Republic of Iraq

Ministry of Higher Education

and Scientific Research

Short Course in

Bacterial Genetics

First Edition

Edited by

Mohammed Faraj Al-Marjani

Department of Biology

College of Science

Al - Mustansiriyah University

Contents

Preface

Historical Perspectives on Bacterial Genetics **1**

Bacterial Structure and Function 5

Nucleic Acid **10**

Nucleic Acid Structure 10

Bacterial Genome **12**

Bacterial chromosome in the nucleoid 16

The nucleoid 16

The genophore 16

DNA super coiling 17

DNA Replication **18**

Gene Expression **21**

RNA Transcription 21

Translation 26

Mutation **31**

Mutagens 31

Type of mutagens 32

Type of mutation 36

Extra chromosomal Agents **47**

Plasmid 47

Transposable elements 53

Horizontal Gene Transfer **62**

Conjugation 62

Transformation 76

Transduction 81

Recombination **86**

Restriction and Modification System **90**

Gene Cloning **98**

Vectors 100

Restriction enzymes 104

DNA ligase 104

Transformation 105

Basic Molecular Biology Tools **109**

The Polymerase Chain reaction 109

DNA Sequences 114

Hybridization 119

Gene Mapping in Bacteria **123**

Preface

This book is written for biological students who are seeking a brief summary of bacterial genetics. As the importance of Bacterial Genetics in our daily lives has grown, so too has the importance of its place in the modern science classroom in the study of Microbiology. Bacterial genetics has become the central science. The purpose in creating this short course is to provide students the basic information for understanding this rapidly changing field. Chapters are divided into concepts, the main mutations and horizontal gene transfer by conjugation , transformation and transduction.

Bacterial genetics is a rapidly growing and changing subject ,the first chapters are devoted to understanding structure of DNA ,replication and gene expression .The book then focuses on mutations , extra chromosomal DNA and gene transfer. The final series of chapters cover the gene cloning ,molecular biology tools and gene mapping .

Finally, I would welcome comment so that we might all learn - as we always do - from the errors.

Mohammed Faraj Al-Marjani

1 Historical Perspectives on Bacterial Genetics

Bacterial genetics is the subfield of genetics devoted to the study of bacteria. Bacterial genetics are different from eukaryotic genetics, however bacteria still serve as a good model for animal genetic studies. One of the major distinctions between bacterial and eukaryotic genetics stems from the bacteria's lack of membrane-bound organelles(this is true of all prokaryotes). Bacterial genetics has led to the establishment of the universal “central dogma” on the transfer of genetic information resulted in a number of discoveries that provided essential clues to understanding complex phenomena in many fields of higher eukaryote.

The differences in morphology and other properties were attributed by Nageli in 1877, to bacterial pleomorphism, which postulated the existence of a single, a few species of bacteria, which possessed a protein capacity for a variation. With the development and application of precise methods of pure culture, it became apparent that different types of bacteria retained constant form and function through successive generations. This led to the concept of monomorphism. Friedrich Miescher in 1871 isolates “nucleic acid” from pus cells.

In the early 1900s, with the rediscovery of Mendel’s work, the factors conveying hereditary traits were named ‘‘genes’’ by Wilhelm Johansson. A large amount of research since then has led to our current understanding about what constitutes a gene and how genes work. In 1902 , Walter Sutton and Theodore Boveri, using dyes synthesized by the German organic chemistry industry, observed that "colored bodies" in cells behaved in ways parallel to the hypothetical agents of heredity proposed by Mendel. These bodies were called chromosomes. The word "genetics" is coined by William Bateson in 1905.

Archibald Garrod pointed the field of inborn errors of metabolism in 1904 ,He also discovered alkaptonuria which is a human disease characterize by black urine due to accumulation of unusual metabolic products , understanding its inheritance.

The discovery of transformation was one of the seminal events in biology, as it led to experiments demonstrating that DNA was the genetic material . This is discovery became the cornerstone of molecular biology and modern genetics, The first evidence of bacterial transformation was obtained by the British scientist Frederick Griffith in the late 1928. Griffith was working with *Streptococcus pneumoni*e , virulent strain had capsule and non virulent strain lack this capsule and are thus unable to cause infection called R strains because their colonies appear rough on agar, in contrast to the smooth appearance of encapsulated strains, called S strains. A mouse infected with only a few cells of an S strain succumbs in a day or two to a massive pneumococcus infection. By contrast, even large numbers of R cells do not cause death when injected. Griffith showed that if heat-killed S cells were injected along with living R cells, the mouse developed a fatal infection and the bacteria isolated from the dead mouse were of the S type . Since the S cells isolated in such an experiment always had the capsule type of the heat-killed S cells, Griffith concluded that the R cells had been transformed into a new type. The process had all the properties of a genetic event and so set the stage for the discovery of DNA.

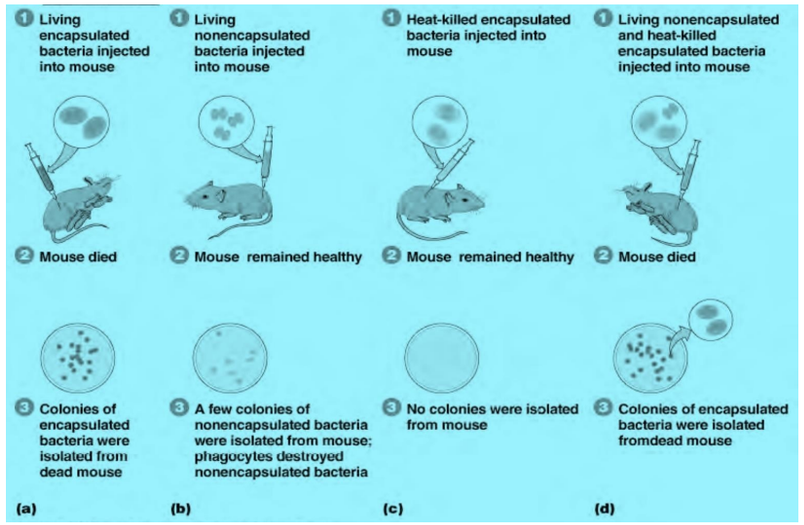


Figure 1 : Evidence for Bacterial Transformation (Griffith Experiment)[copyright 2001 Benjamin Cummings ,an imprint of Addison Wesley Longman.Inc].

The molecular explanation for the transformation of pneumococcus types was provided by Oswald T. Avery and his associates at the Rockefeller Institute in New York in a series of studies carried out during the 1930s, culminating in the classic paper by Avery, C. M. MacLeod, and M. McCarty in 1944. Avery and his coworkers showed that the transformation process could be carried out in the test tube rather than the mouse and that a cell-free extract of heat killed cells could induce transformation. In a series of painstaking biochemical experiments, the active fraction of cell-free extracts was purified and was shown to be DNA.
The transforming activity of purified DNA preparations was very high, and only a very small amount of material was necessary. Subsequently, others at Rockefeller showed that transformation could occur in pneumococcus not only for capsular characteristics but for other genetic characteristics of the organism as well, such as antibiotic resistance and sugar fermentation.

Rosalind Franklin and Maurice Wilkins in 1950 , using X-ray crystallographic equipment to solve the DNA problem determine the Three dimensional structure of the DNA or protein. Alfred Hershey and Martha Chase in 1952 confirming that DNA was the genetic material, which had first been demonstrated in the 1944 Avery–MacLeod–McCarty experiment .

Beadle and Tatum's key experiments (1946-1956) involved exposing the bread mold *Neurospora crassa* to x-rays causing mutations to cause changes in specific enzymes involved in metabolic pathways .This showed that specific genes code for specific proteins, leading to the "one gene, one enzyme" hypothesis. James Watson and Francis Crick in 1953, announced their model of the structure of DNA, providing a theoretical framework for how DNA could serve as the genetic.

An experiment by Matthew Meselson and Franklin Stahl in 1958 which supported the hypothesis that DNA replication was semiconservative.
Gobind Khorana in 1960 show how the nucleotides in nucleic acids, which carry the genetic code of the cell, control the cell's synthesis of proteins.In 1970 ,the restriction enzymes studied by Arber and Meselson were type I restriction enzymes which cleave DNA randomly away from the recognition site.

Cohen, Paul Berg and Boyer made what would be one of the first genetic engineering experiments, in 1973. They demonstrated that the gene for frog ribosomal RNA could be transferred into bacterial cells *Escherichia coli* by using a vector (plasmid) then expressed by them .

Kary Mullis in 1983 start using a pair of primers to synthesis a desired DNA sequence and to copy it using DNA polymerase, a technique which would allow a small strand of DNA to be copied almost an infinite number of times. In 1995, the first bacterial genome sequence, *Haemophilus influenzae*, was completely determined.

The modern study of genetics at the level of DNA is known as molecular genetics and the synthesis of molecular genetics with traditional Darwinian evolution is known as the modern evolutionary synthesis.

Genes can be defined in two different ways: the gene as a ‘‘unit of inheritance’’, or the gene as a physical entity with a fixed position on the chromosome that can be mapped in relation to other genes (the genomic locus). While the latter is the more traditional view of a gene the former view is more suited to our current understanding of the genomic architecture of genes. A gene gives rise to a phenotype through its ability to generate an RNA (ribonucleic acid) or protein product.
Prokaryotes are haploid-each gene is present in only a single copy (for the most part). Not only that, but bacteria multiply (divide) by binary fission (asexual reproduction), so in principle, each progeny cell is a clone of the parent. It would be logical to think that the term allele would not make senses when applied to bacteria. But in reality, mutations happen. Progeny cells may not be clones. This is just a way of saying that progeny bacterial cells might have variant genes (alleles).
They then chose certain bacteria as appropriate objects to study the inheritance because they have short generation times, which shorten the time necessary for experimentation and easily handled without special skills. They grow on simple culture media, which reduces necessary expense for supplies. Also, the simple organization of the cell structure seemed to be advantageous in interpreting experimental results. These features of bacteria are quite appropriate for analysis of heredity through not only genetic but also biochemical approaches. This attitude of scientists yielded a new discipline of biological science, i.e. microbial genetics, which later developed to “molecular genetics”.

2 Bacterial Structure and Function

Bacteria differ from other single-cell microorganisms in both their cell structure and size, which varies from 0.3–5 μm. The smallest bacteria approximate the size of the largest viruses (poxviruses), whereas the longest bacilli attain the same length as some yeasts and human red blood cells . They have three basic forms: cocci, straight rods, and curved or spiral rods. The nucleoid consists of a very thin, long, circular DNA molecular double strand that is not surrounded by a membrane.

The cytoplasmic membrane harbors numerous proteins such as permeases, cell wall synthesis enzymes, sensor proteins, secretion system proteins, and, in aerobic bacteria, respiratory chain enzymes. The membrane is surrounded by the cell wall, the most important element of which is the supporting murein skeleton. The cell wall of Gram-negative bacteria features a porous outer membrane into the outer surface of which the lipopolysaccharide responsible for the pathogenesis of Gram-negative infections is integrated.

The cell wall of Gram-positive bacteria does not possess such an outer membrane. Its murein layer is thicker and contains teichoic acids and wall-associated proteins that contribute to the pathogenic process in Gram-positive infections. Many bacteria have capsules made of polysaccharides that protect them from phagocytosis.

Attachment pili or fimbriae facilitate adhesion to host cells. Motile bacteria possess flagella.Some bacteria produce spores, dormant forms that are highly resistant to chemical and physical noxae.

- **Nucleoid (Nucleus Equivalent)**

The “cellular nucleus” in prokaryotes consists of a tangle of double-stranded DNA, not surrounded by a membrane and localized in the cytoplasm .

In *E. coli* (and probably in all bacteria), it takes the form of a single circular molecule of DNA. The genome of *E. coli* comprises 4.63 X 10^6 base pairs

(bp) that code for 4288 different proteins. The genomic sequence of many bacteria is known.

- **Cytoplasm**

The cytoplasm contains a large number of solute low- and high-molecular weight substances, RNA and approximately 20 000 ribosomes per cell.

The cytoplasm is also frequently used to store reserve substances (glycogen depots, lipids and polymerized metaphosphates).

The most important bacterial cytoplasmic membrane proteins permeases Active transport of nutrients from outside to inside against a concentration gradient. Biosynthesis enzymes required for biosynthesis of the cell wall, e.g., its murein. The enzymes that contribute to the final murein biosynthesis steps are for the most part identical with the "penicillin-binding proteins" (PBPs).

Transmit information from the cell's environment into its interior. The so-called receiver domain extends outward, the transmitter domain inward. The transmission activity is regulated by the binding of signal molecules to a receiver module. In two component systems, the transmitter module transfers the information to a regulator protein, activating its functional module.

- **Ribosomes**

Ribosomes are the sites of protein synthesis. Bacterial ribosomes differ from those of eukaryotic cells in both size and chemical composition. They are organized in units of 70S, compared with eukaryotic ribosomes of 80S. These differences are the basis of the selective action of some antibiotics that inhibit bacterial, but not human, protein synthesis.

- **The Cytoplasmic Membrane**

This elementary membrane, also known as the plasma membrane, is typical of living cells. It is basically a double layer of phospholipids with numerous proteins integrated into its structure. The most important of these membrane proteins are permeases, enzymes for the biosynthesis of the cell wall,

transfer proteins for secretion of extracellular proteins, sensor or signal proteins, and respiratory chain enzymes.

In electron microscopic images of Gram-positive bacteria, the mesosomes

appear as structures bound to the membrane.

- **Cell Wall**

The tasks of the complex bacterial cell wall are to protect the protoplasts from external noxae, to withstand and maintain the osmotic pressure gradient between the cell interior and the extracellular environment (with internal pressures as high as 500–2000 kPa), to give the cell its outer form and to facilitate communication with its surroundings.

The most important structural element of the wall is murein, a netlike polymer material surrounding the entire cell (sacculus). It is made up of polysaccharide chains crosslinked by peptides.

- **Outer membrane proteins**

The outer membrane is the salient structural element. It contains numerous proteins (50% by mass) as well as the medically critical lipopolysaccharide.

- OmpA (outer membrane protein A) and the murein lipoprotein form a

bond between outer membrane and murein.

- Porins, proteins that form pores in the outer membrane, allow passage of

hydrophilic, low-molecular-weight substances into the periplasmic space.

- Outer membrane-associated proteins constitute specific structures that

enable bacteria to attach to host cell receptors.

- A number of OmpS are transport proteins. Examples include the LamB

proteins for maltose transport and FepA for transport of the siderophore

ferric (Fe3+) enterochelin in *E. coli* .

Lipopolysaccharide (LPS) This molecular complex, also known as endotoxin, is comprised of the lipid A, the core polysaccharide, and the O-specific polysaccharide chain .

- **Capsule**

Many pathogenic bacteria make use of extracellular enzymes to synthesize a polymer that forms a layer around the cell . The capsule of most bacteria consists of a polysaccharide ,it protects bacterial cells from phagocytosis. The bacteria of a single species can be classified in different capsular serovars (or serotypes) based on the fine chemical structure of this polysaccharide.

- **Flagella**

Flagella (singular flagellum) give bacteria the ability to move about actively. The flagella are made up of a class of linear proteins called flagellins. Flagellated bacteria are described as monotrichous, lophotrichous, or peritrichous, depending on how the flagella are arranged . The basal body traverses the cell wall and cytoplasmic membrane to anchor the flagellum and enables it to whirl about its axis like a propeller.

- **Pili**

Many Gram-negative bacteria possess thin micro fimbrils made of proteins (0.1– 1.5 nm thick, 4–8 nm long), the attachment pili. They are anchored in the outer membrane of the cell wall and extend radially from the surface. Using these structures, bacteria are capable of specific attachment to host cell receptors .

The conjugation pili (syn. sex pili) in Gram-negative bacteria are required

for the process of conjugation and thus for transfer of conjugative plasmids.

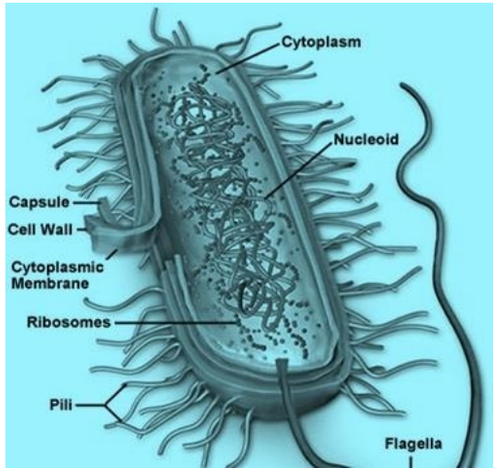


Figure 2 : Prokaryotic cell structure[molecular .magnet.fsu.edu]

- **All cells have 3 main components:**
  - **DNA (‘nucleoid”) [genetic instructions]**
  - **surrounding membrane (“cytoplasmic membrane”)**

    [limits access to the cell’s interior]
  - **cytoplasm, between the DNA and the membrane**

    [where all metabolic reactions occur]

    [especially protein synthesis, which occurs on the ribosomes]
- **Bacteria also often have these features:**
  - **cell wall (resists osmotic pressure)**
  - **flagella (movement)**
  - **pili (attachment)**
  - **capsule (protection and biofilms)**

3 Nucleic Acid

3-1 Nucleic Acid Structure

DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are polymers of nucleotides linked in a chain through phosphodiester bonds. In biological systems, they serve as information-carrying molecules or, in the case of some RNA molecules, catalysts.

Nucleotides (building blocks of nucleic acids) have a distinctive structure composed of three components covalently bound together:

- a 5-carbon sugar - ribose or deoxyribose
- a nitrogen base : pyrimidine (one ring) or purine (two rings)
- a phosphate group

The combination of a base and sugar is called a *nucleoside*. Nucleotides also exist in activated forms containing two or three phosphates, called nucleotide diphosphates or triphosphates. If the sugar in a nucleotide is deoxyribose, the nucleotide is called a deoxynucleotide; if the sugar is ribose, the term ribonucleotide is used.

There are five common bases, and four are generally represented in either DNA or RNA. Those bases and their corresponding nucleosides are described in the following table:

Abbr.	Base	Nucleoside	Nucleic Acid
A	Adenine	Deoxyadenosine	DNA
		Adenosine	RNA
G	Guanine	Deoxyguanosine	DNA
		Guanosine	RNA
C	Cytosine	Deoxycytidine	DNA
		Cytidine	RNA
T	Thymine	deoxythymidine (thymidine)	DNA
U	Uracil	Uridine	RNA

DNA and RNA are synthesized in cells by DNA polymerases and RNA polymerases by forming phosphodiester bonds between the 3' carbon of one nucleotide and the 5' carbon of another nucleotide. This leads to formation of the so-called "sugar-phosphate backbone".

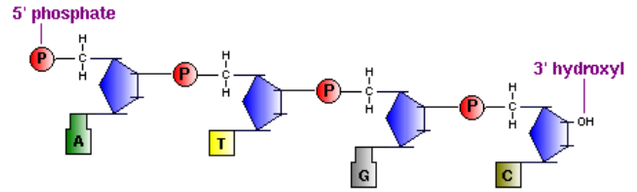


Figure 3 : Polynucleotide showing a 3 -5 phosphodiester bond.

Most DNA exists in the famous form of a double helix, in which two linear strands of DNA are wound around one another. The major force promoting formation of this helix is complementary base pairing: Adenine form hydrogen bonds with Thymine (or Uracil in RNA), and Guanine form hydrogen bonds with Cytosine.

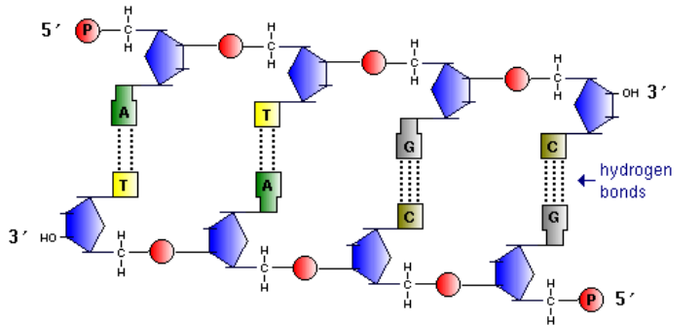


Figure 4 : Base-pairing in double-stranded DNA.

G-C base pairs have Three hydrogen bonds, whereas A-T base pairs have Two hydrogen bonds. In the Watson-Crick model, the bases are in the interior of the helix aligned at a nearly 90 degree angle relative to the axis of the helix.

The double helix of DNA has been shown to exist in several different forms, depending upon sequence content and ionic conditions of crystal preparation. The B-form of DNA prevails under physiological conditions of low ionic strength and a high degree of hydration. Regions of the helix that are rich in C and G dinucleotides can exist in a novel left-handed helical conformation termed Z-DNA. This conformation results from a 180 degree change in the orientation of the bases relative to that of the more common A- and B-DNA.

RNAs are usually single stranded, but many RNA molecules have secondary structure in which intramolecular loops are formed by complementary base pairing. Base pairing in RNA follows exactly the same principles as with DNA: the two regions involved in duplex formation are antiparallel to one another, and the base pairs that form are A-U and G-C.

3 -2 Bacterial Genome

The genome of an organism can be defined as the total DNA content of the cell , and as such it contains all the genetic information required to direct the growth and development of the organism. For all multicellular organisms this growth and development starts from a single cell.

Bacterial genomes are generally smaller and less variant in size between species when compared with genomes of single cell eukaryotes and animals . Bacterial genomes can range in size anywhere from 139 kbp to 13,000 kbp. Recent advances in sequencing technology led to the discovery of a high correlation between the number of genes and the genome size of bacteria, suggesting that bacteria have relatively small amounts of junk DNA.

Studies have since shown that a large number of bacterial species have undergone genomic degradation resulting in a decrease in genome size from their ancestral state. Over the years, researchers have proposed several theories to explain the general trend of bacterial genome decay and the

relatively small size of bacterial genomes. Compelling evidence indicates that the apparent degradation of bacterial genomes is owed to a deletional bias.

In prokaryotes, most of the genome (85-90%) is non-repetitive DNA, which means coding DNA mainly forms it, while non-coding regions only take a small part. In bacteria there is very little wastage of DNA and little repetitive DNA as seen in higher eukaryotes. Only about 10% of the genome is non-coding and even this 10% in most cases plays a critical role to the cell since it contains the control signals required to co-ordinate patterns of gene expression. Most biological entities that are more complex than a virus sometimes or always carry additional genetic material besides that which resides in their chromosomes. In some contexts, such as sequencing the genome of a pathogenic microbe, "genome" is meant to include information stored on this auxiliary material, which is carried in plasmids. In such circumstances then, "genome" describes all of the genes and information on non-coding DNA that have the potential to be present.

Amongst species of bacteria, there is relatively little variation in genome size when compared with the genome sizes of other major groups of life. Genome size is of little relevance when considering the number of functional genes in eukaryotic species. In bacteria however, the strong correlation between genome size and the number of genes makes the size of bacterial genomes an interesting topic for research and discussion. The general trends of bacterial evolution indicate that bacteria started as free-living organisms.

The lifestyles of bacteria play an integral role in their respective genome sizes. Free-living bacteria have the largest genomes out of the three types of bacteria; however, they have fewer pseudogenes than bacteria that have recently acquired pathogenicity. Facultative and recently evolved pathogenic bacteria exhibit a smaller genome size than free-living bacteria, yet they have more pseudogenes than any other form of bacteria. Obligate bacterial symbionts or pathogens have the smallest genomes and the fewest number of pseudogenes of the three groups. The relationship between life-styles of bacteria and genome size raises questions as to the mechanisms of bacterial genome evolution.

Researchers have developed several theories to explain the patterns of genome size evolution amongst bacteria. One theory predicts that bacteria have smaller genomes due to a selective pressure on genome size to ensure

faster replication. The theory is based upon the logical premise that smaller bacterial genomes will take less time to replicate. Subsequently, smaller genomes will be selected preferentially due to enhanced fitness.

Deletional bias selection is but one process involved in evolution. Two other major processes (genetic drift and mutation) can be used to explain the genome sizes of various types of bacteria.

Evidence of a deletional bias is present in the respective genome sizes of free-living bacteria, facultative and recently derived parasites and obligate parasites and symbionts. Free-living bacteria tend to have large population sizes and are subject to more opportunity for gene transfer. As such, selection can effectively operate on free-living bacteria to remove deleterious sequences resulting in a relatively small number of pseudogenes. Continually, further selective pressure is evident as free-living bacteria must produce all gene-products independent of a host. Given that there is sufficient opportunity for gene transfer to occur and there are selective pressures against even slightly deleterious deletions, it is intuitive that free-living bacteria should have the largest bacterial genomes of all bacteria types. Recently formed parasites undergo severe bottlenecks and can rely on host environments to provide gene products. As such, in recently formed and facultative parasites, there is an accumulation of pseudogenes and transposable elements due to a lack of selective pressure against deletions. The population bottlenecks reduce gene transfer and as such, deletional bias ensures the reduction of genome size in parasitic bacteria. Most genes code for a protein product: the gene is transcribed to RNA and this RNA messenger is then translated to the protein product. In addition to genes which encode proteins, there are many genes which encode stable RNAs such as ribosomal RNA and transfer RNA. The number of genes contained within the genome of an organism ranges from around 500 for the bacterium *Mycoplasma genitalium* to over 50,000, predicted to be present in most plants.

Many bacterial genomes have been sequenced including that of *Escherichia coli*, which is not only a model organism but is the organism which is by far the most widely used for gene cloning. The genomes of many pathogenic bacteria have been sequenced including *Mycobacterium tuberculosis* (which

causes TB), *Rickettsia prowazekii* (which causes typhus), *Treponema pallidum* (which causes syphilis), *Vibrio cholera* (which causes cholera) and *Yersinia pestis* (which causes plague).

Comparative analysis of bacterial genomes is proving to be a powerful tool in enhancing our understanding of genome organization in bacteria. The size of bacterial genomes varies considerably. However, for bacteria, genome size correlates well with gene number, which in turn correlates with morphological, physiological or metabolic complexity. Bacteria with small genomes encode a small number of genes and tend to be restricted to growth in relatively few specialized niches; they are often parasites. Bacteria with large genomes encode many genes and have much higher metabolic diversity. So, for example, the 580 kb genome of the intracellular human parasite *Mycoplasma genitalium* is thought to encode just 500 genes, whereas the 7000 kb genome of the nitrogen-fixing bacterium *Mesorhizobium loti*, which colonizes leguminous plants, encodes approximately 6800 genes. The average size of a gene in all bacteria is around 1000 bp; therefore, given the examples mentioned above a simple calculation tells you that bacterial genomes are densely packed with genes, i.e. the 500 genes of *M. genitalium* account for 500 kb of its 580 kb genome.

The fact that bacterial genomes are packed with genes is clearly demonstrated.

The relatively small genome size of bacterial genomes, together with the fact that they contain very little non-coding or repeat DNA and that they do not contain introns, has made bacterial genomes ideal candidates for whole-genome sequencing projects. This is because a relatively quick and successful strategy for whole-genome sequencing involves sequencing randomly generated overlapping fragments of only a few kilobases. These fragments are then assembled into the correct order to yield the finished genome sequence. As bacterial genomes contain very few repeat sequences assembly is relatively routine.

Identification of potential genes within bacterial genomes is also much more reliable, because firstly bacterial genes do not contain introns and secondly the genes are much more closely packed. Potential genes are readily identified by computer analysis.

3-3 Bacterial Chromosome in the Nucleoid

The Nucleoid

The nucleoid (meaning nucleus-like) is an irregularly-shaped region within the cell of a prokaryote that contains all or most of the genetic material. In contrast to the nucleus of a eukaryotic cell, it is not surrounded by a nuclear membrane. The genome of prokaryotic organisms generally is a circular, double-stranded piece of DNA, of which multiple copies may exist at any time. The length of a genome varies widely, but is generally at least a few million base pairs.

The nucleoid can be clearly visualized on an electron micrograph at high magnification, where it is clearly visible against the cytosol. Sometimes even strands of what is thought to be DNA are visible. The nucleoid can also be seen under a light microscope by staining it with the Feulgen stain, which specifically stains DNA. The DNA-intercalating stains DAPI and ethidium bromide are widely used for fluorescence microscopy of nucleoids.

Experimental evidence suggests that the nucleoid is largely composed of about 60% DNA, plus a small amount of RNA and protein. The latter two constituents are likely to be mainly messenger RNA and the transcription factor proteins found regulating the bacterial genome. Proteins helping to maintain the supercoiled structure of the nucleic acid are known as nucleoid proteins or nucleoid-associated proteins, and are distinct from histones of eukaryotic nuclei. In contrast to histones, the DNA-binding proteins of the nucleoid do not form nucleosomes, in which DNA is wrapped around a protein core. Instead, these proteins often use other mechanisms, such as DNA looping, to promote compaction.

The Genophore

A genophore is the DNA of a prokaryote. It is commonly referred to as a prokaryotic chromosome. The term "chromosome" is misleading, because the genophore lacks chromatin. The genophore is compacted through a mechanism known as supercoiling, but a chromosome is additionally compacted through the use of chromatin. The genophore is circular in most prokaryotes, and linear in very few. The circular nature of the genophore allows replication to occur without telomeres. Genophores are generally of a much smaller size than Eukaryotic chromosomes. A genophore can be as

small as 580,073 base pairs (*Mycoplasma genitalium*). Many eukaryotes (such as plants and animals) carry genophores in organelles such as chloroplasts and mitochondria . These organelles are very similar to true prokaryotes.

DNA Supercoiling

DNA supercoiling refers to the over- or under-winding of a DNA strand, and is an expression of the strain on that strand . Supercoiling is important in a number of biological processes, such as compacting DNA. Additionally, certain enzymes such as topoisomerases are able to change DNA topology to facilitate functions such as DNA replication or transcription. Mathematical expressions are used to describe supercoiling by comparing different coiled states to relaxed B-form DNA.

DNA supercoiling is important for DNA packaging within all cells. Because the length of DNA can be thousands of times that of a cell, packaging this genetic material into the cell or nucleus (in eukaryotes) is a difficult feat .Supercoiling of DNA reduces the space and allows for much more DNA to be packaged. In prokaryotes, plectonemic supercoils are predominant, because of the circular chromosome and relatively small amount of genetic material. In eukaryotes, DNA supercoiling exists on many levels of both plectonemic and solenoidal supercoils.

3 - 4 DNA Replication

DNA replication employs a large number of proteins and enzymes, each of which plays a critical role during the process. One of the key players is the enzyme DNA polymerase, which adds nucleotides one by one to the growing DNA chain that are complementary to the template strand. In prokaryotes, three main types of polymerases are known: DNA pol I, DNA pol II, and DNA pol III. DNA pol III is the enzyme required for DNA synthesis; DNA pol I and DNA pol II are primarily required for repair.

There are specific nucleotide sequences called origins of replication where replication begins. In *E. coli*, which has a single origin of replication on its one chromosome. The origin of replication is recognized by certain proteins that bind to this site. An enzyme called helicase unwinds the DNA by breaking the hydrogen bonds between the nitrogenous base pairs. ATP hydrolysis is required for this process. As the DNA opens up, Y-shaped structures called replication forks are formed. Two replication forks at the origin of replication are extended bi-directionally as replication proceeds. Single-strand binding proteins coat the strands of DNA near the replication fork to prevent the single-stranded DNA from winding back into a double helix. DNA polymerase is able to add nucleotides only in the 5' to 3' direction (a new DNA strand can be extended only in this direction). It also requires a free 3'-OH group to which it can add nucleotides by forming a phosphodiester bond between the 3'-OH end and the 5' phosphate of the next nucleotide. This means that it cannot add nucleotides if a free 3'-OH group is not available. Another enzyme, RNA primase, synthesizes an RNA primer that is about five to ten nucleotides long and complementary to the DNA, priming DNA synthesis. A primer provides the free 3'-OH end to start replication. DNA polymerase then extends this RNA primer, adding nucleotides one by one that are complementary to the template strand.

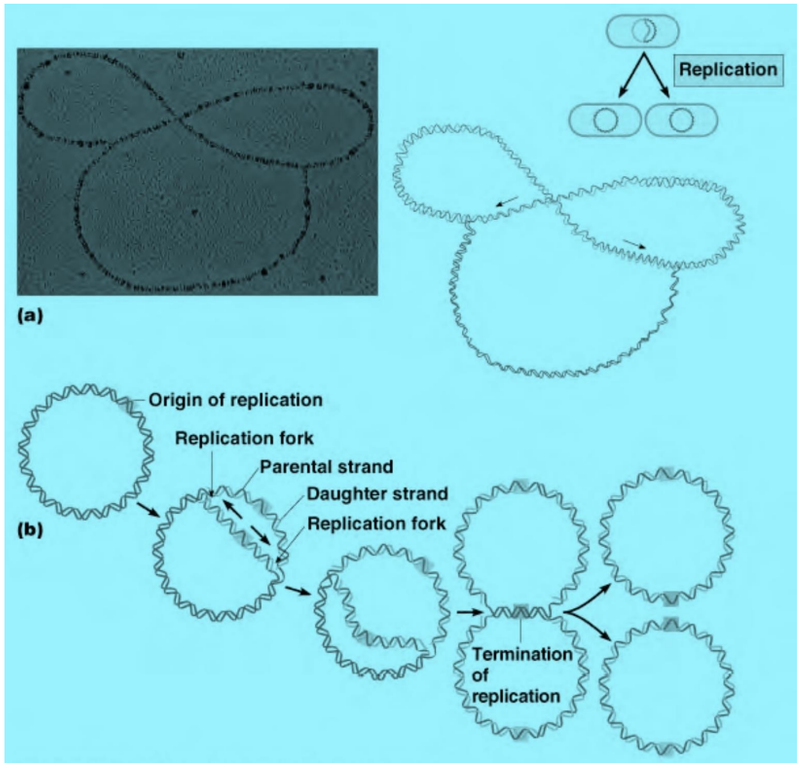


Figure 5: Bidirectionally replication of Bacterial chromosomes : one origin of replication with two replication forks moving in opposite directions [Pearson Education .Inc publishing as Benjamin Cummings].

