarities.
  Uses set of homologous regulatory regions from multiple related species
  Alternatively, co-expressed genes in the same species
  Identify best conserved motifs in these regions, taking phylogenetic relationships into account.
  Should find some island of conservation where the transcription factors bind

  There should be evolutionary pressure that conserves those patterns. It is a computational and mathematical approach.



- **RNA MODIFICATION AND PROCESSING**

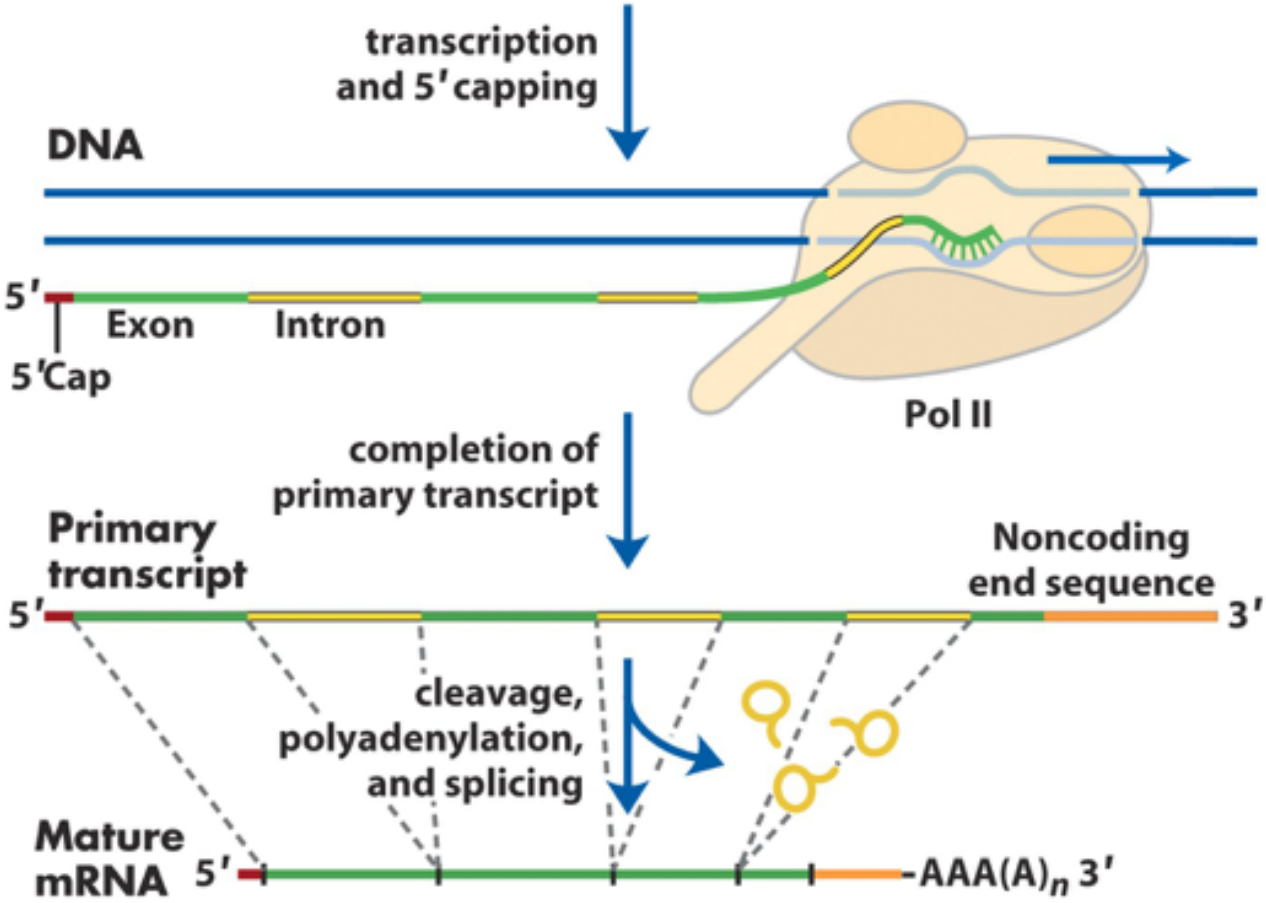
  Nearly all eukaryotic transcripts are **processed**

  - **Modification** of **bases**
  - **Deletions** and **additions** to the **5'** and **3'** ends
  - Removal of sequences (**introns**) within the primary transcript

Eukaryotic mRNAs:

- Have methyl guanosine **CAP** added to **5'** end
- Have polyadenyl tails (polyA) added to **3'** end
- Have introns removed by **splicing**

Introns are spliced out. There are splicing enhancers and silencers. The splicing complex is massive, which binds to mRNA and controls splicing. There are motifs and patterns that control the splicing.



TRANSLATION

Crick and Brenner showed it was triplet by mutational analysis
Proflavin causes mutations by inserting itself between DNA bases
Proflavin-induced mutations of the T4 bacteriophage gene rIIB were created
Mutants with single nucleotide insertion produced non-functional protein due to frameshift
Mutants with 2 or 4 nucleotide mutations also produced non-functional protein
Mutants with a 3 nucleotide insertion produced a functional protein

The presence of **start** (**AUG**) and **stop** (**UAA**, **UGA** and **UAG**) codons determines a sequence of codons
Called an **Open Reading Frame** (**ORF**)
Can be **predicted by computers** in **genome sequence**
In any mRNA sequence, can have **three possible reading frames**
Usually find 2 out of 3 have more stop codons

- **Transfer RNA:**

Serves as the adapter to recognize the **triplet codon** and link it to a particular **amino acid**

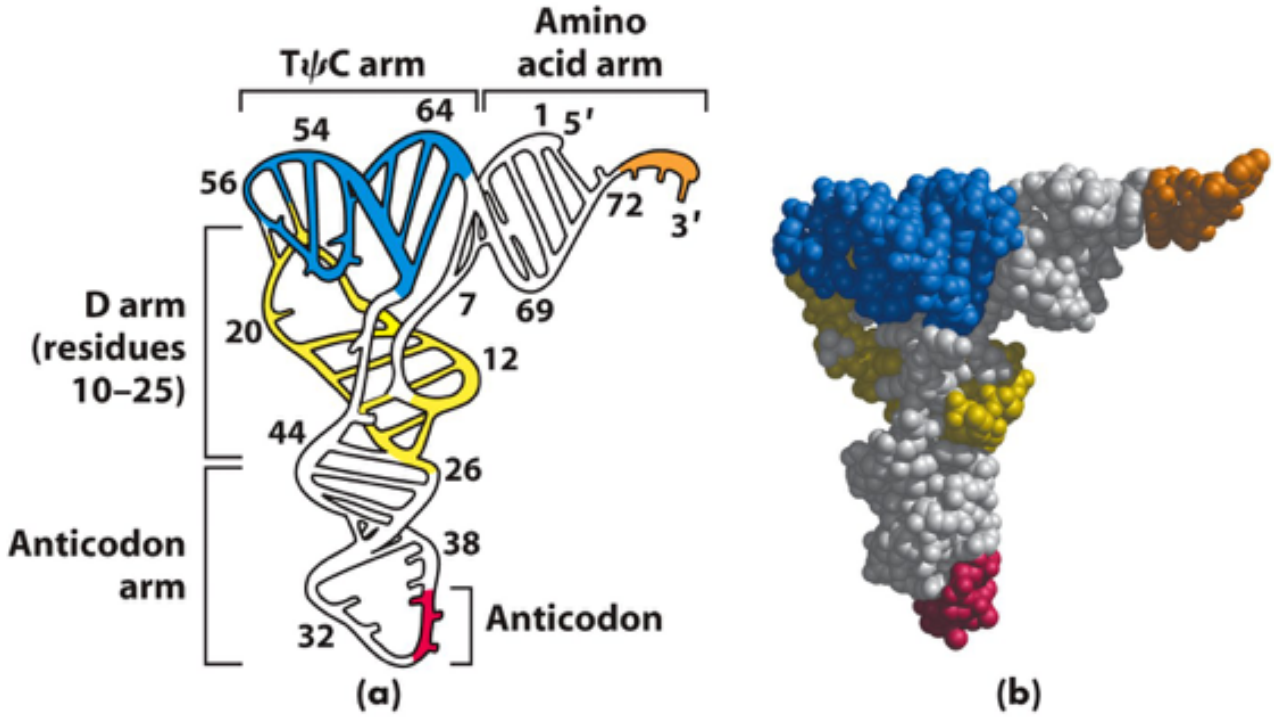
Specific **base pairing** between a **triplet codon** in the **mRNA** and three bases in **tRNA** - the **anticodon (**is specific for the codon and transfer amino acid)

Specific **amino-acid covalently linked to the 3' end** (called also anticodon because it is complementary to the codon)

Structures of different tRNAs tend to be similar

L shaped: anticodon at one end, amino acid at other

Directed by intramolecular base pairing and base stacking



The main secondary structure is conserved, even if there is variation (for specific amino acids). ACC triplets at the 3' end is conserved.
There are modified nucleosides in the tRNA sequence

- **WOBBLE BASE PAIRS**

Inosine in anticodon?
The 3rd base in an mRNA codon can undergo non-Watson-Crick base pairing with the 1st base of a tRNA anticodon
Allows for more flexible base pairing
Wobble base pairs due to multiple codons coding for a single amino acid
One tRNA molecule can recognise and bind to more than one codon
Limited number of tRNA in a cell so ensures efficient use of those present
Not necessary in many cases to have the exact tRNA

- **COMMON FEATURES OF tRNA**

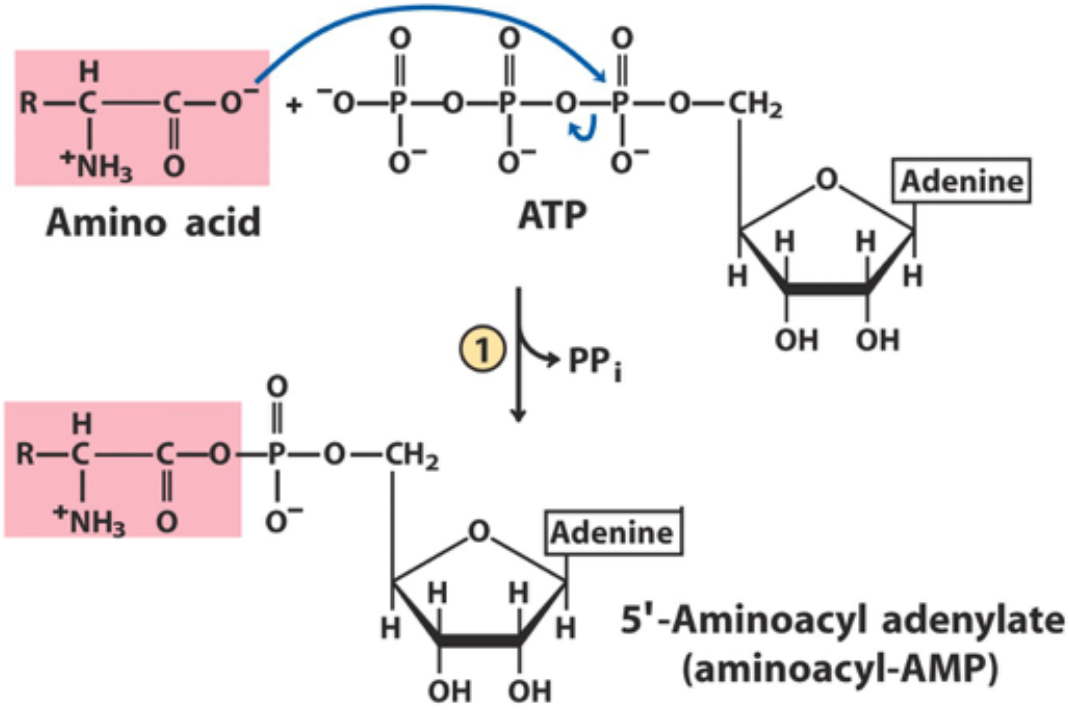
Small: 73 to 93nts (~25 kd)
Contain (~7-15) modified bases
Enzymatic modification (often methylation) of precursor
~$^1/_2$ of the bases **dsRNA**
5' end is phosphorylated, usually pG
Acitvated amino acid attached to the -OH group of the invariant 3' end : **CCA**

- **tRNA CHARGING:** covalent linkage of amino acid to tRNA

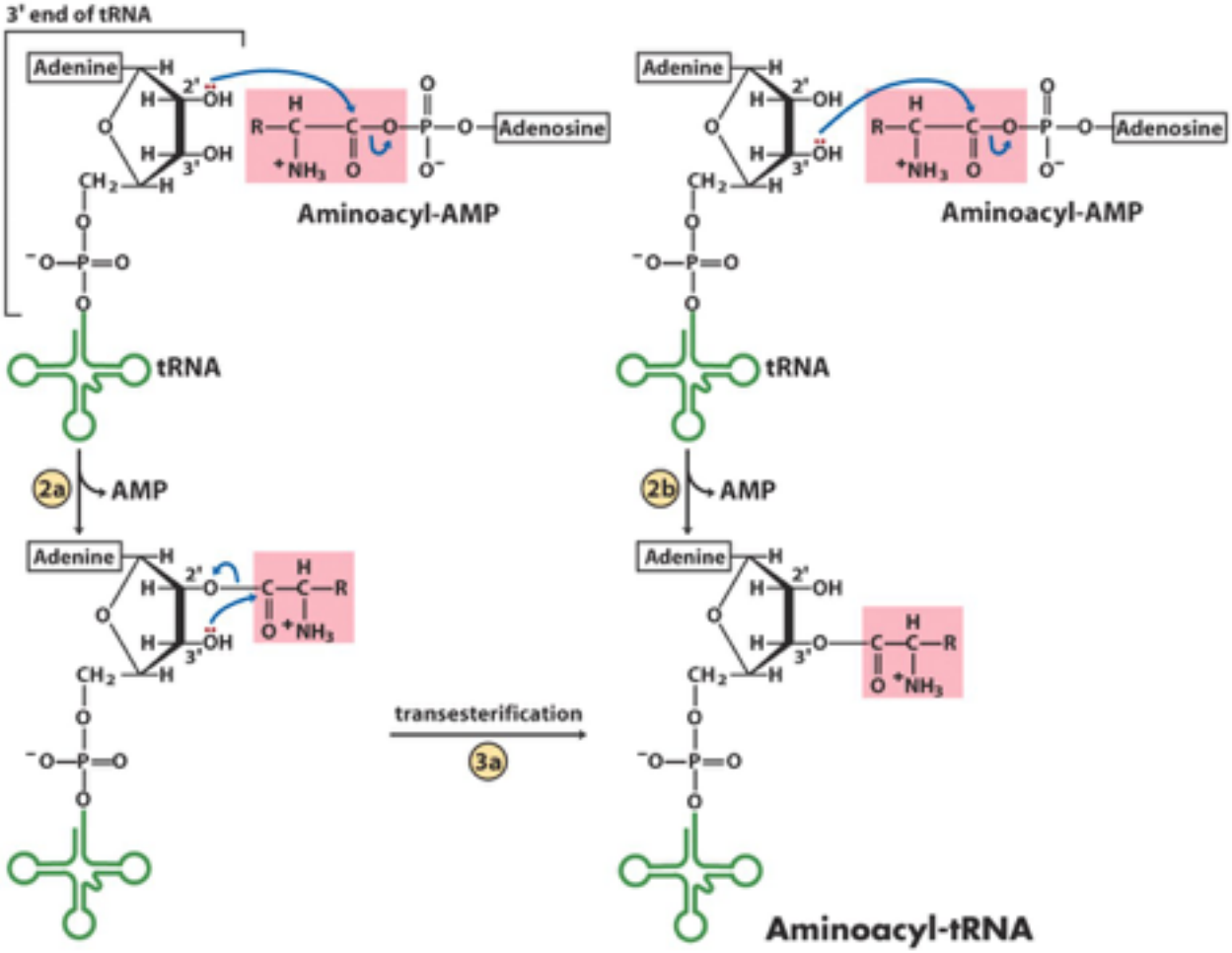
A **crucial** step in protein synthesis
Does not rely upon Watson-Crick base pairing
Specificity comes from the exquisite ability of enzymes (**aminoacyl tRNA synthetases**) to recognize even subtle differences in amino acid structure
Two, similar mechanisms
Involves formation of an AMP-amino acid **intermediate**

1. First stage is the formation of an aminoacyl adenylate,

Aminoacyl adenylate remains bound to the active site



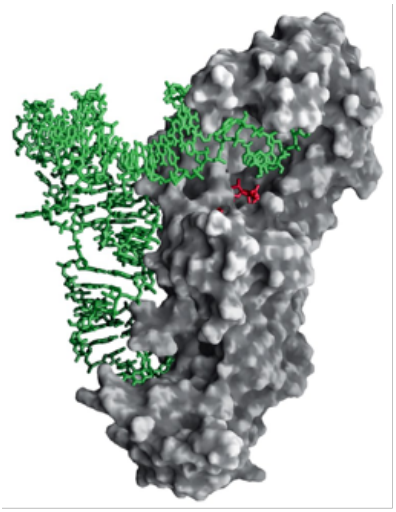
2. In the second step the aminoacyl group is transferred to the tRNA. The mechanism for this step is somewhat different for the two classes of aminoacyl-tRNA synthetases. For class I enzymes (2a), the aminoacyl group is transferred initially to the 2'-hydroxyl group of the 3'-terminal A residue, then (3a) to the 3'-hydroxyl group by a transesterification reaction. For class II enzymes (2b) the aminoacyl group is transferred directly to the 3'-hydroxyl group for the terminal adenylate.



