

CHAPTER 1

BIOTECHNOLOGY: AN INTRODUCTION

What is biotechnology?

Biotechnology is not by itself a science; it is a multidisciplinary approach, which involves several disciplines and sciences (biology, biochemistry, genetics, virology, agronomy, engineering, chemistry, medicine, veterinary and others). Several definitions are found to describe biotechnology. In general terms, biotechnology is the use of living organisms or of substances obtained from living organisms, to make products of value to man. As such, biotechnology has been practiced by human society since the beginning of recorded history in such activities as baking bread, brewing alcoholic beverages, or breeding food crops or domestic animals. Historically, biotechnology meant the use of organisms to perform a task or function. Biotechnology, by this definition has been around for a long time. Processes like the production of beer, wine, cheese and yogurt involve the use of bacteria or yeast in order to convert a natural product, such as milk or grape juice, into a more desirable fermentation product like yogurt or wine. Biotechnology has many traditional applications. One simple example is composting, which builds soil fertility by encouraging microorganisms to break down crop residues. Other applications include the production and use of vaccines to control human and animal diseases. In the food-processing industry cheese and wine making are among the many practical uses of biotechnology.

Modern biotechnology are a variety of tools derived from research in cellular and molecular biology that can be applied in any industry that involves microorganisms, as well as animal and plant cells. The technique allows transforming agriculture. They also have importance for other carbon-based industries such as energy, chemicals, pharmaceuticals and waste management. The potential impact is enormous, because research in the biological sciences is making very rapid advances and the results not only affect a range of sectors but also promote greater linkage between them. A research success in fermenting energy form agricultural wastes, for example, would affect both the energy sector and the economics of food production.

A narrower and more specific definition of "modern" biotechnology is " the commercial application of living organisms or their products, which involves the deliberate manipulation of their DNA molecules". This definition implies a set of laboratory technique developments within the last decades that have been responsible for the tremendous scientific and commercial interest in biotechnology, the founding of many new companies and the redirection of research efforts and financial resources among established companies and universities.

Biotechnology consists of a gradient of technologies, ranging from the long-established and widely used techniques of traditional biotechnology (e.g. food fermentation, biological control) through to modern biotechnology, which is based on the use of new techniques of recombinant-

DNA technology (often called genetic engineering), monoclonal antibodies, and new cell- and tissue-culture methods.

How has been the development and applications of biotechnological advances?

Biotechnology is certainly an important scientific topic. In the last decades it has contributed to a transformation of many parts of the chemical industry, of agriculture and of medicine –a transformation that has emerged from the laboratory into practical application with quite remarkable speed-. Biotechnology is not new: the origins of biotechnology lie deep in the mists of human history. Our early ancestors originated the practice of using living organisms and their products about 10 000 years ago, during the Stone Age when they first began to keep domestic animals and to grow crop plants for food, instead of depending solely on what they could hunt or gather in the wild. One of the oldest industries in the world, the brewing of beer, depends on a typical biotechnological process. The breeding of domesticated animals is biotechnology too, if we accept the usual definition of biotechnology as the exploitation of other living organisms for the benefit of man.

In general, man is not satisfied with the productivity of other organisms in the wild state. Thus, breeding is required to effect a permanent change in the hereditary make-up of the organism to increase the productivity of the desired product (e.g. protein, carbohydrate or alcohol). Historically, breeding has been the limiting factor in improving organisms, because primitive methods, including some described in the Old Testament, are slow and empirical and proceed by trial and error. The possibility that gives "modern biotechnology" is that it offers radically new ways of altering the genetic properties of organisms in a directed manner. This power depends on the discoveries and advances of molecular biology techniques, the knowledge about DNA as the material of heredity, of the genetic code, of methods of reading the genetic message by sequencing genes, on the use of restriction enzymes, with which is possible to cut and splice together sections of DNA in a deliberate and directed way.

The organisms used today in biotechnology can be complex, multicellular ones as dairy cattle or as simple as the yeast needed for brewing beer and baking bread. Even single-microorganisms are valuable because they provide drugs, including antibiotics such as streptomycin and penicillin, as well as other complex chemical products that are difficult or much more expensive to obtain by synthesis in the laboratory.

Therefore, biotechnology is not a new science; rather, it is a new term that has been given to the recent evolution of the science of genetics. This science originated in the late 19th century with the pioneering work of Gregor Mendel. Although most of the information that has made possible the development of recombinant DNA technology and, therefore, the advances in modern biotechnology has been garnered during the past 4-5 decades, the story really began more than 130 years ago, with the independent investigations of Charles Darwin and Gregor Mendel. The contributions of Darwin (considered by some as the father of modern biology) received the most immediate recognition, even though that recognition was not always favorable. Darwin, in his

studies concluded that species are not fixed and unchanging but are capable of evolving over time to produce new species. In addition, Darwin provided a possible explanation for how this evolution could occur. He noted that individual members of a given species show a great deal of variation and proposed that some of them would be fit for the environment in which they found themselves, than others. The more fit individuals would thus produce more offspring than those less fit. Eventually this process, called by Darwin as natural selection (published in 1859), would cause a shift in the characteristics of the population as those traits that favored survival and reproduction were maintained and propagated, while less favorable traits became less common or died out. In plant or animal breeding something similar happens, although the breeder, rather than nature, provides the selective force by choosing the traits to be maintained.

At the same time, the monk Gregor Mendel was working obscurely in the garden of his monastery in Austria. Mendel was laying the foundations of the science of modern genetics, even though it would be another 35 years before the scientific community appreciated the significance of his work. Mendel's research uncovered the basic rules governing heredity. Mendel studied and followed the inheritance of some selected characteristics in successive generations in pea plants. From his results, published in 1865, he deduced that hereditary traits are carried and transmitted to the progenies as discrete units. In essence, he originated the concept of the "gene", although this term was not used until the early 1900s. Mendel's experiments also led him to conclude that each individual carries two units for a given characteristic, but passes just one copy to each progeny. Furthermore, some variants of a particular trait are dominant over others. If they are inherited together the dominant character is expressed, while the other, recessive variant is not seen. These postulates fully explained the inheritance patterns in the pea plants. Mendel's reports went largely unnoticed; even Darwin was unaware of them although they bore directly on his own research. The hereditary units described by Mendel are the raw material for the variation that is acted by natural selection. De Vries and Correns who were doing similar breeding experiments eventually rediscovered Mendel's research around the turn of the century. It was about this time that the basic unit of heredity became known as the gene and the science of heredity was given the name of 'genetics'

Although Mendel described the essential behavior of genes, his experiments did not reveal the chemical nature of the hereditary units. That happened around the middle of the 20th century and involved many works of different scientists around the world, during several decades. The identification of the genetic material as DNA and the description and understanding of its structure and functions required a great deal of work.

During the 1970s scientists developed new methods for combining portions of DNA (deoxyribonucleic acid, the biochemical material in all living cells that conveys the instructions that govern hereditary characteristics) and for moving portions of DNA from one organism to another. This set of enabling techniques is referred to as recombinant-DNA technology or genetic engineering. Over the past two decades there has been an exponential increase in the number of significant advances in genetics. It is this increase in new techniques for understanding and modifying the genetics of living organisms that has led to the greatly increased interest and investments in biotechnology.

CHAPTER 2

MILESTONES IN DNA RESEARCH

- 1866 Austrian botanist and monk Gregor Mendel proposes basic laws of heredity based on cross-breeding experiments with pea plants. His findings, published in a local natural-history journal, are largely ignored for more than 30 years.
- 1882 While examining salamander larvae under a microscope, German embryologist Walther Flemming spots tiny threads within the cells' nuclei that appear to be dividing. The threads will later turn out to be chromosomes.
- 1883 Francis Galton, a cousin of Charles Darwin's and an advocate of improving the human race by means of selective breeding, coins the word eugenics.
- 1910 U.S. biologist Thomas Hunt Morgan's experiments with fruit flies reveal that some genetically determined traits are sex linked. His work also confirms that the genes determining these traits reside on chromosomes.
- 1926 U.S. biologist Hermann Muller discovers that X rays can cause genetic mutations in fruit flies.
- 1932 Publication of Aldous Huxley's novel *Brave New World*, which presents a dystopian view of genetic engineering.
- 1944 Working with pneumococcus bacteria, Oswald Avery, Colin MacLeod and Maclyn McCarty prove that DNA, not protein, is the hereditary material in most living organisms.
- 1950 British physician Douglas Bevis describes how amniocentesis can be used to test fetuses for Rh-factor incompatibility. The prenatal test will later be used to screen for a battery of genetic disorders.
- 1953 American biochemist James Watson and British biophysicist Francis Crick announce their discovery of the double-helix structure of DNA, the molecule that carries the genetic code.
- 1964 Stanford geneticist Charles Yanofsky and colleagues prove that the sequence of nucleotides in DNA corresponds exactly to the sequence of amino acids in proteins.
- 1969 A Harvard Medical School team isolates the first gene: a snippet of bacterial DNA that plays a role in the metabolism of sugar.
- 1970 University of Wisconsin researchers synthesize a gene from scratch.
- 1973 American biochemists Stanley Cohen and Herbert Boyer insert a gene from an African clawed toad into bacterial DNA, where it begins to work. Their experiment marks the beginning of genetic engineering.
- 1975 Scientists at an international meeting in Asilomar, Calif., call for guidelines for recombinant-DNA research.
- 1976 The first genetic-engineering company, Genentech, is founded in South San Francisco.
- 1977 Walter Gilbert and Allan Maxam devise a method for sequencing DNA.
- 1977 Scientists from Genentech and a Duarte, Calif., medical center clone the gene for human insulin.
- 1978 David Botstein discovers RFLP analysis

- 1980 Researchers successfully introduce a human gene--one that codes for the protein interferon--into a bacterium.
- 1980 U.S. Supreme Court rules that life forms can be patented. Kary Mullis develops PCR. Sells patent for \$300M in 1991
- 1980 Martin Cline and co-workers create a transgenic mouse, transferring functional genes from one animal into another.
- 1980 The U.S. Supreme Court in *Diamond v. Chakrabarty* rules that genetically altered life forms can be patented. The decision allows the Exxon Oil Company to patent an oil-eating microorganism.
- 1981 First transgenic mice produced
- 1982 The USFDA approves sale of genetically engineered human insulin
- 1982 The U.S. Food and Drug Administration approves the first genetically engineered drug, a form of human insulin produced by bacteria.
- 1983 Researchers locate a genetic marker for Huntington's disease on chromosome 4. Their achievement leads to a screening test, but the disorder remains incurable. The gene itself will be found 10 years later.
- 1983 While driving along a California highway, Kary Mullis, a biochemist at Cetus Corp., conceives of the so-called polymerase chain reaction, or PCR, a technique that will enable scientists to rapidly reproduce tiny snippets of DNA.
- 1983 An automated DNA sequencer is developed
- 1983 A screening test for Huntington's disease is developed using restriction fragment length markers.
- 1984 Alec Jeffreys, of Britain's University of Leicester, develops "genetic fingerprinting," which uses unique sequences of DNA to identify individuals.
- 1984 First use of genetic fingerprinting in a criminal investigation.
- 1985 Genetically engineered plants resistant to insects, viruses, and bacteria are field tested for the first time
- 1985 The NIH approves guidelines for performing experiments in gene therapy on humans
- 1985 The FDA approves the first genetically engineered vaccine for humans, for hepatitis B.
- 1986 Invention of YACs (yeast artificial chromosomes) as expression vectors for large proteins
- 1986 Craig Venter developed the shotgun sequencing technique for whole genome sequencing
- 1987 The first field tests of genetically engineered crops (tobacco and tomato) are conducted in the United States.
- 1988 Harvard University is awarded the first patent for a genetically altered animal, a mouse that is highly susceptible to breast cancer.
- 1989 Creation of the National Center for Human Genome Research, headed by James Watson, which will oversee the \$3 billion U.S. effort to map and sequence all human DNA by 2005.
- 1990 Formal launch of the international Human Genome Project.
- 1990 American geneticist W. French Anderson performs the first gene therapy on a four-year-old girl with an immune-system disorder called ADA deficiency.
- 1990 UCSF and Stanford issued their 100th recombinant DNA patent and earning \$40 million from the licenses by 1991.

- 1990 First gene therapy attempted on girl with immune deficiency
- 1991 Analyzing chromosomes from women in cancer-prone families, Mary-Claire King, of the University of California, Berkeley, finds evidence that a gene on chromosome 17 causes the inherited form of breast cancer and also increases the risk of ovarian cancer.
- 1992 The U.S. Army begins collecting blood and tissue samples from all new recruits as part of a "genetic dog tag" program aimed at better identification of soldiers killed in combat.
- 1992 Calgene's Flavr Savr tomato, engineered to remain firm for a longer period of time, is approved for commercial production by the US Department of Agriculture.
- 1992 American and British scientists unveil a technique for testing embryos *in vitro* for genetic abnormalities such as cystic fibrosis and hemophilia.
- 1993 After analyzing the family trees of gay men and the DNA of pairs of homosexual brothers, biochemists at the U.S. National Cancer Institute report that at least one gene related to homosexuality resides on the X chromosome, which is inherited from the mother.
- 1993 George Washington University researchers clone human embryos and nurture them in a Petri dish for several days. The project provokes protests from ethicists, politicians and critics of genetic engineering.
- 1993 An international research team, led by Daniel Cohen, of the Center for the Study of Human Polymorphisms in Paris, produces a rough map of all 23 pairs of human chromosomes.
- 1994 The Flavr Savr tomato gains FDA approval
- 1994 The first linkage map of the human genome appears
- 1994 The European Union's first genetically engineered crop, tobacco, is approved in France.
- 1995 Researchers at Duke University Medical Center report that they have transplanted hearts from genetically altered pigs into baboons. All three transgenic hearts survived at least a few hours, proving that cross-species operations are possible.
- 1995 The first full gene sequence of a living organism is completed for *Hemophilus influenzae*.
- 1995 Former football player O.J. Simpson is found not guilty in a high-profile double-murder trial in which PCR and DNA fingerprinting play a prominent but apparently unpersuasive role.
- 1996 Genome of *Saccharomyces cerevisiae* is sequenced
- 1997 Researchers at Scotland's Roslin Institute, led by embryologist Ian Wilmut, report that they have cloned a sheep--named Dolly--from the cell of an adult ewe.
- 1998 Biologist Craig Venter announces ambitious plans to decode the entire human genome by 2001.
- 1998 University of Hawaii scientists, using a variation of Wilmut's technique, clone a mouse, creating not only dozens of copies but three generations of cloned clones.
- 1998 Two research teams succeed in growing embryonic stem cells.
- 1998 Scientists at Japan's Kinki University clone eight identical calves using cells taken from a single adult cow.
- 1998 The first animal genome (*C. elegans*) is sequenced
- 1999 1,274 biotechnology companies in the United States. At least 300 biotechnology drug products and vaccines currently in human clinical trials
Human Genome Project is on time and under budget, the complete human genome map expected in five years or less

- 2000 Biosafety Protocol is approved by 130 countries at the Convention on Biological Diversity in Montréal, Canada. The protocol agrees upon labeling of genetically engineered crops, but still needs to be ratified by 50 nations before it goes into effect.
- 2000 Celera Genomics completes sequencing of the Human Genome.

CHAPTER 3

GENOME STRUCTURE AND ORGANISATION- GENERAL

The eukaryotic genome is safeguarded by the nuclear membrane from exposure to cytoplasm. The constitution of nuclear genome of different organisms has different sequences, variable amount of DNA and the number of chromosomes. However, certain features are unique to eukaryotic genomes compared with prokaryotic genomes.

Major differences between prokaryotic and eukaryotic genomes

Characteristic	Prokaryotes	Eukaryotes
Genome size	600kb to 9.5 Mb	3 Mb to 140000 Mb
Average gene size/number	950bp/4300	2500bp/19000
Operon-like regulatory unit	General	No
Horizontal gene transfer	Significant	Negligible
Rate of non-coding sequences	Low	High
Intron	Rare	General
Redundant gene number	Rare	General
Ploidy level	Haploidy	Haploidy to Polyploidy
Chromosome number	One	More than one
Heterozygosity	No	Yes

The integrity of individual genes is being interrupted by multiple and identical copies of particular sequences and large blocks of non-coding sequences. The characteristics of eukaryotic genomes can be resolved by adopting various options to estimate the genome size and expose the sequence complexity.

What is a genome?

The conventional definition of genome is number of chromosomes representing haploidy. But the recent definition on a genome by Singer and Berg (1991) states the following: The term genome is used to describe the totality of chromosomes (in molecular term DNA) unique to a particular organism or any cell within the organism, as distinct from genotype, which is the information contained within those chromosomes.

The first genomes to be studied in depth at the molecular level were those of the bacterium *Escherichia coli* and its bacteriophage. The second group of genomes to be studied in great detail at the molecular level were those of mitochondria and chloroplasts. Finally attention has now turned to a detailed study of the nuclear genomes of selected eukaryotes: yeast (*Saccharomyces cerevisiae*), a nematode (*Caenorhabditis elegans*), a mammal (*Homo sapiens*) and plants (*Arabidopsis thaliana*) and rice (*Oryza sativa*).

Chromosome numbers and genome

Chromosomes are constructed of two broad classes of DNA - euchromatin and heterochromatin. Euchromatin is really only a functional description usually assumed as genetically active and less contracted than heterochromatin. Heterochromatin is region which has repeated sequences which are transcriptionally inactive, but are important to the functioning of chromosomes. These chromosomes constitute the genome of individual species and the genomic constitution varies in different organisms. Cytological studies revealed that not all chromosomes are similar and vary morphologically giving an option to categorize them into different types.

With this variation in chromosome structure, the number of chromosomes in each organism, otherwise called as ploidy level also varies. Double the dose of a genome [haploidy (n) - single set of chromosomes] is expected as the normal ploidy level and the ploidy of an organism having two sets of chromosomes is called as diploid ($2n$). The organisms having chromosome numbers deviating from the diploid numbers, are heteroploid. Organisms are classified as diploids and polyploids depending upon their chromosomes number. Those having exactly the triple or higher multiples of the haploid number are called polyploids [triploid ($3n$), tetraploid ($4n$), hexaploid ($6n$)]. Individuals with an increase or decrease of one or two chromosomes over the multiples of haploid number are called with prefixes hyper and hypo along the terms used to indicate their ploidy respectively ($3n+1$ - hypertriploid and $4n-1$ - hypotetraploid). Since this system of classification was confusing, more refined way of classifying the individuals based on chromosomes number came into existence and is given below.

Diploids	Individuals having chromosomes double the number of haploidy
Heteroploids	Individuals with chromosomes deviating the diploid number
Polyploids	Individuals with chromosomes multiples of haploid number, but more than diploid number ($3n$, $4n$, $5n...$)
Aneuploids	Individuals with chromosomes deviating from the exact multiples of the basic number ($x+1$, $x+2$, $x-1$, $x-2$)

Whatever may be the number of chromosomes, the basic organization of the chromosome is common for all organisms. DNA and basic proteins viz., histones are two major components of chromosomes and the cell nucleus. During the process of evolution, a gradual increase in the quantity of DNA, with minor adjustments (increase or decrease), has been speculated at every step of evolution with increasing complexity. It is observed that DNA

content per nucleus in organisms at various levels of evolution increased with increasing complexity

Nuclear DNA content and genomes

The total amount of DNA in the (haploid) genome is a characteristic of each living species known as its **C** value. The fact that the DNA content of an organism is much greater than that required to code for and regulate the production of all necessary proteins has been termed the **C** value paradox. In eukaryotes, the **C** value is defined as the amount of DNA per genome (1 **C** = haploid nucleus, 2**C** = diploid nucleus and 4**C** = nucleus which is just about to divide by mitosis). There is enormous variation in the range of **C** values, from as little as a mere 10^6 bp for a mycoplasma to as much as 10^{11} bp for some plants and animals. The following table summarizes the range of **C** values found in different plant species.

Species	pg/2C	Mbp/1C
<i>Arabidopsis thaliana</i>	0.30	145
<i>Arachis hypogaea</i>	5.83	2813
<i>Brassica juncea</i>	2.29	1105
<i>Cicer arietinum</i>	1.53	738
<i>Glycine max</i>	2.31	1115
<i>Gossypium hirsutum</i>	4.39	2118
<i>Helianthus annuus</i>	5.95	2871
<i>Hordeum vulgare</i>	10.10	4873
<i>Lycopersicon esculentum</i>	1.88	907
<i>Oryza sativa</i>	0.87	419
<i>Phaselous vulgaris</i>	1.32	637
<i>Pisum sativum</i>	8.18	3947
<i>Saccharum officinarum</i>	5.28	2547
<i>Triticum aestivum</i>	33.09	15966
<i>Vigna mungo</i>	1.19	574
<i>Vigna radiata</i>	1.20	579
<i>Zea mays</i>	4.75	2292

1 picogram (pg) = 965 Million base pairs (Mbp)

DNA sequences and genomes

DNA sequences can be divided into three groups depending on the number of times a sequence appears in the genome.

1. **Low or single copy DNA:** These are sequences encoding most enzyme functions. In general, these can constitute upto 50% of the total DNA.
2. **Middle repeat DNA:** These sequences code for most of the structural components of a cell. Thirty to forty percent of the genome may comprise middle repeat DNA

3. ***Highly repetitive DNA***: This is simple sequence DNA and is frequently non-coding. In mammals, this can constitute 20-50% of the genome, but in many plants this figure can exceed upto 80 percent. In general, DNA in the genome is arranged with single copy sequences interspersed with either repetitive or middle repeat DNA.

The complexity of any genome is assessed based on the proportion of above-mentioned sequences.

CHAPTER 4

STRUCTURE OF DEOXYRIBO NUCLEIC ACID

The fundamental chemical building block of deoxyribonucleic acid (DNA) is the nucleotide. A nucleotide consists of three parts: (1) a deoxyribose sugar, (2) a phosphate group that acts as a bridge between adjacent deoxyribose sugars, and (3) a nitrogen-containing pyrimidine or purine base.

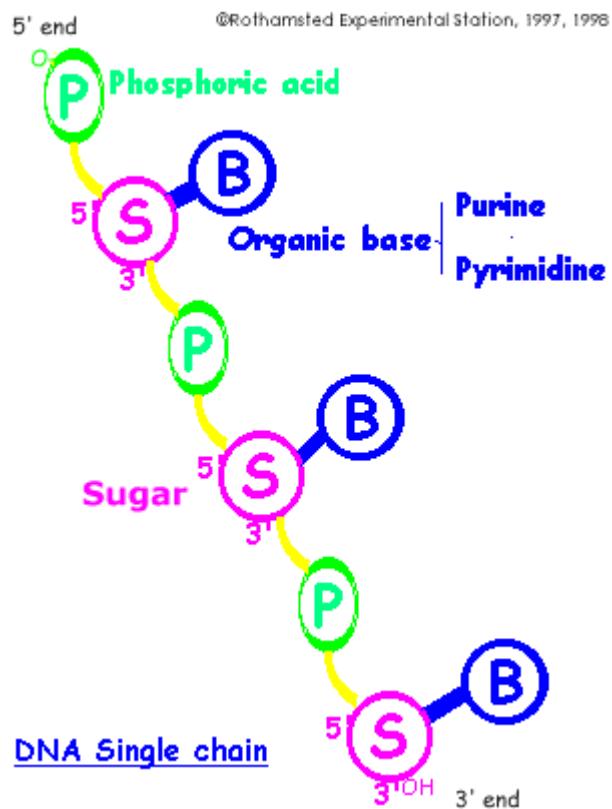


Fig 1. Components of DNA

Each deoxyribose sugar unit contains five carbon atoms joined in a ring structure with an oxygen atom. The carbon atoms of the deoxyribose sugar are designated by numbering them sequentially from one to five. The first carbon atom, the 1' carbon, is by definition the carbon atom covalently attached to one of four organic bases: guanine (G), adenine (A), thymine (T), or cytosine (C). Adenine and guanine are purines, and cytosine and thymine are pyrimidines. Phosphate groups are attached to the third (3') and fifth (5') carbon atoms. In DNA, the term nucleotide refers to the complete assemblage of a nitrogenous base (A, G, C, or T), a five-carbon deoxyribose sugar, and a phosphate group.

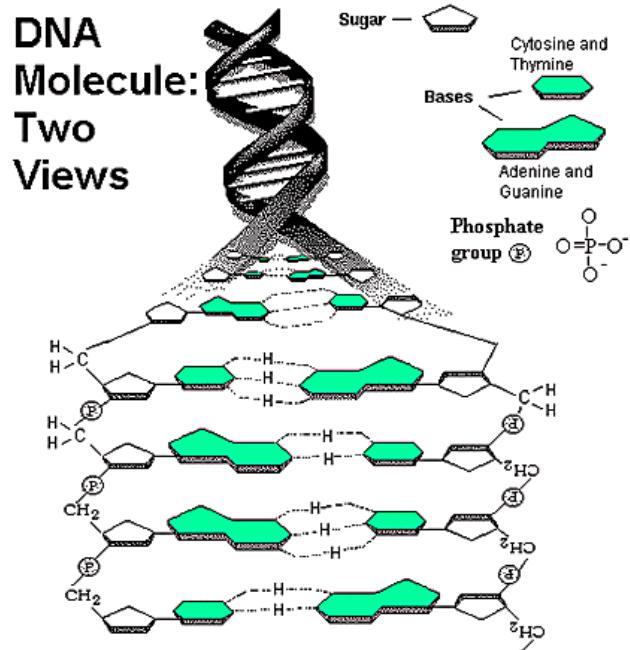


Fig 2. Components of DNA forming a double strand

The diameter of the DNA helix is 2 nm and the vertical rise per base pair is 0.34 nm . A DNA molecule is composed of two unbranched polynucleotide chains (strands) that wind about each other into a structure called a double helix. The structure of the DNA helix is stabilized by van der Waals forces, hydrogen bonds between complementary organic bases (a base pair), and hydrophobic interactions between the nitrogenous bases and the surrounding sheath of water. The alternating sugar-phosphate groups in each DNA strand form the so-called *backbone* of DNA, and they also confer a directionality. Suppose the double helical structure of DNA could be unwound and both strands of DNA laid side-by-side. Then, the sugar-phosphate linkages in one strand would proceed from the 5' to 3' carbon, and the sugar-phosphate linkages in the other strand would proceed in 3' to 5' direction.

The helical structure described by Watson and Crick, called B-DNA, is only one of several possible conformations. Other DNA conformations use the same nucleotides and molecular bonds, but the three-dimensional structure of the helix is different. At least six different DNA conformations (designated A, B, C, D, E, and Z) have been identified, but only the A, Z, and B conformations are found in nature. B-DNA is the most common form of DNA found in living organisms. The average diameter of B-DNA is about 2.0 nm. Other DNA conformations have diameters that range from about 1.8 to 2.3 nm. The average distance between adjacent nucleotides in the same strand of DNA (the *vertical rise*) is between 0.321 to 0.337 nm. The B-DNA helix makes one full revolution approximately every 10.1 to 10.6 nucleotides. The following table summarizes the features of the different forms of DNA.

Form	Direction	Bases/ 360° Turn	Helix Diameter
A	Right	11.0	23A
B	Right	10.0	19A
C	Right	9.3	19A
Z	Left	12.0	18A

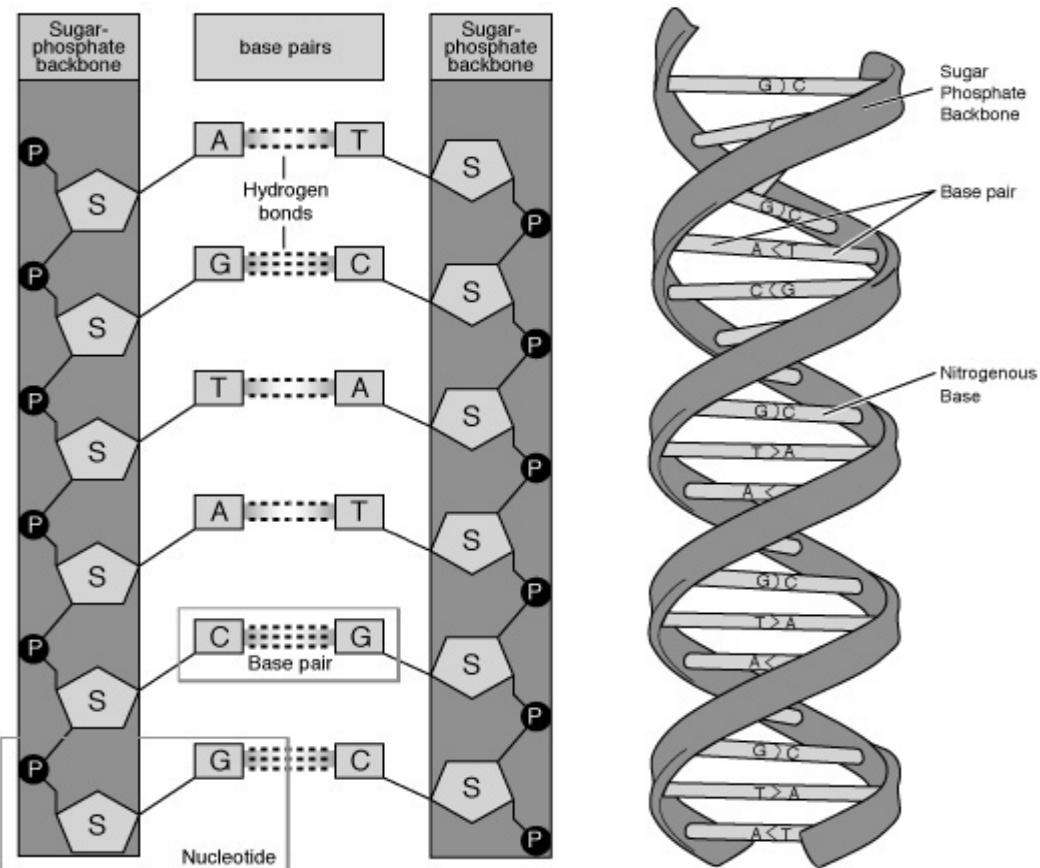


Fig 3. Model of DNA double helix

The hydrogen bonding between complementary base pairs is such that the most energetically stable DNA configuration is achieved when adenine pairs with thymine and guanine pairs with cytosine. Thus although the spatial requirements of B-DNA potentially allow four complementary base pairs to be formed (i.e., G-T, G-C, A-T, and A-C), only the G-C and A-T base pairs are normally found in DNA. Three hydrogen bonds stabilize G-C base pairs and two hydrogen bonds stabilize A-T base pairs. Because hydrogen bonding between base pairs

contributes to the stability of the DNA double helix, base sequence affects the stability of DNA. Namely, regions of the DNA with abundance in G-C base pairs are more stable than A-T rich regions of the DNA. In some prokaryotes, the greater mechanical stability offered by G-C rich sequences make it possible for them to exist near thermal springs too hot for most terrestrial organisms.

The most basic features of the DNA molecule are:

- two strands of DNA are arranged to form a right-handed helix
- the two strand polarities are opposite each other
- the sugar-phosphate backbone is oriented to the outside of the molecule
- the nitrogenous bases are in the middle
- there is a 2-fold axis of symmetry
- the bases are perpendicular to the axis of symmetry
- a major and a minor groove is located between the backbones on opposite strands of the helix.

Chromatin

The fundamental structural unit of chromatin is an assemblage, called the **nucleosome**, composed of five types of histones (designated H1, H2A, H2B, H3, and H4) and DNA. A nucleosome consists of approximately 1.8 turns of DNA wound around a core particle of histone proteins. The core particle is a roughly heart-shaped octamer of 4 types of histones: two each of the H2A, H2B, H3, and H4 proteins. The elliptically shaped histone core particle has a mass of approximately 108,600 Da and maximal dimensions of 11.0' 6.0 to 6.5 nm . Approximately 166 base pairs are bound to the nucleosome: 146 (± 1) base pairs are tightly bound to the core particle and the remaining 20 base pairs are associated with the H1 histone. This nucleosome structure is closely similar in all eukaryotes.

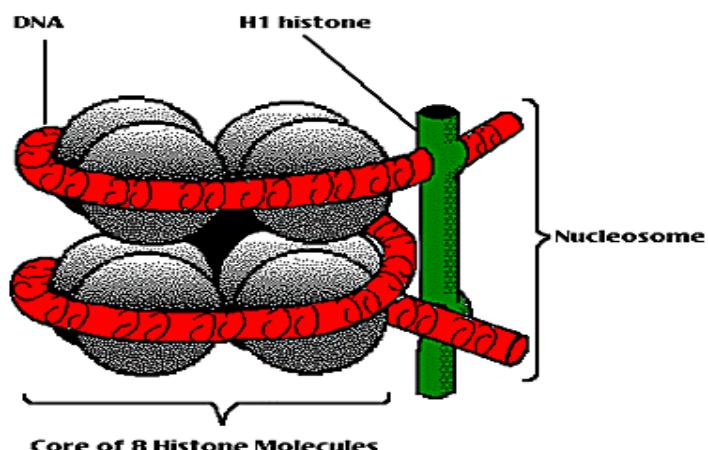


Fig 4. Nucleosome

DNA is bound to the histones through electrostatic forces between the negatively charged phosphate groups in the DNA backbone and positively charged amino acids (e.g., lysine and arginine) in the histone proteins. At times, histone proteins are modified by the addition of acetyl, methyl, or phosphate groups, and this alters the strength of the bonding between the histones and DNA. Modifications such as these are usually associated with the regulation of biological processes such as DNA replication, gene expression, chromatin assembly and condensation, and cell division.

The DNA between two nucleosomes is called the linker segment. This linker segment gives unfolded chromatin a *beads-on-a-string* appearance (the nucleosomes are the beads). The average linker length is variable in different species and for different cells within the same species , but they range from zero to a maximum of about 80 to 100 base pairs. Wolfe suggests that linker lengths tend to be shortest in the lower eukaryotes (e.g., yeast), intermediate in plants, and longest in the higher eukaryotes (i.e., animals). A correlation has also been noted in some species between increasing chromatin repeat length, which is the linker length plus 166 ± 1 base pairs (the number of base pairs bound to a nucleosome), and greater transcriptional activity. That is, the average repeat length of the chromatin usually increases as the level of transcriptional activity increases.

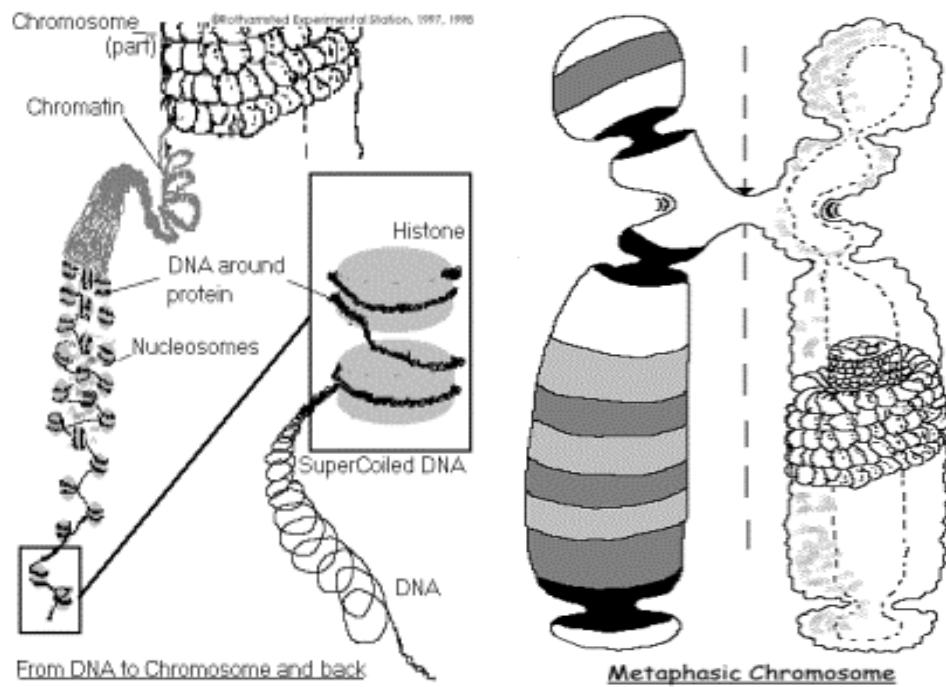


Fig 5. Packing of DNA into chromosomes

Nucleosomes appear to take up primarily random positions along the DNA without regard to the underlying nucleotide sequence. However in some cases, nucleosomes are passively restricted to certain sites in the DNA because regulatory (non-histone) proteins recognize and bind to specific sequences of DNA. When a regulatory protein binds to a specific sequence of DNA, it forces the nucleosomes on either side of the protein to align at regular intervals. This passive alignment of nucleosomes on either side of a regulatory protein is called nucleosome phasing. Because of random variations in the length of the linker segment of the chromatin, the effects of nucleosome phasing gradually decrease as the distance from a regulatory protein increases. Because nucleosome phasing is caused by the binding of regulatory proteins to the DNA, it is typically limited to a few locations in the DNA, usually near non-coding regions involved in gene regulation.

The attachment of regulatory proteins to the DNA can also create short segments of DNA completely free of nucleosomes . These segments are sensitive to chemical and enzymatic attack because they are poorly, if at all, protected by histone proteins (apparently, non-histone proteins offer little or no protection against chemical or enzymatic attack). These segments, called hypersensitive sites to indicate their extreme sensitivity to enzymatic digestion, typically appear and disappear in patterns that are coordinated with gene activity, i.e., more hypersensitive sites appear as gene activity increases.

When chromatin is isolated from cells under conditions similar to those expected in a cellular environment, chromatin does not have a beads-on-a-string structure. Instead, the majority of the chromatin appears to be composed of relatively smooth fibers about 25 to 35 nm in diameter. The folding mechanism and higher-order structure of the chromatin fiber is uncertain. But in one of the more widely accepted models, beads-on-a-string chromatin winds into a regular coil called the solenoid. The solenoid model of the chromatin fiber is proposed to contain six to eight nucleosomes per turn and has a total outside diameter of about 34 nm.

Although chromatin is organized at levels beyond the chromatin fiber (solenoid), the molecular arrangement of the highest levels of chromatin organization is not well understood. However, it has been suggested that chromatin fibers are divided into functional units called domains, and when the chromatin condenses, groups of proteins link these chromatin domains together and arrange the chromatin fibers into loops.

When DNA is stained, for example, with a fluorescent compound, it can be visually divided into condensed chromatin or unfolded chromatin. The condensed regions of chromatin are referred to as *heterochromatin*, and the unfolded regions are termed *euchromatin*. Although a visual system of chromatin classification such as this does not provide quantitative information about the structure of chromatin, it remains useful because some important biochemical processes occur in euchromatin but are absent or proceed very slowly in heterochromatin. In particular, highly compact heterochromatin is usually transcriptionally inert whereas transcriptionally active DNA has a more open euchromatin structure.

The ultimate answer for the structural-functional motif of chromatin organization is that DNA folded into heterochromatin is relatively inaccessible to the polymerases and proteins involved in DNA transcription. Thus before a polymerase can transcribe a section of DNA into ribonucleic acid (RNA), the heterochromatin must unfold into a more open euchromatin structure. Although the conversion of heterochromatin into euchromatin is a necessary step for DNA transcription, not all euchromatin is transcriptionally active; this likely indicates that the chromatin structures of transcriptionally active and inactive euchromatin are different.

There are two types of heterochromatin: constitutive heterochromatin and facultative heterochromatin. Facultative heterochromatin is condensed chromatin that unfolds and becomes transcriptionally active during some portion of the cell cycle. Constitutive heterochromatin refers to the chromatin that remains transcriptionally inert during the entire cell cycle. The bulk of constitutive heterochromatin is composed of non-coding DNA sequences that are repeated hundreds of thousands or even millions of times.

Because DNA replication requires polymerases and proteins similar to those required for DNA transcription, condensed heterochromatin most likely unfolds into euchromatin before replication proceeds. Because DNA with an unfolded beads-on-a-string structure is more accessible for replication than DNA folded into a chromatin fiber (solenoid), DNA about to be replicated and transcriptionally active DNA most likely have an unfolded beads-on-a-string (euchromatin) structure. Transcriptionally inactive euchromatin is tentatively identified as chromatin folded into a solenoid fiber, and heterochromatin is composed of solenoid fibers that are folded and condensed into a tightly packed mass of indeterminate structure.

CHAPTER 5

DIFFERENT TYPES OF DNA SEQUENCES

Eukaryotic genomes are much more complex than prokaryotic genomes. And further, plant genomes are more complex than other eukaryotic genomes. The total amount of DNA in the genome is a characteristic of each living species known as its 'C' value. There is enormous variation in the range of 'C' values, from as little as mere 10^6 bp for a mycoplasma to as much as 10^{11} bp for some plants and animals and is generally called as 'C' value paradox. The name came because of its puzzling variation among the organisms. The 'C' value is quite small. But in other cases the 'C' value is highly variable, thus making the assessment of genome complexity not accurate.

The complexity of any genome is assessed based on the proportion of 1) low or single copy DNA, 2) middle repeat DNA and 3) highly repetitive DNA above mentioned sequences. The proportion of different types of DNA sequences can be estimated by a technique popularly known as **Cot technique**. The Cot technique is based on the observation that extremes of temperature and pH can dissociate the double helix into two separate strands. Given suitable condition these strands will reanneal with their complementary sequences to form stable double helices and is called **Reassociation Kinetics**.

Genome complexity and Reassociation kinetics

Complexity of a genome is described as the total length of different sequences that are present. This is usually given in base pairs. This complexity estimated by measuring the kinetics of DNA reassociation.

When double stranded DNA is heated above the Tm (melting temperature) or treated with alkali, the hydrogen bond is disrupted and the strands separate. Conversely, when the temperature is lowered or the pH restored to near neutrality, complementary strands are capable of reassociating. The rate of reassociation is dependent on temperature, salt concentration and for complex sequences, on the length of DNA fragments.

If these parameters are controlled, then the reassociation rate is proportional to the concentration of reacting sequences and time and follows second order kinetics. This rate of reassociation is denoted by a value called **Cot** value. **Cot** value is inversely proportional to the reaction rate and for simple genome with no repetitive DNA is related to genome size.

The reassociation rate will be faster, when the denatured sequences are of repetitive types. The **Cot** value for these sequences will be between 10^{-4} and 2×10^{-2} with a $\text{Cot}_{1/2}$ value of 0.0013. These sequences represent 25 per cent of the total genome.

The next fraction is called the intermediate fraction. This represents 30 per cent of the DNA. It renatures at a moderate rate with **Cot** values between 0.2 and 100 with a $\text{Cot}_{1/2}$ value of 1.9. These DNA sequences are called as moderately repetitive sequences.

The last fraction to renature is called the slow fraction. This is 45% of the total DNA; it extends over a Cot value ranging from 100 to 10000, with a $\text{Cot}_{1/2}$ value of 630. This category of DNA sequences are generally of non-repetitive sequences.

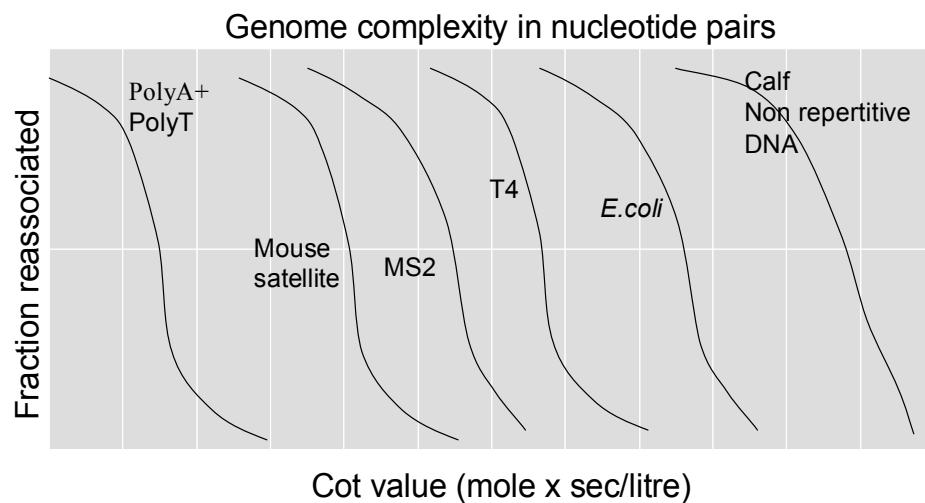


Fig 1: Effect of genome complexity on DNA renaturation rates.

From the above statements, it is clear that the reassociation kinetics of different DNAs depend on sequence complementarity. The reassociation rate is usually denoted by $\text{Cot}_{1/2}$ i.e. value indicating half-reassociation. Since the $\text{Cot}_{1/2}$ is the product of the concentration and time required to proceed halfway, a greater $\text{Cot}_{1/2}$ implies a slower reaction.

The $\text{Cot}_{1/2}$ is directly related to the amount of DNA in the genome. This reflects the situation in which, as the genome becomes more complex, there are fewer copies of any particular sequence within a given mass of DNA. For example, if the initial concentration (C_0) of DNA is 12 pg, it will contain 3000 copies of each sequence in a bacterial genome whose size is 0.004 pg. But the DNA of same concentration from a eukaryotic genome will have only 4 copies of each sequence whose size is 3 pg. It means, the same concentration of DNA from a eukaryotic genome will have a specific sequence 750 times (3000/4) lower compared to the bacterial genome.

Classification of DNA based on Reassociation Kinetics

Based on the reassociation kinetics, DNA sequences are divided into three classes. The three classes are termed the **fast, intermediate, and slow components**.

Single copy DNA sequences

Single-copy sequences are interspersed throughout the plant genome. These sequences are bounded by repeat sequences. The length of the single-copy regions varies widely among plant species. In general, two types of arrangements are recognized.

Short Period Interspersion - single copy sequences of 300-1200 bp are interspersed as islands among short lengths of repeat sequences

Long Period Interspersion - single copy sequences of 2000-6000 bp are interspersed as islands among repeat sequences

Moderately repetitive DNA sequences

- Present at between 10 - 10^5 copies per genome. Found throughout the euchromatin; average 300bp in size
- May be classified as : microsatellites/minisatellites (VNTR, DNA 'fingerprints') dispersed-repetitive DNA, mainly transposable elements (LINEs/ SINES)
- Also includes 'redundant' genes for histones, and ribosomal RNA and proteins, (gene-products present in cell in large numbers).
- Many moderately-repetitive sequences may be involved in regulation of gene expression. This is supported by their interspersion with single-copy sequences and location adjacent to structural genes.

Highly repetitive DNA sequences (satellite DNA)

- Present at $>10^6$ copies per genome
- Occurs as variable length motifs (5-100 bp), in long tracts of up to 100 Mb
- Most is located in heterochromatic regions around the centromere/ telomere.
- Postulated functions include structural or organisational roles, role in chromosome pairing, involvement in cross-over or recombination, junk.
eg. alpha-satellite DNA

This is a highly repetitive sequence, each centromere contains a tandem array of alpha-satellite repeats that extend for millions of base pairs and are arranged in a hierarchy of higher order repeats. These vary between 100-5000 on different chromosomes (0.2-10Mb). Some contain 17bp binding sites for the centromere-specific DNA binding protein. They have been recently cloned and used to construct artificial human chromosomes. **Much of the genome consists of moderately-repetitive sequences interspersed with single copy sequences.**

DNA Repetitive sequences

A conspicuous feature of eukaryotic genomes is repeated sequences. These may range in length from two to thousands of base pairs. They may be repeated in tens, hundreds or thousands of time within the genome. Repeated sequences (repetitive sequences) are often classified into two general types viz., moderately repeated and highly repeated.

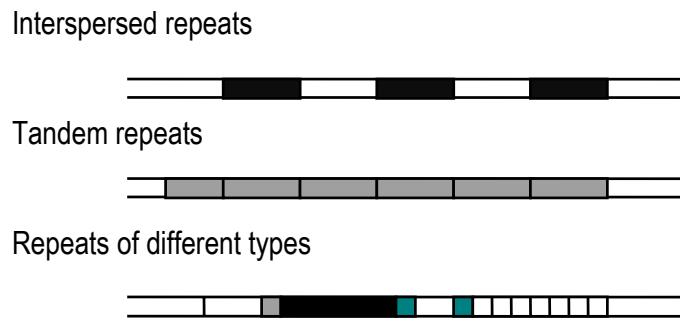


Fig 2. Various classes of DNA repeat sequences

Some repeated sequences have been shown to be interspersed between single copy DNA sequences. Evidence for interspersed repeats come from studies on the reassociation of DNA fragments of different lengths. When DNA is renatured to appropriate Cot values, only the repeated sequences form duplexes.

Inverted or reverse repeat sequences have also been detected by DNA reassociation; these sequences can form hairpin loops within one DNA strand. Interspersed repeats are, on average, less than 1000 base pairs long and are located between single copy regions of the order of 1000 - 2000 base pairs long. Adjacent reverse repeats of short nature are not detectable by HAP or S1 nuclease treatment, but are detectable by sequencing.

Other types of repeats occur in large number either as tandem repeats of one or a few sequences or as a more varied combination of several different sequences.

Many of these repeated sequences have been mapped on chromosomes by hybridization of cloned nucleic acid probes. The location of repetitive sequences is quite variable, it can be in telomeric, centromeric or interstitial heterochromatin sites. In addition, closely related sequences are often found at many separate locations among chromosomes and there is strong evidence that they can duplicate and change position during evolution.

Interspersed repeats

Interspersed repeats located at different regions. They may be called ***mobile elements*** or ***transposable elements***. Sometimes abbreviated as **TIREs** (Transposable and Interspersed Repetitive Elements). **Barbara McClintock** first discovered these elements in 1940s from the studies of corn. Subsequently, they were found in all kinds of organisms. Transposable elements are DNA sequences that can be inserted into many different sites in chromosomes which have the capacity of moving around the genome , either on their own or by the influence of other sequences. These elements are also called as controlling elements or insertion elements or mobile elements. For this reason they are sometimes called as jumping genes. Their movements can be detected physically, because when a mobile element is inserted in a structural gene, the gene may be inactivated. Alternatively, insertion of a transposable element into a control region can modify expression of a neighbouring gene. There are also cases where a transposable element has been shown to be excised from an inactive gene which then becomes active again.

Transposable elements of bacteria and their phage and plasmids encode a gene or genes required for transposition and exhibit inverted terminal repeat sequences also required for transposition. Transposable elements in plants and animals are of three main types: 1) mobile elements which encode a transposase and contain inverted repeats, 2) retrotransposons which encode reverse transcriptase and reproduce through an RNA intermediate and 3) transposable element found in mammals is the retrogenes which do not code their own reverse transcriptase but are generated through retrotranscription by a reverse transcriptase from another source. These mobile elements usually occur as repeated sequences dispersed throughout the genome. They can lead to small and large scale genetic rearrangements through homologous and non-homologous recombination. In mammals, the most common mobile elements are **Long Interspersed Elements (LINEs)** and **Short Interspersed Elements (SINEs)**.

Much of moderately-repeated DNA consists of transposable elements. The two major families, the long and short interspersed nucleotide elements (LINEs and SINEs), are represented in humans mainly by L1 and Alu elements respectively. Both types of element are considered to be retrotransposable (*i.e.* can replicate via an RNA copy reinserted as DNA by reverse transcription) and they have significant roles in genomic function and evolution. The most common examples in humans are L1 and Alu elements which are thought to have arisen by retrotransposition.

LINE (e.g L1)

- Repeated approx. 50,000x in the human genome (0.5% of total)
- Only 3000 of these are full length; the remainder is truncated, mostly at the 5' end.
- Complete element is 6kb in size and contains two open reading frames, one of which encodes a reverse transcriptase.
- AT-rich region is located near the 3' end of the element,
- Element is flanked by two short direct repeats

SINE (e.g. Alu family or Alu repeat- so named as they usually contain a target for the restriction enzyme Alu I)

- $5 \times 10^5 - 10^6$ copies in the haploid genome, with an average of one repeat every 4kb (1 - 10 % total)
- Not found within coding regions but often present in the transcription unit, within introns and occasionally in non-translated regions of the mRNA.
- All contain 290bp consensus sequence which consist of two tandem repeats of a 130bp sequence, one of which has a 32bp deletion.
- Elements are flanked by direct repeats
- Each repeat unit has an AT-rich region that suggests a poly A tail
- 5' end resembles a pol III promoter region.
- The origin of Alu can be traced to the terminal portions of the signal recognition particle 7SL RNA, hence it is probably a pseudogene formed by loss of the 7SL RNA central fragment.