DNA polymerase can only extend in the 5' to 3' direction, which poses a slight problem at the replication fork. As we know, the DNA double helix is anti-parallel; that is, one strand is in the 5' to 3' direction and the other is oriented in the 3' to 5' direction. One strand (the leading strand), complementary to the 3' to 5' parental DNA strand, is synthesized continuously towards the replication fork because the polymerase can add nucleotides in this direction. The other strand (the lagging strand),

complementary to the 5' to 3' parental DNA, is extended away from the replication fork in small fragments known as Okazaki fragments, each requiring a primer to start the synthesis. Okazaki fragments are named after the Japanese scientist who first discovered them.

The leading strand can be extended by one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments. The overall direction of the lagging strand will be 3' to 5', while that of the leading strand will be 5' to 3'. Topoisomerase prevents the over-winding of the DNA double helix ahead of the replication fork as the DNA is opening up; it does so by causing temporary nicks in the DNA helix and then resealing it. The primers are removed by the exonuclease activity of DNA pol I, while the gaps are filled in by deoxyribonucleotides.

Replication origins of *E. coli* and other bacteria also carry recognition sites for proteins that modulate DnaA interactions within *oriC*. These modulators ensure properly timed activation and inactivation of origin activity during the cell cycle via several modes of action, including bending origin DNA into configurations that promote or inhibit DnaA subassemblies, and by directly blocking DnaA access to low-affinity recognition sites .

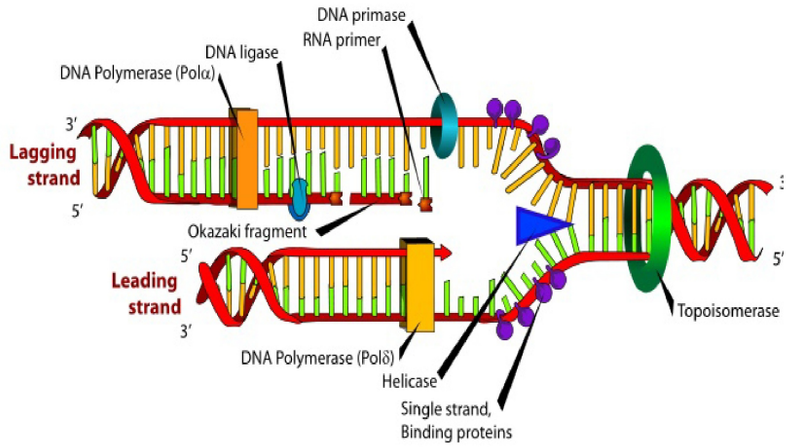


Figure 6 : DNA Replication Events
[http://www.wpclipart.com/medical/anatomy/DNA_replication.png.html]

3-5 Gene Expression

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, via Two main steps Transcription and translation, but in non-protein coding genes such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA.

In genetics, gene expression is the most fundamental level at which the genotype gives rise to the phenotype, i.e. observable trait.

3-5-1 RNA Transcription

Prokaryotic transcription is the process in which messenger RNA transcripts of genetic material in prokaryotes are produced, to be translated for the production of proteins.

Transcription like replication need free 3 end to add the complementary nucleotide ,also the direction of movement and polymerization of RNA polymerase from 5→3 .The transcript sequence(complete gene) start with promoter region where the DNA will opened and the RNA polymerase bind to start transcription .After promoter region another region come called +1 (beginning of transcription) then the coding region followed by Terminator .All the mentioned region represent complete structure for one gene.

Promoter a regulatory nucleotide sequences of DNA (40-60 nts) at the beginning of every gene located upstream (towards the 5' region) of a gene .In prokaryotes, the promoter is recognized by RNA polymerase (δ sigma sub unite). The promoter consists of two short sequences know as -10 box and -35 box positions upstream from the transcription start site. Coding region : Nucleotide sequences which will detriment the genetic code then will translated to Amino acid .It start with ATG triplet initiation codon (AUG in m RNA) .The length of coding region depend on type of produced protein .Terminator :Nucleotide sequences exist after the coding region rich with poly G followed by poly C then poly A .

The RNA polymerase enzyme consist from core (5 units) and another 6th unit called sigma δ , after core binding with sigma it will convert to holoenzyme .sigma subunit play significant role in recognition promoter region then it will released leaving core continue transcription RNA from template.

The Promoter and Terminator are directions for RNA polymerase to indicate the location of the gene to be transcribed. The start and stop codons are directions for the ribosome to indicate where the amino acid information for translation begins and ends . The ORF is the "coding" region of the gene: it begins at the start codon and contains in order all the codons , for all the amino acids in the resulting protein. (3 bases of DNA = 1 codon, each codon indicates one of the 20 amino acids) .

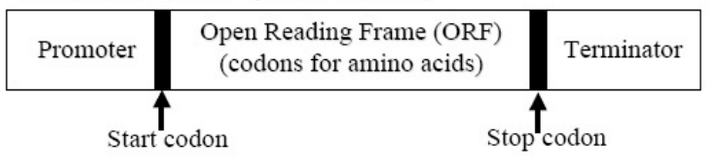


Figure 7 : Gene structure [Tortora, G.J. Microbiology An Introduction]

RNA polymerase binding in bacteria (**in Initiation stage**) involves recognizing core promoter region by sigma factor then binding to -35 box forming closed complex also the α subunit C-terminal domain help in recognizing promoter upstream elements. Usually in *E. coli*, σ^{70} is expressed under normal conditions and recognizes promoters for genes required under normal conditions (house-keeping genes), while σ^{32} recognizes promoters for genes required at high temperatures (heat-shock genes).

After binding to the DNA, the RNA polymerase switches from a closed complex to an open complex at -10 box . This change involves the separation of the DNA strands to form an unwound DNA strand of approximately 13 bp, referred to as the transcription bubble. Ribonucleotides

are base-paired to the template DNA strand, according to Watson-Crick base-pairing interactions.

Transcription elongation involves the further addition of ribonucleotides and the change of the open complex(at -10 box) to the transcriptional complex. RNA Polymerase cannot start forming full length transcripts because of its strong binding to the promoter so it must leave promoter region and further progress . Transcription at this stage primarily results in short RNA fragments of around 9-10 bp . As transcription progresses, ribonucleotides are added to the 3' end of the RNA transcript and the RNAP complex moves along the DNA. The enzyme is highly possessive ,it can add 30 nts \sec .

Although RNAP does not seem to have the 3'exonuclease activity that characterizes the *proofreading* activity found in DNA polymerase, there is evidence of that RNAP will halts) at mismatched base-pairs and correct it.

Notice that all T nitrogen base will replaced by U in the transcribed RNA .

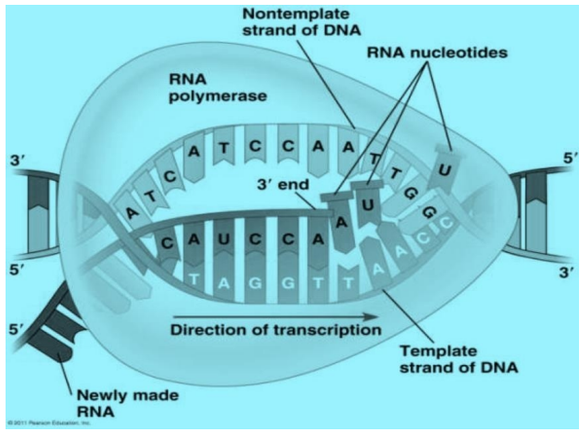


Figure 8 : RNA Polymerase in Transcription Process .

In prokaryotes the termination can be rho-independent or rho-dependent

Intrinsic termination (also called Rho-independent termination) is a mechanism in prokaryotes that causes RNA transcription to be stopped without the aid of rho protein. When Transcription process reached a region called TERMINTATOR or attenuator (a Palindrome region of DNA)this will causes the formation of a "hairpin" structure from the RNA transcription looping and binding upon itself. The bonds between uracil and adenine are very weak. A protein (*nusA*) binds to RNA polymerase and the stem-loop structure tightly enough to cause the polymerase to temporarily stall. The weak Adenine-Uracil bonds lower the energy of destabilization for the RNA-DNA duplex, allowing it to unwind and dissociate from the RNA polymerase.

Rho-dependent termination : Rho factor is a prokaryotic ATP-dependent unwinding enzyme involved in termination transcription. consist from 6 subunits arranged as opened hexameric ring and binds to the transcription terminator sit by moving along the newly forming RNA molecule towards its 3' end and unwinding it from the DNA template thus it will release RNA polymerase from the transcription elongation complex leaving RNA molecules free.

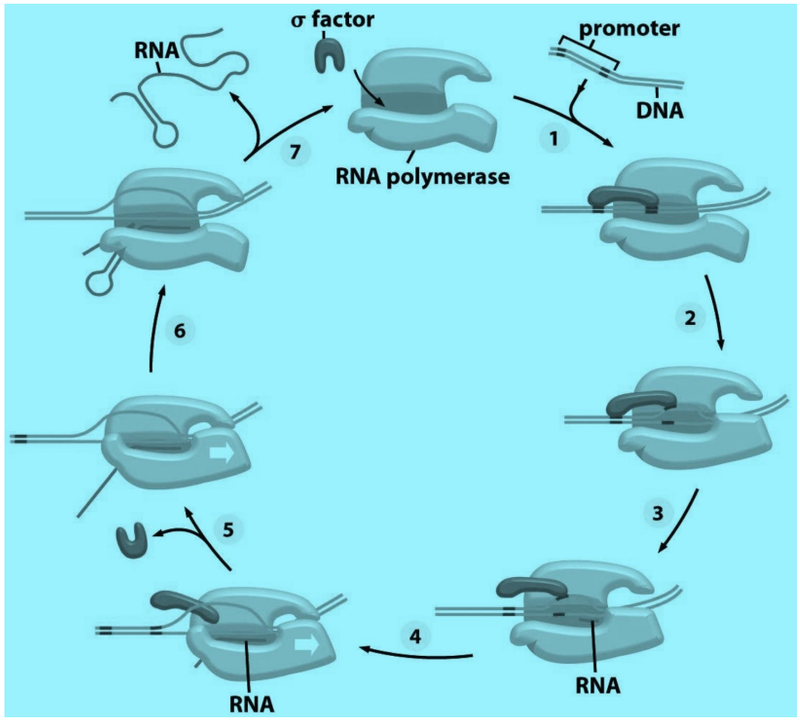


Figure 9 : Transcription Process [Molecular Biology of the cell, Garland Science 2008].

3-5-2 Translation

The translation of genetic information from the 4-letter (A ,C ,G and T) of polynucleotides into the 20-amino acid of proteins is a complex process.
The information in the sequence of a messenger RNA molecule is read out in groups of three nucleotides at a time: each triplet of nucleotides, or codon, specifies (codes for) a single amino acid in a corresponding protein. Since there are 64 possible codons, but only 20 amino acids, there are necessarily many cases in which several codons correspond to the same amino acid. The code is read out by a special class of small RNA molecules, the transfer RNAs (tRNAs). Each type of tRNA becomes attached at one end to a specific amino acid, and displays at its other end a specific sequence of three nucleotides an anticodon that enables it to recognize, through base pairing, a particular codon or subset of codons in mRNA.

mRNA

The mRNA is a sequence of nucleotides, it codes for a sequence of amino acids. To do this, every 3 nucleotide codes for one amino acid. These triplets of nucleotides are called codons. A single mRNA contains many codons. Codons are continuous, non-overlapping and degenerate. Non-overlapping because the 3 nucleotides that consist of one codon never serve as part of another codon. Degenerate because more than one codons code for a given amino acid.

Anticodon: the 3 bases on the "tip" of the tRNA. A single tRNA contains a single anticodon at the "tip" and the corresponding amino acid at the "tail". During translation, codons pair with anticodons so that the correct amino acids can be linked to a given codon.

There are 64 codons, but only 20 amino acids. One amino acid will have multiple codons. The genetic code is said to be degenerate for this reason. Each codon specifies the amino acid (one of 20) to be placed.

- Missense codon: mutated codon that results in a different amino acid.
- Nonsense codon: mutated codon that results in something other than an amino acid. For example, a stop codon.

- Initiation codon (AUG): signals the start of translation. Lies just downstream of the Shine Dalgarno sequence .

- Termination codon (UAG,UGA,UAA): signals the end of translation. Unlike other codons, tRNA are not involved. Instead a protein called "release factor" comes along and terminates translation.

- mRNA stands for messenger RNA. It's the product of transcription and the template for translation. Prokaryotic mRNAs don't have the 5' cap or polyA tail.

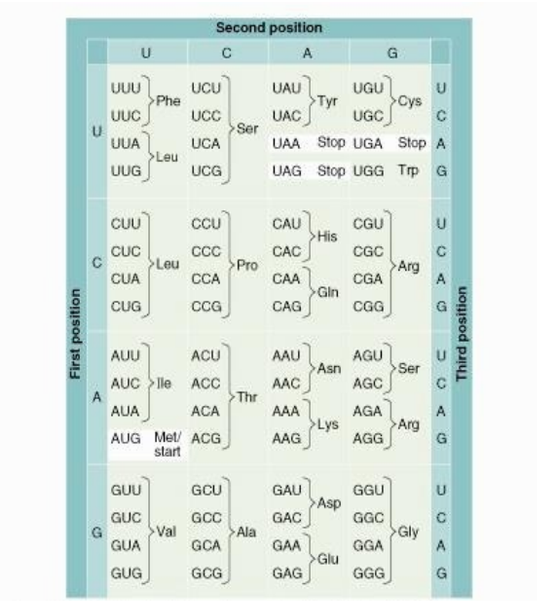


Figure 10 : Use the genetic code chart to decode the amino acid sequence of any mRNA [Tortora, G.J. Microbiology An Introduction 8th, 9th, 10th ed.]

tRNA

- Both tRNA (transfer RNA) and rRNA (ribosomal RNA) are products of transcription. However, they do not serve as the template of translation. tRNA is responsible for bringing in the correct amino acid during translation. rRNA makes up the ribosome, which is the enzyme responsible for translation. tRNA is made of nucleotides, many of which is modified for structural and functional reasons. At the 3' end of the tRNA, the amino acid is attached to the 3'OH via an ester linkage. tRNA structure: clover leaf structure with anticodon at the tip, and the amino acid at the 3' tail.

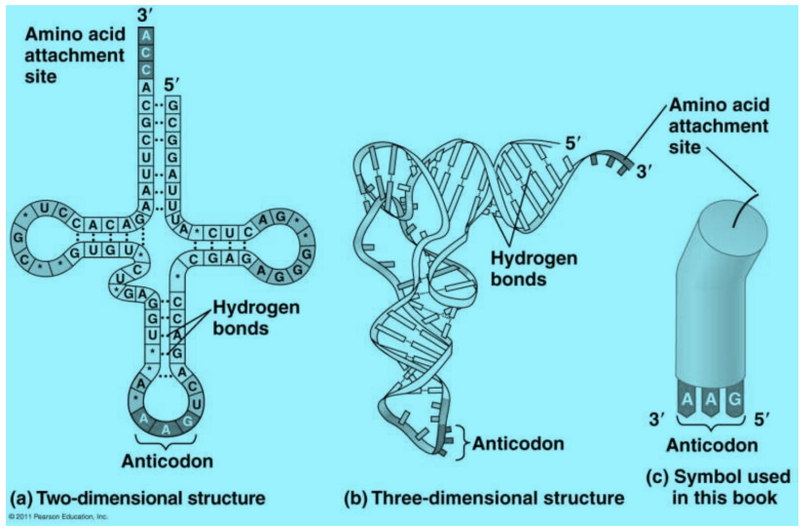


Figure 11: Two and Three dimensional structure of tRNA [Pearson Education inc,2011].

Ribosomes

Ribosomes are the sites of protein synthesis. Bacterial ribosomes differ from those of eukaryotic cells in both size and chemical composition. They are organized in units of 70S.

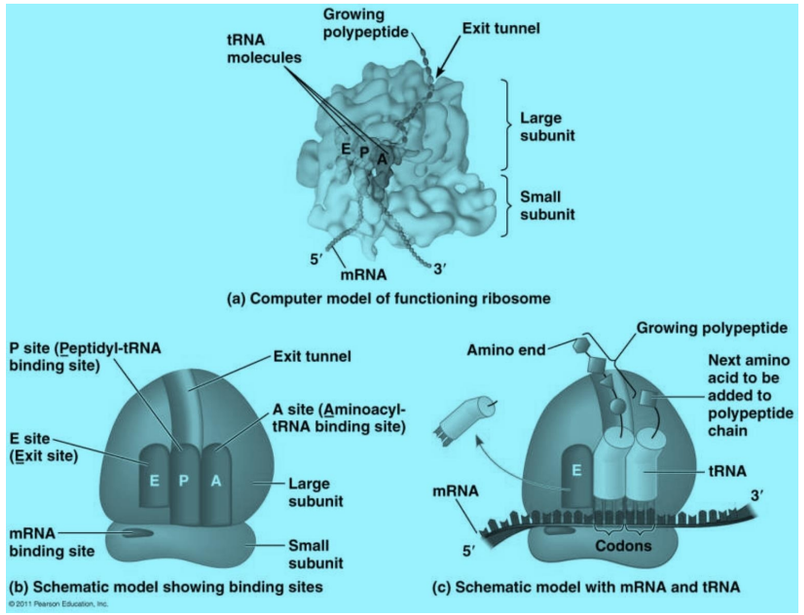
rRNA is made of nucleotides, many of which is modified for structural and functional reasons. rRNA is highly structured because it contains the active site for catalysis. The rRNA of the large ribosomal subunit is responsible for catalyzing peptide bond formation, and can do this even without ribosomal proteins.

To begin translation, you need to form the initiation complex. The initiation complex is basically an assembly of everything needed to begin translation. This includes mRNA, initiator tRNA (fmet), and the ribosome (initiation factors, and GTP aids in the formation of the initiation complex). The initiation complex forms around the initiation codon (AUG), which is just down stream of the Shine-Dalgarno sequence.

Protein in chain **elongation** is made from the N terminus to the C terminus. mRNA codons are read from the 5' to the 3' end.

When a stop codon is encountered, proteins called release factors, bound to GTP, come in and blocks the A site. The peptide chain gets cleaved from the tRNA in the P site. Peptide chain falls off, and then the whole translation complex falls apart.

Aminoacyl-tRNA synthetases attach the correct amino acids to their corresponding tRNAs.



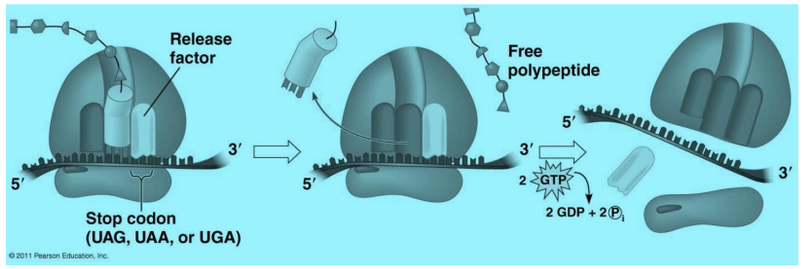


Figure 12: Schematic model of translation process [a- functioning ribosome, b-binding site , c- mRNA and tRNA]. [Pearson Education inc,2011].

4 Mutation

A mutation is a heritable change in the base sequence of the nucleic acid in the genome of an organism. strain carrying such a change is called a mutant. A mutant by definition differs from its parental strain in genotype, the nucleotide sequence of the genome. But in addition, the observable properties of the mutant its phenotype may also be altered relative to the parental strain. One refers to this altered phenotype as a mutant phenotype It is common to refer to a strain isolated from nature as a wild type strain.

The cornerstone of bacterial genetics has been, until recently, the isolation of specific mutants, i.e. strains in which the gene concerned is altered . This alteration shows up as a change in the corresponding characteristics of the organism. It is this change in the observable properties of the organism (the phenotype) that is used to follow the transmission of the gene. The genetic nature of the organism (the genotype) is inferred from the observable characteristics.

Any population of bacteria is far from homogeneous. A culture of *E. coli* (under optimal conditions) will grow from a single cell to its maximum cell density (commonly about 10^{9} cells ml_1) in about 10–15 h, having passed through 30 generations. Within that culture, there will be some variation from one cell to another, which can be due to physiological effects or to genetic changes. Physiological variation means that, due to differences in the environment, growth history or stage of growth at any one moment, the cells may respond differently to some external influence. The key difference between physiological and genetic variation is whether the altered characteristic can be inherited.

Mutations can be either spontaneous or induced. Spontaneous mutations can occur as a result of exposure to natural radiation (cosmic rays, and so on) that alters the structure of bases in the DNA , Also, oxygen radicals can affect DNA structure by chemically modifying DNA.

Mutagens

Mutagen is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level. As many mutations can cause cancer, mutagens are therefore also likely to be carcinogens. Not all mutations are

caused by mutagens: so-called "spontaneous mutations" occur due to spontaneous hydrolysis, errors in DNA replication, repair and recombination.

Mutagens cause changes to the DNA that can affect the transcription and replication of the DNA, which in severe cases can lead to cell death. The mutagen produces mutations in the DNA, and deleterious mutation can result in aberrant, impaired or loss of function for a particular gene, and accumulation of mutations may lead to cancer.

Different mutagens act on the DNA differently. Powerful mutagens may result in chromosomal instability, causing chromosomal breakages and rearrangement of the chromosomes such as translocation, deletion, and inversion. Such mutagens are called clastogens.

Mutagens may also modify the DNA sequence; the changes in nucleic acid sequences by mutations include substitution of nucleotide base-pairs and insertions and deletions of one or more nucleotides in DNA sequences. Although some of these mutations are lethal or cause serious disease, many have minor effects as they do not result in residue changes that have significant effect on the structure and function of the proteins. Many mutations are silent mutations, causing no visible effects at all, either because they occur in non-coding or non-functional sequences, or they do not change the amino-acid sequence due to the redundancy of codons.

Types of mutagens

Mutagens may be of physical, chemical or biological origin. They may act directly on the DNA, causing direct damage to the DNA, and most often result in replication error. Some however may act on the replication mechanism and chromosomal partition. Many mutagens are not mutagenic by themselves, but can form mutagenic metabolites through cellular processes. Such mutagens are called promutagens.

Physical mutagens

- Ionizing radiations such as X-rays, gamma rays and alpha particles may cause DNA breakage and other damages. The most common sources include cobalt-60 and cesium-137.

- Ultraviolet radiations with wavelength above 260 nm are absorbed strongly by bases, producing pyrimidine dimers, which can cause error in replication if left uncorrected.
- Radioactive decay, such as ^{14}C in DNA which decays into nitrogen.

Chemical mutagens

A large number of chemicals may interact directly with DNA.

- Reactive oxygen species (ROS) – These may be superoxide, hydroxyl radicals and hydrogen peroxide, and large number of these highly reactive species are generated by normal cellular processes, for example as a by-products of mitochondrial electron transport, or lipid peroxidation.
- Deaminating agents, for example nitrous acid which can cause transition mutations by converting cytosine to uracil.
- Polycyclic aromatic hydrocarbon (PAH), when activated to diol-epoxides can bind to DNA and form adducts.
- Alkylating agents such as ethylnitrosourea. The compounds transfer methyl or ethyl group to bases or the backbone phosphate groups. Guanine when alkylated may be mispaired with thymine. Some may cause DNA crosslinking and breakages. Nitrosamines are an important group of mutagens found in tobacco, and may also be formed in smoked meats and fish via the interaction of amines in food with nitrites added as preservatives. Other alkylating agents include mustard gas and vinyl chloride.
- Aromatic amines and amides have been associated with carcinogenesis since 1895 . originally used as a pesticide but may also be found in cooked meat, may cause cancer of the bladder, liver, ear, intestine, thyroid and breast.
- Alkaloid from plants, such as those from Vinca species may be converted by metabolic processes into the active mutagen or carcinogen.
- Bromine and some compounds that contain bromine in their chemical structure.
- Sodium azide, an azide salt that is a common reagent in organic synthesis and a component in many car airbag systems

- Psoralen combined with ultraviolet radiation causes DNA cross-linking and hence chromosome breakage.
- Benzene, an industrial solvent and precursor in the production of drugs, plastics, synthetic rubber and dyes.

Base analogs

- Base analog, which can substitute for DNA bases during replication and cause transition mutations.

Intercalating agents

- Intercalating agents, such as ethidium bromide and proflavine, are molecules that may insert between bases in DNA, causing frameshift mutation during replication. Some such as daunorubicin may block transcription and replication, making them highly toxic to proliferating cells.

Metals

Many metals, such as arsenic, cadmium, chromium, nickel and their compounds may be mutagenic, they may however act via a number of different mechanisms. Arsenic, chromium, iron, and nickel may be associated with the production of ROS, and some of these may also alter the fidelity of DNA replication. Nickel may also be linked to DNA hypermethylation and histone deacetylation, while some metals such as cobalt, arsenic, nickel and cadmium may also affect DNA repair processes such as DNA mismatch repair, and base and nucleotide excision repair.

Biological agents

- **Transposon**, a section of DNA that undergoes autonomous fragment relocation/multiplication. Its insertion into chromosomal DNA disrupt functional elements of the genes. Transposons have the ability to move (transpose) from one site to another. The most widely studied transposons carry antibiotic resistance genes and have played a key role in the evolution and spread of antibiotic resistance.

Elements known as Insertion Sequences (IS) have a specific ability to insert into other DNA sequences, thus generating insertion mutations. A substantial proportion of spontaneous mutations may be due to inactivation of genes by insertion of a copy of an IS element rather than by replication errors.

- **Virus ,** Virus DNA may be inserted into the genome and disrupts genetic function. Infectious agents have been suggested to cause cancer as early as 1908 by Vilhelm Ellermann and Oluf Bang.

In summary mutagens can be divided into Three classes based on the ways they cause mutation:

* **Base analogs**. Examples: (2-aminopurine,) (5-bromouracil.) These base analogs are incorporated into DNA where they mispair with other bases. 5BU can pair with adenine and guanine. These result in transition mutations.

* **Base modifications causing mispairs** Examples: ethyl methane sulfonate (EMS). These mutagens modify bases on DNA such that they mispair. EMS alkylate the O6 of guanine, which is highly mutagenic and causes mispairing with thymine ,and show preference for GC to AT transitions. However, they also alkylate bases at many positions with other effects. Nitrous acid and hydroxylamine deaminate cytosine to yield uracil (see deamination above) resulting in transition mutations.

* **Base modifications which destroy pairing**; SOS-dependent mutagens. Examples: UV light, benzo(a)pyrene, aflatoxin B1 (i.e. most carcinogens) These mutagens or their metabolites modify DNA so that no specific pairing is possible; replication cannot proceed past the lesion. Unrepaired AP sites also elicit this response.

* **Frameshift mutagens**. Examples: ICR-191, ethidium bromide. These compounds are often intercalating agents, planar ring compounds that can form "stacking" interactions between nucleotide bases.

Types of mutations

Point mutations

There are many ways in which the structure of the genetic material may change. In point mutations the sequence of the DNA has been altered at a single position. Where this change consists of replacing one nucleotide by another, it is known as a base substitution. The consequence of such a change depends both on the nature of the change and its location. If the change is within the coding region of a gene (i.e. the region which ultimately is translated into protein), it may cause an alteration of the amino acid sequence which may affect the function of the protein. The alteration may of course have little or no effect, either because the changed triplet still codes for the same amino acid or because the new amino acid is sufficiently similar to the original one for the function of the protein to remain unaffected.

For example, the triplet UUA codes for leucine; a single base change in the DNA can give rise to one of nine other codons. Two of the possible changes (CUA , UUG) are completely silent, as the resulting codons still code for leucine. These are known as synonymous codons. Two further changes (AUA and GUA) may well have little effect on the protein since the substituted amino acids (isoleucine and valine respectively) are similar to the original leucine (they are all hydrophobic amino acids). Phenylalanine (UUC or UUU codons) is also hydrophobic but is more likely to cause a significant change in the structure of the protein at that point. The significance of the change to UCA, resulting in the substitution of serine (which is considerably different) for leucine will depend on the role played by that amino acid (and its neighbours) in the overall function or conformation of the protein.

The final two changes (UGA , UAA) are referred to as stop or termination codons (as is a third codon, UAG), since they result in termination of translation; there is normally no tRNA molecule with the corresponding anticodon. The occurrence of such a mutation (also known as a nonsense mutation) will result in the production of a truncated protein; such a protein may or may not be functional, depending on the degree of shortening. The UAG codon was named ‘amber’ which is a literal translation of the German word ‘Bernstein’, the name of one of the investigators who discovered it; subsequently the joke was continued by calling the UAA and UGA codons ‘ochre’ and ‘opal’ respectively, although the latter two names are less commonly used.

Types of point mutations

- A **nucleotide-pair substitution:** replaces one nucleotide and its partner with another pair of nucleotides
- **Silent mutations** have no effect on the amino acid produced by a codon because of redundancy in the genetic code
- **Missense mutations** still code for an amino acid, but not the correct amino acid
- **Nonsense mutations** change an amino acid codon into a stop codon, nearly always leading to a nonfunctional protein

A different kind of mutation still involving a change at a single position, consists of the deletion or addition of a single nucleotide (or of any number other than a multiple of three). This is known as a **frameshift mutation**, since it results in the reading frame being altered for the remainder of the gene. Since the message is read in triplets, with no punctuation marks (the reading frame being determined solely by the translation start codon), an alteration in the reading frame will result in the synthesis of a totally different protein from that point on.

In fact, protein synthesis is likely to be terminated quite soon after the position of the deletion. For most genes the two alternative reading frames are blocked by termination codons, which serves to prevent the production of aberrant proteins by mistakes in translation. Non-translated regions also often contain frequent stop codons, so that the existence of a long region of DNA that does not contain a stop codon in one reading frame (i.e. it has an Open Reading Frame or ORF) .

If a point mutation results in premature termination of translation, it may also affect the expression of other genes downstream in the same operon. This effect is known as polarity, and needs to be considered in genetic analysis.

Point mutations are typically reversible by the process of reversion. A revertant is a strain in which the wild-type (or prototrophic, depending on the strain) phenotype that was lost in the mutant is restored. Revertants can be of two types. In same-site revertants, the mutation that restores activity occurs at the same site at which the original mutation occurred. (If the back mutation is not only at the same site but also leads to the wild-type sequence,

it is called a true revertant.) In second-site revertants, the mutation occurs at a different site in the DNA. Second-site mutations cause restoration of a wild-type phenotype because they function as suppressor mutations.

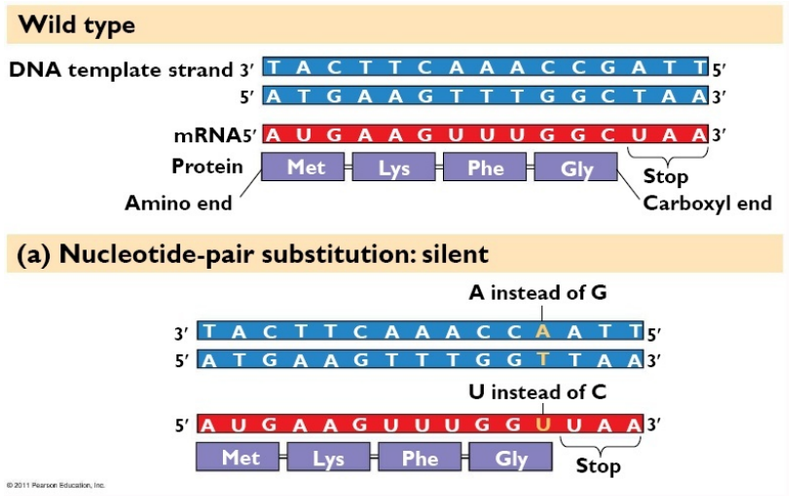


Figure 13: Nucleotide –pair substitution : silent [Pearson Education inc,2011].

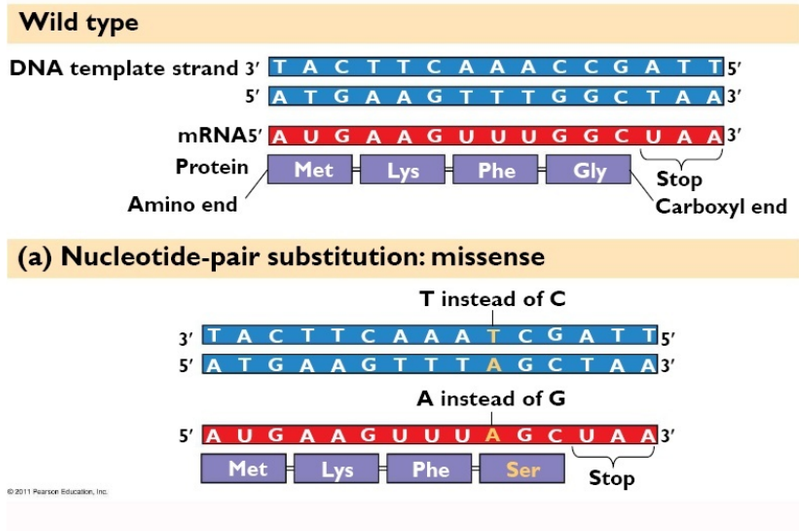


Figure 14: Nucleotide –pair substitution : missense [Pearson Education inc,2011].

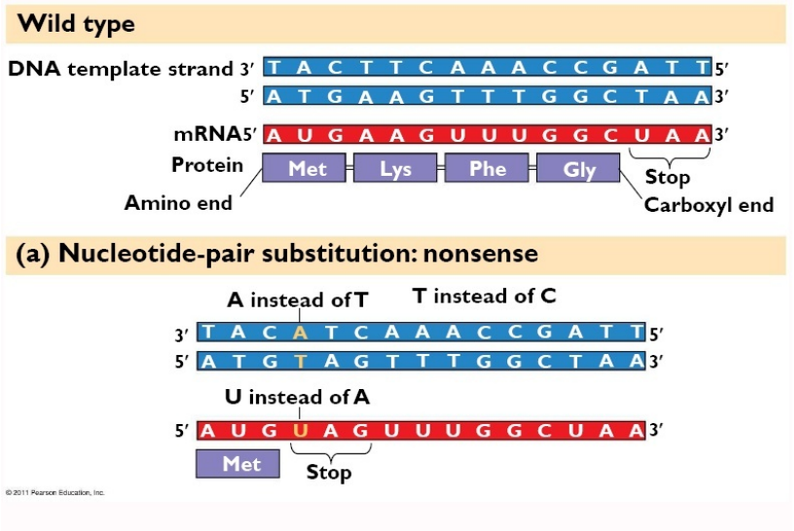


Figure 15: Nucleotide –pair substitution : nonsense [Pearson Education inc,2011].