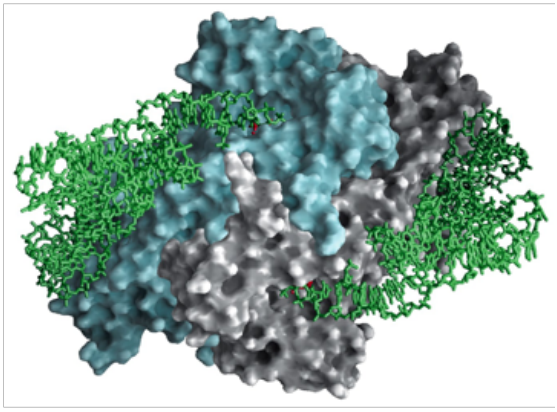
Class I enzymes:
Aminoacyl group transferred to the 2'-hydroxyl group of the 3'-terminal A residue Then to the 3'-hydroxyl group by a transesterification reaction

Class II enzymes:
Aminoacyl group transferred directly to the 3'-hydroxyl group for the terminal adenylate

The end result is exactly the same
Both bound to their cognate tRNAs and bound ATP (red) shows active site



Gln tRNA synthetase, a typical monomeric class I synthetase, from E.coli



Asp tRNA synthetase, a typical dimeric class II synthetase, from yeast

Bound ATP (red) highlights the active site near the end of the aminoacyl arm. Note the intimate contact each enzyme makes with the parts of the tRNA molecule specific to each tRNA (amino acyl arm, anti codon and variable loop or "extra arm") compared to the invariant parts.
Class I are usually monomeric or dimeric and class II are usually dimeric or tetrameric.

- Some tRNA synthetases proofread to be sure that the message transmitted is correct:

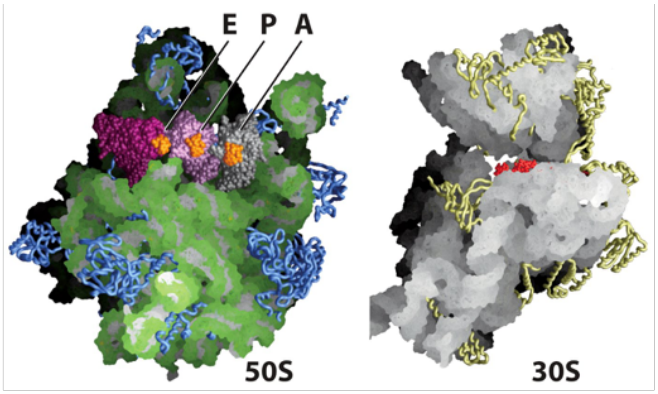
Chemically link $tRNA^{Thr}$ with Ser instead ("mischarged" tRNA)
Incubate with Threonyl tRNA synthetase
Results in rapid hydrolysis of mischarged tRNA to Ser and free tRNA
Suggests if wrong amino acid incorporated, editing function activated and amino acyl - tRNA bond hydrolysed (like the proof reading by DNA polymerase)

- **THE RIBOSOME:**

A "factory" for **translation**
About $^1/_4$ dry weight of *E.coli* - ~15,000 per cell
Enormous structures (~18nm diameter) consisting of rRNA (~65% in *E.coli*) and protein (~35%)
Sizes usually expressed in terms of **sedimentation rates**

Two subunits –

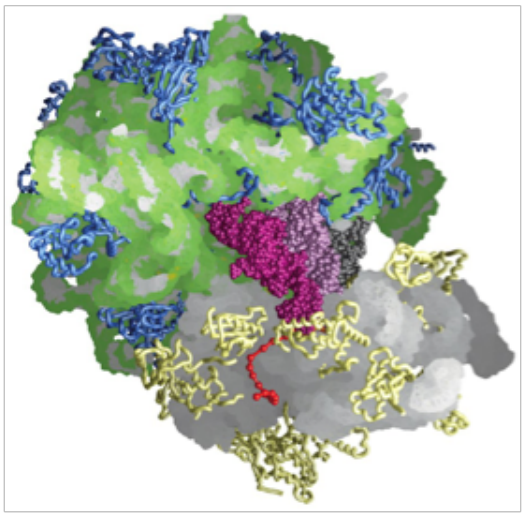
Large (50S bacteria, 60S eukaryotes)
Small (30S bacteria, 40S eukaryotes)



tRNAs(purple, mauve and grey) bound to sites E, P and A

The part of the mRNA that interacts with the tRNA anticodons is shown in red. The rest of

The 50S and 30S bacterial subunits, split apart to visualize the surfaces that interact in the active ribosome. The structure on the left is the 50S subunit with tRNAs(purple, mauve and gray); bound to sites E, P and A; the tRNA anticodons are orange. Proteins appear as blue wormlike structures; the rRNA as a blended space filling representation designed to highlight surface features, with the bases in white and the backbone in green. The structure on the right is the 30S subunit. Proteins are yellow and the rRNA is white. The part of the mRNA that interacts with the tRNA anticodons is shown in red. The rest of the mRNA winds through grooves or channels on the 30S subunit surface.



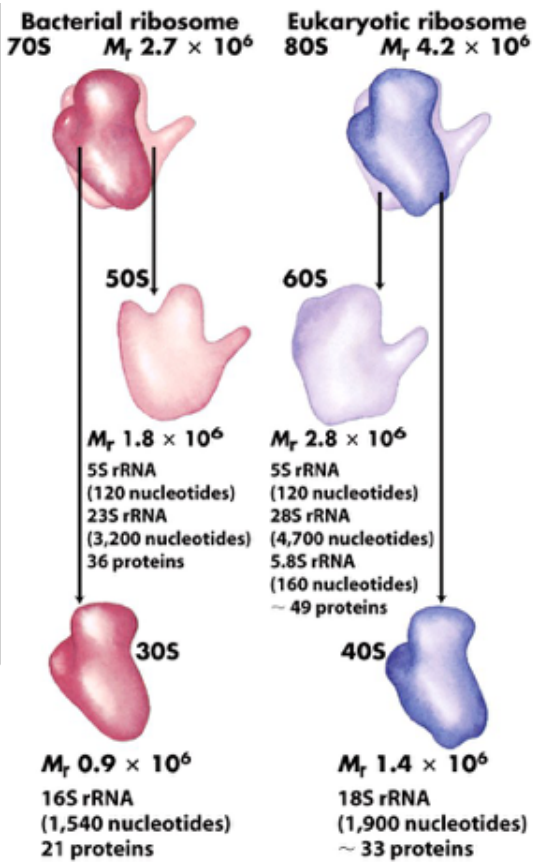


TABLE 27–6 RNA and Protein Components of the *E. coli* Ribosome

Subunit	Number of different proteins	Total number of proteins	Protein designations	Number and type of rRNAs
30S	21	21	S1-S21	1 (16S rRNA)
50S	33	36	L1-L36*	2 (5S and 23S rRNAs)

*The L1 to L36 protein designations do not correspond to 36 different proteins. The protein originally designated L7 is in fact a modified form of L12, and L8 is a complex of three other proteins. Also, L26 proved to be the same protein as S20 (and not part of the 50S subunit). This gives 33 different proteins in the large subunit. There are four copies of the L7/L12 protein, with the three extra copies bringing the total protein count to 36.

The ribosome is a ribozyme (has enzymatic properties but is not a protein)

Visualization of the ribosome highlights ability of **RNA** to form **complex 3D structures** for **chemical catalysis:**

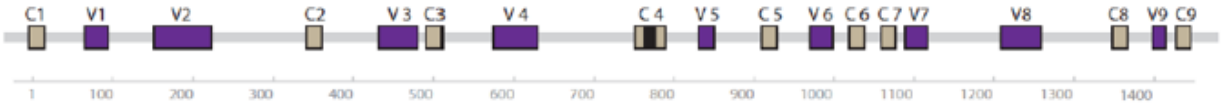
- Just like proteins - folded polypeptide chains
- Not like DNA - a constant(ish) structure

RNA containing structures like the ribosomes called **ribozymes** to distinguish from protein only **enzymes**

Argues for the idea that biochemical life began in an **"RNA world"**

TRANSLATION: THE MECHANISM

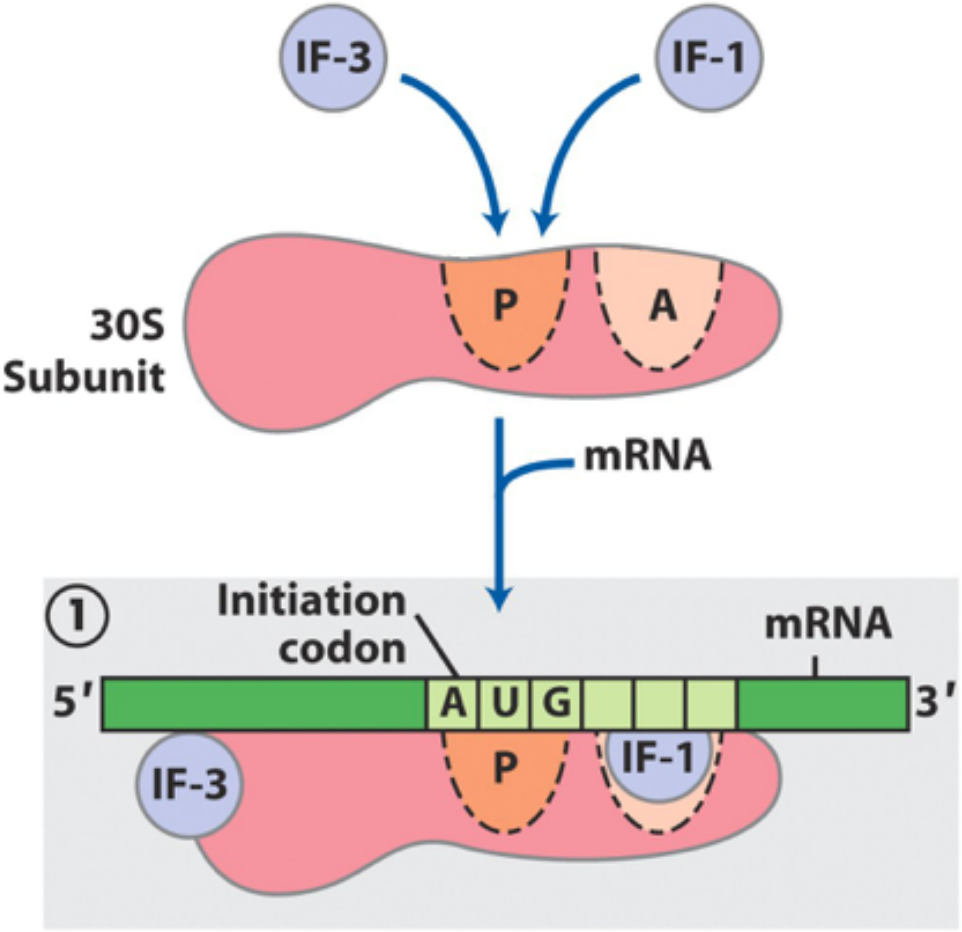
16sRNA can be used to uniquely identify bacteria
Includes interspersed conserved and variable regions, which makes it well suited for PCR amplification and sequencing
Probes are designed to hybridize to the conserved regions, which allows for amplification and sequencing of the variable regions



Schematic representation of the 16S rRNA gene. Location of variable (purple) and conserved (brown) regions in a canonical bacterial 16S rRNA. The black region is invariable in all bacteria.

- All organisms have two tRNAs for the AUG methionine codon
    - one for initiating translation
    - other for "internal" methionine residues
- In bacteria, initiator tRNA has formyl-methionine linked to it; formed enzymatically after Met-tRNA synthetase links Met to $tRNA^{fMet}$
- In eukaryotes both have methionine

- **FORMATION OF THE INITIATION COMPLEX IN BACTERIA**

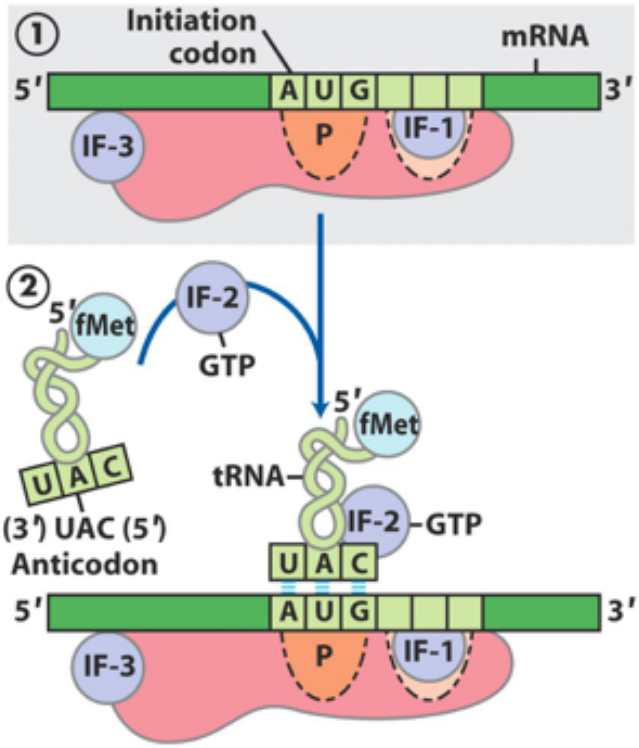


- The 30S ribosomal subunits binds two initiation factors, IF-1 and IF-3
- Factor IF-3 prevents the 30S and 50S subunits from combining prematurely
- The mRNA then binds to the 30S subunit
- The initiating AUG is guided to its correct position by the Shine-Dalgarno sequence (four to nine purine residues) in the mRNA
- Base-pairs with pyrimidine-rich sequence near 3' end 16S rRNA of the 30S subunit
- Positions the initiating AUG sequence of the mRNA in the precise position on the 30S subunit required for initiation of translation
- AUG where fMet-tRNA^{fMet} is to be bound is distinguished from other methionine codons by proximity to the Shine-Dalgarno sequence

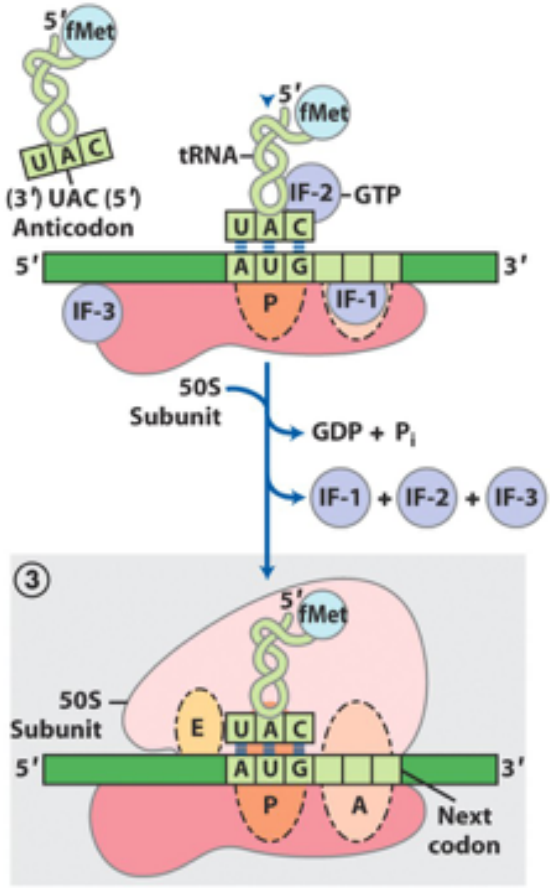
- The 30S ribosomal subunits binds two initiation factors, IF-1 and IF-3. Factor IF-3 prevents the 30S and 50S subunits from combining prematurely. The mRNA then binds to the 30S subunit. The initiating (5') AUG is guided to its correct position by the Shine-Dalgarno sequence (named for Australian researchers John Shine and Lynn Dalgarno, who identified it) in the mRNA. This consensus sequence is an initiation signal of four to nine purine residues, 8 to 13 bp to the 5' side of the initiation codon. The sequence base-pairs with a complementary pyrimidine-rich sequence near the 3' end of the 16S rRNA of the 30S ribosomal subunit. This mRNA-rRNA interaction positions the initiating (5') AUG sequence of the mRNA in the precise position on the 30S subunit where it is required for initiation of translation. The particular (5') AUG where fMet-tRNA^{fMet} is to be bound is distinguished from other methionine codons by its proximity to the Shine-Dalgarno sequence in the mRNA.
  Bacterial ribosomes have three sites that bind aminoacyl tRNAs, the aminoacyl (A) site, the peptidyl (P) site and the exit (E) site. Both the 30S and the 50S subunits contribute to the characteristics of the A and P sites, whereas the E site is largely confined to the 50S subunit. The initiating (5')AUG is positioned at the P site, the only site to which fMet-tRNA^{fMet} can bind. The fMet-tRNA^{fMet} is the only aminoacyl-tRNA that binds first to the P site; during the subsequent elongation stage, all other incoming aminoacyl-tRNAs bind first to the A site and only subsequently to the P and E sites. The E site is the site from which the "uncharged" tRNAs leave during elongation. Factor IF-1 binds at the A site and prevents tRNA binding at this site during initiation.