Tandem repeats

Tandem repeats are an array of consecutive repeats. They include three subclasses: satellites, minisatellites and microsatellites. The name "satellites" comes from their optical spectra. Subclasses are defined on the basis of the size of a repetitive region, as discussed below.

The size of a satellite DNA ranges from 100 kb to over 1 Mb. In humans, a well known example is the **alploid DNA** located at the centromere of all chromosomes. Its repeat unit is 171 bp and the repetitive region accounts for 3-5% of the DNA in each chromosome. Other satellites have a shorter repeat unit. Most satellites in humans or in other species are located at the centromere.

Minisatellites

The size of a minisatellite ranges from 1 kb to 20 kb. One type of minisatellites is called **variable number of tandem repeats (VNTR)**. Its repeat unit ranges from 9 bp to 80 bp. They are located in non-coding regions, and thus the number of repeats for a given minisatellite may differ between individuals. This feature was first discovered by A. J. Jeffreys in 1985, and used for **DNA fingerprinting (= DNA profiling/DNA typing)**, which has found many applications.

Microsatellites

Microsatellites are also known as **short tandem repeats (STR)** or **Simple Sequence Repeats (SSR)**, because a repeat unit consists of only 1 to 6 bp and the whole repetitive region spans less than 150 bp. Similar to minisatellites, the number of repeats for a given microsatellite may differ between individuals. Therefore, microsatellites can also be used for DNA fingerprinting. In addition, both microsatellite and minisatellite patterns can provide information about paternity.

Repetitive DNA sequences are just a junk?

The existence of large amounts of repetitive DNA sequences in eukaryotes led to explore their role in eukaryotic genome organization and expression. In humans, up to 97% of the genome remains as repetitive DNA which are non-coding. The recent studies indicated that the repetitive DNA sequences [e.g. (gt)_n(ga)_m] have some major role in genome organization and gene expression. The repetitive DNA sequences bind nuclear proteins and show characteristics of a specific DNA-protein interaction. What functions these DNA-protein interactions exhibit has not been determined. However, there are many other examples of DNA-protein interaction, which exhibit regulatory control of DNA transcription. More definitive studies have shown that non-coding DNA provides structure to DNA so that it can perform many functions, which would be impossible without some form of structure.

One of the readily apparent differences between prokaryotic and eukaryotic DNA is that eukaryotic DNA is organized into chromosomes, which is further organized into chromatin code. This kind of structure does not "just happen" for DNA - it requires specific design. The

coding regions of DNA are concentrated in the chromosomal regions, which are the richest in G (guanine) and C (cytosine) and seem to correspond to the telomeric regions of certain chromosome arms (T-bands). Scientists have genetically modified and therefore removed a single telomere of one chromosome in yeast cells. The elimination of the telomere caused cell cycle arrest (stopping of cell division), indicating that telomeres help cells to distinguish intact chromosomes from damaged DNA. In the cells that recovered from the arrest the chromosome was eventually lost, demonstrating that telomeres are essential for maintaining chromosome stability. Therefore, non-coding DNA is absolutely necessary for chromosomal structure and function.

The chromosomes are replicated and segregated during mitosis (cell division). Complex interactions occur between the centromeres of chromosomes and the spindles to which they attach. These centromeres form an integrated protein/DNA complex, which is required for chromosomal movement during mitosis. In interphase nuclei, orderly transcription and replication involve highly folded chromosomal domains containing hundreds of kilobases of DNA. Specific non-coding DNA sequences within selected chromosome domains participate in more complex levels of chromosome folding, and index different genetic compartments for orderly transcription and replication

Recent advances have demonstrated that non-protein-coding DNA provides the structural basis of the metaphase chromosomal banding pattern. CpG islands, DNA loops, and matrix attachment sites form the basis of G versus R banding patterns, revealing how non-coding DNA forms the basis of chromosomal structure.

It appears that heterochromatin, composed of what was once thought to be junk DNA, may have some role in suppression of gene(s) and/or spreading of inactivation, if genes are embedded within the heterochromatic region. In a recent study, investigators examined, through genetic engineering, the relationship between exon (protein coding DNA) and intron (non-coding DNA) size in pre-mRNA (messenger RNA, from which protein translation is accomplished) processing. Exons were placed in vertebrate genes along with small and large introns. Both exon and intron size influenced splicing phenotype, such that when introns were large, large exons were skipped; when introns were small, the same large exons were included. These results indicated that non-coding introns can control the recognition and transcription of exons (protein-coding DNA).

There is growing evidence that noncoding DNA plays a vital role in the regulation of gene expression during development. Therefore, non-coding DNA regulates the vital roles of development and embryogenesis.

A recent study has shown that genes (as many as five at a time) are found within the introns of other genes. This kind of arrangement results in the simultaneous expression of all of these genes during transcription of the gene in question. Such regulatory control is rather remarkable, suggesting intelligent designed as opposed to random chance. Some of the non-coding DNA is loop code for single-stranded RNA-protein interactions. The codes are degenerate and corresponding messages are not only interspersed but also actually overlap, so that some nucleotides belong to several messages simultaneously. Tandemly repeated sequences frequently considered as functionless "junks" are found to be grouped into certain

classes of repeat unit lengths, indicating functional involvement of these sequences. It is likely these tandem repeats play the role of weak enhancer-silencers that modulate, in a copy number-dependent way, the expression of proximal genes.

Well over 700 studies (over 100 in the last year) have demonstrated the role of non-coding DNA as enhancers for transcription of proximal genes. Another 60+ studies have demonstrated the role of non-coding DNA as silencers for suppression of transcription of proximal genes. In addition, there are 3' and 5' untranslated regions (UTR) which regulate translation of proteins. Certain trans-acting binding proteins bind to the 3' and 5' UTRs of proximal and distal genes to regulate their translation. Another role for the 3' and 5' UTR is to regulate the rate of mRNA decay, which has now been shown to be a precise process dependent on a variety of specific cis-acting sequences and trans-acting factors.

The roles of non-coding DNA are so numerous and pervasive that evolutionary studies are now looking at these sequences for patterns of "**concerted evolution**" In summary, the non-coding DNA, contrary to statements by evolutionists, is not useless, but is, in fact, **required** for genomic functionality, therefore providing evidence of intelligent design. Thus, the "junk" DNA is really some rather amazing "junk."

CHAPTER 6

ORGANELLE GENOMES

Eukaryotic genomes are multicomponent. The majority of genetic information is located in the nucleus and inherited according to Mendel's laws. However, there are small but essential genomes located in the cellular organelles - the mitochondria and plastids. The **mitochondria**, which have a role in the oxidative degradation of nutrient molecules, also contain DNA, called the **mitochondrial DNA (mtDNA)**. Eukaryotic cells that are capable of photosynthesis contain **chloroplasts with chloroplast DNA (cpDNA)**. Plant cells, therefore, have three separate genetic systems.

Organelle genomes differ from nuclear genomes in a number of important ways. They are small relative to the nuclear genome. There are multiple organelles per cell and each organelle contains from 20 to 20,000 organelle genomes depending upon the cell type. Organelle genomes are organized into structures called nucleoids. Variations observed among the size of three different genomes are given below.

Type of DNA	Organism	size in base pairs
chromosomal DNA	mammals	6×10^9
	plants	$2 \times 10^8 - 2 \times 10^{11}$
	fungi	$2 \times 10^7 - 2 \times 10^8$
mitochondrial DNA	animals	$16 \times 10^3 - 19 \times 10^3$
	higher plants	$150 \times 10^3 - 250 \times 10^4$
	fungi	$17 \times 10^3 - 78 \times 10^3$
	green alga	16×10^3
chloroplast DNA	protozoa	$22 \times 10^3 - 40 \times 10^3$
	higher plants	$120 \times 10^3 - 200 \times 10^3$
	green alga	180×10^3

Organelle genomes are inherited in a non-Mendelian fashion. While inheritance is usually through the maternal parent only, there are a number of interesting exceptions in the higher plants. Mitochondria and plastids are inherited independently. Any combination is possible. For example: *Medicago* (alfalfa) plastids are inherited from both parents (biparental) and mitochondria are inherited from the maternal parent. In some conifers, both plastids and mitochondria are paternally transmitted. A variety of mechanisms exist to actively exclude organelles of one parent or the other.

Organelle genomes encode necessary but insufficient information for organelle biogenesis and function. Many nuclear gene products are required to express organelle genes and to assemble the organelle enzyme complexes. These nuclear gene products are translated on cytosolic ribosomes and imported into the organelle post-translationally.

Organelle structure and function

Before diving into the organelle genomes, their origin, evolution and expression, it's worthwhile to consider the organelle structures and functions that these genomes support.

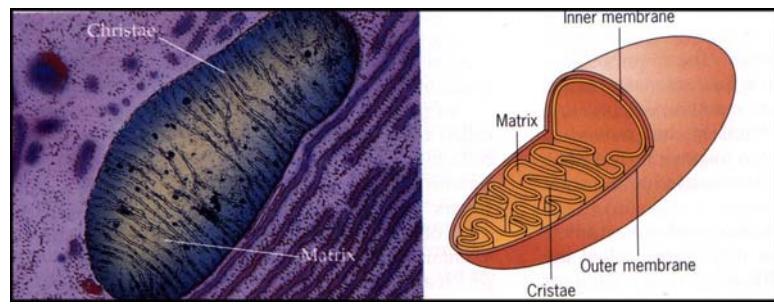


Figure 1. Structure of a mitochondrial

Mitochondria are the site of the complementary process of respiration in which carbohydrates and O₂ are consumed to extract energy with CO₂ evolved. Mitochondria contain an inner and an outer membrane. The inner membrane is in-folded to form structures called cristae. Respiratory electron transfer takes place on the inner membrane and the tricarboxylic acid (TCA) cycle reactions take place in the matrix. Mitochondria are a primary site for folate and thymidylate (purine and pyrimidine precursor) biosynthesis in plants. Mitochondria also carry out a key enzymatic step in the synthesis of ascorbate, one of the primary anti-oxidants of plant cells and were recently shown to be an important site of iron-sulfur center assembly in the plant cell.

Plastids, in particular, can take on a number of different forms and/or functions. They are the primary sites of anabolic metabolism in the plant cell. They are the primary sites of amino acid biosynthesis. They can become specialized for the synthesis and storage of starch (amyloplast), fatty acids (leucoplasts) and carotenoids (chromoplasts).

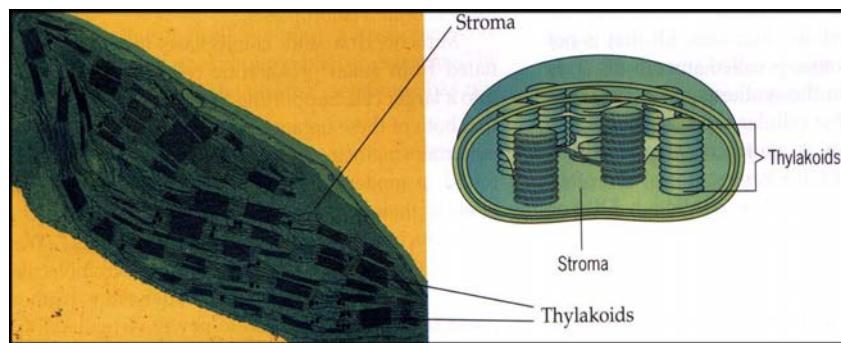


Figure 2. Structure of a chloroplast

Chloroplasts (type of plastids with chlorophyll) are best known as the site of photosynthesis in which light energy, CO₂ and H₂O are used to produce complex carbohydrates with O₂ evolved. Chloroplasts are surrounded by a double membrane, the site of protein import machinery. Light reactions of photosynthesis occur on internal membranous structures - the thylakoids. The dark or carbon reactions take place in the stroma.

The processes of respiration and photosynthesis are obligatory in higher plants. There are unicellular organisms that are facultative anaerobes (eg. *Saccharomyces cerevesiae*) in which respiration is expendable and facultative heterotrophs (eg. *Chlamydomonas reinhardtii*) in which photosynthesis is expendable.

The endosymbiont origin of organelles

A vast body of evidence from molecular and cell biology supports the endosymbiont hypothesis of organelle origin. Plastids and mitochondria evolved from separate endosymbiotic events in which the endosymbionts became highly specialized to perform the processes of photosynthesis and respiration, respectively. Ribosomal RNA phylogenies indicate the closest living relatives of the plastids are the cyanobacteria and the closest living relatives of the mitochondria are the alpha division of the purple bacteria (alpha-proteobacteria). Interestingly, these include modern-day symbionts such as rhizobium and obligate intracellular parasites such as rickettsia.

Endosymbiotic events also made major contributions to plant nuclear genomes, including genes that have nothing to do with organelle function.

Recent evidence demonstrates that the modern-day eukaryotic genome contains significant coding information derived from **both** archaeabacterial genomes (originally considered to be the ancestral nuclear genome) **and** the alpha-proteobacteria. Genes derived from alpha-proteobacteria include many having nothing to do with mitochondrial function. These observations support the "big-bang" theory of eukaryotic origin in which acquisition of the endosymbiont destined to become the mitochondria was the defining event in the origin of the modern-day eukaryotic cell.

Mitochondrial DNA

Mitochondrial DNA was discovered in the 1960s, initially through electron micrographs that revealed DNA-like fibres within the mitochondria. Later, these fibres were extracted and characterised by physical and chemical procedures. Mitochondrial DNA molecules vary enormously in size, from about 16 to 17 kb in vertebrate animals to 2500 kb in some of the flowering plants. Each mitochondrion appears to contain several copies of the DNA, and because each cell usually has many mitochondria, the number of mtDNA molecules per cell can be very large. In a vertebrate oocyte, for example, it has been estimated that as many as 108 copies of the mtDNA are present. Somatic cell, however, have fewer copies, perhaps less than 1000.

Most mtDNA molecules are circular, but in some species, such as *Chlamydomonas reinhardtii* and the ciliate *Paramecium aurelia*, they are linear. The circular mtDNA molecules appear to be organised in many different ways. The simplest arrangement is that seen in the vertebrates, where 37 distinct genes are packed into a 16-17 kb circle leaving little or no space between genes. The most complex arrangement exist in some of the flowering plants, where an unknown number of genes are dispersed over a very large circular DNA molecule hundreds or thousands of kilobases in circumference.

In fact, in these plants the mitochondrial genes may become separated onto different circular molecules by a process of intramolecular recombination. This recombination is mediated by repetitive sequences located in the mtDNA. An exchange between two of the repetitive sequences can partition the master DNA circle into two smaller circles. In some species, several DNA circles of different sizes are formed by recombinations between pairs of repetitive sequences located at different positions around the master DNA circle.

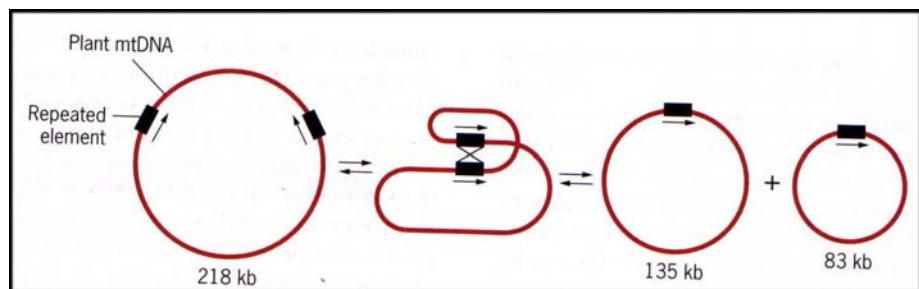


Figure 3. Intramolecular recombination leading to the formation of smaller mtDNA molecules

The fine details of mtDNA organization have been studied by DNA sequencing. Animal mtDNA is small and compact. In human beings, for example, the mtDNA is 16,659 base pairs long and contains 37 genes, including two that encode ribosomal RNAs, 22 that encode transfer RNAs and 13 that encode polypeptides involved in oxidative phosphorylation. In mice, cattle, and frogs, the mtDNA is similar to that of human beings- an indication of a basic conservation of structure within the vertebrate subphylum. Invertebrate mtDNA is about the same size as vertebrate mtDNA, but it has a somewhat different genetic organisation. These differences seem to have been caused by structural arrangements of the genes within the circular mtDNA molecule.

In fungi, the mtDNA is considerably larger than it is in animals. Yeast, for example, possesses circular mtDNA molecules 78 kb long. These molecules contain at least 33 genes, including two that encode transfer RNAs, 23 to 25 that encode transfer RNAs, one that encodes a ribosomal protein and seven that encode different polypeptides involved in oxidative phosphorylation. The yeast mtDNA is larger than animal mtDNA because several of its genes contain introns and there are long noncoding sequences between some of the genes. Animal mtDNA does not contain introns.

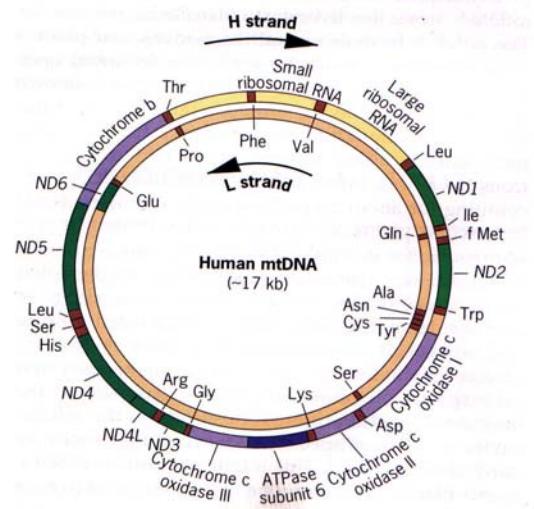


Figure 4. The fine structure of human mtDNA

Plant mtDNA is much longer than the mtDNA of other organisms. It is also more variable in structure. The mtDNA from liverwort, *Marchantia polymorpha*, is 186 kb circular molecule with 94 substantial open reading frames (ORFs), some corresponding to known genes and others having unassigned genetic functions. These latter ORFs are therefore called URFs, for unassigned reading frames. Thirty two distinct introns have been found in the *Marchantia* mtDNA, accounting for about 20 per cent of the entire molecule. In vascular plants, the mtDNA is larger than it is in *Marchantia*; for example it is 570 kb in maize and a 300 kb circle in the watermelon. Higher plant mtDNA molecules contain many noncoding sequences, including some that are duplicated.

Chloroplast DNA

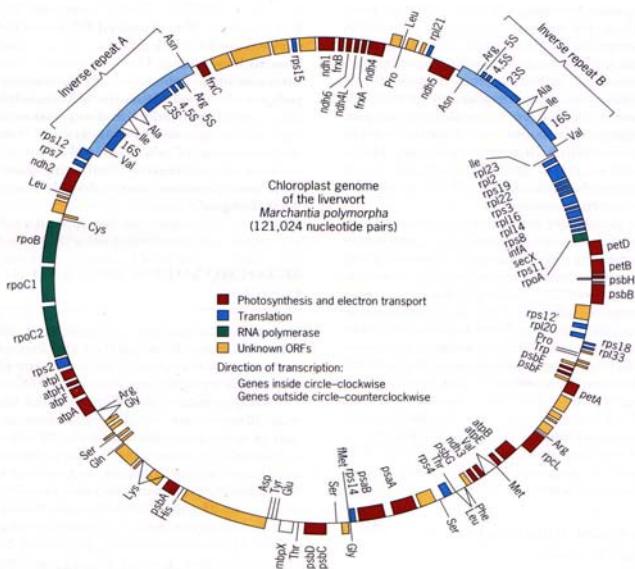


Figure 5. The fine structure of *Marchantia* cpDNA

In higher plants, cpDNA typically range from 120 to 160 kb in size and in algae, from 85 to 292 kb. In few species of green algae in the genus *Acetabularia*, the cpDNA is much larger, about 2000 kb. Among the 200 or so species of plants whose chloroplast DNA has been at least partially characterised, the cpDNA seems to be organised as a covalently closed circular molecule. However, in some species, especially those with large cpDNAs, a linear arrangement cannot be ruled out.

The number of cpDNA molecules in a cell depends on two factors: the number of cpDNA molecules within each chloroplast. For example, in the unicellular alga *Chlamydomonas reinhardtii* there is only one chloroplast per cell, and it contains about 100 copies of the cpDNA. In *Euglena gracilis*, another unicellular organism, there are about 15 chloroplasts per cell and each contains about 40 copies of the cpDNA.

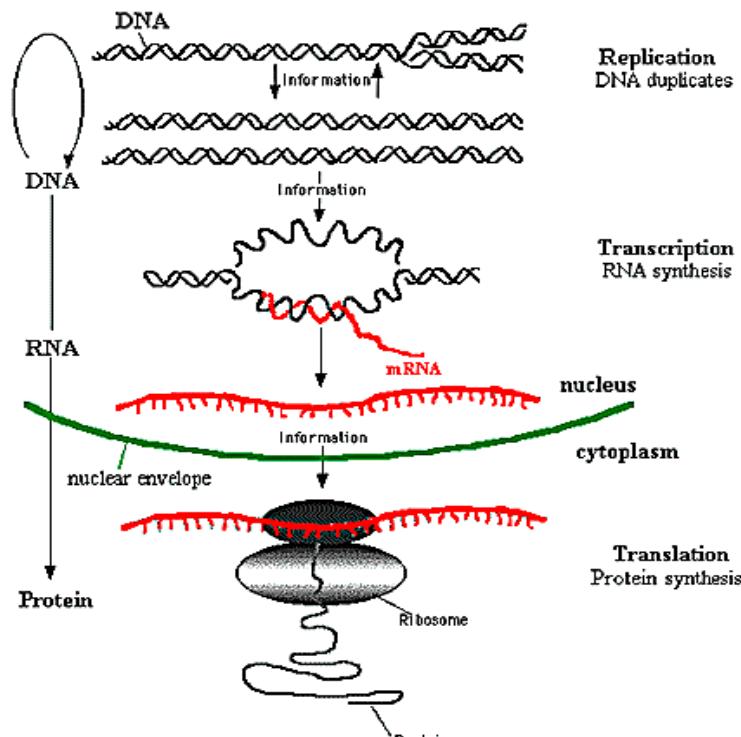
All cpDNA molecules carry basically the same set of genes, but in different species of plants these genes are arranged in different ways. The basic gene set includes genes for ribosomal RNAs, transfer RNAs, some ribosomal proteins, various polypeptide components of the photosystems that are involved in capturing solar energy, the catalytically active subunit of the enzyme ribulose 1,5-diphosphate carboxylase and four subunits of a chloroplast specific RNA polymerase.

The sequencing of cpDNA of *Nicotiana tabacum* and *Marchantia polymorpha* indicated the presence of 1,55,844 and 1,21,024 bp respectively.

CHAPTER 7

CENTRAL DOGMA OF MOLECULAR BIOLOGY: DNA REPLICATION

The processes of replication DNA, Transcription of DNA to RNA and translation of RNA to protein are together described as the central dogma of molecular biology. This dogma forms the backbone of molecular biology and is represented by four major stages.



1. The DNA replicates its information in a process that involves many enzymes: **replication**
2. The DNA codes for the production of messenger RNA (mRNA) during **transcription**
3. In eucaryotic cells, the mRNA is **processed** (essentially by splicing) and migrates from the nucleus to the cytoplasm.
4. Messenger RNA carries coded information to ribosomes. The ribosomes "read" this information and use it for protein synthesis. This process is called **translation**.

Proteins do not code for the production of protein, RNA or DNA. They are involved in almost all biological activities, structural or enzymatic.

DNA replication

During DNA replication, each strand of the existing DNA molecule acts as a template for the production of new strand, and the sequence of nucleotides of the synthesised (growing) strand is determined by base complementarity. This mechanism of DNA replication is called semiconservative replication, because each of the strands of the parental double helix is conserved (or the double helix is half-conserved) during the process.

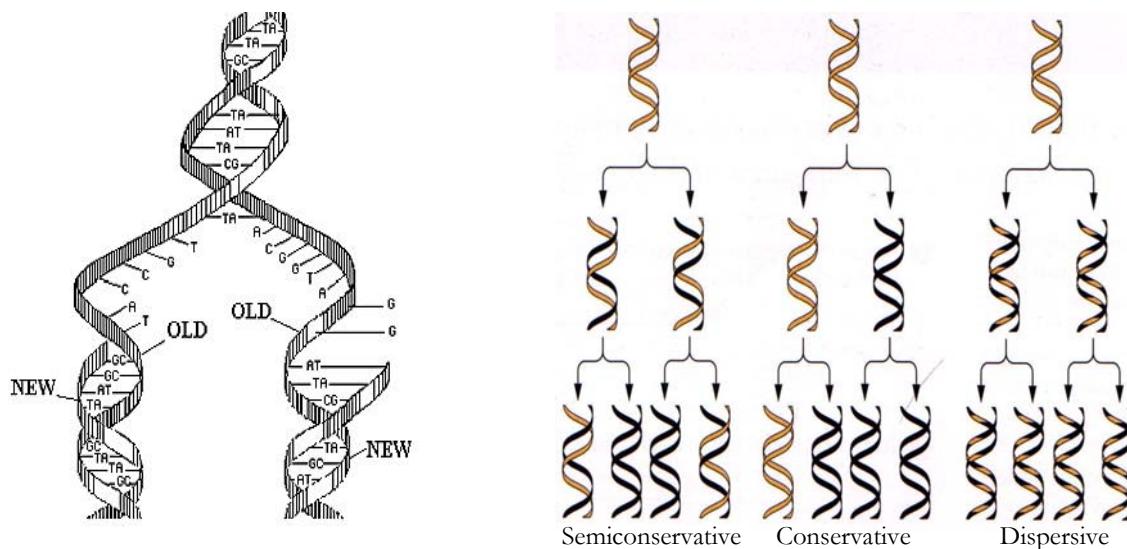


Figure 2. DNA replication: different options

Hypothetically, DNA replication could occur by any of three different mechanisms. In addition to the **semiconservative mechanism** proposed by Watson and Crick, replication could occur by either a **conservative mechanism**, with the parental double helix being conserved and directing the synthesis of a new progeny double helix, or a **dispersive mechanism**, with segments of parental and progeny strands interspersed as a result of the synthesis and rejoicing of short segments of DNA.

DNA replication is semiconservative: Proof

Experimental evidence to establish the semiconservative replication of DNA came from Mathew Meselson and Franklin Stahl in the year 1958. Meselson and Stahl grew *Escherichia coli* cells for many generations in a medium in which ^{15}N (heavier) had been substituted for the normal ^{14}N (lighter). This made the bacterial DNA denser than normal because atoms of ^{15}N were used instead of the normal lighter ^{14}N atoms in all the nitrogenous bases of DNA. They then transferred the multiplying cells to fresh growth medium containing normal ^{14}N . During the cultivation period, they harvested bacteria at regular intervals for analysis. They assumed for each round of bacterial reproduction the DNA would be duplicated. They predicted that the DNA of cells grown on medium containing ^{15}N would have a greater density than the DNA of cells grown on medium containing

^{14}N . These heavier and lighter molecules can be separated by Cesium Chloride Equilibrium Density Gradient Centrifugation.

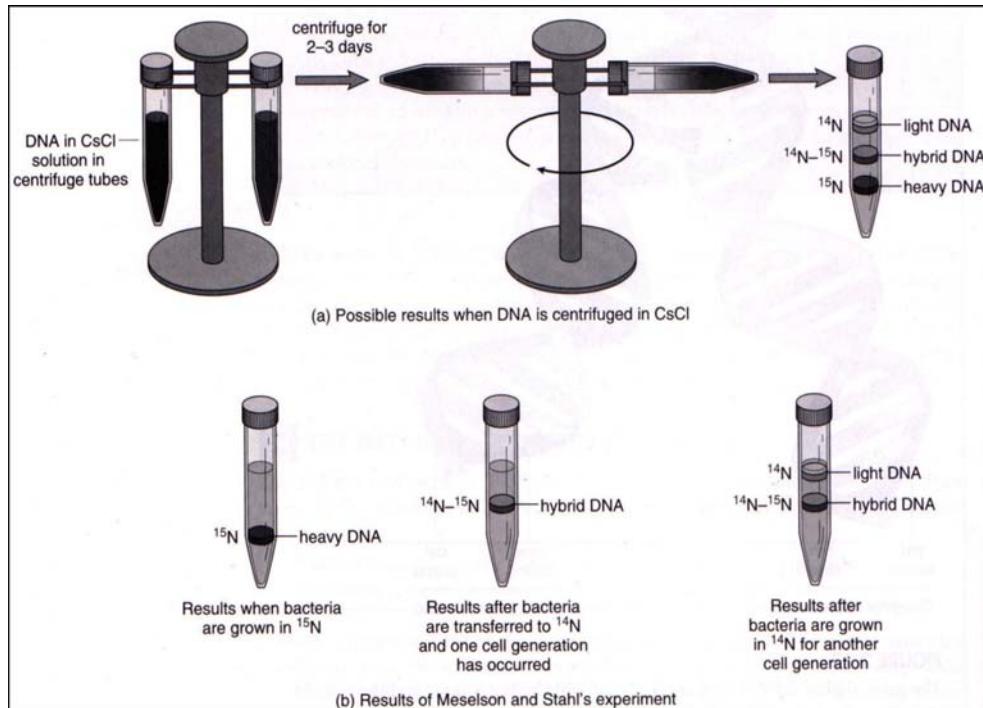


Figure 3. Meselson and Stahl's Cesium Chloride Equilibrium Density Gradient Centrifugation experiment

Meselson and Stahl observed that the DNA of the parental bacteria (grown on ^{15}N containing medium) was heavier. When these bacteria were transferred to ^{14}N containing medium, the DNA harvested from them was lighter than the DNA of the bacteria from ^{15}N containing medium. Subsequent culture of the first generation bacteria on ^{14}N containing medium and the analysis of DNA from them indicated that there were two types of DNA molecules—one having the density similar to the first generation bacteria and the other lighter than the DNA of both parental and first generation bacteria. The following table summarises the results of Meselson and Stahl experiment conducted for establishing that the DNA replication is semiconservative.

Generation	Medium	DNA Density	Reason
I	With ^{15}N	All heavy	DNA had ^{15}N atoms
II	With ^{14}N	Intermediate	DNA had ^{15}N and ^{14}N atoms
III	With ^{14}N	Intermediate	DNA had ^{15}N and ^{14}N atoms
		Lighter	DNA had ^{14}N atoms

From the above experiment, Meselson and Stahl concluded that during the replication of the DNA each parent DNA strand was serving as a template for a new, complementary DNA

strand, which then combined with the parent DNA strand to form a double helix. The method of DNA replication determined by Meselson and Stahl has been termed semiconservative because one strand of the original DNA (semi) is brought over (conserved) to each new double helix.

Enzymes involved in DNA replication

Replication of DNA occurs during the process of normal cell division cycles. Because the genetic complement of the resultant daughter cells must be the same as the parental cell, DNA replication must possess a very high degree of fidelity. The entire process of DNA replication is complex and involves multiple enzymatic activities. The mechanics of DNA replication was originally characterized in the bacterium, *E. coli* which contains 3 distinct enzymes capable of catalyzing the replication of DNA. These have been identified as DNA polymerase (pol) I, II, and III.

Function	DNA polymerase I	DNA polymerase II	DNA polymerase III
5'→3' Polymerase activity	Yes	Yes	Yes
5'→3' Exonuclease activity	Yes	No	No
3'→5' Exonuclease activity	Yes	Yes	Yes

Pol I, otherwise called as “Kornberg’s Enzyme”, is the most abundant replicating activity in *E. coli* but has as its primary role to ensure the fidelity of replication through the repair of damaged and mismatched DNA. Replication of the *E. coli* genome is the job of pol III. This enzyme is much less abundant than pol I, however, its activity is nearly 100 times that of pol I.

DNA Polymerase's Polymerase activity

The enzyme requires the 5'- triphosphates of each of the four nucleotides (dATP, dTTP, dGTP and dCTP), Mg²⁺ and preexisting DNA. This DNA should provide two essential components, one serving a primer function and the other a template function.

- The primer DNA provides a terminus with a free 3'-OH to which nucleotides are added during DNA synthesis. DNA polymerase cannot initiate the synthesis of DNA *de novo*. It has an absolute requirement for 3'-hydroxyl on preexisting DNA chain. DNA polymerase catalyses the formation of a phosphodiester bonds between the 3'-OH at the end of the primer DNA chain and the 5'phosphate of incoming deoxyribonucleotide.
- The template DNA provides the nucleotide sequence that specifies the complementary sequence of the new DNA chain. DNA polymerase requires a DNA template whose base sequence dictates, by its base pairing potential, the synthesis of a complementary base sequence in the strand being synthesised.

DNA Polymerase's Exonuclease activity

The main enzymatic activity of DNA polymerases is the $5' \Rightarrow 3'$ synthetic activity. However, DNA polymerases possess two additional activities of importance for both replication and repair. These additional activities include a $5' \Rightarrow 3'$ exonuclease function and a $3' \Rightarrow 5'$ exonuclease function. The $5' \Rightarrow 3'$ exonuclease activity allows the removal of ribonucleotides of the RNA primer, utilized to initiate DNA synthesis, along with their simultaneous replacement with deoxyribonucleotides by the $5' \Rightarrow 3'$ polymerase activity. The $5' \Rightarrow 3'$ exonuclease activity is also utilized during the repair of damaged DNA. The $3' \Rightarrow 5'$ exonuclease function is utilized during replication to allow DNA polymerase to remove mismatched bases. It is possible (but rare) for DNA polymerases to incorporate an incorrect base during replication. These mismatched bases are recognized by the polymerase immediately due to the lack of Watson-Crick base-pairing. The mismatched base is then removed by the $3' \Rightarrow 5'$ exonuclease activity and the correct base inserted prior to progression of replication.

The ability of DNA polymerases to replicate DNA requires a number of additional accessory proteins. The combination of polymerases with several of the accessory proteins yields an activity identified as **DNA polymerase holoenzyme**. These accessory proteins include:

1. Helicase
2. Single strand-binding proteins
3. Topoisomerases
4. Primase
5. Processivity accessory proteins
6. DNA ligase
7. Uracil-DNA N-glycosylase

Complex Replication Apparatus

The DNA replication process mediated by the component such as 1) Helicase, 2) Single strand-binding (SSB) proteins, 3) Topoisomerases, 4) Primase, 5) Processivity accessory proteins, 6) DNA ligase and 7) Uracil-DNA N-glycosylase has following stages:

1. Unwinding of DNA with helicases
2. Binding of SSB proteins
3. Transient breaking of DNA by topoisomerases
4. RNA priming and formation of primosomes
5. Formation of replisomes
6. Covalent closure of nicks by DNA ligase

Unwinding of DNA with helicases

Semiconservative replication of DNA requires that the two strands of a parental DNA molecule be separated during the synthesis of new complementary strands. This separation needs an unwinding mechanism. The unwinding process is catalysed by enzymes called **DNA helicases**. DNA helicases unwind DNA molecules using energy derived ATP.

Binding of SSB proteins

Once the DNA strands are unwound by DNA helicase, they must be kept in an extended single stranded form for replication. The binding of **SSB proteins** to single stranded DNA is cooperative; that is, the binding of the first SSB monomer stimulates the binding of additional monomers at contiguous sites on the DNA chain. Without the SSB protein binding, the complementary strands could renature or form intrastarnd hairpin loops by hydrogen bonding between short segments of complementary nucleotide sequences.

Transient breaking of DNA by topoisomerases

During the unwinding of the DNA molecule (*E.coli* DNA spins around 3000 revolutions per minute to allow the unwinding) there is every possibility to form DNA supercoils. The formation of supercoils is prevented by the required axes of rotation (swivel) provided by enzymes called **DNA topoisomerases**. The topoisomerases catalyse the transient breaks in DNA molecules but use covalent linkages to themselves to hold on to the cleaved molecules.

RNA priming and formation of primosomes

Once the replication fork has formed, the synthesis of new DNA strands is initiated by RNA primers synthesised by **DNA primase**. A single RNA primer is sufficient for the continuous replication of the leading strand, but the discontinuous replication of the lagging strand requires an RNA primer to start the synthesis of each Okazaki fragment. The initiation of Okazaki fragments on the lagging strand is carried out by the **primosome**, a protein complex containing DNA primase and DNA helicase.

Formation of replisomes

The complete replication apparatus that moves along the replication fork is called **replisome**. The replisome contains the DNA polymerase III holoenzyme, one catalytic core replicates the leading strand, the second catalytic core replicates the lagging strand, and the primosome unwinds the parental DNA molecule and synthesises the RNA primers needed for the discontinuous synthesis of the lagging strand.

Covalent closure of nicks by DNA ligase

During the replication of the lagging strand, RNA primers are used. The DNA polymerase I replaces these RNA primers, and the single strand breaks left by polymerase are sealed by DNA ligase.

Mechanics of DNA Replication

The process of DNA replication begins at specific sites in the chromosomes termed **origins of replication**, requires a primer bearing a free 3'-OH, proceeds specifically in the $5' \Rightarrow 3'$ **direction** on both strands of DNA concurrently and results in the copying of the template strands in a **semiconservative** manner. The semiconservative nature of DNA replication means that the newly synthesized daughter strands remain associated with their respective parental template strands.

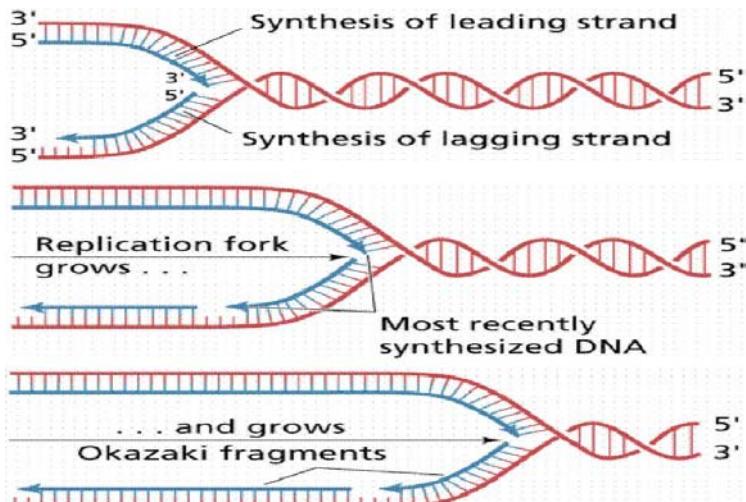


Figure 1. Diagrammatic representation of DNA replication

The large size of eukaryotic chromosomes and the limits of nucleotide incorporation during DNA synthesis, make it necessary for multiple origins of replication to exist in order to complete replication in a reasonable period of time. The precise nature of origins of replication in higher eukaryotic organisms is unclear. However, it is clear that at a replication origin the strands of DNA must dissociate and unwind in order to allow access to DNA polymerase. Unwinding of the duplex at the origin as well as along the strands as the replication process proceeds is carried out by **helicases**. The resultant regions of single-stranded DNA are stabilized by the binding of **single-strand binding proteins**. The stabilized single-stranded regions are then accessible to the enzymatic activities required for replication to proceed. The site of the unwound template strands is termed the **replication fork**.

In order for DNA polymerases to synthesize DNA they must encounter a free 3'-OH which is the substrate for attachment of the 5'-phosphate of the incoming nucleotide. During repair of damaged DNA the 3'-OH can arise from the hydrolysis of the backbone of one of the two strands. During replication the 3'-OH is supplied through the use of an **RNA primer**, synthesized by the **primase activity**. The primase utilizes the DNA strands as templates and synthesizes a short stretch of RNA generating a primer for DNA polymerase.

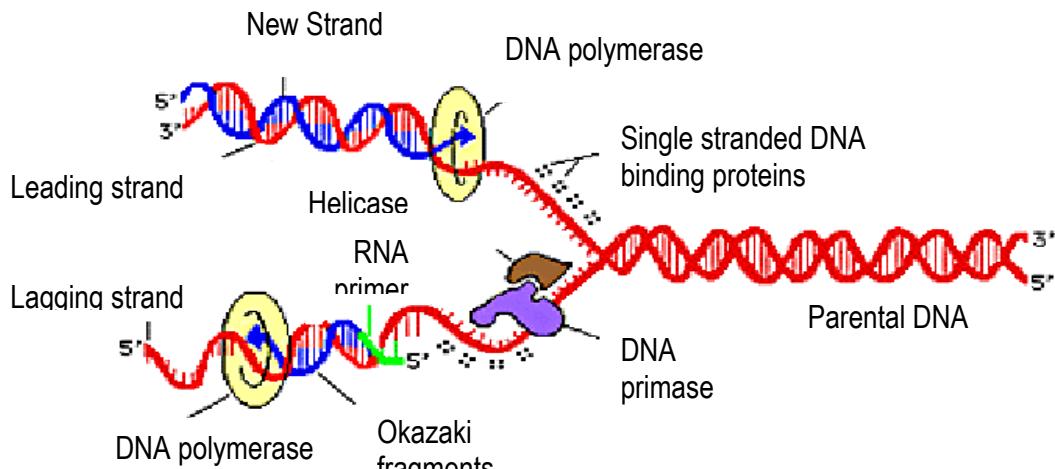


Figure 2. Diagram showing major components of replication apparatus

Synthesis of DNA proceeds in the $5' \Rightarrow 3'$ direction through the attachment of the 5'-phosphate of an incoming dNTP to the existing 3'-OH in the elongating DNA strands with the concomitant release of pyrophosphate. Initiation of synthesis, at origins of replication, occurs simultaneously on both strands of DNA. Synthesis then proceeds bidirectionally, with one strand in each direction being copied continuously and one strand in each direction being copied discontinuously. During the process of DNA polymerases incorporating dNTPs into DNA in the $5' \Rightarrow 3'$ direction they are moving in the $3' \Rightarrow 5'$ direction with respect to the template strand. In order for DNA synthesis to occur simultaneously on both template strands as well as bidirectionally one strand appears to be synthesized in the $3' \Rightarrow 5'$ direction. In actuality one strand of newly synthesized DNA is produced discontinuously.

The strand of DNA synthesized continuously is termed the **leading strand** and the discontinuous strand is termed the **lagging strand**. The lagging strand of DNA is composed of short stretches of RNA primer plus newly synthesized DNA approximately 100-200 bases long (the approximate distance between adjacent nucleosomes). The lagging strands of DNA are also called **Okazaki fragments**. The concept of continuous strand synthesis is somewhat of a misnomer since DNA polymerases do not remain associated with a template strand indefinitely. The ability of a particular polymerase to remain associated with the template strand is termed its' **processivity**. The longer it associates the higher the processivity of the enzyme. DNA polymerase processivity is

enhanced by additional protein activities of the replisome identified as **processivity accessory proteins**.

How is it that DNA polymerase can copy both strands of DNA in the $5' \Rightarrow 3'$ direction simultaneously? A model has been proposed where DNA polymerases exist as dimers associated with the other necessary proteins at the replication fork and identified as the **replisome**. The template for the lagging strand is temporarily looped through the replisome such that the DNA polymerases are moving along both strands in the $3' \Rightarrow 5'$ direction simultaneously for short distances, the distance of an Okazaki fragment. As the replication forks progress along the template strands the newly synthesized daughter strands and parental template strands reform a DNA double helix. This means that only a small stretch of the template duplex is single-stranded at any given time.

The progression of the replication fork requires that the DNA ahead of the fork be continuously unwound. Due to the fact that eukaryotic chromosomal DNA is attached to a protein scaffold the progressive movement of the replication fork introduces severe torsional stress into the duplex ahead of the fork. This torsional stress is relieved by **DNA topoisomerases**. Topoisomerases relieve torsional stresses in duplexes of DNA by introducing either double- (topoisomerases II) or single-stranded (topoisomerases I) breaks into the backbone of the DNA. These breaks allow unwinding of the duplex and removal of the replication-induced torsional strain. The nicks are then resealed by the topoisomerases.

The RNA primers of the leading strands and Okazaki fragments are removed by the repair DNA polymerases simultaneously replacing the ribonucleotides with deoxyribonucleotides. The gaps that exist between the 3'-OH of one leading strand and the 5'-phosphate of another as well as between one Okazaki fragment and another are repaired by **DNA ligases** thereby, completing the process of replication.

CHAPTER 8

TRANSCRIPTION: THE PROCESS OF RNA SYNTHESIS

A DNA molecule is a long chain of building blocks, small molecules called nucleotides. A DNA molecule is small to be seen, but its exact shape has been ingeniously worked out by indirect means. It consists of a pair of nucleotide chains twisted together in an elegant spiral; the double helix; the immortal coil. The nucleotide building blocks come in four different kinds, whose names may be shortened to A, T, C and G. These are the same in all animals and plants. This DNA can be regarded as a set of instructions how to make the body, written in the A, T, C and G alphabet of nucleotides.

DNA molecules do two important things. Firstly they replicate, that is to say they make copies of themselves. This has gone non-stop ever since the beginning of life and the DNA molecules are very good at it indeed. At every replication the DNA messages were faithfully copied, with scarcely any mistakes. The second important thing the DNA does is supervising how the message in the DNA is used to produce different kinds of molecules- the protein. These are the vital molecules of every organism, the so called survival machine.

If this survival machine to function properly, the coded message of the DNA, written in the four lettered nucleotide alphabet, is to be written in a simple mechanical way to another language of alphabets, messenger RNA (mRNA). Of course, both the forms are the languages of nucleotides. To make the molecules needed for the survival machine, the nucleotide language has to be translated into the language of amino acids. This is the first step in the direction of the survival machine construction. These protein molecules not only constitute physical fabric of the body; they also exert sensitive control over all the chemical processes inside the units of survival machines, the so called organisms. The first step, the copying of the DNA information into RNA, is designated **transcription**. The second step, in which amino acids are polymerized in response to the RNA information, is called **translation**. There are subtle differences in these two processes between prokaryotes and eukaryotes, the two forms of survival machines on the earth.

What is transcription?

Transcription is a process by which the coded message available in the DNA is rewritten in the form of mRNA. During this process, one strand of the DNA double helix is used as a template to synthesize a mRNA mediated by the enzyme RNA polymerase. The mRNA, thus produced is called as primary transcript. The primary transcripts of prokaryotes are different from the ones of eukaryotes and this difference is due to the differences in genome organization of prokaryotes and eukaryotes.

During the process of transcription a single strand of RNA is synthesized using a double stranded DNA molecule as a template. The two strands of the DNA molecule are separated from

one another, exposing the nitrogenous bases. Only one strand is actively used as a template in the transcription process, this is known as the **template strand**, otherwise called as **antisense strand or noncoding strand**. The **sense/coding or nontemplate strand** is complementary to the antisense strand, and is not transcribed. The RNA sequence that is made is a complementary copy of the nitrogenous bases in the sense strand. If an guanine (G) base is part of the sequence on the sense DNA strand, then the RNA molecule has a cytosine (C) base added to its sequence at that point. In the RNA molecule uracil (U) substitutes for thymine (T).

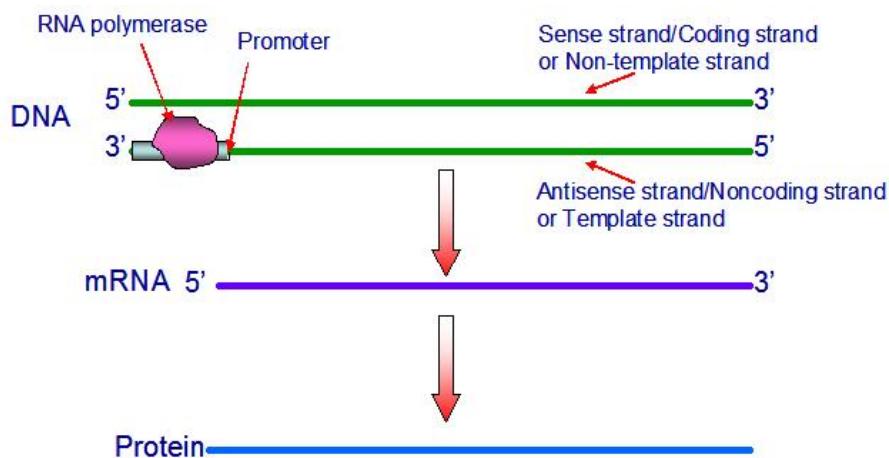


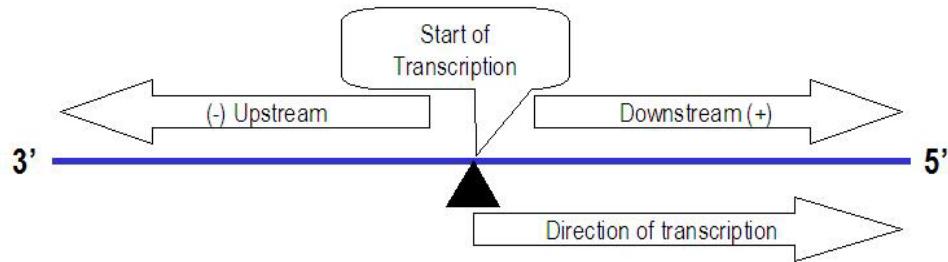
Figure 1. Processes of transcription and translation

The process of transcription is catalyzed by a multisubunit enzyme called RNA polymerase, which needs as substrates double stranded DNA, and the ribonucleotides ATP, UTP, CTP and GTP. One at a time, this enzyme adds ribonucleotides to a growing RNA strand by joining incoming ribonucleotide triphosphates to the ribose sugar molecule of the last nucleotide of the growing RNA strand. Two of the phosphate groups are removed from the triphosphate and a covalent bond is formed between the remaining phosphate and the third carbon atom of the ribose sugar at the end of the RNA strand.

Initiation of the transcription process begins with binding of the RNA polymerase enzyme to the DNA molecule at a region known as the promoter site. The RNA synthesis takes place within locally unwound segment of DNA, sometimes called as **restriction bubble**. This site is right in front of a gene where transcription will begin. In bacteria, this region contains two short sequences of bases that appear to be the same in all, or most, promoters. The RNA polymerase enzyme does not copy the promoter into the RNA, but begins the synthesis of the RNA at a specific nucleotide sequence called the start signal or initiation site which is often the bases GTA on the DNA (which then become the bases CAU on the RNA molecule).

The basic features of transcription are the same in both prokaryotes and eukaryotes, but many of the details such as promoter sequences – are different. A segment of DNA that is transcribed to produce one RNA molecule is called **transcription unit**. Transcription units may be

equivalent to individual genes or they may include several contiguous genes. The process of transcription can be divided into three stages: 1) **initiation** of a new RNA chain, 2) **elongation** of the chain and 3) **termination** of transcription and release of nascent RNA molecule. The point of initiation of transcription is usually called as **transcription site**.



Direction along a DNA molecule, in relation to transcription, is designated using the terms **upstream** and **downstream**. **Upstream** refers to a position on a DNA strand in a direction opposite to the direction of transcription. **Downstream** refers to a position located in the same direction as transcription. Numbers with +/- signs are used to designate upstream and downstream positions. For example, -10 means 10 nucleotides upstream from some reference point, and +5 means 5 nucleotides downstream from the same reference point.

Prokaryotic transcription

In prokaryotes, genes tend to be clustered in coordinately-regulated groups called operons. The genes are transcribed together on a single transcript and each protein within the cluster is translated separately. The prokaryotic transcription is catalyzed by a single RNA polymerase on the template DNA strand in the presence of substrate ribonucleotide triphosphates (NTPs): ATP, CTP, GTP, UTP.