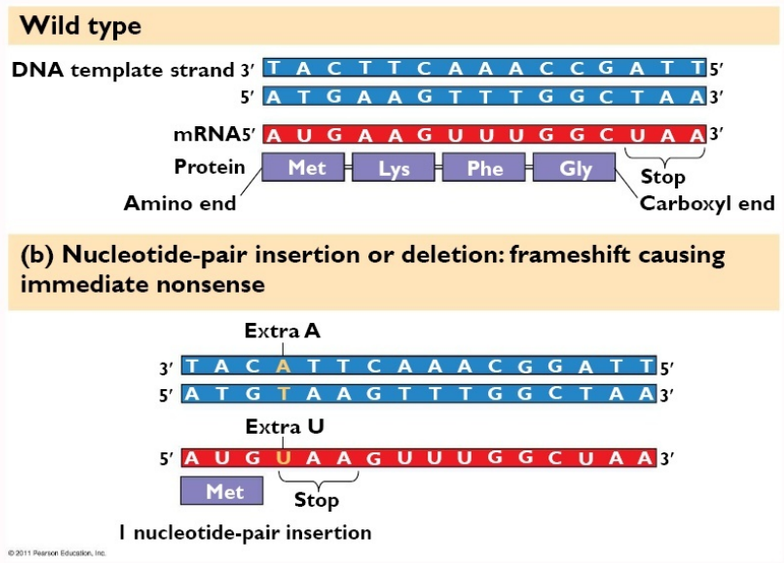


Figure 16: Nucleotide –pair insertion or deletion: immediate nonsense [Pearson Education inc,2011].

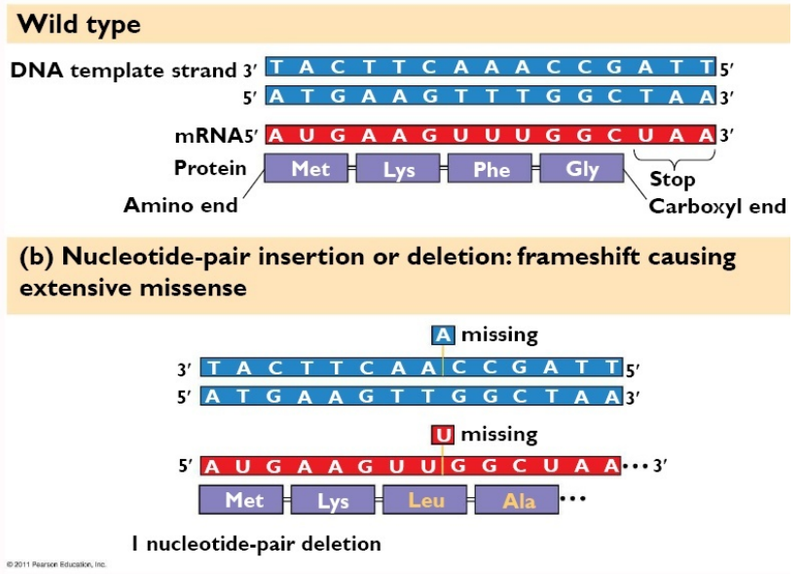


Figure 17: Nucleotide –pair insertion or deletion: extensive missense [Pearson Education inc,2011].

Conditional mutants:

There are many genes that do not affect resistance to antibiotics or bacteriophages, biosynthesis of essential metabolites or utilization of carbon sources. Some of these genes are indispensable and any mutants defective in those activities would die (or fail to grow). Since this includes a wide range of genes that control the essential functions of the cell, such as DNA replication. This means that the gene functions normally under certain conditions while the defect is only apparent when the conditions are changed. One very useful type of conditional mutation confers temperature sensitivity on the relevant function.

DNA modification repair pathways

Most the DNA damage repair pathways operate in similar ways to the mismatch repair pathways . Some share the same components as mismatch repair pathways for instance, many require AP endonuclease Some lesions are repaired by several pathways operating in the same cell. Repair pathways that remove DNA modifications have Three basic mechanisms: direct repair, base-excision repair and nucleotide-excision repair.

A. **Direct chemical reversal.**

1. Photolyase, AKA photo-reactivation. UV light induces the formation of pyrimidine dimers between adjacent C or T bases in DNA. The photolyase enzyme break the cyclobutane dipyrimidine bond. To do so, the enzyme must absorb visible light, hence the name photo-reactivation. *E. coli* and the yeast *Saccharomyces cerevisiae* have such an enzyme.

2. Methyltransferase. The methyl groups from mutagenic O6-methylguanine (O6-MeG is particularly mutagenic) and O4-methylthymine can be removed directly by this enzyme. The methyl groups are transferred to the protein itself at a cysteine residue--the protein is consumed in the process, it cannot be demethylated. This methyltransferase is the product of the *ada* gene of *E. coli* and is part of the "adaptive response", repair functions induced by treatment with methylating agents.

B. base-excision:

Glycosylase + AP endonuclease. Many modified bases are recognized by specific N-glycosylases that cleave the modified base. The resulting AP site is repaired by AP endonucleases. Examples in *E. coli*: hypoxanthine-DNA glycosylase, 3-methyladenine glycosylase . formamidopyrimidine glycosylase, hydroxymethyl glycosylase. T4 phage and bacterium *M. luteus* encode a cyclobutane pyrimidine dimer glycosylase/endonuclease. Such enzymes appear to be absent in other bacteria and mammals. Several mammalian glycosylase/endonucleases have been identified that remove oxidized bases caused by exposure to oxidizing agents or ionizing radiation.

c. Nucleotide excision repair.

In this form of DNA repair, the damaged bases are removed from DNA as an oligonucleotide and the resulting gap is repaired by resynthesis. This pathway is used to remove many bulky adducts in DNA, including cross-links and UV-induced pyrimidine dimers. In *E. coli* 4 genes are required: *uvrA*, *uvrB*, *uvrC* and *uvrD*. A UvrAB complex binds to DNA on both sides of lesion; UvrC then binds and two incisions on each side of the lesion are made. Polymerase I displaces the 12-13 base oligonucleotide and fills the gap; UvrD helicase may aid in dissociation of the UvrABC complex from the DNA or in the dissociation of the oligonucleotide from the DNA. Excision repair pathways have been identified in many organisms.

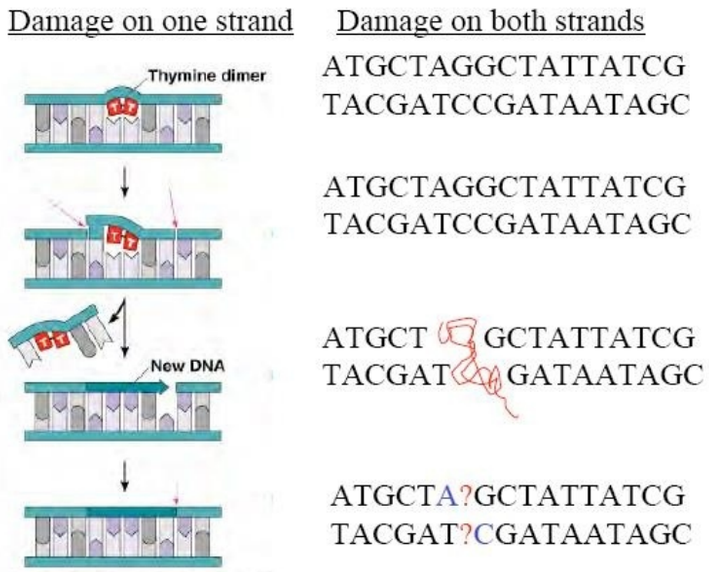


Figure 17: Nucleotide excision repair = enzymes that function to cut out and replace DNA damage [Pearson Education inc,2004]

In both *E. coli* , yeast and human cells, there seems to be a mechanism that targets excision repair to lesions on transcribed strand (i. e. the template strand for transcription) of the damaged gene.

Some special DNA repair pathways: recombinational repair and lesion by pass repair. These are pathways that do not remove DNA adducts but overcome their block to DNA replication. Especially important for DNA damage that occurs in single-stranded regions of the replication fork that cannot be repaired by excision. Recombinational repair also functions to repair ds breaks in DNA which are caused by ionizing radiation (e. g. X-rays, gamma rays) or which may accompany repair reactions.

SOS response : a number of genes involved with DNA repair and related functions are induced by the presence of unrepaired DNA. An alternative

strategy, when faced with overwhelming levels of DNA damage preventing normal replication, is to temporarily modify or abolish the specificity of the DNA polymerase. This enables it to continue making a new DNA stranddespite the absence of an accurately paired 30 end. Although the new strand is produced, it is obviously likely to contain many mistakes and the process is therefore described as 'error-prone'. SOS repair is the cause of mutations arising from ultraviolet irradiation.

Mutation Rate

Mutation rate is a probability that gene will mutate when cell divides . A mutation rate has evolved in cells that is very low yet detectable. This allows organisms to balance the need for genetic stability with that for evolutionary improvement.

The fact that organisms as phylogenetically disparate as hyper thermophilic Archaea and *Escherichia coli* have about the same mutation rate might make one believe that evolutionary pressure has selected for organisms with the lowest possible mutation rates. However, this is not so. The mutation rate in an organismis subject to change. For example, mutants of some organisms have been selected in the laboratory that are hyperaccurate in DNA replication and repair.

However, in these strains, the improved proofreading repair mechanisms has a significant metabolic cost; thus, a hyperaccurate mutant might actually be at a disadvantage in its natural environment. On the other hand, some organisms seem to benefit from a hyperaccurate phenotype that enables them to occupy particular niches in nature. A good example is the bacterium *Deinococcus radiodurans* . This organism is 20 times more resistant to UV radiation and 200 times more resistant to ionizing radiation than *E. coli.*

This resistance, dependent in part upon redundant DNA repair systems and on a mechanism for exporting damaged nucleotides, allows the organism to survive in environments where other organisms cannot, such as near concentrated sources of radiation or on the surfaces of dust particles exposed to intense sunlight.

In contrast to the hyperaccurate phenotype, some organisms actually benefit from an increased mutation rate. DNA repair systems are themselves encoded by genes and thus subject to mutation. For example, the protein subunit of DNA polymerase III involved in proofreading is encoded by the gene *dna*Q. Certain mutations in this gene lead to mutant strains that are still

viable but have an increased rate of mutation. Such a strain is said to be hypermutable or a mutator. Mutations leading to a mutator phenotype are known in severalother DNA repair systems as well. The mutator phenotype is apparently selected for in complex and changing environments, since strains of bacteria with mutator phenotypes appear to be more abundant under these conditions.

Presumably, whatever disadvantage an increased mutation rate may have in such environments is offset by the ability to generate greater numbers of useful mutations. These mutations ultimately increase evolutionary fitness of the population and make the organism more successful in its ecological niche.

A mutator phenotype may be induced in wild-type strains by stress situations. For instance, the SOS response induces error-prone repair. Therefore, when the SOS response is activated, the mutation rate increases. While in some cases this may be a necessary by-product of DNA repair, in other cases the increased mutation rate may itself be of value. That is, the organism may generate adaptive mutations. Adaptive mutations lead to a phenotype in the mutant that allows it to survive a particular stress. Selection is obviously strong under these conditions because, depending on the stress condition, the only option to potential survival from an increased mutation rate may be death.

5 Extra chromosomal Agents

5-1 Plasmid

Plasmids are nonessential extrachromosomal elements that control their own replication. They are found mainly in bacteria and in some eukaryotic microbes such as yeast and algae. Plasmids are normally circular double stranded DNA molecules that range in size from as little as 1 kb to over 100 kb although linear plasmids have been observed in *Borrelia burgdorferi*, the causative agent of Lyme disease. Relative to the size of the bacterial chromosome plasmids are very small.

The number of copies of a plasmid in each cell is tightly controlled with a general rule that small plasmids tend to have a high copy number, sometimes over 100 copies per cell, whereas larger plasmids may be present in one or a few copies per cell.

Certain classes of plasmids are capable of horizontal transmission from one bacterium to another. In most cases the recipient bacterium does not have to be the same species as the donor bacterium and can, in fact, be distantly related. By definition, the coding potential of a plasmid is nonessential for the bacterial life cycle but plasmids frequently encode functions that give the bacteria a selective advantage in certain environments, the most infamous example being the class of genes which encode antibiotic resistance.

The fact that plasmids can confer resistance to antibiotics and that horizontal transmission can occur between distantly related bacteria has major implications for medicine, since it has allowed rapid spread of antibiotic resistance throughout the microbial world. plasmids have been absolutely key to the development of gene cloning.

The term plasmid was first introduced by the American molecular biologist Joshua Lederberg in 1952, originally to describe any bacterial genetic material that exists in an extrachromosomal state for at least part of its replication cycle. Later, to distinguish it from viruses, the definition was narrowed to genetic elements that exists exclusively or predominantly outside of the chromosome and can replicate autonomously.

In order for the plasmids to replicate independently within a cell, they must possess a stretch of DNA that can act as an origin of replication. The self-

replicating unit, in this case the plasmid, is called a replicon. A typical bacterial replicon may consist of a number of elements, such as the gene for plasmid-specific replication initiation protein (Rep), repeating units called iterons, DnaA boxes, and an adjacent AT-rich region. Smaller plasmids make use of the host replicative enzymes to make copies of themselves, while larger plasmids may carry genes specific for the replication of those plasmids. A few types of plasmids are also capable of inserting into the host chromosome, and these integrative plasmids are sometimes referred to as episomes in prokaryotes.

Plasmids almost always carry at least one gene. Many of the genes carried by the plasmid are beneficial for the host cells, for example enabling the host cell to survive in an environment that would otherwise be lethal or restrictive for growth. Some of these genes encode traits for antibiotic resistance or resistance to heavy metal, while others may produce virulence factors that enable a bacterium to colonize a host and overcome its defences, or have specific metabolic functions that allow the bacterium to utilize a particular nutrient, including the ability or to degrade recalcitrant or toxic organic compounds. Plasmids can also provide bacteria with the ability to fix nitrogen. Some plasmids, however, have no observable effect on the phenotype of the host cell or its benefit to the host cells cannot be determined, and these plasmids are called cryptic plasmids.

Classifications and types

Plasmids may be classified in a number of ways. Plasmids can be broadly classified into conjugative plasmids and non-conjugative plasmids. Conjugative plasmids contain a set of transfer or *tra* genes which promote sexual conjugation between different cells. In the complex process of conjugation, plasmid may be transferred from one bacterium to another via sex pili encoded by some of the *tra* genes . Non-conjugative plasmids are incapable of initiating conjugation, hence they can be transferred only with the assistance of conjugative plasmids. An intermediate class of plasmids are mobilizable, and carry only a subset of the genes required for transfer. They can parasitize a conjugative plasmid, transferring at high frequency only in its presence.

Plasmids can also be classified into incompatibility group. A microbe can harbour different types of plasmids, however, different plasmids can only

exist in a single bacterial cell if they are compatible. If two plasmids are not compatible, one or the other will be rapidly lost from the cell. Different plasmids may therefore be assigned to different incompatibility group depending on whether they can coexist together. Incompatible plasmids normally share the same replication or partition mechanisms.

Another way to classify plasmids is by function. There are five main classes:

- Fertility F-plasmids, which contain *tra* genes. They are capable of conjugation and result in the expression of sex pilli.
- Resistance (R) plasmids, which contain genes that provide resistance against antibiotics or poisons. Historically known as R-factors, before the nature of plasmids was understood.
- Col plasmids, which contain genes that code for bacteriocins, proteins that can kill other bacteria.
- Degradative plasmids, which enable the digestion of unusual substances, e.g. toluene and salicylic acid.
- Virulence plasmids, which turn the bacterium into a pathogen.

Plasmids can belong to more than one of these functional groups.

R Plasmid:

The most widely studied plasmid-borne characteristic is that of drug resistance.

R plasmids were first discovered in Japan in strains of enteric bacteria that had acquired resistance to a number of antibiotics (multiple resistance) and have since been found throughout the world. The emergence of bacteria resistant to several antibiotics is of considerable medical significance and was correlated with the increasing use of antibiotics for the treatment of infectious diseases.

Many bacteria can become resistant to antibiotics by acquisition of a plasmid, although plasmid-borne resistance to some drugs such as nalidixic acid and rifampicin does not seem to occur. (In those cases, resistance usually occurs by mutation of the gene that codes for the target protein).

The antibiotic resistance genes themselves are many and varied, ranging from plasmid-encoded betalactamases which destroy penicillins to membrane proteins which reduce the intracellular accumulation of tetracycline. The ability of plasmids to be transferred from one bacterium to

another, even sometimes between very different species , has contributed greatly to the widespread dissemination of antibiotic resistance genes. Bacteria can become resistant to a number of separate antibiotics, either by the acquisition of several independent plasmids or through acquiring a single plasmid with many resistance determinants on it.

Transposons are thought to have played a major role in the development of drug resistance plasmids, by promoting the movement of the genes responsible between different plasmids or from the chromosome of a naturally resistant organism onto a plasmid.

It should be appreciated that other mechanisms of antibiotic resistance also occur and that such resistance is not always due to plasmids: indeed many of the bacteria that are currently causing problems of hospital cross-infection are either inherently resistant or owe their antibiotic resistance to chromosomal genes.

Plasmid RI00 can betransferred between enteric bacteria of the genera *Escherichia, Klebsiella, Proteus, Salmonella,* and *Shigella,* but doesnot transfer to the non enteric gram-negative bacterium *Pseudomonas.*

Different R plasmids with genes for resistance to most antibiotics are known. Many drug-resistant elements on R plasmids, such as those on RI00, are also transposable elements and this, plus the fact that many of these plasmids are conjugative, have made them a serious threat to traditional antibiotic therapies.

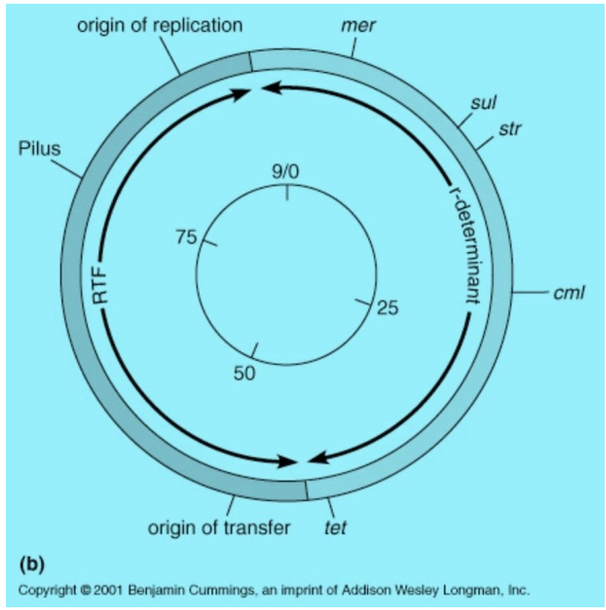


Figure 18 : Genetic map of the resistance plasmid R100 [Benjamin cumming ,an imprint of Addison Wesley Longman,Inc,2001].

Colicins and bacteriocins

Another property conferred by some plasmids that has been widely studied is the ability to produce a protein which has an antimicrobial action, usually against only closely-related organisms. One group of such proteins, produced by strains of *E. coli*, are capable of killing other *E. coli* strains, and are hence referred to as colicins, and the strains that produce them are colicinogenic. The colicin gene is carried on a col plasmid , together with a second gene that confers immunity to the action of the colicin, thus protecting the cell against the lethal effects of its own product. However, colicin E2 is a DNA endonuclease that can cleave DNA, and colicin E3 is a nuclease that cuts at a specific site in 16S rRNA and inactivates ribosomes. Col plasmids can be either conjugative or nonconjugative.

The bacteriocins or bacteriocin-like agents of gram positive bacteria are quite different from the colicins but are also often encoded by plasmids; some even have commercial value. For instance, lactic acid bacteria produce the bacteriocin Nisin A, which strongly inhibits the growth of a wide range of gram-positive bacteria and is used as a preservative in the food industry.

Virulence determinants

In some bacterial species toxin genes are carried on plasmids rather than phages. For example, some strains of *E. coli* are capable of causing a disease that resembles cholera (although milder). These strains produce a toxin known as LT (labile toxin – to distinguish it from a different, heat-stable, toxin known as ST). The LT toxin is closely related to the cholera toxin, but whereas the gene in *Vibrio cholerae* is carried by a prophage, the LT gene in *E. coli* is found on a plasmid.
Plasmids can also carry other types of genes that are necessary for (or enhance) virulence. One of the most dramatic examples of this is the 70-kb virulence plasmid of Yersinia species. This plasmid which is found in species of Yersinia (including *Yersinia pestis*, the causative organism of plague) .

Plasmids in plant-associated bacteria

A different type of pathogenicity is seen with the plant pathogen *Agrobacterium tumefaciens*, which causes a tumour-like growth known as a crown gall in some plants. Again, it is only strains that carry a particular type of plasmid (known as a Ti plasmid, for Tumour Inducing) that are pathogenic; in this case however, pathogenicity is associated with the transfer of a specific part of the plasmid DNA itself into the plant cells. This phenomenon has additional importance because of its application to the genetic manipulation of plant cells .
Members of the genus Rhizobium also 'infect' plants, although in this case the relationship is symbiotic rather than pathogenic. These bacteria form nodules on the roots of leguminous plants. Under these conditions the bacteria are able to fix nitrogen and supply the plant with a usable source of reduced nitrogen, a process of considerable ecological and agricultural importance. The genes necessary for both nodulation and nitrogen fixation are carried by plasmids.

Metabolic activities

Plasmids are capable of expanding the host cell's range of metabolic activities in a variety of other ways. For example, a plasmid that carries genes for the fermentation of lactose, if introduced into a lactose non-fermenting strain, will convert it to one that is able to utilize lactose. Such plasmids can cause problems in diagnostic laboratories where organisms are often identified on the basis of a limited set of biochemical characteristics. Commonly the potentially pathogenic Salmonella genus is differentiated from the (usually) non-pathogenic *E. coli* species primarily

Plasmid replication and control

Many plasmids are replicated as double stranded circular molecules. The overall picture with such plasmids is basically similar to that of the chromosome, in that replication starts at a fixed point known as oriV (the vegetative origin, to distinguish it from the point at which conjugative transfer is initiated, oriT), and proceeds from this point, either in one direction or in both directions simultaneously until the whole circle is copied.
However there are some aspects of replication that differ from that of the chromosome, especially for the multicopy plasmids. Two examples that have been studied intensively are ColE1 and R100. Other plasmids with quite different modes of replication are dealt with later on.

5-2 Transposable elements

Mobile genetic elements are segmentso f DNA able to move from site to site in the genome, or between genomes in the same cell. Found in both prokaryotes and eukaryotes, they are a diverse group differing in structure, mechanism of mobilization, distribution, freedom of movement .
Those elements which cannot replicate outside the host genome, but which control their own mobility within it, are termed transposable elements . The transposable elements are the largest elements and demonstrate great diversity in structure.

5-2-1 Insertion sequences

Insertion sequences (ISs) constitute an important component of most bacterial genomes. Insertion of a DNA fragment into a gene will usually result in the inactivation of that gene, and it is by the loss of that function that such events were initially recognized. A number of genetic elements, including some phages and plasmids , can be inserted into the bacterial chromosome.
The simplest of these genetic elements are known as Insertion Sequences (IS).

Bacterial transposable elements were discovered as a consequence of their ability to cause unstable but strongly polar mutations in *E. coli.* Hybridization analysis showed that small family of inserted DNA sequences was responsible for many of served mutations, and these were termed insertions equences (IS elements). About 100 different IS elements have been identified, the majority in enteric bacteria and their plasmids.

Structure of insertion sequences

There are many IS elements known. They differ in size and other details, but the overall structure of most such elements is similar. IS is 768 bases long but many other IS elements are longer (usually 1300–1500 bases). The central region of an IS element codes for a protein (known as a transposase) which is necessary for the movement of the element from one site to another. At the ends of the insertion sequence are almost perfect inverted repeat (IR) sequences, which in IS1 consist of 23 nucleotides. A minority of elements, such as IS900 from *Mycobacterium paratuberculosis* do not have inverted repeat ends. It must be stressed that reference to an inverted repeat of a DNA sequence does NOT mean that the sequence on an individual strand is repeated backwards, but that the sequence from left to right on the 'top' strand is repeated from right to left on the 'bottom' strand so that reading either copy of the IR in the 50 to 30 direction will result in the same sequence of bases. Since DNA sequences are often presented as just one of the two strands, an inverted repeat of the sequence CAT will appear as ATG. In addition to the inverted repeats, inspection of a DNA region containing an insertion sequence usually shows a further short sequence that is duplicated – but this sequence is repeated in the same orientation and is therefore referred to as a direct repeat (DR). This is not part of the IS, but arises from duplication of the DNA at the insertion site and therefore different copies of

IS1 will have different target sequence repeats depending on the point of insertion. Transposition of IS1 generates rather long direct repeats (nine base pairs). With other insertion sequences, the direct repeats are commonly as short as two to three base pairs.

Insertion sequences have been identified in most bacterial genera, although the presence and the number of copies of any one element often varies from strain to strain.

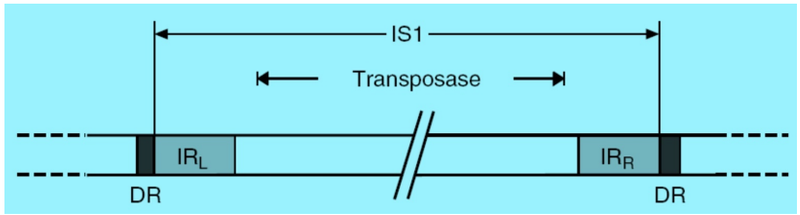


Figure 19 : Structure of the insertion sequence ISl. DR, direct repeat (duplicated target sequence); IR, inverted repeats [Dale and Park,2004].

5-2-2 Transposons

Transposons were originally defined as bacterial transposable elements which carried not only those genes required for transposition, but also nonessential genes, e.g. for antibiotic resistance.
When resistance plasmids were first discovered, there was much speculation as to how a single element could have evolved to carry a number of different antibiotic resistance genes and in particular how apparently related plasmids could have different combinations of such genes (or, conversely, how otherwise dissimilar plasmids could carry related resistance genes). It was assumed that a basic plasmid, having the ability to replicate independently but not carrying any other information, had somehow picked up a resistance gene from the chromosome of a resistant host strain. Transfer of this plasmid to an otherwise sensitive strain then produces a selective advantage for that strain, and therefore indirectly a selective advantage for this 'new' plasmid. As the plasmid moves from one organism to another it has the opportunity to

acquire additional resistance genes, thus giving rise to a family of plasmids containing different combinations of resistance genes.
Since this model implies that unrelated plasmids could pick up the same gene independently, this would explain the widespread distribution of certain resistance genes, notably a type of *b*-lactamase (the enzyme that destroys penicillin and hence confers resistance to penicillins). This particular enzyme, the TEM *b*-lactamase, is the commonest type amongst plasmids in the Enterobacteriaceae and is also present in many members of the genus Pseudomonas. The same gene has also been found in connection with plasmid-mediated penicillin resistance in species as diverse as *Haemophilus influenzae* and *Neisseria gonorrhoeae*.
The reason behind the ubiquity of the TEM *b*-lactamase became apparent from the discovery that this gene could move (transpose) from one plasmid to another.
When a transposable element becomes inserted into a target DNA, a short sequence in the target at the site of integration is duplicated during the insertion process. The duplication arises because single-stranded DNA breaks are generated by the transposase. The transposon is then attached to the single-stranded ends that have been generated, and repair of the single-strand portions results in the duplication.
If the duplicated DNA is sufficient to have duplicated an entire gene or group of genes, the organism will contain multiple copies of these particular genes. Such gene-duplication events are thought to fuel microbial evolution. This is because mutations occurring in one copyof the gene(s) do not affect the other copy; the function of the faulty protein is still "covered" by the product of the unmutated duplicate gene. However, beneficial mutations in one copy can lead to production of a protein that has superior properties to the normal protein and thereby increase fitness. In this way, evolution can "experiment" with one copy of the gene while the identical copy provides the necessary backup function. Genomic analyses have revealed numerous examples of protein-encoding genes that were clearly derived from gene duplication.

Structure of Transposons

The structure of a simple transposon, Tn3 consists of about 5000 base pairs and has a short (38 bp) inverted repeat sequence at each end. It is therefore analogous to an insertion sequence, the distinction being that a transposon carries an identifiable genetic marker – in this case the ampicillin resistance gene (bla, *b*-lactamase). Tn3 codes for two other proteins as well: a

transposase (TnpA), and TnpR, a bifunctional protein that acts as a repressor and is also responsible for one stage of transposition known as resolution (this is explained more fully later on). As with the insertion sequences, there is a short direct repeat at either end of the transposon (five base pairs in the case of Tn3).

Figure 20: Structure of the transposon Tn3. DR, five-base pair direct repeat (target duplication); IR, 38-base pair inverted repeats; res, resolution site; tnpA, transposase; tnpR, resolvase; bla, *b*-lactamase (ampicillin resistance) [Dale and Park,2004].

Some transposable elements have a more complex structure than Tn3. These composite transposons consist of two copies of an insertion sequence on either side of a set of resistance genes. For example the tetracycline resistance transposon Tn10, which is about 9300 bp in length, consists of a central region carrying the resistance determinants flanked by two copies of the IS10 insertion sequence in opposite orientations. IS10 itself is about 1300 bp long with 23-bp inverted repeat ends and contains a transposase gene.
Composite transposons may have their flanking IS regions in inverted orientation or as direct repeats. For example Tn10 and Tn5 both have inverted repeats of an IS (IS10 and IS50 respectively) at their ends, while Tn9 has direct repeats of IS1. The transposition behaviour of such composite elements can be quite complex; the insertion sequences themselves may transpose independently or transposition of the entire region may occur. Furthermore, recombination between the IS elements can occur, leading to deletion or inversion of the region separating them .
Even more complex arrangements can occur. For example, Tn4 appears to be related to Tn21 but contains a complete copy of Tn3 within it. The ampicillin resistance gene of Tn4 can thus be transposed as part of the complete Tn4 transposon or by transposition of the Tn3 element.

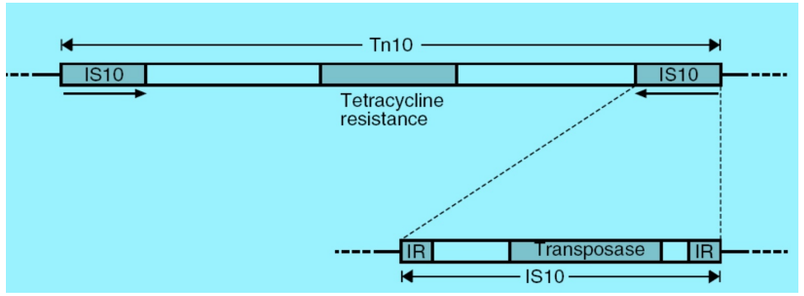


Figure 21:Structure of a composite transposon Tn10[Dale and Park,2004].

The composite transposons such as Tn10 are also known as class I transposons, while transposons such as Tn3, flanked by inverted repeats rather than IS elements, are referred to as class II transposons.

5-2-3 Integrons

Integrons are independent, mobile elements that encode genes for protein functions, and encode additional DNA to guarantee the integron's expression and integration into the bacterial genome.
The integron consists of a regulatory region, an open reading frame encoding a site-specific recombinase and a target site for the recombinase, *attI,* which can integrate a variety of antibiotic resistance gene cassettes carrying a 59-base element, which is also recognized by the recombinase.
Integrons effectively generate widespread antibiotic resistance by donating antibiotic resistance genes to any strain of bacteria. Ironically, it is widely believed that integrons evolved only recently in response to antibiotic selection pressure. In other words, the use of antibiotics advanced the widespread occurrence of antibiotic resistance.
Extremely complex and large transposons can also be built up by insertion of additional genes within an existing transposon. Many large transposons have been identified which are related to Tn21 which has a structure analogous to class II transposons such as Tn3: it has inverted repeats (38 bp) at each end and carries genes for transposition functions . Tn21 may have

developed from a smaller transposon (such as Tn2613) by acquisition of additional genes.
Tn2603 and Tn1696 (and a family of other transposons) are also very similar to Tn21 but contain additional resistance genes.
It is now known that the transposons in the Tn21 family have acquired resistance genes by a specific mechanism. Each individual gene has been inserted separately, as a gene cassette which contains a single gene and a recombination site. Tn21 contains a site known as an integron into which such gene cassettes can be inserted by site-specific recombination. The integron region in Tn21 also contains a gene coding for an integrase which is responsible for the site-specific recombination . After insertion of a gene cassette into the integron, the recombination site remains available for insertion of a further gene cassette, enabling the build-up of an array of several cassettes within the integron. A further twist to the story is that the gene cassettes do not normally contain a promoter. However there is a promoter region within the integron itself, upstream from the insertion site, so each of the gene cassettes is transcribed from the integron promoter. Integrons are thus a naturally-occurring analogy to the
Integrons are not only found in the Tn21 family of transposons, several classes of recognizably distinct integrons have been found in other transposons as well as in plasmids which do not carry functional transposons. They therefore represent a significant additional mechanism for the evolution of bacterial plasmids and the spread of antibiotic resistance.
Integrons were discovered as the result of systematic examination of resistance plasmids and transposons that carry overlapping sets of resistance genes but that otherwise are unrelated.