- Bacterial ribosomes have three sites that bind aminoacyl tRNAs, the aminoacyl (A) site, the peptidyl (P) site and the exit (E) site
- Both the 30S and the 50S subunits contribute to the characteristics of the A and P sites, whereas the E site is largely confined to the 50S subunit
- Initiating AUG is positioned at the P site, only site to which fMet-tRNA^{fMet} can bind
- The fMet-tRNA^{fMet} is the only aminoacyl-tRNA that binds first to the P site
- During subsequent elongation stage, all other incoming aminoacyl-tRNAs bind first to the A site and only subsequently to the P and E sites
- The E site is the site from which the "uncharged" tRNAs leave during elongation
- Factor IF-1 binds at the A site and prevents tRNA binding at this site during initiation

- The complex consisting of the 30S ribosomal subunit, IF-3, and mRNA is joined by both GTP-bound IF-2 and the initiating fMet-tRNA^{fMet}
  The anticodon of this tRNA now pairs correctly with the mRNA's initiation codon

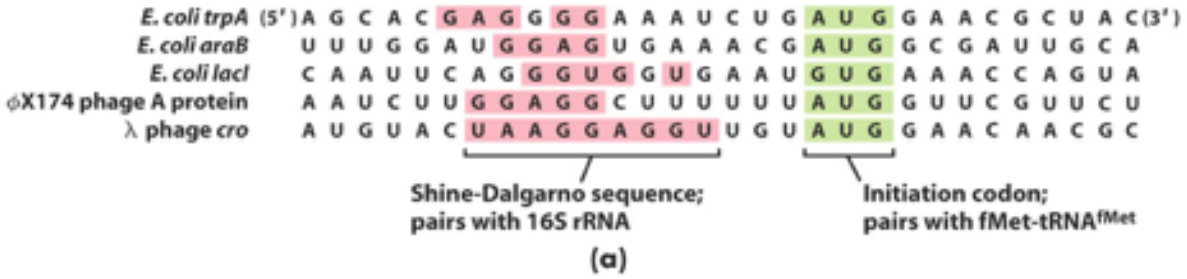


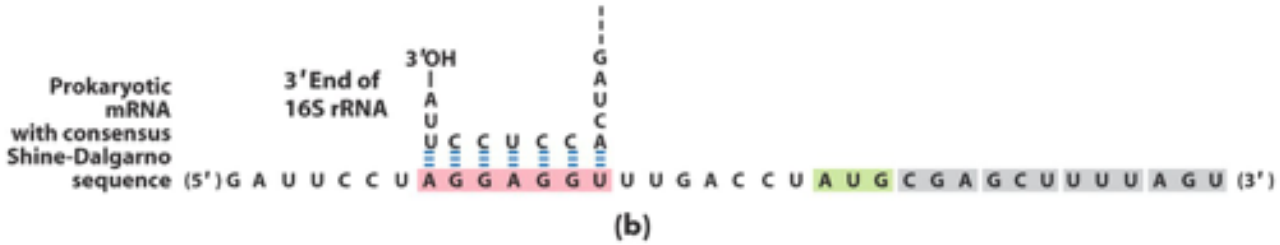
- This large complex combines with the 50S ribosomal subunit; simultaneously, the GTP bound to IF-2 is hydrolyzed to GDP and P_i, which are released from the complex. All three initiation factors depart from the ribosome at this point.
  Correct binding of the fMet-tRNA^{fMet} to the P site in the 70S initiation complex is assured by at least three points of recognition and attachment: the codon-anticodon interaction involving the initiation AUG fixed in the P site; interaction between the Shine-Dalgarno sequence in the mnRNA and the 16S rRNA; and the binding interactions between the ribosomal P site and the fMet-tRNA^{fMet}.



- Correct binding of the fMet-tRNA^{fMet} to the P site in the 70S initiation complex is assured by at least three points of recognition and attachment:
    - The codon-anticodon interaction involving the initiation AUG fixed in the P site
    - Interaction between the Shine-Dalgarno sequence in the mRNA and the 16S rRNA
    - Binding interactions between the ribosomal P site and the fMet-tRNA^{fMet}.

- IF-1 and IF-3 bind the Small 30S subunit
  IF-3 prevents premature assembly with 50S
  mRNA binds to the 30S subunit using its Shine-Dalgarno sequence
  base pairs with 16S rRNA in 30S subunit
  fMet-tRNA^{fMet} brought into the P (peptidyl) site with IF-2GTP
  Combines with 50S subunit after GTP hydrolysis and departure of initiating factors



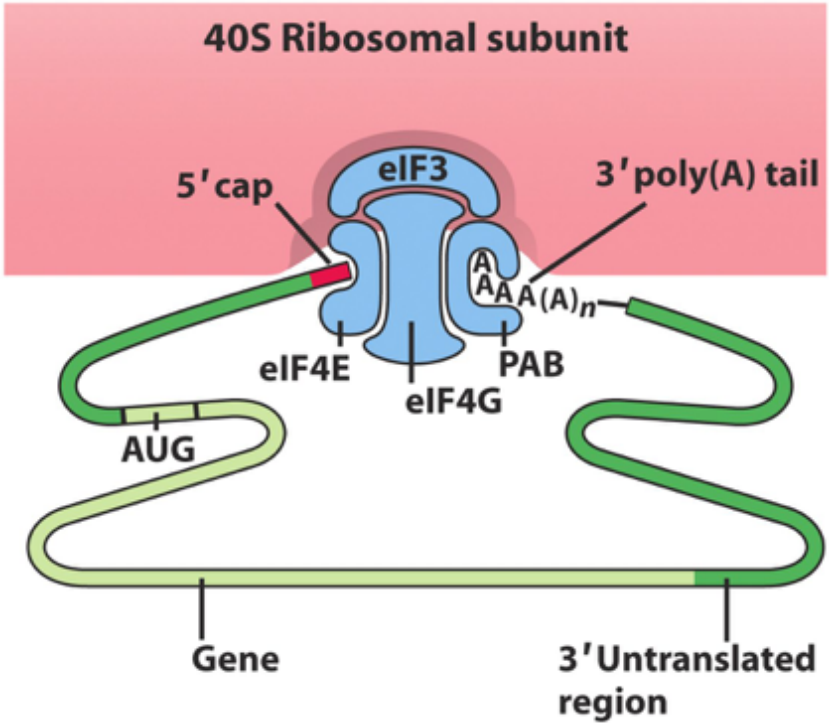


- **STARTING AT THE RIGHT PLACE**

Correct incorporation of fMet-tRNA^{fMet} from:
interaction of mRNA and 16S rRNA
interaction of fMet-tRNA^{fMet} with AUG codon
interaction of fMet-tRNA^{fMet} with P site

- **EUKARYOTIC INITIATION**

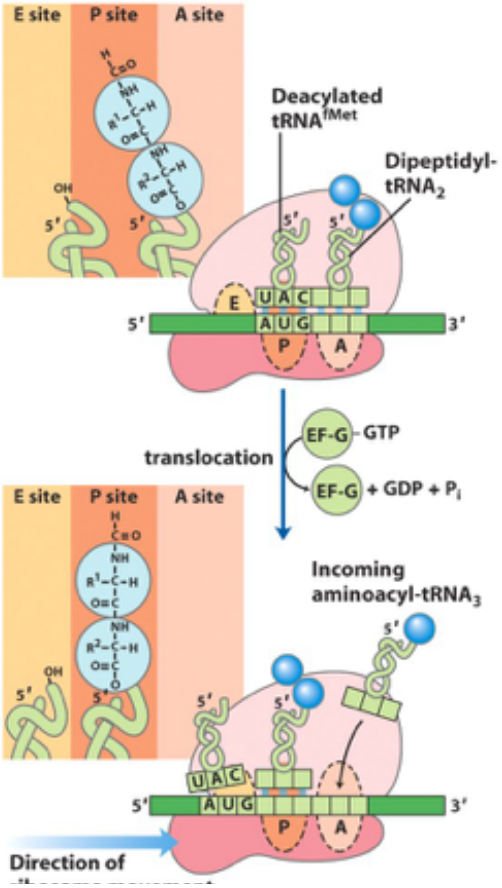
Similar mechanism but more protein **initiation factors** required
No Shine-Dalgarno sequence
Positioning relies upon interaction with both **ends** - the 5’ **cap** and the 3’ **polyA** tail
Eukaryotic mRNAs are usually monocistronic: no need for internal initiation
Ribosomes ‘scan’ the mRNA in a 5’-3’ direction till they encounter the first AUG codon



3 initiation factors, bind polyA mRNA --> there is only one start methionine there, it does not have to worry about scanning the sequence to find the Shine-Algarno sequence.

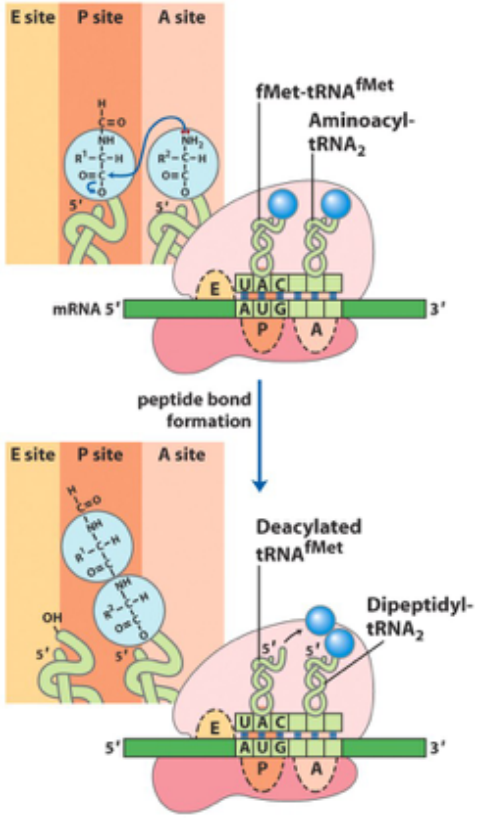
- **ELONGATION AND THE PEPTIDE BOND:**

  - Requires factors **EF-Tu**, **EF-Ts** and **EF-G**
    Aminoacyl-tRNA brought in to A site by **EF-TuGTP**
    GTP **hydrolyzed** and EF-TuGDP release
    EF-TuGTP **regenerated** using **EF-Ts**



- **Peptide bond formation** catalyzed by **23SrRNA**
  Nucleophilic attack by α-amino group of amino acid in A site on carbonyl group of peptide in P site

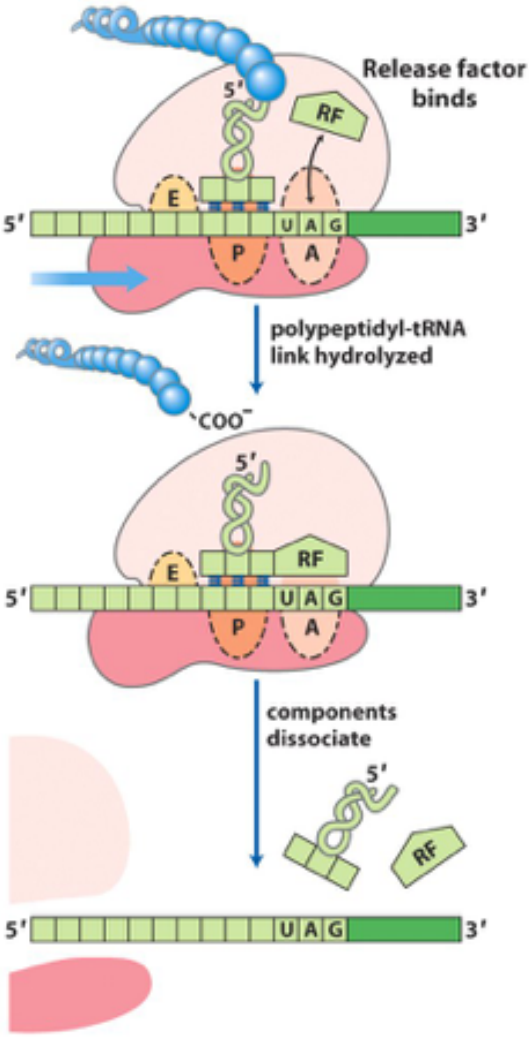
- **Translocation** requiring **EF-G** and **GTP hydrolysis**
  moves ribosome **one codon along** mRNA (**5' to 3'**)
  moves newly synthesized peptidyl-tRNA into P site
  moves "uncharged" tRNA into E site



Elongation mechanism in eukaryotes is basically the same, there are analogous elongation factors; however is not interchangeable (take trascription factors from bacteria and elongate translation in eukaryotes)

- **TERMINATION AND PEPTIDE RELEASE:**

No tRNA for recognizing **UAA**, **UGA** or **UAG**
Release factors (bind to stop codon):
RF1 recognizes **UAA** and **UAG**
RF2 recognizes **UAA** and **UGA**
Encourage ribosome to transfer peptide to **water** rather than an aminoacyl tRNA
Eukaryotes have a **single** factor **eRF**



- **INHIBITION OF PROTEIN SYNTHESIS**

Number of naturally occurring toxins and antibiotics inhibit translation implying structural differences between the eukaryotic and prokaryotic ribosome

- **Puromycin**
  Similar to 3’ end of a tRNA, binds A site and causes premature termination
- **Tetracyclines**
  Block A site
- **Chloramphenicol**
  Blocks peptidyltransferase
- **Cycloheximide**
  Blocks eukaryotic peptidyltransferase
- **Streptomycin**
  Causes misreading of codons

GENE CLONING 1 : INTRODUCTION

- Early 19th century (pre-genetics era) → Traits are inherited as "particles", offspring receive a "particle" from each parent **(Blending inheritance)**
  - a pink flower that results from the mating of a red flower with a white one

- 1866 Mendel → rules to explain inheritance of biological characteristics. Each heritable property is controlled by a **gene** **(Mendelian inheritance)**

- 1869 Friedrich Miescher – identified **chromosomes** as the carriers of genetic material
  Boveri–Sutton chromosome theory states simply that "chromosomes, which are seen in all dividing cells & pass from one generation to the next, are the basis for all genetic inheritance"

- 1922 Thomas Hunt Morgan - analysis of relative position of >2000 genes on 4 chromosomes of *Drosophila* **gene mapping**

- 1903 Walter Sutton – isolated and identified phosphate-rich substances called nuclein (now known as nucleic acids)

- 1885-1901 Albrecht Kossel - isolated the five organic compounds in nucleic acid: adenine, cytosine, guanine, thymine, and uracil.

- 1910 Phoebus Levene - formulated tetranucleotide hypothesis which first proposed that DNA was made up of equal amounts of adenine, guanine, cytosine, and thymine. Identified the difference between DNA and RNA, i.e. deoxyribose and ribose. Found that DNA contained bases, deoxyribose, and a phosphate group.

- 1944 Avery, MacLeod & McCarty & 1952 Hershey & Chase – reported that DNA is the substance that causes bacterial transformation i.e. genetic material is deoxyribonucleic acid **DNA**

- 1950 Maurice Wilkins & Raymond Gosling - X-ray diffraction images of DNA

- 1952 Rosalind Franklin & Raymond Gosling - Photo 51 "The results suggest a helical structure (which must be very closely packed) containing probably 2, 3, or 4 coaxial nucleic acid chains per helical unit and having the phosphate groups near the outside"

- 1953 Arthur Kornberg - isolated **DNA polymerase I**

- 1958 Rosalind Franklin & Aaron Klug – revealed the structure of tobacco mosaic virus

- 1970 Hamilton Smith - isolated HindII **(the first type II restriction enzyme)**

- 1973 Berg, Boyer & Cohen – inserted DNA into a **plasmid** & made the first **recombinant DNA molecules**

- 1977 Frederick Sanger – dideoxy **DNA sequencing**

- 1984 Kary Mullis – developed **polymerase chain reaction (PCR)**

- 1986 Randall Saiki – used Thermophilus aquaticus (Taq) DNA polymerase for PCR

- 1996 Roslin Institute cloned the sheep 'Dolly' **reproductive cloning**
- 2003 **"Complete" human genome sequence**

GENE CLONING:

- When DNA is extracted from an organism all its genes are obtained
- In gene (DNA) cloning a particular gene is copied (cloned)
- Cloning a gene means isolating an exact copy of a single gene from the entire genome of an organism
- Involves copying the DNA sequence of that gene into a smaller, more accessible piece of DNA e.g. plasmid

WHY?

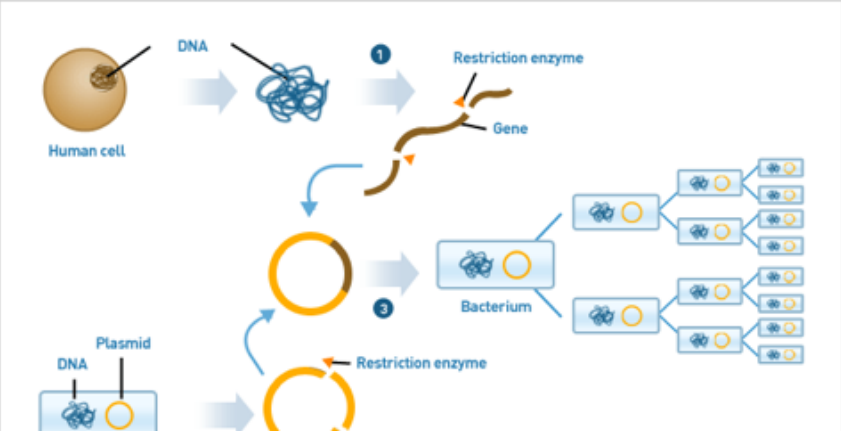
- To obtain a pure sample of an individual gene separated from all other genes in the cell
- Determine the nucleotide sequence of specific genes
- So the specific DNA can be amplified
- So the specific protein can be expressed
- So the protein's function can be investigated
- Organisms can be 'engineered' for specific purposes
  e.g. insulin production, insect resistance, lab research etc.

OVERVIEW OF GENE CLONING

- DNA is purified from a cell. Fragment of the DNA that contains a gene of interest is isolated using a restriction enzyme or PCR

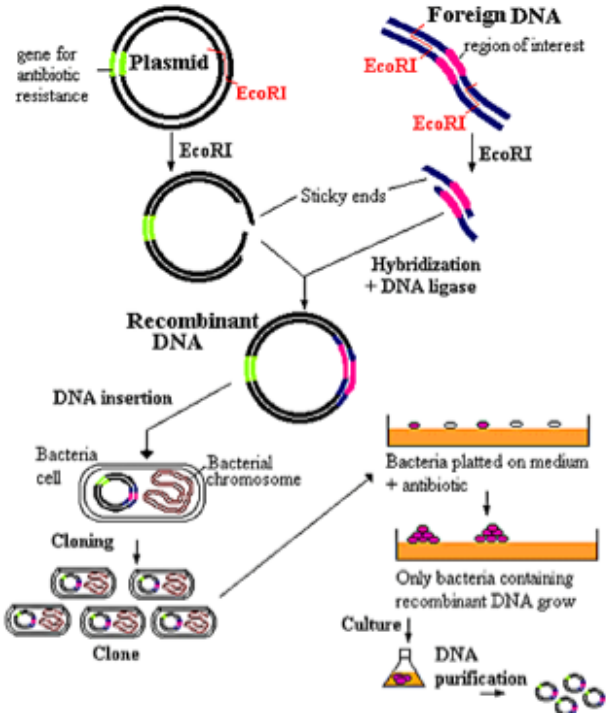
- The DNA fragment is inserted into a circular DNA molecule, vector, in this case a plasmid (from bacterium) to produce a recombinant DNA molecule

3. Transform host cells with the vector which then replicates producing numerous identical copies of itself & the gene that it carries
   When host cell divides copies of recombinant DNA are passed to progeny & further vector replication takes place
   After large number of cell divisions a colony or clone of identical host cell is produced. Identify clone.