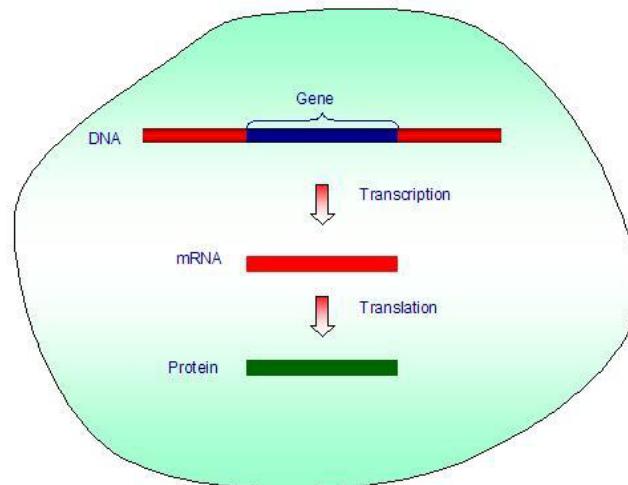


Figure 2. Processes of prokaryotic transcription and translation

RNA polymerase

RNA polymerase that catalyzes prokaryotic transcription is complex, multimeric protein. The *E. coli* RNA polymerase has a molecular weight of about 480,000 and consists of five polypeptides. Two of these polypeptides are identical; thus the enzyme contains four distinct polypeptides. The complete RNA polymerase molecule, the **holoenzyme**, has the composition of α_2 , β , β' and σ . The α subunits are involved in the assembly of the **tetrameric core** ($\alpha_2\beta\beta'$) of RNA polymerase. The β subunit contains the ribonucleotide triphosphate binding site and the β' subunit harbours the DNA template binding site. One subunit of **σ factor**, is involved only in the initiation of transcription, it plays no role in chain elongation. After RNA chain initiation has occurred, the σ factor is released, and chain elongation is catalyzed by the core enzyme ($\alpha_2\beta\beta'$). The function of σ factor is to recognize and bind RNA polymerase to the transcription initiation or **promoter sites** in DNA.

Initiation of RNA chains

Initiation of RNA chains involves three steps: 1) binding of the RNA polymerase holoenzyme to a promoter binding in DNA, 2) the localized unwinding of the two strands of DNA by RNA polymerase, providing a template strand free to base pair with incoming ribonucleotides in the nascent RNA chain. The holoenzyme remains bound at the promoter region during the synthesis of the first eight or nine bonds, then the σ factor is released, and the core enzyme begins the elongation phase of RNA synthesis downstream from the promoter.

Elongation of RNA chains

Elongation of RNA chains is catalyzed by the RNA polymerase core enzyme, after the release of the σ factor. The covalent extension of RNA chains takes place within the transcription bubble, a locally unwound segment of DNA. The RNA polymerase molecule contains both DNA unwinding and DNA rewinding activities. RNA polymerase continuously unwinds the DNA double helix ahead of the polymerization site and rewinds the complementary DNA strands behind the polymerization site as it moves along the double helix.

Termination of RNA chains

Termination of RNA chains occurs when RNA polymerase encounters a termination signal. When this occurs, the transcription complex disassociates, releasing the nascent RNA molecule. There are two types of transcription terminators in *E. coli*. One type results in the termination only in the presence of a protein called ***rho* (ρ) factor**; therefore, such termination sequences are called ***rho-dependent terminators***. The other type results in the termination of transcription without the involvement of *rho*-factor; such sequences are called ***rho-independent terminators***.

In prokaryotes, the transcription, translation and degradation mRNA occur simultaneously. The polypeptide synthesizing machinery is not separated by a nuclear envelope from the site of

mRNA synthesis. Therefore, once the 5' end of mRNA has been synthesized, it can immediately be used as a template for polypeptide synthesis. Indeed, transcription and translation often are tightly coupled in prokaryotes.

Eukaryotic transcription

Although the overall process of RNA synthesis is similar in prokaryotes and eukaryotes, the process is considerably more complex in eukaryotes. In eukaryotes, RNA is synthesized in the nucleus and RNAs that encode proteins must be transported to the cytoplasm for translation on ribosomes. The complexity of eukaryotic transcription is evident from the involvement of three different RNA polymerases in the process of RNA synthesis.

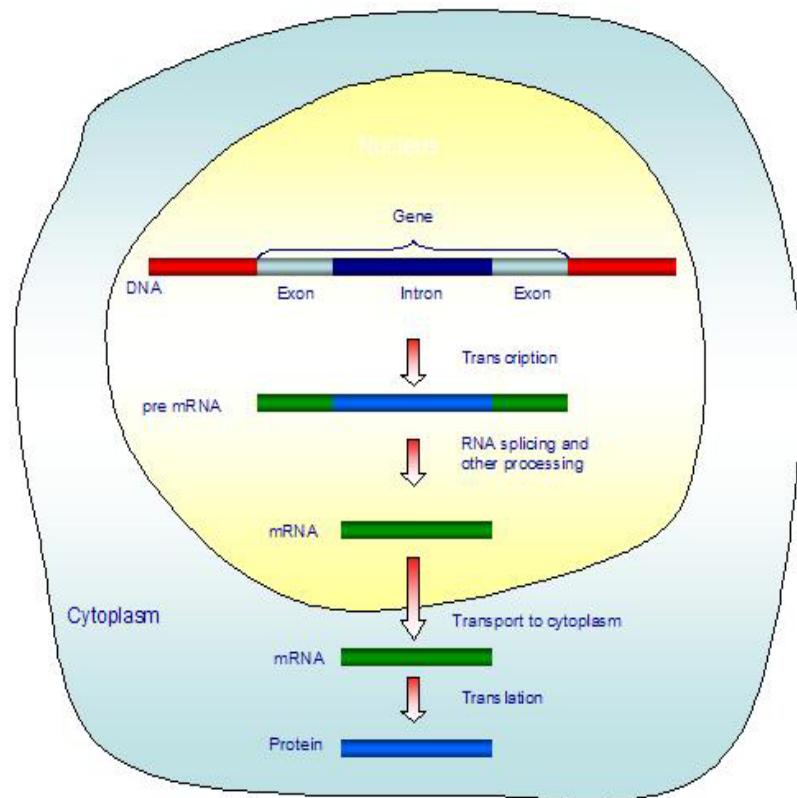


Figure 3. Processes of eukaryotic transcription and translation

RNA polymerases

Eukaryotic transcription process is catalyzed by three different types of RNA polymerases and designated as **RNA polymerases I, II and III**. These enzymes are more complex, with 10 or more subunits, than the *E. coli* RNA polymerase. Moreover, unlike the *E. coli* RNA polymerase, all the three enzymes require the assistance of other proteins called transcription factors. In order to initiate the synthesis of RNA chains. The key features of three RNA polymerases are given below.

Enzyme	Location	Function	Products
RNA polymerase I	Nucleolus	Catalyzes the synthesis of ribosomal RNAs except the small 5S RNA	Ribosomal RNA excluding 5S rRNA
RNA polymerase II	Nucleus	Transcribes nuclear genes that encode proteins and perhaps other genes specifying hnRNA	Nuclear pre mRNA
RNA polymerase III	Nucleus	Catalyzes the synthesis of transfer RNA molecules, the 5S rRNA molecules and small nuclear RNA	tRNA, 5S rRNA, and other small nuclear RNA

Initiation of RNA chains

Unlike prokaryotic RNA polymerase, eukaryotic RNA polymerases cannot initiate transcription by themselves. All three eukaryotic RNA polymerases require the assistance of additional factors, otherwise called as transcription factors to start the initiation of mRNA synthesis. These transcription factors must bind to a promoter region in DNA and form appropriate initiation complex before RNA synthesis. Different promoters and transcription factors are used by the RNA polymerases I, II and III. In all cases, the initiation of transcription involves the formation of a locally unwound segment of DNA, providing a strand that is free to function as a template for the synthesis of complementary strand of mRNA. The initiation of transcription by RNA polymerase II requires the assistance of several basal transcription factors. The basal transcription factors must interact with promoters in the correct sequence to initiate transcription effectively. Each transcription factor is denoted TFIIX (Transcription Factors for polymerase II), where X is a letter identifying the individual factor). Some of the transcription factors and their role in initiation are given below.

TFIID	It contains TATA Binding protein and several small TBP associated proteins
TFIIF	Has DNA unwinding activity Releases the RNA polymerase from the promoter region and so that it can move downstream and mediate RNA chain elongation.
TFIIE	Phosphorylates RNA polymerase

All the above mentioned transcription factors and others such as TFIIA, TFIIB, TFIIE and TFIIE combine together and form an initiation complex.

Elongation of RNA chains

Once the RNA polymerases are released from their initiation complexes, they catalyze RNA chain elongation by the same way as the RNA polymerases of prokaryotes

Termination of RNA chains

The 3' ends of RNA transcripts synthesized by RNA polymerase II are produced by endonucleolytic cleavage of the primary transcripts. The cleavage event that produces termination of RNA chain occurs at a site 11 to 30 bases downstream from a conserved sequence **AAUAAA**.

Other events occurring during eukaryotic RNA synthesis

Capping

Early in the elongation process, the 5' ends of eukaryotic pre mRNAs are modified by the addition of 7-methyl guanosine (7 MG) caps. These 7MG caps are added when the growing RNA chains are only about 30 nucleotides long. The 7 MG caps are recognized by protein factors involved in the initiation of translation and also help protect the growing RNA chains from degradation by nucleases.

Polyadenylation

After the termination, a stretch of adenosine monophosphate residues are added to a length of 200 nucleotides to the 3' ends of transcripts. This poly (A) addition is catalyzed by the enzyme poly(A) polymerase. The poly(A) tails of eukaryotic mRNAs enhance the stability and play an important role in their transportation from the nucleus to the cytoplasm.

RNA editing

Normally genetic information is not altered in the mRNA intermediary. Sometime, it happens that exceptions do occur. These exceptions in transcribing the genetic information are rectified by two processes: 1) by changing the structures of individual bases and 2) by inserting or deleting uridine monophosphate residues. This normalization is called as **RNA editing**.

RNA splicing

A eukaryotic gene consists of coding regions called **exons** interspersed among non-coding regions called **introns**. Although absent from prokaryotic genes, introns have been found in some tRNA genes, a few rRNA genes, in genes from chloroplast and mitochondrial DNA, and in some viruses that attack bacteria (bacteriophages). Introns must be removed prior to translation, as they do not specify useful information needed to synthesize a protein. In essence, introns are "noise," and their function (if any) in genes is not fully understood. The removal of introns takes place within the nucleus, prior to the translocation of the mRNA to the ribosomes.

Intron removal is carried out by a molecular complex made of RNA and proteins, or **small nuclear ribonucleoproteins**, known as a **spliceosome**. Each small nuclear ribonucleoprotein consists of protein and small nuclear RNA (snRNA). The process of removal of introns in RNA can occur by three means: 1) tRNA precursor splicing, 2) autocatalytic splicing and 3) pre-mRNA splicing.

Thus, the genetic information available in DNA is transcribed to mRNA in different forms in different survival machines *viz.* prokaryotes and eukaryotes.

CHAPTER 09

TRANSLATION: THE PROCESS OF PROTEIN SYNTHESIS

Translation refers to the actual process of protein synthesis. Translation converts the nucleotide sequence of the mRNA into the sequence of amino acids comprising a protein, according to the specifications of a coded language. This is a complex process, requiring the functions of a large number of macromolecules. These include 1) over 50 polypeptides and 3 to 5 RNA molecules present in the ribosome, 2) at least 20 amino acid-activating enzymes, 3) from 40 to 60 different tRNA molecules and 4) numerous soluble proteins involved in polypeptide chain initiation, elongation and termination.

Different types of RNA

RNA has the same primary structure as DNA . It consists of a sugar-phosphate backbone, with nucleotides attaches to the 1' carbon of the sugar. The differences between DNA and RNA are that:

- RNA has a hydroxyl group on the 2' carbon of the sugar (thus, the difference between deoxyribonucleic acid and ribonucleic acid).
- Instead of using the nucleotide thymine, RNA uses another nucleotide called uracil
- Because of the extra hydroxyl group on the sugar, RNA is too bulky to form a stable double helix. RNA exists as a single-stranded molecule. However, regions of double helix can form where there is some base pair complementation (U and A, G and C), resulting in hairpin loops. The RNA molecule with its hairpin loops is said to have a secondary structure.
- In addition, because the RNA molecule is not restricted to a rigid double helix, it can form many different tertiary structures. Each RNA molecule, depending on the sequence of its bases, can fold into a stable three-dimensional structure.

There are several different kinds of RNA made by the cell.

mRNA - messenger RNA is a copy of a gene. It acts as a photocopy of a gene by having a sequence complementary to one strand of the DNA and identical to the other strand. The mRNA acts as a carrier to carry the information stored in the DNA in the nucleus to the cytoplasm where the ribosomes can make it into protein.

rRNA - ribosomal RNA is one of the structural components of the ribosome. It has sequence complementarity to regions of the mRNA so that the ribosome knows where to bind to an mRNA it needs to make protein from.

snRNA - small nuclear RNA is involved in the machinery that processes RNAs as they travel between the nucleus and the cytoplasm.

tRNA - transfer RNA is a small RNA that has a very specific secondary and tertiary structure such that it can bind an amino acid at one end, and mRNA at the other end. It acts as an adaptor to carry the amino acid elements of a protein to the appropriate place as coded for by the mRNA.

The Genetic Code

How does an mRNA specify amino acid sequence? The answer lies in the genetic code. It would be impossible for each amino acid to be specified by one nucleotide, because there are only 4 nucleotides and 20 amino acids. Similarly, two nucleotide combinations could only specify 16 amino acids. The final conclusion is that each amino acid is specified by a particular combination of three nucleotides, called a **codon**, thus there are 64 codons in total. Deciphering these triplet codes will answer the questions such as 1) which codons specify each of the amino acids?, 2) how many of the 64 possible triplet codons are used? and 3) how are they used?

Properties of the Genetic Code

The main features of the genetic code were worked out during the 1960s. They include

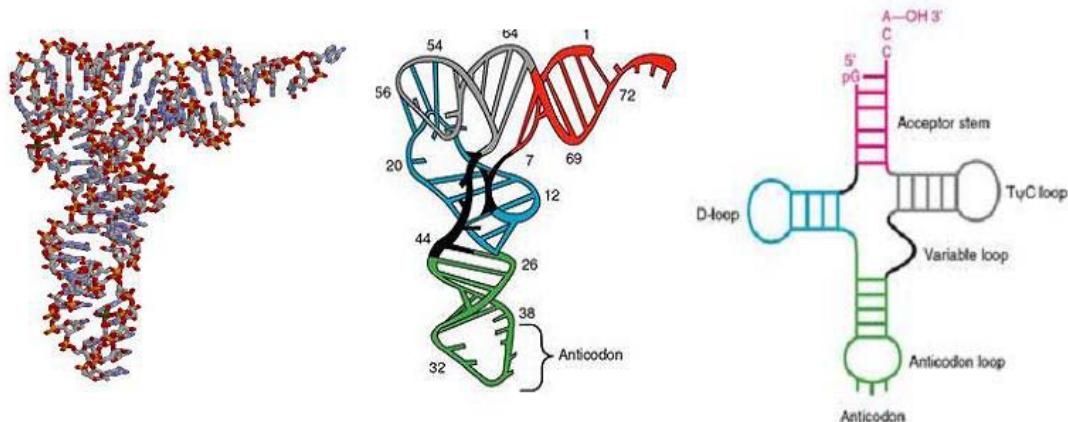
- **The genetic code is composed of nucleotide triplets:** Three nucleotides in mRNA specify one amino acid in the polypeptide product; thus each codon contains three nucleotides.
- **The genetic code is non-overlapping:** Each nucleotide in mRNA belongs to just one codon except in rare cases where genes overlap.
- **The genetic code is punctuation free:** There are no punctuations within the coding regions of mRNA molecules. During the translation codons are read consecutively.
- **The genetic code is degenerate:** All but two of the amino acids are specified by more than one codon.
- **The genetic code is ordered:** Multiple codons for a given amino acid and codons for amino acids with similar chemical properties are closely related, usually differing by a single nucleotide.
- **The genetic code contains start and stop codons:** Specific codons are used to initiate and to terminate polypeptide chains.
- **The genetic code is universal:** With minor exceptions, the codons have the same meaning in all living organisms, from viruses to humans.

The following table gives the details of what the 64 codons do. Note the degeneracy of the genetic code. Each amino acid might have up to six codons that specify it.

	U	C	A	G	
U	UUU Phenylalanine (Phe)	UCU Serine (Ser)	UAU Tyrosine (Tyr)	UGU Cysteine (Cys)	U
C	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
A	UUA Leucine (Leu)	UCA Ser	UAA STOP (Ochre)	UGA STOP (Opal)	A
	UUG Leu	UCG Ser	UAG STOP (Amber)	UGG Tryptophan (Trp)	G
C	CUU Leucine (Leu)	CCU Proline (Pro)	CAU Histidine (His)	CGU Arginine (Arg)	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Glutamine (Gln)	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Isoleucine (Ile)	ACU Threonine (Thr)	AAU Asparagine (Asn)	AGU Serine (Ser)	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lysine (Lys)	AGA Arginine (Arg)	A
	AUG Methionine (Met) or START	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Valine Val	GCU Alanine (Ala)	GAU Aspartic acid (Asp)	GGU Glycine (Gly)	U
	GUC (Val)	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glutamic acid (Glu)	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

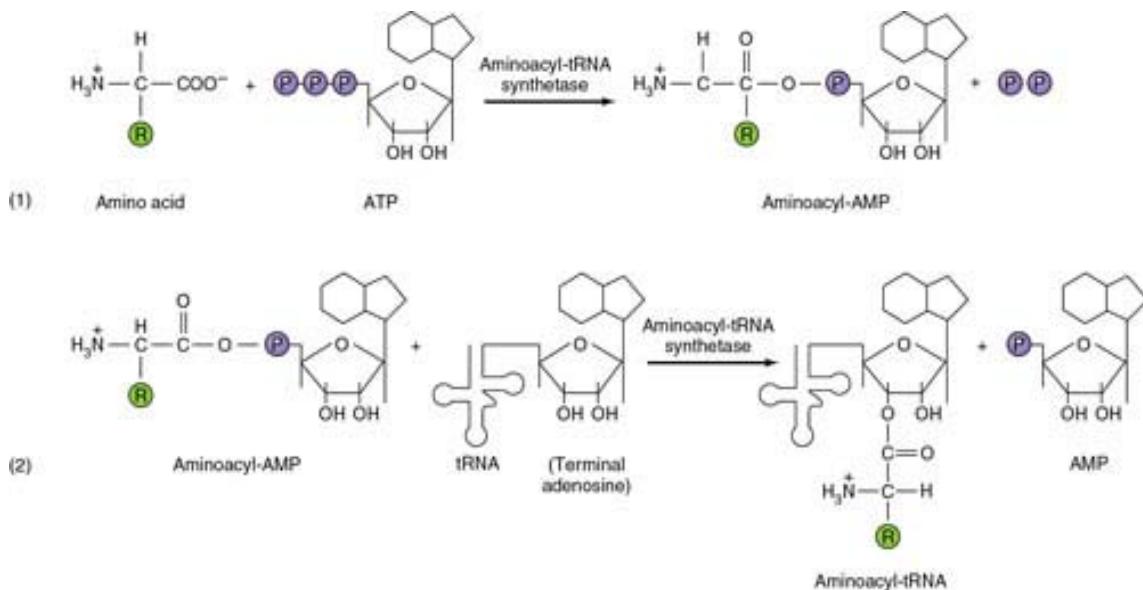
tRNA- the translator

The translation of a sequence of nucleotides in mRNA into the correct sequence of amino acids in the polypeptide product requires the accurate recognition of codons by tRNA molecules. How do tRNAs recognize to which codon to bring an amino acid? The tRNA has an **anticodon** on its mRNA-binding end that is complementary to the codon on the mRNA. Each tRNA only binds the appropriate amino acid for its anticodon.



The tRNA molecules are clover leaf shaped. There are four major arms, named for their structure and function. The acceptor arm consists of a base-paired stem that ends in an unpaired sequence whose free 2' – or 3' –OH group is aminoacylated. The other arms consist of base-paired **stems** and unpaired **loops**. The **T Ψ C arm** is named for the presence of this triplet sequence, the **anticodon arm** always contains the anticodon triplet in the centre of the loop, and the **D arm** is named for its content of the base dihydrouridine (Ψ stands for pseudouridine and D stands for dihydrouridine). The overall length of tRNA lengths is from 74 to 95 bases. The tRNA are considered as adapter molecules since they carry amino acids. There are from one to four tRNAs for each of the 20 amino acids.

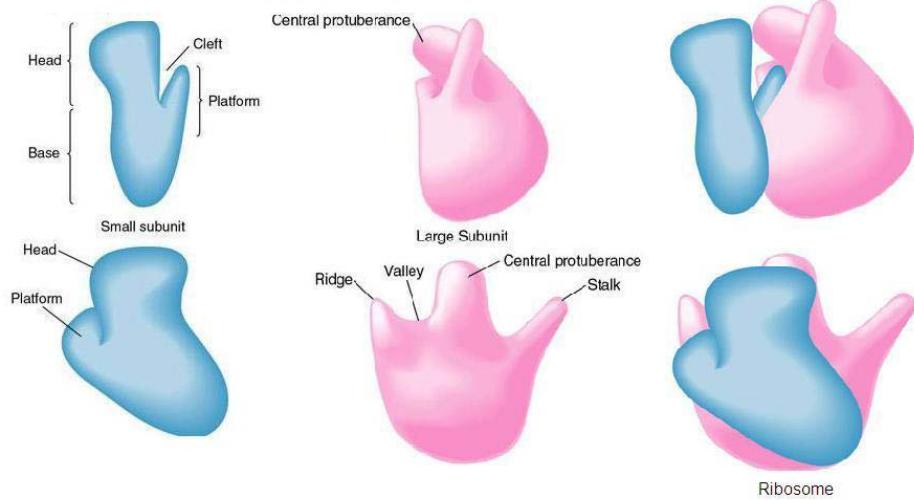
The amino acids are attached to the tRNAs by high-energy bonds between carboxyl group of the amino acids and 3'-hydroxyl termni of the tRNAs. The tRNAs are activated or charged with amino acids in a two-step process, with both the reactions catalyzed by the same enzyme, **aminoacyl tRNA synthetase**. First, the amino acid reacts with ATP to form aminoacyl~adenylate, releasing pyrophosphate. Energy for the reaction is provided by cleaving the high energy bond of the ATP. Then the activated amino acid is transferred to the tRNA, releasing AMP. There is at least one aminoacyl tRNA synthetase for each of the 20 amino acids. The tRNAs properly recognized by a synthetase are described as its **cognate tRNAs**.



Ribosomes- the site of protein synthesis

Ribosomes are multimolecular aggregates containing RNAs and proteins. They can be separated into two subunits, a large and a small subunit. Ribosome sizes are most frequently expressed in terms of their rates of sedimentation during centrifugation, in Svedburg units. The prokaryotic ribosome has size of 70S. The ribosomes of eukaryotes are larger (usually 80S).

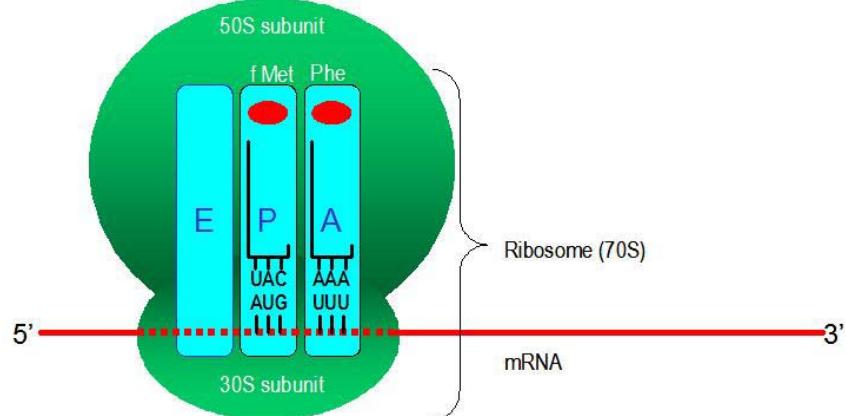
The complete 70S ribosome has the asymmetrical construction. The partition between the head and body of the small subunit is aligned with the notch of the large subunit, so that platform of the small subunit fits into the large subunit. The RNAs constitute the major part of the mass of the ribosome.



The macromolecular composition of prokaryotic and eukaryotic ribosomes are given below.

Prokaryotic ribosome		Eukaryotic ribosome	
31 ribosomal proteins + 5SrRNA+ 23SrRNA	21 ribosomal proteins +16SrRNA	49 ribosomal proteins + 5SrRNA + 28SrRNA	33 ribosomal proteins + 18SrRNA
50S subunit	30S subunit	60S subunit	40S subunit
70S ribosome		80S ribosome	

The ribosomes contain binding sites for mRNA, for tRNAs and for a variety of protein factors.



Ribosomes are also the site of a catalytic activity that is probably responsible for the formation of the peptide bond during protein synthesis and have several active centers. Each ribosome has two domains: **translational domain** and **exit domain**. The translational domain has **aminoacyl site** (A sites) and **peptidyl site** (P Site) and the exit domain has **exit site** (E site).

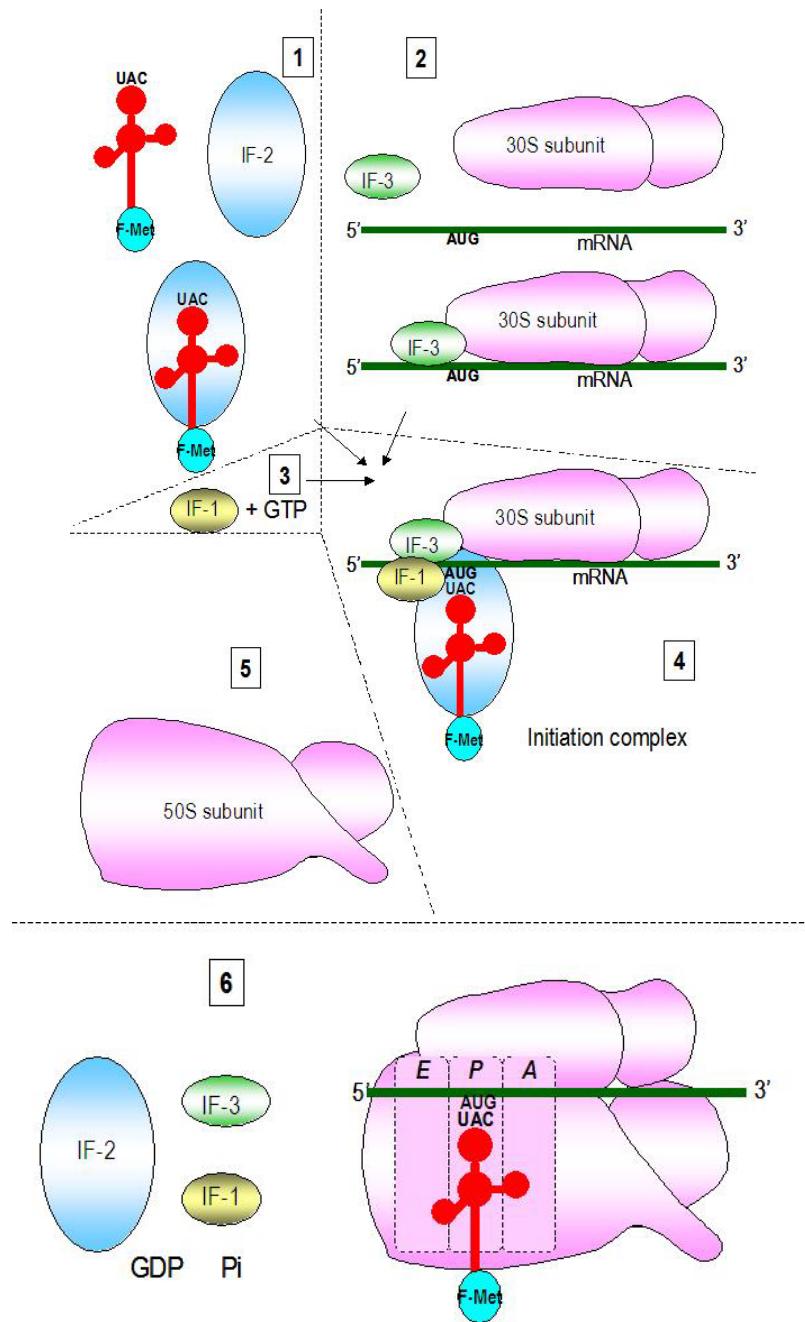
Translation

Translation is the process of converting the sequence of nucleotide residues in mRNA into the sequence of amino acid residues in a polypeptide. The messenger RNA (mRNA), the product of transcription and splicing, carries the code for the order of insertion of amino acids. This genetic code is a triplet code without punctuation. The mRNA also contains signals for initiation and termination of a polypeptide chain. The initiation signals are near its 5' end, while the termination signals are near its 3' end. Thus, the translation process has three steps *viz.* 1) polypeptide chain **initiation**, 2) chain **elongation** and 3) chain **termination**. A variety of protein factors, usually loosely bound to ribosomes, are also required. Many are needed for proper initiation of polypeptide chain synthesis, while others are required for binding incoming aminoacyl tRNAs to the ribosome and moving the ribosome relative to the mRNA (translocation) after each addition step.

Initiation

The initiation of translation includes all events that precede the formation of a peptide between the first two amino acids of the new polypeptide chain. The initiation process involves the 30S subunit of the ribosome, a special initiator tRNA, an mRNA molecule, three soluble proteins **initiation factors: IF-1, IF-2 and IF-3** and one molecule of GTP. In the first stage of the initiation of translation, a free 30S subunit interacts with an mRNA molecule and initiation factors. The 50S subunit joins the complex to form the 70S ribosome in the final step of the initiation process. The synthesis of polypeptides is initiated by a special tRNA, designated tRNA_f^{Met}. This means that all polypeptides begin with methionine during synthesis. The tRNA_f^{Met} interacts with protein initiation factor, IF-2 to begin the initiation process. Polypeptide chain initiation begins with formation of two complexes: 1) one contains IF-2 and methionyl tRNA_f^{Met} and 2) the other contains mRNA molecule, a 30S ribosomal subunit and IF-3. The 30S subunit/mRNA complex will form only in the presence of IF-3; this IF-3 controls the ability of the 30S subunit to begin the initiation process.

The IF-2 and methionyl tRNA_f^{Met} complex and the mRNA/30S subunit/IF-3 complex then combine with each other and with IF-1 and one molecule of GTP to form complete 30S initiation complex. The final step in the initiation of translation is the addition of 50S subunit to the 30S initiation complex to produce the 70S ribosome. The addition of 50S subunit results in the release of all the initiation factors. The addition of 50S ribosomal subunit positions the initiator tRNA, methionyl tRNA_f^{Met}, in the peptidyl site (P site) with the anticodon of the tRNA aligned with the AUG initiation codon of the mRNA. Methionyl tRNA_f^{Met} is the only aminoacyl-tRNA that can enter the P-site directly, without first passing through the aminoacyl site (A site). With the initiator AUG positioned in the P site, the second codon in the mRNA is in the A site, dictating the amino-acyl tRNA binding specificity at that site and setting the stage for the second phase in polypeptide synthesis, chain elongation.



1-3. Formation of IF-2/tRNA_{Met} and IF-3/mRNA/30S subunit complex.

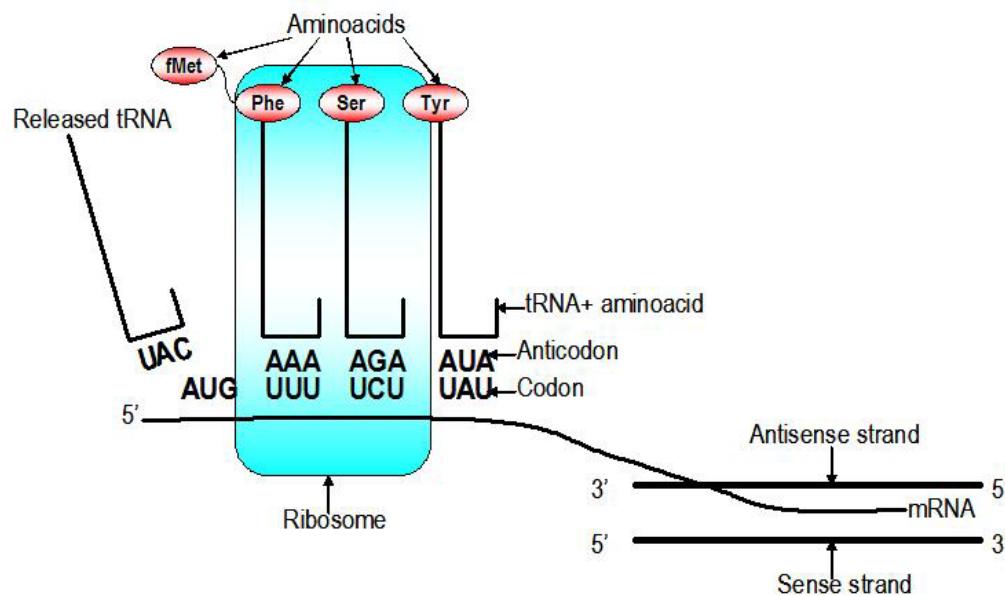
4. The complexes formed in step 1 combine with each other, IF-1 and GTP to form the 30S initiation complex.

5-6. After IF-3 is released, the 50S subunit joins the initiation complex; GTP is cleaved and IF-1 and If-2 are released.

Elongation

The process of polypeptide chain elongation involves addition of amino acid to the growing polypeptide and has three steps: 1) binding of an aminoacyl-tRNA to the A site of the ribosome, 2) transfer of the growing polypeptide chain from the tRNA in the P site to the tRNA in the A site by the formation of a new polypeptide bond and 3) translocation of the ribosome along the mRNA to position the next codon in the A site. During step 3, the charged tRNA is translocated from A site to P site and the uncharged tRNA is moved from P site to E site. These three steps are repeated in a cyclic manner throughout the elongation process.

The binding of an aminoacyl-tRNA to the A site of the ribosome requires an elongation factor Tu and a molecule of GTP (EF-TuGTP). The second step of polypeptide bond formation between the amino acids in P site and A site is catalyzed by **peptidyl transferase**, an enzymatic activity built into the 50S subunit of the ribosome. During the third step in chain elongation, the peptidyl-tRNA in the A site of the ribosome is translocated to the P site, and the uncharged tRNA in the P site is translocated to the E site, as the ribosome moves three nucleotides toward the 3' end of the mRNA molecule. The translocation step requires GTP and elongation factor G (EF-G). The translocation of peptidyl-tRNA from the A site to the P site leaves the A site unoccupied and the ribosome ready to begin the next cycle of chain elongation.



Termination

Polypeptide chain elongation undergoes termination when any of three chain-termination codons (UAA, UAG or UGA) enters the A site on the ribosome. These three stop codons are recognized by soluble proteins called **release factors (RFs)**. The presence of a release factor in the A site alters the activity of peptidyl transferase such that it adds a water molecule to the

carboxyl terminus of the nascent polypeptide. This releases the polypeptide from the tRNA molecule in the P site and triggers the translocation of the free tRNA to the E site. Termination is completed by the release of the mRNA molecule from the ribosome and the dissociation of the ribosome into its subunits. The ribosomal subunits are then ready to initiate another round of protein synthesis, as previously discussed.

Comparison between prokaryotic and eukaryotic translation

Prokaryotes	Eukaryotes
The initiation of translation is in prokaryotes, involves three initiation factors.	The initiation of translation is more complex in eukaryotes, involving several soluble initiation factors.
The aminogroup of the methionine on the initiator tRNA is formylated.	The aminogroup of the methionine on the initiator tRNA is not formylated.
The initiation complex forms at the Shine-Dalgarno sequence (AGGAGG).	The initiation complex forms at the 5'terminus of the mRNA.
Initiation complex scans the mRNA, starting at Shine-Dalgarno sequence, searching for AUG translation-initiation codon.	Initiation complex scans the mRNA, starting at 5'end, searching for AUG translation-initiation codon.
5' terminus of the mRNA is free of 7-methyl guanosine cap	A cap binding protein (CBP) binds to the 7-methyl guanosine cap at the 5' terminus of the mRNA.
In prokaryotes, two release factors, RF-1 and RF-2 present. RF-1 recognizes termination codons UAA and UAG; RF-2 recognizes UAA and UGA.	In eukaryotes, a single release factor (eRF) recognizes all three termination codons.

CHAPTER 10

DNA REPAIR MECHANISMS

Transformed phenotypes in living organisms arise due to mutations in DNA. These mutations can be due to several factors and DNA damage is one among them. DNA damage can occur as the result of exposure to environmental stimuli such as alkylating chemicals or ultraviolet or radioactive irradiation and free radicals generated spontaneously in the oxidizing environment of the cell. These phenomena can, and do, lead to the introduction of mutations in the coding capacity of the DNA. Mutations in DNA can also, but rarely, arise from the spontaneous tautomerization of the bases. Modification of the DNA bases by alkylation (predominately the incorporation of -CH₃ groups) predominately occurs on purine residues. Methylation of G residues allows them to base pair with T instead of C. A unique activity called **O⁶-alkylguanine transferase** removes the alkyl group from G residues.

Mutations in DNA are of two types. **Transition mutations** result from the exchange of one purine, or pyrimidine, for another purine, or pyrimidine. **Transversion mutations** result from the exchange of a purine for a pyrimidine or *visa versa*. These mutations are supposed to be kept at a tolerable level. The multiplicity of repair mechanisms that have evolved in organisms ranging from cateria to humans emphatically documents the importance of keeping mutaion, both somatic and germ-line at a tolerable level. For example, *E. coli* cells possess at least five distinct mechanisms for the repair of defects in DNA: 1) light-dependent repair or photoreactivation, 2) excision repair, 3) mismatch repair, 4) postreplication repair and 5) error-prone repair. Mammals seem to possess all of the repair mechanisms found in *E. coli* except photoreactivation.

Light-dependent repair

Light dependent repair or photoreactivation of DNA in bacteria is carried out by a light activated enzyme called **DNA photolyase**. When DNA is exposed to UV light, thymine dimmers are produced by covalent cross-linkages between adjacent thymine residues. DNA photolyase binds to thymine dimmers in DNA and uses light energy to cleave the covalent cross-links (Figure 1). Photolyase will bind to the thymine dimmers in DNA in the dark, but it cannot catalyze cleavage of the bonds joining the thymine moieties without energy derived from visible light, specifically light within blue region of the spectrum.

Excision repair

Excision repair of damaged DNA involves at least three steps. In step1, a DNA repair endonuclease or endonuclease-containing enzyme complex recognizes, binds to, and excises the damaged base or bases in DNA. In step 2, a DNA polymerase fills in the gap by using the undamaged complementary strand of DNA as template. In step3, the enzyme DNA ligase seals the break left by DNA polymerase to complete the repair process. There are two major types of excision repair: **base excision repair** systems remove abnormal or chemically modified bases

from DNA, whereas **nucleotide excision repair** pathways remove larger defects like thymine dimmers.

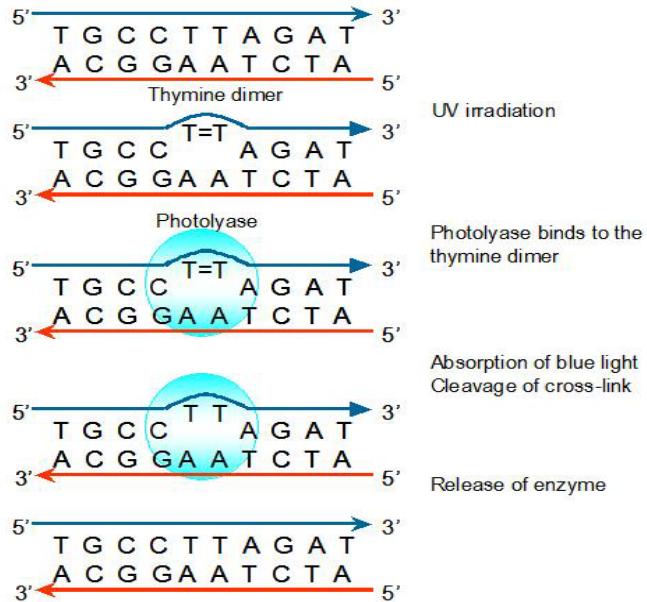


Figure 1. Cleavage of thymine-dimer cross-links by light activated photolyase.

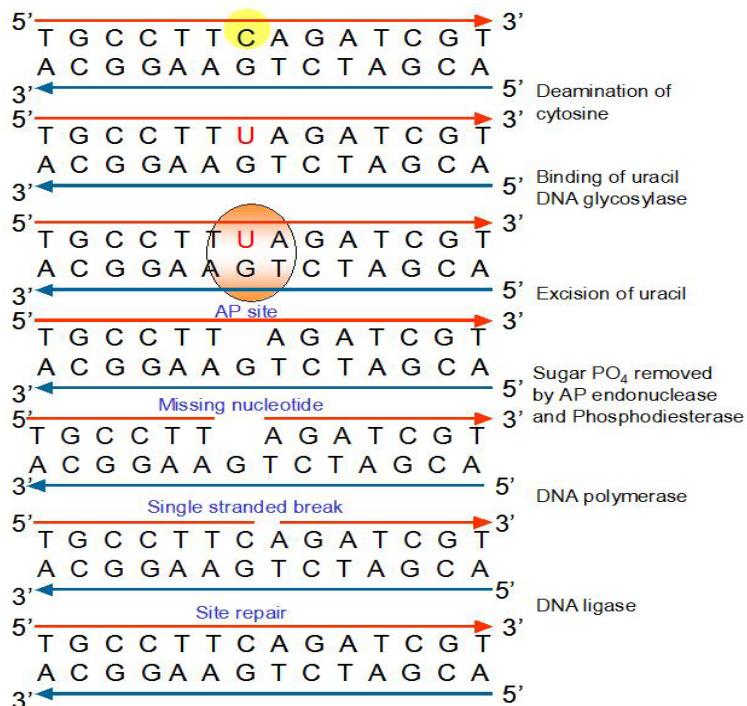


Figure 2. Repair of DNA by the base excision pathway

Base excision repair

Base excision repair can be initiated by any group of enzymes called DNA glycosylases that recognize abnormal bases in DNA. Each glycosylase recognizes a specific type of altered base, such as deaminated bases, oxidized bases and so on. The glycosylases cleave the glycosidic bond between the abnormal base and 2-deoxyribose, creating **apurinic or apyrimidinic sites (AP sites)** with missing bases. These AP sites are recognized by AP endonucleases, which act together with phosphodiesterases to excise the sugar-phosphate groups at sites where no base is present. DNA polymerase then replaces the missing nucleotide according to the specifications of the complementary strand, and DNA ligase seals the nick.

Nucleotide excision repair

Nucleotide excision repair removes larger lesions like thymine dimers and bases with bulky side groups from DNA. In nucleotide excision repair, a unique excision nuclease activity produces cuts on either side of the damaged nucleotide(s) and excises an oligonucleotide containing the damaged base(s). This nuclease called an **exinuclease** to distinguish it from the endonucleases and exonucleases.

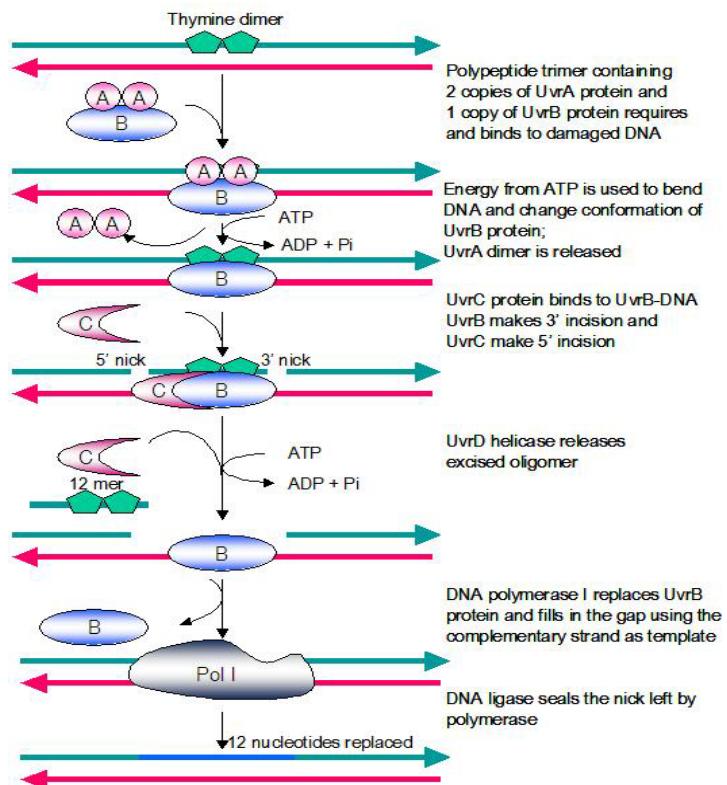


Figure 3. Repair of DNA by the nucleotide excision pathway in *E. coli*

The nucleotide excision repair in *E. coli* needs the products of three genes, *uvrA*, *uvrB* and *uvrC*. A trimeric protein containing two UvrA polypeptides and one UvrB polypeptide recognizes the defect in DNA, binds to it and uses energy from ATP to bend the DNA at the damaged site. The UvrA dimer is then released and the UvrC protein binds to the UvrB-DNA complex. The UvrB protein cleaves the phosphodiester bond from the damaged nucleotide(s) on the 3' side, and the UvrC protein hydrolyzes the eight phosphodiester linkage from the damage on the 5' side. The *uvrD* gene product, DNA helicase II, releases the excised dodecamer. In the last two steps of the pathway, DNA polymerase I fills in the gap, and DNA ligase seals the remaining nick in the DNA molecule.

Mismatch repair

The mismatch repair is carried out by the 3'→5' exonuclease activity built into DNA polymerase. DNA polymerase proofreads DNA strands during their synthesis, removing any mismatched nucleotides at the 3' termini of growing strands. Mismatches often involve the normal four bases in DNA. For example, a T may be mispaired with a G. Because both T and G are normal components of DNA, mismatch repair systems need some way to determine whether the T or G is the correct base at a given site. The repair system makes this distinction by identifying the template strand, which contains the original nucleotide sequence, and the newly synthesized strand, which contains the misincorporated base (the error). This distinction can be made based on the pattern of methylation in newly replicated DNA.

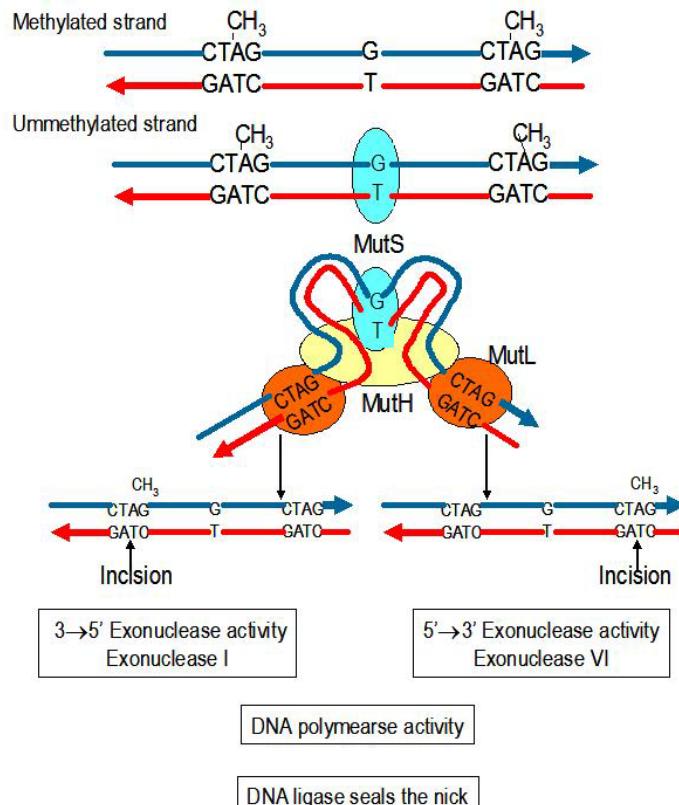


Figure 4. Mismatch repair in *E. coli*

In *E. coli*, the A in GATC sequences is methylated subsequent to its synthesis. Thus an interval occurs during which the template strand is methylated and the newly synthesized strand is unmethylated. This mismatch repair system uses this difference in methylation state to excise the mismatched nucleotide in the nascent strand and replace it with the correct nucleotide by using the methylated parental strand of DNA as template.

Mismatch repair of DNA in *E. coli* requires the products of four genes, *mutH*, *mutL*, *mutS* and *mutU*. The MutS protein recognizes mismatches and binds to them to initiate the repair process. MutH and MutL proteins then join the complex. MutH contains a GATC-specific endonuclease activity that cleaves the unmethylated strand at hemimethylated GATC sites either 5' or 3' to the mismatch. The incision sites may be 1000 nucleotide pairs or more from the mismatch. The subsequent excision process requires MutS, MutL, DNA helicase II (MutU) and an appropriate exonuclease. If the incision occurs at GATC sequence 5' to the mismatch, a 5'→3' exonuclease like *E. coli* exonuclease VI is required. If the incision occurs 3' to the mismatch, a 3→5' nuclease activity like that of *E. coli* exonuclease I required. After the excision process has removed the mismatched nucleotide from the unmethylated strand, DNA polymerase fills the gap and DNA ligase seals the nick.

Postreplication repair

During the DNA replication, the presence of a thymine dimer in a template strand block the operation of DNA polymerase III, thus preventing the addition of nucleotides in the new strand. DNA polymerase restarts DNA synthesis at some position past the dimer, leaving a gap in the nascent strand opposite the dimer in the template strand.

At this point, the original nucleotide sequence has been lost from both strands of this progeny double helix. The damaged DNA molecule is repaired by a recombination-dependent repair process mediated by the *E. coli* *recA* gene product. The RecA protein, which is required for homologous recombination, stimulates the exchange of single strands between homologous double helices.

During postreplication repair, the RecA protein binds to the single strand of DNA at the gap and mediates pairing with the homologous segment of the sister double helix. The gap opposite the dimer is filled with the homologous DNA strand from the sister DNA molecule. The resulting gap in the sister double helix is filled in by DNA polymerase, and the nick is sealed by the DNA ligase. The thymine dimer remains in the template strand of the original progeny DNA molecule, but the complementary strand is now intact. If the thymine dimer is not removed by the nucleotide excision repair system, this postreplication repair must be repeated after each round of DNA replication.

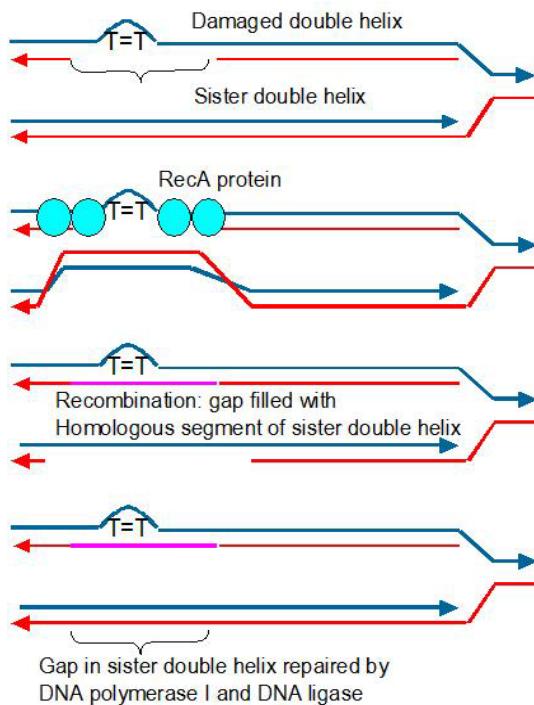


Figure 5. Postreplication repair of DNA in *E. coli*.

Error prone repair system

The DNA repair systems described so far are quite accurate. However, when the DNA of *E. coli* cells is heavily damaged by mutagenic agents such as UV light, the cells take some drastic steps in their attempt to survive. They go through the so called **SOS response**, during which a whole battery of DNA repair, recombination and replication proteins are synthesized. Two of these proteins, encoded by the *umuC* and *umuD* (UV mutable genes), encode proteins that allow DNA replication to proceed across damaged segments of template strands, even though the nucleotide sequences in the damaged region cannot be accurately replicated. This **error-prone repair** system eliminates gaps in the damaged nucleotides in the template strands but, in so doing, sharply increases the frequency of replication errors. The SOS response appears to be a somewhat desperate and risky attempt to escape lethal effects of heavily damaged DNA. When error-prone repair system is operative, mutation rates increase sharply.