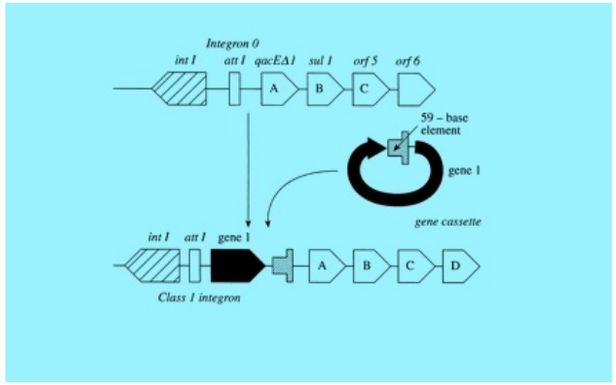


Figure 22 : Class 1 integron structure (Bennett,2015)

Mechanisms of transposition

Replicative transposition

In considering mechanisms of transposition, it is not necessary to distinguish between insertion sequences and transposons as the same mechanisms apply to both types of transposable elements. Transposons such as Tn3 transpose by a replicative mechanism: a copy of the element is inserted at a different site (on the chromosome or on a plasmid) while the original copy is retained. The outcome of this stage is the formation of a larger plasmid known as a cointegrate, which consists of the complete sequence of both plasmids fused together, but now with two copies of the transposon, in the same orientation. With some naturally-occurring plasmids, cointegrate molecules such as this can be readily isolated; in other cases, the intermediate is rapidly resolved into two separate plasmids, each of which contains a copy of the transposon, with the target sequence on the recipient plasmid being duplicated on either side of the inserted transposon.
Since the two copies of the transposon are in the same orientation (i.e. it is a direct repeat), resolution of the cointegrate can occur by recombination between the two copies . This can be achieved by general recombination using host recombination systems. However, some transposons (including

Tn3) encode their own resolution system. The tnpR gene of Tn3 codes for a resolvase which mediates a site-specific recombination at the resolution site within the transposon, thus ensuring an efficient resolution of the cointegrate, independent of host recombination activity.

Non-replicative (conservative) transposition

Not all transposons show exactly the same behaviour. In particular, some transposons and some insertion sequences, do not replicate when they transpose, exhibiting a mode of transposition known as conservative (or non-replicative) transposition. This occurs with the insertion sequence IS10 and the related transposon Tn10.

6 Horizontal Gene Transfer

Horizontal gene transfer , the transmission of DNA between different genomes , occur between different species . Acquisition of DNA through horizontal gene transfer is distinguished from the transmission of genetic material from parents to offspring during reproduction, which is known as vertical gene transfer. Horizontal gene transfer is made possible in large part by the existence of mobile genetic elements, such as plasmids (extrachromosomal genetic material), transposons ("jumping genes"), and bacteria-infecting viruses (bacteriophages). These elements are transferred between organisms through different mechanisms, which in prokaryotes include transformation, conjugation, and transduction. In transformation, prokaryotes take up free fragments of DNA, often in the form of plasmids, found in their environment. In conjugation, genetic material is exchanged during a temporary union between two cells, which may entail the transfer of a plasmid or transposon. In transduction, DNA is transmitted from one cell to another via a bacteriophage.

In horizontal gene transfer, newly acquired DNA is incorporated into the genome of the recipient through either recombination or insertion. Recombination essentially is the regrouping of genes, such that native and foreign (new) DNA segments that are homologous are edited and combined. Insertion occurs when the foreign DNA introduced into a cell shares no homology with existing DNA. In this case, the new genetic material is embedded between existing genes in the recipient's genome.

6 – 1 Conjugation

Bacterial conjugation is the transfer of genetic material between bacterial cells by direct cell-to-cell contact or by a bridge-like connection between two cells. Discovered in 1946 by Joshua Lederberg and Edward Tatum.
In most cases, this involves the transfer of plasmid DNA, although with some organisms chromosomal transfer can also occur. As with other modes of gene transfer in bacteria.
In the simplest of cases, conjugation is achieved in the laboratory by mixing the two strains together and after a period of incubation to allow conjugation to occur, plating the mixture onto a medium that does not allow either parent

to grow, but on which a transconjugant that contains genes from both parents will grow. Plasmid transfer can be readily detected even if only, say, 1 in 10^6 recipients have received a copy of it.
Conjugation is most easily demonstrated amongst members of the Enterobacteriaceae and other Gram-negative bacteria (such as Vibrios and Pseudomonads).

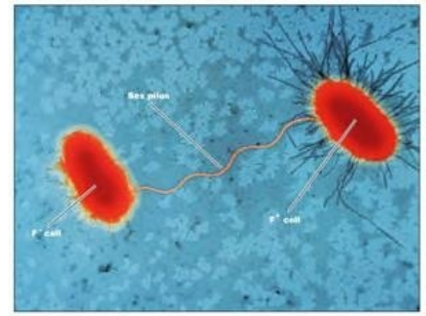


Figure 23 : Direct contact between two conjugating bacteria is first made via a pilus. [Czura, Microbial Genetics,Chapter 8)

Several genera of Gram-positive bacteria possess reasonably well-characterized conjugation systems; these include Streptomyces species, which are commercially important as the major producers of antibiotics, the lactic streptococci, which are also commercially important because of their application to various aspects of the dairy products industry, and medically important bacteria such as *Enterococcus faecalis* .
The most obvious significance of conjugation is that it enables the transmission of plasmids from one strain to another. Since conjugation is not necessarily confined to members of the same species, this provides a route for genetic information to flow across wide taxonomic boundaries. One practical consequence is that plasmids that are present in the normal gut flora can be transmitted to infecting pathogens, which then become resistant to a range of different antibiotics.

Mechanism of conjugation

Formation of mating pairs

In the vast majority of cases, the occurrence of conjugation is dependent on the presence, in the donor strain, of a plasmid that carries the genes required for promoting DNA transfer. In *E. coli* and other Gram-negative bacteria, the donor cell carries appendages on the cell surface known as pili. These vary considerably in structure – for example the pilus specified by the F plasmid is long, thin and flexible, while the RP4 pilus is short, thicker and rigid. The pili make contact with receptors on the surface of the recipient cell, thus forming a mating pair.
The pili then contract to bring the cells into intimate contact and a channel or pore is made through which the DNA passes from the donor to the recipient
Interestingly, this mechanism has much in common with a protein secretion system which is used by some bacteria to deliver protein toxins directly into host cells.

Transfer of DNA

The transfer of plasmid DNA from the donor to the recipient is initiated by a protein which makes a single-strand break (nick) at a specific site in the DNA, known as the origin of transfer (oriT). A plasmid-encoded helicase unwinds the plasmid DNA and the single nicked strand is transferred to the recipient starting with the 5^- end generated by the nick. Concurrently, the free 3^- end of the nicked strand is extended to replace the DNA transferred, by a process known as rolling circle replication which is analogous to the replication of single stranded plasmids and bacteriophages.The nicking protein remains attached to the 5^- end of the transferred DNA. DNA synthesis in the recipient converts the transferred single strand into a double stranded molecule. Note that this is a replicative process. Thus although there is said to be a transfer of the plasmid from one cell to another, what is really meant is a transfer of a copy of the plasmid. The donor strain still has a copy of the plasmid and can indulge in further mating with another recipient. It is also worth noting that after conjugation the recipient cell has a copy of the plasmid and it can transfer a copy to another recipient cell. The consequence can be an epidemic spread of the plasmid through the mixed population.

Mobilization and chromosomal transfer

Not all plasmids are capable of achieving this transfer to another cell unaided; those that can are known as conjugative plasmids. In some cases a conjugative plasmid is able to promote the transfer of (mobilize) a second otherwise nonconjugative plasmid from the same donor cell. This does not happen by chance and not all non-conjugative plasmids can be mobilized.
In order to understand mobilization the plasmid ColEI can be taken as an example . Mobilization involves the mob gene, which encodes a specific nuclease, and the bom site (¼oriT, the origin of transfer), where the Mob nuclease makes a nick in the DNA. ColE1 has the genes needed for DNA transfer but it does not carry the genes required for mating-pair formation. The presence of another (conjugative) plasmid enables the donor to form mating pairs with the recipient cell and ColE1 can then use its own machinery to carry out the DNA transfer.
Some plasmids which can be mobilized do not carry a *mob* gene. Mobilization then depends on the ability of the Mob nuclease of the conjugative plasmid to recognize the bom site on the plasmid to be mobilized. This only works if the two plasmids are closely related. On the other hand, the bom site is essential for mobilization. This is an important factor in genetic modification as removal of the bom site from a plasmid vector ensures that the modified plasmids cannot be transferred to other bacterial strains .

In most cases, the DNA that is transferred from the donor to the recipient consists merely of a copy of the plasmid. However, some types of plasmids can also promote transfer of chromosomal DNA. The first of these to be discovered, and the best known, is the F (fertility) plasmid of *E. coli*, but similar systems exist in other species, notably *Pseudomonas aeruginosa*.
However, in many cases chromosomal transfer occurs without any stable association with the plasmid, possibly by a mechanism analogous to mobilization of a non-conjugative plasmid.
When a plasmid is transferred from one cell to another by conjugation, the complete plasmid is transferred. In contrast, chromosomal transfer does not involve a complete intact copy of the chromosome. One reason for this is the time required for transfer. The process is less efficient than normal DNA replication and transfer of the whole chromosome would take about 100 min (in *E. coli*). The mating pair very rarely remains together this long. In contrast, a plasmid of say 40 kb is equivalent to 1 per cent of the length of

the chromosome, thus the transfer of the plasmid would be expected to be completed in 1 min.

DNA synthesis by the rolling circle mechanism replaces the transferred strand in the donor, while the complementary DNA strand is made in the recipient. Therefore, at the end of the process, both donor and recipient possess completely formed plasmids. For transfer of the F plasmid then, an F-containing cell, which is designated F+, can mate with a cell lacking the plasmid, designated F-, to yield two F+ cells

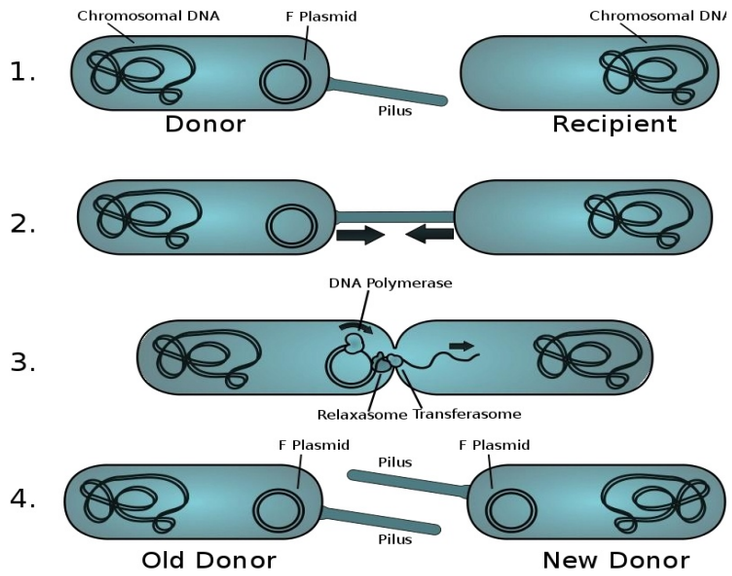


Figure 24: Schematic drawing of bacterial conjugation[Czura, Microbial Genetics,Chapter 8)

The F plasmid is an episome, a plasmid that can integrate into the host chromosome . When the F plasmid is integrated into the chromosome, the chromosome becomes mobilized and can lead to transfer of chromosomal genes. Following genetic recombination between donor and recipient, lateral gene transfer by this mechanism can be very extensive. Cells possessing an unintegrated F plasmid are called F+. Those that have a chromosome-integrated F plasmid are called Hfr (for high frequency of recombination).
A high-frequency recombination cell (also called an Hfr strain) is a bacterium with a conjugative plasmid (often the F-factor) integrated into its genomic DNA. The Hfr strain was first characterized by Luca Cavalli-Sforza. Unlike a normal F^+ cell, Hfr strains will, upon conjugation with a F^- cell, attempt to transfer their entire DNA through the mating bridge, not to be confused with the pilus. This occurs because the F factor has integrated itself via an insertion point in the bacterial chromosome. Due to the F factor's inherent tendency to transfer itself during conjugation, the rest of the bacterial genome is dragged along with it, thus making such cells very useful and interesting in terms of studying gene linkage and recombination. Because the genome's rate of transfer through the mating bridge is constant, molecular biologists and geneticists can use Hfr strain of bacteria (often *E. coli*) to study genetic linkage and map the chromosome. The procedure commonly used for this is called interrupted mating.

An important breakthrough came when Luca Cavalli-Sforza discovered a derivative of an F^+ strain. On crossing with F^- strains this new strain produced 1000 times as many recombinants for genetic markers as did a normal F^+ strain. Cavalli-Sforza designated this derivative an Hfr strain to indicate a high frequency of recombination. In $Hfr \times F^-$ crosses, virtually none of the F^- parents were converted into F^+ or into Hfr. This result is in contrast with $F^+ \times F^-$ crosses, where infectious transfer of F results in a large proportion of the F^- parents being converted into F^+. It became apparent that an Hfr strain results from the integration of the F factor into the chromosome.

Now, during conjugation between an Hfr cell and a F^- cell a part of the chromosome is transferred with F. Random breakage interrupts the transfer before the entire chromosome is transferred. The chromosomal fragment can then recombine with the recipient chromosome.

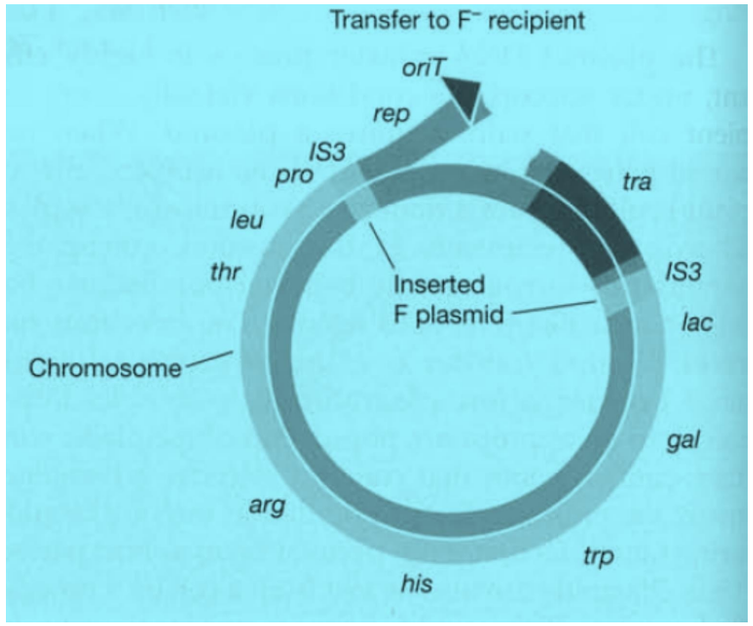


Figure 25: Breakage of the Hfr chromosome at the origin of transfer and the beginning of DNA transfer to the recipient

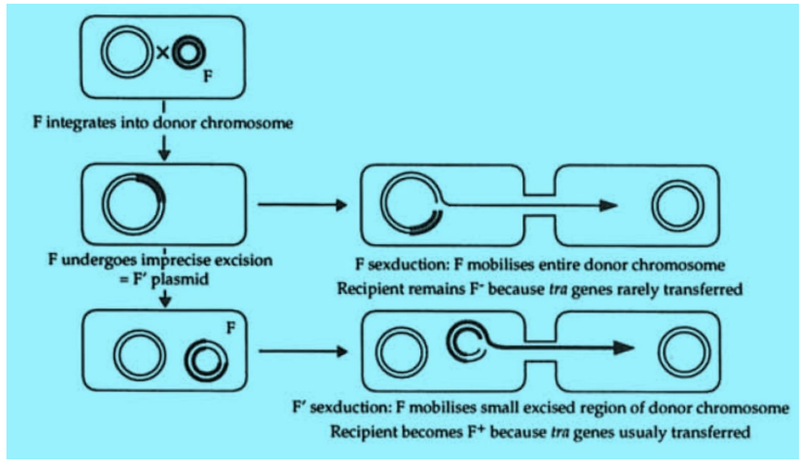


Figure 26 : F sexduction :F mobilizes small excised region of donor hromosome.

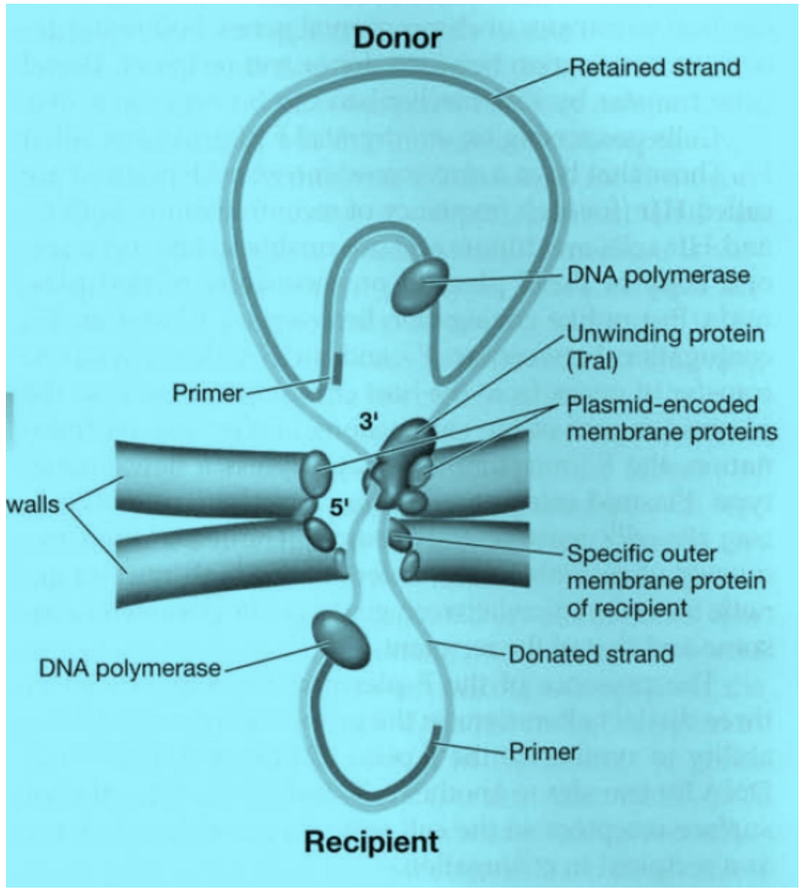


Figure 27: Details of the replication and transfer process.

F plasmids containing chromosomal genes are called F' *(Fprime)plasmids.*F' plasmids differ from normal F plasmids in that they contain identifiable chromosomal genes, and they transfer these genes at high frequency to recipients. F' -mediated transfer resembles specialized transduction in that only a restricted group of chromosomal genes can be transferred.

Transferring a known F' into a recipient allows one to establish diploids (two copies of each gene) for a limited region of the chromosome. Such partial diploids are called *merodiploids.*

F plasmid

The F plasmid was originally discovered during attempts to demonstrate genetic exchange in *E. coli* by mixed culture of two auxotrophic strains, so that plating onto minimal medium would only permit recombinants to grow. It was shown quite early on that the recombinants were all derived from one of the parental strains and that a one-way transfer of information was therefore involved, from the donor ('male') to the recipient ('female'). The donor strains carry the F plasmid (F) while the recipients are F^-. One feature of this system which must have seemed curious at the time is that co-cultivation of an F and an F^- strain resulted in the 'females' being converted into 'males'. This is due to the transmission of the F plasmid itself which occurs at a high frequency, in contrast to the transfer of chromosomal markers which is very inefficient with an F donor.

The usefulness of conjugation for genetic analysis was enormously enhanced by the discovery of donor strains in which chromosomal DNA transfer occurred much more commonly. These Hfr (High Frequency of Recombination) strains arise by integration of the F plasmid into the bacterial chromosome. An additional characteristic of an Hfr strain is that chromosomal transfer starts from a defined point and proceeds in a specific direction. The origin of transfer is determined by the site of insertion of the F plasmid and the direction is governed by the orientation of the inserted plasmid.

since transfer does not start from a defined point on the chromosome. The combination of the partial transfer of chromosomal DNA with the ordered transfer of genes made conjugation an important tool in the mapping of bacterial chromosomes, Integration and excision of F: formation of F plasmids Integration of the F plasmid occurs by recombination between a sequence on the plasmid and a chromosomal site.

The lac operon has therefore become incorporated into the plasmid (at the same time generating a deletion in the chromosome) and will be transferred with the plasmid to a recipient strain. This is one mechanism where by a plasmid can acquire additional genes from a bacterial chromosome and transfer them to another strain or species.

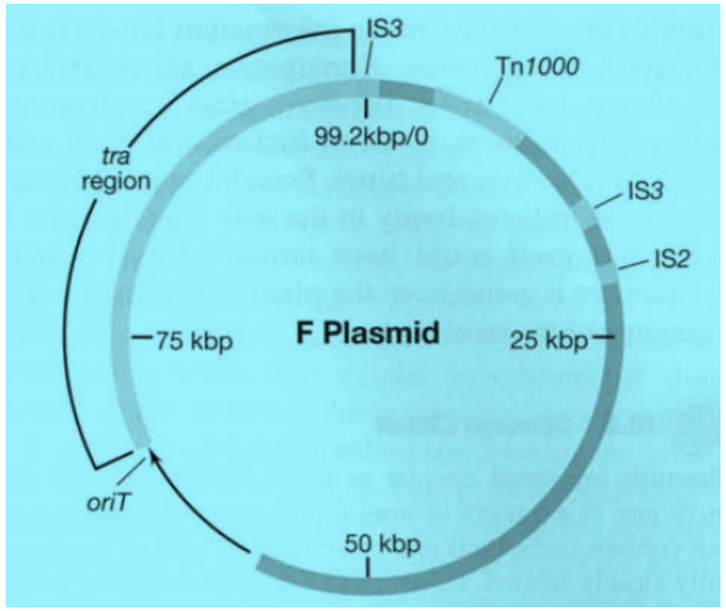


Figure 28 : Genetic map of the F plasmid of *Escherichia coli.*

Before the advent of gene cloning, F plasmids were useful in a number of ways, including the isolation of specific genes and their transfer to other host strains.
This enabled the creation of partial diploids, i.e. strains with one copy of a specific gene on the plasmid in addition to the chromosomal copy. The use of partial diploids for the study of the regulation of the lac operon especially in distinguishing regulatory genes that operate in trans (i.e. they influence the expression of a gene on a different molecule) from those that only affect the genes to which they are attached (i.e. they operate in cis). Such experiments are still relevant, but would now use recombinant plasmids produced *in vitro*.

Conjugation in other bacteria

Many Gram-positive species, ranging from Streptomyces to Enterococcus, also possess plasmids that are transmissible by conjugation and in many cases the mechanism of DNA transfer is quite similar to that described above.
In general, the number of genes required for conjugative transfer, in some cases as few as five genes, is very much less than in Gram-negative bacteria where 20 or more genes are needed. Conjugative plasmids in Gram-positive bacteria can therefore be considerably smaller. One reason for a smaller number of genes being required is that there seems to be no need for production of a pilus. This is probably, at least in part, a reflection of the different cell-wall architecture in Gram-positive bacteria which lack the outer membrane characteristic of the Gram-negatives.
One group of Gram-positive bacteria where conjugation systems have been studied in detail are the enterococci, principally *Enterococcus faecalis*. Some strains of *E. faecalis* secrete diffusible peptides that have a pheromone-like action that can stimulate the expression of the transfer (*tra*) genes of a specific plasmid in a neighbouring cell. Note that, rather surprisingly, it is the recipient cell that produces the pheromones. The donor cell, carrying the plasmid, has a plasmid encoded receptor on the cell surface to which the pheromone binds. Different types of plasmid code for different receptors and are therefore stimulated by different pheromones. However the recipient produces a range of pheromones and is therefore capable of mating with cells carrying different plasmids.
After the pheromone has bound to the cell-surface receptor it is transported into the cytoplasm, by a specific transport protein, where it interacts with a protein called TraA. This protein is a repressor of the *tra* genes on the plasmid and the binding of the peptide to it relieves that repression, thus stimulating expression of the tra genes. One result is the formation of aggregation products which cause the formation of a mating aggregate containing donor and recipient cells bound together. A further consequence of expression of the tra genes is stimulation of the events needed for transfer of the plasmid which occurs by a mechanism similar to that described previously.
One advantage of this system is that the cells containing the plasmid do not express the genes needed for plasmid transfer unless there is a suitable recipient in the vicinity. Not only does this reduce the metabolic load on the

cell but it also means that they are not expressing surface antigens (such as conjugative pili) that could be recognized by the host immune system.

Conjugative transposons *E. faecalis* also provides an example of an exception to the general rule that conjugation is plasmid-mediated. Some strains of *E. faecalis* contain a transposon known as Tn916. Transposons are able to move from one DNA site to another. What sets Tn916 apart from other transposons is its ability to transfer from one cell to another by conjugation.

Conjugative transposons such as Tn916 differ from plasmids in that they are replicated and inherited as part of the chromosome. There is no stable independently replicating form as there is with a plasmid. However, closer inspection of the method of transfer

In particular, Tn916 contains a origin of transfer (oriT) which is quite similar to that found in many plasmids. The first step in transfer is the excision of the transposon from the chromosome, using transposon-encoded enzymes (Int and Xis) which are related to those responsible for the integration and excision of bacteriophage lambda . This produces a circular molecule that resembles a plasmid in all but one vital feature – it does not have an origin of replication so is unable to be copied in the normal way. However since it does have an oriT site and carries the *tra* genes needed for conjugal transfer, it can be transferred to a recipient cell.

transfer of Tn916 involves single-stranded DNA synthesis initiated at oriT and transfer of the displaced strand to the recipient. The transferred single strand is then circularized and converted to a double-stranded circular form which is inserted randomly into the recipient chromosome by the action of the integrase.

Transfer would start from oriT and would have to work right round the chromosome before reaching the rest of the transposon. This does not seem to happen. The reason is that the promoter for expression of the tra genes is found towards the left-hand end of the transposon and faces away from the *tra* genes.

In the integrated linear form the tra genes will not be expressed. However, when the transposon is excised from the chromosome and circularized, this brings the promoter into the correct position and orientation for transcription of the *tra* genes. They will therefore be expressed from the circular intermediate, but not from the integrated form. This ensures that the transfer system will only be activated after excision has occurred.

Tn916 is the prototype of a family of related conjugative transposons that are especially widespread in Gram-positive cocci, although related elements also occur in Gram-negative bacteria (e.g. Bacteroides). For many of these

elements, including Tn916, conjugative transmission is promiscuous in that they can transfer to other species or genera. It can be assumed therefore that conjugative transposons have played a significant role in the dissemination of genetic material, including antibiotic resistance genes, throughout the bacterial kingdom. In particular. many of these transposons, including Tn916, carry a tetracycline resistance gene (*tet*M) which is found in a wide range of bacterial species, suggesting that they have played a role in the dispersal of this particular gene.

Conjugation in *Mycobacteria smegmatis*, like conjugation in *E. coli*, requires stable and extended contact between a donor and a recipient strain, is DNase resistant, and the transferred DNA is incorporated into the recipient chromosome by homologous recombination. However, unlike *E. coli* Hfr conjugation, mycobacterial conjugation is chromosome rather than plasmid based. Furthermore, in contrast to *E. coli* Hfr conjugation, in *M. smegmatis* all regions of the chromosome are transferred with comparable efficiencies. The lengths of the donor segments vary widely, but have an average length of 44.2kb. Since a mean of 13 tracts are transferred, the average total of transferred DNA per genome is 575kb. This process is referred to as Distributive conjugal transfer. Gray et al. found substantial blending of the parental genomes as a result of conjugation and regarded this blending as reminiscent of that seen in the meiotic products of sexual reproduction.

Bacteria related to the nitrogen fixing *Rhizobia* are an interesting case of inter-kingdom conjugation. For example, the tumor-inducing (Ti) plasmid of *Agrobacterium* and the root-tumor inducing (Ri) of *A. rhizogenes* contain genes that are capable of transferring to plant cells. The expression of these genes effectively transforms the plant cells into opine-producing factories. Opines are used by the bacteria as sources of nitrogen and energy. Infected cells form crown gall or root tumors, respectively. The Ti and Ri plasmids are thus endosymbionts of the bacteria, which are in turn endosymbionts (or parasites) of the infected plant.

The Ti and Ri plasmids can also be transferred between bacteria using a system (the *tra*, or transfer, operon) that is different and independent of the system used for inter-kingdom transfer (the *vir*, or virulence, operon). Such transfers create virulent strains from previously a virulent strains.

Several conjugative plasmids have also been found in *Sulfolobus,* a genus of ***Archaea***. Little is known about conjugation in *Sulfolobus,*although it is known that cell pairing occurs before plasmid transfer and that transfer is unidirectional. However, with one exception, the genes involved seem to have little similarity to those in gram-negative *Bacteria.* The exception is a gene similar to *traG,* whose protein product in F plasmid-mediated conjugation seems to be involved in stabilizing mating pairs. It thus seems likely that the mechanism of conjugation in *Archaea* is quite different from that in *Bacteria.*

6- 2 Transformation

Transformation describes the uptake of naked DNA from the surrounding medium, andth e change in genotype thusc onferred upon the recipient cell the transformant. Transformation occurs naturally in many bacteria (e.g. *Bacillus, Streptomyces* and *Haemophillus* spp.) although competence (ability to take up exogenous DNA) is usually transient, being associated with a particular physiological state and requiring the expression of specific competence factors. Other species of bacteria, including *E. coli,* are refractory to natural transformation, but a state of competence can be induced artificially which allows DNA uptake; this has facilitated the use of *E. coli* for molecular cloning .

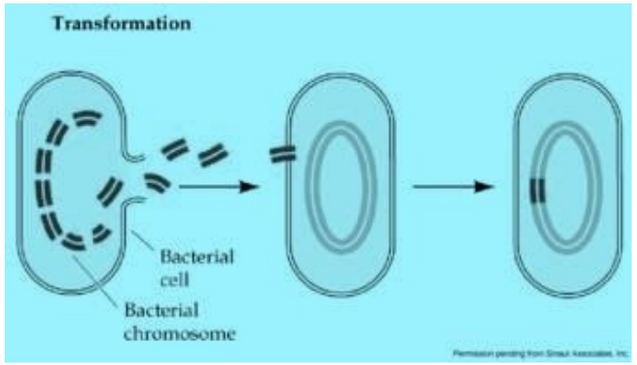


Figure 29: Transformation process

In a sense, it was the discovery of transformation, the uptake of DNA by a bacterial cell, that initiated the study of bacterial genetics and molecular biology as we know it today. It was in 1928 that Fred Griffith, working with pneumococcus (*Streptococcus pneumoniae*) discovered that a virulent strains could be restored to virulence by incubation with an extract from killed virulent cells. Sixteen years later, Avery, MacLeod and McCarty demonstrated that the 'transforming principle' was DNA, which established the role of DNA as the hereditary material of the bacterial cell. Transformation has been important in genetic analysis of some species and more recently (and to a much greater extent) because of its key role in gene cloning.

With the pneumococcus, cells spontaneously become competent to take up DNA. Such naturally-occurring transformation has been most studied in *Bacillus subtilis* and *Haemophilus influenzae* (as well as *S. pneumoniae*) and was for some time thought to be limited to these and related species. It is now known to be much more widespread. In particular, transformation contributes extensively to the antigenic variation observed in the gonococcus (*Neisseria gonorrhoeae*) through the transfer of *pil* genes coding for the major protein subunit of the surface appendages (pili) by which the bacteria attach to epithelial cells. Although the number of species in which natural transformation has been demonstrated is still quite limited, it is likely that it occurs, albeit at a low level, in many other bacteria. The details of the process vary between species, but some generalizations are possible. Competence generally occurs at a specific stage of growth, most commonly in late log phase, just as the cells are entering stationary phase. This may be a response to cell density rather than (or as well as) growth phase. For example, in *Bacillus subtilis*, some of the genes involved in the development of competence are also involved in the early stages of sporulation. The development of competence at this stage is associated not only with nutrient depletion but also with the accumulation of specific secreted products (competence factors) which act via a two component regulatory system to stimulate the expression of other genes required for competence. Since the level of these competence factors is dependent on cell concentration, competence will only develop at high cell density. This is a form of quorum sensing, in which the response of an individual cell is governed by the concentration of bacteria in the surrounding medium.

Following the development of competence, double-stranded DNA fragments bind to receptors on the cell surface, but only one of the strands enters the cell. In some species, the process is selective for DNA from the same species, through a requirement for short species-specific sequences. For

example, the uptake of DNA by the meningococcus (*Neisseria meningitidis*) is dependent on the presence of a specific 10-bp uptake sequence. The genome of *N. meningitidis* contains nearly 2000 copies of this sequence, which will only occur infrequently and by chance, in other genomes. Similarly, transformation of *Haemophilus influenzae* is facilitated by the presence of a 29-bp uptake sequence which occurs approximately 1500 times in the genome of *H. influenzae*. These organisms will therefore only be transformed efficiently with DNA from the same species.

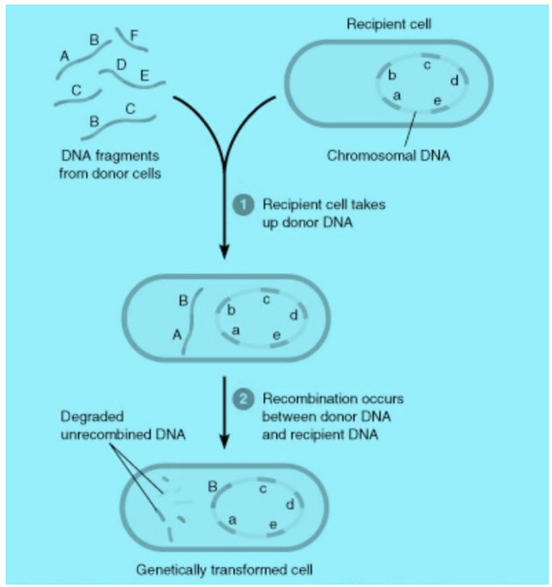


Figure 30 : Transformation steps[Benjamin cumming ,an imprint of Addison Wesley Longman,Inc,2001].

On the other hand, *B. subtilis* and *S. pneumoniae* can take up virtually any linear DNA molecule. But taking up the DNA is only the start. If the cell is to become transformed in a stable manner, the new DNA has to be replicated and inherited. As here fragments of chromosomal DNA (rather than plasmids) are being considered, replication of the DNA will only happen if

the incoming DNA is recombined with the host chromosome. This requires homology between the transforming DNA and the recipient chromosome. This does not constitute an absolute barrier to transformation with DNA from other species. Provided there is enough similarity in some regions of the chromosome, those segments of DNA can still undergo recombination with the recipient chromosome. The closer the taxonomic relationship, the more likely it is that they will be sufficiently similar.