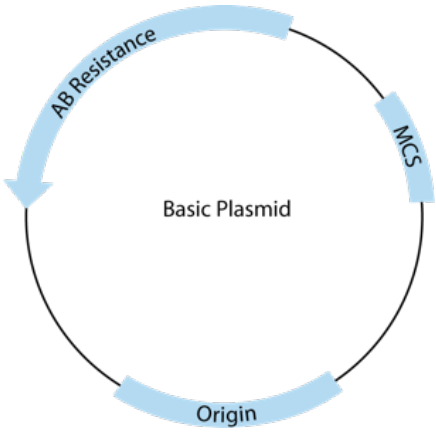


GENE CLONING 2

- Foreign DNA is digested with a restriction enzyme
- Bacterial plasmids (small circular DNA additional to regular DNA of bacteria) are cut with the same restriction enzyme
- A fragment of DNA can thus be inserted into the plasmid DNA using DNA ligase to form a "recombinant DNA molecule"
- Incorporate plasmids into bacteria host cells by transformation
- Each cell contains a different segment of DNA from the original organism – DNA library
- Cells can now be plated out on agar medium. Colonies of cells containing the desired gene can then be identified & isolated



- **Isolate DNA** from organism/cell that contains gene of interest
    - Lyse cell by chemical & physical methods e.g. sonication, homogenization
    - Remove membrane lipids with detergent
    - Remove proteins by adding a protease
    - Remove RNA with RNase
    - Precipitate DNA with alcohol
- DNA is fragmented with restriction enzymes (endonucleases) e.g. EcoRI, HindIII etc. cut the gene into small pieces

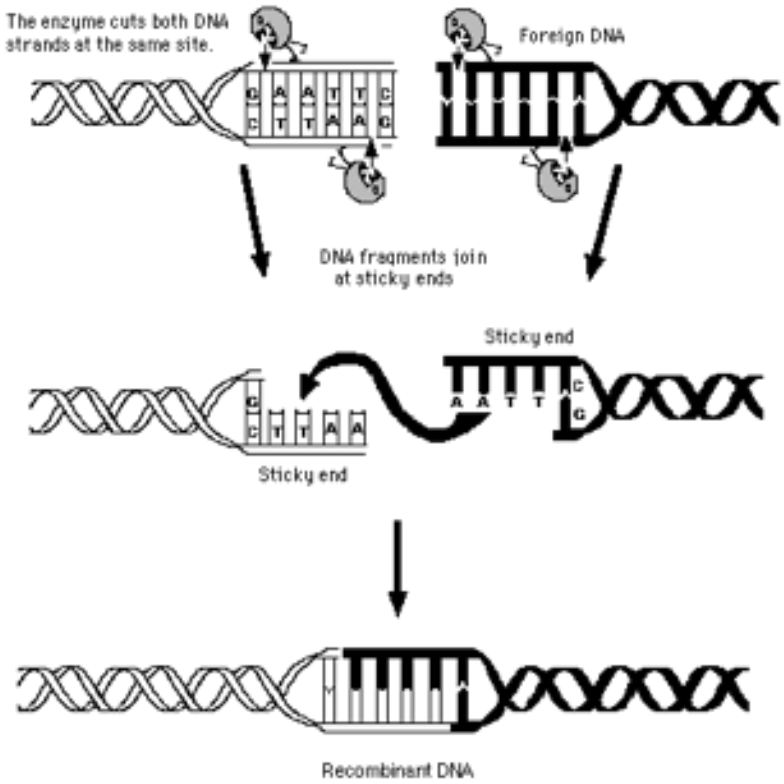


- Plasmid DNA consists of
    - Origin of replication
    - Antibiotics resistance gene
    - Multiple cloning site (MCS)

- Different DNA pieces cut with the same restriction enzyme can join or recombine
- Restriction enzymes creates staggered cuts in specific sequences → cohesive "sticky" or blunt ends
- Sticky ends anneal (hybridize)
- DNA ligation

- DNA ligase requires ATP
- Joins DNA strands together by catalysing the formation of a phosphodiester bond

Restriction Enzyme Action of EcoRI



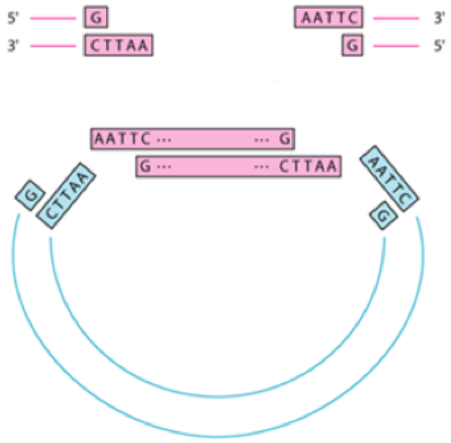
DNA CLONING STRATEGIES

- **Single restriction enzyme**
    - Non-directional (50% chance)
    - Self-ligation of the vector occur
        - the vector needs to be dephosphorylated to minimise self-ligation

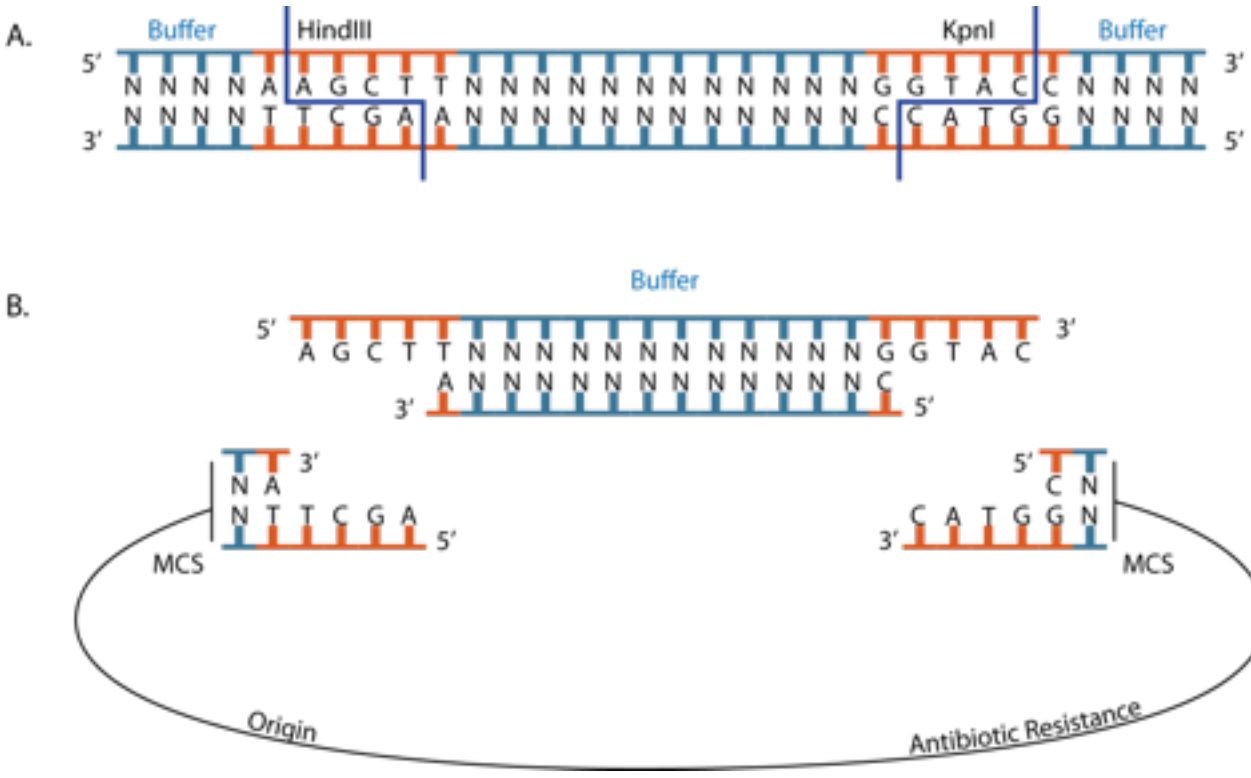
        Alkaline phosphatase

        Shrimp alkaline phosphatase (SAP)

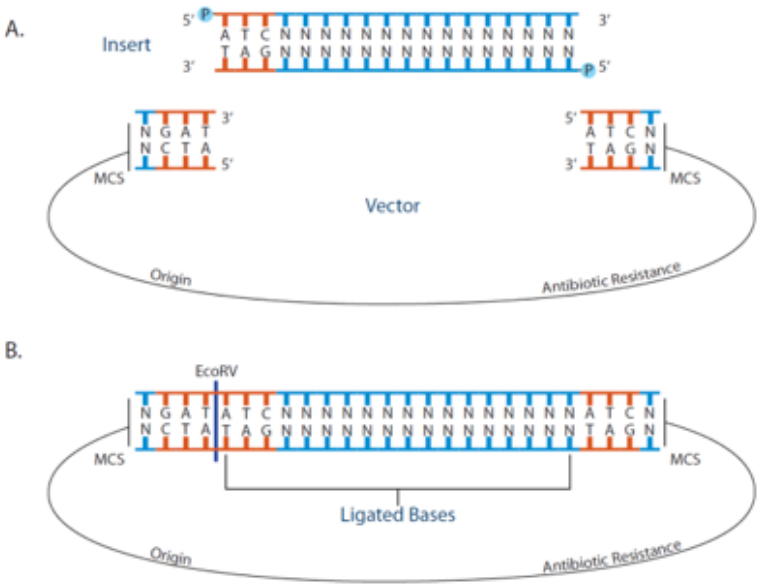
        Calf-intestinal alkaline phosphatase (CIP)



- **Two restriction enzymes: (directional)**



- **Blunt end ligation:** Blunt end ligation occurs 100X slower than sticky-end ligation, it's not directional



- DNA is fragmented with **restriction enzymes (endonucleases)** e.g. EcoRI, HindIII etc. cut the gene into small pieces

  Frequency of cutting: 6 base cutter (¼)^6 = 1/4096

	Genome size	No. of fragments (6 base cutter)
Drosophila	1.8 x 10^8	45,000
Humans	3.2 x 10^9	800,000

- **Extraction of DNA fragments from agarose gel:**
  - Cutting out target bands from gels
  - Melting agarose gels at around 50°C
    - Low melting point agarose
    - Lowering of melting point by NaI
  - Purification of the DNA fragments
    - Glass beads (silica)

- **PCR BASED CLONING:** directional

  PCR requires some sequence information about 2 regions of DNA of interest to synthesize the appropriate **primers**
  Primers are oligonucleotides **complementary** to different regions on the 2 strands of DNA template (**flanking** the region to be amplified)
  Primers ~15-20 nucleotides designed to be ~200-2000 bp apart
  One hybridizing to one strand of dsDNA, the other hybridizing to the other strand such that both primers are oriented with their 3' ends pointing towards each other

Primer acts as a starting point for DNA synthesis
The oligo is **extended** from its 3' end by **DNA polymerase**

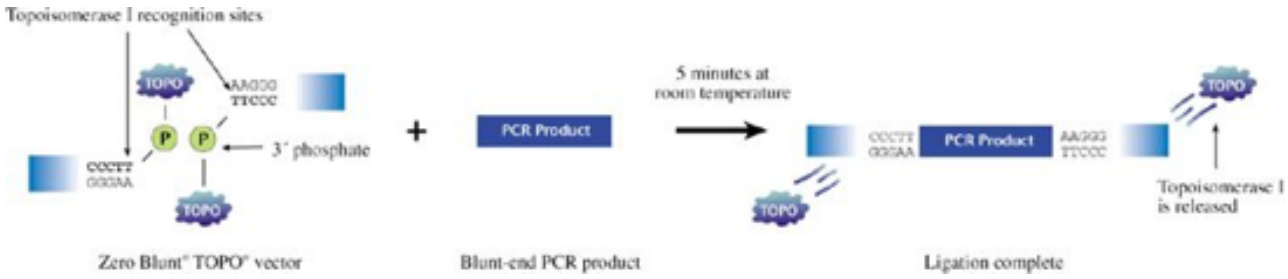
- **TA CLONING**

  High fidelity DNA polymerases (such as Pfu DNA polymerase) possesses 3' to 5' exonuclease proofreading activity to correct nucleotide-misincorporation errors, thus no adenine is added to the 3' end
  DNA fragments amplified by High fidelity DNA polymerases can still be used for TA cloning by adding A overhangs to the blunt PCR product with Taq DNA Polymerase and dATP
  Taq DNA polymerase adds an adenine to the 3' end of the product

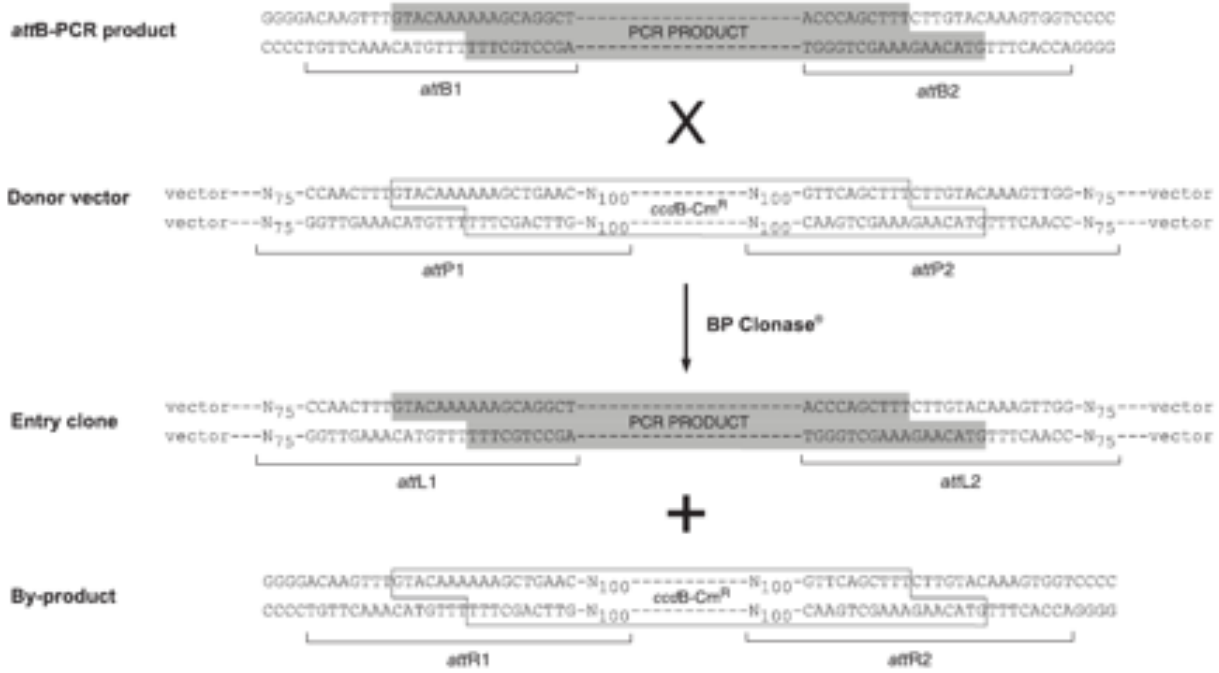
- **DNA LIGATION**

  - DNA ligase: T4 DNA ligase (from bacteriophage T4)
  - DNA topoisomerase I
  - DNA recombinase

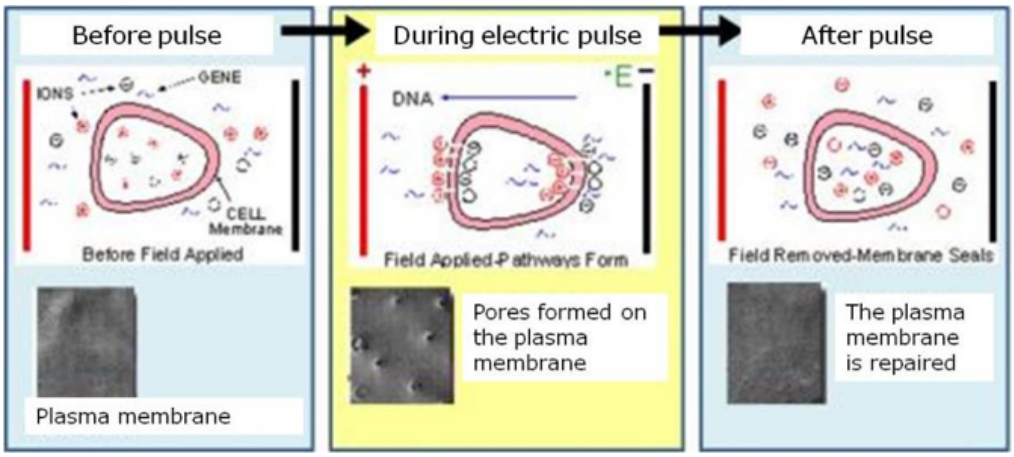
- **TOPO cloning:**
  Vaccinia virus topoisomerase I specifically recognises and digests DNA sequence (C/T)CCTT, and unwinds the DNA and re-ligates it again at the 3' phosphate group of the last thymidine.
  TOPO vectors carry (C/T)CCTT at the two linear ends. The linear vector DNA already has the topoisomerase enzyme covalently attached to both of its strands' free 3' ends.
  PCR is performed to amplify target DNA fragment.
  Once the PCR products and TOPO vectors are mixed, the topoisomerase catalyse ligation of the two ends at r.t. in 5 min.