CHAPTER 11

GENE STRUCTURE AND REGULATION

The number of genes responsible for different phenotypes varies from organism to organism. The number may be from 4000 genes in the typical bacterial genome to 100,000 genes in the human genome. Of the 4000 genes in the bacterial genome or the estimated 100,000 genes in the human genome, only a fraction is expressed at any given time. The differential expression of the genes in an organism results in different gene products. The degree and type of regulation naturally reflect the function of the gene product of the gene. Some gene products are required all the time and their genes are expressed at a more or less constant level in virtually all the cells of a species or organism. These genes are generally called as **housekeeping genes**.

The expression genes can be classified into various categories: constitutive, inducible and repressible. Constant, seemingly unregulated expression of a gene is called **constitutive gene expression**. Sometimes, the genes are regulated to have their expression by some of the signal molecules. The expression of a gene under the specific molecular signal with increased level of gene product is called as **inducible** and the process of increasing the expression of the gene is called **induction**. Conversely, gene products, that decrease in concentration in response to a molecular signal is referred to as **repressible** and the decrease in gene expression is called **repression**.

Although housekeeping genes are expressed constitutively, the proteins they encode are present in widely varying amounts. The expression of all these genes (constitutive, inducible or repressible) in the form of transcription expression starts due to the binding of RNA polymerase to specific sequences called **promoter sequences**. In other words, these promoter sequence regions control the gene expression and are called as **control region**.

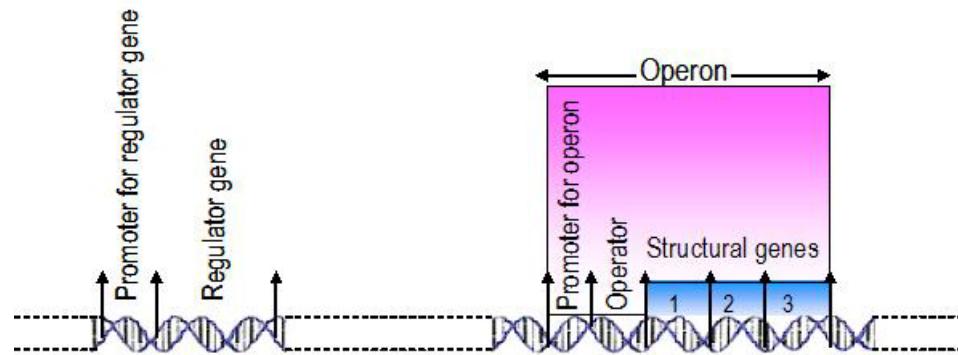
In the case of inducible or repressible gene expression, several types of regulatory proteins mediate the regulation. These proteins affect the interaction between RNA polymerase and the promoters.

The other classification of genes is based on the gene products for which they are responsible. The genes responsible for the synthesis of polypeptide sequences or structural proteins are called as **structural genes** and the genes regulating the expression of structural genes by way producing signal molecules (either inducers or repressors) are called as **regulatory genes**.

Gene regulation in prokaryotes

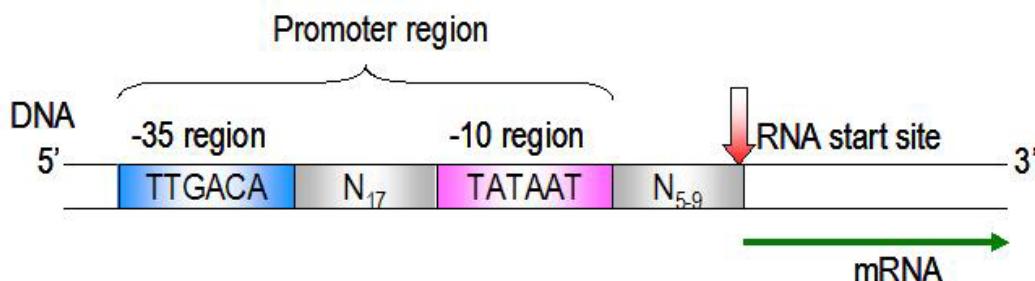
Gene structure

The structural and regulatory model for a gene was first proposed by **F. Jacob and J. Monod** in the year 1961 and is popularly known as **Operon model**. They proposed that the transcription of one or a set of contiguous genes (genes specifying amino acid sequences of enzymes or structural proteins) is controlled by two regulatory elements, the regulator or the repressor gene, encodes a protein called the repressor.



In bacteria, genes are clustered into **operons**: gene clusters that encode the proteins necessary to perform coordinated function, such as biosynthesis of a given amino acid. RNA that is transcribed from prokaryotic operons is **polycistronic** a term implying that multiple proteins are encoded in a single transcript.

In bacteria, control of the rate of transcriptional initiation is the predominant site for control of gene expression. As with the majority of prokaryotic genes, initiation is controlled by two DNA sequence elements that are approximately 35 bases and 10 bases, respectively, upstream of the site of transcriptional initiation and as such are identified as the -35 and -10 positions. These 2 sequence elements are termed **promoter sequences**, because they *promote* recognition of transcriptional start sites by **RNA polymerase**. The consensus sequence for the -35 position is TTGACA, and for the -10 position, TATAAT. (The -10 position is also known as the **Pribnow-box**.) These promoter sequences are recognized and contacted by **RNA polymerase**.

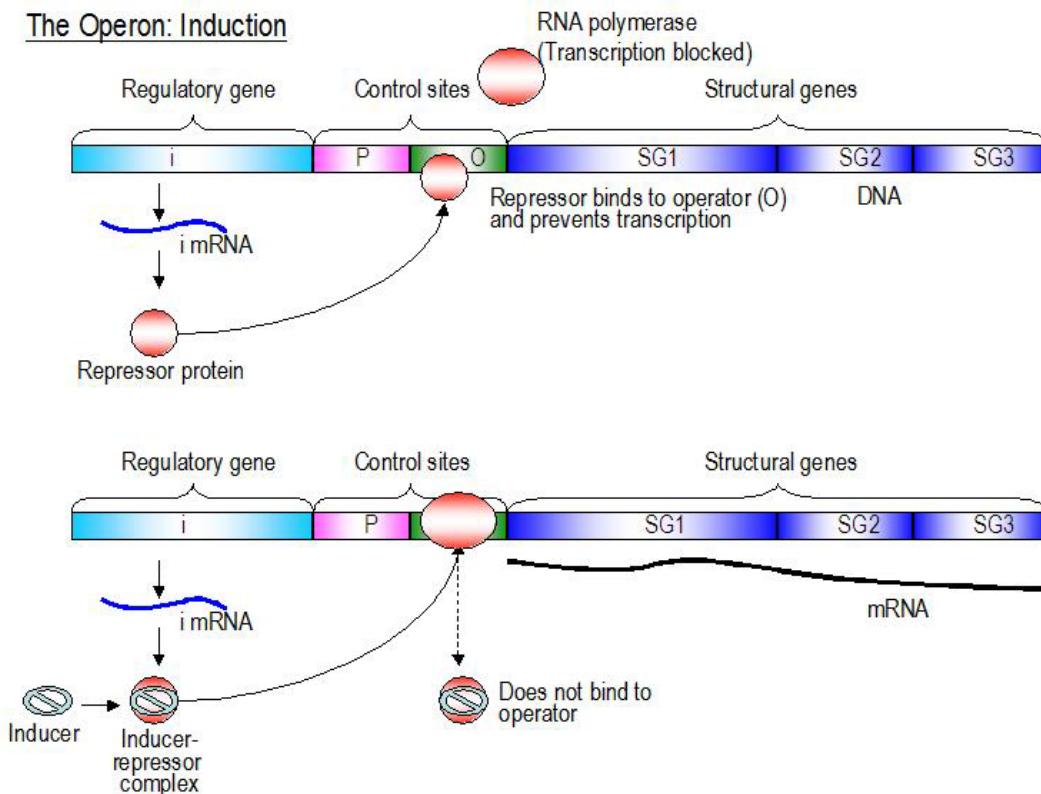


The activity of **RNA polymerase** at a given promoter is in turn regulated by interaction with accessory proteins, which affect its ability to recognize start sites. These regulatory proteins can act both positively (**inducers**) and negatively (**repressors**). The accessibility of promoter regions of prokaryotic DNA is in many cases regulated by the interaction of proteins with sequences termed **operators**. The operator region is adjacent to the promoter elements in most operons and in most cases the sequences of the operator bind a repressor protein. However, there are several operons in *E. coli* that contain overlapping sequence elements, one that binds a repressor and one that binds an activator.

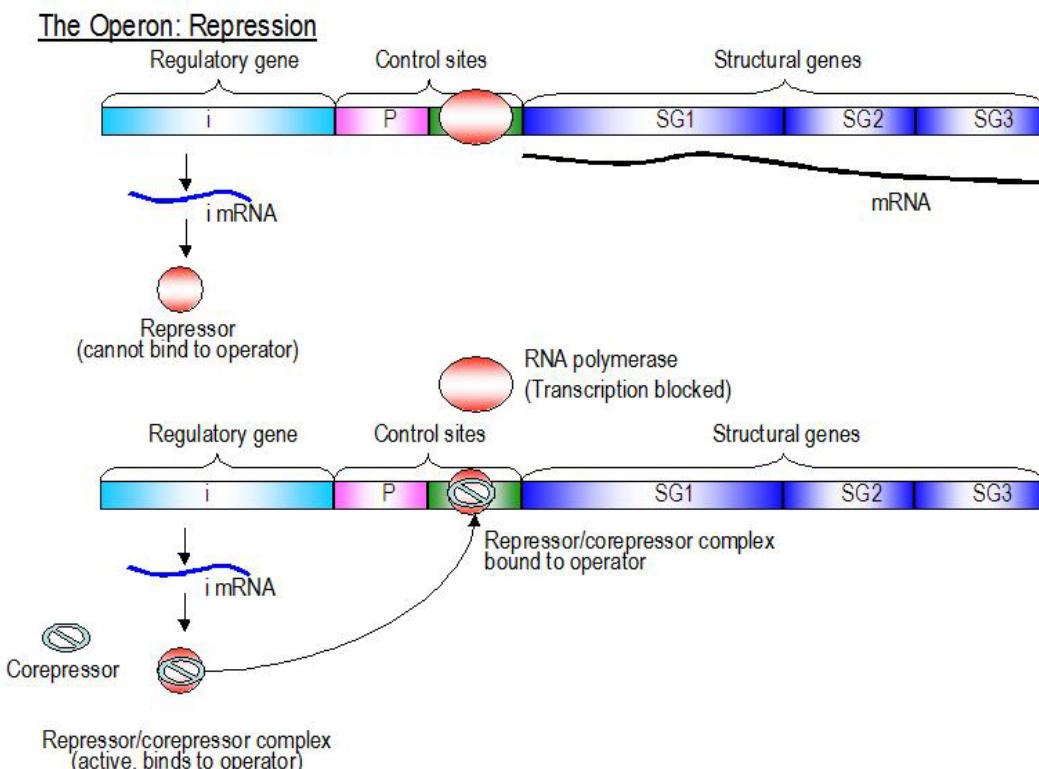
Gene regulation

As indicated above, prokaryotic genes that encode the proteins necessary to perform coordinated function are clustered into operons. Two major modes of transcriptional regulation function in bacteria (*E. coli*) to control the expression of operons. Based on this operons can be classified as inducible operons and repressible operons.

In the case of inducible operon, the free repressor binds to the operator, turning-off transcription. When the effector molecule (the inducer) is present, it is bound to the repressor, causing the repressor to be released from the operator; that is repressor-inducer complex cannot bind to the operator. Thus the addition of inducer turns on or induces the transcription of the structural genes in the operon.



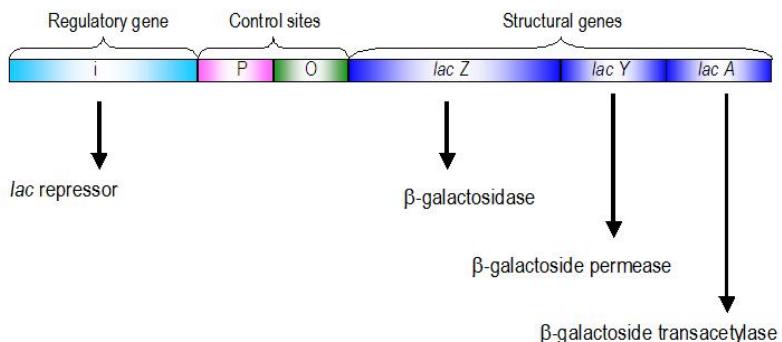
In the case of repressible operon, the situation is just reversed. The free repressor cannot bind to the operator. Only the repressor-effector molecule (co-repressor) complex is active in binding to the operator. Thus the transcription of the structural genes in a repressible operon is turned on in the absence of and turned off in the presence of the effector molecule (the co-repressor).



The *lac* Operon

The *lac* operon consists of one regulatory gene (the *i* gene) and three structural genes (*z*, *y*, and *a*). The *i* gene codes for the repressor of the *lac* operon. The *z* gene codes for **β -galactosidase**, which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units, galactose and glucose. The *y* gene codes for **β -galactoside permease**, which increases permeability of the cell to β -galactosides. The *a* gene encodes a **β -galactoside transacetylase**. The *lac* operon is regulated by two modes: induction and repression.

The *lac* Operon



Induction of *lac* operon

The *lac* operon is an inducible protein; the *lac Z*, *lac Y* and *lac A* are expressed only in the presence of lactose. The *lac* regulator gene designated as *i* gene, encodes a repressor that is 360 amino acids long. In the absence of inducer, the repressor binds to the *lac* operator sequence, which, in turn, prevents RNA polymerase from catalyzing the transcription of three structural genes. A few molecules of β -galactosidase, β -galactoside permease and β -galactoside transacetylase are synthesized in the uninduced state, providing a low background enzyme activity. This background enzyme activity is essential for induction of the *lac* operon, because the inducer of the operon, allolactose, is derived from lactose in a reaction catalyzed by β -galactosidase. One formed, allolactose is bound by the repressor, causing the release of the repressor from the operator. In this way, allolactose induces the transcription of *lac Z*, *lac Y* and *lac A* structural genes.

Repression of *lac* operon

The *lac* operon is repressed, even in the presence of lactose, if glucose is present in the medium. This repression is maintained until the glucose supply is exhausted. The repression of the *lac* operon under these conditions is termed **catabolite repression** (or the **glucose effect**) and is a result of the low levels of cAMP that result from an adequate glucose supply.

The catabolite repression of the *lac* operon is mediated by a regulatory protein called **Catabolite Activator Protein (CAP)** and small effector molecule called **cyclic AMP or cAMP** (adenosine-3'5'-phosphate). These two molecules form a complex called **cyclic AMP Receptor Protein**. Only the CAP-cAMP complex binds to the *lac* promoter; in the absence of cAMP, CAP does not bind. Thus cAMP acts as the effector molecule, determining the effect of CAP on *lac* operon transcription. High concentrations of glucose cause sharp decreases in the concentration of cAMP. In the presence of low concentration of cAMP, CAP cannot bind to the *lac* operon promoter. In turn, RNA polymerase cannot bind efficiently to the *lac* promoter in the absence of bound CAP-cAMP complex.

Attenuation

The *trp* operon encodes the genes for the synthesis of tryptophan. This cluster of genes, like the *lac* operon, is regulated by a repressor that binds to the operator sequences. The activity of the *trp* repressor for binding the operator region is enhanced when it binds tryptophan; in this capacity, tryptophan is known as a **corepressor**. Since the activity of the *trp* repressor is enhanced in the presence of tryptophan, the rate of expression of the *trp* operon is graded in response to the level of tryptophan in the cell.

Expression of the *trp* operon is also regulated by **attenuation**. The attenuator region, which is composed of sequences found within the transcribed RNA, is involved in controlling transcription from the operon after RNA polymerase has initiated synthesis. The attenuator of sequences of the RNA are found near the 5' end of the RNA termed the leader region of the RNA. The leader sequences are located prior to the start of the coding region for the first gene of the operon (the *trpE* gene). The attenuator region contains codons for a small leader polypeptide, that contains tandem tryptophan codons which produce truncated try transcript.

Gene regulation in eukaryotes

In eukaryotic cells, the ability to express biologically active proteins comes under regulation at several points:

Chromatin structure: The physical structure of the DNA, as it exists compacted into chromatin, can affect the ability of transcriptional regulatory proteins (termed **transcription factors**) and RNA polymerases to find access to specific genes and to activate transcription from them. The presence of the histones and CpG methylation most affect accessibility of the chromatin to RNA polymerases and transcription factors.

Transcriptional initiation: This is the most important mode for control of eukaryotic gene expression. Specific factors that exert control include the strength of **promoter elements** within the DNA sequences of a given gene, the presence or absence of **enhancer sequences** (which enhance the activity of RNA polymerase at a given promoter by binding specific transcription factors), and the interaction between multiple activator proteins and inhibitor proteins.

Transcript processing and modification: Eukaryotic mRNAs must be capped and polyadenylated, and the introns must be accurately removed. Several genes have been identified that undergo tissue-specific patterns of alternative splicing, which generate biologically different proteins from the same gene.

RNA transport: A fully processed mRNA must leave the nucleus in order to be translated into protein.

Transcript stability: Unlike prokaryotic mRNAs, whose half-lives are all in the range of 1–5 minutes, eukaryotic mRNAs can vary greatly in their stability. Certain unstable transcripts have sequences (predominately, but not exclusively, in the 3'-non-translated regions) that are signals for rapid degradation.

Translational initiation: Since many mRNAs have multiple methionine codons, the ability of ribosomes to recognize and initiate synthesis from the correct AUG codon can affect the expression of a gene product. Several examples have emerged demonstrating that some eukaryotic proteins initiate at non-AUG codons. This phenomenon has been known to occur in *E. coli* for quite some time, but only recently has it been observed in eukaryotic mRNAs.

Post-translational modification: Common modifications include glycosylation, acetylation, fatty acylation, disulfide bond formations, etc.

Protein transport: In order for proteins to be biologically active following translation and processing, they must be transported to their site of action.

Control of protein stability: Many proteins are rapidly degraded, whereas others are highly stable. Specific amino acid sequences in some proteins have been shown to bring about rapid degradation.

Control of eukaryotic transcription initiation

Transcription of the different classes of RNAs in eukaryotes is carried out by three different polymerases. RNA pol I synthesizes the rRNAs, except for the 5S species. RNA pol II synthesizes the mRNAs and some small nuclear RNAs (snRNAs) involved in RNA splicing. RNA pol III synthesizes the 5S rRNA and the tRNAs. The vast majority of eukaryotic RNAs are subjected to post-transcriptional processing.

The most complex controls observed in eukaryotic genes are those that regulate the expression of RNA pol II-transcribed genes, the mRNA genes. Almost all eukaryotic mRNA genes contain a basic structure consisting of coding exons and non-coding introns and basal promoters of two types and any number of different transcriptional regulatory domains. The basal promoter elements are termed **CCAAT-boxes** (pronounced **cat**) and **TATA-boxes** because of their sequence motifs. The TATA-box resides 20 to 30 bases upstream of the transcriptional start site and is similar in sequence to the prokaryotic Pribnow-box (consensus TATA^{T/A}AT^A, where ^{T/A} indicates that either base may be found at that position).

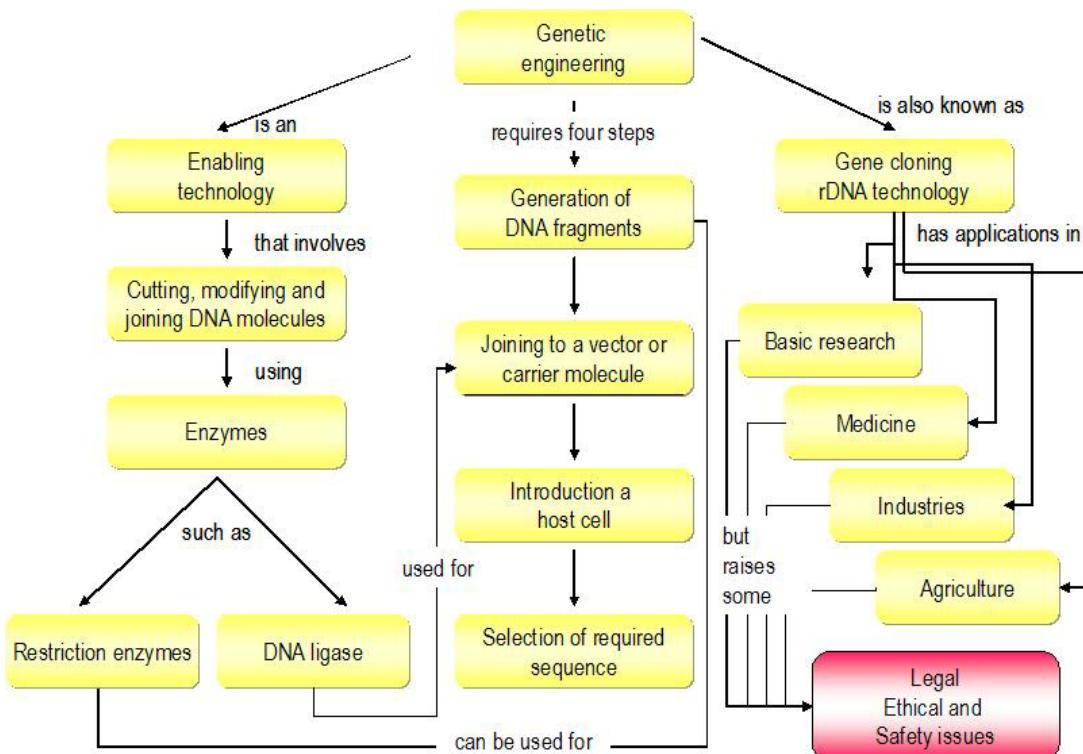
Numerous proteins identified as **TFIIB**, **B**, **C**, etc. (for transcription factors regulating RNA pol II), have been observed to interact with the TATA-box. The CCAAT-box (consensus GG^{T/C}CAATCT) resides 50 to 130 bases upstream of the transcriptional start site. The protein identified as **C/EBP** (for **CCAAT-box/Enhancer Binding Protein**) binds to the CCAAT-box element.

There are many other regulatory sequences in mRNA genes, as well, that bind various transcription factors. These regulatory sequences are predominantly located upstream (5') of the transcription initiation site, although some elements occur downstream (3') or even within the genes themselves. The number and type of regulatory elements to be found varies with each mRNA gene. Different combinations of transcription factors also can exert differential regulatory effects upon transcriptional initiation. The various cell types each express characteristic combinations of transcription factors; this is the major mechanism for cell-type specificity in the regulation of mRNA gene expression.

CHAPTER 12

RECOMBINANT DNA TECHNOLOGY: TOOLS

The development of **recombinant DNA technology (rDNA technology)** permitting the transfer of genetic material between widely divergent species has opened a new era of research into the structure and function of the genome. The rDNA technology is defined as "the formation of new combinations of heritable material by the insertion of nucleic acid molecules, produced by whatever means outside the cell, into any virus, bacterial plasmid or other vector system so as to allow their incorporation into a host organism in which they do not naturally occur but in which they are capable of continued propagation. The rDNA technology has provided the means to achieve: 1) the fractionation of individual DNA components of complex genomes, 2) the amplification of cloned genes, 3) the opportunity to study the expression of individual genes thus cloned and 4) the potential to create new genetic combinations. There are several other terms that can be used to describe the technology, including **gene manipulation**, **gene cloning**, **genetic modification** and **genetic engineering**. The term genetic engineering is often thought to be rather emotive or trivial, yet it is probably the label that most people would recognize.



Steps involved in rDNA technology

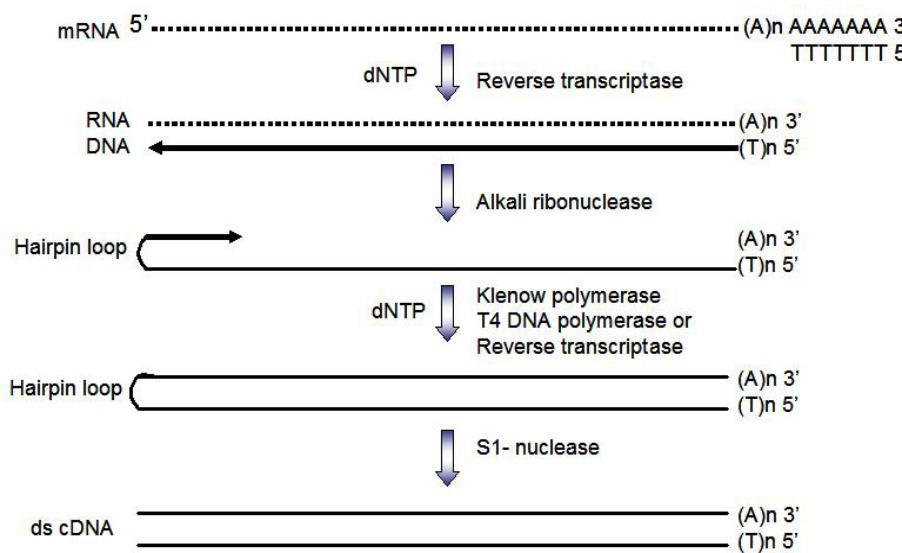
Any rDNA experiment has four essential steps: 1) generating DNA fragments, 2) cutting and joining the DNA fragments to vector DNA molecules, 3) introducing the vectors carrying the foreign DNA into host cells where they can replicate and 4) selecting the clone(s) of recipient cells that have acquired the recombinant DNA molecules.

Generating DNA fragments

One of the most important problems prior to rDNA experiment is to separate the DNA fragments from the total genomic DNA. This is normally accomplished either by fragmentation of DNA or synthesis of new DNA molecule. The fragmentation of DNA molecule can be achieved by mechanical shearing. The long thin threads which constitute duplex DNA molecules are sufficiently rigid to be very easily broken by shear forces in solution. In this method, high molecular weight DNA is sheared to population of molecules with a mean size of about 8kb pairs by stirring at 1500 rpm for 30 minutes. Breakage occurs essentially at random with respect to DNA sequence producing termini consisting of short single stranded regions which may be repaired later. The other sophisticated technique available to generate DNA fragments involves using restriction endonucleases about which discussion is made in the subsequent section. Other two possible sources for generating DNA fragments for cloning are complementary DNA (cDNA) synthesis using mRNA as a template and artificial synthesis of DNA molecule.

cDNA synthesis

Fundamental differences exist between the genomes of prokaryotes and eukaryotes. In prokaryotes, the coding sequences (exons) are not intervened by non-coding sequences (introns) whereas in eukaryotes the genes are generally split; the coding regions are interspersed with non coding DNA. This makes the expression of eukaryotic genes in prokaryotes a tough task.



Synthesis of cDNA from mRNA

To overcome this problem, cDNA synthesis or artificial DNA synthesis can be well exploited. In cDNA synthesis, the eukaryotic mRNA is used as a template to generate DNA. This can be achieved by making a complementary copy of the mRNA using the enzyme reverse transcriptase and whose function is to synthesize DNA upon an RNA template. At first, the enzyme was called **RNA dependent DNA polymerase**.

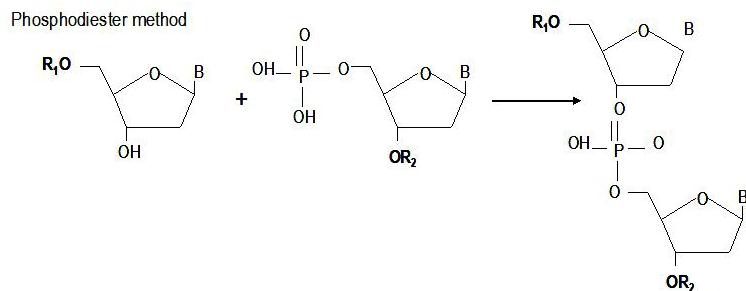
A DNA copy is made by hybridizing oligo-T primers, 10 to 20 nucleotides in length, to the 3' end of purified mRNA. Avian myeloblastosis virus (AMV) reverse transcriptase is used to synthesise a cDNA copy of the primed molecule. In the resultant RNA-DNA hybrid, the RNA can be destroyed by alkaline hydrolysis to which DNA is resistant. Thus, a single stranded cDNA is obtained which can be converted into a double stranded form in second DNA polymerase reaction. In the 3' end of the cDNA self complementarity occurs thus producing a hair pin or snap-back structure. This acts as a primer for duplex DNA synthesis by DNA polymerase. The hair-pin loop is trimmed away by treatment with single strand specific nuclease S1, giving rise to a fully duplex molecule. The power of this technique is that only a fraction of the genome (that fraction which is transcribed into mRNA) is copied. The resulting cDNA clones can be subsequently be used as probes to identify genomic fragments contained in a genomic library.

Chemical synthesis of DNA

Although the methods for generating DNA fragments mentioned above are those most commonly used, the chemical synthesis is considered as an increasingly important method for generating DNA molecules. The chemical synthesis of specific gene sequences, regulatory sequences, oligonucleotide probes, primers and linkers is a technique in which solid phase synthesis is adopted. In chemical synthesis of DNA, two important strategies adopted are described below.

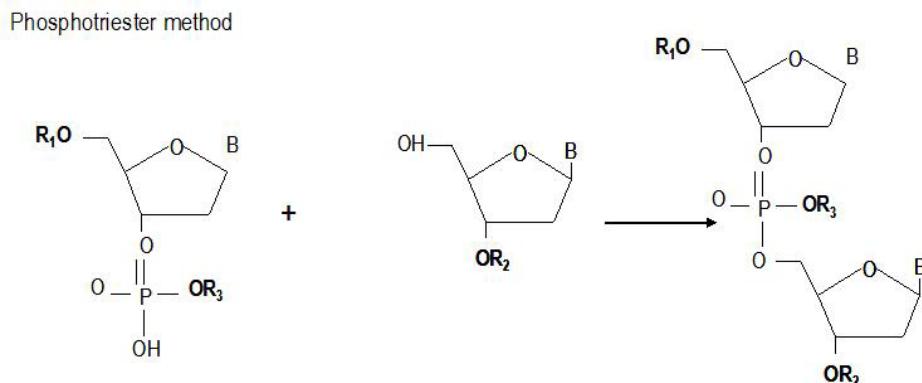
Phosphodiester method

In the phosphodiester method, 3' and 5' hydroxyl groups of deoxyribose are protected (R_1 and R_2). In this method, the phosphorus group between the two nucleosides is unprotected. These compounds are therefore soluble in organic solvents to a limited extent. The first significant successes, such as the synthesis of the genes for alanine and tyrosine suppressor tRNA for yeast and *E. coli* respectively were gained with the phosphodiester method.



Phosphotriester method

The phosphotriester method for the synthesis of oligodeoxyribonucleotides proceeds essentially in two steps: 1) preparation of suitably protected monomers and 2) coupling of the monomers in the desired sequence by an appropriate phosphorylation procedure.



In both protocols the 3' and 5' hydroxyl groups of the deoxyribose sugar are suitably protected (R_1 and R_2). In the phosphotriester method a third protecting group (R_3) is used for the hydroxyl group at the internucleotide bond.

Chemical synthesis of DNA has found an extraordinary number of applications in gene technology which include synthesis of partial or total gene sequences, primers for DNA and RNA sequencing, hybridization probes for the screening of RNA, DNA and cDNA or genomic libraries and adapters and linkers for gene cloning

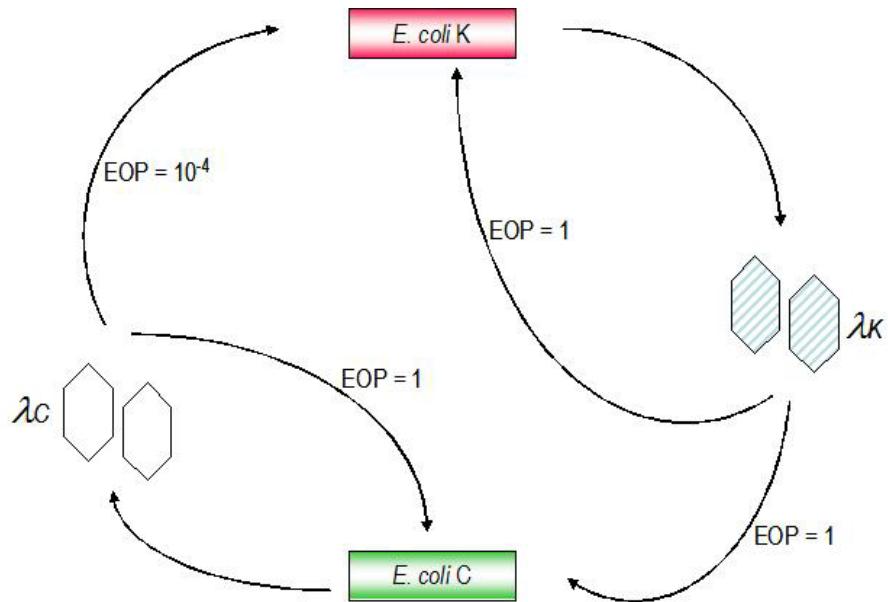
Cutting and joining the DNA fragments to vector DNA molecules

Restriction endonucleases: Tool for cutting DNA molecules

Techniques for cutting of DNA molecules into discrete fragments by specific enzymes were virtually unknown until the late sixties. A solution to this fundamental problem eventually grew from long standing research into the phenomenon of **host controlled restriction and modification system**.

Host controlled restriction and modification phenomenon can be well explained with the following example. If a stock preparation of phage is allowed to grow upon *E. coli* strain C and this stock is then tried upon *E. coli* C and *E. coli* K, the titres observed on these two strains will differ by several orders of magnitude, the titre on *E. coli* K being the lowest. The phages are said to be restricted by the second host strain (*E. coli* K) and the phenomenon is called **restriction**. When those phage that do result from the infection of *E. coli* K are now replated on *E. coli* K they are no longer restricted; but if they are first cycled through *E. coli* C they are once again restricted when plated upon *E. coli* K. The non-heritable change conferred upon the phage by the second host

strain (*E. coli* K) that allows it to be replicated on that strain without further restriction is called **modification**. These processes can occur whenever DNA is transferred from one bacterial strain to another. Conjugation, transduction, transformation and transfection are all subject to the constraint of host controlled restriction and this process is made possible by the enzymes called **restriction endonucleases**.



Restriction endonucleases

Nomenclature

Every restriction enzyme would have a specific name which would identify it uniquely. The first three letters, in italics, indicate the biological source of the enzymes, the first letter being the initial of the genus and the second and third being the first two letters of the species name. Thus restriction enzymes from *Escherichia coli* are called *Eco*; *Haemophilus influenzae* becomes *Hin*; *Diplocococcus pneumoniae* *Dpn* and so on. Then comes a letter that identifies the strain of bacteria; *Eco R* for strain R. Finally there is a roman numeral for the particular enzyme if there are more than one in the strain in question; *Eco RI* for the first enzyme from *E. coli* R, *Eco RII* for the second.

Types

The restriction endonucleases can be divided into three groups as type I, II and III. Types I and III have an ATP dependent restriction activity and a modification activity resident in the same multimeric protein. Both these types recognize unmethylated recognition sequences in DNA. Type I enzymes cleave the DNA at random site, whereas Type III cleave at a specific site. Type II

restriction modification system possess separate enzymes for endonuclease and methylase activity and are the most widely used for genetic manipulation. The properties of three types of restriction endonucleases and a list of enzymes are given below.

Property	Type I	Type II	Type III
Structure	Enzyme complex of 500-600 k dal composed of three separate subunits	Normally homodimers of 20-70 k dal	Heterodimers with subunits of 70 and 100 k dal
Composition	Multienzyme complex with R (endonuclease), M (methylase) and S (specificity) subunits	Separate enzymes; endonuclease is a homodimer, methylase a monomer	M subunit provides specificity on its own; functions as methylase; as heterodimer with R subunit; functions as methylase-endonuclease
Cofactors	Mg ²⁺ , ATP, S-adenosylmethionine (SAM) (needed for cleavage as well as methylation)	Mg ²⁺ , SAM (for methylation only)	Mg ²⁺ , ATP (for cleavage), SAM (needed for methylation: stimulate cleavage)
Recognition sites	Asymmetric, bipartite, may be degenerate; 13-15 base pairs containing interruption of 6 to 8 base pairs	Asymmetric, may be bipartite, may be degenerate; 4 to 8 base pairs normally 180° rotational symmetry	Asymmetric, uninterrupted, 5-6 nucleotide long with no rotational symmetry
Cleavage	Non-specific, variable distance (100-1000 nucleotides) from recognition site	Precise cleavage within recognition site at defined distance	Precise cleavage at a fixed distance; 25-27 nucleotides from recognition site
Example	EcoK	EcoRI	EcoP1

Some restriction endonucleases and their recognition sites

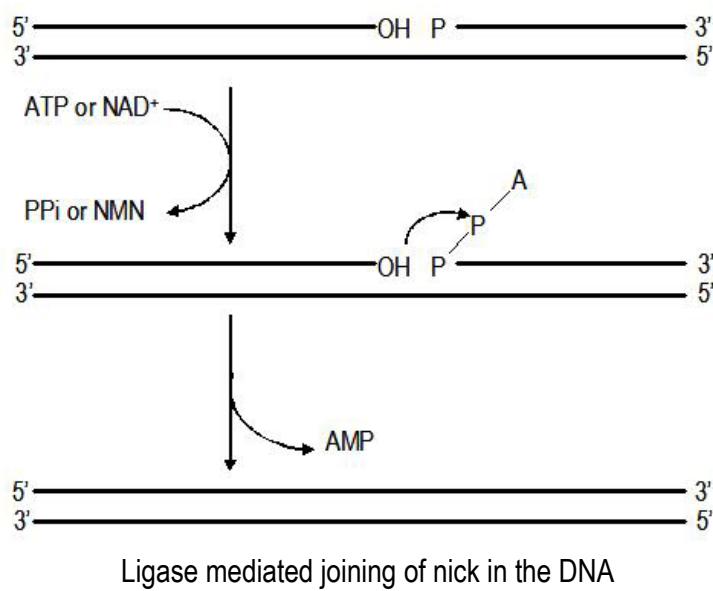
Enzyme	Recognition site	Enzyme	Recognition site
4-base cutters		6-base cutters	
<i>Mbo</i> I, <i>Dpn</i> I, <i>Sau</i> 3AI	/GATC	<i>Bgl</i> I	A/GATCT
<i>Msp</i> I, <i>Hpa</i> II	C/CGG	<i>Clal</i>	AT/CGAT
<i>Alu</i> I	AG/CT	<i>Pvu</i> II	CAG/CTG
<i>Hae</i> III	GG/CC	<i>Pvu</i> I	CGAT/CG
<i>Tai</i> I	AC/GT	<i>Kpn</i> I	GGTAC/C
		8-base cutters	
		<i>Not</i> I	GC/GGCC
		<i>Sbf</i> I	CCTGCA/GG

Ligases: Tool for joining DNA fragments

Joining DNA fragments of various types is yet another fundamental step in rDNA technology. This process is otherwise called as ligation and is achieved by the catalytic reaction of enzymes called ligases. These enzymes catalyses the formation of phosphodiester bonds between DNA molecules. The ligase enzymes of *E. coli* and phage T4 have the ability to seal the single stranded nicks between nucleotides in a duplex DNA.

Ligation

Although the reactions catalyzed by the enzymes of *E. coli* and T4 infected *E. coli* are similar, they differ in their cofactor requirements. The T4 enzyme requires ATP, while the *E. coli* enzymes require NAD⁺. In each case the cofactor is split to form an **enzyme-AMP complex**. The complex binds to the nick, which must expose a 5'- phosphate and 3'-OH group, and makes a covalent bond in the phosphodiester chain.



The other enzyme having utility in ligation is terminal deoxynucleotidyl-transferase. This adds an entire nucleotide to 3' end of the chain. It requires a source of energized nucleotides and simply adds them to the growing chain. This means that, if some DNA is mixed with terminal transferase and just one nucleotide, say the adenine nucleotide, the chain will grow as succession of adenines at the 3' end of the strand. If another chain is incubated with terminal transferase and thymine nucleotides it will have a protruding strand that is all thymine. If the above two strands are mixed together the complementary base pairing between the protruding strands will give duplex DNA.

Prevention of self-ligation

In rDNA technology, prevention of self-ligation in vector DNA molecules or passenger DNA molecules is considered more important. Generally vector DNA molecules are highly susceptible to self ligation thus forming recircularised DNA molecules. The presence of self ligated molecules reduces the probability of recovering desired recombinant clones. Self ligation can be reduced to some extent by adopting homopolymer tailing. Wherever homopolymer tailing is undesirable other strategies like directional cloning and dephosphorylation of termini can be followed.

Directional cloning: Directional cloning is otherwise called forced cloning. This is possible in a vector having two or more target sites in a non essential portion of the DNA. Cleavage at these sites cause the removal of the non essential DNA and produce a vector molecule with two different termini which are not complementary so that the individual vectors cannot recircularise.

Dephosphorylation of termini: The main function of DNA ligase is to produce a phosphodiester bond between adjacent nucleotides if one contains a 5' PO₄ group and the other a 3' -OH group. Thus, removal of terminal 5' PO₄ groups from the cleaved DNA will prevent self ligation. Dephosphorylation of termini can be carried out by treating linearised DNA with bacterial alkaline phosphatase. The dephosphorylated DNA molecules can be religated with phosphorylated passenger DNA to produce a functional recombinant DNA molecules.

DNA Vectors: The carriers of DNA molecules

DNA vectors and their properties

One of the most important elements in gene cloning is the vector, which in conjunction with the passenger DNA forms the recombinant DNA which can be propagated in suitable host cells. In order to perform its function, a vector must possess the following properties:

- They should be capable of autonomous replication in at least one host organism.
- They should be of small size, since this aids the preparation vector DNA and reduces the complexity of analyzing recombinant molecules.
- They should be capable of amplifying the cloned sequence by occurring in multiple copies. High copy number facilitates in maximizing expression of cloned genes.
- There should be a unique cleavage site for a range of restriction endonucleases. Occurrence of multiple cleavage sites reduces the likelihood of functional recombinant DNA formation.
- They should possess one or more genetic markers enabling easy selection of cloned molecules.
- They should permit detection by simple genetic tests, of the presence of passenger DNA inserted at cloning site.

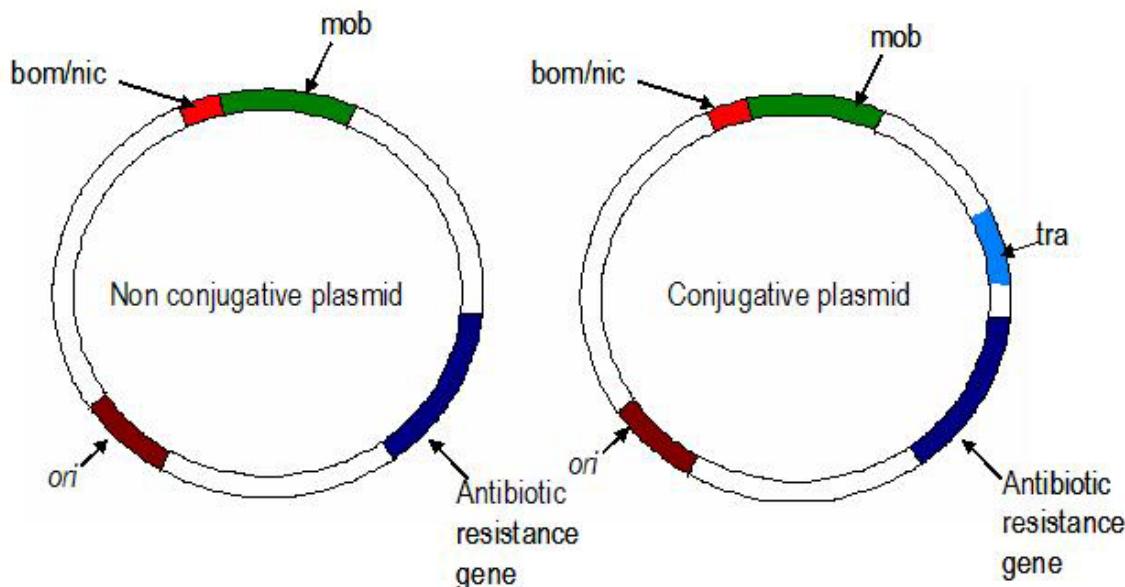
- They should have appropriate transcriptional and translational signals located adjacent to cloning sites for better expression of cloned DNA sequences.
- They should have host specificity when there is biological containment for a vector.

A variety of different cloning vectors have been developed by using the items mentioned above as guidelines. They are as follows: **plasmids, phages, cosmids, phasmids, shuttle vectors, expression vectors** and **single stranded DNA**

Plasmids

Plasmids are self replicating, double stranded, circular DNA molecules that are maintained in bacteria as independent extra chromosomal entities. These are also found in some yeast but not in higher eukaryotes. Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 to greater than 200×10^6 Da and are generally dispensable.

Plasmids can be grouped into two major types: **conjugative** and **non-conjugative**. In conjugative plasmids transfer genes (*tra*) and mobilizing genes (*mob*) are present whereas in non-conjugative plasmids *tra* genes absent. The non-conjugative plasmids can be mobilized by another conjugative plasmid present in the same cell, if the *mob* gene is intact.



General structure of plasmids.

Non-conjugative differ from conjugative plasmids by the absence of *tra* gene

Plasmids can also be categorized on the basis of their being maintained as multiple copies per cell (**relaxed plasmids** or **high copy number plasmids**) or as limited copies per cell (**stringent plasmids** or **low copy number plasmids**). The replication of stringent plasmids is

coupled to chromosome replication, hence their low copy number. Generally conjugative plasmids are of low molecular weight and present in multiple copies per cell. An exception is the conjugative plasmids RBK which has a molecular weight of 25×10^6 daltons and is maintained as relaxed plasmid.

Plasmids that carry specific sets of genes for the utilization of unusual metabolites are called as **degradative plasmids**. Some plasmids will not have any apparent functional coding genes and are called **cryptic plasmids**. Some of the plasmids do not coexist in the same host cell in the absence of selection pressure and are called **incompatible plasmids**. Some plasmids are capable of promoting their own transfer to a wide range of host. These plasmids are called as **promiscuous plasmids**. Plasmids can also be grouped into **narrow-host range plasmids** and **wide host range plasmids** based on their nature of infectivity. Based on the origin of plasmids, they can be grouped into **naturally occurring plasmids** and **synthetic plasmids**.

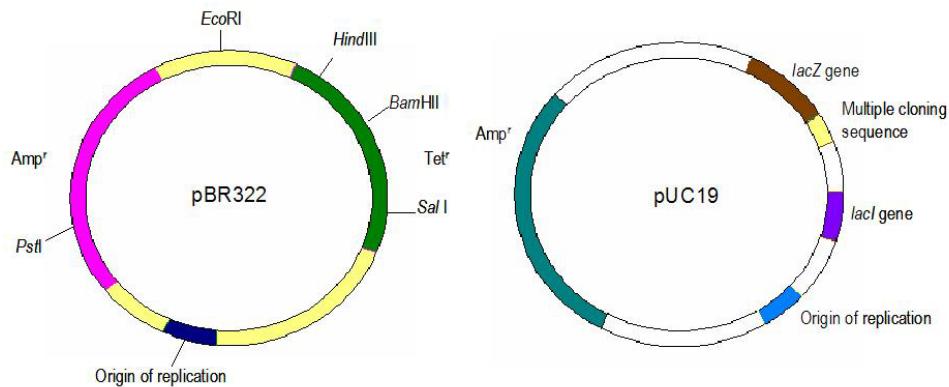
A list of naturally occurring plasmids and their properties are furnished below.

Plasmid	Size (kb)	Conjugative	Copy number	Amplifiable	Selectable marker
ColE1	7.0	-	10 – 15	+	E1 ^{imm}
RSF1030	9.3	-	20 – 40	+	Apr
CloDF13	10.0	-	10	+	DF13 ^{imm}
pSC101	9.7	-	1 -2	-	Tc ^r
R6K	42	+	10 – 40	-	Ap ^r Sm ^r
F	103	+	1 – 2	-	-
R1	108	+	1 – 2	-	Ap ^r Cm ^r Sn ^r
RK2	56.4	+	3 – 5	-	Ap ^r Km ^r Tc ^r

In general plasmid cloning vectors are designated by a lowercase 'p' which stands for plasmid, and some abbreviations that may be descriptive.

pBR322 plasmid

Plasmid pBR322 is the one of the best studied and most often used "general purpose" plasmids. The BR of the pBR322 recognizes the work of the researchers F. Bolivar and R. Rodriguez, who created the plasmid and 322 is a numerical designation that has relevance to these workers. pBR322 is 4362 base pair long and completely sequenced. pBR322 carries two antibiotic resistance genes. One confers resistance to ampicillin (Amp^r) and the other confers resistance to tetracycline (Tet^r). There are eleven known enzymes which cleave pBR 322 at unique sites. For three of the enzymes, *Hind* III, *Bam* HI and *Sal* I, the target site lies within the Tet^r genes and for another two, *Pst* I and *Pru* I, they lie in Amp^r genes. Thus cloning in pBR 322 with the aid of these enzymes results in insertional inactivation where the inserted DNA disrupts the function of the gene containing the cloning site. Where the cloning site is within in an antibiotic resistance gene, such insertional inactivation results in transformants sensitive to the appropriate antibiotic. Thus, insertional inactivation helps in the selection of recombinants.



General structure of pBR322 and pUC19 plasmids.

pUC19 plasmid

Plasmid pUC19 is 2686 bp long and contains an ampicillin resistance (*Amp'*) gene, a regulatable segment of β - galactosidase gene (*lacZ*) of the lactose operon of *E. coli*, *lacI* gene that produces a repressor protein that regulates the expression of *lacZ* gene, a short sequence with multiple cloning sites (*EcoRI*, *SacI*, *KpnI*, *XmaI*, *Smal*, *BamHI*, *XbaI*, *Sall*, *HincII*, *AccI*, *BspMI*, *PstI*, *SphI* and *HindIII*) and the origin of replication from pBR322. The presence of *lac Z* and *lacI* genes allows to select the recombinants based on the β - galactosidase production in the presence of isopropyl- β -D-thiogalactopyranoside (IPTG), an inducer of the *lac* operon. (UC in pUC stands for University of California).

Phages

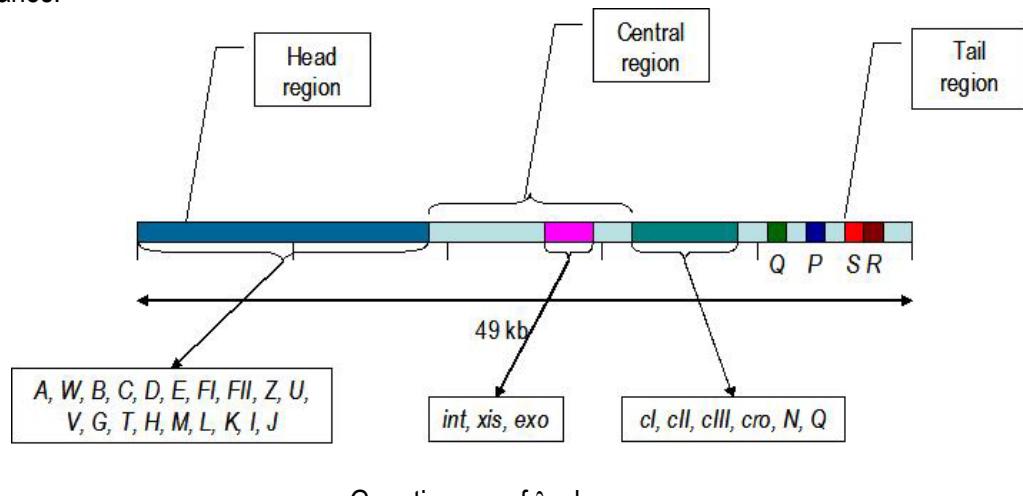
Derivatives of phage have been developed as cloning vectors since the early days of gene technology. The phage derivatives are considered to be the most suitable cloning vehicles for cloning genomic eukaryotic DNA because of the following advantages over the plasmids.