One example of this, with considerable practical significance, is the development of resistance to penicillin in *S. pneumoniae*. This appears to have occurred by the replacement of part of the genes coding for the penicillin target enzymes with corresponding DNA from naturally-resistant oral streptococci.

Natural transformation is of limited usefulness for artificial genetic modification of bacteria, mainly because it works best with linear DNA fragments rather than the circular plasmid DNA that is used in genetic modification. For introducing foreign genes into a bacterial host, various techniques are used to induce an artificial state of competence. Alternatively, a mixture of cells and DNA may be briefly subjected to a high voltage which enables the DNA to enter the cell (a process known as electroporation). Although the mechanisms involved are quite different, they all share the characteristic feature of the uptake of 'naked' DNA by the cells and are therefore also referred to as transformation.

Bacteria can be transformed with DNA extracted from a bacterial virus rather than from another bacterium. This process is known as **transfection.**

If the DNA is from a lytic bacteriophage, transfection leads to virus production and can be measured by the standard phage plaque assay . Transfection is useful for studying the mechanism of transformation and recombination because the small size of phage genomes allows for the isolation of a nearly homogeneous population of DNA molecules. By contrast, in conventional transformation, the transforming DNA is typically a random assortment of chromosomal DNA of various lengths, and this tends to complicate experiments designed to study the mechanism of transformation.

In summary:

Bacterial transformation may be referred to as a stable genetic change brought about by the uptake of naked DNA (DNA without associated cells

or proteins) to increase DNA quantity; competence refers to the state of being able to take up exogenous DNA from the environment. There are two forms of transformation and competence: natural and artificial:

Natural transformation :Natural transformation is a bacterial adaptation for DNA transfer that depends on the expression of numerous bacterial genes whose products appear to be responsible for this process. In general, transformation is a complex, energy-requiring developmental process. In order for a bacterium to bind, take up and recombine exogenous DNA into its chromosome, it must become competent, that is, enter a special physiological state. Competence development in *Bacillus subtilis* requires expression of about 40 genes. The DNA integrated into the host chromosome is usually (but with rare exceptions) derived from another bacterium of the same species, and is thus homologous to the resident chromosome.

Transformation, as an adaptation for DNA repair : Competence is specifically induced by DNA damaging conditions. For instance, transformation is induced in *Streptococcus pneumoniae* by the DNA damaging agents mitomycin C (a DNA crosslinking agent) and fluoroquinolone (a topoisomerase inhibitor that causes double-strand breaks). In *B. subtilis*, transformation is increased by UV light, a DNA damaging agent. In *Helicobacter pylori*, ciprofloxacin, which interacts with DNA gyrase and introduces double-strand breaks, induces expression of competence genes, thus enhancing the frequency of transformation.

6-3 Transduction

Transduction is the process by which cellular genes can be transferred from a donor to a recipient cell by a virus particle (phage-mediated transfer), the recipient being known as a transductant following transfer .
This process is normally highly specific for phage DNA. However, with some phages, errors can be made and fragments of bacterial DNA (produced by phage-mediated degradation of the host chromosome) are occasionally packaged by mistake leading to phage-like particles that contain a segment of bacterial genome . These transducing particles arecapable of infecting a recipient cell, since the information necessary for attachment and injection of DNA is carried by the proteins of the phage particle, irrespective of the nucleic acid it contains. The transduced segment of DNA will therefore be injected into the new host cell.
Not all bacteriophages are capable of carrying out transduction. The basic requirements of an effective transducing phage are that infection should result in an appropriate level of degradation of the chromosomal DNA to form suitably sized fragments at the right time for packaging and that the specificity of the packaging process should be comparatively low.
In some cases, the transduced DNA is a bacterial plasmid, in which case the injected DNA molecule is capable of being replicated and inherited. More commonly the DNA incorporated into the transducing particle is a fragment of chromosomal DNA which will be unable to replicate in the recipient cell. For it to be replicated and inherited, it must be incorporated into the recipient chromosome (by homologous recombination), as is the case with other mechanisms of gene transfer.
This process is known as generalized transduction since essentially any gene has an equal chance of being transduced.

6-3.1 Specialized transduction

some phages (temperate phages) are able to establish a state known as lysogeny, in which expression of phage genes and replication of the phage is repressed. In many cases the prophage is inserted into the bacterial DNA and replicates as part of the chromosome. When lysogeny breaks down and the phage enters the lytic cycle, it is excised from the chromosome by recombination between sequences at each end of the integrated prophage. If this recombination event happens in the wrong place, an adjacent region of bacterial DNA is incorporated into the phage DNA. All the progeny of this phage will then contain this bacterial gene which will therefore be transduced at a very high frequency (effectively 100 per cent per phage particle) once the transducing phage has been isolated. Since the DNA transferred is limited to a very small region of the chromosome, the phenomenon is known as specialized (or restricted) transduction.
This is very similar to the formation of F- plasmids referred to earlier . As with the F- plasmids, it is now much easier to add genes to DNA by creating recombinants *in vitro* .
Another phage that has been employed in a similar way is the phage Mu which has the advantage of inserting at multiple sites in the chromosome by a transposon-like mechanism. It is therefore much easier to create a wide range of specialized transducing phages with Mu which can be used both in genetic mapping and in mutagenesis.

When a normal temperate phage (that is, a nondefective phage) lysogenizes a cell and its DNA is converted to the prophage state, the cell is immune to further infection by the same type of phage. This acquisition of immunity can be considered a change in phenotype. However, other phenotypic alterations can often be detected in the lysogenized cell that are unrelated to phage immunity. Such a change, which is brought about through lysogenization by a normal temperate phage, is called phage conversion.
Two cases of phage conversion have been especially well studied. One involves a change in structure of a polysaccharide on the cell surface of *Salmonella anatum* on lysogenization with bacteriophage 815.The second involves the conversion of non toxin-producing strains of *Corynebacterium diphtheriae* (the bacterium that causes the disease diphtheria) to toxin-producing (pathogenic) strainsupon lysogenization with phage *f3* . In both of these situations, the genes encoding the necessary molecules are an integral part of the phage genome and hence are automatically (and exclusively) transferred upon infection by the phage and lysogenization.

Lysogeny probably carries a strong selective value for the host cell because it confers resistance to infection by viruses of the same type. Phage conversion may also be of considerable evolutionary significance because it results in efficient genetic alteration of host cells. Many bacteria isolated from nature are natural lysogens. It seems reasonable to conclude, therefore, that lysogeny is a common condition and may often be essential for survival of the host in nature.

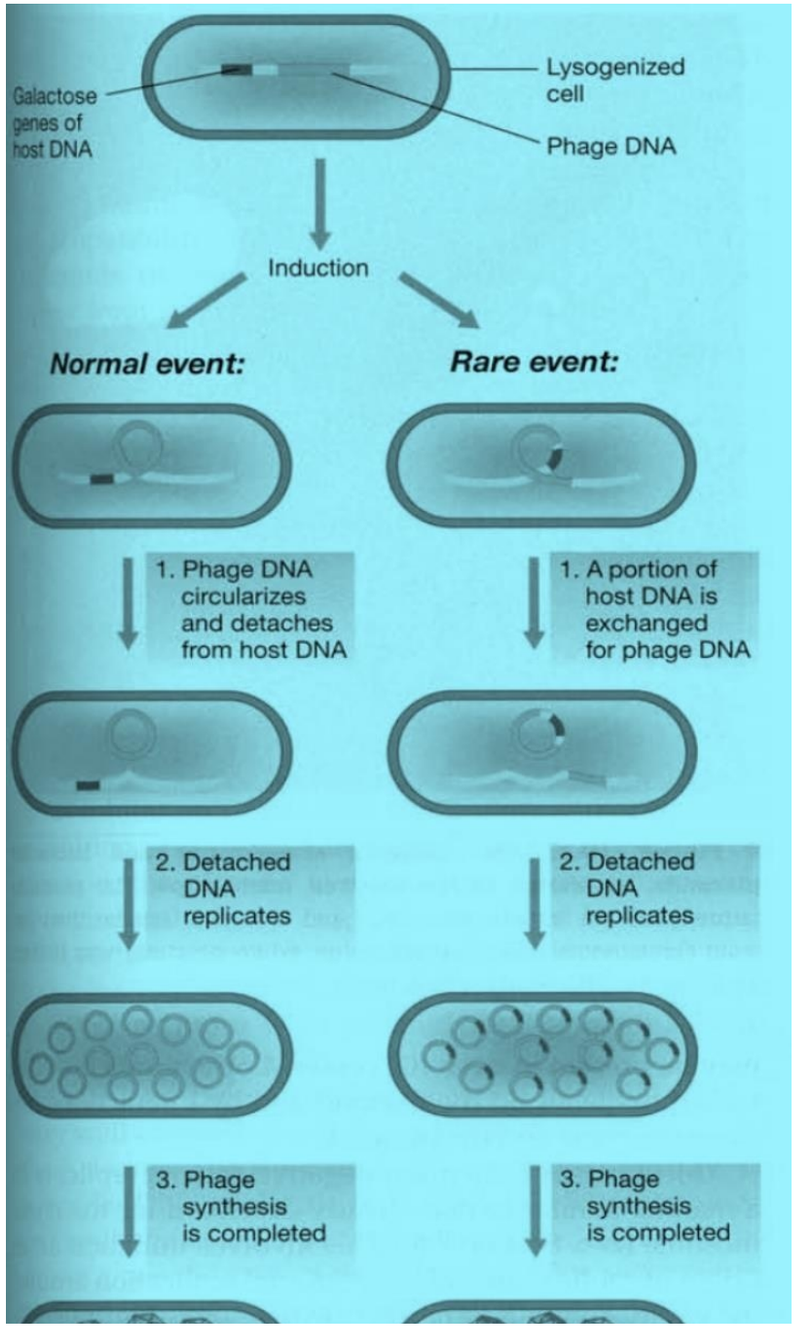


Figure 31 : Specialized transduction. Normallytic events, and the production of particles transducing the galactose genes in an*Escherichia coli* cell containing a lambda prophage.

6-3-2 Genaralized transduction

Transduction is the phage-mediated transfer of genetic material. The key step in transduction is the packaging of DNA into the phage heads during lytic growth of the phage . This process is normally highly specific for phage DNA. However, with some phages, errors can be made and fragments of bacterial DNA (produced by phage-mediated degradation of the host chromosome) are occasionally packaged by mistake leading to phage-like particles that contain a segment of bacterial genome . These transducing particles are capable of infecting a recipient cell, since the information necessary for attachment and injection of DNA is carried by the proteins of the phage particle, irrespective of the nucleic acid it contains. The transduced segment of DNA will therefore be injected into the new host cell.
Not all bacteriophages are capable of carrying out transduction. The basic requirements of an effective transducing phage are that infection should result in an appropriate level of degradation of the chromosomal DNA to form suitably sized fragments at the right time for packaging and that the specificity of the packaging process should be comparatively low.
The transduced DNA is a bacterial plasmid, in which case the injected DNA molecule is capable of being replicated and inherited. More commonly the DNA incorporated into the transducing particle is a fragment of chromosomal DNA which will be unable to replicate in the recipient cell. For it to be replicated and inherited, it must be incorporated into the recipient chromosome (by homologous recombination), as is the case with other mechanisms of gene transfer.

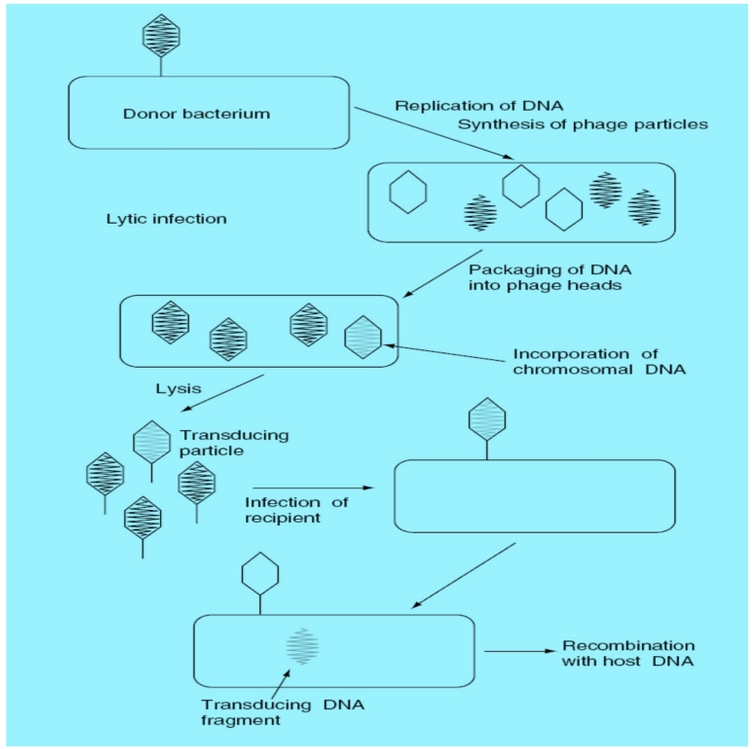


Figure 32 : Generalized transduction [Dale and Park ,2004]

6-4 Recombination

General (homologous) recombination

A common feature of all the forms of gene transfer between bacteria, except for the transfer of plasmids (which can replicate independently), is the requirement for the transferred piece of DNA to be inserted into the recipient chromosome by breaking both DNA molecules, crossing them over and rejoining them. This process, known as recombination. There are several

different forms of recombination, but the mechanisms that require the presence of homologous regions of DNA which must be highly similar but do not have to be identical are of specific interest in this context. It is therefore known as homologous recombination. Of the alternative forms of recombination, site-specific recombination is particularly important.
It should be noted that recombination mechanisms have other roles within the cell apart from the incorporation of foreign DNA. In particular, recombination mechanisms are involved with some types of DNA repair . These may actually be of more fundamental importance to the cell and may be the real reason why bacteria have evolved to contain several mechanisms for recombining DNA molecules.

Homologous recombination is divided into four stages: synapsis, strand transfer, repaira nd resolution.
(1) In synapsis, homologous duplexes are aligned.
(2) During strand transfer , a single DNA strand is transferred from one duplex to the other. The first strand transfer marks the initiation of recombination as it invades the homologous duplex and (if the recipient duplex is intact) displaces a resident strand. This process may generate a short region of hetero duplex DNA duplex DNA comprising strands from different parental molecules which may contain base mismatches reflecting sequence differences (different alleles) in the parental duplexes. If the recipient duplex is intact, the displaced resident strand is able to pair with the free strand of the initiating duplex. The two transferred strands cross each other, forming a structure termed a cross bridge, cross branch or Holliday junction.The site of the Holliday junction may move in relation to its original position by progressive strand exchange between duplexes. This is branch migration, and may increase or decrease the amount of hetero duplex DNA.
(3) Repair and resolution do not occur in a fixed order as this depends upon the recombining partners and the availability of appropriate enzymes. Repair refers to three different processes. In the simplest case, the recombining duplexes are intact (i.e. there is no genetic information missing from either duplex) and repair involves religation of the broken strands. This is conservative recombination .The Holliday junction can then be resolved in either of two planes to generate one of two products a patch of heteroduplex DNA in a non recombinant background, or a splice of heteroduplex DNA with recombination of flanking markers.
However, if genetic information is missing from either duplex(i.e. if there is a single-strand gap, or a break) DNA repair synthesis replaces the missing

information using information from the homologous duplex as a template . Recombination including the synthesis of new DNA is non conservative recombination. In the extreme case where an entire chromosome segment is missing, resolution of the Holliday junction yields a replication fork which can duplicate the missing segment .

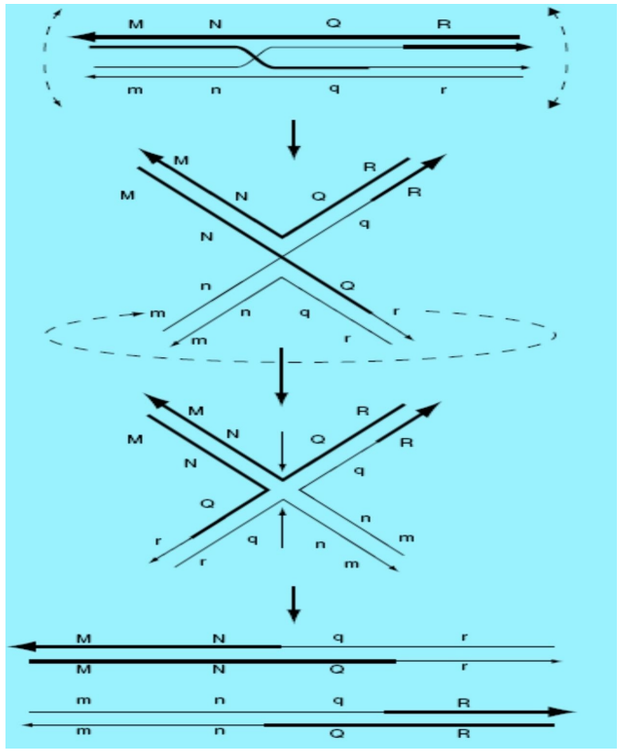


Figure 33: The Holliday model of homologous recombination between intact duplexes[Dale and Park ,2004]

Molecular basis of homologous recombination**.** In *E.* coli, proteins responsible for all the major stages of recombination - synapsis, strand exchange, branch migration and Holliday junction resolution have been identified . The remaining requirement or recombination, single stranded DNA, appears to be provided by a number of overlapping pathways.

The RecBCD pathway is the major source of recombinogenic substrates in *E.* coli. The products of the genes recB, recC and recD combine to form a large complex, the RecBCD enzyme or exonuclease V, which has three activities: (1) ATP-dependent exonuclease activity; (2) ATP-enhanced endonuclease activity; (3) ATP-dependent helicase activity.

Recombination is stimulated by cis-acting sites termed chi (GCTGGTGG), which occur approximately one every **5** kbp in the *E.* coli genome and represent preferential cleavage sites for the endonuclease. RecBCD cuts the strand containing the chi sequence approximately f ive bases to the 3' side in an orientation dependent manner. Chi sites are thus recombination hotspots (q.v.). Blunt ended linear duplex DNA is the substrate for the helicase activity. The enzyme binds to the DNA and progressively unwinds it, producing paired single-stranded loops in its wake. These loops are occasionally cleaved by the endonuclease activity of the enzyme, and following cleavage further unwinding generates the single-stranded tail necessary for homologous recombination.

Although the RecBCD pathway is important for recombination, *rec*BC mutants retain up to 10% homologous recombination activity in, dicating the presence of other pathways for producing recombinogenic DNA. Additionally, almost full recombination proficiency can be restored to *rec*BdC mutants by mutations at other loci, notably *sbc*A, *sbc*B and *sbc*C. The analysis of mutant strains with the genotype *rec*BC *sbc*BC has facilitated the identification of further recombination genes. These have overlapping functions, and collectively define the RecF pathway. Another system, revealed b y studying *rec*BC *sbc*A mutants, involves upregulation of the ***recE*** gene

Site-specific and non-homologous (illegitimate) recombination

Does not require homology between recombining partners. The proteins mediating this process (site-specific recombinases) recognize short, specific DNA sequences in the donor and recipient molecules, and interaction between the proteins facilitates recombination. Homology often exists between the donor and recipient sites because the same recombinase protein binds to both recognition sites.

7 Restriction and Modification System

The restriction/modification system in bacteria is a small-scale immune system for protection from infection by foreign DNA.

The restriction modification system (RM system) is used by bacteria, and perhaps other prokaryotic organisms, to protect themselves from foreign DNA, such as the one borne by bacteriophages. It was first discovered by Salvatore Luria and Mary Human in 1952 and 1953. They found that bacteriophage growing within an infected bacterium could be modified, so that upon their release and re-infection of a related bacterium the bacteriophage's growth is restricted (inhibited). Later work by Daisy Dussoix and Werner Arber in 1962 and many other subsequent workers led to the understanding that restriction was due to attack and breakdown of the modified bacteriophage's DNA by specific enzymes of the recipient bacteria. Daniel Nathans and Hamilton Smith in 1975 work resulted in the discovery of the class of enzymes now known as Restriction enzymes. When these enzymes were isolated in the laboratory they could be used for controlled manipulation of DNA, thus providing the foundation for the development of genetic engineering. Werner Arber, Daniel Nathans, and Hamilton Smith were awarded the Nobel Prize in Physiology or Medicine in 1978 for their work on restriction-modification.

Bacteria have restriction enzymes, also called restriction endonucleases, which cleave double stranded DNA at specific points into fragments, which are then degraded further by other endonucleases. This prevents infection by effectively destroying the foreign DNA introduced by an infectious agent (such as a bacteriophage). Approximately one quarter of known bacteria possess RM systems and of those about one half have more than one type of system.

As the sequences recognized by the restriction enzymes are very short, the bacterium itself will almost certainly contain some within its genome. In order to prevent destruction of its own DNA by the restriction enzymes, methyl groups are added. These modifications must not interfere with the DNA base-pairing, and therefore, usually only a few specific bases are modified on each strand.

Endonucleases cleave internal/non-terminal phosphodiester bonds. Restriction endonucleases cleave internal phosphodiester bonds only after recognising specific sequences in DNA which are usually 4-6 base pairs long, and often palindromic.

Restriction endonucleases (restriction enzymes) cut DNA molecules at specific positions. The discovery of a large number of restriction and modification systems called for a uniform system of nomenclature. A suitable system was proposed by Smith and Nathans (1973) , the designation of these enzymes comes from the source organism, using the first letter of the genus name plus the first two letters of the species with additional letters/numbers to indicate the specific enzyme (since one species may produce several different restriction enzymes).

Recognition sequences
Most, but not all, type II restriction endonucleases recognize and cleave DNA within particular sequences of four to eight nucleotides which have a twofold axis of rotational symmetry. Such sequences are often referred to as palindromes because of their similarity to words that read the same backwards as forwards.
For example, the restriction and modification enzymes R. *Eco*RI and M.*Eco*RI recognize the sequence:

5-GAA T TC-3
3-CT T AAG-5

The position at which the restricting enzyme cuts is usually shown by the symbol '/' and the nucleotides methylated by the modification enzyme are usually marked with an asterisk. For *Eco*RI these would be represented thus:

5-G/AA* T T C-3
3-C TT A*A/G-5

For convenience it is usual practice to simplify the description of recognition sequences by showing only one strand of DNA, that which runs in the 5′□to 3′□direction. Thus the *Eco*RI recognition sequence would be shown as G/AATTC. From the information shown above we can see that *Eco*RI makes single-stranded breaks four bases apart in the opposite strands of its target sequence so generating fragments with protruding 5′□termini:

5′-G 5′-AATTC-3′
3′-CTTAA-5′ G-5′

These DNA fragments can associate by hydrogen bonding between overlapping 5′ termini, or the fragments can circularize by intra molecular reaction.
For this reason the fragments are said to have sticky or cohesive ends. In principle, DNA fragments from diverse sources can be joined by means of the cohesive ends and, the nicks in the molecules can be sealed to form an intact artificially recombinant DNA molecule.
Not all type II enzymes cleave their target sites like *Eco*RI. Some, such as *Pst*I (CTGCA/G), produce fragments bearing 3′ overhangs, while others, such as *Sma*I (CCC/GGG), produce blunt or flush ends.
To date, over 10 000 microbes from around the world have been screened for restriction enzymes. From these, over 3000 enzymes have been found representing approximately 200 different sequence specificities.
Restriction enzymes with the same sequence specificity and cut site are known as isoschizomers. Enzymes that recognize the same sequence but cleave at different points, for example *Sma*I (CCC/GGG) and *Xma*I C/CCGGG), are sometimes known as neoschizomers.
Under extreme conditions, such as elevated pH or low ionic strength, restriction endonucleases are capable of cleaving sequences which are similar but not identical to their defined recognition sequence.
This altered specificity is known as *star* activity. The most common types of altered activity are acceptance of base substitutions and truncation of the number of bases in the recognition sequence. For example, *Eco*RI* (*Eco*RI star activity) cleaves the sequence N/AATTN, where N is any base, whereas *Eco*RI cleaves the sequence GAATTC.

If each of the bases occurs equally often in the DNA (i.e. the GC content is 50 percent, which is approximately true for *E. coli* but not always for other organisms) and if the distribution of the bases is random, then a six base sequence would be expected to occur, on average, every 46 bases (¼ 4096 bases). So an enzyme like *Eco*RI would be expected to cut the DNA into fragments with an average size of 4 kb. For enzymes recognizing a four-base site, the expected average fragment size is 44 (¼256) bases.
Note that the recognition sites are usually palindromic; in other words, if the double stranded sequence is turned round (maintaining the polarity of the strands), it is exactly the same.

Examples of restriction endonucleases [Dale and Park,2004].

Enzyme	Recognition site	Number of bases	Ends generated	Source of enzyme
*Eco*RI	G/AATTC	6	5′ sticky	*Escherichia coli*
*Hin*dIII	A/AGCTT	6	5′ sticky	*Haemophilus influenzae*
*Bam*HI	G/GATCC	6	5′ sticky	*Bacillus amyloliquefaciens*
Pst I	CTGCA/G	6	3′ sticky	*Providencia stuartii*
*Sma*I	CCC/GGG	6	Blunt	*Serratia marcescens*
*Sau*3A	/GATC	4	5′ sticky	*Staphylococcus aureus*
*Alu*I	AG/CT	4	Blunt	*Arthrobacter luteus*

There are five kinds of restriction modification system: type I, type II, type IIS, type III and type IV, all with restriction enzyme activity and a methylase activity. They were named in the order of discovery, although the type II system is the most common.

Type I systems are the most complex, consisting of three polypeptides: R (restriction), M (modification), and S (specificity). The resulting complex can both cleave and methylate DNA. The type I systems were the first to be characterized and a typical example is that from *E. coli* K12. The active enzyme consists of two restriction subunits, two modification (methylation) subunits and one recognition subunit. These subunits are the products of the *hsd*R, *hsd*M and *hsd*S genes
Both reactions require ATP, and cleavage often occurs a considerable distance from the recognition site. The S subunit determines the specificity of both restriction and methylation. Cleavage occurs at variable distances from the recognition sequence, so discrete bands are not easily visualized by gel electrophoresis.

Type II systems are the simplest and the most prevalent. Instead of working as a complex, the methyltransferase and endonuclease are encoded as two separate proteins and act independently (there is no specificity protein). Both proteins recognize the same recognition site, and therefore compete for activity. The methyltransferase acts as a monomer, methylating the duplex one strand at a time. The endonuclease acts as a homodimer, which

facilitates the cleavage of both strands. Cleavage occurs at a defined position close to or within the recognition sequence, thus producing discrete fragments during gel electrophoresis. For this reason, Type II systems are used in labs for DNA analysis and gene cloning.

Neisseria meningitides has multiple type II restriction endonuclease systems that are employed in natural genetic transformation. Natural genetic transformation is a process by which a recipient bacterial cell can take up DNA from a neighboring donor bacterial cell and integrate this DNA into its genome by recombination. Although early work on restriction modification systems focused on the benefit to bacteria of protecting themselves against invading bacteriophage DNA or other foreign DNA, it is now known that these systems can also be used to restrict DNA introduced by natural transformation from other members of the same, or related species.

In the pathogenic bacterium *Neisseria meningitides* (meningococci), competence for transformation is a highly evolved and complex process where multiple proteins at the bacterial surface, in the membranes and in the cytoplasm interact with the incoming transforming DNA. Restriction-modification systems are abundant in the genus *Neisseria*. *N. meningitides* has multiple type II restriction endonuclease systems. The restriction modification systems in *N. meningitides* vary in specificity between different clades. This specificity provides an efficient barrier against DNA exchange between clades. Restriction-modification appears to be a major driver of sexual isolation and speciation in the meningococci. Caugant and Miden suggested that restriction-modification systems in meningococci may act to allow genetic exchange among very close relatives while reducing (but not completely preventing) genetic exchange among meningococci belonging to different clonal complexes and related species.

Type III systems have R and M proteins that form a complex of modification and cleavage. The M protein, however, can methylate on its own. Methylation also only occurs on one strand of the DNA unlike most other known mechanisms. The heterodimer formed by the R and M proteins competes with itself by modifying and restricting the same reaction.

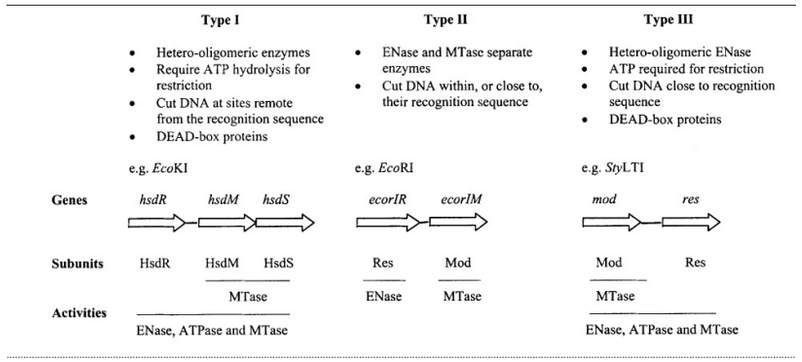


Figure 34 : The characteristics and organization of the genetic determinants and subunits of different types of R-M systems. ENase, endonuclease activity; MTase, methyltransferase activity [King and Murray ,1994].

DNA modification marks and protects the chromosome of a restriction-proficient bacterium, but there are situations where unmodified targets could become exposed to a restriction enzyme and thereby jeopardize the integrity of the bacterial chromosome. An obvious example of this problem is encountered when a bacterium acquires genes that encode a different R-M system from any already present within the cell. One simple solution is to delay production of the restriction enzyme until the modification enzyme has had time to modify all the targets in the bacterial chromosome .This process, however, takes many generations following the acquisition of the genes specifying *Eco*KI, because unmethylated DNA is a very poor substrate for modification

A methyltransferase, M.*Sss*I, that methylates the dinucleotide CpG has been isolated from *Spiroplasma*. This enzyme can be used to modify *in vitro* restriction endonuclease target sites which contain the CG sequence. Some of the target sequences modified in this way will be resistant to endonuclease cleavage, while others will remain sensitive. For example, if the sequence CCGG is modified with *Sss*I, it will be resistant to *Hpa*II but sensitive to *Msp*I. Since 90% of the methyl groups in the genomic DNA of many animals, including vertebrates and echinoderms, occur as 5-methylcytosine in the sequence CG, M.*Sss* can be used to imprint DNA from other sources with a vertebrate pattern.

The Dam and Dcm methylases of *E. coli*

Most laboratory strains of *E. coli* contain three sitespecific DNA methylases. The methylase encoded by the *dam* gene transfers a methyl group from *S*-adenosylmethionine to the N6 position of the adenine residue in the sequence GATC. The methylase encoded by the *dcm* gene (the Dcm methylase, previously called the Mec methylase) modifies the internal cytosine residues in the sequences CCAGG and CCTGG at the C5 position

In DNA in which the GC content is 50%, the sites for these two methylases occur, on average, every 256–512 bp. The third methylase is the enzyme M.*Eco*KI but the sites for this enzyme are much rarer and occur about once every 8 kb.

These enzymes are of interest for two reasons. First, some or all of the sites for a restriction endonuclease may be resistant to cleavage when isolated from strains expressing the Dcm or Dam methylases. This occurs when a particular base in the recognition site of a restriction endonuclease is methylated. The relevant base may be methylated by one of the *E. coli* methylases if the methylase recognition site overlaps the endonuclease recognition site. For example, DNA isolated from Dam□ *E. coli* is completely resistant to cleavage by *Mbo*I, but not *Sau*3AI, both of which recognize the sequence GATC. Similarly, DNA from a Dcm □□strain will be cleaved by *Bst*NI but not by *Eco*RII, even though both recognize the sequence CCATGG. It is worth noting that most cloning strains of *E. coli* are Dam□

Dcm□□ but double mutants are available .

The second reason these methylases are of interest is that the modification state of plasmid DNA can affect the frequency of transformation in special situations. Transformation efficiency will be reduced when Dam-modified plasmid DNA is introduced into Dam− *E. coli* or Dam- or Dcm-modified DNA is introduced into other species . When DNA is to be moved from *E. coli* to another species it is best to use a strain lacking the Dam and Dcm methylases.

R-M systems are major players in the co-evolutionary interaction between mobile genetic elements (MGEs) and their hosts. Genes encoding R-M systems have been reported to move between prokaryotic genomes within MGEs such as plasmids, prophages, insertion sequences/transposons, integrative conjugative elements (ICEs) and integrons. However, it was recently found that there are relatively few R-M systems in plasmids, some in prophages, and practically none in phages. On the other hand, all these MGEs encode a large number of solitary R-M genes, notably MTases. In

light of this, it is likely that R-M mobility may be less dependent on MGEs and more dependent, for example, on the existence of small genomic integration hotspots. It is also possible that R-M systems frequently exploit other mechanisms such as natural transformation, vesicles, nanotubes, gene transfer agents or generalized transduction in order to move between genomes.

8 Gene Cloning

Molecular (gene) cloning provides the foundation for genetic-engineering and most molecular genetics procedures and has greatly facilitated the detailed analysis of genomes.
Gene cloning can be divided into several steps as summarized here:

1. Isolation and fragmentation of the source DNA.This can be total genomic DNA from an organism of interest, DNA synthesized from an RNA template by reverse transcriptase , a gene or genes amplified by the polymerase chain reaction , or even totally synthetic DNA made *in vitro* . If genomic 'DNA is the source, it is cut with restriction enzymes first to give a mixture of manageable-sized fragments

2. Joining the DNA fragments to a cloning vector with DNA ligase. Cloning vectors are small, independently replicating genetic elements used to replicate genes, and most are derived from plasmids or viruses. Cloning vectors are typically designed to allow *in vitro* insertion of foreign DNA at a restriction site that cuts the vector in a way that does not affect its replication . If the source DNA and the vector are cut with the same restriction enzyme, joining can be mediated by annealing of the single-stranded regions called "sticky ends" . "Blunt ends" generated by some restriction enzymes can also be joined using synthetic DNA linkers or adapters. DNA ligase is required to seal the final phosphodiester bond.