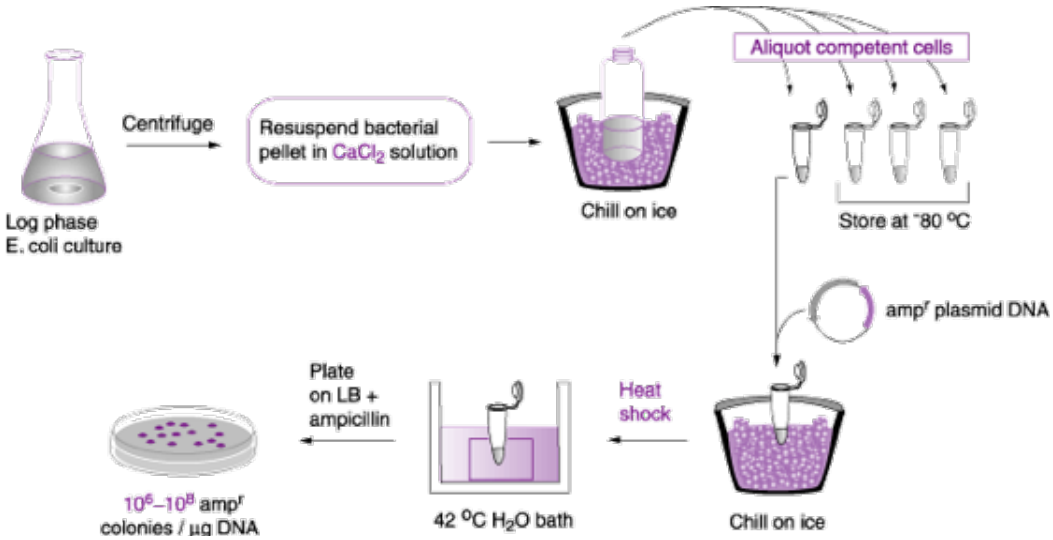
- **DNA RECOMBINASE**
  Cloning by homologous recombination:
  PCR is performed with primers containing overlap sequences with vectors.
  DNA recombinases recognise overlapped sequences between vectors and PCR products, and catalyse the recombination and insertion of the PCR product into the vector.



- **Bacterial transformation (electroporation)**
- DNA can be introduced into the bacterial cells through the pores created by an electrical field



- **$CaCl_2$/Heat Shock method**
- The cells become competent when incubated with $CaCl_2$ in cold condition, due to changes of the cell surface structure thus making it more permeable to DNA
- The heat-pulse creates a thermal imbalance on either side of the cell membrane, which forces the DNA to enter the cells through pores

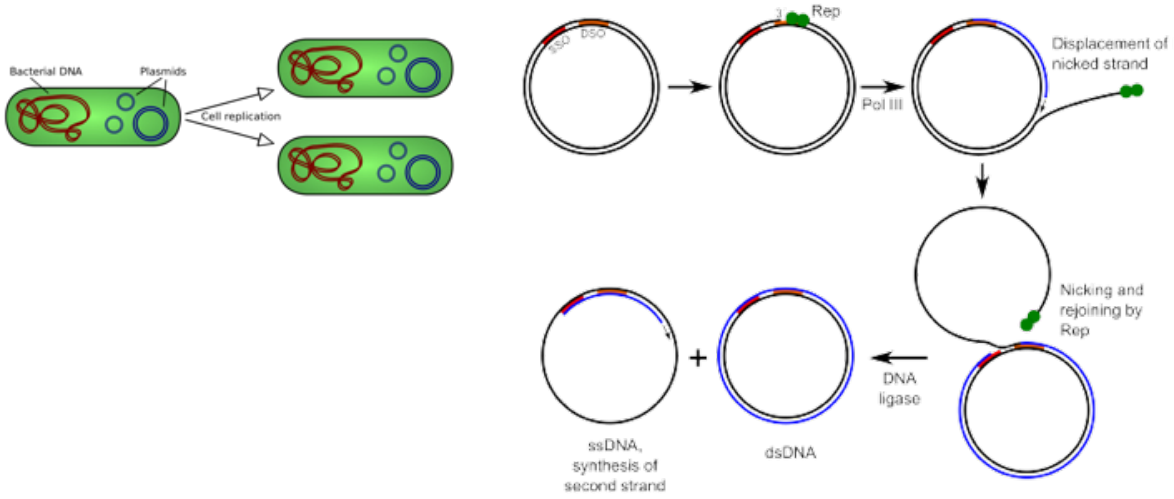


Strategies for introducing new DNA into animal & plant cells:

- precipitation of DNA onto cell surface with Ca phosphate
- introduction by liposomes
- transformation of plant protoplasts i.e. plant cell after wall has been degraded
- microinjection i.e. inject DNA into nucleus
- biolistics i.e. transformation with micro projectiles (gene gun)
- electroporation

- **PLASMID REPLICATION**

- Plasmid DNA can remain in circular shape in bacteria
- Plasmid DNA replicates in rolling circle way
- Thus linear plasmid DNA cannot be replicated



Antibiotics resistance genes

- Code proteins exhibit resistance to antibiotics
- Act as selectable markers to identify bacteria with a particular plasmid
  i.e. in presence of ampicillin only cells expressing the protein for antibiotic resistance can grow
- Most popular antibiotic selections: Ampicillin & Kanamycin

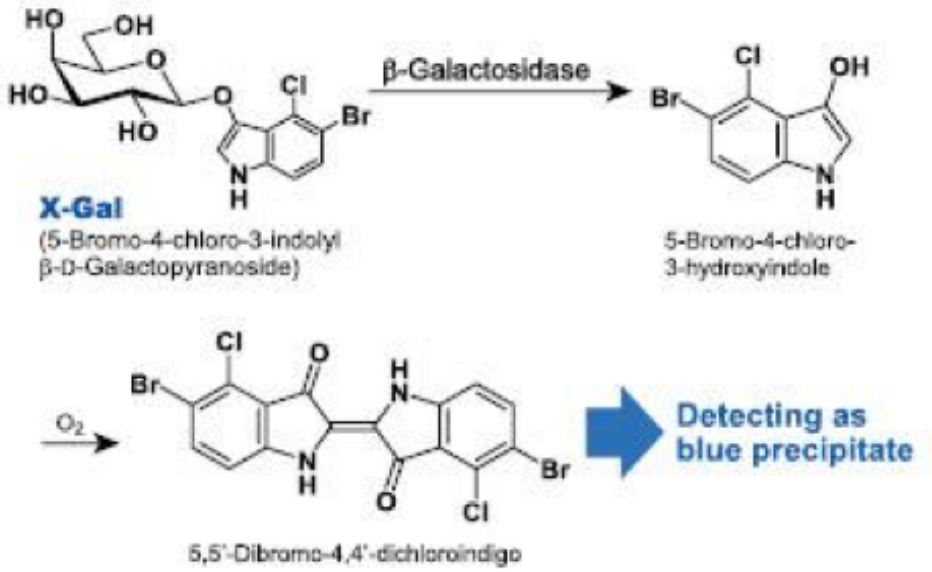
- Ampicillin is an irreversible inhibitor of the enzyme transpeptidase, which is needed by bacteria to make their cell walls.
  Ampicillin causes cell lysis by inhibiting bacterial cell wall synthesis.
  β-lactamase provides antibiotic resistance by breaking the β-lactam antibiotics' structure such as ampicillin.
  β-lactamase gene can often be called ampicillin resistance gene, or simply amp^R

- Kanamycin interacts with the 30S subunit of prokaryotic ribosomes.
  Kanamycin induces substantial amounts of mistranslation and indirectly inhibits translocation during protein synthesis, thus causing cell death.
  Aminoglycoside 3'-phosphotransferase (also known as Neomycin phosphotransferase II) enzyme, which inactivates by phosphorylation a range of aminoglycoside antibiotics such as kanamycin.

Aminoglycoside 3'-phosphotransferase gene can often be called kanamycin resistance gene, or simply kan^R

- After transformation cells are plated onto ampicillin medium → colonies
- All colonies are transformants (harbours a plasmid with amp^R).
- Untransformed cells are ampicillin-sensitive → no colonies

- A plasmid vector that contains ***lacZ*** gene which codes for part of the enzyme **b-galactosidase**
- Some strains of *E. coli* have a **modified *lacZ*** gene & only synthesise the enzyme when a plasmid which harbours the **missing *lacZ* segment is present**



X-Gal hydrolysis with β-galactosidase.

- A plasmid vector that contains ***lacZ*** gene which codes for part of the enzyme **b-galactosidase**
- Some strains of *E. coli* have a **modified *lacZ*** gene & only synthesise the enzyme when a plasmid which harbours the **missing *lacZ* segment is present**
- Multiple cloning site is in the middle of LacZ gene
- Insertion of external DNA fragment into MCS causes disruption of LacZ gene, thus no functional b-galactosidase expressed, resulting white colony – blue/white selection

- **Cloning vector:** a foreign DNA fragment (usually <10 kb) can be inserted into a vector
- Must replicate in host cell → many copies of recombinant DNA & passed to daughter cells
- **Plasmids**: replication origin, MCS, antibiotic resistance
- Bacteriophages (e.g. l phage) are used for cloning larger DNA fragments
- For much larger DNA fragments, bacterial artificial chromosomes (BACS) & yeast artificial chromosomes (YACs) are used

GENE CLONING 3

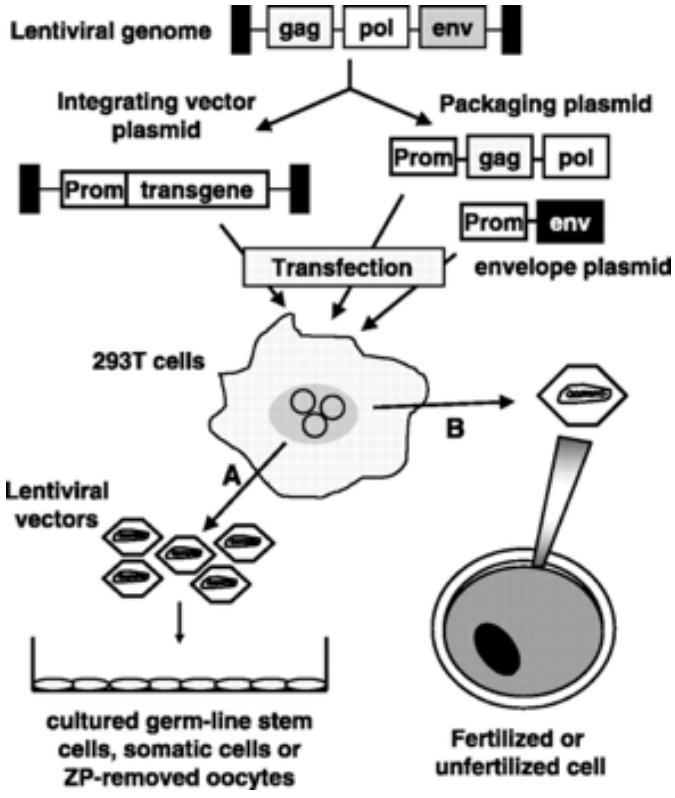
- **Bacteriophages (λ phage) as vector:**

    - Genes encoding head & tail proteins & other genes
    - involved in host cell lysis are clustered in distinct regions
    - in ~ 50 kb l phage genome
    - Genes irrelevant for survival/growth can be **deleted**
    - from the phage genome & replaced by other DNA
    - sequences of interest
    - Insert size is **limited to ~ 25 kb** due to the requirement
    - that the DNA has to fit into the phage head

- **Viruses as cloning vectors**

    - **Adenoviruses:** will take DNA fragments up to 8kb. Induce inflammatory genes e.g. IFNg in transduced cells
    - **Papillomaviruses:** High capacity for inserted DNA. Does not cause death of infected mouse cells & BPV molecules are passed to daughter cells → permanently transformed cell line
    - **Adeno-associated viruses*:** Inserts into host DNA at same position within human chromosome 19
    - **Retroviruses (e.g. Lentivirus)*:** commonly used in gene therapy. Inserts at random positions → very stable integrants
      *potentially useful in gene therapy

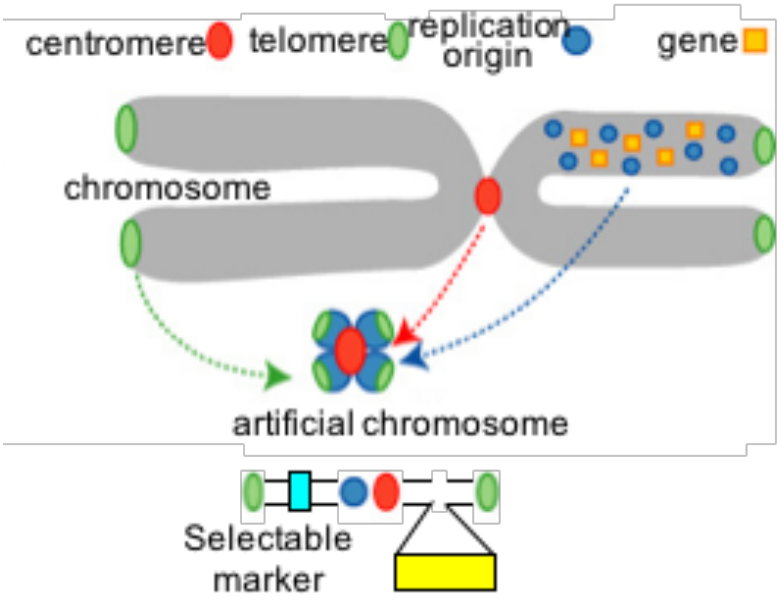
- The wild-type lentiviral genome is made up of packaging genes (gag, pol), envelope gene (env), and long-terminal repeats (shown in black)
- The genome can be separated into 3 plasmids, packaging, envelope, and integrating vector for transgene
- The 3 plasmids are transfected into a cell line to produce the viral vectors, which are collected in the supernatant



- **YAC (yeast artificial chromosome):**

    - **YAC libraries** useful for cloning very large DNA fragments (> **1 Mb**) cloning large genes (e.g. 250kb cystic fibrosis gene) & creating libraries of large overlapping clones e.g. for individual chromosomes (chromosomal libraries)
    - **YAC**s are **hybrids** of bacterial plasmid DNA & yeast DNA. Components required for replication/segregation of natural yeast chromosomes combined with *E. coli* plasmid DNA
    - **YAC**s are grown in yeast *Saccharomomyces cerevesiae* & contain **selectable markers.** Yeast selectable markers allow growth of transformant on selective media lacking specific nutrients (non-transformants unable to grow)
    - Yeast strains used are **auxotrophic** i.e. unable to make a specific compound e.g. Trp1 mutants make no tryptophan only grow on media supplemented with tryptophan. If mutant strain is transformed with **YAC** containing intact Trp1 gene → compensation for inactive gene & transfected cell grows on media lacking tryptophan

    - **YAC**s are unstable and frequently lose parts of the DNA during propagation



- **DNA LIBRARY:**

Genomic library

- A collection of the total genomic DNA from a single organism
- Inserted into a vector
- Used for sequencing and cloning applications

Particular gene is only a small part of the genome e.g. if entire genome is 10^6 kbp thus 10 kbp gene represents only 0.001% of total nuclear DNA
Not practical to recover rare sequences directly from isolated nuclear DNA due to large amount of extraneous DNA sequences
Genomic library is prepared by isolating total DNA from the organism, digesting it into fragments of suitable size & cloning them into a vector i.e.

shotgun cloning –does not target a specific gene but seeks to clone all the genes at one time hoping that at least 1 recombinant clone will contain part of the gene of interest
Isolated DNA is only partially digested by the chosen restriction enzyme so that not every restriction site is cleaved in every DNA molecule
Even if the gene of interest contains a susceptible restriction site some intact genes might still be found in the digest

Genomic library is a collection of clones sufficient to contain every single gene present in a particular organism
Genomic libraries are prepared by:

- Extracting & purifying total cell DNA
- Making a partial restriction digest
- Cloning DNA fragments into a suitable vector
- Transform bacteria with recombinant DNA
- Characterize the library

- How many clones are needed to make a representative library?

Compare 2 genomes: required library size for 20 kbp fragments & 99% probability
***E. coli* genome 4.6 x 10^6** **Human genome 3 x 10^9 bp**

$N_{E.coli} = \frac{\ln(1-0.99)}{\ln[1-(2 \times 10^4/4.6 \times 10^6]}$ $N_{human} = \frac{\ln(1-0.99)}{\ln(1-(2 \times 10^4/3.0 \times 10^9)}$

$= 1.1 \times 10^3$ recombinants $= 6.9 \times 10^5$ recombinants

$(4.6 \times 10^6)/(2 \times 10^4) = 2.3 \times 10^2$ $(3 \times 10^9)/(2 \times 10^4) = 1.5 \times 10^5$

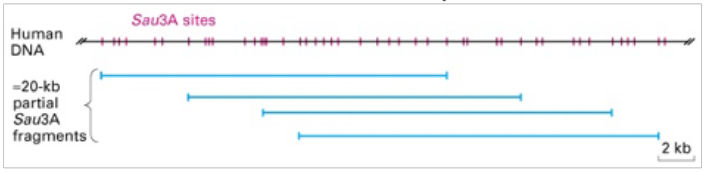
In practice for human genome:

Vector	Maximum insert size allowed	Approx no of clones needed in library
l phage	20kb	5×10^5
YAC (yeast artificial chromosome)	1000kb	1×10^4

To clone fragments of genomic DNA which are even as small as 20kb we cannot just cut the DNA to completion with a restriction enzyme – **partial digestion**
For average enzyme distance between sites is given by: **D = 4^n**
D is the distance in base pairs; **n** is number of bases in the recognition site & **4** because there are 4 different bases in DNA
4 bases in recognition site → average distance between sites of 256
6 → 4096

--------NGATCN---------*Sau*3AI
--------NCTAGN----------
N--------
NCTAG

*Sau*3AI cuts 1/256bp so ~ 10^7sites in human genome
If only 1/80 of sites are cut THEN fragments will on average be then ~20,000 bp



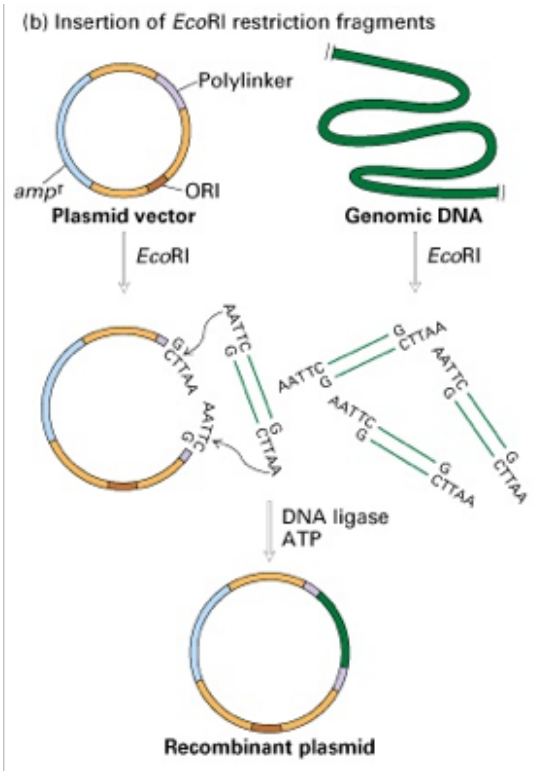
Partial digestion as a means of isolating longer overlapping DNA fragments (set of overlapping clones or sequences is a **contig**)