- Thousands of phage plaques can be obtained in a single petridish.
- Selection by DNA-DNA hybridisation is possible
- In vitro* packaging into empty phage head is possible thus increasing phage infectivity
- Size selection of the packaged DNA is possible
- Millions of independently cloned virus particle can be constituted to form a gene library.

Bacteriophage is a genetically complex but very extensively studied virus of *E. coli*. The DNA of phage, in the form in which it is isolated from the phage particle is a linear duplex molecule of 48502 bp (~49kb) in length. The DNA isolated from virus particles is a double stranded linear molecule with short complementary single stranded projections of 12 nucleotides at its 5' ends.

These cohesive termini, also referred to as **cos sites**, allow the DNA to be circularized after infection of the host cell.

The genetic map of phage λ comprises approximately 40 genes which are organized in functional clusters. Genes coding for head and tail are proteins (genes A-J) are on the left of the linear map. The central region contains genes, such as *int*, *xis*, *exo* etc. which are responsible for lysogenisation i.e the process leading to the integration of viral DNA and other recombination events. Much of this central region is not essential for lytic growth. Genes to the right of the central region comprise six regulatory genes, two genes (O and P) which are essential for DNA replication during lytic growth and two more genes (S and R) which are required for the lysis of the cellular membranes.



Genetic map of λ phage

In the phage DNA, larger central region is not essential for phage growth and replication. This region of phage can be deleted or replaced without seriously impairing the phage growth cycle. Using this non-essential region of phage λ , several phage vector derivatives have been constructed for efficient gene cloning.

Types of phage vectors

Wild type phage DNA itself cannot be used as a vector since it contains too many restriction sites. Further, these sites are often located within the essential regions for phage's growth and development. From these wild phages, derivatives with single target sites and two target sites have been synthesized. Phage vectors which contain single site for the insertion of foreign DNA have been designated as **insertional vectors**; vectors with two cleavage sites, which allow foreign DNA to be substituted for the DNA sequences between those sites, are known as **replacement vectors**. Apparently if too much non-essential DNA is deleted from the genome it cannot be packaged into phage particles efficiently. For both types of vector, the final recombinant genome must be between 39 and 52 kb of the wild type phage genome, if they are to be packaged into infectious particles. Insertion vectors must therefore be at least 39 kb in length to maintain their viability. This places an upper limit of about 12 kb for the size of foreign DNA fragments which can

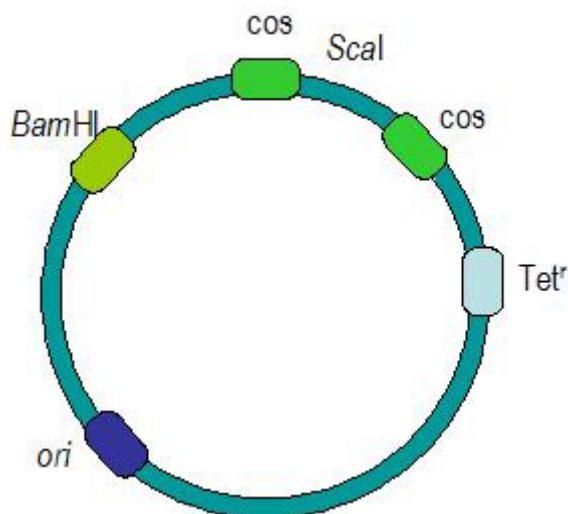
be inserted. Replacement vectors have a larger capacity because the entire non-essential region can be replaced, allowing the cloning of the fragments upto 22 kb. Several types of vectors have been developed which allow direct screening for recombinant phages and are useful for cloning specific DNA fragments. A list of phage vectors with their characteristics is given below.

Phage vector	Size (kb)	Enzyme	Size of insertion (kb)
Charon 4A	45.3	EcoRI, <i>Xba</i> I	7-20
λ L47.1	40.6	EcoRI, <i>Hind</i> III, <i>Bam</i> HI	8.6-21.6
λ Dam sr1 λ 3	38.3	EcoRI	13
λ 1059	44.0	<i>Bam</i> HI	6.3-24.4

Cosmids

Plasmids containing phage **cos** sites are known as **cosmids**. Cosmids can be used to clone large fragments of DNA by exploiting the phage *in vitro* packaging system. Since cosmids have advantages of both plasmids and phage vectors they can be delivered to the host by the more efficient infection procedures rather than by transformation. Cloning with cosmid vectors has widened the scope of plasmid cloning in the following ways.

- The infectivity of plasmid DNA packaged in phage head is at least three orders of magnitude higher than that of pure plasmids DNA.
- The process almost exclusively yields hybrid clones so that a subsequent selection for recombinant DNA becomes unnecessary.
- In contrast to normal plasmid transformations, the system strongly selects for clones containing large DNA inserts. It is therefore, particularly well suited for generating genomic libraries.



General structure of a cosmid vector

The following table provides a list of cosmid vectors and their structural features.

Cosmid	Size (kb)	Cleavage sites	Size of insertion (kb)
MUA3	4.76	EcoRI/PstI/PvuII/PvuI	40 – 48
pJB8	5.40	BamHI	32 – 45
Homer I	5.40	EcoRI/Clal	30 – 47
Homer II	6.38	SstI	32 – 44
pJC79	6.40	EcoRI/Clal/BamHI	32 – 44

Phasmids

Phasmids, also called as phagemids, are hybrids formed between small multicopy plasmids and bacteriophages. A phasmid can be propagated as a plasmid or lytically as a phage. Lytic functions of phasmid can be switched off by propagation in the appropriate lysogene where the plasmid origin of replication is used for maintenance. The phasmid may replicate as phage if propagated in a non-lysogenic strain. In the case of phasmids based on λ , such as λ 1130, the temperature sensitive gene, cl_{857} carried by the vector may be used to switch between replication modes, simply by growing the host at the permissive (plasmid mode) or restrictive (phage mode) temperature.

Phasmids are particularly useful in the generation and analysis of mutations exhibiting non-selectable or lethal phenotypes, such as those affecting the replication of plasmids. Phasmids may also be used as phage replacement vectors and for directing the high level expression of protein from cloned sequences by replication in the phage mode.

Shuttle vectors

Shuttle vectors normally comprise an *E. coli* plasmid or part of such plasmid (e.g., pBR 322), ligated *in vitro* to a plasmid or virus replicon from another species. Shuttle vectors can be made, for example, for *E. coli/B. subtilis*, *E. coli/yeast* or *E. coli/mammalian* cells. The shuttle vector strategy permits the exploitation of the many manipulative procedures, such as amplification, available in *E. coli* (or other genetically well characterized species such as *B. subtilis* or *S. cerevisiae*) backgrounds. The ability to transfer cloned genes across species boundaries is of potential value in the genetic manipulations of industrially important species and this can be achieved by using shuttle vectors.

Expression vectors

In DNA cloning experiments all the genes cloned are not expressed fully because of weak promoters in vector DNA. This can be dramatically improved by placing such genes downstream of strong promoters. An additional problem in maximizing expression of cloned genes in *E. coli* which is frequently encountered with genes from a heterologous source is that the gene carries no translation start signal which can be efficiently recognized by the *E. coli* translation system. This problem may arise for heterologous genes cloned into any host. Thus, even though the gene can

be transcribed from a promoter within the vector, the resulting mRNA is poorly translated and little or no protein product will be synthesized. In such cases alternative strategies available are fusing the gene to amino terminal region of vector gene that is efficiently translated in the host or coupling the gene to a DNA fragment carrying both strong promoter and a ribosomal binding site. Vectors with this additional feature are called **expression vectors**.

Host systems for cloned vectors

***E. coli* system**

Vectors and their hosts form integrated system for constructing and maintaining recombinant DNA molecules. The choice of a particular **host -vector system** depends on a variety of factors, including ease and safety of manipulations and the likelihood of expression of cloned genes. Among the host system *E. coli* system remains well exploited one. Several strains, such as x1776, have been disabled for use as safe host in potentially hazardous cloning experiments. Most cloning experiments can, however, be carried out with strains that are considerably less disabled and hence more easily handled than other hosts.

***Bacillus subtilis* system**

Bacillus subtilis is the best characterized of all Gram positive bacteria. It has a well defined genetic map and efficient systems for transformation and transfection. In addition, *B subtilis* is commercially important since procedures for the synthesis of peptide antibiotic and extracellular enzymes, such as proteases are made available. Further, the species is nonpathogenic which makes it a safe host for cloning potentially hazardous genes. However, *B. subtilis* does sporulate readily, thus increasing the probability that cloned genes would survive outside the laboratory or fermentor. Asporogenous mutants with increased autolytic activity may however, be used as high containment host strains. Several other cloning systems such as systems of streptococci, staphylococci, streptomycetes, etc. are developed for gene manipulation experiments.

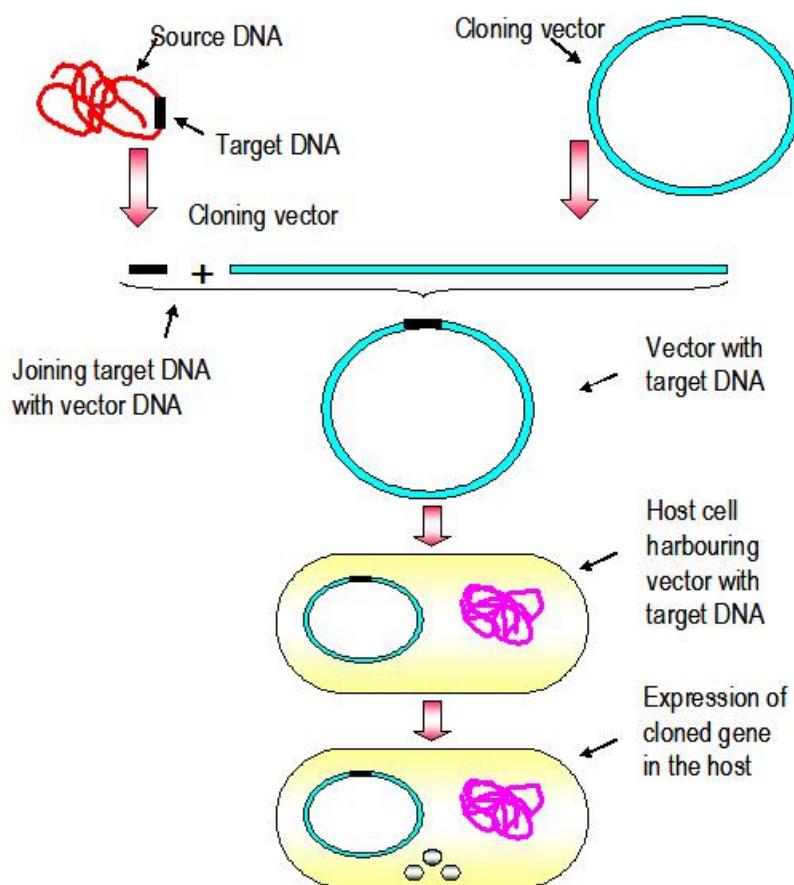
Yeast host system

Actinomycetes host system is interesting for a number of reasons. The actinomycetes synthesize a wide range of metabolites which provide the majority of medically and agriculturally important antibiotics. Actinomycetes genes may also be the primary source of clinically important antibiotic resistance determinants. Finally they have a complex morphological development cycle which involves a series of changes from vegetative mycelial growth to spore formation. The real interest in gene cloning in actinomycetes is that it would facilitate the development of industrial strains which give increased antibiotic yields.

CHAPTER 13

RECOMBINANT DNA TECHNOLOGY: METHODS

Progress in any scientific discipline is dependent on the availability of technologies, tools and methods that extend the range and sophistication of experiments, which may be performed. Over the last 30 years or so, this has been demonstrated in a spectacular way by the emergence of genetic engineering. This field has grown rapidly to the point where, in many laboratories around the world, it is now routine practice to isolate a specific DNA fragment from the genome of an organism, determine its base sequence and assess its function. The technology is also now used in many other applications, including forensic analysis of scene-of-crime samples, paternity disputes, medical diagnosis, genome mapping and sequencing and the biotechnology industry. What is particularly striking about the technology of gene manipulation is that it is readily accessible to individual scientists with well established laboratory protocols and methodologies.

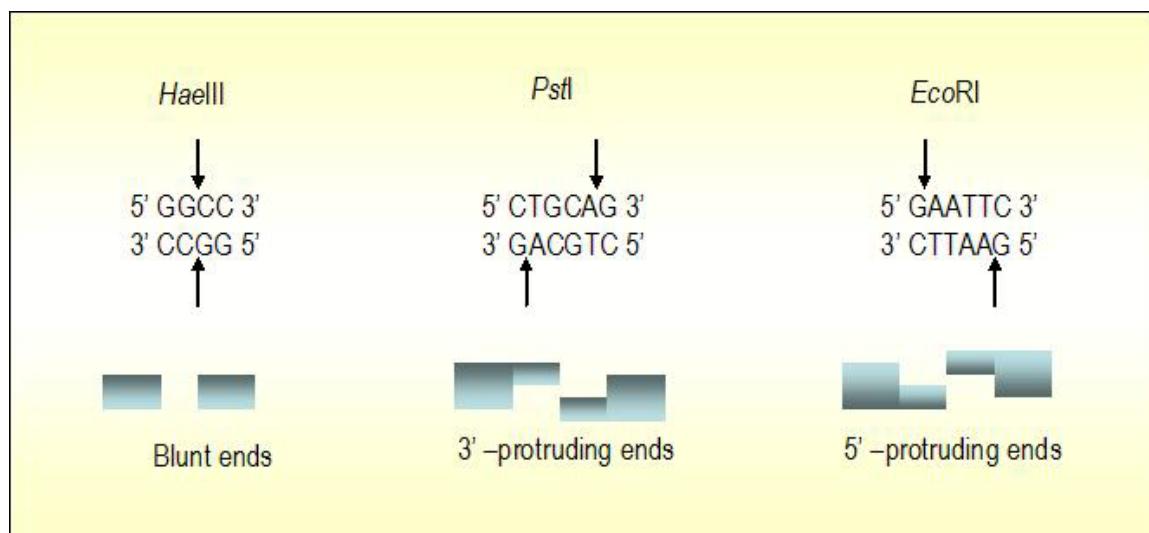


Basic steps involved in a rDNA technology

Although there are many, diverse and complex techniques involved, the basic principles of gene manipulation are reasonably simple. The premise on which the technology is based is that genetic information, encoded by DNA and arranged in the form of genes, is a resource that can be manipulated in various ways to achieve certain goals in both pure and applied science by adopting the following steps: 1) generating DNA fragments using restriction endonucleases, 2) ligation strategies, 3) cloning strategies, 4) modes of introduction of cloned vectors into hosts and 5) selection strategies for cloned genes

Generating DNA fragments using restriction endonucleases

Restriction endonucleases are enzymes that recognize specific sequences within duplex DNA molecules and cut the DNA at or near these sites. More than 500 different restriction endonucleases have been discovered. These enzymes can be grouped into three types viz. Type I, II and III. For practical purposes, the Type I and III restriction enzymes are not much used in rDNA technology. The real precision scissors are the Type II enzymes. Type II restriction endonucleases recognize and cut DNA within particular sequences of tetra, penta, hexa or hepta nucleotides which have an axis of rotational symmetry. In the following examples, different restriction enzymes cut the DNA at specific sequences as indicated by arrows.



Among the restriction enzymes, some enzymes cut the DNA molecules to give **blunt end** fragments otherwise termed as **flush end** DNA fragments and some others produce DNA molecules where one of the strands will have protruding 5' or 3' termini. These fragments are called fragments with **cohesive ends** or **sticky ends**. The majority of the recognition sequences for restriction endonucleases are **palindromic**, that is the sequence is the same if read from 5' to 3' from both complementary strands.

The sites of cut made by endonucleases are called **target sites** or **cleavage sites** and the number of these sites in a DNA molecule depends on the size of the DNA, its base composition and the GC content of the recognition site. The number and size of the fragments generated by a restriction enzyme depends on the frequency of occurrence of the target site in the DNA to be cut. Assuming a DNA molecule with a 50 percent G+C content and a random distribution of four bases, a restriction enzyme recognizing a particular tetranucleotide sequence will be able to cut the DNA molecules into fragments at once in every 4^4 (i.e. 256) nucleotide pairs. If the enzyme is having the property of making cuts in hexanucleotide sequences means, the cuts will be made at every 4^6 (i.e. 4096) nucleotide pairs and an eight nucleotide recognition sequence 4^8 (65536) base pairs.

Restriction enzymes that have the same recognition sequences can be isolated from different bacterial species. Such enzymes are called **isoschizomers**. An example is provided by **MboI** (*Moraxella boris*) and **Sau3A** (*Staphylococcus aureus*), both of which recognize the sequence **GATC**. Furthermore, some restriction enzymes generate cohesive ends that can reanneal with identical termini produced by other enzymes. For instance, DNA cleaved with *BamHI* (GGATCC) has compatible ends with DNA cleaved with *BglII*, *MboI*, *Sau3A*, etc.

The number of restriction fragments made by an enzyme would be reduced if there is a methylation of restriction sites. In some cases, the DNA recognition by an enzyme will not be altered by methylation and enzymes of this nature are said to be **enzymes with star activity** e.g. *EcoRI*, *BarnHI* and *SaI*.

Ligation strategies

In rDNA technology, sealing discontinuities in the sugar-phosphate chains, otherwise called as ligation, is vital step. This process is catalyzed by DNA ligase by repairing broken phosphodiester bonds. During ligation, the enzyme's activity is influenced by factors such as 1) substrate specificity, 2) temperature and 3) salt concentration

Substrate specificity

The DNA ligase isolated from *E. coli* and T4 differ not only in their cofactor requirements but also in their substrate specificities. The physiological substrate for both enzymes is breakage point at phosphodiester bond between neighbouring 3' hydroxyl and 5' phosphate ends still held together by an intact complementary strand. Another substrate for either enzyme contains the open and staggered phosphodiester bonds formed through reassociation of the protruding termini of different DNA molecules generated by digestion with certain type II restriction endonucleases.

Among the two ligases, T4 DNA ligase is capable of ligating nicks in the RNA chains of double stranded RNA-DNA hybrids and annealing RNA termini with DNA strands. One of the most remarkable properties of T4 DNA ligase, which distinguishes from the bacterial DNA ligase, is its ability to accomplish blunt end ligation of double stranded DNA molecules.

Temperature

Reaction temperature is another important parameter which influences ligase activity. The optimal ligation temperature is 37°C, but at this temperature the hydrogen bond joint between the sticky end is unstable. The optimum temperature for ligating the cohesive termini is therefore a compromise between the rate of enzyme action and association of the termini and has been found by experiment to be in the range 4 -15°C.

Concentration

At constant ionic strength, preferences for intra-or-intermolecular reactions depend on the length of DNA fragments and the DNA concentration. The smaller the DNA fragment at a given DNA concentration, more is the intramolecular reactions, leading to circularization of DNA. At constant lengths, the probability of circularization increase with decreasing concentrations.

Ligation methods

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process which has been extensively used to create artificial recombinants. If the termini of DNA fragments are not compatible, there are other methods to ligate the fragments.

Cohesive end ligation

The cohesive end ligation is possible when both the foreign DNA to be cloned and the vector DNA possess the same molecular ends. The compatible sticky ends have been generated by cleavage with the same enzyme on the same recognition sequences of both foreign DNA and vector DNA. Using DNA ligase, these molecules can be ligated without any problem.

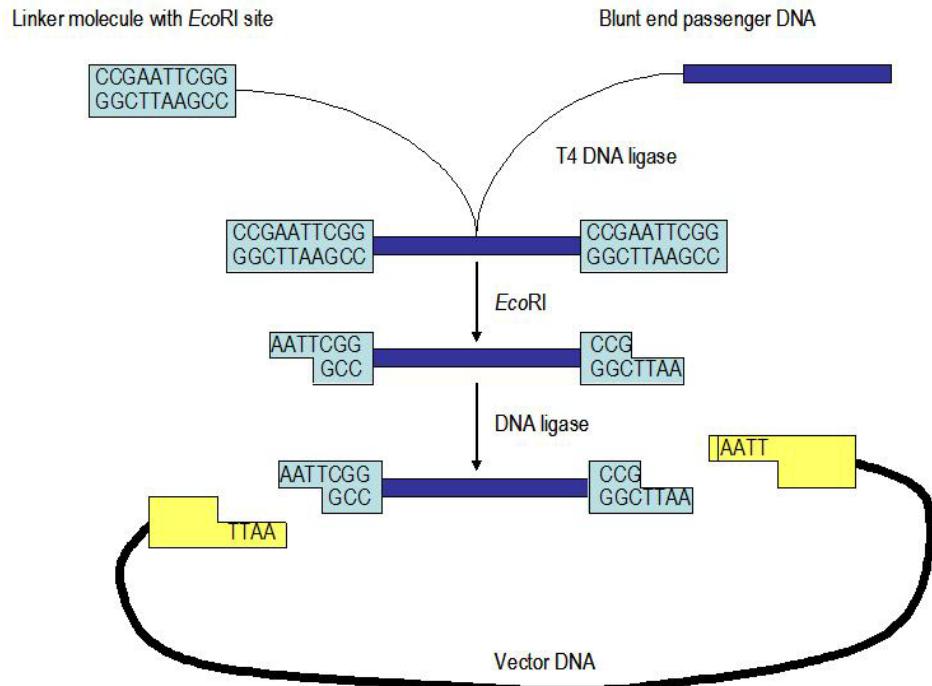
Very often it is necessary to ligate DNA fragments with different and non-compatible ends, or blunt ends with either staggered 3' or 5' ends. Incompatible DNA fragments with recessed ends can be ligated by modifying their ends by any one of the following methods *viz.*, (i) filling in recessed 3' termini and (ii) renewal of 5' protruding termini.

Blunt end ligation

The E. coli DNA ligase will not catalyze blunt end ligation except under special reaction conditions of macromolecular crowding. The unique property of T4 DNA ligase was used to ligate DNA fragments with blunt ends involving short decameric oligonucleotides called **linkers**.

Using linkers

Short oligonucleotides (decamers) which contain sites for one or more restriction enzymes are used to facilitate the ligation process among the DNA fragments with blunt ends.



Joining of blunt end DNA to a vector using linkers

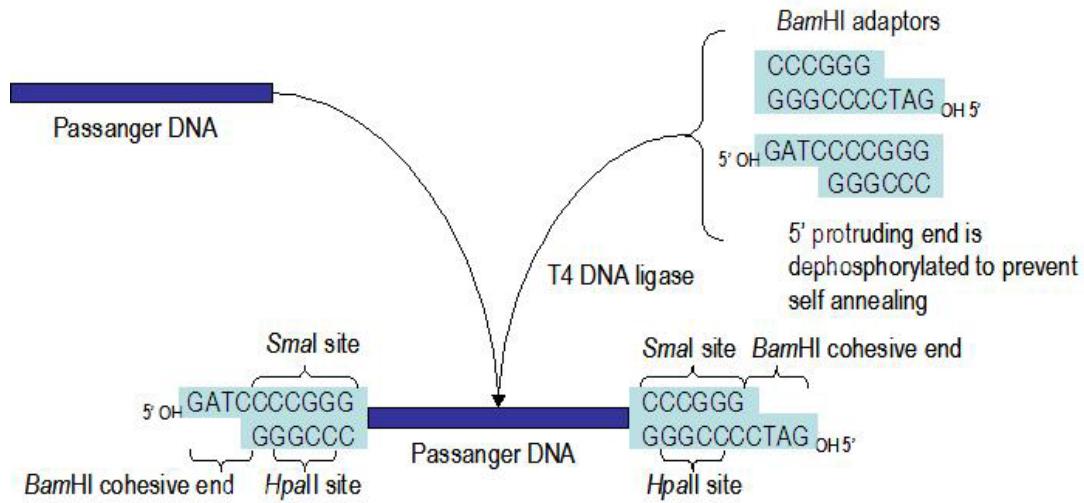
The linker molecules can be ligated to both ends of the foreign DNA to be cloned and then treated with restriction endonuclease to produce sticky end fragments which can be incorporated into a vector molecule that has been cut with the same restriction endonuclease. Insertion by means of the linker creates restriction sites at each end of the foreign DNA, and thus enables the foreign DNA excised and recovered after cloning and amplification in the host bacterium.

Using adaptors

The other strategy adopted for ligating DNA fragments with blunt ends is using **adaptors**. The adaptor molecules are synthetic deoxynucleotides that can be used to join two incompatible cohesive ends, two blunt ends or a combination of both. Such adaptors are of several types viz., **preformed, conversion and single stranded adaptors**.

Preformed adaptors

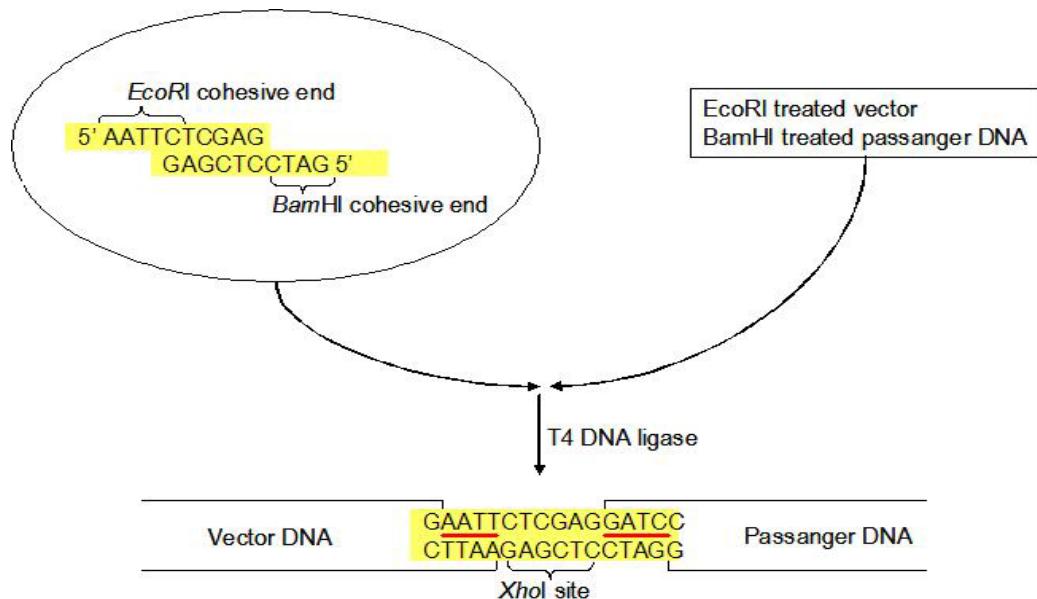
Preformed adaptors are short DNA duplexes with at least one cohesive end. The problem of internal cleavage of the insert DNA can be overcome by using a preformed adaptor that will introduce a new restriction site. For example, an adaptor having *Bam*HI cohesive ends and sites *Hpa*II and *Sma*I can be attached to passenger DNA and inserted into a *Bam*HI vector. After cloning, passenger DNA can be excised from the hybrid by using any one of the enzymes that recognize the restriction sites within the adaptor region.



Use of preformed adaptors

Conversion adaptors

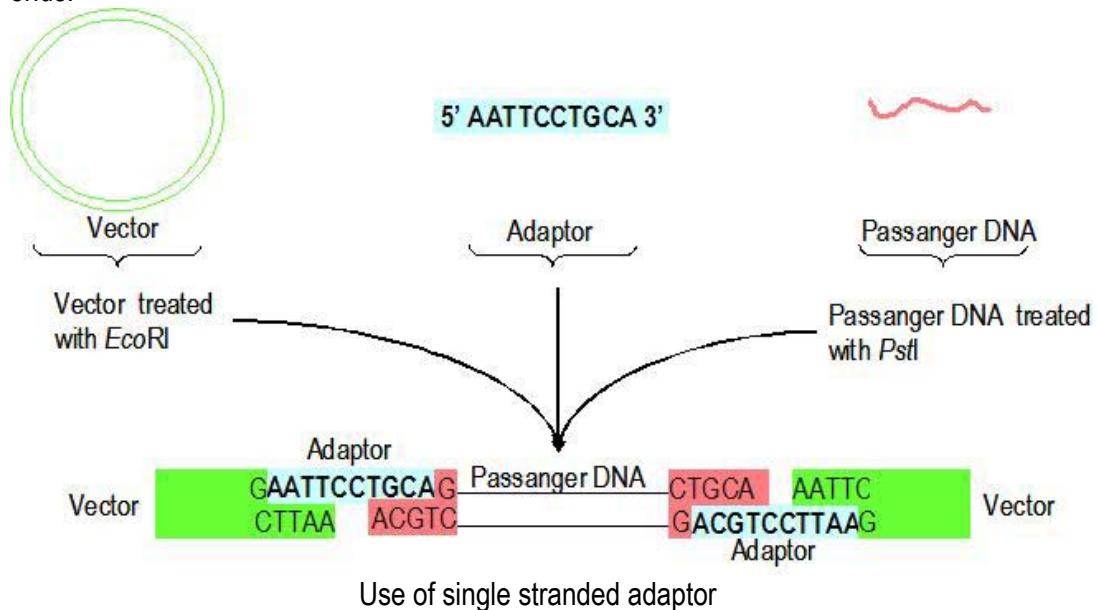
Conversion adaptors are synthetic oligonucleotides bearing different cohesive restriction termini. Such adaptors enable vector molecules that have been cleaved with one endonuclease to be joined to passenger fragments that have been cleaved with another. Often these adaptors contain internal restriction sites that permit recovery of the passenger fragment, for example, the *Eco*RI-*Bam*HI adaptor contains a site for *Xba*I.



Use of conversion adaptor

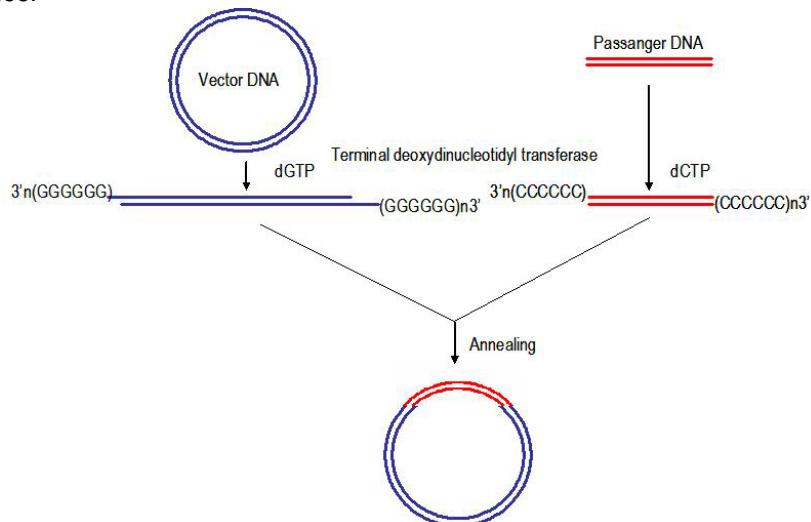
Single stranded adaptors

Single stranded adaptors can be used to make 3'-protruding cohesive ends compatible with 5' protruding ends. Such adaptors permit the insertion of passenger fragments into sites on vectors from which they would otherwise be precluded because of incompatible cohesive ends.



Homopolymer tailing

Homopolymer tailing is the other method adopted to clone blunt DNA molecules, especially cDNA molecules.



Homopolymer tailing

The addition of several nucleotides of single type to the 3' blunt end of DNA molecule is catalyzed by the enzyme **terminal deoxynucleotidyl transferase**. The terminal transferase permits the addition of complementary homopolymer tails (50 to 150 dA or dT long and about 20 dG or dC long) to 3' end of plasmid vector and passenger DNA. These tails can reanneal to form open circular hybrid molecules, which can be ligated *in vitro* or more commonly *in vivo* following transformations to produce functional recombinant molecules.

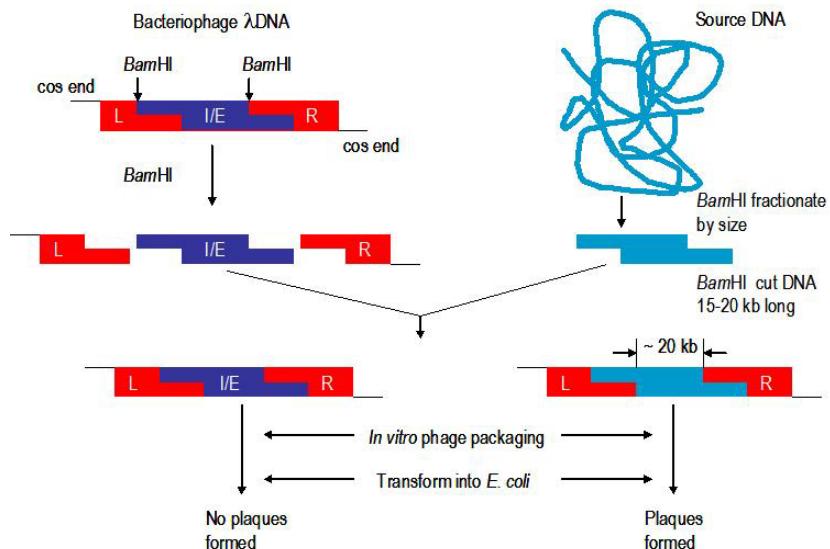
Cloning strategies

In vitro packaging of phage or cosmid DNA

The phage or cosmid vectors can be introduced into the host cells by packaging the DNA into phage capsids. The important events in the *in vitro* packaging are as follows: 1) encapsidating the DNA with the phage head precursor and capsid protein and 2) incorporating the products of genes responsible for head and tail. The principle of packaging *in vitro* is to supply the ligated recombinant DNA with high concentrations of phage head precursor, packaging proteins and phage tails. A principal difficulty of this *in vitro* packaging system is that the endogenous DNA from the phage lysates competes with exogenously added recombinant DNA.

In vitro packaging of phage DNA

Most bacteriophage cloning vectors have been constructed from the bacteriophage λ chromosome. The complete 48,502 nucleotide-pair sequence of the wild type genome is known along with the functions of all of its genes. The central one-third of the λ chromosome contains genes that are required for lysogeny, but not for lytic growth.

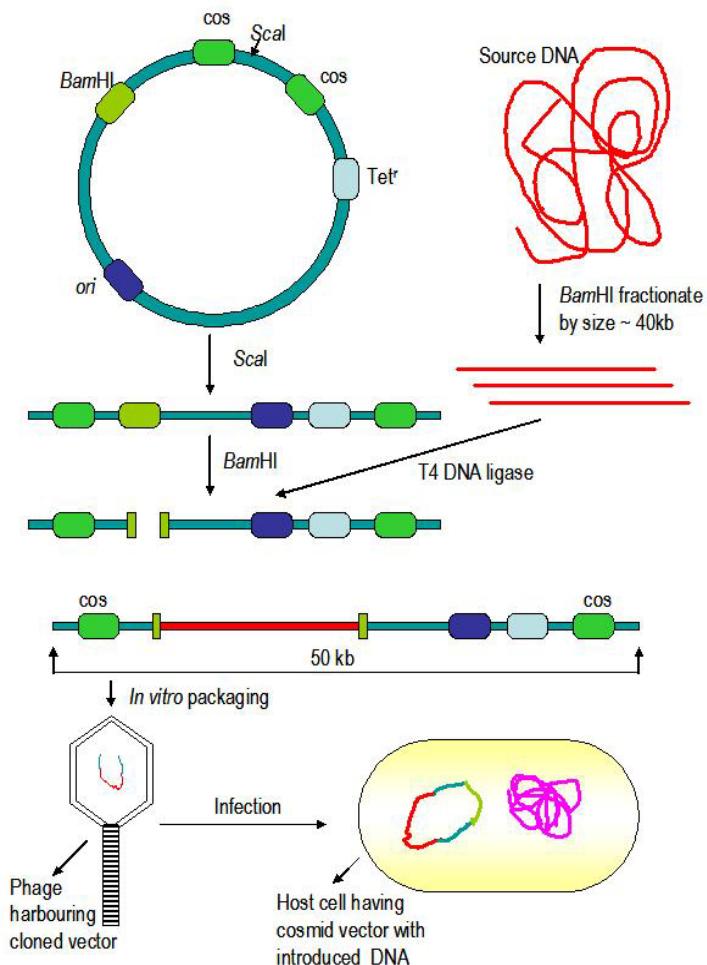


Using bacteriophage as cloning vector

Thus the central part (about 15 kb in length) of the λ chromosome can be excised with restriction enzymes and replaced with foreign DNA. The resulting recombinant DNA molecules can be packaged in phage heads *in vitro*. The phage particles can inject the recombinant DNA molecules into *E. coli* cells, where they will replicate and produce clones of the recombinant DNA molecules.

In vitro packaging of cosmid DNA

Some eukaryotic genes are larger than 15 kb in size and cannot be cloned intact in either plasmid or lambda phage cloning vectors. For this and other reasons, scientists have developed vectors that can accommodate larger DNA insertions. The first such vectors, called **cosmids** (for cos site and plasmid), were hybrids between plasmids and the phage λ chromosome. Cos stands for cohesive site, in reference to 12 base complementary single stranded termini in the mature λ chromosome.



Using cosmid as a cloning vector

The cos site is recognized by the λ DNA packaging apparatus, which makes staggered cuts at this site during packaging to produce complementary cohesive ends of the mature lambda chromosome. Cosmids combine the key advantages of plasmid and phage vectors; they possess 1) the plasmid's ability to replicate autonomously in *E. coli* cells and 2) the *in vitro* packaging capacity of λ chromosome. The cosmid vector is capable of accommodating inserts of 35 – 45 kb in size.

Modes of introduction of cloned vectors into hosts

The foreign DNA materials are incorporated into the hosts by exploiting phenomena viz., **transformation, conjugation, transduction and transfection**.

Transformation

Transformation is the process by which microorganisms take up naked DNA and subsequently acquire an altered genotype. Transformation has been reported in various bacteria, in yeast and in some filamentous fungi. This process involves the binding of DNA to competent cells, uptake of the DNA and its establishment within recipients (either as a replicon or by recombination with resident replicon). The transforming DNA may be of chromosomal or plasmid origin. The ability to introduce plasmids into cells by transformation is an important prerequisite for gene cloning *in vitro*. Accordingly, transformation systems are now being developed for a wide range of microorganisms that have not previously been explored genetically.

Conjugation

Conjugation is the process of gene transfer that requires cellular contact between donors and recipients. The plasmids in bacteria having the property of conjugation are called as conjugative plasmids. These plasmids have *tra* gene which controls the transfer of genes. Besides *tra* gene some conjugative plasmids possess *mob* gene which transfers some of the non-conjugative plasmids that are resident in the same cell. Nature of conjugation varies with gram negative and gram positive bacteria.

Transduction

Transduction is the process of gene transfer that is mediated by a bacteriophage. There are two types of transductions viz., generalized transduction and specialized transduction. In generalized transduction random parts of donor DNA are packaged in phage capsids and transferred to recipient cells. In specialized transduction, specific regions of bacterial DNA in covalent union with phage DNA are packaged and transduced. Both types of transductions are normally mediated by temperate phages, although certain virulent phages can mediate generalized transduction. The amount of bacterial DNA that can be transduced by a phage can be increased by using phage derivatives that carry deletions in non-essential regions of the phage genome.

Transfection

The uptake and infection of cells with naked viral DNA is generally termed transfection.

Expression of cloned genes

The expression of cloned DNA in the form of functional protein depends on the factors *viz.*, 1) transcription of the appropriate gene, 2) efficient translation of the mRNA and 3) post translational processing and compartmentalization of the synthesized protein. A failure to perform correctly in any one of these processes results in the failure of a given gene to be expressed. However, the expression of cloned genes can be improved by constructing the vectors which give improved transcription of inserts and positioning cloned inserts in the correct translational reading frame.

Selection of recombinants

Making recombinant molecules is a game with very long odds against success. Even when the bits of DNA have been joined up and inserted into cells, only very few cells out of many tens of thousands will contain the recombinant molecule and all the technical expertise in the world is no use whatsoever unless one can find the cell that contain the recombinant DNA. Techniques for selecting the few valuable cells from the mass of useless ones are thus of paramount importance. In the following section the most commonly used method for the selection of recombinant DNA are discussed.

Directional selection

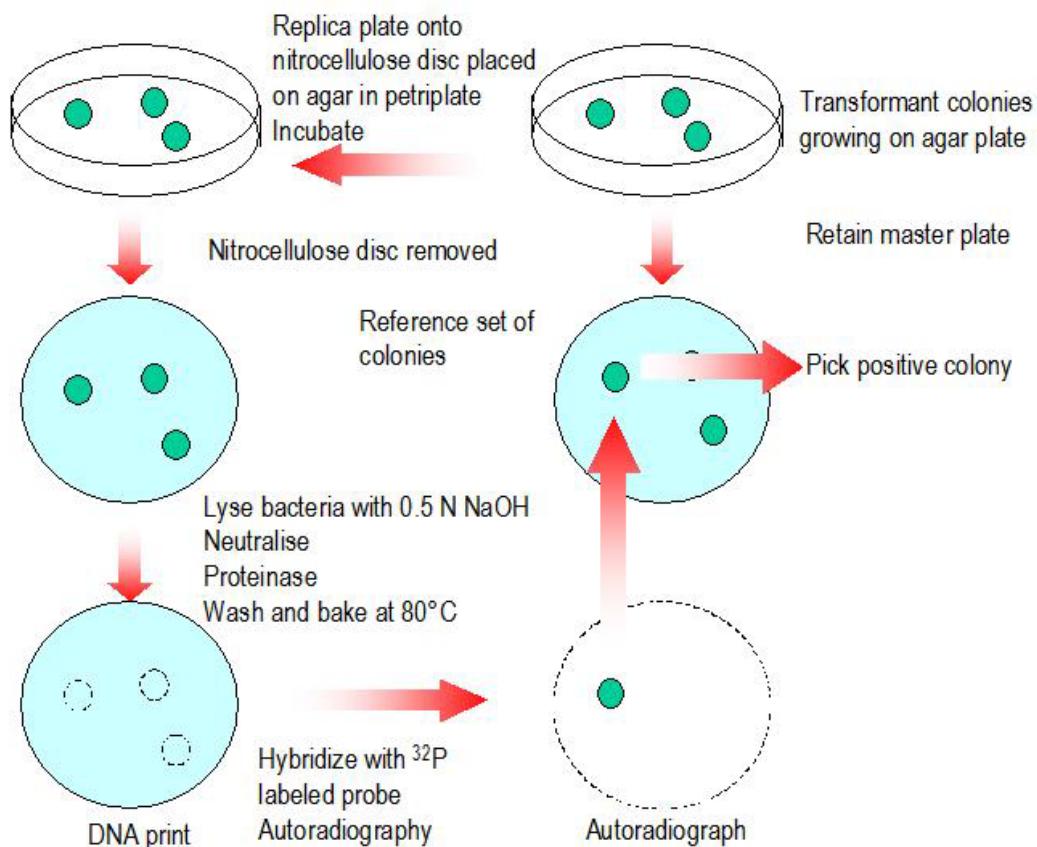
The phenotypes conferred by the cloned genes on the host are used as markers of selection. All useful vector molecules carry a selectable genetic marker or have a genetically selectable property. Plasmid vectors generally possess drug resistance or nutritional markers and in phage vectors the plaque formation itself is the selectable property.

Insertional inactivation

The technique depends upon homologous recombination between DNA cloned and the host genome. If the cloned sequence lacks both promoter and sequences encoding essential regions of the carboxyl terminus of the protein, recombination with homologous genomic sequences will cause gene disruption and produce a mutant genotype. On the other hand, if the cloned fragment contains appropriate transcriptional and translational signals, homologous recombination will result in synthesis of a functional mRNA transcript, and no mutant phenotype will be observed.

Colony hybridization

Various recombinant detection methods employing hybridization with DNA isolated and purified from the transformed cells have been developed of which screening procedure to detect DNA sequences in transformed colonies by hybridization *in situ* with radioactive probe RNA remains as the best choice. In this method, the colonies to be screened are at first replica plated onto a nitrocellulose filter disc that has been placed on the surface of an agar plate prior to inoculation. A reference set of these colonies on the master plate is retained. The filter paper bearing the colonies is lysed and their DNA are denatured. The filter paper is then treated with protease K to remove protein and leave denatured DNA bound to the nitrocellulose, for which it has a high, affinity in the form of a DNA print of the colonies. The DNA is fixed firmly by baking at 80°C.

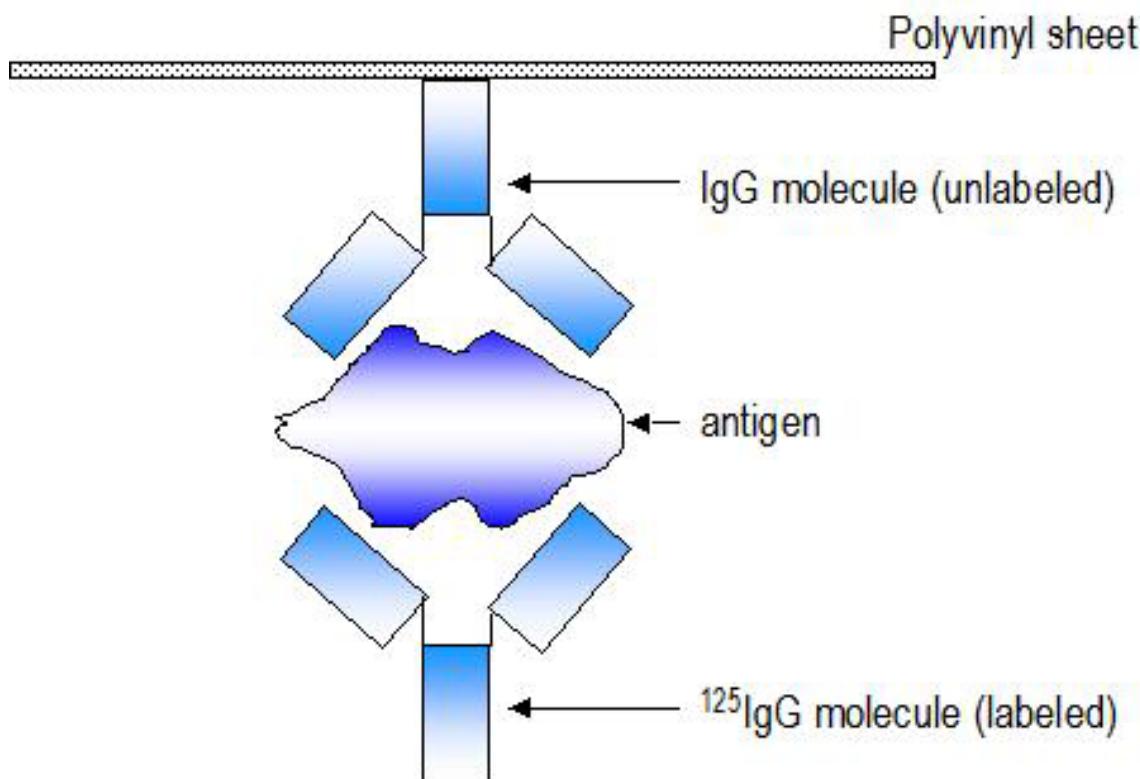


Selection of recombinant clones by colony hybridization

The labeled probe is hybridized to this DNA, and the result of this hybridization is monitored by autoradiography. A colony whose DNA print gives positive autoradiographic result can then be picked from the reference plate.

Immunological screening

Immunological screening involves the use of antibodies that specifically recognize antigenic determinants on the polypeptide synthesized by a target clone. This is one of the versatile expression cloning strategies, because it can be applied to any protein for which an antibody is available. Furthermore, the protein need not be functional.

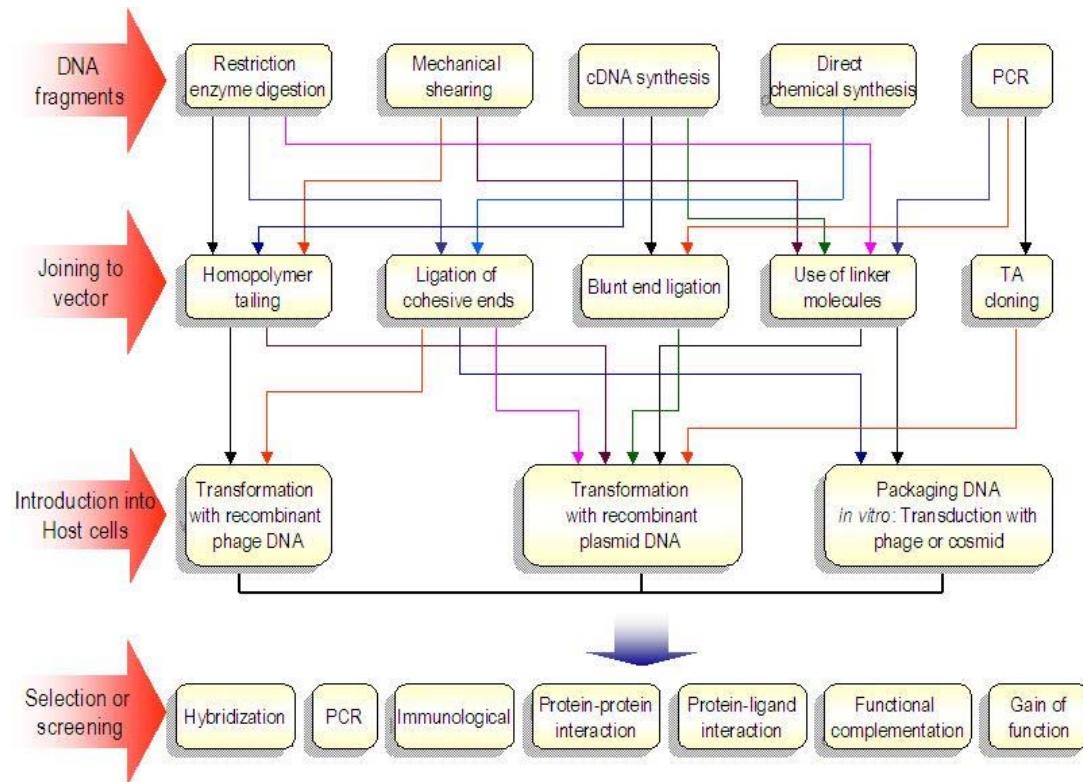


Antigen-antibody complex formation in the immunological screening

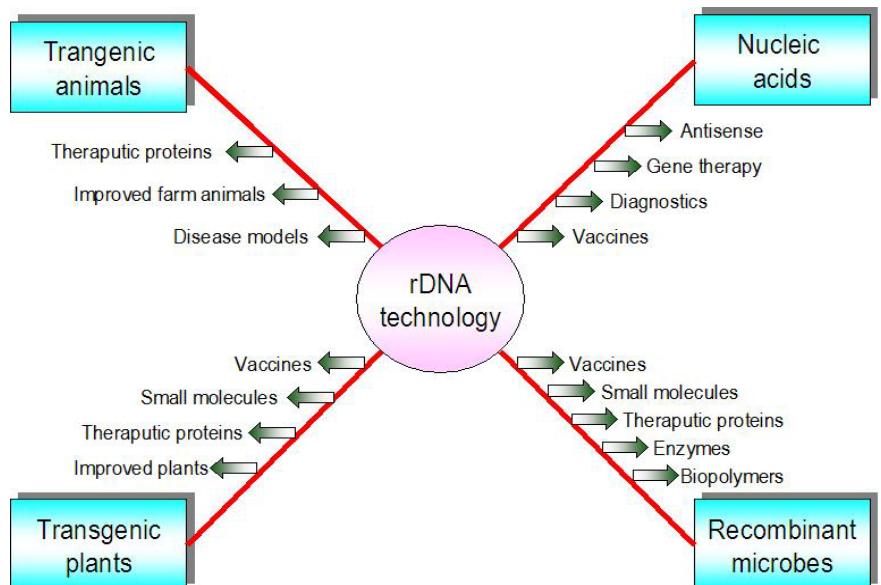
In the immunological screening method, first transformed cells are plated on agar in a conventional petridish. A replica plate must also be prepared because subsequent procedures kill these colonies. The bacterial colonies are then lysed in one of a number of ways. The lysis releases the antigens from positive colonies. A sheet of polyvinyl that has been coated with the appropriate antibody (unlabelled) is applied to the surface of the plate, whereupon the antigen complexes with bound antibody (IgG unlabelled). The sheet is removed and exposed to ¹²⁵I-labelled IgG. The ¹²⁵IgG can react with the bound antigen via antigenic determinants at sites other than those involved in the initial binding of the antigen to IgG coated sheet. Washing the sheet and making an autoradiographic image detect positively reacting colonies. The required clones can then be recovered from the replica plate.