3. Introduction and maintenance of the cloned DNA in a host organism. The recombinant DNA molecule made in a test tube is introduced into a host organism, for example, by the process of transformation , where it can then replicate . Transfer of the DNA into the host usually yields a mixture of clones. Some cells contain the desired cloned gene, whereas other cells contain other clones generated by joining the source DNA to the vector. Such a mixture is known as a DNA library or a gene library because many different clones can be purified from the mixture, each containing different cloned DNA segments from the source organism, Making a gene library by cloning random fragments of a genome is called shotgun cloning, and is a widely practiced technique in gene cloning and genomic analyses.

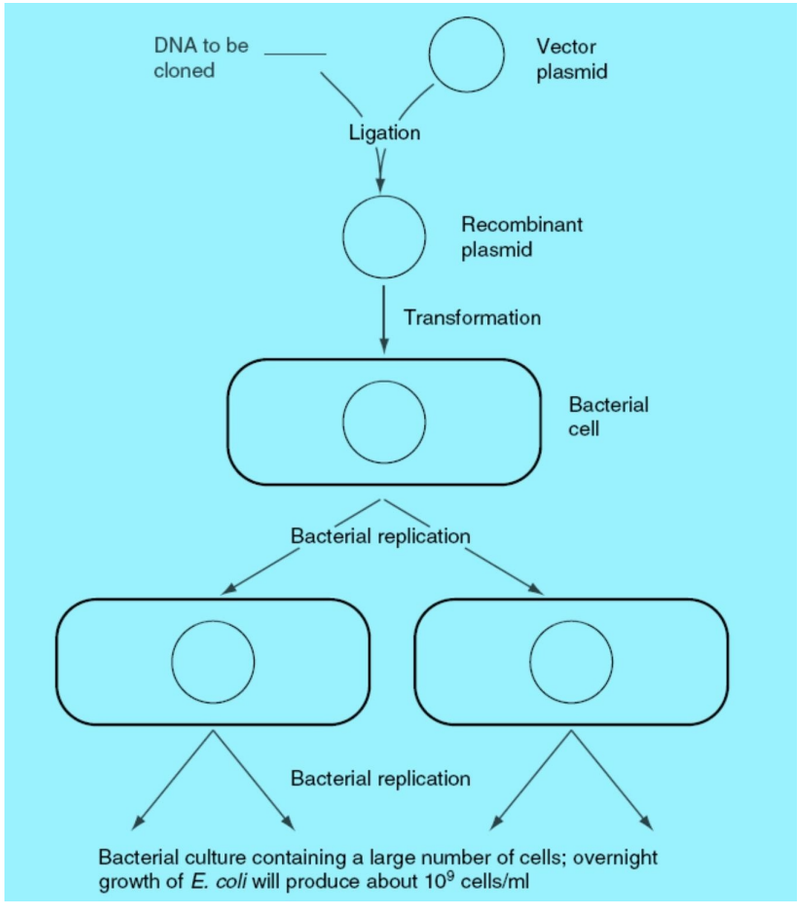


Figure 35: Basic outline of gene cloning [Dale and von Schantz,2002]

8-1 Vectors

The most commonly used vectors for gene cloning are **plasmids** . These are small circular DNA molecules found in many types of bacteria. Plasmids have an "origin of replication" which directs the replication of the plasmid and ensures that the cell contains many copies of the plasmid which are distributed between the daughter cells when the cell divides.
As long as the gene that you have cloned is part of a DNA molecule with an origin of replication, that is, cloned into a plasmid, it will also be copied when the plasmid is copied.
There are a number of other features of plasmids that are useful in gene cloning. Naturally occurring plasmids can be quite large: some are more than 100 kb in size, but the ones used in routine gene cloning tend to be less than 10 kb. This makes them easy to purify and to manipulate. Plasmids commonly used in cloning contain a selectable marker, usually an antibiotic resistance gene. This means that you can tell which bacteria contain the plasmid simply by spreading them onto an agar plate containing the antibiotic. Those that contain the plasmid will grow and eventually form a visible colony, and all the cells within that colony will carry copies of the plasmid. Any bacteria that do not contain the plasmid will be killed by the antibiotic, and so cannot give rise to a colony.
An example of one of these modified plasmid cloning vectors is pBR322, which replicates in *Escherichia coli* . **Plasmid pBR322** has a number of characteristics that make it suitableas a cloning vehicle:

1. It is relatively small, only 4361 bp.

2. It is stably maintained in its host *(Escherichia coli)* in relatively high copy number, 20-30 copies per cell.

3. It can be amplified to a very high number (1000-3000 copies per cell, about 40% of the genome) by inhibiting protein synthesis with the antibiotic chloramphenicol.

4. It is easy to isolate in the supercoiled form using a variety of routine techniques.

5. A reasonable amount of foreign DNA can be inserted into it, although inserts of more than 10 kbp lead to plasmid instability.

6. The complete base sequence of this plasmid is known, making it possible to identify all restriction enzyme cut sites.

7. There are single cleavage sites for various restriction enzymes such as *PstI, Sa*II, *EcoRI,* HindIII, and *BamHI.* It is crucial that only a single recognition site for a given restriction enzyme exist on the cloning vector so that treatment with that enzyme linearizes the vector but does not cut it into pieces. Individual sites for each of several restriction enzymes increase the versatility and usefulness of the vector.

8. It has genes conferring ampicillin resistance and tetracycline resistance on its host. These permit ready selection of hosts containing the plasmid because such hosts are resistant to both antibiotics. The sites recognized by some of the restriction enzymes are within one or the other of these resistance genes, facilitating the identification of plasmids carrying cloned DNA .

9. It can be inserted into cells easily by transformation or artificial transformation .

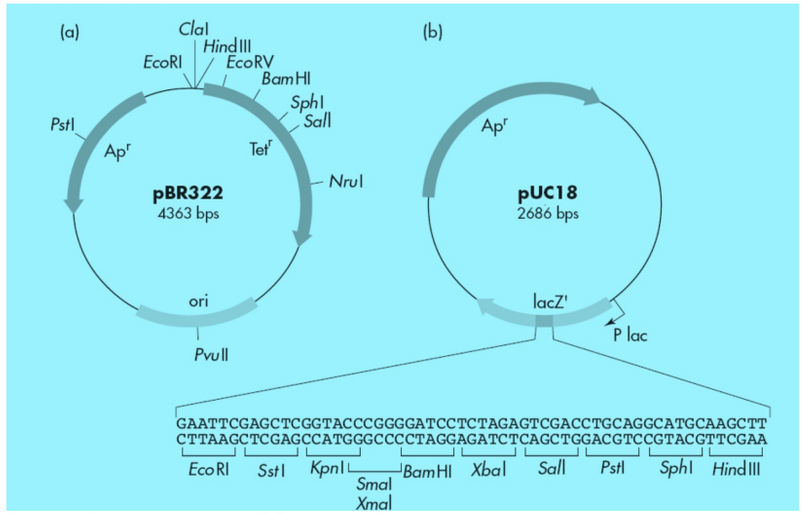


Figure 36 : a) A map of pBR322 showing the positions of the ampicillin resistance (Apr) and tetracycline resistance (*Tet*) genes b) A map of pUC18 showing the position of the ampicillin resistance [Lodge et al.,2007] .

The use of plasmid pBR322 in gene cloning : *BamHI* restriction site is within the gene for tetracycline resistance, and the *PstI* site is within the gene for ampicillin resistance. If foreign DNA is inserted into one of these sites, the antibiotic resistance conferred by the gene containing this site is lost, a phenomenon called insertional inactivation. Insertional inactivation is used to detect the presence of foreign DNA within the plasmid. Thus, when pBR322 is digested with *BamHI,* linked with foreign DNA, and transformed bacterial clones isolated, those clones that are both ampicillin resistant and tetracycline resistant lack the foreign DNA . On the other hand, those cells resistant to ampicillin but sensitive to tetracycline contain the plasmid with inserted foreign DNA. Since ampicillin resistance and tetracycline resistance can be determined independently on agar plates using replica plating , isolation of bacteria containing the desired clones and elimination of cells not containing the plasmid can readily be accomplished.

Bacteriophage Lambda as a Cloning Vector

One phage that is used as a specialized transducing phage is bacteriophage lambda . During specialized transduction lambda acts as a vector, but the recombination occurs in the cell. Lambda is a particularly useful cloning vector because its biology is well understood, it can hold larger amounts of DNA than most plasmids, and DNA can be efficiently packaged into phage particles *in vitro.* These can be used to infect suitable host cells, and infection is much more efficient than transformation (transfection). Phage lambda has a complex genetic map and a large number of genes. Of particular significance for its use as a cloning vector, however, is the fact that the central third of the lambda genome, the region between genes J and N, is unessential for infectivity and can be replaced with foreign DNA . This allows relatively large DNA fragments, up to about 20 kb, to be cloned into lambda. This is twice the cloning capacity of a small plasmid vector such as pBR322.

When the vectors replicate in a lactose-negative (Lac-) strain of *Escherichia coli,* B-galactosidase is synthesized from the phage gene and the presence of lactose positive (Lac+) plaques can be detected by using a color Indicator agar . However, if a foreign gene is inserted into the B-galactosidase gene, the Lac+ character is lost. Such Lac-plaques can be readily detected as colorless plaques among a background of colored plaques.

Cloning with lambda replacement vectors involves the following steps :

1.Isolating the vector DNA from phage particles and digestion with the appropriate restriction enzyme.

2.Connecting the two lambda fragments to fragments of foreign DNA using DNA ligase.

3. Packaging of the DNA by adding cell extracts containing the head and tail proteins and allowing the formation of viable phage particles to occur spontaneously.

4. Infecting *E. coli* and isolating phage clones by picking plaques on a host strain.

5. Checking recombinant phage for the presence of the desired foreign DNA sequence using nucleic acid hybridization procedures, DNA sequencing, or observation of genetic properties.

Cosmids

Cosmids are plasmid vectors containing foreign DNA plus the *cas* (cohesive end) site from the lambda genome . These *cas* sites are required for packaging DNA into lambda virions. Cosmids are constructed from plasmids containing cloned DNA by ligating the lambda *cas* region to the plasmid DNA. The modified plasmid can then be packaged into lambda virions *in vitro* and the phage particles used to transduce *Escherichia coli.* Cosmid construction avoids the necessity of having to transform *E. coli,* which at best is an inefficient process .

One additional advantage of cosmids is that they can be used to clone large fragments of DNA, with inserts as large as 50 kbp accepted by the system. Therefore, with bigger inserts, fewer clones are needed to obtain representation of the whole genetic element. Cosmids also permit storage of the DNA in phage particles instead of in plasmids. Phage particles are much more stable than plasmids, so the recombinant DNA can be kept for long periods of time.

8- 2 Restriction Enzymes

Cut of DNA can be done using enzymes, which are naturally produced by bacteria, and which cut DNA whenever a particular sequence of bases occurs. These are called restriction enzymes or restriction endonucleases, a name that derives from the normal function of these enzymes in the bacteria from which they are isolated . One of the most commonly used restriction enzymes is by treatment of DNA with restriction enzymes are often called restriction fragments.

Many bacteria have been examined for the presence of restriction enzymes and a large number of these enzymes have been isolated. They have a range of different properties, and recognize a wide variety of different sequences, many of which are useful in gene cloning.

8-3 DNA Ligase

If the DNA has been cut up using a restriction enzyme like *Eco*RI, which produces sticky ends, then when two molecules with the same sticky ends come into contact, hydrogen bonding between the complementary bases will cause the molecules to stick together. This is in fact why these molecules are said to have sticky ends. This is not a very stable arrangement and the two molecules will soon drift apart again. For gene cloning, you need to be able to covalently link the two molecules.

The enzyme that is capable of doing this is called DNA ligase. When two restriction fragments with sticky ends are transiently held together by hydrogen bonding there are in effect two single-stranded breaks in a double-stranded molecule ; DNA ligase repairs these single-stranded breaks. DNA ligase catalyzes the formation of a covalent phosphodiester bond between the 5′□phosphate on one DNA strand and a 3′□hydroxyl on another. This process requires energy. The most commonly used DNA ligase is a protein produced by a bacteriophage (a virus that infects bacteria) called T4. It uses ATP as an energy source.

Any molecule with an *Eco*RI sticky end can anneal to any other molecule with the same sticky end, so many fragments of genomic DNA will anneal to each other, in random order. However, these molecules will have neither an origin of replication nor an antibiotic resistance marker and so even if they can be introduced into *E. coli*, they will not lead to the formation of a colony. Another possibility is that the ends of the vector molecules may

anneal with each other either reforming the original plasmid or forming larger molecules with more than one copy of the plasmid. A third possibility, that one vector molecule will be joined to one of the genomic DNA fragments and will circularize to form a new recombinant molecule, is the desired outcome from the cloning experiment . It is possible to arrange the conditions in the ligation reaction so that the likelihood of the formation of such a recombinant molecule is favored. In a dilute solution, the chances of the two ends of the vector molecule coming into contact with each other are higher than the chances of an interaction between two different molecules. Ligation reactions are carried out at high DNA concentrations, typically with a molar ratio of 3:1 of insert to vector (i.e. three times as many insert molecules as vector molecules), to increase the likelihood of the correct recombinant being formed.

8- 4 Transformation

The final step required in gene cloning is to introduce the new recombinant plasmid into *E. coli*. This process is called transformation, and it involves two steps. First, we need to get the DNA into the bacterial cell and then, because this is an inefficient process, we need to select those cells which contain the plasmid. Some species of bacteria such as *Neisseria gonorrhoea* naturally take up DNA from their environment, and they are described as being naturally competent for transformation. *E. coli*, however, is not naturally competent, and *E. coli* cells need to be treated in a special way to enable them to take up DNA. There are two basic methods for introducing DNA into *E. coli*: chemical treatment and electroporation.

Chemical treatment for the preparation of competent *E. coli*

To prepare competent *E. coli* a culture is grown and then harvested when it is in log phase, at which stage the bacteria are dividing rapidly. The cells are harvested by centrifugation and washed several times in a chilled buffer containing divalent cations, typically CaCl2. The bacteria are finally suspended in a small volume of the buffer so that they are present at a high density. To introduce DNA (for instance, a recombinant plasmid molecule) into these cells, a small sample of this suspension is mixed and incubated on ice with the ligation mixture; it is then heat-shocked at 42°C for about 1 min This technique has been widely used for many years, although the precise mechanism by which it causes *E. coli* cells to take up DNA is only now

becoming clear. During transformation the DNA associates with the lipopolysaccharide on the outer surface of the competent cells, uptake of this DNA is associated with damage to the cell walls caused, in part, by the Ca2+ ions and with the heat shock.

Electroporation

An alternative to chemical treatment of *E. coli* to make competent cells is called electroporation. In this process the bacteria are harvested and washed as before but in cold distilled water or a buffer with a very low ionic strength. A small sample of a dense suspension of these bacteria is then mixed with the DNA in a special cuvette and a short pulse of a very high voltage current is passed through the bacterial suspension. Again, it is not entirely clear why this technique works but it is thought that the high voltage pulse makes the bacterial membrane more permeable and possibly moves the DNA into the bacterium by a process akin to electrophoresis . Electroporation is a much more efficient process than chemical transformation and it can be used with bacteria other than *E. coli*. Electroporation can also be used to transform mammalian and plant cells Unfortunately, transformation, even by electroporation, is a very inefficient process: only about one in a million of the bacterial cells will successfully take up the plasmids. This is one of the reasons why it is important to have a selectable marker on cloning vectors. Whichever of the two methods above is used, after the DNA has been introduced into the cells bacterial growth medium is then added to the sample and the culture allowed to recover at 37°C for 30 to 60 min. During this time expression of the antibiotic resistance gene will begin. The culture is then spread onto agar plates, containing an antibiotic, and incubated at 37°C until colonies have formed: typically, this will be overnight. Only those bacteria which have been transformed (i.e. have taken up the plasmid DNA) and which are expressing the antibiotic resistance marker will be able to grow on the agar containing the antibiotic.

After incubation at 37°C for a suitable length of time, colonies will be seen on the agar plate. Each of these colonies results from a single bacterium that has divided many times. Because each colony results from a single bacterium and has the same genetic makeup it is called a clone . If a given colony results from a single *E. coli* transformed with a recombinant plasmid, each individual bacterium in the colony will contain a copy of the same plasmid.

It is often necessary to purify plasmid DNA from *E. coli*. For example, after doing a ligation and transformation experiment, you will need to check the plasmid to ensure that it contains the DNA fragment that you are interested in. Further manipulations of the plasmid DNA often follow a successful cloning experiment, and these will also require plasmid purification.

Fortunately, the purification of plasmid DNA from bacteria is a routine procedure. It involves growing a culture of cells from a single colony containing the plasmid, harvesting these cells and then breaking them open (referred to as "cell lysis"), removing non-nucleic acid components, and then selectively recovering the nucleic acid by precipitation with ethanol. For some purposes, it is also necessary to separate the plasmid DNA from other nucleic acids (RNA and chromosomal DNA). Traditionally, this second step was achieved by centrifugation on a cesium chloride gradient in the presence of ethidium bromide, a technique sometimes still used for the purification of chromosomal DNA. One of the most straightforward and reliable techniques for purification of plasmid DNA is called alkaline lysis. This technique can be used as a quick method for isolating plasmid DNA for analytical purposes or can be refined to produce large amounts of high quality DNA for cloning.

8 – 5 Measuring the Size of DNA Fragments

The most widely used technique that allows you measure the size of DNA fragments is called gel electrophoresis. In this technique, DNA molecules are separated according to their size in a gel, and actual size is estimated by comparison with marker DNAs of known size. The main material used for gels is called agarose, although polyacrylamide is also often used, especially when the DNA fragments are small. DNA molecules are charged at neutral pH because each phosphate in the DNA backbone contributes one negative charge. In solution, if an electric current is applied, they will move towards the positive electrode or anode. However, because the charge to mass ratio is the same for all DNA molecules, this will not result in separation of different size fragments. However, if instead of a solution the current is applied to a gel with the DNA molecules at one end, the gel will impede the movement of the molecules. Under these conditions the size of the DNA molecules becomes the most important factor as they move through the gel. A gel is essentially a tangled network of polymeric molecules with a series of pores in it; smaller molecules are able to thread their way through the gel at a faster rate than larger molecules so the molecules will be separated according to their relative sizes .

Agarose gels are used for separating DNA fragments from 100 bp to 20 kb. Agarose, a carbohydrate polymer, is a highly purified form of agar; it is supplied as a powder. The powder is dissolved in a buffer containing ionic compounds, by heating to boiling point. The molten agarose solution is poured into a gel-forming tray and a comb inserted . As the agar cools it forms a gel. The comb is then removed to reveal wells in the gel into which samples can be loaded. The gel is placed into a horizontal electrophoresis tank and submerged in buffer. DNA samples are heated briefly in loading buffer to ensure that there are no intermolecular interactions. Loading buffer contains dye, which helps with monitoring the progress of electrophoresis, and a dense liquid such as glycerol, which ensures that the sample settles into the bottom of the well.

The DNA samples are loaded into the wells and a current applied. The DNA molecules thread their way through the gel and are separated according to size. The DNA is visualized by staining with ethidium bromide, which fluoresces under ultraviolet light. This dye intercalates between the bases of DNA, which concentrates it in the gel where the DNA is. As a result these bands fluoresce brightly under ultraviolet illumination. This staining technique is very sensitive . Despite the fact that ethidium bromide is a mutagen and ultraviolet light can cause severe burns this is still the most commonly used technique for visualization of DNA.

The distance that the band has migrated from the well is a measure of the size of the DNA fragments. Gels are calibrated by running a sample of DNA, which contains a series of restriction fragments of known size, in parallel to the experimental samples. A commonly used DNA size marker is DNA from the bacteriophage lambda, cut with *Hin*dIII, which gives a recognizable pattern of eight bands ranging in size from 125 bp to 23.13 kb . The relationship between the size of the DNA fragments and the distance the bands migrate is not a linear one; rather, the distance migrated through the gel is proportional to the log10 of the molecular weight. If you plot the log10 of the size of the DNA fragments against the distance migrated, you will obtain a straight line (with some deviation at the extreme ends of the size range, owing to other factors).

9 Basic Molecular Biology Tools

9-1 The Polymerase Chain Reaction

There have been a number of key developments in molecular biology techniques, but one that has had the most impact in recent years has been the polymerase chain reaction, or PCR. One of the reasons for the adoption of the PCR is the elegant simplicity of the reaction and relative ease of the practical steps? It is frequently one of the first techniques used when analysing DNA, and it has opened up the analysis of cellular and molecular processes to those outside the field of molecular biology.

The PCR is used to amplify a precise fragment of DNA from a complex mixture of starting material, usually termed the template DNA, and in many cases requires little DNA purification. It does require knowledge of the DNA sequence information which flanks the fragment of DNA to be amplified (target DNA). From this information two oligonucleotide primers can be chemically synthesized, each complementary to a stretch of DNA to the 3' side of the target DNA, one oligonucleotide for each of the two DNA strands .

Further reagents required for the PCR include a DNA polymerase, each of the four nucleotide dNTP building blocks of DNA in equimolar amounts (50-200 pM) and a buffer appropriate for the enzyme. This is usually optimized for Mg^{+2} concentration (0.5-5 mM). This is a critical component of the PCR, affecting not only the enzyme, but also primer/template binding and effective incorporation of dNTPs which together form a soluble complex. The PCR may be thought of as a technique analogous to the DNA replication process that takes place in cells since the outcome is the same - the generation of new complementary DNA stretches based upon the existing ones. It is also a technique that has, in many cases, replaced the traditional DNA cloning methods since it fulfills the same function, the production of large amounts of DNA from limited starting material, but achieves this in a fraction of the time needed to clone a DNA fragment. Although not without its drawbacks, the PCR is a remarkable development which is changing the approach of many scientists to the analysis of nucleic acids and continues to have a profound impact on core biosciences and biotechnology.

Stages and Components of the PCR

The PCR consists of three defined sets of times and temperatures termed steps: (i) denaturation, (ii) annealing and (iii) extension. Each of these steps is repeated 30-40 times, termed cycles. In the first cycle the double-stranded template DNA is denatured by heating the reaction to above 90°C. The region within the complex DNA which is to be specifically amplified (target DNA) is made accessible. The reaction mixture is then cooled to 40°C. The precise temperature is critical and each PCR system has to be defined and optimized. Reactions that are not optimized may give rise to other DNA products in addition to the specific target, or may not produce any amplified products at all. The annealing step allows the hybridization of the two oligonucleotide primers, which are present in excess, to bind to their complementary sites that flank the target DNA. The annealed oligonucleotides act as primers for DNA synthesis, since they provide a free 3' hydroxyl group for DNA polymerase. The DNA synthesis step is termed extension and is carried out by a thermostable DNA polymerase, most commonly *Taq* DNA polymerase. DNA synthesis proceeds from both 3' ends of the primers until the new strands have been extended along and beyond the target DNA. It is important to note that, since the new strands extend beyond the target DNA they will contain a region near their 3' ends which is complementary to the other primer. Thus, if another round of DNA synthesis is allowed to take place, the new strands, as well as the original strands, will be used as templates. Most interestingly, the products obtained from the new strands will have a precise length, delimited exactly by the two regions complementary to the primers.

As the system is taken through successive cycles of denaturation, annealing and extension all the new strands will act as templates and so there will be an exponential increase in the amount of DNA produced. The net effect is to selectively amplify the target DNA and the primer regions flanking it. One problem with early PCR reactions was that the temperature needed to denature the DNA also denatured the DNA polymerase.

However the availability of a thermostable DNA polymerase enzyme isolated from the thermophilic bacterium *Thermus aquaticus* found in hot springs provided the means to automate the reaction. *Taq* DNApolymerase has a temperature optimum of 72°C and survives prolonged exposure to temperatures as high as 96°C and so is still active after each of the denaturation steps. The widespread utility of the technique is also due to the ability to automate the reaction and, as such, many thermal cyclers have

been produced in which it is possible to program in the temperatures and times for a particular PCR reaction.

A number of thermostable DNA polymerases other than *Thermus aquaticus* have been discovered and marketed commercially. Those isolated from *Thermococcus litoralis* (VentTM), found in deep ocean floors, are highly thermostable and capable of extending templates in excess of 12 Kb pairs; they also have proofreading ability and so have a high degree of fidelity in comparison to *Taq* DNA polymerases. Further manipulation of VentTM has resulted in derivatives with greater thermostability (Deep-VentTM) and a derivative lacking the exonuclease component (ex0 -). *Pfu* DNA polymerase isolated from a marine bacterium also has proofreading activity and incorporates radio labelled nucleotides and analogues efficiently. Hence it is particularly useful when producing radio labelled gene probes or performing techniques such as cycle sequencing. One minor problem with these polymerases and another DNA polymerase isolated from *Thermotoga maritima* (UITmaTM) is their exonuclease activity which has been reported to cause modification and degradation of primers under initial sub-optimal conditions. This may be overcome by using various exonuclease deficient forms (ex0 -) of the enzyme.

One further interesting thermostable DNA polymerase is that isolated from *Thermus thermophilus (Tth)* which at 70°C and in the presence of Mn2 is able to carry out reverse transcription reactions. Following cDNA synthesis and chelation of Mn2+ the polymerase is able to carry out polymerization of the template. This dual activity of *Tth* DNA polymerase allows RNA-PCR to be carried out in a single tube and obviates the need for a separate cDNA synthesis reaction.

Primer Design in the PCR

The key to the PCR lies in the design of the two oligonucleotide primers. These have to not only be complementary to sequences flanking the target DNA but also must not be self-complementary or bind each other to form dimers since both prevent DNA amplification. They also have to be matched in their GC content and have similar annealing temperatures, usually at a concentration of I pM. The increasing use of information from the internet and the sequences held in gene databases are useful starting points when designing primers and reaction conditions for the PCR. A number of software packages such as Oligo, Primer, *etc.,* have allowed the process of primer design to be less troublesome. It is also possible to include more than

one set of primers in a PCR. This method, termed multiplex PCR, allows the amplification of more than one product in a single reaction tube and is especially useful in molecular diagnosis of clinical disorders.
It is also possible to design primers with additional sequences at their *5′* end such as restriction endonuclease target sites or promoter sequences. This allows further manipulation such as cloning to be undertaken following amplification . However, modifications such as these require that the annealing conditions be altered to compensate for the areas of non-homology in the primers. In general primers are usually designed to be between 20 and 30 bases in length. It is best to balance the melting temperature of the primer pair and to have a GC content of 40-60%. A number of factors are best avoided in primer sequences, such as the potential for secondary structure formation, primer complementarily or dimer formation and runs of purines, pyrimidines or repetitive motifs. The precise times and temperatures for a PCR may be calculated manually or by computer programs, however in practice these are usually a guide in determining the final optimal parameters for amplification.
One useful technique which may be adopted in order to amplify the desired product rather than unwanted ones is touchdown PCR. Here a thermal cycler is programmed to lower the annealing temperature incrementally from a value above the expected melting temperature to a value below it. This allows optimal hybridization of the primers to the target and thus leads to amplification of the correct products over any undesired ones.
A number of PCR methods have been developed where either one or both of the primers are short and random in sequence.This gives rise to arbitrary priming in genomic templates but interestingly may give rise to discrete banding patterns when analysed by gel electrophoresis. In many cases this technique may be used reproducibly to identify a particular organism or species. This is sometimes referred to as rapid amplification of polymorphic DNA (RAPDs) and has been used successfully in the detection and differentiation of a number of pathogenic strains of bacteria.
The PCR may be used to amplify DNA from a variety of sources or templates. It is also a highly sensitive technique and requires only one or two molecules for successful amplification. Unlike many manipulation methods used in current molecular biology the PCR technique is sensitive enough to require very little template preparation. Indeed the extraction of DNA from many prokaryotic and eukaryotic cells may involve a simple boiling step. However, the components of many DNA extraction techniques such as SDS, phenol, ethanol and proteinase K may adversely affect the PCR at certain concentration. In contrast, a number of reagents have been shown to

improve the efficiency of amplification. In particular, the inclusion of tetramethyl ammonium chloride (TMAC) appears to improve the specificity of primer annealing whilst betaine binds and stabilizes AT sequences but destabilizes GC regions and is therefore useful in the amplification of GC-rich sequences.
Nucleotide analogues such as 7-deaza-2'-deoxyguanosine triphosphate are also a useful addition when amplifying sequences with the potential to form secondary structures.
The PCR may also be used to amplify RNA, a process termed RTPCR (reverse transcriptase-PCR) . Initially a reverse transcription reaction which converts mRNA to cDNA is first carried out. This reaction normally involves the use of the enzyme reverse transcriptase although some thermostable DNA polymerases used in the PCR, *e.g. Tth,* have a reverse transcriptase activity under certain buffer conditions. This allows mRNA transcription products to be effectively analysed. It may be also be used to differentiate latent viruses (detected by standard PCR) or active viruses which replicate and thus produce transcription products which are detectable by RT-PCR. In addition, the PCR may be extended to determine relative amounts of a transcription product.

Sensitivity of the PCR

The enormous sensitivity of the PCR system is also one of its main drawbacks since the very large degree of amplification makes the system vulnerable to contamination. Even a trace of foreign DNA, such as that contained in dust particles, may be amplified to significant levels and may give misleading results. Hence cleanliness is paramount when carrying out PCR and dedicated equipment and, in some cases, laboratory areas dedicated to pre- and post-PCR stages are used. It is possible that amplified products may also contaminate the PCR although this may be overcome by UV irradiation to degrade the already amplified products so that they cannot be used as templates. A further interesting solution is to introduce uracil into the PCR by incorporating dUTP. Following analysis, the PCR products are treated with the enzyme uracil-N-glycosylase (UNG) which degrades the N-glycosidic bond in DNA containing uracil, and PCR products are fragmented and rendered useless as templates.

9 - 2 DNA sequencing

The ability to determine the sequence of bases in DNA is a central part of modern molecular biology and provides what might be considered the ultimate structural information. There are two main methods for sequencing DNA. In one method, developed by Allan Maxam and Walter Gilbert, chemicals are used to cleave the DNA at certain positions, generating a set of fragments that differ by one nucleotide. The same result is achieved in a different way in the second method, developed by Fred Sanger and Alan Coulson, which involves enzymatic synthesis of DNA strands that terminate in a modified nucleotide. Analysis of fragments is similar for both methods and involves gel electrophoresis and autoradiography (assuming that a radioactive label has been used). The enzymatic method (and variants of the basic technique) has now almost completely replaced the chemical method as the technique of choice, although there are some situations where chemical sequencing can provide useful data to confirm information generated by the enzymatic method.

9- 2-1 Maxam–Gilbert (chemical) sequencing

The Maxam-Gilbert Method, developed in the late 1970's, was the first method to determine the sequence of a DNA molecule of up to 500 bp. It involves chemical cleavage at different specific nucleotides of four samples of an end-labeled DNA restriction fragment. A defined fragment of DNA is required as the starting material. This need not be cloned in a plasmid vector, so the technique is applicable to any DNA fragment. The DNA is radiolabelled with ^{32} P at the 5 $^{-}$ ends of each strand, and the strands are denatured, separated, and purified to give a population of labelled strands for the sequencing reactions.

The next step is a chemical modification of the bases in the DNA strand. This is done in a series of four or five reactions with different specificities, and the reaction conditions are chosen so that, on average, only one modification will be introduced into each copy of the DNA molecule. The modified bases are then removed from their sugar groups and the strands cleaved at these positions using the chemical piperidine. The theory is that, given the large number of molecules and the different reactions, this process will produce a set of nested fragments.

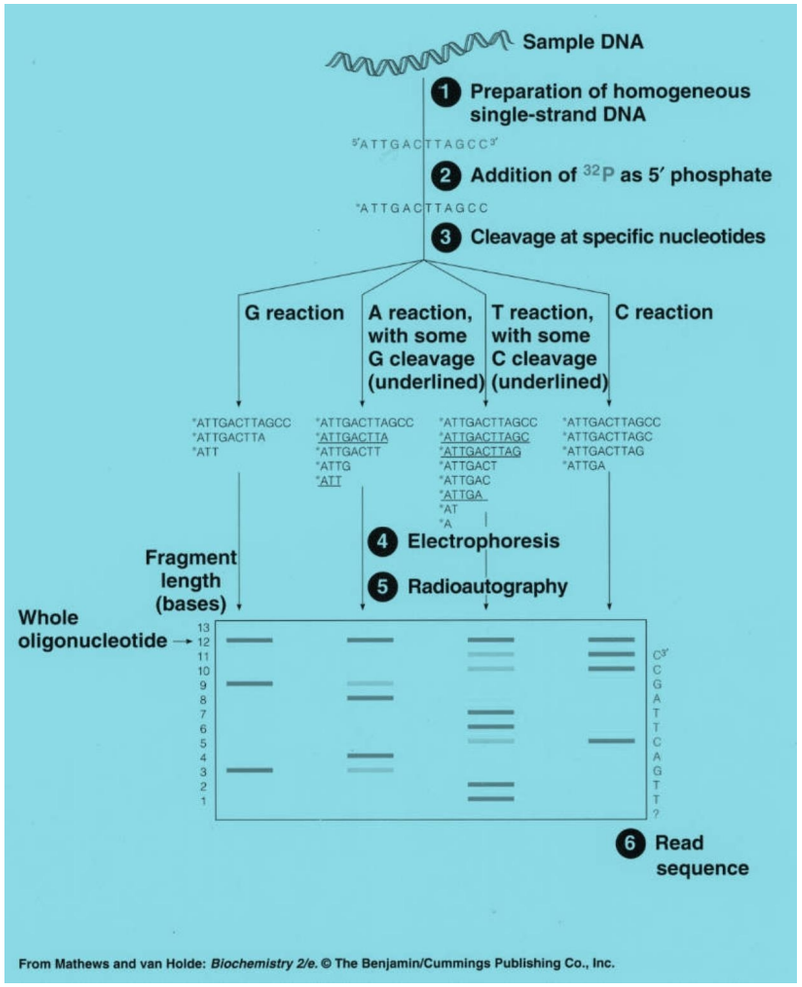


Figure 37 : The double-stranded fragment to be sequenced is labeled at the 5 ends . The label is removed from one end, and the fragment then is denatured. Four identical samples of the prepared fragment are subjected to four different sets of chemical reactions that selectively cut the DNA backbone at G, G + A, C + T, or C residues. The reactions are controlled so that each labeled chain is likely to be broken only once.[Agarwal ,2003].