- Neither of these distances is long enough to be useful in a complete digestion so a partial digestion is used to generate suitable large fragments of DNA
- May get non-random distribution of fragment sizes due to non-random distribution of enzyme sites

• **Isolation** & **ligation** of **DNA fragments**

• **Purify fragments** of required size by agarose **gel electrophoresis** & **elute** fragments from gel

• **Ligate** fragments into a vector **pre-cut** with a **restriction enzyme** using **DNA ligase**

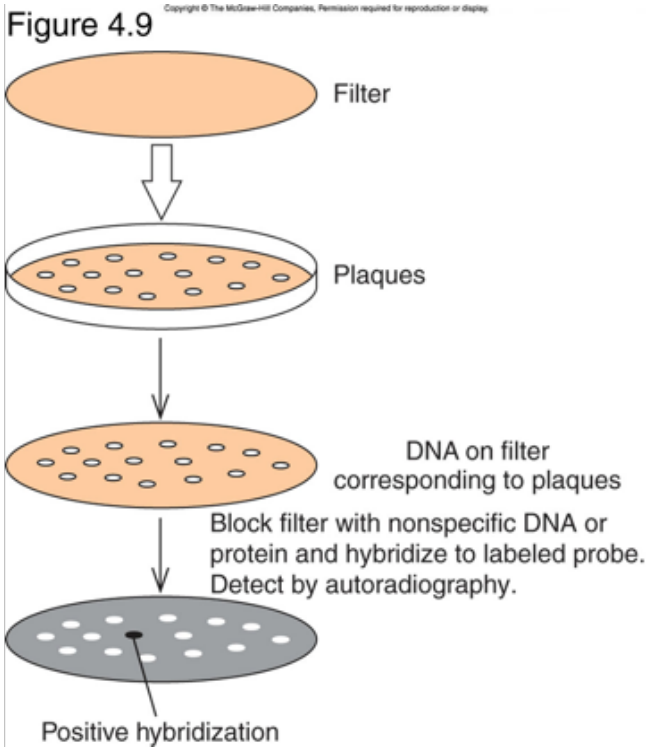


• Phage particles & extracellular viral genomic DNA transferred to nitrocellulose filter

• Nitrocellulose filter is incubated with Radiolabelled (or biotinylated) probe for gene of interest

• Autoradiograph will locate the clone with gene of interest

• Clone can be isolated & inoculated into new host cells for further amplification



cDNA library

- A collection of cloned cDNA (complementary DNA) fragments from particular cells or tissues
- cDNA is produced from mRNA, thus contains only the expressed genes of cells/tissues
- Does not contain untranslated region, e.g. promoters, introns
- Inserted into a vector
- Represents the transcriptome of the cells/tissues
- Used for screening high expression or functional important genes and the isolation of coding sequences of genes

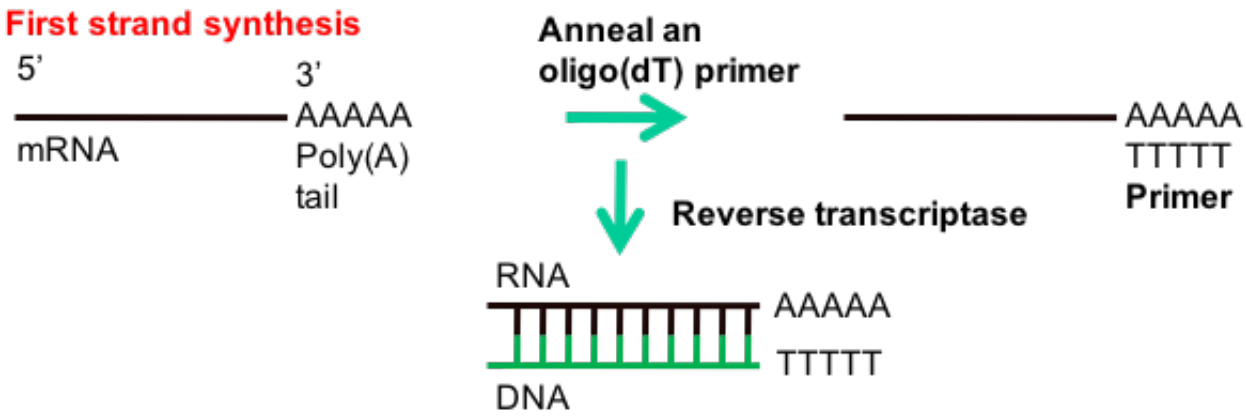
Multi-cellular organisms have specialization of individual cells e.g. liver cells, brain cells etc
Each cell contains the same complement of genes but in different cell types different sets of genes are switched on while others are silent
If messenger RNA (mRNA) cloned only those genes being expressed will be cloned
If mRNA is the starting material then resulting clones comprise only a selection of the total number of genes in the cell
mRNA can be cloned as cDNA

cDNA CLONING 1

mRNA cannot be cloned directly but a DNA, a copy of the mRNA, can be cloned
This conversion is accomplished by the action of reverse transcriptase & DNA polymerase
Reverse transcriptase makes a single-stranded DNA copy of the mRNA
Second DNA strand is generated by DNA polymerase & double- stranded product is introduced into an appropriate plasmid or l vector

cDNA CLONING 2

Key to cDNA cloning procedure is synthesis of cDNA from mRNA template by **reverse transcriptase**
mRNA obtained & purified from other RNAs by trizol extraction & column purification
Reverse transcriptase cannot initiate DNA synthesis without a primer
Use oligo(dT) as primer – this is complementary to poly(A) tail at the 3'-end of most eukaryotic mRNAs
Oligo(dT) binds to poly(A) at 3'-end of mRNA & primes DNA synthesis using the mRNA as a template

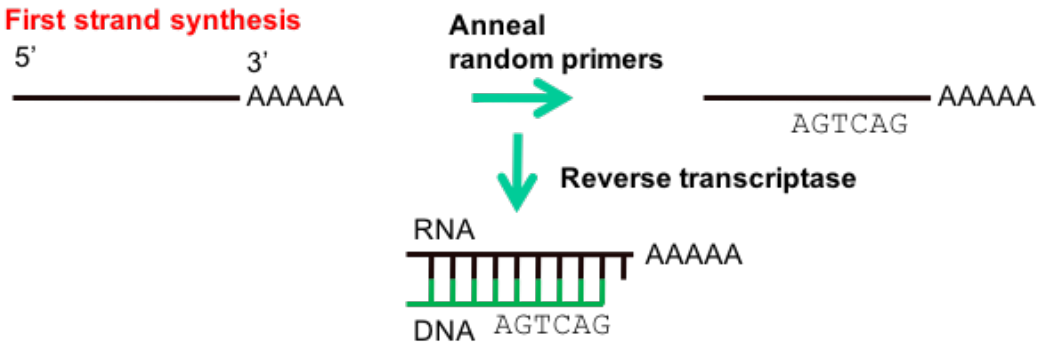


• Random primer can also be used

AGTCAG	AGTCTG	AGTCCG	AGTCGG	AGTTAG
AGTCAT	AGTCTT	AGTCCT	AGTCGT	AGTTAT
AGTCAC	AGTCTC	AGTCCC	AGTCGC	AGTTAC
AGTCAA	AGTCTA	AGTCCA	AGTCGA	AGTTAA

.....

4x4x4x4x4x4=4096

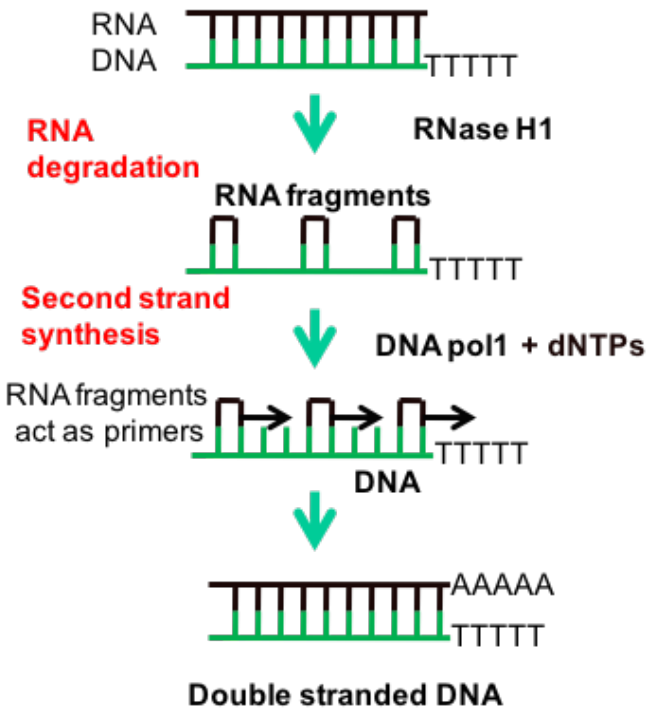


cDNA CLONING 3

• After mRNA copied → single-stranded DNA ("first strand") the RNA is partially degraded with ribonuclease (RNase) H1. Degrades RNA strand of RNA/DNA hybrid

• Remaining RNA fragments act as primers for DNA pol1 "second strand synthesis" which uses the first as the template

• Result is double-stranded cDNA

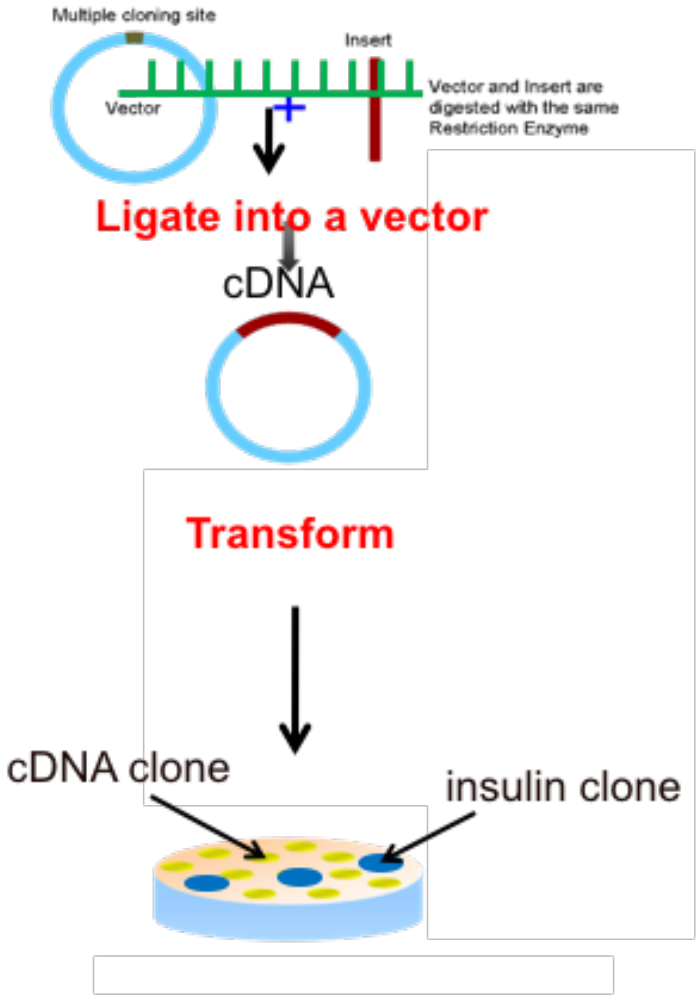


cDNA CLONING 4

- Ligate cDNA into a vector
- cDNAs have no sticky ends so ligate blunt ends or attach sticky ends (for more efficient ligation)
- cDNA clones are representative of mRNA present in original preparation
- cDNA library would contain a large proportion of clones representing insulin mRNA (other clones will also be present)

- Identify clones by hybridization of specific probe

GENE CLONING 4



Application of gene cloning:

- **Research tool**
  e.g. production of transgenic animals to study biological questions
- **Agriculture**
  e.g. genetically modified crops such as increased resistance of maize to corn borer

3. **Medicine**
   e.g. bulk production of compounds such as factor VIII, gene therapy etc
4. **Forensic science**
   e.g. identification of crime suspects e.g. Colin Pitchfork who in 1988 was the first criminal caught & convicted using DNA profiling

Transgenic mice:

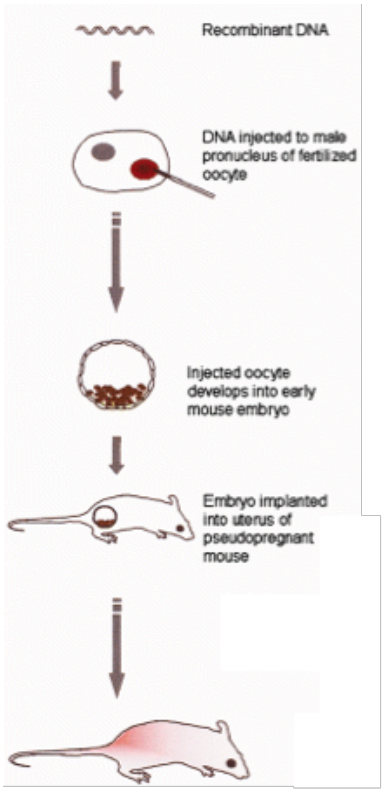
- Transgenic mouse contains additional artificially-introduced genetic material in every cell
- Used to study gene function/regulation – gain of function e.g. mouse may produce a new protein or loss of function if the integrated DNA interrupts another gene
- Transgenic mouse is a useful system for studying mammalian genes because analysis is carried out on the whole organism
- Transgenic mice also used to model human diseases that involve the over- or mis-expression of a particular protein

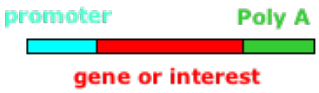
Example:

- Normal mice cannot be infected with polio virus since they lack the cell-surface molecule (CD155) that in humans is the receptor for the virus. So normal mice cannot serve as a model for studying the disease

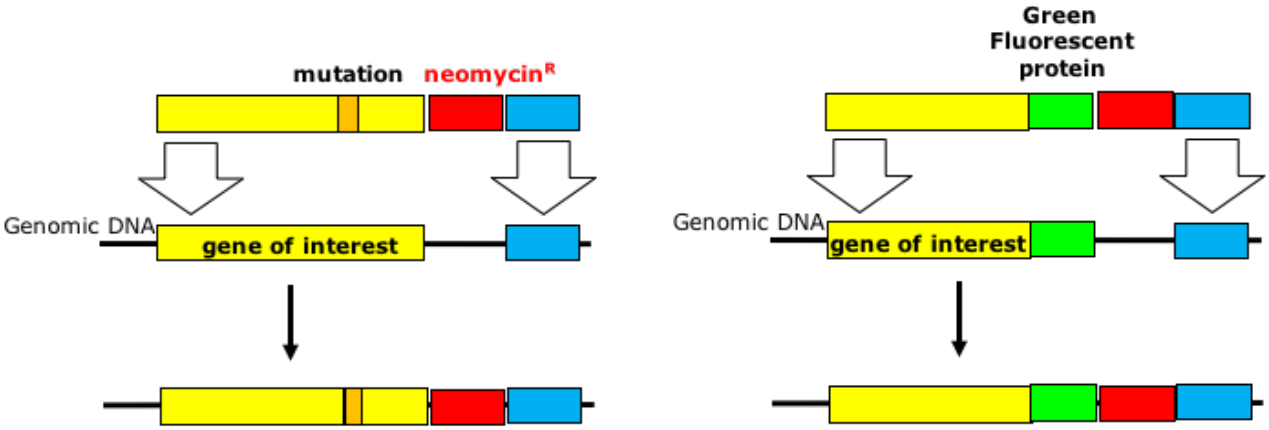
- Transgenic mice expressing the human gene for the polio virus receptor can be infected by polio virus → paralysis & other pathologies similar to human disease

- Foreign DNA may → gain of function e.g. production of a new protein or the expression of an existing protein at a higher level or in a different range of cells
- Useful in studying gene function/regulation & to model human diseases caused by dominantly acting mutant proteins e.g. Alzheimer's disease
- Foreign DNA can also → loss of function if it interrupts/disturbs expression of an existing gene. A strategy → knockout mice

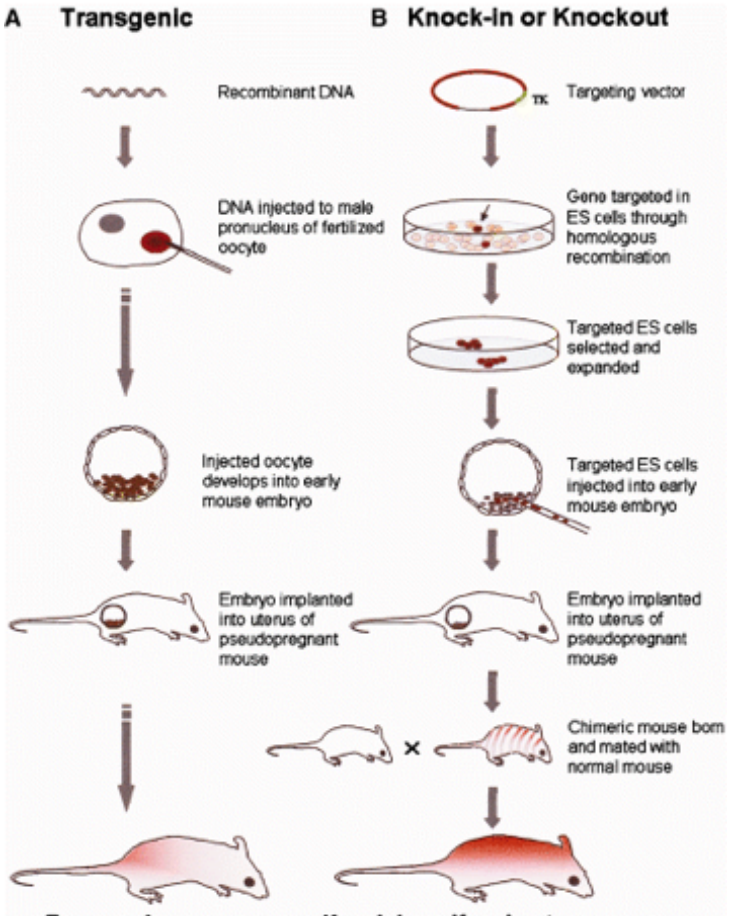


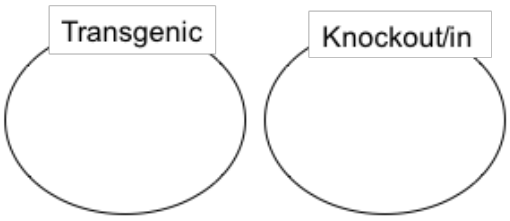


- Embryonic stem (ES) cells are derived from very early mouse embryo & can differentiate into all types of cell when introduced into another embryo
- DNA → ES cells may integrate randomly but if introduced DNA is similar in sequence to part of mouse genome it may undergo 'homologous recombination' & integrate as a single copy at a specific site
- ES cells will colonise host embryo & contribute to germ line→ production of some sperm carrying the extra DNA
- When these transgenic sperm fertilise normal egg → transgenic mouse with same foreign DNA in every cell



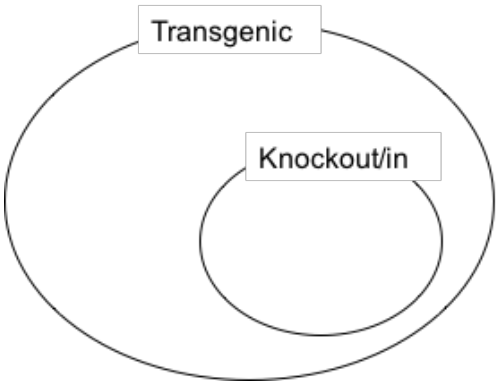
Gene of interest is not deleted. Instead, additional function (e.g. mutation, chimera with foreign protein) is added.