The generalized overview of various cloning strategies and various applications of rDNA technology are schematically represented below.

Cloning strategies



Applications of rDNA technology



CHAPTER 14

APPLICATIONS OF rDNA TECHNOLOGY IN MEDICINE

One of the earliest commercial applications of gene manipulation techniques was the production in bacteria of proteins with therapeutic values. Not surprisingly, the first such products were recombinant versions of protein already used as the therapeutics: human growth hormone and insulin. Prior to the advent of genetic engineering human growth hormone was produced from pituitary glands removed from cadavers. Not only limit this supply of the hormone but, in some cases it resulted in recipients contracting Creutzfeld-Jacob syndrome. The recombinant approach resulted in unlimited supplies of safe material. This safety aspect was extended to various products. As the methods for cloning genes became more and more sophisticated, an increasing numbers of drugs, therapeutic agents and molecular diagnostics are introduced in medicine. A number of these were shown to have therapeutic potential and found their way into clinical medicine. Products of modern biotechnology include artificial blood vessels from collagen tubes coated with a layer of the anticoagulant heparin. Gene therapy – altering DNA within cells in an organism to treat or cure a disease – is one of the most promising areas of biotechnology research. New genetic therapies are being developed to treat diseases such as cystic fibrosis, AIDS and cancer. DNA fingerprinting is the process of cross matching two strands of DNA, a vital tool of molecular diagnostics. In criminal investigations, DNA from samples of hair, bodily fluids or skin at a crime scene are compared with those obtained from the suspects. In practice, it has become one of the most powerful and widely known applications of biotechnology today.

Some recombinant proteins that are used therapeutically

Year	Product	Clinical indication
1982	Human insulin	Diabetes
1985	Human growth hormone	Pituitary dwarfism
1986	Hepatitis B vaccine	Prevention of hepatitis B infection
1986	Interferon α_{2a}	Hairy cell leukemia
1987	Tissue plasminogen activator	Acute myocardial infarction
1990	Interferon γ_{1b}	Chronic granulomatus disease
1996	Interferon β_{1a}	Multiple sclerosis

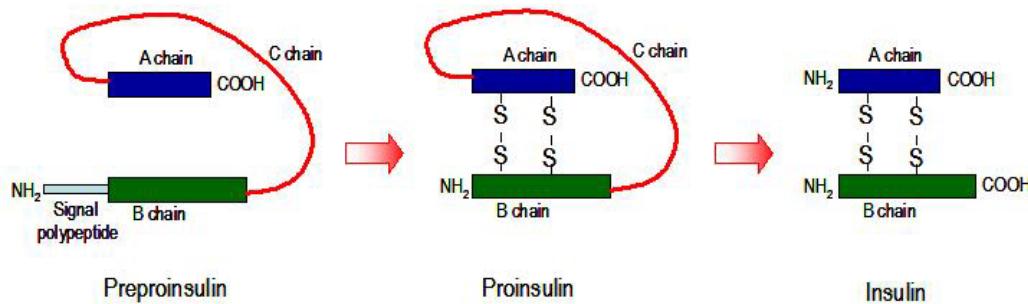
Pharmaceuticals

Insulin

Insulin, the synthesized by the β -cells of the islets of Langerhans in pancreas controls the level of glucose in the blood. Individuals with insulin deficiency manifest the condition as *diabetes mellitus*. These individuals require daily injections of insulin. Prior to 1982, this insulin was isolated from the pancreatic glands of cattle or pig obtained from meat packing plants. However, about five percent of the individuals with diabetes are allergic to these insulins, which were slightly different

from human insulin. In 1982, this problem was solved by producing the human insulin using rDNA technology.

Insulin is a small peptide hormone composed of two chains (A and B) joined by disulfide bonds. The A and B chains are 30 and 21 amino acids long respectively. These two chains are synthesized as part of a larger gene product called preproinsulin, which contains an amino-terminal signal sequence that is excised during its secretion from β -cells and a central region (C chain) that is subsequently removed by polypeptide cleavage. The primary secretion product, proinsulin is converted to insulin by the proteolytic excision of the C chain and the formation of three disulfide bonds.



Artificial synthesis of human insulin

Human insulin was initially produced in *E. coli* cells by recombinant DNA techniques in 1979. The DNA sequences encoding the A and B chains of human insulin were chemically synthesized and joined to the *E. coli* gene encoding β -galactosidase. The chimeric genes were inserted into plasmid pBR322 and introduced into *E. coli* cells by transformation. The fusion proteins specified by these chimeric genes contained a methionine residue between the bacterial β -galactosidase and the human insulin portions. This facilitated the removal of the insulin from the fusion protein by treatment with cyanogen bromide (which cleaves peptide chains on the carboxyl sides of methionine residues).

Human growth hormone

The second benefit to arise from rDNA technology was the availability of another polypeptide hormone, human growth hormone (HGH). This protein is used to treat adolescents suffering pituitary dwarfism to enable them to achieve a normal height. HGH was initially produced in *E. coli* cells harbouring a modified gene composed of the coding sequence for HGH fused to synthetic bacterial regulatory elements. This chimeric gene was constructed *in vitro* and introduced into *E. coli* by transformation. In 1985, HGH produced in *E. coli* became the second pharmaceuticals produced of genetic engineering.

Interferons

Interferons are proteinaceous substances produced by the body for defense against viral infections. They are also produced in response to many non-antigenic chemicals including polysaccharides, endotoxins, DNA, RNA etc. The interferons affect the cellular membranes thus, the viral particles are not disseminated to other cells. Moreover, interferons interfere with virus multiplication by inhibiting their protein synthesis.

The human interferon (IFN) proteins, which include the α , β and γ interferons (IFN- α , IFN- β and IFN- γ) are naturally occurring proteins; each one might have a therapeutic use. A number of different strategies have been used to isolate either the genes or cDNAs for human proteins. In the case of IFN proteins, a technique called cDNA enrichment technique was adopted which includes the following steps.

- Isolating mRNA from human leucocytes. The isolated mRNA were reverse transcribed and cloned to vectors.
- Identifying transformants and hybridization to crude interferon mRNA preparation.
- Isolation of IFN candidate mRNAs and translation in a cell free protein synthesis system
- Assaying for IFN antiviral activity

Whenever large quantities of the IFN are required, the IFN cDNAs can be subcloned into an *E. coli* expression vector and expressed at high levels.

After the isolation of the first IFN gene in the early 1980s, it was determined that there are a number of different IFNs. The IFN- α is encoded by a family of 15 non-allelic genes, whereas IFN- β and IFN- γ are each encoded by a single gene.

Interferon	Disorder
Interferon α_{2a}	Hepatitis C, Hairy cell leukemia
Interferon α_{2b}	Bladder cancer, Head and neck cancer
Interferon α_{n3}	AIDS, Cervical dysplasia
Interferon β_{1a}	Multiple sclerosis
Interferon β_{1b}	Chronic progressive multiple sclerosis
Interferon γ_{1b}	Renal cell carcinoma

Recently scientists developed technologies to synthesize hybrid IFNs, by combining gene sequences from different IFN genes. These hybrid IFNs' antiproliferative activity was found to be greater than that of the parental molecules. The creation of these hybrid IFNs demonstrates that new potential therapeutic molecules can be constructed by combining functional domains from related genes.

Vaccines

Vaccination protects a recipient from pathogenic agents by establishing an immunological resistance to infection. An injected or oral vaccine induces the host to generate antibodies against the disease causing organism; therefore, during future exposures, the infectious agent is inactivated (neutralized or killed), its proliferation is prevented, and the disease state is not established. Most of the modern vaccines typically consist of either a killed (inactivated) or a live, non-virulent (attenuated) form of an infectious agent. Within the last decade, rDNA technology has provided a means of creating a new generation of vaccines that overcome the drawbacks of traditional vaccines. The availability of gene cloning has enabled researchers to contemplate various novel strategies for vaccine development.

Subunit vaccines

Vaccines generally consist of either killed or attenuated forms of the whole pathogenic agent. The antibodies elicited by these vaccines initiate an immune response to inactivate (neutralize) pathogenic organisms by binding to proteins on the outer surface of the agent. For disease causing viruses, it has been shown that purified outer surface of the viral proteins, either capsid or envelope proteins are alone sufficient for eliciting neutralizing antibodies in the host organism. Vaccine that use components of a pathogenic organism rather than the whole organism are called “subunit” vaccines. The primary requirement for creating any subunit vaccine is identification of the components of the infectious agent that elicits antibodies that react against the intact form of the infectious agent.

Herpes Simplex Virus

Herpes simplex virus (HSV) has been implicated as a cancer causing agent (oncogenic)- in addition to its more common roles in causing sexually transmitted disease, severe eye infections, and encephalitis. The protection against HSV would be best achieved by a subunit vaccine, which would not be oncogenic. The HSV type 1 (HSV-1) envelope glycoprotein D (gD) is elicits antibodies that neutralize intact HSV. The HSV-1 gD gene was isolated and then cloned into a mammalian expression vector and expressed in Chinese hamster ovary (CHO) cells, which, unlike the E. coli system, allows foreign proteins to be properly glycosylated.

Foot and mouth disease virus

Foot and mouth disease virus (FMDV) has a devastating impact on cattle and swine and is extremely virulent. Formalin killed FMDV can be used as a vaccine to control foot and mouth disease. Research on FMDV found that the major antigenic determinant inducing neutralizing antibody is the capsid viral protein (VP1). Thus, the gene for VP1 became the target for cloning. The product of the VP1 coding sequence was identified immunologically as part of a fusion protein (the bacteriophage MS2 replicase protein + protein of FMDV VP1 protein).

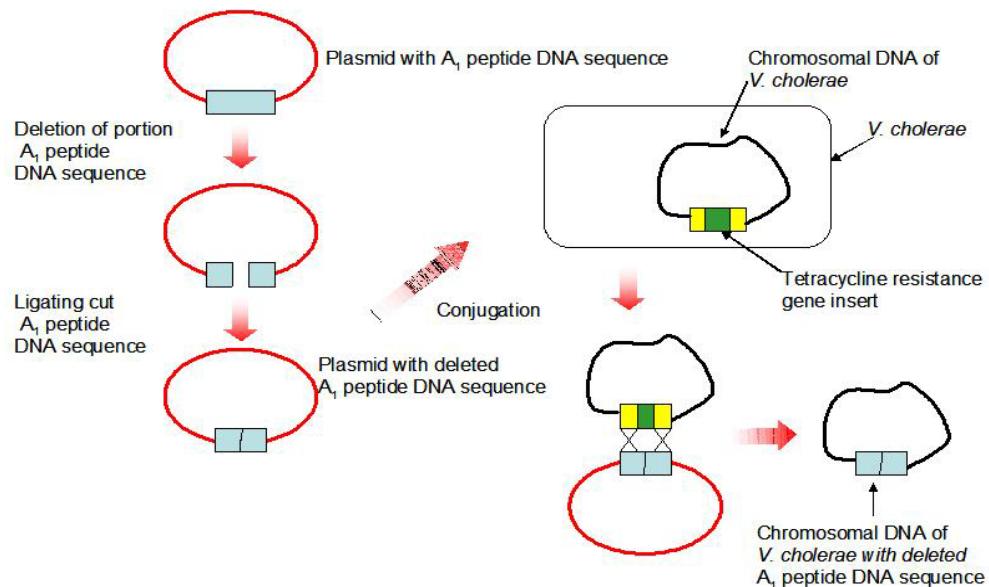
The analysis on the FMDV VP1 protein for its potential to produce revealed that the peptides corresponding to amino acids 141-160 elicited sufficient antibody to protect the animal against subsequent challenges with FMDV. Hepatitis B core protein (HBcAg) is used as a carrier molecule for the short synthetic peptides. This kind of vaccines are called as **peptide vaccines**.

Attenuated vaccines

In some instances, genetic manipulation may be used to construct modified organisms (bacteria or viruses) that are used as live recombinant vaccines. These vaccines are either nonpathogenic organisms that have been engineered to carry and express antigenic determinants from a target pathogenic agent or engineered strains of pathogenic organisms in which the virulence genes have been modified or deleted. The antigenic determinants from the engineered pathogens elicit a weak immunological response. If it is possible to develop a live vaccine, it is usually advantageous to do so, because live vaccines are generally much more effective than killed or subunit vaccines. The major requirement for alive vaccine is that no virulent forms be present in the inoculation material.

Cholera

The bacterium *Vibrio cholerae*, the causative agent of cholera, colonizes the small intestine and secretes large amounts of a hexameric enterotoxin, which is the actual pathogenic agent. Among the six subunits (one unit of A and five units of B). The A subunit has two functional domains: the A₁ peptide, which contains the toxic activity, and the A₂ peptide, which joins A subunit to the B subunits. Previous studies had indicated that a subunit vaccine consisting of inactivated hexameric cholera enterotoxin was not effective in providing immunity against *V. cholerae*. Since *V. cholerae* colonizes the surface of the intestinal mucosa, it is thought that an effective cholera vaccine should probably be directed to this structure and should therefore be administered orally.



With this in mind, a strain of *V. cholerae* in which part of the coding sequence for A₁ peptide was deleted. The strain with disrupted A₁ peptide DNA sequence cannot produce enterotoxin; therefore, it is a good candidate for a live vaccine.

Vector vaccines

The important features of a vector vaccine are the delivery and expression of cloned genes encoding antigens that elicit neutralizing antibodies against the pathogenic agents. Vaccinia virus, a member of poxvirus family, can be a representative for developing vector vaccines. Vaccinia virus has wide host range (invertebrates to vertebrates including humans). However, the vaccinia virus genome is very large and lacks unique restriction sites. Therefore, it is not possible to insert additional DNA directly into the viral genome. Of necessity, the genes for specific antigens must be introduced into the viral genome by *in vivo* homologous recombination.

Some of the human disease agents for which recombinant vaccines are currently available are given below.

Pathogenic agent	Disease
Dengue virus	Hemorrhagic fever
Hepatitis A	Liver damage
Hepatitis B	Liver damage
Herpes simplex virus type2	Genital ulcers
Influenza A and B viruses	Acute respiratory disease
Rabies virus	Encephalitis
<i>Vibrio cholerae</i>	Cholera
<i>Haemophilus influenzae</i>	Meningitis
<i>Mycobacterium leprae</i>	Leprosy
<i>Bordetella pertussis</i>	Whooping cough
<i>Clostridium tetani</i>	Tetanus
<i>Mycobacterium tuberculosis</i>	Tuberculosis
<i>Salmonella typhi</i>	Typhoid fever
<i>Plasmodium</i> spp	Malaria
<i>Wuchereria bancrofti</i>	Filariasis

Antibiotics

Since the discovery of penicillin in the late 1920s, more than 6,000 antibiotics with different specificities and a variety of modes of action have been isolated from various microorganisms. The universal use of antibiotics to treat bacterial diseases has resulted in an enormous improvement in human health and has undoubtedly saved millions of lives. The vast majority of the most important antibiotics have been isolated from the gram-positive soil bacterium *Streptomyces*, although fungi and other gram-positive and gram-negative bacteria are also sources of antibiotics.

The biosynthesis of an antibiotic may include 10 to 30 separate enzyme catalyzed steps, so that cloning all the genes for the synthesis of a particular antibiotic is not an easy task. One strategy for isolating the complete set of antibiotic biosynthesis genes consists of transforming one or more mutant strains that are unable to synthesize the antibiotic with DNA from a clone bank constructed from wild type chromosomal DNA. Following the introduction of clone bank into mutant cells, transformants are screened for their ability to produce antibiotic. Then the plasmid DNA from the clone that supplies a functional gene and gene product, i.e provides the missing function of (complements) a mutant strain, is used as a DNA hybridization probe to screen the another clone bank of wild type chromosomal DNA and isolate clones with regions that overlap the probe sequence. All these overlapping sequences are put together to identify genes responsible for antibiotic biosynthesis.

Some clinically important antibiotics

Antibiotic	Producer organism	Activity	Mode of action
Penicillin	<i>Penicillium chrysogenum</i>	Gram-positive bacteria	Wall synthesis
Cephalosporin	<i>Cephalosporium acremonium</i>	Broad spectrum	Wall synthesis
Griseofulvin	<i>Penicillium griseofulvum</i>	Dermatophytic fungi	Microtubules
Bacitracin	<i>Bacillus subtilis</i>	Gram-positive bacteria	Wall synthesis
Polymyxin B	<i>Bacillus polymyxa</i>	Gram-negative bacteria	Cell membrane
Amphotericin B	<i>Streptomyces nodosus</i>	Fungi	Cell membrane
Erythromycin	<i>Streptomyces erythreus</i>	Gram-positive bacteria	Protein synthesis
Neomycin	<i>Streptomyces fradiae</i>	Broad spectrum	Protein synthesis
Streptomycin	<i>Streptomyces griseus</i>	Gram-negative bacteria	Protein synthesis
Tetracycline	<i>Streptomyces rimosus</i>	Broad spectrum	Protein synthesis
Vancomycin	<i>Streptomyces orientalis</i>	Gram-positive bacteria	Protein synthesis
Gentamicin	<i>Micromonospora purpurea</i>	Broad spectrum	Protein synthesis
Rifamycin	<i>Streptomyces mediterranei</i>	Tuberculosis	Protein synthesis

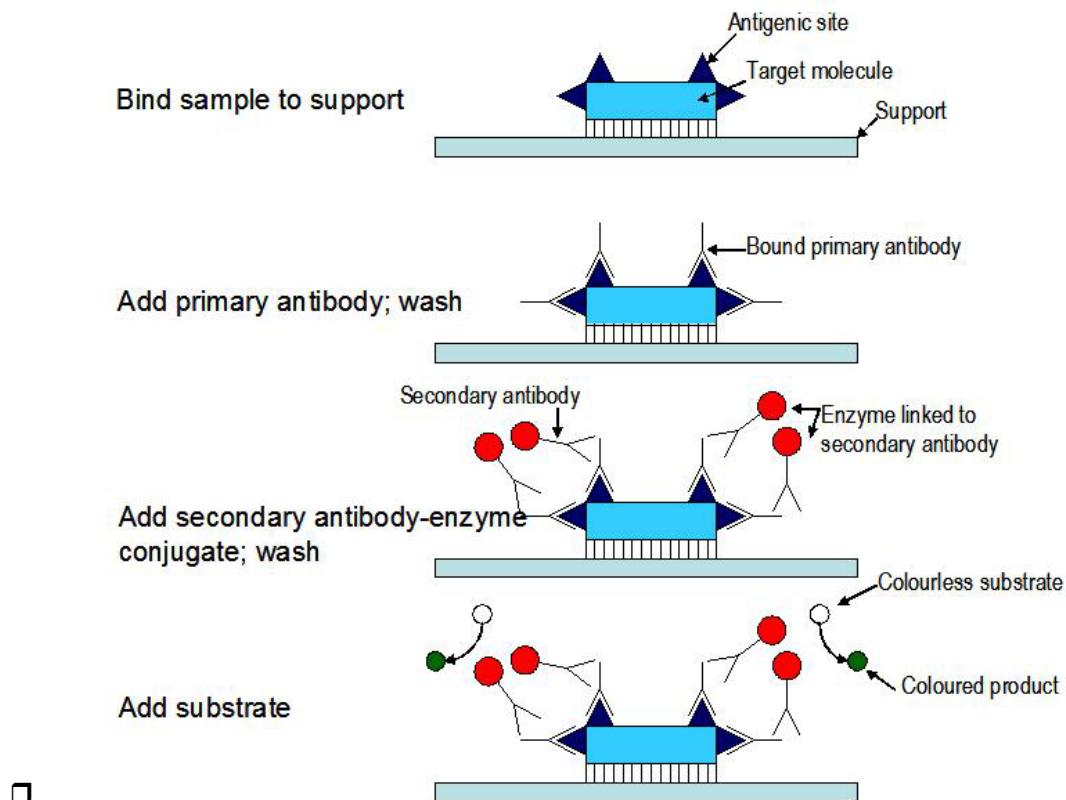
Molecular diagnostics

The successes of modern medicine and agriculture often depends on the ability of the workers in these fields to detect the presence of specific viruses, bacteria, fungi, parasites, proteins and small molecules in humans, animals, plants, water and soil. For example, the prevention, control or treatment of infectious disease is generally facilitated by the early and accurate identification of the causative pathogenic organism. In general, any useful detection strategy must be specific, sensitive and simple. Specificity means that the assay must yield a positive response for the only target organism or molecule. Sensitivity means that the diagnostic test must identify very small amounts of the target organism or molecule, even in the presence of other potentially interfering organisms or substances. Simplicity is required for the test to be run efficiently, effectively and inexpensively on a routine basis. Several diagnostic tools have developed involving protein and DNA molecules. The protein based diagnostics involve immunological response from the organisms and the presence of specific organism or molecule is based solely on identifying specific antibody-antigen complexes.

Enzyme Linked Immunosorbent Assay (ELISA)

There are number of different ways to determine whether an antibody has bound to its target antigen. The ELISA is one method, and it is frequently used for diagnostic detection. A generalized ELISA protocol has the following steps:

- Bind the sample being tested for the presence of a specific molecule or organism to a solid support
- Add a marker specific antibody (primary antibody) to the bound material, and then wash the support to remove rebound primary antibody
- Add a second antibody (secondary antibody) that bind specifically to the primary antibody and not to the target molecule. Bound to the secondary antibody is an enzyme such as alkaline phosphatase, peroxidase or urease, which can catalyze a reaction that converts a colourless substrate into a coloured product. Wash the mixture to remove any unbound secondary antibody-enzyme conjugate.
- Add the colourless substrate; observe or measure the amount of coloured product.

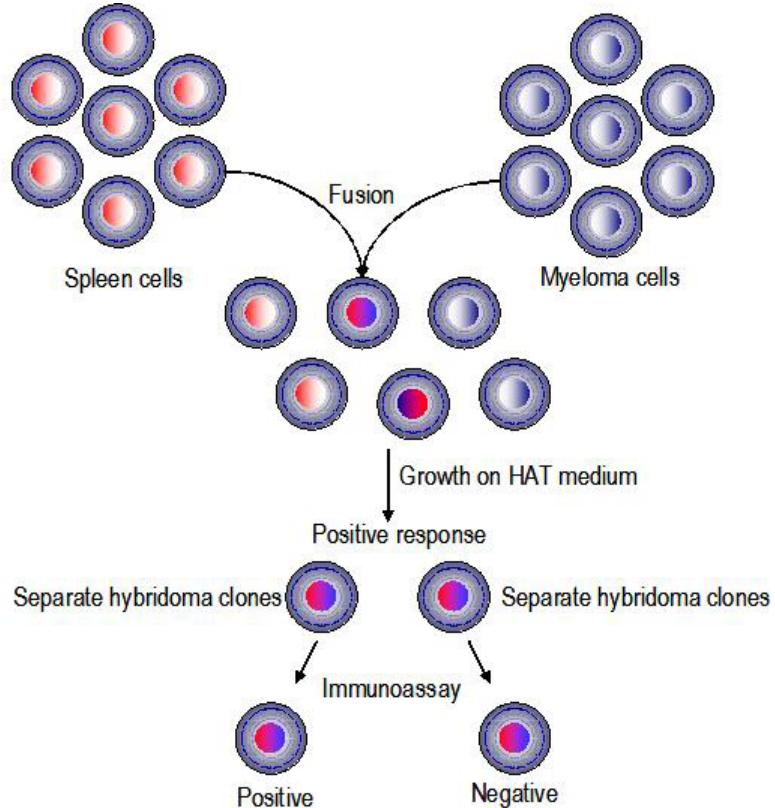


The principal feature of an ELISA system is the specific binding of the primary antibody to the target site. If the target molecule is, for example, a protein, then a purified preparation of this protein is generally used to generate antibodies that will be used to detect the target. If there are several antibodies that detect several antigenic determinants, then the antibodies are called **polyclonal antibodies**. If there is a single antibody for specific antigen determinant, then it is known as **monoclonal antibody**.

Hybridoma technology

The fundamental objective for the applied use of antibodies was to discover how to create a cell line that could be grown in culture and with a high affinity for specific target antigen. Such a cell line would provide a consistent and continual source of identical antibody molecules. Unfortunately, the B lymphocytes (B cells) which synthesize antibodies do not grow in culture. However, it was envisioned that a hybrid cell type could be created to solve this problem. This hybrid cell type would have the B-cell genetic components for producing antibodies and the cell division functions of a compatible cell types to enable the cells to grow in culture. It was known that normal B lymphocytes sometimes become cancer cells (myelomas) that acquire the ability to grow in culture while retaining many of the attributes of B cells. Thus, myeloma cells, especially those that did not produce antibody molecules, became candidates for fusion with antibody producing B cells. In the mid 1970s, these ideas became reality as **Hybridoma technology**.

The spleenic cell suspension of B cells is mixed with a suspension of myeloma cells that are genetically defective for the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT⁻). The combined cell suspensions are mixed with 35% polyethylene glycol (PEG) for a few minutes and then transferred to a growth medium containing hypoxanthine, aminopterin and thymidine (HAT medium).



Unfused spleen cells and spleen-spleen fusion cells	Cannot grow in any culture medium
Unfused myeloma cells and myeloma-myeloma fusion cells	Cannot grow in the HAT medium; cannot use hypoxanthine as a precursor for the biosynthesis of guanine and adenine; aminopterin in the medium will not allow the myeloma cells to use dihydrofolate reductase, an enzyme that is responsible for purine biosynthesis
Spleen-myeloma fusion cells	Survive in HAT medium because the spleen cells contribute a functional HGPRT

The PEG treatment facilitates fusion between cells resulting in myeloma-spleen fusion cells, myeloma-myeloma fusion cells and spleen-spleen fusion cells with spleen cells and myeloma cells. The HAT medium, however, allows only the myeloma-spleen fusion cells to grow, because none of the other cell types can proliferate in this medium. After the clones have been cultured, their media are tested again to determine which cell lines (hybridomas) produce monoclonal antibody molecules that recognize the target antigen. If more than one specific hybridoma is isolated, further tests are conducted to determine whether the different clones produce antibody against the same antigenic determinant. The hybridoma technology helps to increase the specificity of an ELISA protocol considerably since specific monoclonal antibodies can be obtained from this technique.

Transgenic animals and plants as bioreactors

Pharming is the play on words that refers to the use of transgenic animals and plants to produce recombinant therapeutic proteins. The production of growth hormone in the serum of transgenic mice provided the first evidence that recombinant proteins could be produced, continuously, in the body fluids of animals. Five years later, several groups reported the secretion of recombinant proteins in mouse milk. The first proteins produced in this way were sheep β -lactoglobulin and human tissue-plasminogen activator (tPA). Extremely high levels of (30g/l) of human α_1 - antitrypsin (AAT) was produced in the milk of transgenic ewe. There have over 100 reports since these early experiments. Using similar protocols, tPA was produced in the milk of transgenic goat.

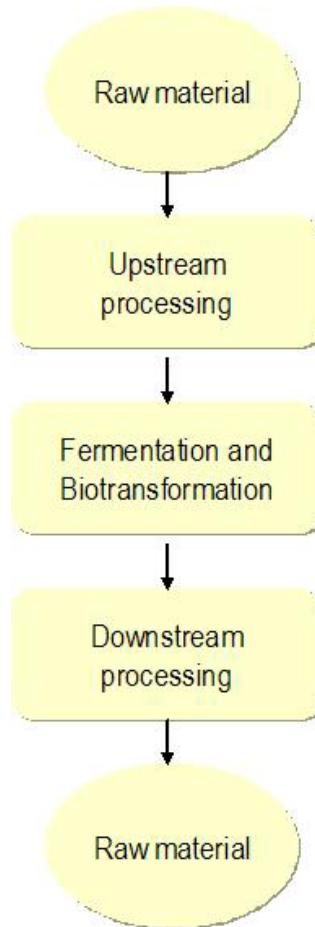
Plants are a useful alternative to animals for recombinant protein production because they are inexpensive to grow and scale-up from laboratory testing to commercial production is easy. The first report of human therapeutic agent in plants was the expression of human growth hormone, as a fusion with *Agrobacterium* nopaline synthase enzyme, in transgenic tobacco. Antibody production in plants was first demonstrated in tobacco by producing immunoglobulins. For example, a fully humanized antibody against herpes simplex virus-2 (HSV-2) has been expressed in soybean. Plants producing recombinant slgA against the oral pathogen *Streptococcus mutans* have been generated, and these plant derived antibodies (plantibodies) have been recently commercialized.

CHAPTER 15

INDUSTRIAL APPLICATIONS OF rDNA TECHNOLOGY

Industrial biotechnology applies the techniques of modern molecular biology to improve the efficiency and reduce the environmental impacts of industrial processes like textile, paper and pulp, and chemical manufacturing. All these processes use microorganisms for producing commercial products and typically have three key stages.

1. Upstream processing: preparation of a raw material so that it can be used as food source for the target microorganism.
2. Fermentation and transformation: growth (fermentation) of the target microorganism in a large bioreactor with the consequent production (biotransformation) of a desired compound, which can be for example, an antibiotic, an amino acid, or a protein.
3. Downstream processing: purification of the desired compound from either the cell medium or cell mass.



Biotechnology research is dedicated to maximizing the overall efficiency of each of these steps and to finding microorganisms that make products that are useful as foods, food supplements and drugs. During the 1960s and 1970s, this research focused on upstream processing, bioreactor design and downstream processing. These studies led to enhanced bioinstrumentation for monitoring and controlling the fermentation process and to efficient large-scale growth facilities that increased the yields of various products.

The biotransformation component of the overall process was the most difficult phase to optimize. Commodity production by naturally occurring microbial strains on a large scale was often considerably less than optimum. The traditional genetic improvement regimens were tedious, time consuming and costly because of the large numbers of colonies that had to be selected, screened and tested. The nature of industrial biotechnology was changed forever by the development of rDNA technology. With these techniques, the optimization of biotransformation phase of industrial biotechnology process was achieved more directly.

Large scale production of rDNA products

The production of commercial products that are synthesized by genetically engineered microorganisms requires the partnership of two kinds of experts. Molecular biologists are responsible for isolating, characterizing, modifying and creating effectively expressed genes in microorganisms that can be used for industrial production, and biochemical engineers ensure that the genetically engineered forms of a microorganism can be grown in large quantities under conditions that give optimal product yields. The large scale multiplication of microorganisms is carried out by the process called **fermentation** in specially designed structures called **bioreactors**.

Fermentation

Microorganisms can be grown in batch, fed-batch or continuous culture. In **batch fermentation**, the sterile growth medium is inoculated with the appropriate microorganisms and the fermentation proceeds without the addition of fresh growth medium. In **fed-batch fermentation**, nutrients are added incrementally at various times during the fermentation reaction; no growth medium is removed until the end of the process. In the **continuous fermentation** process, fresh growth medium is added continuously during the fermentation but there is concomitant removal of an equal volume of spent medium containing suspended microorganisms. For each types of fermentation, oxygen (which is usually provided in the form of air), an antifoaming agent, and, if required, acid or base are injected into the bioreactor as needed. Though the batch fermentation is a commonly used process for growing microorganisms, fed-batch fermentation is specifically used for producing human therapeutic proteins. Continuous fermentation is used for the commercial production of single cell protein, antibiotics and organic solvents.

Bioreactors

Bioreactors are specialized growth chambers (fermenters, fermentation vessels) where microorganisms are grown for harvesting specific products. There is limitless number of bioreactor designs. However, all the designs fall into three fundamental classes:

- Stirred tank reactors (STR), which have internal mechanical agitation.
- Bubble columns, which rely on the introduction of air or another gas (sparging) for agitation.
- Airlift reactors, which have either an internal or an external loop; mixing and circulation of the culture fluid in these reactors are the result of the motion of an introduced gas (usually air), which caused density differences within the different parts of the bioreactor.

The traditional and by far the most commonly used bioreactor is the STR. This type of bioreactor has several advantages over the other types: it has highly flexible operating conditions; it is readily available commercially; it provides efficient gas transfer to the growing microbial cells.

After the fermentation process, when the product is present in the cells, the cells can be harvested by either centrifugation or cross-flow filtration and lysed chemically or mechanically. The product is then fractionated from the cell lysate. Various devices and processes have been created to ensure that these steps are conducted effectively. When the product is in the culture medium, the concentration and purification procedures are simpler.

Biomass conversion

Biomass is the name given to any recent organic matter that has been derived from plants as a result of the photosynthetic conversion process. Biomass energy is derived from plant and animal material, such as wood from forests, residues from agricultural and forestry processes, and industrial, human or animal wastes.

The energy value of biomass from plant matter originally comes from solar energy through the process known as photosynthesis. The chemical energy that is stored in plants and animals (that eat plants or other animals), or in the wastes that they produce, is called bioenergy. During conversion processes such as combustion (burning), biomass releases its energy, often in the form of heat, and the carbon is reoxidised to carbon dioxide to replace that which was absorbed while the plant was growing. Essentially, the use of biomass for energy is the reversal of photosynthesis.

In nature, all biomass ultimately decomposes to its elementary molecules with the release of heat. Therefore the release of energy from the conversion of biomass into useful energy imitates natural processes but at a faster rate. Therefore, the energy obtained from biomass is a form of renewable energy. Utilising this energy recycles the carbon and does not add carbon dioxide to the environment, in contrast to fossil fuels. Of all the renewable sources of energy, biomass is unique in that it is effectively stored solar energy. Furthermore, it is the only renewable source of carbon, and is able to be processed into convenient solid, liquid and gaseous fuels.

Biomass can be used directly (e.g. burning wood for heating and cooking) or indirectly by converting it into a liquid or gaseous fuel (e.g. ethanol from sugar crops or biogas from animal waste). The net energy available in the biomass when it is combusted ranges from about 8MJ/kg for green wood, to 20MJ/kg for oven dry plant matter, to 55MJ/kg for methane; compared with about 23 - 30MJ/kg for coal. The efficiency of the conversion process determines how much of the actual energy can be practically utilised.

Biomass Use

Biomass is a renewable resource that includes waste from trees, forest residues, food processing, animals, and organic industrial and municipal sources. The conversion of these waste products begins with a process called hydrolysis; sugars are first produced from the biomass and then converted into environmentally safe fuel and other chemical products.

Modern biomass now represents only three percent of primary energy consumption in industrialised countries. However, much of the rural population in developing countries, which represents about 50 percent of the world's population, is still reliant on traditional biomass, mainly in the form of wood, for fuel. Traditional biomass accounts for 35 percent of primary energy consumption in developing countries, raising the world total to 14 percent of primary energy consumption.

The Earth's natural biomass replacement represents an energy supply of around 3,000EJ (3×10^{21} J) a year, of which just under two percent is currently used as fuel. It is not possible, however, to use all of the annual production of biomass in a sustainable manner. One analysis carried out by the United Nations Conference on Environment and Development (UNCED) estimates that biomass could potentially supply about half of the present world primary energy consumption by the year 2050.

In the future, biomass has the potential to provide a cost-effective and sustainable supply of energy, while at the same time aiding countries to meet their greenhouse gas reduction targets. By the year 2050, it is estimated that 90 percent of the world population will live in developing countries. It is critical therefore that the biomass processes used in these countries are sustainable. The modernization of biomass technologies, leading to more efficient biomass production and conversion, is one possible direction for biomass use in developing countries.

In industrialized countries, the main biomass processes utilized in the future are expected to be the direct combustion of residues and wastes for electricity generation, bio-ethanol and biodiesel as liquid fuels, and combined heat and power production from energy crops. In the short to medium term, biomass waste and residues are expected to dominate biomass supply, to be substituted by energy crops in the longer term. The future of biomass electricity generation lies in biomass integrated gasification/gas turbine technology, which offers high-energy conversion efficiencies and will be further developed to run on biomass-produced fuels.

The production of biofuels such as ethanol and biodiesel has the potential to replace significant quantities of fossil fuels in many transport applications. The widespread use of ethanol in Brazil has shown that biofuels are technically feasible on a large scale. In the USA and Europe biofuel production (ethanol and biodiesel) is increasing, with most of the products being marketed in fuel blends, e.g. E20 is 20 percent ethanol and 80 percent petrol and has been found to be suitable for most spark ignition engines without any modifications. At present this production is supported by government incentives, but in the future, with the increased growth of energy crops, and economies of scale, cost reductions may make biofuels competitive in their own right.

Ethanol production

Historically, production of ethanol was limited to using sources of sugar that were available in soluble forms, such as sugar (sucrose), molasses from sugar cane, or fructose from the corn plant. Since these soluble sugars are edible (i.e. suitable for human consumption), their relative value tends to be higher than for the rest of the plant (leaves, stalks, etc.) which is inedible and usually has a much lower value. In many cases, the inedible portions of the plants are considered to be waste materials.

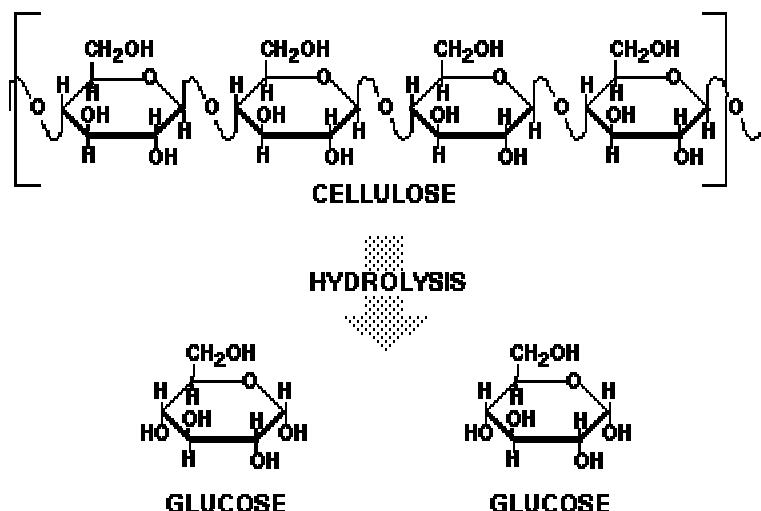
However, new technologies have been developed which now allow for the production of ethanol from "lignocellulosic biomass." Lignocellulosic biomass is the leafy or woody part of plants: corn stover, bagasse, yard and wood waste, paper pulp, etc. This is very significant: for example, where one acre of sugarcane produces about ten tons of edible sugar and three tons of molasses, it also produces (in the form of leaves and stalks) an additional twenty to twenty-five tons of non-edible materials. Lignocellulosic biomass also refers to energy grasses or tree crops. Starting with a list of twenty crops and sources of biomass, a short list of feedstocks was developed. The most promising crops were sugarcane, leucaena, eucalyptus, napier grass, and sweet sorghum. Waste paper and organic waste were also identified as potentially promising feedstocks for ethanol production.

The primary components of most plant material are commonly described as lignocellulosic biomass. The biomass is principally composed of the compounds cellulose, hemicellulose, and lignin. Cellulose, a primary component of most plant cell walls, is made up of long chains of the 6-carbon sugar, glucose, arranged in bundles. (Often described as crystalline bundles). Cellulose is a primary component of paper. In the plant cell wall, the cellulose molecules are interlinked by another molecule, hemicellulose. The hemicellulose is primarily composed of the 5-carbon sugar, xylose. Another molecule called lignin is also present in significant amounts and gives the plant its structural strength. Improvements in technology have recently provided a variety of methods of extracting and dissolving the cellulose and hemicellulose to produce the component sugars in a form that can be converted to ethanol. Appropriate pre-treatment can free the cellulose and hemicellulose from the plant material. Further treatment using chemicals, enzymes or microorganisms can be used to liberate simple sugars from the cellulose and hemicellulose making them available to microorganisms for fermentation to ethanol.

Steps involved in ethanol production

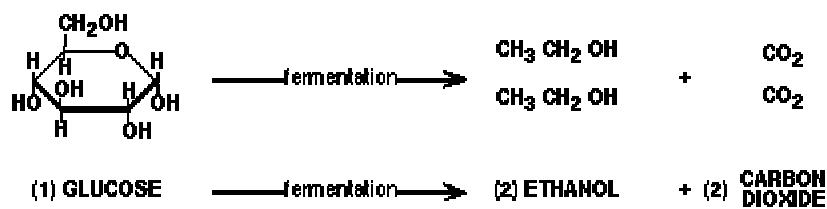
Hydrolysis

The first step involves hydrolysis: splitting the bonds in the cellulose to produce the sugar glucose.



Glucose fermentation

Once the large molecules are extracted they can be broken down into their component sugars using enzymes or acids. The sugars then can be converted to ethanol using appropriately selected microorganisms in a process called fermentation. The formation of ethanol from 6 carbon sugars is illustrated in Figure II-3.



One molecule of glucose produces 2 molecules of ethanol and 2 molecules of carbon dioxide. An examination of the molecular weights of the molecules reveals that the weight of ethanol produced is equal to about half the weight of the starting material (glucose).

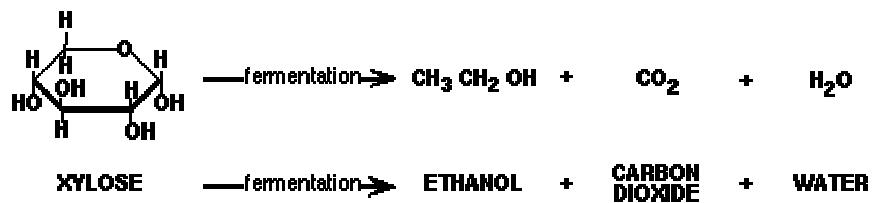
Glucose	$\text{C}_6\text{H}_{12}\text{O}_6$	Molecular Weight = 180
Ethanol	$\text{C}_2\text{H}_5\text{OH}$	Molecular Weight = $46 \times 2 = 92$
Carbon Dioxide	CO_2	Molecular Weight = $44 \times 2 = 88$

The maximum weight (%) of ethanol from the process would be $92/180 = 51\%$. Almost half the weight of the glucose $88/180$ (49%) is converted to carbon dioxide. Hemicellulose is made up

of the 5 carbon sugar xylose arranged in chains with other minor 5 carbon sugars interspersed as side chains. Just as with cellulose, the hemicellulose can be extracted from the plant material and treated to release xylose which, in turn, can be fermented to produce ethanol. As reviewed by Roberts and Hilton, xylose fermentation is not straight forward. Depending on the microorganism and conditions, a number of fermentations are possible. The array of products can include ethanol, carbon dioxide, and water.

Xylose fermentation

Actually, three conversions have been documented with yields of ethanol ranging from 30 to 50 percent of the weight of the starting material (weight ethanol/weight xylose).



However, laboratory results have indicated a wide range of variation. In the discussion of potential yields of ethanol from various materials, a range of hemicellulose-to-xylose conversion efficiencies and a range of xylose- to-ethanol conversion efficiencies have been combined to provide an assumed overall conversion efficiency of hemicellulose to ethanol of about 50 percent.

Recently, genetically modified *E. coli*, *Erwinia* or *Klebsiella* microorganisms with a set of genes viz. alcohol dehydrogenase II, and pyruvate decarboxylase genes from *Zymomonas mobilis* are used to enhance ethanol production by diverting pyruvate to ethanol during growth under either aerobic or anaerobic conditions. This allows lactose, glucose, xylose, arabanose, galactose and mannose to be converted to ethanol without producing organic acids.

CHAPTER 16

APPLICATIONS OF rDNA TECHNOLOGY IN ENVIRONMENTAL SCIENCE

The quality of life on Earth is linked inextricably to the overall quality of the environment. In early times, it is believed that we had an unlimited abundance of land and resources; today, however, the resources in the world show, in greater or lesser degree, our carelessness and negligence in using them. The problems associated with contaminated sites now assume increasing prominence in many countries. Contaminated lands generally result from past industrial activities when awareness of the health and environmental effects connected with the production, use, and disposal of hazardous substances were less well recognized than today. The problem is worldwide, and the estimated number of contaminated sites is significant. It is now widely recognized that contaminated land is a potential threat to human health, and its continual discovery over recent years has led to international efforts to remedy many of these sites, either as a response to the risk of adverse health or environmental effects caused by contamination or to enable the site to be redeveloped for use. The conventional techniques used for remediation have been to dig up contaminated soil and remove it to a landfill, or to cap and contain the contaminated areas of a site. The methods have some drawbacks. The first method simply moves the contamination elsewhere and may create significant risks in the excavation, handling, and transport of hazardous material. Additionally, it is very difficult and increasingly expensive to find new landfill sites for the final disposal of the material. The cap and contain method is only an interim solution since the contamination remains on site, requiring monitoring and maintenance of the isolation barriers long into the future, with all the associated costs and potential liability. A better approach than these traditional methods is to completely destroy the pollutants if possible, or at least to transform them to innocuous substances. Some technologies that have been used are high-temperature incineration and various types of chemical decomposition (e.g., base-catalyzed dechlorination, UV oxidation). They can be very effective at reducing levels of a range of contaminants, but have several drawbacks, principally their technological complexity, the cost for small-scale application, and the lack of public acceptance, especially for incineration that may increase the exposure to contaminants for both the workers at the site and nearby residents.

Environmental biotechnology is not a new field; composting and wastewater treatments are familiar examples of old environmental biotechnologies. However, recent studies in molecular biology and ecology offer opportunities for more efficient biological processes. Notable accomplishments of these studies include the clean-up of polluted water and land areas.

Bioremediation

Bioremediation is an option that offers the possibility to destroy or render harmless various contaminants using natural biological activity. As such, it uses relatively low-cost, low-technology techniques, which generally have a high public acceptance and can often be carried out on site. It will not always be suitable, however, as the range of contaminants on which it is effective is limited, the time scales involved are relatively long, and the residual contaminant levels achievable may not

always be appropriate. Although the methodologies employed are not technically complex, considerable experience and expertise may be required to design and implement a successful bioremediation program, due to the need to thoroughly assess a site for suitability and to optimize conditions to achieve a satisfactory result. Because bioremediation seems to be a good alternative to conventional clean-up technologies research in this field, especially in the United States, rapidly increasing. Bioremediation has been used at a number of sites worldwide, including Europe, with varying degrees of success. Techniques are improving as greater knowledge and experience are gained, and there is no doubt that bioremediation has great potential for dealing with certain types of site contamination.

Principles of bioremediation

Bioremediation is defined as the process whereby organic wastes are biologically degraded under controlled conditions to an innocuous state, or to levels below concentration limits established by regulatory authorities. By definition, bioremediation is the use of living organisms, primarily microorganisms, to degrade the environmental contaminants into less toxic forms. It uses naturally occurring bacteria and fungi or plants to degrade or detoxify substances hazardous to human health and/or the environment. The microorganisms may be indigenous to a contaminated area or they may be isolated from elsewhere and brought to the contaminated site. Contaminant compounds are transformed by living organisms through reactions that take place as a part of their metabolic processes. **Biodegradation** of a compound is often a result of the actions of multiple organisms. When microorganisms are imported to a contaminated site to enhance degradation we have a process known as **bioaugmentation**.

For bioremediation to be effective, microorganisms must enzymatically attack the pollutants and convert them to harmless products. As bioremediation can be effective only where environmental conditions permit microbial growth and activity, its application often involves the manipulation of environmental parameters to allow microbial growth and degradation to proceed at a faster rate. Like other technologies, bioremediation has its limitations. Some contaminants, such as chlorinated organic or high aromatic hydrocarbons, are resistant to microbial attack. They are degraded either slowly or not at all, hence it is not easy to predict the rates of clean-up for a bioremediation exercise; there are no rules to predict if a contaminant can be degraded. Bioremediation techniques are typically more economical than traditional methods such as incineration, and some pollutants can be treated on site, thus reducing exposure risks for clean-up personnel, or potentially wider exposure as a result of transportation accidents. Since bioremediation is based on natural attenuation the public considers it more acceptable than other technologies. Most bioremediation systems are run under aerobic conditions, but running a system under anaerobic conditions may permit microbial organisms to degrade otherwise recalcitrant molecules. See Table 1 for a list of contaminants potentially suitable for bioremediation.

Microbial populations for bioremediation processes

Microorganisms can be isolated from almost any environmental conditions. Microbes will adapt and grow at subzero temperatures, as well as extreme heat, desert conditions, in water, with

an excess of oxygen, and in anaerobic conditions, with the presence of hazardous compounds or on any waste stream. The main requirements are an energy source and a carbon source. Because of the adaptability of microbes and other biological systems, these can be used to degrade or remediate environmental hazards. These microorganisms can be divided into the following groups:

Aerobic: In the presence of oxygen. Examples of aerobic bacteria recognized for their degradative abilities are *Pseudomonas*, *Alcaligenes*, *Sphingomonas*, *Rhodococcus*, and *Mycobacterium*. These microbes have often been reported to degrade pesticides and hydrocarbons, both alkanes and polycyclic aromatic compounds. Many of these bacteria use the contaminant as the sole source of carbon and energy.

Anaerobic: In the absence of oxygen. Anaerobic bacteria are not as frequently used as aerobic bacteria. There is an increasing interest in anaerobic bacteria used for bioremediation of polychlorinated biphenyls (PCBs) in river sediments, dechlorination of the solvent trichloroethylene (TCE), and chloroform.

Ligninolytic fungi: Fungi such as the white rot fungus *Phanaerochaete chrysosporium* have the ability to degrade an extremely diverse range of persistent or toxic environmental pollutants. Common substrates used include straw, saw dust, or corn cobs.

Methylotrophs: Aerobic bacteria that grow utilizing methane for carbon and energy. The initial enzyme in the pathway for aerobic degradation, methane monooxygenase, has a broad substrate range and is active against a wide range of compounds, including the chlorinated aliphatics trichloroethylene and 1,2-dichloroethane.

Bioremediation strategies

Different techniques are employed depending on the degree of saturation and aeration of an area. *In situ* techniques are defined as those that are applied to soil and groundwater at the site with minimal disturbance. *Ex situ* techniques are those that are applied to soil and groundwater at the site which has been removed from the site via excavation (soil) or pumping (water). *Bioaugmentation* techniques involve the addition of microorganisms with the ability to degrade pollutants.

In situ bioremediation

These techniques are generally the most desirable options due to lower cost and less disturbance since they provide the treatment in place avoiding excavation and transport of contaminants. *In situ* treatment is limited by the depth of the soil that can be effectively treated. In many soils effective oxygen diffusion for desirable rates of bioremediation extend to a range of only a few centimeters to about 30 cm into the soil, although depths of 60 cm and greater have been effectively treated in some cases. The most important land treatments are:

Bioventing: Bioventing is the most common *in situ* treatment and involves supplying air and nutrients through wells to contaminated soil to stimulate the indigenous bacteria. Bioventing employs low air flow rates and provides only the amount of oxygen necessary for the biodegradation while minimizing volatilization and release of contaminants to the atmosphere. It works for simple hydrocarbons and can be used where the contamination is deep under the surface.

In situ biodegradation: *In situ* biodegradation involves supplying oxygen and nutrients by circulating aqueous solutions through contaminated soils to stimulate naturally occurring bacteria to degrade organic contaminants. It can be used for soil and groundwater. Generally, this technique includes conditions such as the infiltration of water-containing nutrients and oxygen or other electron acceptors for groundwater treatment.

Biosparging: Biosparging involves the injection of air under pressure below the water table to increase groundwater oxygen concentrations and enhance the rate of biological degradation of contaminants by naturally occurring bacteria. Biosparging increases the mixing in the saturated zone and thereby increases the contact between soil and groundwater. The ease and low cost of installing small-diameter air injection points allows considerable flexibility in the design and construction of the system.

Bioaugmentation: Bioremediation frequently involves the addition of microorganisms indigenous or exogenous to the contaminated sites. Two factors limit the use of added microbial cultures in a land treatment unit: 1) nonindigenous cultures rarely compete well enough with an indigenous population to develop and sustain useful population levels and 2) most soils with long-term exposure to biodegradable waste have indigenous microorganisms that are effective degrades if the land treatment unit is well managed.