9- 2- 2 Sanger–Coulson (dideoxy or enzymatic) sequencing

Although the end result is similar to that attained by the chemical method, the Sanger--Coulson procedure is totally different from that of Maxam and Gilbert. In this case a copy of the DNA to be sequenced is made by the Klenow fragment of DNA polymerase .
The template for this reaction is single-stranded DNA, and a primer must be used to provide the 3_ terminus for DNA polymerase to begin synthesising the copy . The production of nested fragments is achieved by the incorporation of a modified dNTP in each reaction. These dNTPs lack a hydroxyl group at the 3_ position of deoxyribose, which is necessary for chain elongation to proceed. Such modified dNTPs are known as **dideoxynucleoside triphosphates** (ddNTPs). The four ddNTPs (A, G, T, and C forms) are included in a series of four reactions, each of which contains the four normal dNTPs. The concentration of the dideoxy form is such that it will be incorporated into the growing DNA chain infrequently. Each reaction, therefore, produces a series of fragments terminating at a specific nucleotide, and the four reactions together provide a set of nested fragments. The DNA chain is labelled by including a radioactive dNTP in the reaction mixture. This is usually [α-35S]dATP, which enables more sequence to be read from a single gel than the 32P-labelled dNTPs that were used previously.
The generation of fragments for dideoxy sequencing is more complicated than for chemical sequencing and usually involves subcloning into different vectors. Many plasmid vectors are now available , and some types can be used directly for DNAsequencing experiments. Another method is to clone the DNA into a vector such as the bacteriophage M13 , which produces single-stranded DNA during infection. This provides a suitable substrate for the sequencing reactions.

9 - 2 -3 Electrophoresis and reading of sequences

Separation of the DNA fragments created in sequencing reactions is achieved by PAGE. For the standard lab procedure (small-scale nonautomated), a single gel system is used. The gels usually contain 6--20% polyacrylamide and 7 M urea, which acts as a denaturant to reduce the effects of DNA secondary structure. This is important because fragments that differ in length by only one base are being separated. The gels are very thin (0.5 mm or less) and are run at high-power settings, which causes them to heat up to 60--70◦C. This also helps to maintain denaturing conditions.

Sometimes two lots of samples are loaded onto the same gel at different times to maximise the amount of sequence information obtained.
After the gel has been run, it is removed from the apparatus and may be dried onto a paper sheet to facilitate handling. It is then exposed to X-ray film. The emissions from the radioactive label sensitize the silver grains, which turn black when the film is developed and fixed . Reading the autoradiograph is straight forward the sequence is read from the smallest fragment upwards . Using this method, sequences of up to several hundred bases may be read from single gels. The sequence data are then compiled and studied using a computer, which can perform analyses such as translation into amino acid sequences and identification of restriction sites, regions of sequence homology, and other structural motifs such as promoters and control regions.

Automation of DNA sequencing

One of the major advances in technology that enabled sequencing to move from single-gel lab-based systems up to large-scale ‘production Whereas a good lab scientist or technician could sequence maybe a few hundred bases per day, this was not going to solve the problem of determining *genome* sequences as opposed to *gene* sequences. Improving the technology by orders of magnitude was required. This was achieved by improvements in sample preparation and handling, with robotic processing enabling high-volume throughput. In a similar way the automation of the sequencing reactions, and linear continuous capillary electrophoresis techniques, enabled scale-up of the sequence-determination stage of the process.
In addition to the challenges of sequence determination, a parallel challenge lay in the need to develop sufficient computing power to deal with the vast amounts of data generated by the newly improved technologies employed in genome sequencing. In fact, it could be argued that the most critical part of the whole process is the data analysis side of things without the ability to interrogate sequence data, the sequence remains essentially silent.

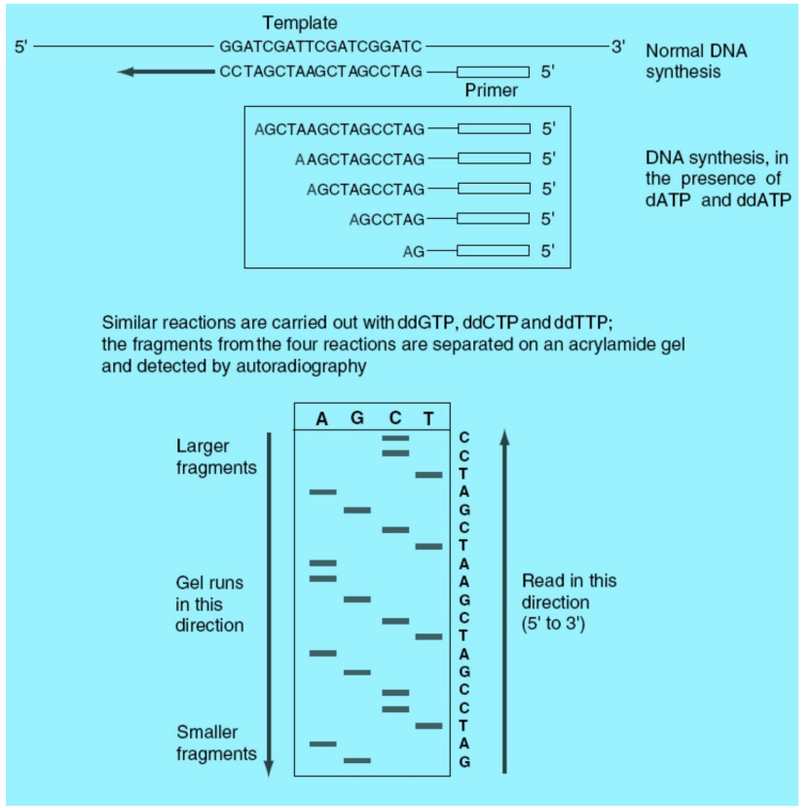


Figure 38 : Determination of DNA Sequencing [Dale and von Schantz ,2002]

9 - 3 Hybridization

Hybridization is a phenomenon in which single-stranded DNA or RNA molecules anneal to complementary DNA or RNA. Though a double-stranded DNA sequence is generally stable under physiological conditions, changing these conditions in the laboratory (generally by raising the surrounding temperature) will cause the molecules to separate into single strands. These strands are complementary to each other but may also be complementary to other sequences present in their surroundings. Lowering the surrounding temperature allows the single-stranded molecules to anneal or "hybridize" to each other.

DNA replication and transcription of DNA into RNA both rely upon nucleotide hybridization, as do molecular biology techniques including Southern blots and Northern blots, the polymerase chain reaction (PCR), and most approaches to DNA sequencing.

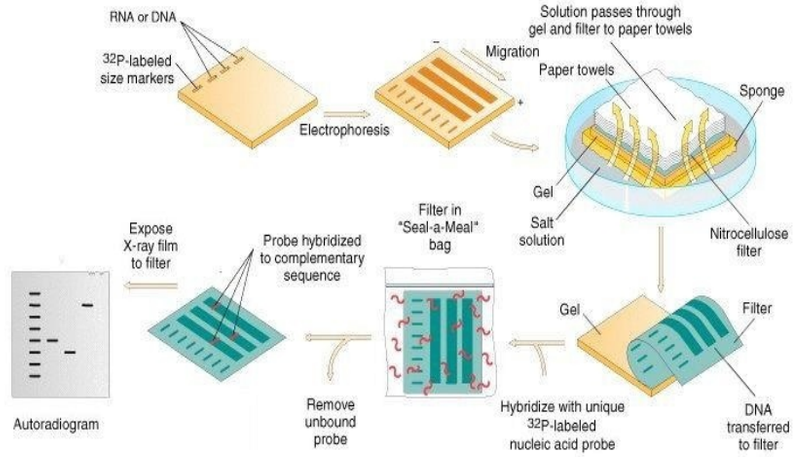
nucleic acid hybridization can be used as an extremely sensitive detection method, capable of picking out specific DNA sequences from complex mixtures. Usually a single pure sequence is labeled with ^{32}P and used as a probe. The probe is denatured before use so that the strands are free to base-pair with their complements. The DNA to be probed is also denatured and is usually fixed to a supporting membrane made from nitrocellulose or nylon. Hybridization is carried out in a sealed plastic bag or tube at 65-68◦C for several hours to allow the duplexes to form. The excess probe is then washed off and the degree of hybridization can be monitored by counting the sample in a scintillation spectrometer or by preparing an autoradiograph, where the sample is exposed to X-ray film.

Southern and Northern blotting technique

Filter hybridization is a method used to determine the presence of homologous or complementary sequences in DNA or RNA, by allowing a denatured radio-labeled 'probe' nucleic acid (in solution) to anneal to the denatured nucleic acid (immobilized to a nitrocellulose filter) to be tested. Using this specific filters containing the immobilized nucleic acid by a technique known as Southern blots, Northern blots. The Southern Blot technique is useful for identifying a DNA sequence that appears only once or twice in the genome, DNA is applied to an agarose gel,

and electrophoresis separates the fragments of DNA according to size. The gel is then placed on a thin sponge wick resting in a dish of salt solution, and a special filter (typically nitrocellulose) is placed on top of the gel. A stack of absorbent material (typically paper towels) is placed on top of this stack. The absorbent material draws the salt solution from the dish into the wick and through the gel by capillary action, which transfers the DNA fragments into the filter. The procedure is called a "Southern transfer" after the filter now contains the DNA fragments in the same pattern as the gel,

The filter is placed in a standard solution of radioactively-labelled DNA probe for a particular gene sequence. The probe binds to the filter only where a complementary DNA sequence is located. After washing to remove unbound probe, a piece of X-ray film is placed over the hybridized filter and left for several hours to several days. The radioactive label produces a black band on the film where it has stuck to the complementary DNA, producing an autoradiogram. If a labelled size marker has been used, the exact sizes of the fragments can be determine. The steps of southern and northern blotting is the same only DNA samples is replaced with RNA in the second procedure



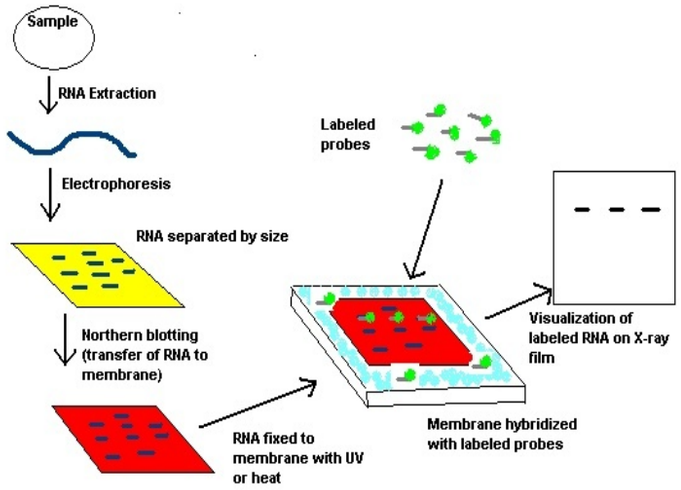


Figure 39: Analysis of DNA by the Southern Blot technique(RNA by northern blotting technique)

10 Gene Mapping in Bacteria

One of the main objectives in genetic analysis is the determination of the position of genes on the chromosome. In bacteria, the classical methods of gene mapping depend on the production of recombinants by gene transfer using conjugation, transformation and transduction.
These methods have now been supplemented, although not entirely supplanted, by methods based on *in vitro* gene technology. Nevertheless, a basic understanding of these methods is valuable for an appreciation of the development of our knowledge of bacterial genetics.

The three mechanisms of genetic exchange transformation, transduction, and conjugation- together with the powerful tools of gene cloning, can be used to map the locations of genes on a bacterial chromosome.
Genes in *Escherichia coli* were initially mapped using conjugation. By using Hfr strains with origins at different sites , it was possible to map the entire *E. coli* genome. However, conjugation does not permit ordering genes that are closely linked. Therefore, generalized transduction was used for fine-structure mapping of the *E. coli* chromosome. Bacteriophage PI can carry small fragments of DNA and proved very useful for detailed mapping of genes in *E. coli.*
The genetic techniques used to map the chromosomes of other prokaryotes were dictated by the efficiency with which genetic transfer occurs in the particular organism. For example, transformation is a very inefficient process in *E. coli* but is a very efficient process in gram-positive bacteria . Thus, transformation proved an effective tool for mapping genes in *Bacillus.*
In addition to conjugation, transduction was also used to map genes in the *E. coli* chromosome. However, because transduction only involves the transfer of a few genes-many fewer than is possible using conjugation-transduction was only useful for fine structure mapping purposes.

The genetic map of *Escherichia coli* strain K-12 . The map distances are given in "minutes" of transfer, with the entire chromosome containing 100 minutes and with "zero time" arbitrarily set as that at which the first genetic transfer (the *threonine* operon) can be detected using the first Hfr strain of *E. coli* that was discovered. The map contains only a few of the many thousands of genes present in the *E. coli* chromosome and shows the location of a few restriction enzyme recognition sites. The size of the chromosome is given in both minutes (the original units determined by

conjugation studies) and in kilo base pairs. The *E. coli* chromosome, like that of many other prokaryotes, has been completely sequenced. Sequencing of relatively small bacterial chromosomes of 3 or 4 mega bases is now done routinely by automated sequencing of random, or shotgun, clones. Because of the enormous amount of genetic information genomic analyses makes available, this information can only be managed through computer databases. This reality has spawned a whole new field of molecular biology called bioinformatics.

In addition to its important role as a model organism, *E. coli* continues to be the organism of choice for both research and applications of genetic engineering.

The strain of *E. coli* whose chromosome was originally sequenced, strain MG1655, is a derivative of *E. coli* K-12, the traditional strain used for genetic studies. The wild-type E. *coli* K-12 is a lysogen of bacteriophage lambda and also contains the F plasmid. However, strain MG1655 had been "cured" of lambda (by radiation) and of the F plasmid before sequencing began. The chromosome of strain MG1655 contains 4,639,221 bp. Analysis of the genomic sequence showed there to be 4288 possibly functional openreading frames (ORFs), which account for about 88% of the genome. Approximately1% of the genome is composed of genes encoding tRNAs and rRNAs, and about 0.5% is noncoding, repetitive sequences. The remaining 10% of the genome contains all of the regulatory sequences: promoters, operators, origin and terminus of DNA replication.

Mapping with Hfr strains

Integration of the F plasmid into the *E. coli* chromosome produces an Hfr strain which is capable of transferring a copy of the chromosome to a suitable recipient. Transfer of the whole chromosome would take about 100 min. For this reason, the *E. coli* genetic map is calibrated from 0 to 100 min, with each gene being assigned a position that corresponds to the time at which it is transferred from an arbitrary origin at the threonine locus (thr, 0 min) with transfer proceeding in a clockwise direction. The actual time at which transfer of a specific gene occurs and the direction of transfer will depend on the Hfr strain used, since the F plasmid can be integrated at different points and/or in a different orientation. However, it is quite rare for the complete chromosome to be transferred. The mating pairs will tend to become separated at randomly distributed times. The longer it takes for

transfer of a gene, the more chance there is that the mating pair will have separated before that gene is transferred. There will therefore be a gradient of transfer corresponding to the position of the genes with respect to the point at which transfer starts.

This provided a convenient way of determining the relative position of genes on the *E. coli* chromosome . If a prototrophic Hfr strain is mated with a multiply auxotrophic recipient (e.g. thr leu trp his arg), the number of recipients that have received each of the markers can be determined by plating aliquots of the mixture on a minimal medium supplemented with four out of the five amino acids. For example, the number of thr recombinants is measured using a medium that contains leucine, tryptophan, histidine and arginine, but not threonine. It is of course necessary to prevent growth of the prototrophic donor, for example by using a streptomycin-resistant recipient and including streptomycin in the medium. Streptomycin in this instance is used as a counter selecting agent. On this medium, the donor will be unable to grow (because of the streptomycin) and the parental recipient will not grow (because of the absence of threonine). The only cells that can grow will be the recombinant recipients that have received the *thr* gene.

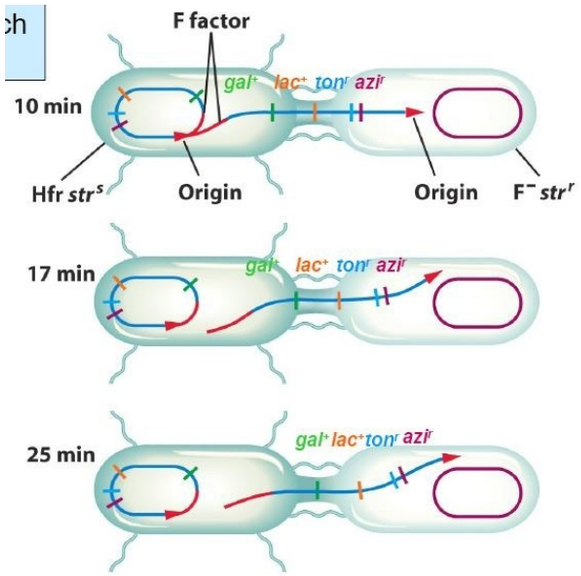


Figure 40 : Transfer of the chromosome by Hfr during conjugation

In this instance, the Hfr H donor has been used, from which the genes are transferred in a clockwise direction starting very close to the thr locus. There is a linear relationship between the logarithm of the number of recombinants and the map position of the genes concerned. If it is assumed that the position of the *trp* gene is not known, determining the number of Trp recombinants will allow the gene to be mapped.

An alternative method for more accurate mapping of genes that are transferred relatively early in mating involves deliberately separating the mating pairs (by violent agitation) in samples of the mixture at different times after the start of mating (interrupted mating). Recombinants that have received a specific gene start to appear at a certain time after the start of mating (the time of entry), which is a measure of the distance of that gene from the origin of transfer.

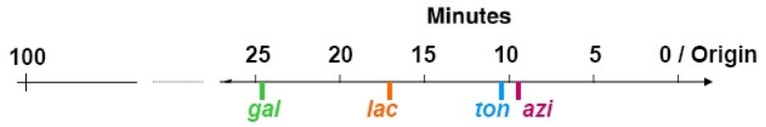


Figure 41 : Genetic map of the genes in bacteria

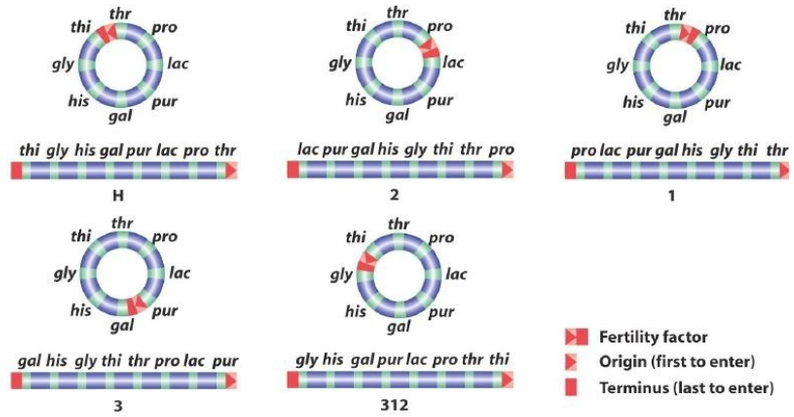


Figure 41 : The order of gene transfer is determined by polarity and site of integration of F-factor in the chromosome…. in each Hfr cell line the polarity and the site are different •hence the exact gene transfer order are not the same •but the relative gene order is the same as in the original chromosome

Co-transformation and co-transduction

Transformation can be used for mapping the relative positions of genes, by selecting recombinants in which one marker has been transferred and then determining the frequency of recombination for a second marker. If the two are close together, they will tend to be inherited together.

If the two markers are labelled A and B: the donor is wild type for both genes (AB), while the recipient is the double mutant (A_B_). After transformation of the recipient with chromosomal DNA extracted from the donor, the cells are plated on a medium that only allows A cells to grow. The colonies that result from this (transformants) can then be tested for the presence of the B gene (the unselected marker). Two recombination events (crossovers) are needed to incorporate a piece of linear DNA into the chromosome. Since transformants that have received gene A have been selected . If A and B are close together, it is unlikely that the second crossover would occur in the short region between them, so both markers will be incorporated together (co-transformation). The further apart the two markers are, the more likely it is that the second crossover will happen in the intervening region and hence the frequency of co-transformation will be lower.

Generalized transduction can be used in a similar way and provides a powerful tool for the short range mapping of the position of genes on the bacterial chromosome. As an example, if the phage P1 is grown on a prototrophic *E. coli* strain, the phage preparation can be used to infect cells of a recipient strain which is auxotrophic for threonine and proline. Plating these cells on a medium that contains threonine (but not proline) will detect transductants that have received the pro gene from the donor by transduction. These cells can then be tested for their ability to grow in the absence of threonine, i.e. whether they have also received the *thr* gene. A high degree of co-transduction indicates closely linked genes.

Appendix
Amino acid notations

	Three-letter notation	One letter notation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

[Dale and Park,2004]

WORKING GLOSSARY

Acriflavine : 3,6-Diamino-10-methylacridinium chloride *or* (according to some authors) a mixture of this compound and 3,6-diaminoacridine (proflavine). Acriflavine is soluble in water and in ethanol, and has been used as an ANTISEPTIC.

Adaptor : A synthetic oligodeoxyribonucleotide which is similar to a linker but which contains more than one type of restriction site and may also have pre-existing STICKY ENDS.

Allosteric effect: binding of a substance to a site on a protein can alter its conformation and hence alter the activity of a distinct site on the protein.

Amber mutation: the introduction of a stop codon (UAG) within the coding sequence of a gene which results in premature termination of translation.

Amplification: (a) increasing the copy number of a plasmid by inhibiting replication of the chromosome while allowing plasmid replication to continue; (b) an increase in the number of copies of a gene due to repeated duplication .

Analogues: unrelated genes with similar function .

Autoradiography : A process to detect radioactively labeled molecules (which usually have been separated in an *SDSPAGE* or *agarose gel*) based on their ability to create an image on photographic or X-ray film.

Annealing: formation of double-stranded DNA from single-stranded DNA .

Anticodon: the region of tRNA which pairs with the codon.

Antimetabolite: an analogue of the end-product of a pathway that causes feedback inhibitionor repression, but cannot replace the genuine product; used for selecting feedback-deficient mutants.

Antisense RNA: RNA that is complementary to the mRNA which can interfere with translation.

Anti-terminator: a protein that allows RNA polymerase to read through a terminator.
Attenuation: (a) reduced expression of those genes in an operon that are situated beyond a certain point; (b) reduction in the virulence of a pathogenic microorganism .

Autogenous control: regulation of the expression of a gene by its own product.

Auxotroph : an organism that has developed a nutritional requirement through mutation.

Back mutation: exact reversal of a mutation to restore the wild-type sequence of the gene.

Base : The purine or pyrimidine component of a nucleotide; often used to refer to a nucleotide residue within a nucleic acid chain.

Base Pair : One pair of complementary nucleotides within a duplex strand of a nucleic acid. Under Watson-Crick rules, these pairs consist of one pyrimidine and one purine: i.e.,C-G, A-T (DNA) or A-U (RNA). However, "noncanonical" base pairs (e.g., G-U) are common in RNA secondary structure.

Bacterial artificial chromosome: use of a vector based on the F plasmid to clone large DNA fragments.

Bacteriocin: a protein/polypeptide with antibiotic activity, usually against a narrow range of closely related bacteria. Usually plasmid mediated.

Bacteriophage lambda: A virus which infects *E. coli* , and which is often used in molecular genetics experiments as a vector, or cloning vehicle. Recombinant phages can be made in which certain non-essential l DNA is removed and replaced with the DNA of interest. The phage can accommodate a DNA "insert" of about 15-20 kb. Replication of that virus will thus replicate the investigator's DNA. One would use phage l rather than a plasmid if the desired piece of DNA is rather large.

Bioinformatics: computer-based analysis of biomolecular data, especially large-scale datasets derived from genome sequencing.

Binding site: A place on cellular DNA to which a protein (such as a transcription factor) can bind. Typically, binding sites might be found in the vicinity of genes, and would be involved in activating transcription of that gene (promoter elements), in enhancing the transcription of that gene (enhancer elements), or in reducing the transcription of that gene (silencers). NOTE that whether the protein in fact performs these functions may depend on some condition, such as the presence of a hormone, or the tissue in which the gene is being examined. Binding sites could also be involved in the regulation of chromosome structure or of DNA replication.

BLAST: Basic Local Alignment Search Tool; a computer program for searching sequence databases for related sequences.

Blotting: A technique for detecting one RNA within a mixture of RNAs (a Northern blot) or one type of DNA within a mixture of DNAs (a Southern blot). A blot can prove whether that one species of RNA or DNA is present, how much is there, and its approximate size. Basically, blotting involves gel electrophoresis, transfer to a blotting membrane (typically nitrocellulose or activated nylon), and incubating with a radioactive probe. Exposing the membrane to X-ray film produces darkening at a spot correlating with the position of the DNA or RNA of interest.

Box: a short sequence of bases in DNA conforming (more or less) to a consensus for that particular type of box; usually has a regulatory function, e.g. by providing a binding site for a regulatory protein.

Catabolite repression: the expression of the gene is turned off by the presence of an easily metabolized substrate such as glucose.

Cap site: Two usages: In eukaryotes, the cap site is the position in the gene at which transcription starts, and really should be called the "transcription initiation site". The first nucleotide is transcribed from this site to start the nascent RNA chain. That nucleotide becomes the 5' end of the chain, and thus the nucleotide to which the cap structure is attached . In bacteria, the CAP site is a site on the DNA to which a protein factor (the Catabolite Activated Protein) binds.

Carboxyl Terminus : Refers to the COOH end of a peptide chain (by custom drawn at the right of a protein sequence)

cDNA : Complementary DNA. A DNA molecule which was originally copied from an RNA molecule by reverse transcription. The term "cDNA" is commonly used to describe double-stranded DNA which originated from a single-stranded RNA molecule, even though only one strand of the DNA is truly complementary to the RNA.

cDNA Library : A collection of cDNA fragments, each of which has been cloned into a separate vector molecule.

Chaperones: proteins that affect the folding of other proteins or the assembly of complex structures.

Chloramphenicol: an antibiotic which interferes with protein synthesis.

Chromosome walking: a technique for identification of DNA regions adjacent to a known marker by sequential hybridization of clones.

Circular permutation: a population of linear DNA molecules carrying genes in the same order on a circular map, but starting at different positions.

Cis-acting: a control region that influences genes on the same DNA molecule only (usually adjacent genes) and has no effect on other DNA molecules
Cistron: a region of DNA that codes for a single polypeptide. No longer in common use except for emphasis .

Clonal: variation only by mutation, without horizontal gene transfer.

Clone: population of organisms descended asexually from a single individual.

Cloning vector : genetic element into which genes can be recombined and replicated

Cloning: obtaining an homogeneous population of cells by repeated single colony isolation.

Codon usage: a measure of the relative use of synonymous codons (i.e. different triplets coding for the same amino acid).

Codon: A group of three bases in mRNA that codes for a single amino acid in a polypeptide chain.

Cohesive ends: Single-stranded ends on DNA molecules that have complementary sequences.

Cointegrate: plasmid formed by the complete fusion of two smaller plasmids.

Colicin: a bacteriocin produced by a strain of *E. coli.*

Competent: a bacterial cell that is able to take up added DNA .

Complementary DNA (cDNA): DNA synthesized using mRNA as a template, a reaction carried out by reverse transcriptase.

Complementary strand: a nucleic acid strand that will pair with a given single strand of DNA(or RNA).

Complementation: Restoration of the wild-type phenotype by the introduction of a second DNA molecule without recombination.

Composite transposon: mobile element formed by two copies of an insertion sequence flanking a region of DNA.

Conditional mutant: the effects of the mutation are only seen under certain conditions, such as elevated temperature. Under other conditions (permissive conditions), the cell/virus behaves normally.

Conjugation : transfer of genes from one prokaryotic cell to another by a mechanism involving cell-to-cell contact and a plasmid.

Conjugative transposon: mobile element that promotes conjugation and transposes from one cell to the other without having any stable independent existence.

Consensus: for some functions, such as promoters, ribosome binding sites, etc., all the known sequences with that role are related to a common sequence, the consensus sequence for that function.

Constitutive: the gene product is formed irrespective of the presence of inducers or repressors.

Contig: DNA sequence built up from a number of smaller overlapping sequences, during a sequencing project.

Copy number: the number of copies of a plasmid per cell; also used to refer to the number of copies of a specific gene (gene copy number).

cos site: the sequence of bases of bacteriophage _ which is cut asymmetrically during packaging, generating an unpaired sequence of 12 bases at each end of the phage DNA (cohesive ends).

Cosmid: a plasmid that contains the cos site of bacteriophage _. After introducing a large insert, the recombinant cosmid forms a substrate for in vitro packaging.

Co-transduction: the simultaneous transfer of two genetic markers by transduction.

Co-transformation: the simultaneous transfer of two genetic markers by transformation.

Cross-feeding: stimulation of growth of a mutant by material liberated from another cell.

Cross-talking: regulation of gene expression by signalling between cells or between compartments within a cell.

Curing: elimination of a plasmid or prophage from a bacterial culture.

Denaturation: (a) separation of the two strands of DNA by disruption of the hydrogen bonds (usually by heat or high pH); (b) disruption of the secondary and tertiary structure of proteins.

Direct repeat: two identical or very similar DNA sequences reading in the same direction .

Directed mutation: mutation that occurs in response to external conditions that favour the mutant; contrast random mutation. Also referred to as adaptive mutation.

Distal: sequences beyond a given point .

Domain: a region of a protein with a (partly) independent structure, connected to other domains by flexible loops.

Dyad symmetry: a short DNA sequence immediately repeated in the opposite direction; often involved in binding regulatory proteins.

Early genes: bacteriophage genes that are expressed soon after infection .
Electrophoresis: separation of DNA molecules (or other molecules, including RNA or protein) by application of an electric field, usually across an agarose or acrylamide gel.

Electroporation: inducing cells to take up DNA by subjecting them to brief electric pulses.

Enhancers: DNA regions that increase the level of transcription; often remote from the transcriptional start, upstream or downstream.

Epigenetic: a change to the DNA that affects expression of a gene but is not inherited.

Episome: originally defined as a genetic element (plasmid or bacteriophage) that can exist either autonomously or integrated into the chromosome. Now often used to refer only to the extrachromosomal state.

Error-prone repair: an inducible repair mechanism, part of the SOS response, which functions by reducing the specificity of replication to allow a damaged region to be copied. The cause of mutations following ultraviolet irradiation.

Excision repair: repair of damaged DNA by removal of that part of the DNA strand containing the affected region, followed by filling in the gap by repair synthesis.

Exon: coding sequence, part of a single gene but separated by an intron.
Expression vector: a cloning vector designed for expression of the cloned insert using regulatory sequences present on the vector.

Exonuclease: An enzyme which digests nucleic acids starting at one end. An example is Exonuclease III, which digests only double-stranded DNA starting from the 3' end.

F^- (F-prime) plasmid: an F plasmid carrying some chromosomal genes. Formed by inaccurate excision of F from the chromosome.

Fimbriae: filamentous appendages on the surface of a bacterial cell. Similar to pili and the use of the two terms overlaps.

Fluctuation test: a procedure for distinguishing between random and directed mutations.

Frameshift: insertion or deletion of bases, other than in multiples of three, changes the reading frame of protein synthesis beyond that point.

Fusion protein: a protein made from all or part of two (or more) different proteins, e.g. by gene cloning or naturally by ribosomal frameshifting.

Gene disruption: use of genetic techniques to inactivate a gene by inserting within it a DNA fragment containing a selectable marker. The inserted fragment is called a cassette, and the process of insertion, cassette IIlllltagenesis.Also called gene knockout.

Gene library: a collection of recombinant clones which together represent the entire genetic material of an organism. More specifically, a genomic library represents the entire genome, while a cDNA library represents the mRNA present in the cells at the time of extraction.

Gene replacement: inactivation of a chromosomal gene by recombination with an homologous sequence, inactivated *in vitro*.

Genetic map : the arrangement of genes on a chromosome.

Genetic code :The 'code' in which information for the synthesis of proteins is contained in the nucleotide sequence of a DNA molecule (or, in certain viruses, of an RNA molecule). During PROTEIN SYNTHESIS, the coded information in DNA is initially transmitted tomRNA (q.v.), i.e., the deoxyribonucleotide sequence is transcribed into a complementary sequence of ribonucleotides.

Genetic engineering : The in vitro manipulation of nucleic acid molecules e.g. to generate new combinations of genes or sequences to place a given gene or genes under the control of a different regulatory system to introduce specific mutations into a molecule.

Genetic footprinting: A procedure which may be used to determine the need for expression of each of a range of specific genes during growth under particular, chosen conditions.

Genome sequencing: determination of the complete sequence of the DNA of an organism.

Genome: the entire genetic material of an organism; in bacteria, usually confined to the chromosome; in eukaryotes, to the nuclear DNA.

Genotype: the precise genetic makeup of an organism; the complete sequence of a cell's chromosome(s) and plasmids if present.

Gel electrophoresis: A method to analyze the size of DNA (or RNA) fragments. In the presence of an electric field, larger fragments of DNA move through a gel slower than smaller ones. If a sample contains fragments at four different discrete sizes, those four size classes will, when subjected to electrophoresis, all migrate in groups, producing four migrating "bands". Usually, these are visualized by soaking the gel in a dye (ethidium bromide) which makes the DNA fluoresce under UV light.

Global regulation: a change in external conditions may affect the regulation of a large number of otherwise unrelated genes.

Hairpin: a region of DNA or RNA that contains a short inverted repeat can form a basepaired structure resembling a hairpin (similar to a stem–loop structure).

Hemi-methylated: DNA in which only one strand is methylated.
Heteroduplex: a double-stranded DNA molecule formed by base pairing between two similar but not identical strands. Since the two strands are not identical, some regions will remain single stranded.

Hfr: high frequency of recombination: conjugation donor in which F plasmid is integrated into chromosome.

Holliday junction: an intermediate in homologous recombination.

Homologous recombination: recombination between two DNA molecules which share an extensive region of homology; requires RecA .

Homologous: DNA (or RNA) molecules with the same sequence or sufficiently similar for complementary strands to hybridize.