Transgenic: DNA is injected into zygote
i.e. without homologous recombination

Knockout/in: DNA is injected into ES cells
i.e. with homologous recombination



Transgenesis: the process of introducing an exogenous gene (transgene) into an organism, so that the organism will transmit that exogenous property to its offspring

- **AGRICOLTURE**:

- Genetically engineered crops
  e.g. replacement for insecticides
- Cloning of genes in animals
  e.g. transgenic farm animals for protein production

3. Cloning of animals
   e.g. Dolly

EUROPEAN CORN BORER:

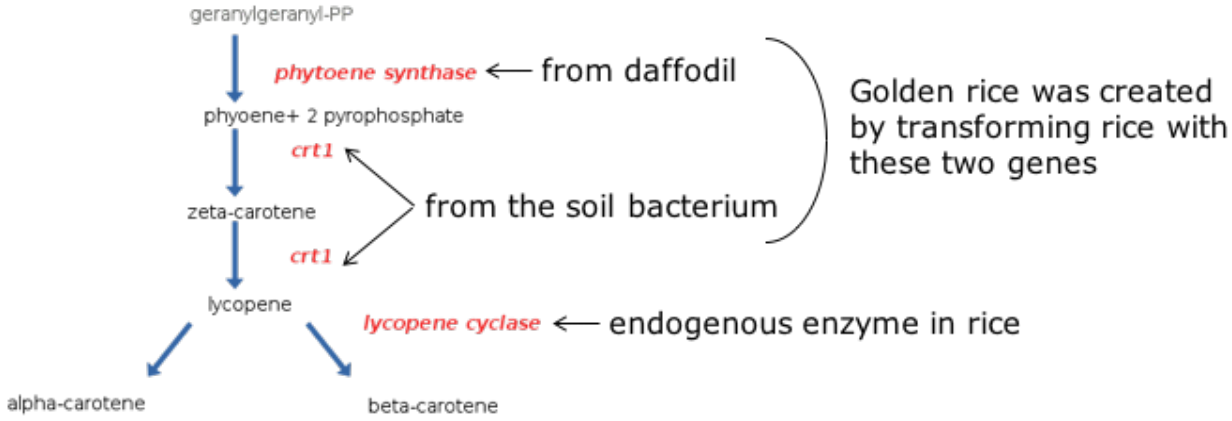
- European corn borer eggs are laid on the underside of leaves
- Hatch in 3-9 days depending upon weather conditions
- Evades the effects of insecticides
- Need to express insecticide directly in the plant

CLONING AND EXPRESSION OF ENDOTOXIN:

- Bacterium *Bacillus thuringiensis* (BT) has evolved defence mechanism to survive in the gut of insects by producing d-endotoxin, CryIA(b)
- Accumulates as an inactive precursor in bacteria but after ingestion by insect protoxin cleaved by proteases in alkaline condition→ active toxin which binds to the epithelium of insect gut and causes cell lysis by the formation of cation-selective channels, which leads to death
- The protoxin cannot be cleaved in human gut, due to the high acidity in stomach
- CryIA(b) protein is 1115 amino acids in length but toxic activity resides in segment 29-607 therefore first 648 codons made by PCR
- Ligated into a vector between promoter & polyadenylation signal (required for production of mature mRNA for translation) from cauliflower mosiac virus
- Introduced into maize embryos by mircroprojectile bombardment

GOLDEN RICE:

- Vitamin A deficiency is a serious problem in developing world, responsible for 1–2 million deaths, 500,000 cases of irreversible blindness annually
- Golden rice was genetically engineered to express β-carotene, a precursor of vitamin A, in the edible parts of rice (The rice plant can naturally produce β-carotene in its leaves, where it is involved in photosynthesisis)



- Although golden rice was developed as a humanitarian tool, it has met with significant opposition from environmental and anti-globalization activists

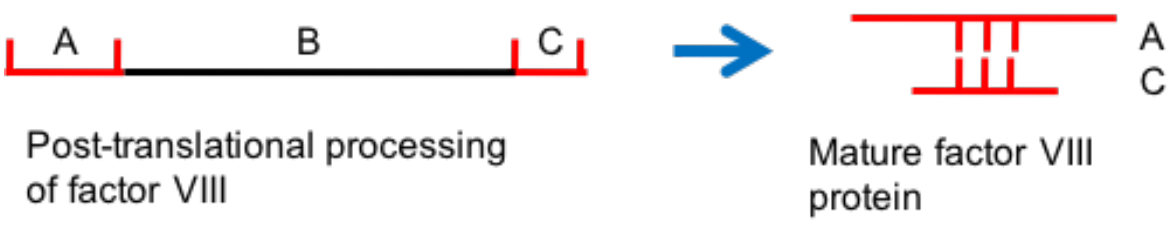
APPLICATION IN MEDICINE:

- Production of recombinant pharmaceuticals
  e.g. insulin, growth hormone, vaccines etc
- Identification of genes causing human diseases
  e.g. cystic fibrosis etc
- Gene therapy & cancer
  e.g. cystic fibrosis, congenital blindness RPE65 defect etc

Factor VIII

- Central role in blood clotting
- Failure to synthesize factor VIII → haemophilia
- Until recently treat haemophilia by injection of purified factor VIII
- Purification - difficult to remove virus particles that may be present (e.g. Hepatitis & HIV have been passed onto haemophiliacs via factor VIII injections)

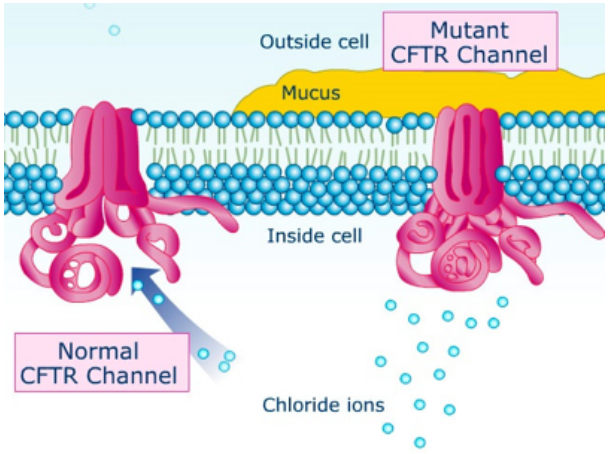
- Factor VIII 186 kb → 2351 amino acid polypeptide → complex
  post-translational processing → dimeric protein with 17 disulphide bonds



- Not possible to synthesize active protein in *E. coli*
- Cloned as 2 subunits – each cDNA fragment ligated into an expression vector between promoter and polyadenylation signal
- Plasmid → hamster cell line → functional factor VIII

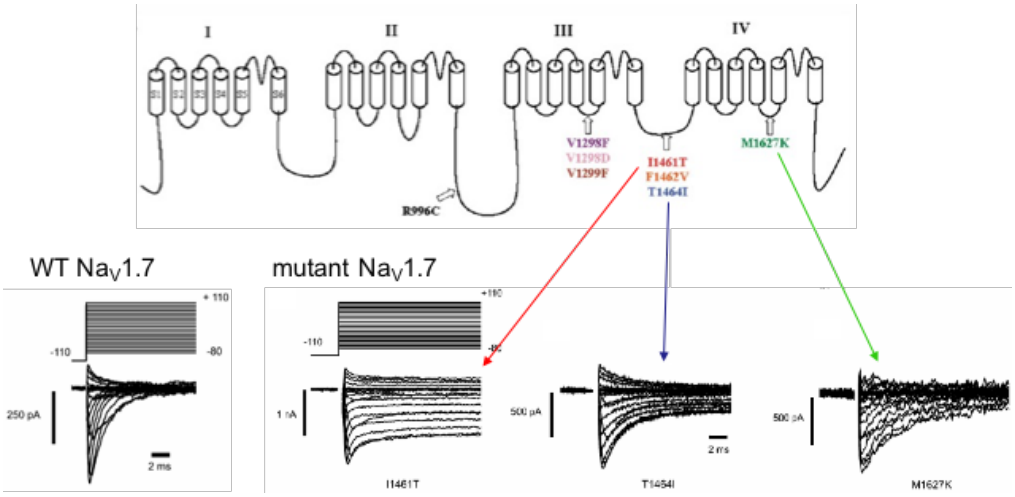
CYSTIC FIBROSIS:

- Cystic fibrosis is an autosomal recessive genetic disorder
- Causes difficulty in breathing, sinus infections, poor growth, and infertility
- Characterised by abnormal transport of chloride and sodium ions across an epithelium, leading to thick, viscous secretions
- Cystic fibrosis is caused by a mutation in the gene for the protein cystic fibrosis transmembrane conductance regulator (CFTR), an ion channel that transports chloride ions



- CFTR regulates the movement of chloride ions across epithelial membranes. Mutations in CFTR gene lead to loss of function, resulting accumulation of chloride ions inside the cells, causing sticky mucus to build up on the outside of the cells

Voltage gated sodium channel Nav1.7 (ability to sense pain)



One-point mutation of sodium channel Nav1.7 causes prolonged channel opening and chronic pain
(gain of function mutation)

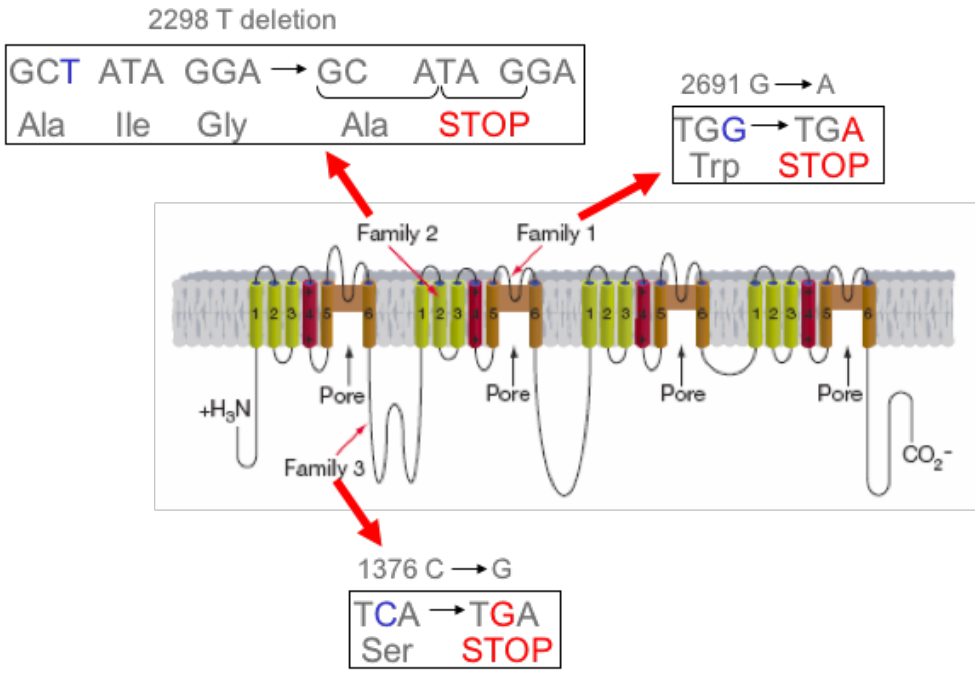
- **GENE THERAPY:**

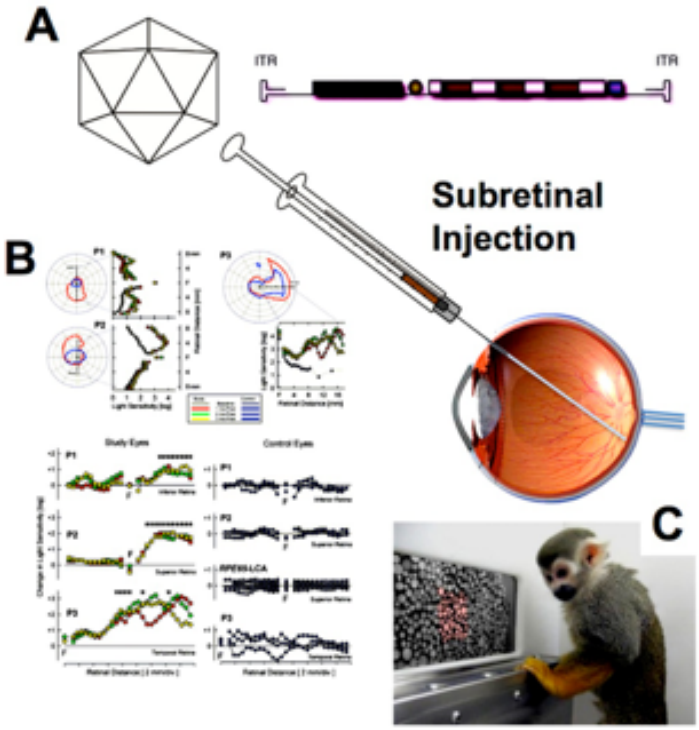
- Basic concept of gene therapy involves introducing gene → target cells to cure, prevent or slow down the progression of disease
- Manipulation of cells removed from organism which are transfected & placed back e.g. stem cells from bone transfected with DNA used in the treatment of blood disorders
- Cancer cell gene therapy → anti-sense RNA to silence oncogene

- Cystic fibrosis is a good candidate as it is caused by mutations in a single gene
- In 1993, the first experimental gene therapy treatment was given to a patient with cystic fibrosis
- A modified common cold virus was used as a delivery vehicle or vector to carry the normal CFTR gene to the cells in the airways of the lung
- This gene therapy may become widely available by 2020

CONGENITAL BLINDNESS:

- One type of congenital blindness is caused by a mutation in the Retinal Pigment Epithelium-specific 65 kDa protein (RPE65)
- RPE65 process a type of vitamin A needed to keep light-sensing photo-receptor cells (i.e. the rods and cones of the retina) in operating order
- Administration of an adeno-associated virus (AAV) carrying the normal RPE65 gene was injected into the eyes of the patients, aiming to replace the non-functioning RPE65 gene with one that works
- The treatment has restored significant vision for some patients

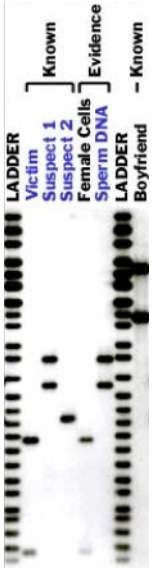




- **FORENSIC SCIENCE:**

- Any type of organism can be identified by examination of DNA sequences unique to that species
- Identifying individuals within a species is less precise but as DNA sequencing technologies develop direct comparison of very large DNA segments (whole genomes) will be possible → precise individual identification

- DNA profiling: molecular genetic analysis that identifies DNA patterns
- Based on the principle that individuals have their unique DNA patterns
- Restriction Fragment Length Polymorphism (RFLP) can be studied against Variable Number Tandem Repeat (VNTR)
- Perform Southern blotting using a probe for the repeat



INTRODUCTION TO GENETICS

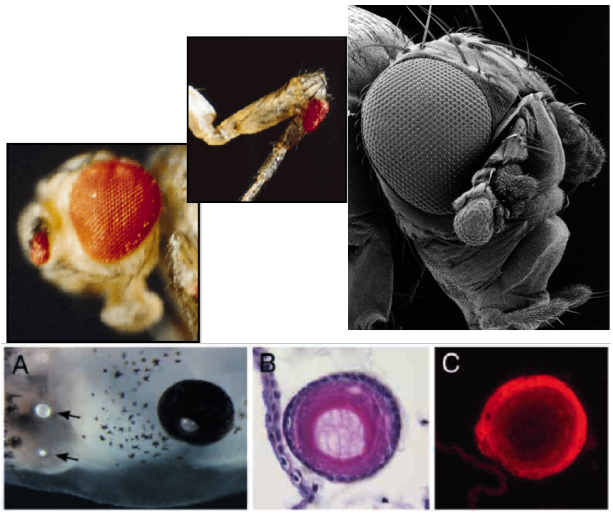
Genetics is concerned with precisely understanding the biological properties that are transmitted from parent to offspring.

Genetics is also concerned with the mechanism of hereditary transmission.