***Ex situ* bioremediation**

These techniques involve the excavation or removal of contaminated soil from ground.

Landfarming: Landfarming is a simple technique in which contaminated soil is excavated and spread over a prepared bed and periodically tilled until pollutants are degraded. The goal is to stimulate indigenous biodegradative microorganisms and facilitate their aerobic degradation of contaminants. In general, the practice is limited to the treatment of superficial 10–35 cm of soil. Since landfarming has the potential to reduce monitoring and maintenance costs, as well as clean-up liabilities, it has received much attention as a disposal alternative.

Composting: Composting is a technique that involves combining contaminated soil with nonhazardous organic amendants such as manure or agricultural wastes. The presence of these organic materials supports the development of a rich microbial population and elevated temperature characteristic of composting.

Biopiles: Biopiles are a hybrid of landfarming and composting. Essentially, engineered cells are constructed as aerated composted piles. Typically used for treatment of surface contamination with petroleum hydrocarbons they are a refined version of landfarming that tend to control physical losses of the contaminants by leaching and volatilization. Biopiles provide a favorable environment for indigenous aerobic and anaerobic microorganisms.

Bioreactors: Slurry reactors or aqueous reactors are used for *ex situ* treatment of contaminated soil and water pumped up from a contaminated plume. Bioremediation in reactors involves the processing of contaminated solid material (soil, sediment, sludge) or water through an engineered containment system. A slurry bioreactor may be defined as a containment vessel and apparatus used to create a three-phase (solid, liquid, and gas) mixing condition to increase the bioremediation rate of soilbound and water-soluble pollutants as a water slurry of the contaminated soil and biomass (usually indigenous microorganisms) capable of degrading target contaminants. In general, the rate and extent of biodegradation are greater in a bioreactor system than *in situ* or in solid-phase systems because the contained environment is more manageable and hence more controllable and predictable. Despite the advantages of reactor systems, there are some disadvantages. The contaminated soil requires pre treatment (e.g., excavation) or alternatively the contaminant can be stripped from the soil via soil washing or physical extraction (e.g., vacuum extraction) before being placed in a bioreactor.

Advantages of bioremediation

- ❑ Bioremediation is a natural process and is therefore perceived by the public as an acceptable waste treatment process for contaminated material such as soil. Microbes able to degrade the contaminant increase in numbers when the contaminant is present; when the contaminant is degraded, the biodegradative population declines. The residues for the treatment are usually harmless products and include carbon dioxide, water, and cell biomass.
- ❑ Theoretically, bioremediation is useful for the complete destruction of a wide variety of contaminants. Many compounds that are legally considered to be hazardous can be transformed to harmless products. This eliminates the chance of future liability associated with treatment and disposal of contaminated material.
- ❑ Instead of transferring contaminants from one environmental medium to another, for example, from land to water or air, the complete destruction of target pollutants is possible.
- ❑ Bioremediation can often be carried out on site, often without causing a major disruption of normal activities. This also eliminates the need to transport quantities of waste off site and the potential threats to human health and the environment that can arise during transportation.
- ❑ Bioremediation can prove less expensive than other technologies that are used for clean-up of hazardous waste.

Disadvantages of bioremediation

- Bioremediation is limited to those compounds that are biodegradable. Not all compounds are susceptible to rapid and complete degradation.
- There are some concerns that the products of biodegradation may be more persistent or toxic than the parent compound.
- Biological processes are often highly specific. Important site factors required for success include the presence of metabolically capable microbial populations, suitable environmental growth conditions, and appropriate levels of nutrients and contaminants.
- It is difficult to extrapolate from bench and pilot-scale studies to full-scale field operations.
- Research is needed to develop and engineer bioremediation technologies that are appropriate for sites with complex mixtures of contaminants that are not evenly dispersed in the environment. Contaminants may be present as solids, liquids, and gases.
- Bioremediation often takes longer than other treatment options, such as excavation and removal of soil or incineration.
- Regulatory uncertainty remains regarding acceptable performance criteria for bioremediation. There is no accepted definition of “clean”, evaluating performance of bioremediation is difficult, and there are no acceptable endpoints for bioremediation treatments.

Phytoremediation

Although the application of microbe biotechnology has been successful with petroleum-based constituents, microbial digestion has met limited success for widespread residual organic and metals pollutants. Vegetation- based remediation shows potential for accumulating, immobilizing, and transforming a low level of persistent contaminants. In natural ecosystems, plants act as filters and metabolize substances generated by nature. Phytoremediation is an emerging technology that uses plants to remove contaminants from soil and water. The term “phytoremediation” is relatively new, coined in 1991. Its potential for encouraging the biodegradation of organic contaminants requires further research, although it may be a promising area for the future.

Phytoremediation techniques can be classified into the following five types based on the contaminant fate: phytoextraction, phytotransformation, phytostabilization, phytodegradation, rhizofiltration, even if a combination of these can be found in nature.

Phytoextraction or phytoaccumulation is the process used by the plants to accumulate contaminants into the roots and aboveground shoots or leaves. This technique saves tremendous remediation cost by accumulating low levels of contaminants from a widespread area. Unlike the degradation mechanisms, this process produces a mass of plants and contaminants (usually metals) that can be transported for disposal or recycling.

Phytotransformation or phytodegradation refers to the uptake of organic contaminants from soil, sediments, or water and, subsequently, their transformation to more stable, less toxic, or less mobile form. Metal chromium can be reduced from hexavalent to trivalent chromium, which is a less mobile and noncarcinogenic form.

Phytostabilization is a technique in which plants reduce the mobility and migration of contaminated soil. Leachable constituents are adsorbed and bound into the plant structure so that they form a stable mass of plant from which the contaminants will not reenter the environment.

Phytodegradation or rhizodegradation is the breakdown of contaminants through the activity existing in the rhizosphere. This activity is due to the presence of proteins and enzymes produced by the plants or by soil organisms such as bacteria, yeast, and fungi. Rhizodegradation is a symbiotic relationship that has evolved between plants and microbes. Plants provide nutrients necessary for the microbes to thrive, while microbes provide a healthier soil environment.

Rhizofiltration is a water remediation technique that involves the uptake of contaminants by plant roots. Rhizofiltration is used to reduce contamination in natural wetlands and estuary areas.

Phytoremediation is well suited for use at very large field sites where other methods of remediation are not cost effective or practicable; at sites with a low concentration of contaminants where only polish treatment is required over long periods of time; and in conjunction with other technologies where vegetation is used as a final cap and closure of the site. There are some limitations to the technology that it is necessary to consider carefully before it is selected for site remediation: long duration of time for remediation, potential contamination of the vegetation and food chain, and difficulty establishing and maintaining vegetation at some sites with high toxic levels.

Biomining

Microbial processes are gaining increasing interest in the mining industry. **Bioleaching** of heavy metals, **biooxidation** of gold ores, desulfurization of coal and oil, tertiary recovery of oil and **biosorption** of metal ions are examples of the wide variety of potential and actual applications of microorganisms in mining and related fields. Currently, bacterial leaching and biooxidation are largescale processes that are being successfully used in copper and gold processing.

The term **biomining** has been coined to refer to the use of microorganisms in mining processes. Biomining encompasses two related microbial processes that are useful in the extractive metallurgy of heavy metals: bacterial leaching, also known as **bioleaching**, and **biooxidation**.

Leaching is the solubilization of one or more components of a complex solid by contact with a liquid phase. In bacterial leaching, the solubilization is mediated by bacteria. So bacterial leaching is a process by which the metal of interest is extracted from the ore by bacterial action, as in the case of bacterial leaching of copper. On the other hand, biooxidation implies the bacterial oxidation of reduced sulfur species accompanying the metal of interest, as in the biooxidation of refractory gold minerals. For many years bioleaching was thought as a technology for the recovery of metals from low-grade ores, flotation tailings or waste material. Today bioleaching is being applied as the main process in large-scale operations in copper mining and as an important pretreatment stage in the processing of refractory gold ores. Metals such as copper and uranium from its ore has been extracted using acidophilic autotrophs e.g., *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans*, *Agrobacterium* in plant transformation, *Bacillus cereus*, *Bacillus polymxa*. These organisms live by oxidizing the sulphur that binds copper, zinc, lead and uranium into their respective sulphide minerals. This process releases the valuable metals.

The main advantages of bacterial leaching of copper and other heavy metals as compared with pyrometallurgy lie in its relative simplicity, mild operation conditions, low capital costs, low energy input, and in its friendliness towards the environment. The biooxidation of refractory gold ores presents similar characteristics when compared with roasting and pressure oxidation.

Bacterial leaching of copper is usually performed in heaps of ground ore or in dumps of waste or spent material. Heaps and dumps are irrigated in closed circuit with an acidic liquor that contains a fraction of the bacterial population, the rest being attached to mineral. When the desired metal concentration is attained, the rich liquor is pumped to the solvent extraction (SE) section and then sent to electrowinning (EW), where the fine metal is recovered. The raffinate from the SE section is recycled to the heap or dump and the spent liquor of the EW section is recycled to the SE operation. Heaps and dumps present a number of advantages such as simple equipment and operation, low investment and operation costs and acceptable yields. On the other hand it must be realized that the operation suffers from some severe limitations: the piled material is very heterogeneous and practically no close process control can be exerted, except for intermittent pH adjustment and the addition of some nutrients. Moreover, the rates of oxygen and carbon dioxide transfer that can be obtained are low, and extended periods of operation are required in order to achieve sufficient conversions.

From a process-engineering standpoint, the complex network of biochemical reactions encompassed in bioleaching would best be performed in reactors. The use of reactors would allow a good control of the pertinent variables, resulting in a better performance. Parameters such as volumetric productivity and degree of extraction can be significantly increased. The main limitation in the use of reactors in biomining is the very large amounts of run-of-mine ore that in most cases is to be treated. The Chuquicamata copper mine in Chile produced 630,000 tons of fine copper in 1999. The production of that amount of metal implied the treatment of around 6 million tons of run-of-mine. If such amount would to be treated in bioreactors, the required equipment volume would of the order of 30 million cubic meters, an unthinkable figure. This limits their application to the treatment of mineral concentrates or when moderate volumes of ore are to be processed. For instance, over 11,000 tons of gold concentrates are biooxidized in reactors every year.

Biosensors

For the detection and monitoring of pollutants, a wide range of biological methods are already in use to detect pollution incidents and for the continuous monitoring of pollutants. Long established measures include: counting the number of plant, animal and microbial species, counting the numbers of individuals in those species or analyzing the levels of oxygen, methane or other compounds in water. More recently, biological detection methods using biosensors and immunoassays have been developed and are now being commercialized.

Microbial **biosensors** are microorganisms, which produce a reaction upon contact with the substance to be sensed. Usually they produce light but cease to do so upon contact with substances, which are toxic to them. Both naturally occurring light emitting microorganisms as well as specially developed ones are used.

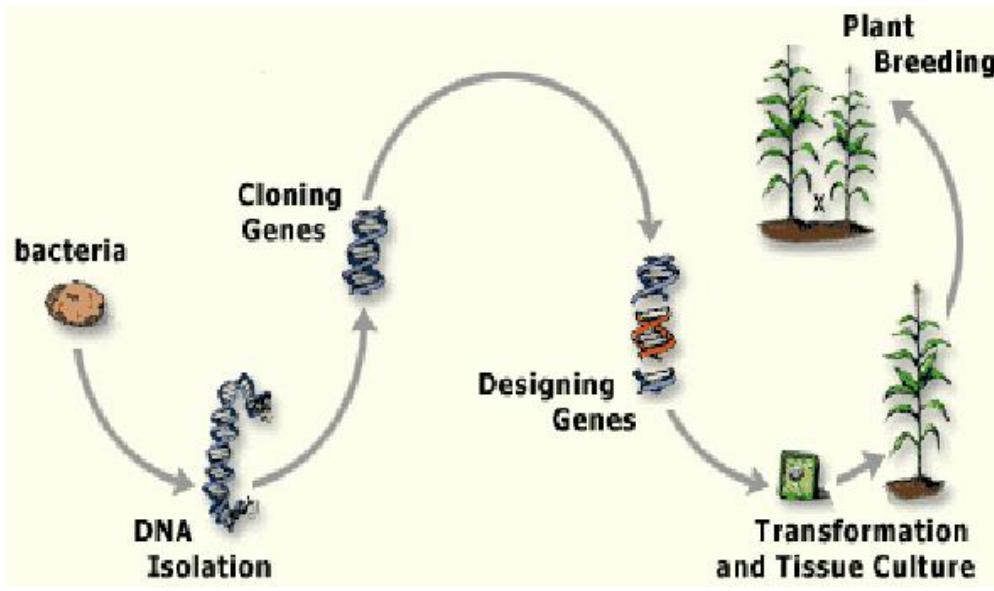
Most biosensors are a combination of biological and electronic devices - often built onto a **microchip**. The biological component might be simply an enzyme or antibody, or even a colony of bacteria, a membrane, neural receptor, or an entire organism. Immobilized on a substrate, their properties change in response to some environmental effect in a way that is electronically or optically detectable. It is then possible to make quantitative measurements of pollutants with extreme precision or to very high sensitivities.

In general terms, biotechnology can be used to assess the well-being ecosystems, transform pollutants into harmless substances, generate biodegradable materials from renewable sources, and develop environmentally safe manufacturing and disposal processes. Researchers are exploring biotechnological approaches to problem solving in many areas of environmental management and quality assurance such as: restoration ecology, contaminant detection, monitoring, and remediation, toxicity screening and conversion of waste to energy.

CHAPTER 17

APPLICATIONS OF rDNA TECHNOLOGY IN AGRICULTURE: PRODUCTION OF TRANSGENIC PLANTS

For several thousand years, farmers have been altering the genetic makeup of the crops they grow. Human selection for features such as faster growth, larger seeds or sweeter fruits has dramatically changed domesticated plant species compared to their wild relatives. Early farmers also discovered long ago that some crop plants could be artificially bred together or **cross-pollinated** to increase yields and to combine desirable characteristics from different parent plants in the offspring. The science of traditional or "classical" plant breeding developed in the 20th century with the application of newly understood genetic principles to this older craft of crop improvement. Plant breeders understood better how to select superior plants and breed them to create new and improved varieties of different crops. This has dramatically increased the productivity of the plants for food, fiber and other purposes.



Steps involved in plant gene transfer

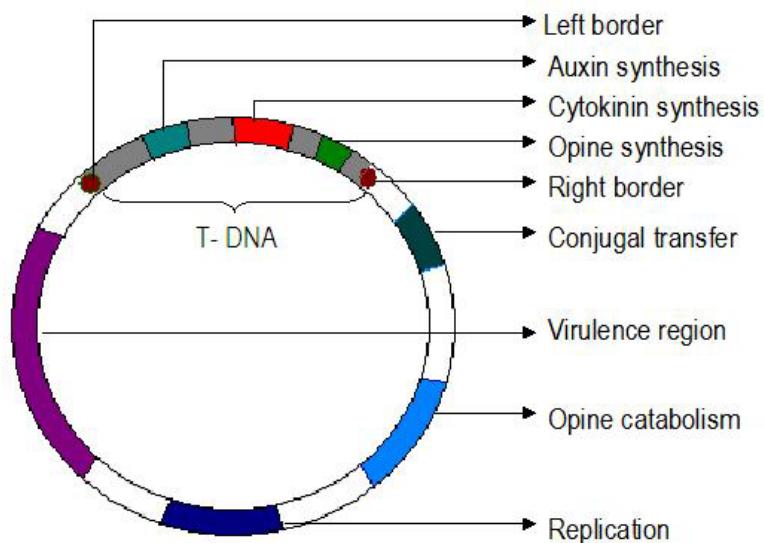
A plant breeder tries to assemble a combination of genes in a crop plant, which will make it as useful and productive as possible. Depending on where and for what purpose the plant is grown, desirable genes may provide features such as higher yield or improved quality, pest or disease resistance, or tolerance to heat, cold and drought. Combining the best genes in one plant is a long and difficult process; especially as traditional plant breeding has been limited to artificially crossing plants within the same species or with closely related species to bring different genes together. For example, a gene for protein in soybean could not be transferred to a completely different crop such as corn using traditional techniques. Transgenic technology enables plant breeders to bring together in one plant useful genes from a wide range of living sources, not just

from within the crop species or from closely related plants. This technology provides the means for identifying and isolating genes controlling specific characteristics in one kind of organism, and for moving copies of those genes into another quite different organism, which will then also have those characteristics. This powerful tool enables plant breeders to do what they have always done - generate more useful and productive crop varieties containing new combinations of genes - but it expands the possibilities beyond the limitations imposed by traditional cross-pollination and selection techniques.

***Agrobacterium tumefaciens*: A natural genetic engineer**

It could be said that the birth of modern plant genetic engineering stems from the recognition of the natural genetic engineering ability of the bacterium *Agrobacterium tumefaciens*. This opportunistic soil phytopathogen causes crown gall tumours in wounded gymnosperms and dicotyledonous angiosperms (dicots). Oncongenic strains contain a single copy of a large (150-250 kb) tumour-inducing (Ti) plasmid. Part of this plasmid DNA (the 'transfer' or T-DNA) is transferred to the wounded plant cell and stably integrated into the genome.

The genes encoded on the T-DNA, whilst bacterial in origin, contain plant (*i.e.* eukaryotic) regulatory signals enabling expression in infected plant cells. The expression of these genes has the following two consequences: (i) the synthesis of phytohormones; and (ii) the synthesis of opines. The former are necessary for the neoplastic transformation of the infected tissue, which proliferates to produce the characteristic tumorous gall of crown gall disease. The latter are amino acid derivatives which diffuse from the tumour into the surrounding soil where they may serve as a sole carbon source for *Agrobacteria* harbouring a Ti plasmid; they also induce the *tra* operon which allows conjugal transfer of the Ti plasmid to other *Agrobacteria*.



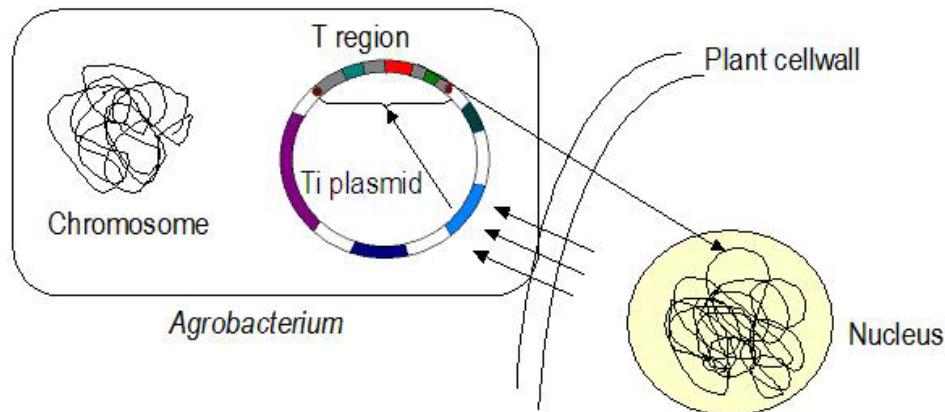
Structure of Ti plasmid

Thus, the Ti plasmid of *A. tumefaciens* has evolved an elegant genetic parasitism, in which infection results in the hijacking of plant metabolic resources to produce food (opines) that are metabolizable only through bacterial products encoded by the Ti plasmid; this plasmid can spread only in the presence of an infected tissue. Furthermore, infected tissue proliferates, thereby increasing the amount of opine available.

Biology of T-DNA transfer and crown gall disease

The infection of a plant cell by oncogenic *Agrobacteria* occurs in the following manner.

- The *Agrobacterium* recognizes some signal molecule exuded by the susceptible (wounded) plant cells and, in a chemo tactic response, moves up the concentration gradient towards these plant cells. The *Agrobacterium* must then become attached to the cell.
- In response to the same, or different, signal molecules, the *vir* regulon on the Ti plasmid is induced to express the gene necessary for T-DNA transfer to the plant cell.



Plant cell - *Agrobacterium* interaction

- The T-DNA is integrated into the plant genome and the T-DNA genes are expressed in the plant cell. The expression of the *orc* genes gives rise to cell proliferation and expression of the opine genes whose products are responsible for the synthesis of the opine amino acid derivatives.
- Finally, the opine(s) diffuses from the tumour tissue and into the surrounding *Agrobacteria*-containing soil, where it induces the two Ti plasmid encoded operons, *tra* and *oc*. The former encodes the genes necessary for conjugal transfer of the Ti plasmid to other *Agrobacteria*, while the latter encodes the genes required for opine catabolism.

What are Transgenic Plants?

A transgenic crop plant contains a gene or genes, which have been artificially inserted instead of the plant acquiring them through pollination. The inserted gene sequence (known as the **transgene**) may come from another unrelated plant, or from a completely different species: transgenic Bt corn, for example, which produces its own insecticide, contains a gene from a bacterium. Plants containing transgenes are often called **genetically modified** or **GM crops** although in reality all crops have been genetically modified from their original wild state by domestication, selection and controlled breeding over long periods of time.

Steps involved in synthesis transgenic plants

Introduction to DNA

The underlying reason that transgenic plants can be constructed is the universal presence of **DNA** (deoxyribonucleic acid) in the cells of all living organisms. This molecule stores the organism's genetic information and orchestrates the metabolic processes of life. Genetic information is specified by the sequence of four chemical bases (adenine, cytosine, guanine, and thymine) along the length of the DNA molecule. **Genes** are discrete segments of DNA that encode the information necessary for assembly of a specific protein. The proteins then function as enzymes to catalyze biochemical reactions, or as structural or storage units of a cell, to contribute to expression of a plant trait. The general sequence of events by which the information encoded in DNA is expressed in the form of proteins *via* an mRNA.

The transcription and translation processes are controlled by a complex set of regulatory mechanisms, so that a particular protein is produced only when and where it is needed. Even species that are very different have similar mechanisms for converting the information in DNA into proteins; thus, a DNA segment from bacteria can be inserted and made to function in a plant cell.

Among the most important tools in the genetic engineer's tool kit are enzymes that perform specific functions on DNA. Other enzymes known as **ligases** join the ends of two DNA fragments. These and other enzymes enable the manipulation and amplification of DNA, essential components in joining the DNA of two unrelated organisms.

Locating Genes for Plant Traits

Identifying and locating genes for agriculturally important traits is currently the most limiting step in the transgenic process. Relatively little is known about the specific genes required to enhance yield potential, improve stress tolerance, modify chemical properties of the harvested product, or otherwise affect plant characters. Usually, identifying a single gene involved with a trait is not sufficient; scientists must understand how the gene is regulated, what other effects it might have on the plant, and how it interacts with other genes active in the same biochemical pathway. Public and private research programs are investing heavily into new technologies to rapidly sequence and determine functions of genes of the most important crop species. These efforts

should result in identification of a large number of genes potentially useful for producing transgenic varieties.

Designing Genes for Insertion

Once a gene has been isolated and cloned (amplified in a bacterial vector), it must undergo several modifications before it can be effectively inserted into a plant.

- A **promoter sequence** must be added for the gene to be correctly expressed (i.e., translated into a protein product). The promoter is the on/off switch that controls when and where in the plant the gene will be expressed. To date, most promoters in transgenic crop varieties have been "constitutive", i.e., causing gene expression throughout the life cycle of the plant in most tissues. The most commonly used constitutive promoter is CaMV35S, from the cauliflower mosaic virus, which generally results in a high degree of expression in plants. Other promoters are more specific and respond to cues in the plant's internal or external environment. An example of a light-inducible promoter is the promoter from the cab gene, encoding the major chlorophyll a/b binding protein.

Promoter	Source	Nature of expression
Nopaline synthase (nos)	Ti plasmid	Constitutive
Octopine synthase (ocs)	Ti plasmid	Constitutive
Mannopine synthase (mas)	Ti plasmid	Constitutive
35S RNA promoter	Cauliflower Mosaic Virus	Constitutive
Ubiquitin-1	Maize	Constitutive
Actin-1	Rice	Constitutive
Zein	Maize	Seeds
Glutelin	Rice	Seeds

- Sometimes, the **cloned gene is modified** to achieve greater expression in a plant. For example, the *Bt* gene for insect resistance is of bacterial origin and has a higher percentage of A-T nucleotide pairs compared to plants, which prefer G-C nucleotide pairs. In a clever modification, researchers substituted A-T nucleotides with G-C nucleotides in the *Bt* gene without significantly changing the amino acid sequence. This process is described as **codon modification**. The result was enhanced production of the gene product in plant cells.
- The **termination sequence** signals to the cellular machinery that the end of the gene sequence has been reached.
- A **selectable marker gene** is added to the gene "construct" in order to identify plant cells or tissues that have successfully integrated the transgene. This is necessary because achieving incorporation and expression of transgenes in plant cells is a rare event, occurring in just a few percent of the targeted tissues or cells. Selectable marker genes

encode proteins that provide resistance to agents that are normally toxic to plants, such as antibiotics or herbicides. Only plant cells that have integrated the selectable marker gene will survive when grown on a medium containing the appropriate antibiotic or herbicide. As for other inserted genes, marker genes also require promoter and termination sequences for proper function.

Marker	Mode of selection
Drug resistance	
<i>aad</i>	Trimethoprim, streptomycin, spectinomycin, sulphonamides
<i>ble</i>	Bleomycin
<i>dhfr</i>	Methotrexate
<i>hpt</i>	Hygromycin
<i>nptii</i> and <i>aphii</i>	Kanamycin, Neomycin, G418
<i>gat</i>	Gentamycin
Herbicide resistance	
<i>bar</i> and <i>pat</i>	Phosphinothrinicin (bialaphos, glufossinate ammonium, Basta)
<i>csr1-1</i>	Chlorsupuron
<i>dhsps (sul)</i>	Sulfonamides
<i>epsps</i>	Glyphosate

Transforming Plants

Non-biological methods

Transformation is the heritable change in a cell or organism brought about by the uptake and establishment of introduced DNA. There are three possible methods of transforming plant cells and tissues viz. physical, chemical and biological methods. The following are the some of the successful methods employed in plant gene transfer

Polyethylene Glycol (PEG) method: The DNA transformation procedure using PEG involves incubating protoplasts and DNA in the presence of PEG and a post-incubation with high Ca^{2+} concentration. This procedure is similar to one used for fusion of protoplasts to produce somatic hybrids. In fusion conditions, PEG is usually present at a concentration of 40 percent for several seconds whereas for DNA transformation the optimal PEG concentration is 13.3 percent for an incubation of 30 minutes. This method was reported by **Krens** and his colleagues in 1982.

Electroporation method: Electroporation is another efficient method for the incorporation of foreign DNA into protoplasts. The principle underlying electroporation is to subject the solutions containing protoplasts and foreign DNA briefly to electric pulse to induce gentle ruptures of the membrane at places to form small reversible pores sufficient to take in DNA molecules. The

voltage range varies from 250-2000V. This method was introduced by **Fromm** and his coworkers in 1986.

Microinjection method: The microinjection method was introduced by two groups of scientists led by **Crossway** and **Reich** in 1986. In this method, the foreign DNA is directly injected into the nucleus of the protoplast, after immobilizing the protoplasts on either agar or on filter papers. Recently a method known as "**holding pipette method**" was introduced. In this, the protoplasts are isolated from cell suspension culture and are placed on a depression slide, by its side with a microdroplet of DNA solution. Using the holding pipette, the protoplast has to be held and the DNA to be injected into the nucleus using the injection pipette. After the micro injection the injected cells are cultured by hanging droplet culture method.

Microparticle bombardment method: In 1987, **Klein** and his colleagues evolved a method by which the delivery of DNA into cells of intact plant organs or cultured cells is done by a process called Projectile Bombardment. The micro-projectiles (small high density particles) are accelerated to high velocity by a particle gun apparatus. These particles with high kinetic energy penetrate the cells and membranes and carry foreign DNA inside of the bombarded cells. This method is otherwise called as "**Biolistics Method**".



Gene gun

Sonication method: This is a simple technique recently (1990) formulated by **Xu** and his coworkers. In this method the explants (especially leaves) are excised and cut into segments, immersed in sonication buffer containing plasmid DNA and carrier DNA in a sterile glass petridish. Then the samples were sonicated with an ultrasonic pulse generator at 0.5 c/cm^2 acoustic intensity for 30 minutes. After 30 minutes, the explants were rinsed in buffer solution without DMSO and transferred to the culture medium.

Biological method

Agrobacterium method: *Agrobacterium tumefaciens* is a remarkable species of soil-dwelling bacteria that has the ability to infect plant cells with a piece of its DNA. When the bacterial DNA is integrated into a plant chromosome, it effectively hijacks the plant's cellular machinery and uses it to ensure the proliferation of the bacterial population. The DNA in an *A. tumefaciens* cell is contained in the bacterial chromosome as well as in another structure known as a Ti (tumor-inducing) plasmid. The Ti plasmid contains a stretch of DNA termed T-DNA (~20 kb long) that is transferred to the plant cell in the infection process. To harness *A. tumefaciens* as a transgene vector, scientists have removed the tumor-inducing section of T-DNA, while retaining the T-DNA border regions and the *vir* genes. The transgene is inserted between the T-DNA border regions, where it is transferred to the plant cell and becomes integrated into the plant's chromosomes.

Selection of successfully transformed tissues

Following the gene insertion process, plant tissues are transferred to a selective medium containing an antibiotic or herbicide, depending on which selectable marker was used. Only plants expressing the selectable marker gene will survive, and it is assumed that these plants will also possess the transgene of interest. Thus, subsequent steps in the process will only use these surviving plants.

Regeneration of whole plants

To obtain whole plants from transgenic tissues such as immature embryos, they are grown under controlled environmental conditions in a series of media containing nutrients and hormones, a process known as tissue culture. Once whole plants are generated and produce seed, evaluation of the progeny begins. This regeneration step has been a stumbling block in producing transgenic plants in many species, but specific varieties of most crops can now be transformed and regenerated.

Genes and transgenics

Transfer of genes for herbicide tolerance

Success has been made in the incorporation of genes conferring tolerance to herbicides. The transgenic plants thus produced show expression of foreign genes resulting in a higher level of herbicide tolerance. The best example is the work of **Shah** and his co-workers. In 1986, they isolated a cDNA clone encoding an enzyme 5-enolpyruvyl-shikimate phosphate (EPSP) synthase from a glyphosate tolerant *Petunia hybrida* cell line. These cell lines over produced the enzyme to the tune of 20 times more. The chimeric EPSP synthase gene was constructed with the use of the cauliflower mosaic virus 35 promoter and introduced into the non-tolerant *Petunia* cell lines. The calli from transformed cell lines showed tolerance to glyphosate and the plants regenerated from the calli showed tolerance to the herbicide whereas the control plants died after spraying the herbicide.

In 1987, **De Block** and his co-workers transferred a gene conferring resistance to bialaphos and phosphinotricin called bar gene isolated from *Streptomyces hygroscopicus*, into tobacco, tomato and potato. This gene encodes for an enzyme, phosphinotricin acetyltransferase which prevents toxicity due to phosphinotricin. This gene has been transferred to tobacco, tomato and potato with the help of 35 S promoter of CaMV. The transformed plants showed high levels of resistance against field dose applications of the bialaphos and phosphinotricin. These results, pave the way to engineer resistance to various herbicides into major crops.

Expression of insect tolerance in transgenic plants

Bacillus thuringiensis is a bacterium that produces proteinaceous crystals during sporulation. These crystal proteins have insecticidal properties especially to lepidopteran insects. The use of *Bacillus thuringiensis* as a microbial insecticide offers advantages over chemical control agents in that the species-specific action of its insecticidal crystal proteins (ICPs) makes it harmless to non-target insects, to vertebrates, to the environment and the user. In 1987, **Fischhoff** and his colleagues constructed chimeric genes containing the CaMV35S promoter and the *B. thuringiensis* crystal protein coding sequences. The cloned *B. thuringiensis* gene has been introduced into tomato and tobacco and the transgenic plants thus produced show an increased level of resistance to lepidopteran insects. The larvae fed on transgenic plants were killed within 48 hours and there was little evidence of feeding damage to leaves of transformants. Thus, the introduction of toxin genes into plants seems to be a practical approach for providing protection against certain insect pests.

Expression of coat protein genes for virus protection

In agriculture cross-protection is a common practice to protect the plants from viruses and the coat protein of viruses have an important role in systemic cross protection. **Abel** and his coworkers introduced a chimeric gene containing a cloned cDNA of the coat protein (CP) gene of TMV into tobacco cells on a Ti plasmid of *A. tumefaciens* from which tumour inducing genes have been removed. Plants regenerated from transformed cells expressed TMV mRNA and CP as a nuclear trait. Seedlings from self fertilized transgenic plants were inoculated with TMV and observed for development of disease symptoms. The seedlings that expressed the CP gene showed delayed symptom development and 10-60 per cent of the transgenic plants failed to develop symptoms. This approach would be useful to develop lines with resistance to viruses where resistant varieties have been difficult to develop through conventional plant breeding.

Expression of antisense RNA in transgenic plants

Antisense RNA occurs naturally in several organisms to control gene expression. It can inhibit expression of a gene by preventing ribosome binding, obstructing transport of mRNA from nucleus, and increasing mRNA degradation. **Rottstein** and his co-workers in 1987 demonstrated the inhibition of the expression of the nopaline synthase (NOS) gene in tobacco. The transgenic plant having the NOS gene was transferred with a NOS antisense gene construct with CaMV 35S

promoter. The transformed plants were analysed for NOS activity and the enzyme activity varied depending on the tissue used. This mechanism can be a viable tool if the plants are transformed with antisense genes for the expression of various undesirable characters.

New plant varieties derived through recombinant DNA technology

2000		
Aventis.	Male sterile corn	The barnase gene from <i>Bacillus amyloliquefaciens</i> .
1999		
Agriotope Inc.	Modified fruit ripening cantaloupe	S-adenosylmethionine hydrolase gene from <i>E. coli</i> bacteriophage T3.
BASF AG	Phytase seed canola	The phytase gene from <i>Aspergillus niger</i> var van Tieghem.
Rhone-Poulenc Ag Company	Bromoxynil tolerant canola	The nitrilase gene from <i>Klebsiella pneumoniae</i> subsp. <i>ozaenae</i> .
1998		
AgrEvo, Inc.	Glufosinate tolerant soybean	Phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .
	Glufosinate tolerant sugar beet	Phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .
	Insect protected and glufosinate tolerant corn	The <i>cry9C</i> gene from <i>Bacillus thuringiensis</i> subsp. <i>tolworthi</i> and the bar gene from <i>Streptomyces hygroscopicus</i> .
	Male sterile or fertility restorer and glufosinate tolerant canola	The male sterile canola contains the barnase gene and the fertility restorer canola contains the barstar gene from <i>Bacillus amyloliquefaciens</i> . Both lines have the phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .
Calgene Co.	Bromoxynil tolerant/insect protected cotton	Nitrilase gene from <i>Klebsiella pneumoniae</i> and the <i>cryIA(c)</i> gene from <i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> .
	Insect protected tomato	The <i>cryIA(c)</i> gene from <i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> .
Monsanto Co.	Glyphosate tolerant corn	A modified enolpyruvylshikimate-3-phosphate synthase gene from corn.
	Insect and virus protected potato	The <i>cryIIIA</i> gene from <i>Bacillus thuringiensis</i> (Bt) sp. <i>tenebrionis</i> and the Potato Leafroll Virus replicase gene.
	Insect and virus protected potato	The <i>cryIIIA</i> gene from <i>Bacillus thuringiensis</i> (Bt) sp. <i>tenebrionis</i> and the Potato Virus Y coat protein gene.
Monsanto Co./Novartis	Glyphosate tolerant sugar beet	The enolpyruvylshikimate-3-phosphate synthase gene from <i>Agrobacterium</i> sp. strain CP4, and a truncated glyphosate oxidoreductase gene from <i>Ochrobactrum anthropi</i> .
Pioneer Hi-Bred	Male sterile corn	The DNA Adenine methylase gene from <i>Escherichia coli</i> .
University of Saskatchewan	Sulfonylurea tolerant flax	Acetolactate synthase gene from <i>Arabidopsis</i> .

1997		
AgrEvo, Inc.	Glufosinate tolerant canola	Phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .
Bejo Zaden BV	Male sterile radicchio rosso	The barnase gene from <i>Bacillus amyloliquefaciens</i>
Dekalb Genetics Corp.	Insect protected corn	The <i>cryIA(c)</i> gene from <i>Bacillus thuringiensis</i> (Bt).
DuPont	High oleic acid soybean	Sense suppression of the GmFad2-1 gene which encodes a delta-12 desaturase enzyme.
Seminis Vegetable Seeds	Virus resistant squash	Coat protein genes of cucumber mosaic virus, zucchini yellow mosaic virus, and watermelon mosaic virus 2.
University of Hawaii/Cornell University	Virus resistant papaya	Coat protein gene of the papaya ringspot virus.
1996		
Agritope Inc.	Modified fruit ripening tomato	S-adenosylmethionine hydrolase gene from <i>E. coli</i> bacteriophage T3.
Dekalb Genetics Corp.	Glufosinate tolerant corn	Phosphinothricin acetyl transferase gene from <i>Streptomyces hygroscopicus</i> .
DuPont	Sufonylurea tolerant cotton	Acetolactate synthase gene from tobacco, <i>Nicotiana tabacum</i> cv. <i>Xanthi</i> .
Monsanto Co.	Insect protected potato	The <i>cryIIA</i> gene from <i>Bacillus thuringiensis</i>
	Insect protected corn	The <i>cryIA(b)</i> gene from <i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> .
	Insect protected corn	The <i>cryIA(b)</i> gene from <i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> .
	Glyphosate tolerant/insect protected corn	The enolpyruvylshikimate-3-phosphate synthase gene from <i>Agrobacterium</i> sp. strain CP4 and the glyphosate oxidoreductase gene from <i>Ochrobactrum anthropi</i> in the glyphosate tolerant lines. The CryIA(b) gene from <i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> in lines that are also insect protected.
Northrup King Co.	Insect protected corn	The <i>cryIA(b)</i> gene from <i>Bacillus thuringiensis</i> (Bt) subsp. <i>kurstaki</i> .
Plant Genetic Systems NV	Male sterile and fertility restorer oilseed rape	The male sterile oilseed rape contains the barnase gene from <i>Bacillus amyloliquefaciens</i> ; the fertility restorer lines express the barstar gene from <i>Bacillus amyloliquefaciens</i> .
Plant Genetic Systems NV	Male sterile corn	The barnase gene from <i>Bacillus amyloliquefaciens</i>
Northrup King Co.	Insect protected corn	The <i>cryIA(b)</i> gene from <i>Bacillus thuringiensis</i> (Bt) subsp. <i>kurstaki</i> .
Plant Genetic Systems NV	Male sterile and fertility restorer oilseed rape	The male sterile oilseed rape contains the barnase gene from <i>Bacillus amyloliquefaciens</i> ; the fertility restorer lines express the barstar gene from <i>Bacillus amyloliquefaciens</i> .
	Male sterile corn	The barnase gene from <i>Bacillus amyloliquefaciens</i>
1995		
AgrEvo Inc.	Glufosinate tolerant canola	Phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .
	Glufosinate tolerant corn	Phosphinothricin acetyltransferase gene from <i>Streptomyces viridochromogenes</i> .

Calgene Inc.	Laurate canola	The 12:0 acyl carrier protein thioesterase gene from California bay, <i>Umbellularia californica</i> .
Ciba-Geigy Corp.	Insect protected corn	The <i>cry1A(b)</i> gene from <i>Bacillus thuringiensis kurstaki</i> .
Monsanto Co.	Glyphosate tolerant cotton	Enolpyruvylshikimate-3-phosphate synthase gene from <i>Agrobacterium</i> sp. strain CP4.
	Glyphosate tolerant canola	Enolpyruvylshikimate-3-phosphate synthase gene from <i>Agrobacterium</i> sp. strain CP4.
	Insect protected cotton	The <i>cry1A(c)</i> from <i>Bacillus thuringiensis</i> (Bt) subsp. <i>kurstaki</i> .
1994		
Asgrow Seed Co.	Virus resistant squash	Coat protein genes of watermelon mosaic virus 2 and zucchini yellow mosaic virus.
Calgene Inc.	Flavr Savr™ tomato	Antisense polygalacturonase gene from tomato.
	Bromoxynil tolerant cotton	A nitrilase gene isolated from <i>Klebsiella ozaenae</i> .
DNA Plant Technology Corp.	Improved ripening tomato	A fragment of the aminocyclopropane carboxylic acid synthase gene from tomato.
Monsanto Co.	Glyphosate tolerant soybean	Enolpyruvylshikimate-3-phosphate synthase gene from <i>Agrobacterium</i> sp. strain CP4.
	Improved ripening tomato	Aminocyclopropane carboxylic acid deaminase gene from <i>Pseudomonas chloraphis</i> strain 6G5.
	Insect protected potato	The <i>cryIIIA</i> gene from <i>Bacillus thuringiensis</i> (Bt) sp. <i>tenebrionis</i> .
Zeneca Plant Science	Delayed softening tomato	A fragment of the polygalacturonase gene from tomato.

Issues on genetically modified crops : Time line

- 1993 International Undertaking (IU).
- 1990 EU Directive 90/220/EEC on deliberate release of GMOs into the environment.
- 1992 Convention on Biodiversity, Rio.
- Early 1990s China grows transgenic tobacco.
- 1994 Calgene launches Flavr Savr Tomato.
- 1995 UNEP International Technical Guidelines for Safety in Biotechnology adopted.
- 1996 Commercial introduction of GM food crops in US.
- 1997 EU Directive 90/220 modified to include labelling. EU regulation on novel food came into force.
- 1998 Asian Food Information Centre established in Singapore.
- 1998–99 Height of GM debate in Europe.
- 1998 EU Council of ministers adopt regulation to introduce more comprehensive measures to include detectable traces of GM protein on DNA (further amended in 2000).
- 1998 Controversial paper published on feeding GM potatoes containing a lectin gene (from snowdrops) to rats. The science was disputed but the incident created furore and moulded public opinion in the UK over the safety of GM products.
- 1999 *De facto* moratorium on product approvals in the EU.
- 1999 Korea Consumer Protection Board announces that 82% of tofu contains traces of GM.

- 1999 Paper published in Nature suggested harmful effect of Bt pollen on monarch butterflies Leads to extensive two-year debate. May 1999 British Medical Association publishes report recommending a moratorium on GM crops.
- 1999 Monsanto's GM milk ruled unsafe by Codex Alimentarius Commission.
- 1999 Taiwan announces that by 2005 labelling on GM foods will be introduced.
- 2000 Egypt, Saudi Arabia and Kuwait ban imports of Thai canned tuna after it was found to be packed in GM oil. Egypt refused to change its ban on imports of Thai tuna, so Thailand requested WTO talks with Egypt on September 22, 2000.
- 2000 Cartagena Protocol on Biosafety.
- 2000 OECD Edinburgh Conference.
- 2000 Council for Biotechnology Information (CBI) formed in US by Aventis, BASF, BIO, Dow, Dupont, Monsanto, Novartis, and Zeneca: US\$50 million public information campaign to promote agro-biotechnology.
- 2000 Aventis introduces Liberty Link canola.
- 2000 Recall of maize products found to be contaminated with Starlink Bt maize.
- 2000 Monsanto announce 'new pledge' committing company to dialogue, transparency, respect, sharing and delivering of benefits and environmental responsibility.
- 2000 Monsanto undertakes to not use animal or human genes in GM crops or to develop sterile gene technologies.
- 2001 UNDP human development report: Making New Technologies Work for Human Development.
- 2001 GM traces found in Gerber baby food (declared non-GM) in Thailand and Philippines (made by Indofoods).
- 2001 USDA releases Guidance for Industry: Voluntary Labelling.
- 2001 Torrey Mesa Research (Syngenta) announces completion of genome sequencing of rice.
- 2001 Golden rice (Vitamin-A-enhanced) starts undergoing trials in IRRI, Los Baños, Philippines.
- 2001 Japan issues guidelines on mandatory labelling of GM foods.
- 2001 Sri Lanka announces proposed ban on all GM imports.
- 2001 Korea introduces labelling of GM foods.
- 2001 OECD Bangkok Conference.
- 2001 New Zealand Royal Commission report published. Sep 2001 EU launch public consultation on Life Sciences and Biotechnology.
- 2001 Sri Lanka indefinitely suspends ban on GM after international pressure from US and WTO.
- 2001 Paper published in Nature suggesting that GM transgenes had been found in maize land races in Mexico, an important in situ centre for genetic diversity.
- 2001 WTO Qatar.
- 2001 Australia and New Zealand amend labelling regulations to become mandatory.
- 2001 ICGEB Risk Monitoring and Public Perception workshop in Venezuela.

- 2001 OECD conference to address environmental impacts of genetically modified organisms.
- 2002 UNEP-GEF launches multi-million-dollar project to help countries develop the scientific and legal skills for evaluating the health and environmental issues surrounding the imports of LMOs.
- 2002 English Nature raise gene stacking concerns in weeds after study commissioned in Canada finds transgene flow from GM canola.
- 2002 Mexican government and independent scientists find GM contamination for the second year running among wild maize varieties in an important in situ centre for maize genetic resources.
- 2002 Codex Task force agrees on draft basic principles for safety analysis of GM foods and safety guidelines for agricultural products are drafted.
- 2002 China and the US agree to a 30-day waiting period and temporary recognition of US safety assurances after stand-off over proposed import restrictions that were to come into force on March 20.
- 2002 EU agrees to lift three-year moratorium on GM approvals by mid-October, defusing long-standing US/EU dispute.

CHAPTER 18

APPLICATIONS OF rDNA TECHNOLOGY IN FOOD INDUSTRY

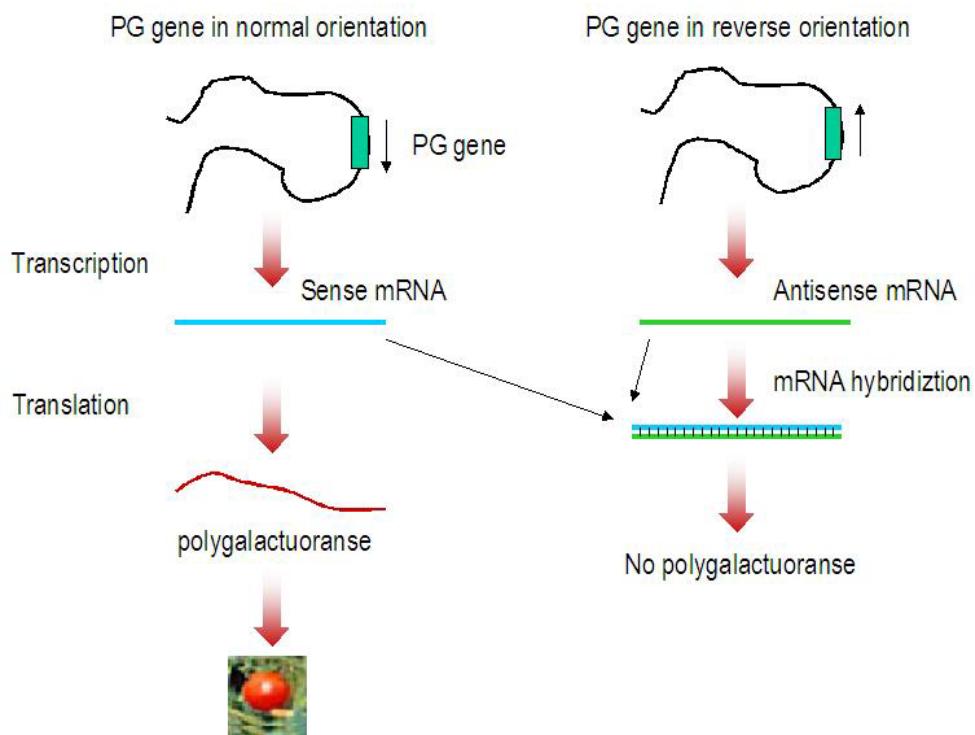
Experts on population growth predict that the greatest challenge in the next century will be to maintain an abundant and safe food supply for the world population. Food biotechnology has been presented as one of the solutions to this problem. It is a technology about which many questions have been asked. Biotechnology has the potential to offer consumers, farmers and food processors a greater variety of safe food with benefits like better quality, higher yields and improved nutritional value. The volume of biotech crops in developmental stages continues to grow. Biotechnology has been used in a number of crops for several years, and more genetically enhanced products are expected to be on the market in the coming years, many of which are similar in nature to products already in the field or on the market. These genetically modified crops (GM crops) or genetically modified organisms (GMO) or living modified organisms (LMO) are expected to give several food products with reduced levels of natural toxins in plants, increased levels of nutritional important compounds or molecules, increased shelf life, increased taste and palatability and so on. Some of the specific products that should soon be on the market as a result of rDNA research include:

- oils, such as soybean and canola oils, developed to contain more stearate, making margarine and shortenings more healthful
- peas grown to remain sweeter and produce higher crop yields
- smaller, seedless melons for use as single servings
- bananas and pineapples with delayed ripening qualities
- peanuts with improved protein balance
- fungal resistant bananas
- tomatoes with a higher antioxidant (lycopene) content than current varieties
- potatoes with a higher solids content (higher starch) than conventional potatoes
- reducing the amount of oil absorbed during processing of foods like French fries or potato chips
- fruits and vegetables fortified with or containing higher levels of vitamins such as C and E, to potentially protect against the risk of chronic diseases such as cancer and heart disease
- rice with increased levels of provitamin A and Fe
- garlic cloves, producing more allicin, possibly helping to lower cholesterol levels
- higher-protein rice, using genes transferred from pea plants
- strawberries, containing increased levels of ellagic acid, a natural-cancer fighting agent
- peppers, strawberries, raspberries, bananas, sweet potatoes and melons that are enhanced for better nutrition and quality
- strawberries with higher crop yields and improved freshness, flavor and texture

Food crops with increased shelf-life

Tomatoes are usually picked green, so that they are able to withstand shipping and transportation without bruising. They are then ripened artificially by using ethylene gas, as ethylene is a key trigger for the ripening process. During the ripening process, polygalacturonase (PG), an enzyme responsible for the digestion of pectin in the cell wall, digests the pectin in the cell wall, and makes the fruit soft. In trying to delay ripening, two approaches have been used. One is to target the production of ethylene itself, thus delaying the onset of normal ripening mechanism. A second approach is to block the production of polygalacturonase. This was achieved by using novel rDNA technology called antisense RNA technology.

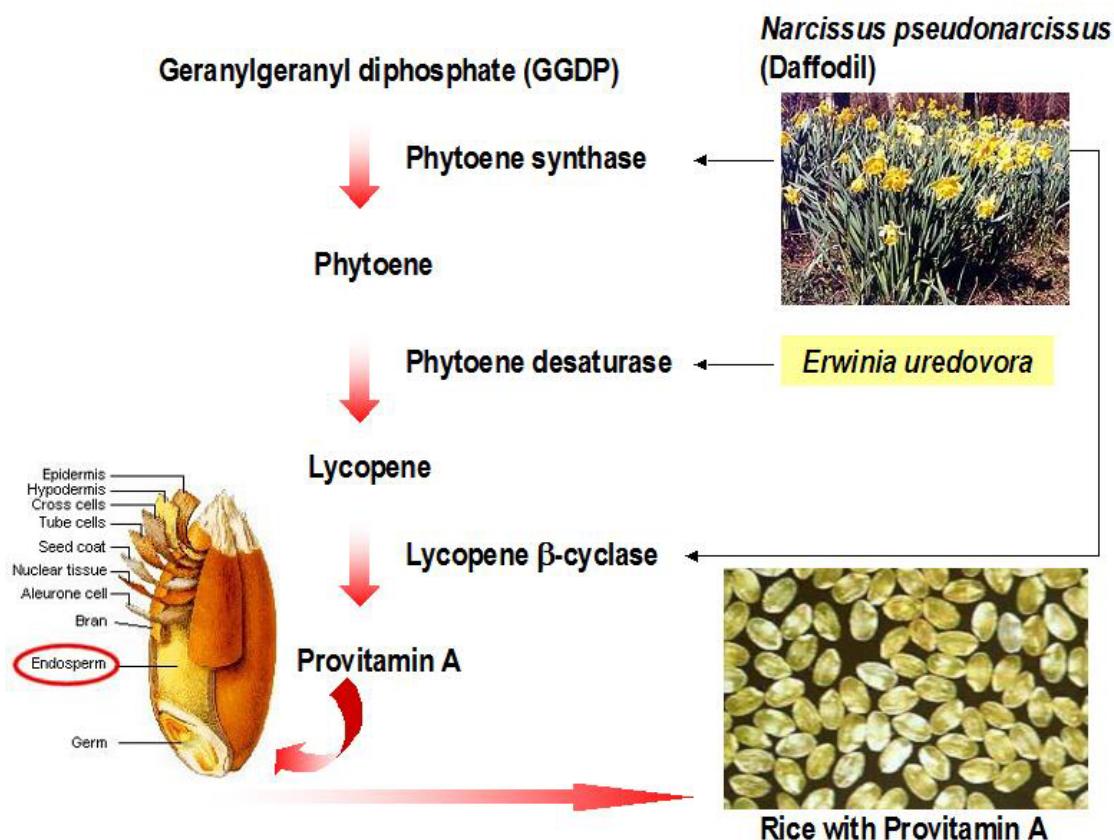
In this antisense RNA technology, the gene for polygalacturonase was targeted. The gene sequence of polygalacturonase is inserted in the opposite orientation, so that on transcription an mRNA that is complementary to the normal mRNA is produced. This antisense mRNA will therefore bind to the sense mRNA in the cell, inhibiting its translation and effectively shutting the expression of the gene.