Homologues: genes (or enzymes) with similar sequence and function, descended from a common ancestor .

Horizontal gene transfer: transfer of DNA from one bacterium to another, to contrast with vertical transfer by normal inheritance.

Hybridization: The formation of double-stranded DNA, RNA or DNA/RNA molecules by the production of hydrogen bonds between wholly or partially complementary sequences.

Hot-spot : A region of DNA which is particularly prone to e.g. transposition or mutation

Hot-start PCR : A form of PCR in which an essential component of the reaction mixture is withheld, or blocked, until the temperature of the mixture has, for the first time, risen above the primer-binding temperature; the object of this procedure is to avoid mis-priming (i.e. binding of primers to inappropriate sequences) – which, in the standard form of PCR, tends to occur primarily in the initial pre-cycling stage, i.e. when all components are present but the mixture is still at room temperature. By avoidance (or

minimization) of mis-priming, the hot-start technique promotes the specificity of a PCR assay; it can also promote the sensitivity of an assay by concentrating the full potential of the system on amplification of the required target sequence. Avoidance or minimization of non-specific products also serves to reduce the background against which the legitimate target is to be detected.

Hydrophilic: ‘water-loving’. Substances, or parts of a structure which interact with water and therefore tend to be exposed to water.

Hydrophobic: ‘water-hating’. Substances, or parts of a structure, that do not interact with water and hence tend to remove themselves from an aqueous environment.

Insertion sequence: a DNA sequence that is able to insert itself or a copy of itself into another DNA molecule; carries no information other than that required for transposition .

Insertion vector: a cloning vector into which DNA can be inserted at a single site .

Insertional inactivation: destruction of the function of a gene by insertion of a foreign DNA fragment, either by transposition or by gene cloning.

Integrity: maintenance of the structure of a plasmid without deletions or rearrangements.

Integron: DNA integration element; a region that is able to acquire additional genes by sites pecific integration.

Intercalation: the action of certain mutagens such as the acridines that can stack between the bases in the centre of a double helix. Commonly results in frameshift mutations.

Intermolecular: interactions between two different molecules.

Intramolecular: interactions between different parts of the same molecule.

Intron: Interruption (intervening sequence) in the coding sequence of a gene; removed by splicing .

Inversion: change in the orientation of a DNA fragment with respect to the sequences either side.

Inverted repeat: two identical or very similar DNA sequences reading in opposite directions.

Isoelectric focusing: separation of proteins by electrophoresis in a stable pH gradient so that each protein will move to its isoelectric point.

Isoelectric point: the pH at which a specific protein has no net overall charge.

Isogenic: strains that are identical in their genetic composition; normally used to mean identical in all genes except the one being studied.

Iteron: DNA region containing a number of short repeated sequences; involved in the replication and copy number control of some types of plasmid.

Kilobase: a nucleic acid region that is 1000 bases long.

Knock-out experiment: A technique for deleting, mutating or otherwise inactivating a gene in a mouse. This laborious method involves transfecting a crippled gene into cultured embryonic stem cells, searching through the thousands of resulting clones for one in which the crippled gene exactly replaced the normal one (by homologous recombination), and inserting that cell back into a mouse blastocyst. The resulting mouse will be chimaeric but, if you are lucky , its germ cells will carry the deleted gene. A few rounds of careful breeding can then produce progeny in which both copies of the gene are inactivated.

Leader: (a) nucleotide sequence at the 50 end of mRNA before the start point for translation of the first structural gene. Often involved in regulation of gene expression; (b) also used to refer to the signal peptide but preferable to limit it to the first meaning.

Leaky: a mutation where the loss of the relevant characteristic is not complete.

Ligation: joining two or more DNA molecules using DNA ligase.

Linking number: a measure of the overall interlinking of two DNA strands; composed of twist and writhe.

Lysogeny: a (more or less) stable relationship between a bacteriophage (prophage) and a host bacterium (lysogen).

Lytic cycle: multiplication of a bacteriophage within a host cell leading to lysis of the cell and re-infection of other sensitive bacteria.

Mapping: determination of the position of genes (genetic map) or of physical features such as restriction endonuclease sites (restriction map, physical map).

Melting: separation of double-stranded DNA into single strands (see denaturation).

Messenger RNA (mRNA): an RNA molecule that is used for translation into a protein.

Microarray: a set of oligonucleotides representing part of each gene of the bacterium, spotted onto a glass slide. Used for genome-wide comparison of gene expression or for identifying genomic deletions.

Microsatellite: A microsatellite is a simple sequence repeat (SSR). It might be a homopolymer ('...TTTTTTT...'), a dinucleotide repeat ('....CACACACACACACA.....'), trinucleotide repeat ('....AGTAGTAGTAGTAGT...') etc. Due to polymerase slip (a.k.a. polymerase chatter), during DNA replication there is a slight chance these repeat sequences may become altered; copies of the repeat unit can be created or removed. Consequently, the exact number of repeat units may differ between unrelated individuals. Considering all the known microsatellite markers, no two individuals are identical. This is the basis for forensic DNA identification and for testing of familial relationships (e.g. paternity testing).

M13: A bacteriophage which infects certain strains of *E. coli* . The salient feature of this phage is that it packages only a single strand of DNA into its capsid. If the investigator has inserted some heterologous DNA into the M13

genome, copious quantities of single-stranded DNA can subsequently be isolated from the phage capsids. M13 is often used to generate templates for DNA sequencing.

Minicells: small, non-replicating, cells that lack DNA; produced at cell division in mutants defective in the min genes.

Mismatch repair: a method for removing and replacing incorrect bases from a new DNA strand.

Mobilization: conjugative transfer of a non-conjugative plasmid in the presence of a conjugative plasmid.

Modification: alteration of the structure of DNA (usually by methylation of specific residues) so that it is no longer a substrate for the corresponding restriction endonuclease .

Molecular cloning : isolation and incorporation of a fragment of DNA into a vector where it can be replicated.

Mosaic genes: genes composed of regions from different sources.

Motif: a short sequence of amino acids or bases which is conserved in proteins or nucleic acid sequences with similar functions.

Multiple cloning site: a short region of a vector containing a number of unique restriction sites into which DNA can be introduced.

Mutagenesis: treatment of an organism with chemical or physical agents so as to induce alterations in the genetic material.

Mutant: a cell (or virus) with altered properties due to a change in its genetic material .

Mutation: an alteration in the genetic material giving rise to a cell (or virus) with altered properties .

Nick translation: A method for incorporating radioactive isotopes (typically 32P) into a piece of DNA. The DNA is randomly nicked by DNase I, and then starting from those nicks DNA polymerase I digests and then replaces a stretch of DNA. Radiolabeled precursor nucleotide triphosphates can thus be incorporated.

Nonsense mutation: base substitution creating a stop codon within the coding sequence causing premature termination of translation.

Northern blot: transfer of RNA molecules from an agarose or acrylamide gel to a filter for hybridization.

Nuclease: An enzyme which degrades nucleic acids. A nuclease can be DNA-specific (a DNase), RNA-specific (RNase) or non-specific. It may act only on single stranded nucleic acids, or only on double-stranded nucleic acids, or it may be non-specific with respect to strandedness. A nuclease may degrade only from an end (an exonuclease), or may be able to start in the middle of a strand (an endonuclease). To further complicate matters, many enzymes have multiple functions; for example, Bal31 has a 3'-exonuclease activity on double-stranded DNA, and an endonuclease activity specific for single-stranded DNA or RNA

Okazaki fragment: short fragment of DNA produced during replication of the lagging strand.

Oligonucleotide: a short nucleic acid sequence (usually synthetic).

Open reading frame (ORF): a nucleic acid sequence with a reading frame that contains no stop codons; it can therefore potentially be translated into a polypeptide.

Operator: a region of DNA to which a repressor protein binds to switch off expression of the associated gene. Usually found adjacent to or overlapping with the promoter.

Operon: a group of contiguous genes that are transcribed into a single mRNA from a common promoter and hence are subject to coordinated induction/repression.

Orthologues: equivalent genes (or enzymes) with similar sequence and function which have evolved from a common ancestor .

Origin of replication: Nucleotide sequences present in a plasmid which are necessary for that plasmid to replicate in the bacterial host

Packaging: the process of incorporating DNA into a bacteriophage particle.

Pathogenicity island: a DNA region carrying virulence determinants; often with a different base composition from the remainder of the chromosome.

Penicillin enrichment: technique for increasing the proportion of auxotrophic mutants.

Phenotype: the observable characteristics of an organism .

Pheromone: secreted chemical used for signalling between two or more individuals.

Phagemid: A type of plasmid which carries within its sequence a bacteriophage replication origin. When the host bacterium is infected with "helper" phage, the phagemid is replicated along with the phage DNA and packaged into phage capsids.

Plasmid: A circular piece of DNA present in bacteria or isolated from bacteria. *Escherichia coli*, the usual bacteria in molecular genetics experiments, has a large circular genome, but it will also replicate smaller circular DNAs as long as they have an "origin of replication". Plasmids may also have other DNA inserted by the investigator. A bacterium carrying a plasmid and replicating a million-fold will produce a million identical copies of that plasmid. Common plasmids are pBR322, pGEM, pUC18.

Pilus : filamentous appendage on the surface of a bacterial cell. Some types of pili are specified by plasmids and are involved in mating-pair formation in conjugation .

Plaque: a region of clearing or reduced growth in a bacterial lawn as a result of phage infection.

Polar mutation: a mutation in one gene may affect the expression of others (e.g. genes downstream in an operon). The phenotypic effect may not be directly caused by the original mutation.
Polarity: differences in the level of translation of genes within an operon, especially the effect of a polar mutation in one gene in reducing or abolishing the expression of subsequent genes.

Polycistronic mRNA: messenger RNA coding for several proteins .

Polymerase chain reaction: enzymatic amplification of a specific DNA fragment using repeated cycles of denaturation, primer annealing and chain extension.

Polymorphism: the existence of different forms of a characteristic within a population.

Post-replication repair: DNA repair process involving the exchange of DNA between damaged and undamaged strands.

Post-transcriptional regulation: Any process occurring after transcription which affects the amount of protein a gene produces. Includes RNA processing efficiency, RNA stability, translation efficiency, protein stability. For example, the rapid degradation of an mRNA will reduce the amount of protein arising from it. Increasing the rate at which an mRNA is translated will increase the amount of protein product.

Post-translational processing: The reactions which alter a protein's covalent structure, such as phosphorylation, glycosylation or proteolytic cleavage.

Post-translational regulation: Any process which affects the amount of protein produced from a gene, and which occurs AFTER translation in the grand scheme of genetic expression. Actually, this is often just a buzz-word for regulation of the stability of the protein. The more stable a protein is, the more it will accumulate.

PRE: Progesterone Response Element: A binding site in a promoter to which the activated progesterone receptor can bind. The progesterone receptor is essentially a transcription factor which is activated only in the presence of progesterone . The activated receptor will bind to a PRE, and transcription of the adjacent gene will be altered. See also "Response element".

Pribnow box: consensus sequence within a promoter, centred at the -10 position with respect to the start of transcription.

Primary structure: the sequence of a nucleic acid or a protein.

Primer: synthesis of a new DNA (but not RNA) strand can only occur by extension of a preexisting partial DNA strand. If a specific oligonucleotide (a primer) is provided, complementary to a defined region of the template strand, all the new DNA strands made will start from that point.

Probe: a DNA or RNA molecule that will hybridize to a specific target sequence. Labelling the probe (using radioactive isotopes or non-radioactive markers) enables it to be used to detect the specific target DNA/RNA.

Promoter: region of DNA to which RNA polymerase binds in order to initiate transcription.

Proof-reading: the ability of DNA polymerase to check the accuracy of the newly made sequence.

Prophage: the repressed form of bacteriophage DNA in a lysogen; it may be integrated into the chromosome or exist as a plasmid.

Protein engineering: altering a gene so as to produce defined changes in the properties of the encoded protein, e.g. thermal stability, substrate profile.

Proteome: the complete content of different proteins in a cell .

Protoplast: formed by complete removal of the cell wall using osmotically-stabilized conditions.

Prototroph: a nutritionally wild-type organism that does not need any additional growth supplement .

Proximal: sequence before a given point, usually referring to the direction of transcription or translation .

Pulsed-field gel electrophoresis: separation of large DNA molecules by application of an intermittently varying electric field; generic term for a number of ways of achieving this.

Purine: one of the two types of bases in nucleic acids (adenine, guanine; see pyrimidine).

Pyrimidine dimer: covalent linkage between adjacent pyrimidines on a DNA strand caused by UV irradiation. Commonly referred to as thymine dimers but not restricted to thymine.

Pyrimidine: one of the two types of bases in nucleic acids (cytosine and thymine in DNA; cytosine and uracil in RNA.

Quorum sensing: mechanism whereby bacteria respond to cell density.

Random mutation: mutation occurring irrespective of its benefit to the cell (contrast directed mutation).

Reading frame: a nucleic acid sequence is translated in groups of three bases (codons); there are three possible ways of reading the sequence (in one direction) depending on where it is started. These are the three reading frames.

Real-time PCR: a PCR technique which allows monitoring of the amplification of the product as it happens. Especially useful for quantitative applications of PCR.

Recombinant: product of recombination (q.v.) using either definition, which leads to recombinant bacteria resulting from some form of gene transfer or recombinant plasmids arising from in vitro manipulation.

Recombination: (a) the production of new strains by mating two genetically distinct parents; (b) the generation of new DNA molecules by breaking and re-joining the original molecules; this may occur naturally within the cell (in vivo) or artificially in vitro. There is considerable overlap between these definitions but they are not always synonymous.

Relaxation: conversion of supercoiled circular DNA to an open circular form.

Relaxed plasmid: (a) open circular structure after nicking one strand of a plasmid; (b) plasmid which replicates to high copy number without being tied to chromosomal replication.

Replica plating: using a velvet pad or some equivalent apparatus to transfer a number of colonies to several different media in order to compare their growth requirements or other characteristics.

Replication: synthesis of a copy of a DNA molecule using the original as a template.

Replicon: (a) a DNA molecule (such as a plasmid) that contains an origin of replication and is capable of autonomous replication within a suitable host cell; (b) the replication control region of a plasmid.

Reporter gene: a gene which codes for a readily detected product (such as *b*-galactosidase); study of a regulatory region of DNA is facilitated by fusion with the reporter gene.

Repression: (a) reduction in transcription of a gene usually due to the action of a repressor protein; (b) also applied to the natural repression of conjugal transfer that occurs with many plasmids and to the establishment of lysogeny with temperate bacteriophages.

Resolution: production of two smaller plasmids from a cointegrate.

Restriction fragment: The piece of DNA released after restriction digestion of plasmids or genomic DNA. See "Restriction enzyme". One can digest a plasmid and isolate one particular restriction fragment (actually a set of identical fragments). The term also describes the fragments detected on a genomic blot which carry the gene of interest.

Reverse genetics: starting with specific alterations to the DNA in vitro and then examining the phenotype; contrasts with classical genetics which relies on selecting mutants on the basis of their phenotype and then studying the nature of the mutation.

Reverse transcriptase: An enzyme which will make a DNA copy of an RNA template - a DNA-dependant RNA polymerase. RT is used to make cDNA; one begins by isolating polyadenylated mRNA, providing oligo-dT as a primer, and adding nucleotide triphosphates and RT to copy the RNA into cDNA.

RFLP: Restriction fragment length polymorphism; the acronym is pronounced "riflip". Although two individuals of the same species have almost identical genomes, they will always differ at a few nucleotides. Some of these differences will produce new restriction sites (or remove them), and thus the banding pattern seen on a genomic Southern will thus be affected. For any given probe (or gene), it is often possible to test different restriction enzymes until you find one which gives a pattern difference between two individuals - a RFLP. The less related the individuals, the more divergent their DNA sequences are and the more likely you are to find a RFLP.

Reversion: a mutation that reverses the effect of the original mutation.

Ribosomal frameshifting: a change in reading frame used by the ribosome resulting in a fusion protein or re-initiation from an adjacent start codon.

Ribosome binding site: the region on an mRNA molecule to which ribosomes initially attach.

RNAi: 'RNA interference' is the mechanism by which small double-stranded RNAs can interfere with expression of any mRNA having a similar sequence. Those small RNAs are known as 'siRNA', for short interfering RNAs. The mode of action for siRNA appears to be via dissociation of its strands, hybridization to the target RNA, extension of those fragments by an RNA-dependent RNA polymerase, then fragmentation of the target. Importantly, the remnants of the target molecule appears to then act as an siRNA itself; thus the effect of a small amount of starting siRNA is effectively amplified and can have long-lasting effects on the recipient cell.

The RNAi effect has been exploited in numerous research programs to deplete the call of specific messages, thus examining the role of those messages by their absence.

RNase: Ribonuclease; an enzyme which degrades RNA. It is ubiquitous in living organisms and is exceptionally stable. The prevention of RNase activity is the primary problem in handling RNA.

rRNA: "ribosomal RNA"; any of several RNAs which become part of the ribosome, and thus are involved in translating mRNA and synthesizing proteins. They are the most abundant RNA in the cell (on a mass basis).

S1 nuclease: An enzyme which digests only single-stranded nucleic acids.

Sequence: As a noun, the sequence of a DNA is a buzz word for the structure of a DNA molecule, in terms of the sequence of bases it contains. As a verb, "to sequence" is to determine the structure of a piece of DNA; i.e. the sequence of nucleotides it contains.

Shotgun cloning: The practice of randomly clipping a larger DNA fragment into various smaller pieces, cloning everything, and then studying the resulting individual clones to figure out what happened. For example, if one was studying a 50 kb gene, it "may" be a bit difficult to figure out the restriction map. By randomly breaking it into smaller fragments and mapping those, a master restriction map could be deduced. See also Shotgun sequencing.

Shotgun sequencing: A way of determining the sequence of a large DNA fragment which requires little brainpower but lots of late nights. The large fragment is shotgun cloned (see above), and then each of the resulting smaller clones ("subclones") is sequenced. By finding out where the subclones overlap, the sequence of the larger piece becomes apparent. Note that some of the regions will get sequenced several times just by chance.

snRNA: Small nuclear RNA; forms complexes with proteins to form snRNPs; involved in RNA splicing, polyadenylation reactions, other unknown functions (probably).

SNP: Single Nucleotide Polymorphism (SNP) - a position in a genomic DNA sequence that varies from one individual to another. It is thought that the primary source of genetic difference between any two humans is due to the presence of single nucleotide polymorphisms in their DNA. Furthermore, these SNPs can be extremely useful in genetic mapping to follow inheritance of specific segments of DNA in a lineage. SNP-typing is the process of determining the exact nucleotide at positions known to be polymorphic.

Solution hybridization: A method closely related to RNase protection. Solution hybridization is designed to measure the levels of a specific mRNA species in a complex population of RNA. An excess of radioactive probe is

allowed to hybridize to the RNA, then single-strand specific nuclease is used to destroy the remaining unhybridized probe and RNA. The "protected" probe is separated from the degraded fragments, and the amount of radioactivity in it is proportional to the amount of mRNA in the sample which was capable of hybridization. This can be a very sensitive detection method.

Southern blot: A technique for analyzing mixtures of DNA, whereby the presence and rough size of one particular fragment of DNA can be ascertained. See "Blotting". Named for its inventor, Dr E. M. Southern.

Taq polymerase: A DNA polymerase isolated from the bacterium *Thermophilis aquaticus* and which is very stable to high temperatures. It is used in PCR procedures and high temperature sequencing.

TATA box: A sequence found in the promoter (part of the 5' flanking region) of many genes. Deletion of this site (the binding site of transcription factor TFIID) causes a marked reduction in transcription, and gives rise to heterogeneous transcription initiation sites.

Tm: The melting point for a double-stranded nucleic acid. Technically, this is defined as the temperature at which 50% of the strands are in double-stranded form and 50% are single-stranded, i.e. midway in the melting curve. A primer has a specific Tm because it is assumed that it will find an opposite strand of appropriate character.

Transcription factor: A protein which is involved in the transcription of genes. These usually bind to DNA as part of their function (but not necessarily). A transcription factor may be general (i.e. acting on many or all genes in all tissues), or tissue-specific (i.e. present only in a particular cell type, and activating the genes restricted to that cell type). Its activity may be constitutive, or may depend on the presence of some stimulus; for example, the glucocorticoid receptor is a transcription factor which is active only when glucocorticoids are present.

Transcription: The process of copying DNA to produce an RNA transcript. This is the first step in the expression of any gene. The resulting RNA, if it codes for a protein, will be spliced, polyadenylated, transported to the cytoplasm, and by the process of translation will produce the desired protein molecule.

Transfection: A method by which experimental DNA may be put into a cultured mammalian cell. Such experiments are usually performed using cloned DNA containing coding sequences and control regions (promoters, etc) in order to test whether the DNA will be expressed. Since the cloned DNA may have been extensively modified (for example, protein binding sites on the promoter may have been altered or removed), this procedure is often used to test whether a particular modification affects the function of a gene.

Transformation (with respect to bacteria): The process by which a bacteria acquires a plasmid and becomes antibiotic resistant. This term most commonly refers to a bench procedure performed by the investigator which introduces experimental plasmids into bacteria.

Transformation (with respect to cultured cells): A change in cell morphology and behavior which is generally related to carcinogenesis. Transformed cells tend to exhibit characteristics known collectively as the "transformed phenotype" (rounded cell bodies, reduced attachment dependence, increased growth rate, loss of contact inhibition, etc). There are different "degrees" of transformation, and cells may exhibit only a subset of these characteristics. Not well understood, the process of transformation is the subject of intense research.

Transgenic mouse: A mouse which carries experimentally introduced DNA. The procedure by which one makes a transgenic mouse involves the injection of DNA into a fertilized embryo at the pro-nuclear stage. The DNA is generally cloned, and may be experimentally altered. It will become incorporated into the genome of the embryo. That embryo is implanted into a foster mother, who gives birth to an animal carrying the new gene. Various experiments are then carried out to test the functionality of the inserted DNA.

Transient transfection: When DNA is transfected into cultured cells, it is able to stay in those cells for about 2-3 days, but then will be lost. During those 2-3 days, the DNA is functional, and any functional genes it contains will be expressed. Investigators take advantage of this transient expression period to test gene function.

Translation: The process of decoding a strand of mRNA, thereby producing a protein based on the code. This process requires ribosomes (which are

composed of rRNA along with various proteins) to perform the synthesis, and tRNA to bring in the amino acids. Sometimes, however, people speak of "translating" the DNA or RNA when they are merely reading the nucleotide sequence and predicting from it the sequence of the encoded protein. This might be more accurately termed "conceptual translation".

tRNA: "transfer RNA"; one of a class of rather small RNAs used by the cell to carry amino acids to the enzyme complex (the ribosome) which builds proteins, using an mRNA as a guide. Fairly abundant.

Upstream activator sequence: A binding site for transcription factors, generally part of a promoter region. A UAS may be found upstream of the TATA sequence (if there is one), and its function is (like an enhancer) to increase transcription. Unlike an enhancer, it can not be positioned just anywhere or in any orientation.

Vector: The DNA "vehicle" used to carry experimental DNA and to clone it. The vector provides all sequences essential for replicating the test DNA. Typical vectors include plasmids, cosmids, phages and YACs.

Western blot: A technique for analyzing mixtures of proteins to show the presence, size and abundance of one particular type of protein. Similar to Southern or Northern blotting , except that (1) a protein mixture is electrophoresed in an acrylamide gel, and (2) the "probe" is an antibody which recognizes the protein of interest, followed by a radioactive secondary probe .

References

Agarwal,P.K.(2003). Genome Sequencing. CPS260/BGT204.1 Algorithms in Computational Biology.

Akhtar A and Gasser SM (2007) .The nuclear envelope and transcriptional control. Nat. Rev. Genet.8, 507–517.

Ambur OH, Frye SA, Nilsen M, Hovland E, Tønjum T (2012). "Restriction and sequence alterations affect DNA uptake sequence-dependent transformation in Neisseria meningitidis". PLoS ONE **7** (7): e39742. doi:10.1371/journal.pone.0039742. PMC 3388099. PMID 22768309.

Batzoglou, D. B. Jaffe, K. Stanley, J. Butler, S. Gnerre, E. Mauceli, B. Berger, J. P. Mesirov, and E. S.Lander. Arachne: a whole-genome shotgun assembler. Genome Research, 12(1):177–189, 2002.

Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P, MuTHER Consortium: Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012, **8:**e1002629.

Bennett ,P.M.(2015) Integrons and gene cassettes: a genetic construction kit for bacteria. Journal of Antimicrobial Chemotherapy ,43(1):1-4.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA: The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 2010, **28:**1045-1048.

Birge, E.A. (2000) Bacterial and Bacteriophage Genetics.An introduction .
Birikh, K. R., Heaton, P. A., and Eckstein, F. (1997) The structure, function and application of the hammerhead ribozyme. *Eur. J. Biochem.* **245,** 1–16.

Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E: Epigenetic predictor of age**.** *PLoS One* 2011, **6:**e14821.

Boundless. “Bacterial Genomes.” Boundless Microbiology. Boundless, 21 Jul. 2015. Retrieved 31 Oct. 2015 from https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/microbial-genetics-7/genes-71/bacterial-genomes-412-10774/

Boundless. “Bacterial Genomes.” Boundless Microbiology. Boundless, 21 Jul. 2015. Retrieved 31 Oct. 2015 from https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/microbial-genetics-7/genes-71/bacterial-genomes-412-10774/

Boundless. “DNA Replication in Prokaryotes.” Boundless Microbiology. Boundless, 21 Jul. 2015. Retrieved 31 Oct. 2015 from https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/microbial-genetics-7/genes-71/dna-replication-in-prokaryotes-413-12940/.

Boundless. “Gene Inversion.” Boundless Microbiology. Boundless, 21 Jul. 2015. Retrieved 31 Oct. 2015 from https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/microbial-genetics-7/genes-71/gene-inversion-414-4570/-6639/

Brock, T.D. (1990) The emergence of bacterial genetics, Cold Spring Harbor Laboratory Press, New York, 346 pp. [A most accurate historical account of bacterial genetics with a complete bibliography before the establishment of recombinant DNA technology.

Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D, Nofroni C, Comanducci M, Riley DR, Daugherty SC, Angiuoli SV, Covacci A, Pizza M, Rappuoli R, Moxon ER, Tettelin H, Medini D (2011). "Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination". Proc. Natl. Acad. Sci. U.S.A. 108 (11): 4494–9. doi:10.1073/pnas.1019751108. PMC 3060241. PMID 21368196.

Bukhari, A. J., Shapiro, J. A. and Adhya, S. L. (1977) DNA Insertion elements, Plasmids, and Episomes. Cold Spring Harbor Laboratory/University of Tokyo Press. 782 pp.

Calladine, C.R. and H. R. Drew. Understanding DNA: The Molecule and How It Works. 2nd edn (1997). Academic Press. (3rd edn due in 2004). An unusual book.

Caugant DA, Maiden MC (2009). "Meningococcal carriage and disease--population biology and evolution". Vaccine. 27 Suppl 2: B64–70. doi:10.1016/j.vaccine.2009.04.061. PMC 2719693. PMID 19464092.

Claus H, Friedrich A, Frosch M, Vogel U (2000). "Differential distribution of novel restriction-modification systems in clonal lineages of Neisseria meningitidis". J. Bacteriol. 182 (5): 1296–303. PMC 94415. PMID 10671450.

Dale ,J.W. and von Schantz)2002.from gene to genome ,concepts and applications of DNA Technology. Wiley and Sons Ltd.

Dieci G, Fiorino G, Castelnuovo M et al. (2007) The expanding RNA polymerase III transcriptome. Trends Genet. 23, 614–622. Haag JR & Pikaard CS (2007) RNA polymerase I: a multifunctional molecular machine. Cell 131, 1224–1225.

Hahn S (2004) Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol.11, 394–403.Jones KA(2007) Transcription strategies in terminally differentiated cells: shaken to the core.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K: Genome-wide methylation profiles reveal quantitative views of human aging rates**.** *Mol Cell* 2013, 49:359-367.

Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ, Singleton AB: Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011, **20:**1164-1172.

Horvath S, Zhang Y, Langfelder P, Kahn R, Boks M, van Eijk K, van den Berg L, Ophoff RA: Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* 2012, **13:**R97.

Illingworth R, Kerr A, Desousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A: A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* 2008, **6:**e22.

Jacob, F. and Wollman, E.L. (1961) "The sexuality and the genetics of bacteria, Academic Press, New York, 374 pp.
Judson, H. F. 1996. The Eighth Day of Creation: The Makers of the Revolution in Biology. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Karam, J. D. (editor-in-chief), Drake, J. W., Kreuger,K. N., Mosig, G, Hall, D. H., Eiserling, F. A., Black, L W., Spicer, E. K., Kutter, E., Carls on, K., and Miller, E. S. (edts) (1994) " Molecular Biology of bacteriophage T4 American Society for Microbiology, Washington D. C., 615 pp.
Kilpatrick, M. W., Phylactou, L. A., Godfrey, M., Wu, C. H., Wu, G. Y., and

Koch C, Wagner W: Epigenetic-aging-signature to determine age in different tissues. *Aging* 2011, **3:**1018-1027.

Kornberg RD (2007) The molecular basis of eukaryotic transcription. Proc. Natl Acad. Sci. USA 104, 12955–12961.

Kornberg, A. and Baker, T. (1992) " DNA Replication W. H. Freeman and Co., New York, 931 pp.

Kumaran RI, Thakar R & Spector DL (2008) Chromatin dynamics and gene positioning. Cell 132, 929–934.Reina JH & Hernandez N (2007) On a roll for new TRF targets. Genes Dev. , 2855–2860.

Lewis, R. 2001. Human Genetics: Concepts and Applications.

Lewontin, R. C. 2000. It Ain't Necessarily So: The Dream of the Human Genome and Other Illusions. New York: New York Review of Books.

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Sun J,

Huang Y, Zheng H, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, *et al*.: The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* 2010, 8**:**e1000533.

Lodge, J.; Lund,P. and Minchin ,S. (2007). Gene cloning : principles and applications . Taylor & Francis GroupI6R. K. Saiki, S. J. Scharf, F. Faloona, K. B. Mullis, G. T Horn, H. A. Erlich and N. Arnheim,
Lodish,H.; A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. Molecular Cell Biology.Scientific American Books, Inc., 4 edition, 2000.
Luria, S.E., Darnell, Jr. J.E., Baltimore, D., and Campbell, A. (1987) General Virology (3rded). John Wiley and Sons, New York, Santa Barbara, Chichester, Brisbane, Tokyo, 578 pp. [A comprehensive text book of bacteriophage, plant, and animal viruses by pioneering scientists of these viruses.
Maroni, G. 2000. Molecular and Genetic Analysis of Human Traits. Malden, MA: Blackwell.

Mugatroyd C, Wu Y, Bockmühl Y, Spengler D: The Janus face of DNA methylation in aging. *Aging* 2010, **2:**107-110.

Myers,E.W. Whole-genome DNA sequencing. IEEE Computational Engineering and Science, 3(1):33–43,1999.

Nathans D, Smith HO (1975). "Restriction endonucleases in the analysis and restructuring of dna molecules". Annu. Rev. Biochem. **44**: 273–93. doi:10.1146/annurev.bi.44.070175.001421.

Neidhardt, F. C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M., and Umbarger,H.E. (edits. 1996) "*Escherichia coli* and *Salmonella typhimurium* – cellular and molecular biology, 2ndedition. American Society of Microbiology Press, Washington DC, Vol. 1, 1539 pp., Vol. 2, 2822 pp.

Numata S, Ye T, Hyde Thomas M, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger Daniel R, Kleinman Joel E, Lipska Barbara K: DNA methylation signatures in development and aging of the human prefrontal cortex**.** *Am J Hum Genet* 2012, **90:**260-272.

Olby, R. C. 1966. Origins of Mendelism. London: Constable.

Orel, V. 1996. Gregor Mendel: The First Geneticist. Oxford, England: Oxford University Press.
Orel, V., and D. L. Hartl. 1994. Controversies in the interpretation of Mendel's discovery. History and Philosophy of the Life Sciences 16: 423.
Phylactou, L. A., Tsipouras, P., and Kilpatrick, M. W. (1998) Hammerhead

Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD: Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010, **20:**434-439.

Rodríguez-Rodero S, Fernández-Morera J, Fernandez A, Menéndez-Torre E,

Fraga M: Epigenetic regulation of aging. *Discov Med* 2010, **10:**225-233.

Sandler, I. 2000. Development: Mendel's legacy to genetics. Genetics 154: 7.
Schneider R & Grosschedl R (2007) Dynamics and interplay of nuclear architecture, genome organization, and gene expression. Genes Dev.21, 3027–3043.

Shimayama, T., Nishikawa, S., and Taira, K. (1995) Generality of the NUX rule: kinetic analysis of the results of systematic mutations in the trinucleotide at the cleavage site of hammerhead ribozymes. *Biochemistry* **34,** 3649–3654.

Sykes, B, ed. 1999. The Human Inheritance. New York: Oxford University Press.

Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM, Barzilai N: Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell* 2010, **9:**506-518.

Tortora, G.J. Microbiology An Introduction 8th, 9th, 10th ed. San Francisco: Pearson Benjamin Cummings, 2004, 2007, 2010.

Tsipouras, P. (1996) Delivery of a hammerhead ribozyme specifically down-regulates the production of fibrillin–1 by cultured dermal fibroblasts. *Hum. Mol. Genet.* **5,** 1939–1944.

Wayengera M (2003). "HIV and Gene Therapy: The proposed [R-M enzymatic] model for a gene therapy against HIV.". Makerere Med J. **38**: 28–30.

Wayengera M, Kajumbula H, Byarugaba W (2007). "Frequency and site mapping of HIV-1/SIVcpz, HIV-2/SIVsmm and Other SIV gene sequence cleavage by various bacteria restriction enzymes: Precursors for a novel HIV inhibitory product". Afr J Biotechnol **6** (10): 1225–1232.

Wilson, G., "Organization of Restriction-Modification Systems,"Nucleic Acids Research (1991), Vol 19, pg2539-2566.

Wilson, G., "Restriction and Modification Systems," Annual Review of Genetics (1991), 25:585-627.

Proof

Printed By Createspace

Digital Proofer