In addition genetics is concerned with the nature of the hereditary material

How variations in the hereditary information arise and what is their nature.

Moreover concerned with the ways that this information functions to dictate the features that constitute any given species



Ectopic expression of the gene eyeless causes the formation of eye structures on a Drosophila antenna and leg. Mouse Pax6 misexpression in antennae.
Fig. 1. Pax6 misexpression results in eye-related phenotypes. Xenopus embryos were injected with 160 pg Pax6 RNA in one animal pole blastomere at the 16-cell stage and fixed at stage 48. (A) Ectopic lenses (arrows) in the absence of retinal tissue. (B) Hematoxylin and eosin stained section showing an isolated ectopic lens adjacent to surface ectoderm. The lens fiber cell mass is surrounded by hematoxylin stained nuclei of the lens epithelial layer. (C) ß-crystallin immunolabeling of the ectopic lens in B.

- "Master" genes involved in fundamental biological and biochemical processes are conserved between species.

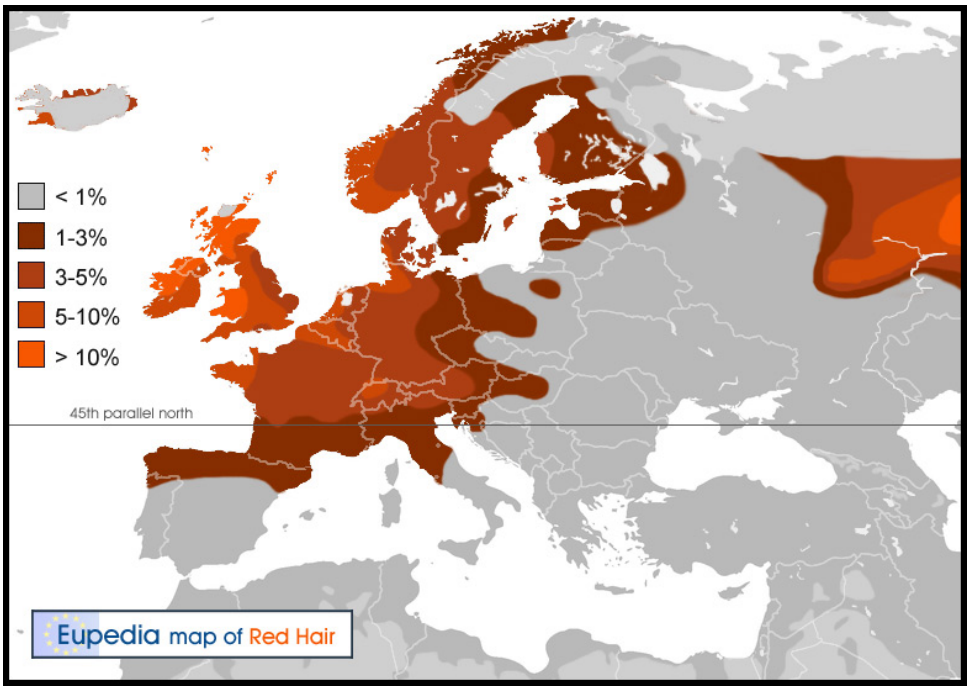
- The PAX-6 / *eyeless* gene tells us that there is an evolutionary connection between all the myriad eye forms in nature (PAX-6 gene is essential and sufficient to make eyes)

- The study of genes in one species is likely to tell us about the biochemistry and biology of all species.

- These genes are conserved in form of function (mouse pax6 genes are ectopically expressed in Drosophila antenna--> antenna grow eyes)

- **GENETIC OF RED HAIR**

Distribution of red hair is not random across Europe. Red hair can be seen particularly in North Western European populations. This is due to inability to make eumelanin (black / brown pigment) so only pheomelanin (red / yellow pigment) is made

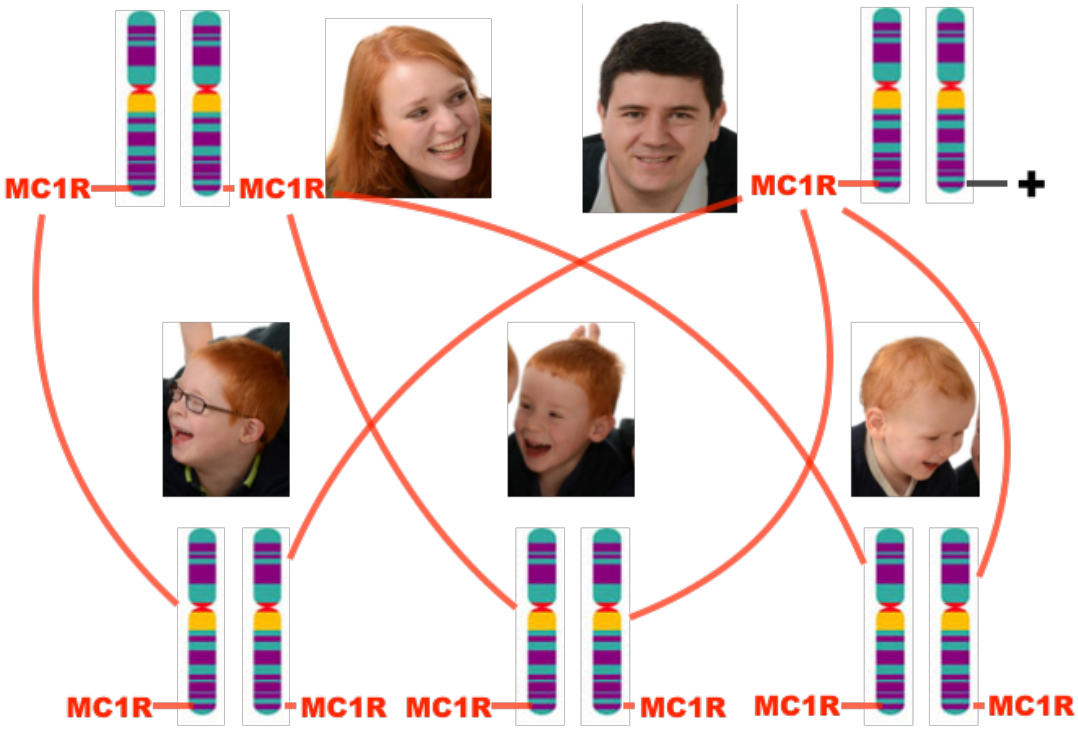


Non random distribution shows that red hair is heritable (and not environmental).

- Manuel Corpas had his exome (the part that gives rise to mRNA) sequenced.
- exome = the DNA sequence complementary to all of the mRNA made from the genome.
- He carries a mutation in the Melanocortin 1 receptor (A instead of G)

```
16 89986091 rs11547464 G A
```

- This mutation on his chromosome 16 has been inherited by his sons.
- Why doesn't Manuel have red hair?
- Manuel has 2 chromosome 16s and 2 MC1R genes
- The MC1R mutation is recessive and Manuel is heterozygous - he has one normal version of the gene and one mutant version
- Each time the fertilisation event leading to his sons occurred, the sperm contained only the mutant form of MC1R



- Gene - they determine a particular characteristic e.g. height, eye colour.

- The particular flavour of that characteristic we call an allele

- Individuals with two different alleles are heterozygous, with two alleles the same are homozygous

- The genetic constitution summing all of the alleles is called the genotype

- The observed form of a character of an organism is the phenotype

MENDELIAN GENETICS

Red hair is an example of what we call Mendelian inheritance after the person who first understood it - Gregor Mendel

A 19th century cleric devoted to teaching and interested in science

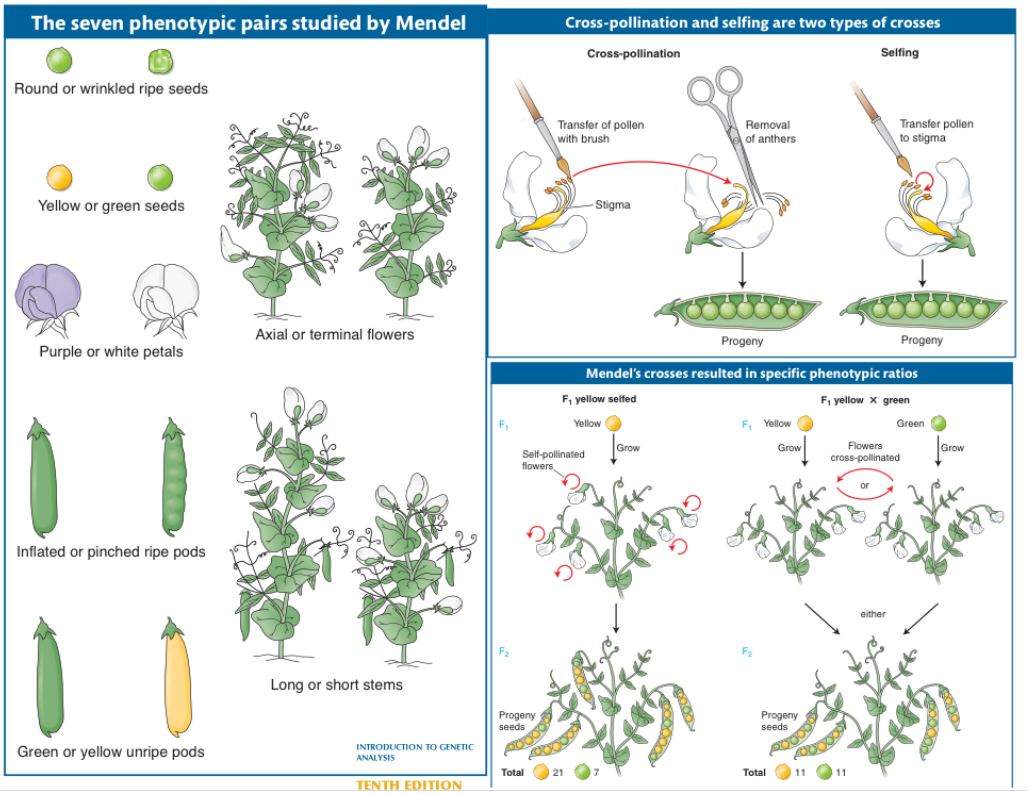
Mendel made some methodological breakthroughs in his work on pea breeding

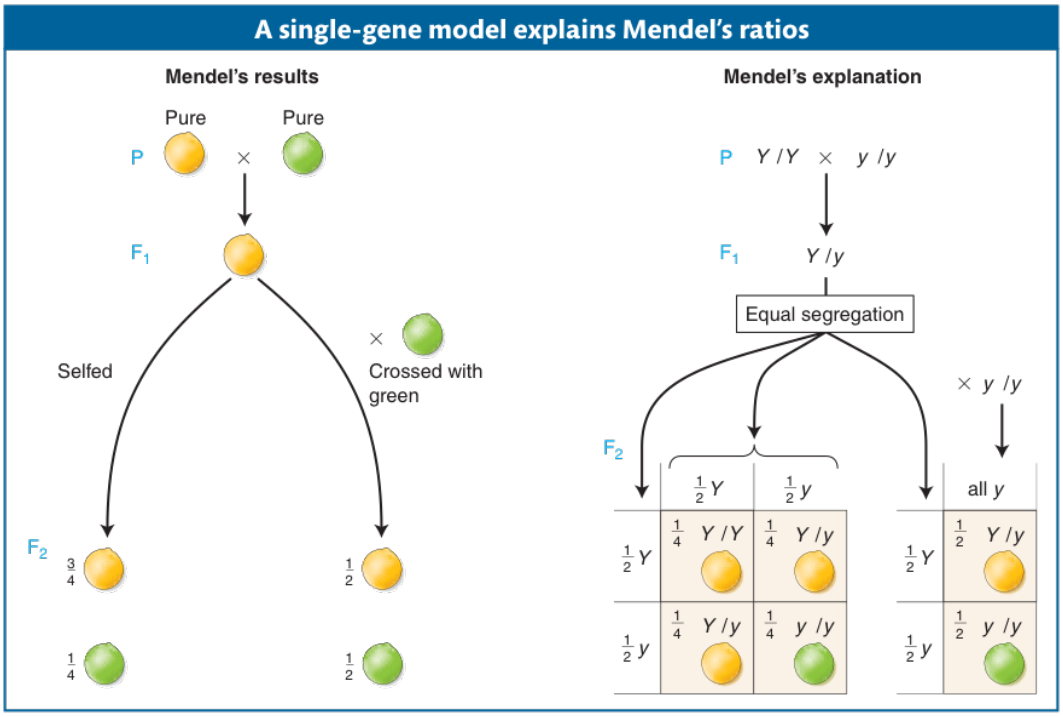
1. He used pure breeding lines: all individuals the same generation after generation

2. He used single pairs of cleanly distinguishable alternatives

3. He collected numerical data

His rigorous collection of significant amounts of data allowed him to proposed 2 rules to predict types of progeny, and proportions of each type, knowing certain information about the parents





The "particular factor" is conserved in the F1 progeny and can reappear. So there are two particular factors for a given trait. One dominant and one recessive.

- 1. Characters are determined by particulate factors (genes) existing in pairs in each organism

- 2. Each reproductive cell (gamete) contains only one of each pair - which one is entirely random - thus these factors segregate

- 3. The union of male and female reproductive cells is a random process that reunites pairs of factors (genes)

- Mendel's work for the first time allows not only types of progeny to be predicted from a particular genetic cross, but also proportions of progeny

- The basis of all three branches of genetics:

    - Transmission genetics

    - Can we identify genes and phenotypes, mapping genes?

    - Molecular genetics

    - What is the molecular basis to heredity - information flow?

    - Population genetics

    - What can we say about the genetic makeup of populations in their natural state?

THE CHROMOSOME BASIS TO MENDEL

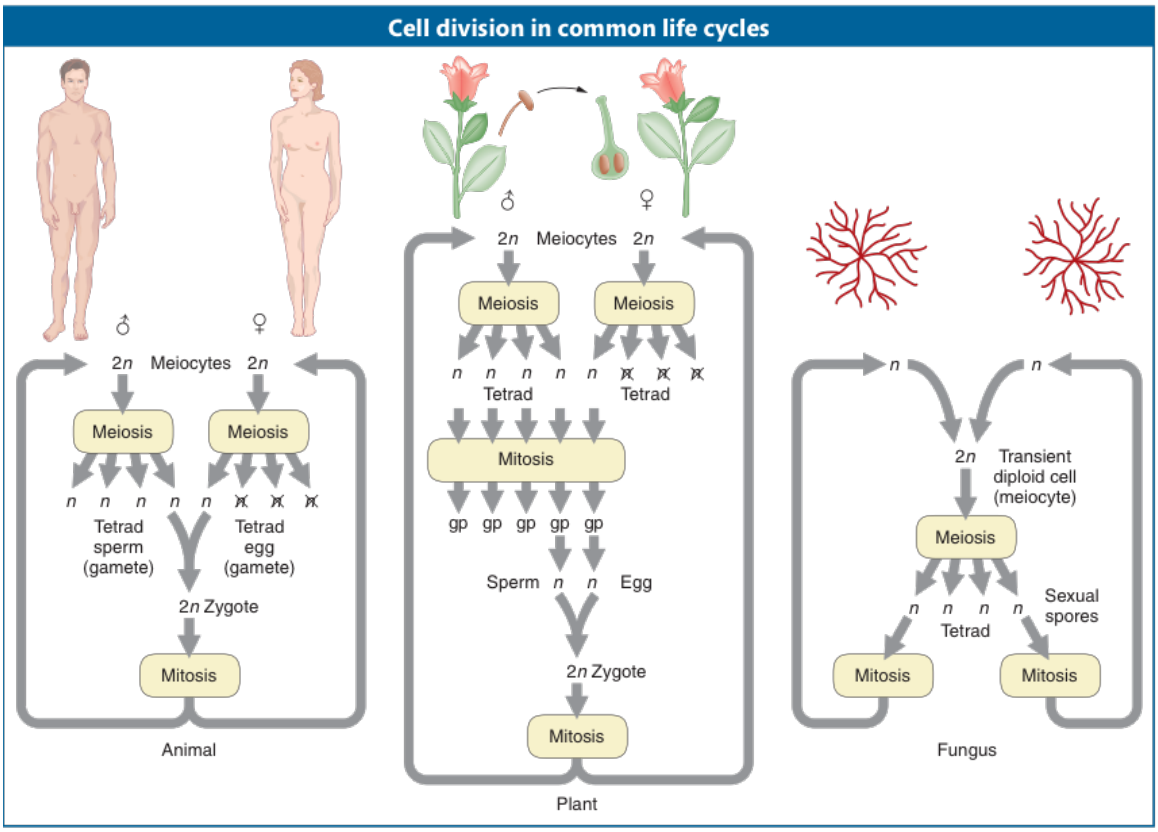
- All cells of the organism are clones - with an identical genome.

- Mitotic cell division permits this to be the case

- A specialised cell division is required to make the gametes such that each gamete (sperm and egg in animals) has half the genetic complement of all other (somatic) cells

- Somatic cells are diploid (2n) but gametes are haploid (n)

- Ability of an organism to derive from a single cell (zygote): this is due to mitosis, but to make the sperm --> meiosis (haploid).



The life cycles of humans, plants, and fungi, showing the points at which mitosis and meiosis take place. Note that in the females of humans and many plants, three cells of the meiotic tetrad abort. The abbreviation n indicates a haploid cell, 2n a diploid cell; gp stands for gametophyte, the name of the small structure composed of haploid cells that will produce gametes. In many plants such as corn, a nucleus from the male gametophyte fuses with two nuclei from the female gametophyte, giving rise
to a triploid (3n) cell, which then replicates to form the endosperm, a nutritive tissue that surrounds the embryo (which is derived from the 2n zygote).

- Homologues synapse

- Chiasma form between homologues

- Homologues segregate at meiosis I

- No replication preceding meiosis II

- Gives 4 cells with haploid complement of chromosomes

Note:

- How chromosomes behave is identical to how genes behave to account for pattern of inheritance:

1. Present in pairs

2. Segregate

3. One member of each pair inherited

4. One received from each parent

- Fungi such as yeast exist primarily as haploid, not diploid organisms

- Can go through a transitory diploid stage which quickly goes into meiosis

- In the haploid state there is no dominant or recessive - all alleles are expressed