The tomato developed from the above technology remains as the first genetically modified food to be sold in the market. The biotech company Calgene developed this tomato and named it as Flavr Savr tomato. However, various problems with the characteristics affecting the growth and picking of the crop led to the failure of the Flavr Savr in commercial terms.

Food crops with improved nutritional quality

The vitamin and mineral content of plants vary from species to species, and from tissue to tissue in a particular plant. Milled cereal grains- basically the endosperm component of the seed- are particularly deficient in essential vitamins and minerals, and yet they represent the staple food for much of the world's population. Golden rice is the result of an effort to develop rice varieties that produce provitamin-A (β - carotene) as a means of alleviating vitamin A (retinol) deficiencies in the diets of poor and disadvantaged people in developing countries.



Provitamin-A is not produced by traditional rice varieties. However, geranylgeranyl diphosphate (GGDP), a compound normally found in rice can be used to produce provitamin-A. Through the work of two European scientists, Ingo Potrykus of the Swiss Federal Institute of Technology in Zurich and Peter Beyer of the University of Freiburg in Germany, rice plants were developed containing two genes from daffodil (*Narcissus pseudonarcissus*) and one gene from bacterium *Erwinia uredovora* that carry out the four steps required for the production of provitamin-A in rice endosperm, the nutritive tissue surrounding the embryo of a seed and makes up the majority of the rice grain that is eaten. These genes were inserted into the rice genome with specific promoters such that the enzymes and the provitamin-A are only produced in the rice endosperm. The resulting plants appear normal except that after milling (to remove the brown bran), their grain is a golden yellow color due to the presence of provitamin-A.

Foods crops with improved quality of lipids, fats and oils

Dietary fats and oils contribute to balanced nutrition, providing energy, fat-soluble vitamins and the essential polyunsaturated fatty acids (linoleic and linolenic) required for growth, cellmembrane function, development and disease prevention. Conversely, fats and oils have been implicated in a diverse group of ailments such as obesity, cancer and heart disease. A whole range of genetically modified oil seeds may be available in the future to promote health and prevent disease. In addition, specialty seeds with novel oil composition may be available to serve unique nutritional needs. For example, plants could be engineered to generate Lorenzo's Oil, a mixture of long-chain, saturated fatty acids used to treat patients suffering from the inherited genetic disorder ALD (adrenoleukodystrophy). Changing the composition of plant based fats and oils can also improve their taste, quality and oxidative stability. Rancidity, a common off-flavor, can be reduced by regulating the amount of polyunsaturated fats or through the accumulation of antioxidants such as vitamin E, which preserves fats and oils and prevents oxidation. For example, walnuts or peanuts could be modified to maintain healthy oils and oxidative stability. Natural margarines could be made in plants through the synthesis of saturated fatty acids like those found in cocoa butter. This would not only improve the taste but also make it possible to produce margarine free of the unhealthy trans fatty acids that accumulate in partially hydrogenated traditional margarines.

In addition, oil-producing seed crops may be modified to make specialty oils for a variety of nonfood products such as soaps, detergents, cosmetics, candles, lubricants, grease, polymers, plasticizers, coatings, inks, printing, textile dyes and surfactants. Unlike their petroleum counterparts, these products would be biodegradable and better for the environment. The economic value of plants modified for pharmaceutical, medical or industrial applications will be much greater.

Foods crops with improved quality of proteins

Plant proteins provide amino acids important for human health. Many plant proteins, like those present in corn seed, do not have the complete complement of essential amino acids. Plant-seed proteins can be modified to express proteins with a more desirable amino-acid composition. This is particularly important for animal feeds, where seeds engineered to produce a higher concentration of sulfur-containing amino acids could improve wool growth in sheep. The amino-acid composition of seed-storage proteins found in seeds, nuts, fruits and tubers can be modified to enhance nutritional value. Plants may also be modified to produce proteins that aid in mineral nutrition, such as hemoglobin to improve iron uptake and other specific proteins to improve calcium uptake. A significant segment of the U.S. population suffers from food allergies, an immunoglobulin (IgE)-mediated hypersensitivity reaction to particular plant proteins. Understanding the structure of the antigenic determinants of proteins that contribute to severe and often life-threatening allergy reactions could lead to the development of plant products that alter these proteins, greatly limiting the problem. Care must be taken in assessing transgenic products for their allergenic potential.

Foods crops with improved quality of carbohydrates

Plants manufacture both polymeric carbohydrates, like starch, and individual sugars, like sucrose. Plant starch is used in a wide range of industrial applications such as coatings for paper and textiles and as a gelling agent in the food industry. It is now possible to make starches that are free of the amylose fraction, resulting in a gelling agent that is clearer and forms a gel at a lower temperature. Starches with higher levels of amylose are more desirable as coating agents and maintain texture at higher temperatures. For example, "sticky rice" has lower amylase content. The presence or absence of amylose greatly influences the physiochemical properties of starch; genetic engineering will result in specialized starches with higher value for specific applications. In the paper industry, starches sometimes must be chemically modified by phosphorylation, in which phosphate mono-ester residues are chemically added to corn starch. Some starches (e.g., potato) are naturally phosphorylated, containing the enzymes to add this residue to the starch molecule. Cloning the gene responsible for phosphorylation of glucans makes it possible to engineer precise levels of phosphorylation, resulting corn starch that is more useful to the paper industry. The chain length and distribution in amylopectin, the more highly structured (branched) component of starch, can be manipulated by altering the starch synthase activity. Genes responsible for the synthesis of fructans can modify plants of agronomic value to produce this polymeric carbohydrate. Fructans are an important ingredient in "functional" foods because they promote health and help to reduce the risk of colon cancer. Sugar beets that accumulate higher levels of fructans could be developed as a source for low-calorie sweeteners.

Foods crops with improved nutritional components

Health conscious consumers are compelling farmers and seed companies to improve the overall nutritional quality of their products. Extensive medical, biochemical and epidemiological research points to specific plant-produced substances (phytochemicals), as well as classes of phytochemicals that offer specific health benefits. Fruits and vegetables are a major source of beneficial phytochemicals. Phytochemical families with clearly beneficial health properties include glucosinolates found in the brassica vegetables including broccoli; carotenoids, such as the tomato fruit pigment lycopene, found in many plant families; flavonoids, such as the isoflavones found in soybeans; and the anthocyanins and flavonols found in many fruits and vegetables. Some foods containing consistently higher levels of these and other plant nutrients should be available through conventional breeding methods within 10 years. The natural variation that would provide the basis of health-enhanced varieties may be present already in breeding populations. Compared with traditional breeding strategies, the application of biotechnology to improve phytonutrient levels in whole foods is more difficult due to the complex array of potentially important chemicals and the complexity of the underlying biosynthetic pathways.

Foods crops with improved quality of flavor, color and sweetness

The ability to transgenically manipulate color intensity and hue was demonstrated more than 10 years ago. In flowers, the altered expression of the enzymes of flavonoid biosynthesis yielded novel floral pigmentation patterns. Such approaches have not been applied to fruits yet, but the potential exists. Anthocyanins are the pigments responsible for color in many fruits, such as

grapes and strawberries. Deeply colored fruits are generally more desirable to consumers. Further, anthocyanins and related flavonoids have antioxidant properties that reduce the risk of cardiovascular disease and cancer. Fruits with consistently higher levels of anthocyanins, produced through genetic modification, could reach the supermarket within 15 years. These will likely be produced by altering the expression of whole biochemical pathways rather than through modulation of specific enzymes. Improved flavor is of major interest to consumers, but it does not receive significant attention from breeders, who work largely to improve production and durability during post-harvest distribution. The complexity of flavor — which includes a balance between sweetness and acidity as well as the compounds that give products their characteristic taste — has discouraged the pursuit of biotechnological approaches to flavor improvement. Biotechnological efforts to improve sweetness have met with little success so far. In some cases, an increase in sweetness leads to a decrease in size that is unacceptable in the marketplace. In addition, attempts to increase sweetness by expressing nonsugar, sweetness-enhancing proteins such as monellin or thaumatin have been tried. Thaumatin is a protein which is isolated from katemfe fruit of west Africa (*Thaumatococcus daniellii* Benth). There may be several related proteins in the plant, but two main forms: thaumatin I and thaumatin II. Thaumatin I has is composed of 207 amino acids, with molecular weight 22,204. Thaumatin II is synthesized as a precursor protein of 235 amino acids; the first 22 amino acids and the last six amino acids are apparently cleaved to produce a protein the same size as thaumatin I (207 amino acids) and 98% identical to thaumatin I. Thaumatin is sweet, with a slow onset, lingering sweetness and a licorice after-taste. Sweetness potency relative to sucrose is about 2000 times on a weight basis, and about 100,000 times sucrose on a molar basis. All these efforts were frustrated because their compounds bind to cellular proteins and are subsequently not available to the sensory system.

CHAPTER 19

BIOSAFETY ISSUES AND rDNA RESEARCH

Although "biotechnology" and "genetic modification" (GM) commonly are used interchangeably, GM is a special set of technologies that alter the genetic makeup of such living organisms as animals, plants, or bacteria. Combining genes from different organisms is known as recombinant DNA technology, and the resulting organism is said to be "genetically modified," "genetically engineered," or "transgenic." GM products (current or in the pipeline) include medicines and vaccines, foods and food ingredients, feeds, and fibers.

GM crops are grown commercially or in field trials in over 40 countries and on six continents. In 2000, about 109.2 million acres were planted with transgenic crops, the principal ones being herbicide- and insecticide-resistant soybeans, corn, cotton, and canola. Other products expected include sweet potato resistant to a virus that could decimate most of the African harvest, rice with increased iron and vitamins that may alleviate chronic malnutrition in Asian countries, and a variety of plants able to survive weather extremes. On the horizon are bananas that produce human vaccines against infectious diseases such as hepatitis B; fish that mature more quickly; fruit and nut trees that yield years earlier, and plants that produce new plastics with unique properties. In 2000, countries that grew 99 percent of the global transgenic crops were the United States (68%), Argentina (23%), Canada (7%), and China (1%). Although growth is expected to plateau in industrialized countries, it is increasing in developing countries. The next decade will see exponential progress in GM product development as researchers gain increasing and unprecedented access to genomic resources that are applicable to organisms beyond the scope of individual projects.

Technologies for genetically modifying (GM) foods offer dramatic promise for meeting some areas of greatest challenge for the 21st century. Like all new technologies, they also pose some risks, both known and unknown. Controversies surrounding GM foods and crops commonly focus on human and environmental safety, labeling and consumer choice, intellectual property rights, ethics, food security, poverty reduction, and environmental conservation.

GM products: benefits and bontroversies

Benefits

Crops

- Enhanced taste and quality
- Reduced maturation time
- Increased nutrients, yields, and stress tolerance
- Improved resistance to disease, pests, and herbicides
- New products and growing techniques

Animals

- Increased resistance, productivity, hardiness, and feed efficiency
- Better yields of meat, eggs, and milk
- Improved animal health and diagnostic methods

Environment

- "Friendly" bioherbicides and bioinsecticides
- Conservation of soil, water, and energy
- Bioprocessing for forestry products
- Better natural waste management
- More efficient processing

Society

- Increased food security for growing populations

Controversies

Safety

- Potential human health impact: allergens, transfer of antibiotic resistance markers, unknown effects
- Potential environmental impact: unintended transfer of transgenes through cross-pollination, unknown effects on other organisms (e.g., soil microbes), and loss of flora and fauna biodiversity

Access and Intellectual Property

- Domination of world food production by a few companies
- Increasing dependence on Industrialized nations by developing countries
- Biopiracy—foreign exploitation of natural resources

Ethics

- Violation of natural organisms' intrinsic values
- Tampering with nature by mixing genes among species
- Objections to consuming animal genes in plants and vice versa
- Stress for animal

Labeling

- Not mandatory in some countries (e.g., United States)
- Mixing GM crops with non-GM confounds labeling attempts

Society

- New advances may be skewed to interests of rich countries

Major technological advances such as biotechnology are seldom implemented without controversy. For some people, a new technology is a blueprint for unimagined disaster. They believe that the traditional ways and values will be eroded. These individuals often feel that new initiatives have inherent hazards that can be avoided only if the development of the technology is prevented. Others view a new technology as a bountiful wellspring of untold benefits for society and argue that any obstacles will forestall, probably irreparably, essential advances. Often those with the latter attitude assume that a new technology is fragile and must be given special concessions to enable it to reach fruition. A third, "neutral" group captures the middle ground between these two extremes. This group believes that nothing can be considered fundamentally new, and, therefore, a new technology forms part of a continuum of discoveries. To this third group, existing regulatory systems are considered sufficient for handling the potential risks, and the expected benefits will inevitably occur as the new technology.

Because biotechnology can potentially affect many aspects of modern society, including agricultural production and medical treatment, there are significant ethical, legal, economic and social issues that need to be considered. For example, since its inception in 1973, serious doubts have been voiced by some individuals about the safety of recombinant DNA technology. These concerns prompted scientists to declare a self-imposed moratorium on certain types of rDNA experiments until the adoption of official regulatory guidelines designed to ensure that recombinant organisms were unable to proliferate outside or cause any other damage to other living organisms environment. The formulation of these regulations took place in 1974-75. Thus, the public became aware of the possibilities of, both negative and positive, of genetically manipulated organisms.

Regulating rDNA technology: The initial time line

1976: The US, National Institute of Health (NIH), the primary US research grant agency in the medical and health sciences, issued guidelines for the conduct of NIH supported research using rDNA technology.

1980: The NIH Recombinant DNA Advisory Committee (NIH-RAC) as a result of experience and specific data from studies that the committee and NIH had sponsored relaxed the original NIH guidelines considerably.

Over the past two decades the number of significant advances made in genetic engineering has increased dramatically. It is this increase in the use of new techniques for understanding and modifying the genetics of living organisms that has led to greatly increased interest and investment in biotechnology. The applications developed from these new methods place them firmly within the continuum of techniques used throughout human history in industry, agriculture and food processing. Thus, modern biotechnology provides powerful new tools and these tools are used to generate new products. Because of modern biotechnology provides enormous potential power that is far reaching, and the number and variety of products is so great, it is important to provide appropriate regulatory mechanisms to ensure that products produced by the use of new techniques are as safe as the products of traditional biotechnology, especially when those products are organisms that might interact with the environment. The regulatory

mechanisms, thus established for rDNA research, are considered as Bio-safety regulations in rDNA research.

Bio-safety regulations in rDNA research

Bio-safety regulations in rDNA research means the policies and procedures adopted to ensure the environmentally safe application of biotechnology. The three most important biosafety-related needs identified are 1) to develop biosafety guidelines, 2) establish a responsive national system and 3) increase capabilities to perform risk analysis of genetically modified organisms. Establishing a system for biosafety review, in itself, has many facets and challenges. These include the formulation and adoption of safety guidelines, establishing national and institutional biosafety committees and ensuring infrastructure for contained and large-scale field testing of genetically modified organisms. An effective biosafety system can be evolved when the following four conditions are met:

- The framework (policies, guidelines) defines clearly the structure of the biosafety system, the roles and responsibilities of those involved, and how the review process is to operate.
- The people involved are knowledgeable and well trained and have the support of their institutions.
- The review process is scientifically sound and uses available information to make decisions that balance benefits and risks.
- Feedback mechanisms are used to incorporate new information and revise the system as needed.

Biosafety

One of the major issues relating to the role and application of biotechnology in agriculture is the safety of organisms with novel traits and the appropriate regulatory measures for research and development, field-testing, and marketing of organisms with novel traits. This is because uncontrolled introduction of organisms with novel traits might cause undesirable changes in ecological or genetic relationships in some communities. Hence careful design and review of organisms with novel traits, along with proper planning and regulation of environmental introductions, is advisable to ensure that organisms with novel traits do not pose unacceptable risks to the environment. In performing risk assessment and risk management a distinction should be made between evaluation of organisms intended for contained use and those for planned introduction into uncontained settings.

Familiarity

Organisms with novel traits intended for introduction may differ from their parental organisms in their ability to survive and reproduce under varying environmental or climatic conditions. Organisms with novel traits should be evaluated on the basis of scientific principles for their potential to interact in unexpected and undesirable ways with local biological communities.

The concept of *familiarity* can assist decision makers in evaluations by providing a context in which to apply accumulated experience with such products. A wealth of experience has been acquired through traditional practices. Experience with the existing cultivars gives valuable insights as to the expected behavior of the new products.

Since the mid-1980s a substantial body of experience has also been accumulating with products to which new characteristics have been added through biotechnology. Field tests and data from laboratories and other facilities provide relevant information about phenotypic expression of the new characteristics in modified organisms and their interactions with the environment. By accumulating such experience it is expected that the performance of entire classes of organisms with novel traits will become familiar enough to require minimal regulatory attention. When familiarity with a plant or microorganism reaches such a level that there is reasonable assurance that the organism is essentially similar to known introductions, and when these present negligible risk, the introduction is assumed to be suitable for field testing, according to established practice.

Familiarity does not necessarily imply that the organism is safe – it does mean that accumulated knowledge and experience suggest appropriate and adequate approaches to risk assessment and management. In evaluating the potential risks associated with products of the new technologies, the aim is to introduce regulatory mechanisms to monitor any risks so that humanity can safely benefit from these new organisms and products.

Risk assessment

Risk assessment is the process of gathering diverse data to identify possible risks in research and development involving genetically modified organisms, plants and animals. Risk assessment should focus on the characteristics of the product itself rather than on the techniques used to produce it, provided standard safety measures are employed during production.

Following the framework put forward by the National Research Council of the U.S. National Academy of Sciences, the following factors are recognized to be important in assessing risks of organisms with novel characteristics:

For the environment

- Properties of the organism and of the environment into which it may be introduced
- Possibility of containing and controlling the organism
- Probable effect on the environment should the organism or genetic trait persist longer than intended or spread to non-target environments
- Risks to human health and the environment that is associated with introduction of organisms with novel traits.

For the organism with the novel trait, taking into account

- The recipient host or parental organism that receives the new trait
- The donor organism from which the trait is derived
- The vector used to transfer the trait from donor to recipient
- The inserted or introduced trait, including potential toxicity of a gene product or its metabolites
- Empirical data on the novel organism
- The intended application, for example, contained use or planned production
- The potential receiving environment

Another way of looking at risk assessment is to distinguish the following two parameters, namely *hazard* and *exposure*. *Hazard assessment* means evaluating whether an organism can be harmful and assessing whether it is a pest or a pathogen or if it will introduce new pests or pathogens or enhance existing ones. *Exposure assessment* involves evaluating the extent to which the environment or humans might be exposed to organisms with novel traits. The degree of exposure depends on the following parameters:

- The route of introduction
- The survival and reproductive potential of the organisms with novel traits
- The mode and rate of dispersal beyond the site of introduction (by wind, water and insects)
- The location and size of any receptive or susceptible population.

Finally, basic questions to be asked in performing risk assessment include:

- Are the potential risks of the organism in the environment acceptable, compared to the relative benefits of such an organism?
- Is the public regulatory mechanism adequate to ensure safety?

Possible types of risk to be assessed in agricultural biotechnology include the likelihood that the plant product or plant material may show any of the following characteristics:

- The potential to become weeds or to transmit weediness properties
- The potential to show undesirable toxicity
- The potential for microorganisms or viruses, including those used in vaccines, to exhibit undesirable pathogenicity
- The potential for insects to become pests.

Data obtained from studies performed in laboratories, greenhouses, and contained animal facilities can provide an indication of the degree of risk, but the inability to fully simulate environmental processes in contained facilities is in itself a justification to move to field trials in the open environment under appropriate conditions of management and containment. In the meantime, accumulating evidence from studies in greenhouses or contained animal facilities, as well as in

field trials, indicates that many of the perceived risks are remote. As a consequence, guidelines should form a flexible structure in order to follow for adaptation in light of learned experience.

Risk management

The type of risk management for contained use and planned introductions of organisms with novel traits depends on the organism involved and the intended application. The process involves reviewing alternatives and selecting the most appropriate regulatory actions based on the findings of the risk assessment. Measures to be taken to minimize risk include physical and biological containment. Questions to be posed include:

- What are the risks?
- How probable is it that they will occur?
- How serious is the damage if they occur?
- What can be done to minimize the risks and contain the damage?
- Do the benefits outweigh the risks?

Containment

The term *containment* is used to describe safe methods for maintaining control over the distribution of organisms with novel traits in the laboratory and in the environment into which they are introduced. The purpose of containment is to minimize unnecessary exposure of laboratory workers and the environment to potentially hazardous organisms.

Biological containment of microorganisms principally involves the use of specific combinations of vector and host in such a way that the probability of transfer of a vector to an unintended host and subsequent survival of the host-vector combination in the environment is limited. The growth of plants, which require special environmental conditions for their survival (for example, biological containment), can be achieved in either greenhouse or field. Similar results can be obtained with studies using contained animal facilities.

Physical containment involves physical constraints on the movement of organisms of uncertain risk or potential hazard. The aim of physical containment is to prevent inappropriate exposure of humans and the environment to organisms. Physical containment is achieved by following the principles of good laboratory practice, occupational safety, and hygiene; by involving well-qualified and competent personnel who follow safe, standard procedures; and by having a working environment designed to prevent the unintended spread to the environment of organisms with novel traits.

Biosafety levels

Biosafety levels are described as a series of constraints on the handling and dissemination of organisms graduated according to the level of potential risk. Different biosafety levels are reached by different combinations of laboratory facilities appropriate for the operations performed,

and the hazards posed by different organisms. Four biosafety levels have been defined, depending on the characteristics of the organisms involved. The proposed safety levels for work with organisms with novel traits take into consideration the results of the risk assessment described above. For microorganisms these levels and conditions are summarized below. Similar levels and conditions have been established for transgenic plants and animals.

- Biosafety level one* requires safety equipment and facilities as appropriate for undergraduate and secondary training laboratories and is suitable for work with strains of viable organisms not known to cause disease in humans, animals or plants.
- Biosafety level two* is similar but includes specific personnel training, limited access to the laboratory and physical containment facilities, and is suitable for work involving moderate potential hazards to personnel and the environment.
- Biosafety level three* is suitable for work with indigenous or exotic agents that may cause serious or potentially lethal disease as a result of exposure.
- Biosafety level four* is required for work with those agents that pose a high individual risk of life-threatening disease.

Biosafety levels three and four are characterized by additional safety measures involving, among other things, further personnel training, strict working practices, qualified supervision, and strict physical containment in specially designed facilities and buildings.

An enabling environment for biotechnology

The first step for a government in creating a suitable environment to realize the potential of biotechnology, to improve conditions and services for humankind, and to mitigate concerns about potential adverse effects to human, animal and environmental safety is to provide a regulatory framework that ensures safe development of biotechnology products in a timely and effective manner. Such principles might include:

- Regulations based on the characteristics of the product
- Science-based risk assessment
- Protection of health and the environment
- Building on existing legislation and experience.

This approach recognizes and builds on the knowledge, expertise, and infrastructure already existing in regulatory areas. Therefore, it is both economically and scientifically sound because it allows regulators to build on existing knowledge and experience and to incorporate additional experiences gained through the application of the regulations to biotechnology products.

Every product goes through several stages of development before it may be introduced for testing in the environment, and before it is licensed, approved, or registered for commercial sale and use. These may include:

- Import
- Laboratory research
- Environmental release
- Determination of product safety under relevant legislation
- Commercial release and use.

According to the nature of the product and its developmental stage, appropriate health and environmental considerations may be regulated under different legislative acts, together with specific regulations and published guidelines, and administered by different agencies. Identification of relevant legislation and responsibilities for its administration by different agencies and the construction of an effective coordinating mechanism at the national level are therefore the first steps in developing an enabling environment for safe use of biotechnology. The number of relevant acts and regulatory agencies varies from country to country, but there is a general need to identify the acts, agencies, and guidelines that relate to the different products of biotechnology and to define the nature of such products.

Given that biotechnology is the application of science and engineering to the direct or indirect use of living organisms (or parts or products of living organisms) in their natural or modified forms, such organisms with novel traits are the products of a process whereby their genetic material is altered by both natural processes and genetic engineering. Legislation to regulate and control the safety of such genetic modification is generally divided into two main areas: (a) the construction and contained use of organisms with novel traits and (b) the deliberate release of organisms with novel traits or the marketing of products containing them.

Therefore, there is a need to establish at the national and institutional levels appropriate policies, regulations, and enforcement mechanisms for the control of introductions, methods for field-testing, export and import, and commercial releases of organisms with novel traits. Furthermore, the potential of biotechnology for creating diverse products calls for an assessment of the product's safety and efficiency and the application of standards and regulatory measures based on public health, food safety and quarantine laws. Existing regulations should be suitably strengthened to render them effective to encourage the use of new products and to ensure human health and environmental safety.

Appropriate regulatory reviews for organisms with novel traits are carried out by a variety of officials in various ministries and agencies in different nations. An effective means of coordinating the various regulatory components, including the development and application of appropriate guidelines, is through the establishment of a single national technical advisory committee, with responsibility for both areas. Some countries, however, have a technical committee for each area. The Australian experience is most useful in this regard in which the Genetic Manipulation Advisory Committee, the equivalent of a national biosafety committee, was

established in 1987 and reported to the minister for administrative services. This was followed by the formation of institutional biosafety committees and the definition of the responsibilities of individual investigators and producers. This straightforward system requires the minimum in terms of new institutions and organizations and recognizes that adapting existing systems and approaches is the easiest way to quickly set up practical review mechanisms and tap existing relevant expertise, which in most countries is substantial. The system lends itself to countries seeking to develop their own regulatory process.

National Biosafety Committees

National biosafety committees (NBCs), supported by scientific subcommittees, are typically responsible to the appropriate minister(s) to:

- Maintain an overview of the biosafety factors associated with innovative genetic manipulation techniques.
- Identify and keep under review classes of work that have undefined risk levels.
- Alert national authorities to the existence of novel risk factors.
- Provide specialist technical advice on specific biosafety matters to organizations and regulatory agencies using these techniques.
- Prepare or assist with the preparation of codes, standards, or guidelines for the assessment and management of biosafety risk factors, whether for the committee to oversee activities or to assist regulatory agencies. These codes, standards and guidelines are intended to assist countries in benefiting from the products of biotechnology without undue risk to human health and the environment.
- Participate in public discussions about the biosafety of these techniques.
- Meet with overseas agencies to ensure that, as far as practicable, national guidelines and regulations are in harmony with international practice.

It is then the responsibility of all organizations and individual engaged in research, development, introductions or application of organisms with novel traits to familiarize themselves and comply with the relevant portions of such guidelines at each stage of their work.

Compliance with biosafety guidelines does not relieve those who wish to use organisms with novel traits from any other obligations, which they may have under legislations dealing with particular types or uses of products. The introduction of new products for trade and commerce will continue to come under the jurisdiction of ministries with relevant sectoral responsibility. It is the responsibility of those engaged in particular fields of work to ascertain and abide by the relevant legal requirements.

Terms of reference for a National Biosafety Committee

In regard to a government's need for disseminating technical and biosafety advice to ministers and other appropriate governmental authorities on the continuing assessment of the risks

and benefits associated with the production and application of biological materials produced in laboratories and which occur in nature, a national biosafety committee shall:

- Establish and review, as necessary, guidelines for physical and biological containment and control procedures appropriate to the level of assessed risk involved in relevant research, development and application activities.
- Review relevant proposals, except those that relate to research under contained laboratory conditions, and recommend any conditions under which this work should either be carried out or not be undertaken.
- Consult with relevant government agencies and other organizations as appropriate.
- Report to the minister and other responsible government authorities at least annually and promptly report any breaches of the above guidelines and any other relevant matters referred to them.
- Establish contact and maintain liaison with such monitoring bodies in other countries and with international organizations as appropriate.
- Advise on personnel training with regard to safety procedures as necessary.
- Collect and disseminate information relevant to the above, having due regard to proprietary information.
- Establish and oversee the work of a scientific subcommittee, whose guidelines follow and whose role and function include not only participation in the relevant items above but also in research performed under contained laboratory conditions.

In support of an NBC a scientific subcommittee shall:

- Be formed to support the work of the NBC. It shall enter into discussions directly with scientists and their host institutions and with funding bodies in determining the conditions under which research should be carried out.
- Review proposals for such research and recommend any conditions under which experiments should be carried out or that work not be undertaken.
- Provide technical advice to NBC and contribute to its functions.

In drawing up relevant codes, standards and guidelines, the NBC and its scientific subcommittee will take cognizance of the large body of expertise which exists worldwide on the development, content and application of such codes, standards and guidelines. These guidelines generally relate to large- and small-scale genetic manipulation work and to the introduction of organisms with novel traits or their products into the environment. More specific guidelines relating to the nature of the product are also often required (for example, the *Guidelines for Regulations of Veterinary Biologics* produced by Agriculture Canada).

Institutional Biosafety Committees

Institutional biosafety committees (IBCs) are essential to the overall monitoring and surveillance of genetic manipulation work and to the administration of the various guidelines. The caliber and experience of members on the IBC should be such that the IBC can competently carry

out its duties. The chair of the committee should be of sufficient standing in the organization to ensure that IBC decisions and advice are implemented effectively. Appropriate arrangements shall be made when the chair is on leave. The NBC will consider the advice and assessment of the IBC to be of fundamental importance in its decision making process.

The IBC shall ensure that staff recruited to work in laboratories, production facilities, or field trials in which work is conducted with organisms with novel traits are informed of potential hazards, have adequate training to ensure that their work is carried out under institutional guidelines, and have access to the IBC for advice. It shall inform new staff members of NBC and institutional guidelines and of the need to comply with them. Organizations of appropriate size conducting such work shall have a biological safety officer.

Composition

The IBC shall include:

- Staff with the requisite knowledge and expertise to assess, evaluate and oversee works being carried out in the institution.
- A biological safety officer
- An individual with expertise in testing biological safety facilities and equipment
- At least one person not associated with the institution who is in a position to consider the wider community interests.

A molecular biologist, a population biologist, or a geneticist shall be included among persons with requisite expertise. IBC membership shall include an ecologist with expertise relevant to the organism if introduction of a live modified organism is envisaged. In general, the scientific disciplines need only to be represented when work falling in that area is performed in the institution. For example, an institution working only on plants need not have an animal geneticist represented. Roles and responsibilities may be combined in the same person when appropriate.

Biological safety officer

It is recommended that institutions either appoint a biological safety officer or, in the case of smaller institutions or organizations, assign such duties to the IBC. The officer should have experience working with the containment conditions. The officer shall be adequately trained and be able to offer advice on, or participate in, training of new staff or laboratory personnel. Appropriate deputizing arrangements shall be made when the officer is on leave.

The biological safety officer or the IBC chair shall act as adviser to the head of the institution or firm in all matters relating to the containment, biological hazards and staff safety. Regular safety audits and the supervision of a regular testing program for appropriate pieces of equipment shall be undertaken by the biological safety officer.

Conflicts of interest

The composition of the IBC is such that they often include members with specialist expertise who originate proposals themselves. Project supervisors should not assess their own proposals as IBC members. It is necessary that the IBC have sufficient scientific members, or add them as required, so that it is not dependent on the advice of the person submitting the proposal. To ensure that no conflicts of interest arise, final decision on proposals shall be made in the absence of the originator.

Monitoring work

The IBC shall ensure that its advice and that of the NBC on specific proposals is conveyed to the principal investigator(s) and is acted upon. Members of the IBC shall inspect the laboratories, facilities or introduction sites from time to time to adequately monitor safety aspects of ongoing projects and production.

The IBC may draft whatever rules it considers appropriate to supplement the guidelines of the NBC or give effect to their intent. It shall have appropriate powers to ensure that the guidelines and rules are observed. Such rules may relate to containment procedures and operations for managing the project and to handling, transport and storage of transgenic plants and animals. The IBC will keep minutes of its discussions and decisions.

Duties

The main functions of the IBC relevant to planned work are to:

- Assess and review all the proposals it receives in order to identify potential hazards to personnel, the community and the environment and provide advice to the project supervisor on these hazards and their management.
- Ensure that all relevant data are included in the project proposal.
- Send the proposal to the NBC for review and assessment when required under the appropriate guidelines and ensure that NBC advice is complied with.
- Ensure that the appropriate institutional animal experimentation ethics committee has been consulted for proposals involving vertebrate animals.
- Monitor any changes to work within the organization and make recommendations to project supervisors from time to time.
- Review the qualifications and experience of personnel involved in projects to ensure that they are adequate for good professional practice and for the supervision of staff.
- Keep records for each planned introduction project.
 - Provide an annual report to the NBC.
- Submit a final report, if required under the appropriate guidelines to the NBC, at the end of the project.

Records

The organization shall ensure that records are kept of all procedures, decisions, and staff involved in respect to each project.

Reporting requirements

At the time of establishment the IBC is to provide the NBC with complete information on its composition and membership.

Accidents and Incidents

If unexpected results arise from a project, the IBC chair or the biosafety officer shall record accidents and any action taken. If the IBC chair believes that an accident or incident occurred that was directly attributable to work with genetic manipulation and was of sufficient concern, he or she shall make a report to the NBC and the head of the institution. An example of such an incident might be a deliberate failure to comply with the appropriate guidelines or any incident or accident that may have resulted in a risk to human health or the environment.

Sanctions

In general, governments require compliance with biosafety guidelines as a condition for funding research. Registration for tax incentives for private sector funding of research may also be conditional on compliance with the guidelines. A substantive breach of the guidelines may result in prosecution under either public health or environmental legislation.

The ultimate responsibility for safe use of biotechnology lies with those who practice the science in the process of research and production. The following section outlines these responsibilities.

Responsibilities of the Principal Investigator and Executive Officer

For each project a principal investigator (or, in the case of production, a production manager) will be identified who accepts full responsibility for all aspects of work. This individual should be thoroughly familiar with the requirements of the guidelines and should ensure that any project for which he or she is responsible involving the use of organisms with novel traits complies with the guidelines and, when appropriate, with NBC advice. In particular, he or she shall:

- Assess the proposal to determine if it falls within the guidelines. If in doubt, the project supervisor shall consult the IBC.
- Provide any information on the proposal and its conduct that the IBC may require for its assessment and monitoring activities.
- Comply with NBC and IBC advice and recommendations on proposals.

- Forward a new proposal to the IBC before any substantial change is made to the procedures or the organisms used in a project.
- Carry out work under the conditions approved by the IBC and as advised by the NBC. For work involving vertebrate animals prior and continuing approval must be received from the institutional animal experimentation ethics committee.
- Ensure that all workers, including subordinates and other co-workers, are aware of the nature of any potential hazards of the work and have received appropriate training in safety and emergency procedures.
- Notify all changes in the project team to the IBC.
- Report all unexpected results, accidents and unexplained illnesses or absences immediately to the IBC.
- Advise the IBC of any intention to import or transport biological material that falls under the guidelines.
- Keep records, as appropriate, for each project.
- Submit a report to the IBC at the end of the project.

CHAPTER 20

INTELLECTUAL PROPERTY RIGHTS: PATENTS IN BIOTECHNOLOGY

Tremendous advances in the life sciences over the three decades have made it possible to modify living organisms for human purpose more rapidly and efficiently than ever before. Expectations of massive profits have already catalyzed major changes in the structure of biological research. Wanting some guarantees of returns on their investments, researchers are increasingly seeking, and winning, patents on their living “inventions through biotechnology”. The rules are changing quickly and there are large discrepancies among countries in the kind and scope of protection offered.

Biotechnological research is increasingly dictated by commercial interests. Strengthened patent protection is closely linked with this trend. This has serious implications for agricultural and other biological research. Scientific informations, even plant and animal varieties, will become increasingly privatized. The resources needed by scientists to develop more productive and efficient agricultural systems involving biotechnological tools are becoming private property. This necessitates the need for patenting of techniques or processes, products and the living organisms under the broader issue of genetic engineering.

Biotechnology: Science of Sciences

The life sciences are changing in their fundamental character, and at a rapid rate. These changes are of two principal kinds. First tremendous technical advances have been realized over the past couple of decades. It is now possible and indeed, is common practice to transfer genetic material between completely dissimilar organisms: from fish to plant, from microorganisms to animals, from humans to other organisms and theoretically from other organisms to humans. It is also possible to isolate and to multiply for commercial use parts of organisms to, for example, mass-produce chemicals that are otherwise produced in much smaller quantities by plants. For instance, vanilla, an expensive flavouring that comes from bean of vanilla orchid, can now also be produced in a vat from masses of vanilla orchid cells. These technical advances have served to blur the distinctions between natural and man-made, between life and chemistry, and between living and nonliving.

The second change, intimately connected with these scientific breakthroughs (to the extent that it is difficult to separate cause and effect) is a strong and escalating trend toward the commercialization of the life sciences. There are large profits to be made. Spurred on by advances in the science, investment capital has been mobilized in a way unprecedented in the history of biological science.

The hype surrounding the new “biotechnologies,” much of it generated in the interests of stimulating investment has led to strong political support. An obvious prerequisite for investment is some assurance that the investment will be rewarded. In biotechnology the investment is in

research and development, and the assurance needed is some form of intellectual property protection.

Patenting Life Forms

The concept of intellectual product as property, that the benefits to creativity and ideas should be protected, is not a new one. As far back as 300 B.C. cooks were granted exclusive rights to prepare any peculiar and excellent dish in order that others might be inducted to labor at excelling in such pursuits. Letters of patent were granted by monarchs in the Middle Ages giving exclusive rights to merchants to sell commodities and inventions (later limited to inventions). The modern patent system traces directly back to a statute from the Republic of Venice. 19th March 1474. The concept of intellectual property protection governing “living inventions” is much newer, however. The first example of patents being applied to living organisms occurred in 1873 when Louis Pasteur was granted patents in the United States on some yeast strains that were “free from organic germs” but institutionalized mechanisms for the protection of plant varieties only really began in the 1930s, and that outside the patent system.

Patenting Products or Processes

Patent is intended to protect a particular product or a process that is the result of inventive thought. The patent permits the holder to forbid commercial exploitation (use, sale, manufacture) of the protected product or process by others in the country or countries where the patent is granted for a limited period (normally 17-20 years, but the periods varies by country and product). Three specific conditions of eligibility must be met for patents. These conditions are fairly standard between countries, and include the following:

- Novelty – the invention must be new
- Utility – it must be useful
- Inventiveness (or non-obviousness)-it must represent a real advance that might not have been reached without the inventor's creative insight.

The patent application demands that the invention be disclosed in a way that enable the skilled public to reproduce it. The scope of the protection granted is proportional to the degree of inventiveness.

Finally, the patent must relate to a technology for which patents are permitted. This is the area in which the most variation occurs between countries patent legislation. Usually, the invention must be capable of industrial application-for example, an industrial process or product, and not merely an idea, discovery, artistic work, or business scheme. Each country has the responsibility to decide on the appropriate scope of patentability to suit its own particular socioeconomic situation.

Such things as food, drugs, and agriculture have at various times and places been excluded to keep prices down (to make it cheap and easy to imitate and adapt existing

technologies and products). Most developing countries currently do not grant patents on pharmaceuticals for this reason, similarly, some prohibit patents on agricultural innovations. Rapid changes, however are occurring in the scope of patentability, with an evident trend to encompass all products and processes, including living organisms.

Forms of Patents

Patents are available of processes used to develop modified organisms or to produce biological products, such patents often described as “process patents”. In general fall within the category of procedures for which patents were designed and are not particularly controversial. Arguments against such patents, which are sometimes raised, are more correctly arguments against biotechnology *per se*.

In the U.S., patents have been granted for specific plant and animals varieties. These varieties could be protected using plant breeders' rights but, claiming a degree of inventiveness, patent protection has been granted. This seems to give the patent holder the authority to restrict use of the patented variety for breeding purpose, protection not offered by plant breeders' rights.

Patent protection is also available in a number of countries for plants that contain a novel gene. For a gene to qualify as something “not found in nature” it must be either novel in and of itself (*i.e.* created by the inventor), or transferred to a species in which it is not found in nature. Such patents (on genes) seem to imply that the holder of the right could prohibit others from engaging in unauthorized commercial activity involving any plant material of the protected species. Protection may extend to cover closely related species to which the transferred gene could be moved using conventional breeding techniques.

Finally, and most controversial, the U.S. patent and trademark office has granted a patent on a plant characteristic. The American Biotechnology Company, Molecular Genetic Inc. developed a variety of maize that produces high levels of the amino acid tryptophan. The company's patent claims a monopoly over any high, tryptophan-producing maize regardless of the process by which this characteristic is achieved. This broad claim remains to be challenged in court. With similar broad claims in the area of chemical production, the courts have tended to find evidence of infringement only when the product in question has been made by the process described in the patent.

Patent Types in Biotechnology

At least four types of patens are currently being issued in the biotechnology area:

One is for innovations typical of other areas of invention. For example, in the hybrid rice area, there are patents on procedures for ensuring that the flowering of female parents are synchronized and on procedures for shaping plants and their leaf systems to increase the likelihood of achieving the desired cross-pollination.

A second type is for varieties, and is being issued in circumstances under which PVP would be available. These patents (which are unavailable in Europe) are currently issued in the United States on inbred lines and hybrids, and are intended to protect the developers of inbred lines against reconstruction of those lines by others. In light of their coverage of the same type of innovation as the PVP system, their validity (which has not yet been tested) is subject to some question.

A third type covers a wide group of plants. It is exemplified by the Agracetus patent on all transgenic cotton (a patent being re-examined by the US Patent and Trademark Office). The obvious problem for the patent examiner with such applications is where to draw the line when the applicant has demonstrated a particular way to put a specific gene in a specific line -and at the time of application it may be rather unclear whether the method will work with all lines of the species.

A final type is also quite broad and is exemplified by patents on research tools. These have become quite significant in the medical sector where, for example, those who have sequenced specific receptors have sought patents claiming the right to use the sequenced protein in drug discovery processes. It is probably only a matter of time before there are similar patents in the agricultural biotechnology area. Table 1 provides common types of patent categories, with examples from recombinant DNA technology

Patents and Concerns of Developing Countries

Many of the concerns over strengthened intellectual property protection discussed in the foregoing are magnified in the cases of developing countries. Public research systems are weak and private-sector agricultural research is virtually nonexistent in many of the developing countries. The ethical issues may be more pronounced, considering the vastly different cultures, and the costs of administration would be significant.

For a developing nation, the central questions (other than those imposed by international law and international politics) in deciding whether to adopt an intellectual property system involve the size of the local market and the size of the local research establishment. Unless the market is large enough to be interest to inventors, and unless there is a local industry strong enough to be worth motivating, there is no value in an intellectual property system.

Plant Variety Protection (PVP) is adequate to support a local sector based on conventional breeding. With a biotechnology sector, a stronger system that allows, for example, a claim on all plants containing a specific gene (inserted by genetic engineering) is desirable. In any case, a developing nation is wise to insert a strong research exemption.

Common types of patent categories, with examples from recombinant DNA technology

Categories	Examples
Product patents	
Substance	Cloned genes, recombinant proteins, monoclonal antibodies, plasmids, promoters, vectors, cDNA sequences and monovalent vaccines
Compositions of matter	Multivalent vaccines, biofertilizers, bioinsecticides, pharmaceutical mixtures, microorganisms and transgenic organisms
Devices	Pulsed Field Gel Electrophoresis apparatus, DNA sequencing apparatus, Microparticle gene gun and Thermalcyclers
Process patents	
Process of preparation	DNA isolation, synthesizing double stranded DNA, vector-insert construction, polymerase chain reaction applications and purification of recombinant protein
Method of working	Nucleic acid hybridization assays, diagnostic procedures, detection systems using PCR and mutant assays
Use	Applying biofertilizers and bioinsecticides, fermentation of genetically modified organisms and nontherapeutic animal treatment systems

The creation of a patent law is just a beginning. It is also essential to build a patent office (which may be done cooperatively with developed nations or other developing nations). It is essential to have a patent bar that can advise inventors, assist in enforcing the law, and assist in negotiating license agreements. Moreover, it is important to have government policies with respect to protection of government inventions and their transfer to the private sector. Finally, there must be arrangements for allocating patent rights between employees and employers.

Intellectual property law in and of itself is far from adequate to provide effective protection. A variety of components must be in place to support the law, including a legal system, and a political and economic system that is conducive to private business and to the protection of private property in general.

Patents in Biotechnology: Litigation Issues

Questions of patent scope are problematic in biotechnology because many biotechnology inventions will still work even after a competitor modifies them somewhat from the form in which the patentee initially created them. For example, minor changes in DNA sequence, or amino acid sequence of a protein, or use of a different recombinant vector or host organism, may still yield a functional competitive product. If these changes are enough to bring the competitive product outside the patent monopoly, the value of many biotechnology products will be severely limited.

There are now a number of decisions on the validity of broad patent claims, and on the proper construction of biotechnology patent claims in infringement actions, that shed some light on this question of proper patent scope and take a somewhat restrictive approach.

Scripps Vs Genentech , decided by the Federal Circuit in 1991, was an infringement action brought by the holder of a patent on a purified protein obtained from plasma, against a firm that was making a recombinant version of the protein. The Federal Circuit rejected the argument to the plasma-derived protein, but suggested that this might be an appropriate case for application of the rarely used “reverse doctrine of equivalents.”

The Federal Circuit also addressed issues of patent scope in Amgen Vs Chugai, a case involving patent rights in erythropoietin (EPO). Amgen’s patent on the DNA sequence included broad, generic claims attempting to cover all possible DNA sequences that will encode any polypeptide having an amino acid sequence sufficiently duplicative of EPO to possess the property of increasing the production of red blood cells. The Federal Circuit held that these broad claims were invalid because they extended to millions of different sequences and were inadequately supported by Amgen’s disclosure, which explained how to make only a small number of analogs of the EPO gene. In separate litigation between the same parties, the Federal Circuit held that Amgen’s patent claims on DNA sequences and recombinant vectors and host organisms were not infringed by the importation into the US of recombinant erythropoietin made abroad through use of these patented starting materials. This decision has caused considerable controversy within the US biotechnology patent community and has triggered proposals for statutory reform.

The Federal Circuit recently took a restrictive approach to application of the doctrine of equivalents in a biotechnology patent infringement lawsuit in Genentech Vs Wellcome Foundation. Genentech and its co-plaintiffs held several patents relating to tissue plasminogen activator (TPA), including a patent directed to the natural protein extracted from human cells, a patent directed to the starting materials needed to make the protein recombinantly, *i.e.* the DNA sequence encoding the protein, the expression vector containing the sequence, and the microorganism or cell culture expressing the protein, and the process of producing the protein through recombinant DNA technology. The defendants Wellcome and Genetics Institute, used recombinant DNA technology to produce two proteins that were structurally distinct from natural TPA. The issue was the extent of the plaintiffs to produce two proteins that were structurally distinct from natural TPA. The issue was

the extent of the plaintiffs patent rights to cover products with structurally differences that had some similar biological activity. The trial court found that the accused products did not literally infringe the patent claims and submitted to the jury the question of whether they infringed under the doctrine of equivalents. The jury found for the plaintiffs, and the Federal Circuit reversed, in effect restricting the plaintiffs patent rights to the particular structures disclosed in their patent specification.

With the advent of the Human Genome Project and in particular with the undertaking of the partial sequencing of thousands of human cDNA molecules from different tissues and organs, the patenting of partial DNA sequences become extremely contentious. In 1991, the issue of patenting gene fragments was broached when US National Institutes of Health (NIH) filed for the patent rights of 315 partially sequenced human cDNAs (expressed sequence tags [EST]). Two additional filings brought the total number of partial sequences to 6,869. In 1994, in a preliminary ruling, the US-PTO notified NIH that it would reject the patent applications on the ground that the functions of the sequences were not known.

Putting these decisions on biotechnology patent scope together, it is understood that a growing body of law that restricts the potential scope of biotechnology patents.

Patents in Biotechnology: Restriction or No Restriction to Research

Intellectual Property Rights (IPR) in biotechnology are having dramatic negative effect on the progress of non-profit research. Although patent policy and law intend to facilitate research progress, the current practice has led to many barriers in access and use of genetic materials and DNA technology. Such barriers restrict the free exchange of information and threaten the health of our nation's scientific enterprise. Policies related to patents and licensing are increasing the cost of research at a time when funds for research are being reduced. But many of the patented technologies in the field of biotechnology are available for non-profit research by the scientific organizations. For example, the RAPD marker technology developed by DuPont Co. scientists has exclusive rights to use the technique. DuPont has decided not to charge fees for research carried out in universities or government laboratories that "has no commercial purpose". This distinction deserves consideration; it could set a precedent for protecting freedom of inquiry.

Patenting: The Indian Scenario

The Patents Act, 1970 provides opportunity for Indian scientists to have the exclusive rights for their inventions in India and its six Convention Countries viz., United Kingdom, New Zealand, Eire, Sri Lanka, Canada and Australia. If an Indian applicant, after filing the application in India, desires to file a corresponding application in any one or more of the above-mentioned countries, he can do so within 12 months from the date of filing the India application. The Indian Patent Act, 1970 has following salient features:

- The Indian Law provides only "process" patents in food, pharmaceuticals and chemical sectors while the GATT requires "product" patents in all branches of technology.

- In the Indian Law the burden of proof on "process" patents lies on the complainant while the GATT places it on the defendant.
- The duration of the patent in the India Law is 7 years in case of food and pharmaceutical sectors and 14 year in case of all other sectors but the GATT provides for a uniform duration of 20 years for all patents.
- The India Law provides for automatic compulsory licensing in case of food, pharmaceutical and chemical sectors without the patent holder being heard while GATT permits compulsory licensing on the merits of each case but the patent holder will have be heard.
- There is no system for protection of plant varieties in India but the GATT requires an effective "*sui-generis*" system for protection of plant varieties.
- The India Law does not allow patenting of life forms but the GATT requires microorganisms to be patented.

The Indian Patent Act of 1970, did not have the system of patenting "products" in the pharmaceutical sector. Only process patents were recognized. By virtue of this, Indian Pharmaceutical Industry was able to adopt different processes and thus manufacture a product, which was earlier invented elsewhere in the world. This helped in various drugs being available in one therapeutic group, and thereby, also helped in making the prices competitive and cheap. This system has worked well, in as much as today, the country is almost self sufficient in the production of formulations, and import mainly the latest generation of drugs.

But a fallout of this system has been that it has not encouraged investments on basic Research and Development (R&D) in the Indian Pharmaceutical Sector, as desired. It is understood that while on an average, companies spend 15-16 per cent of their sales turnover on R&D in the U.S. and Europe, in India, it is between 1-2 per cent only.

Future Focus

The country is on the threshold of a new patent regime, which would be effective for the pharmaceutical sector only from January 1, 2005, that is after completion of the ten years' transitory period, under the TRIPs. This will signal the beginning of a new chapter, as patents on products also, in addition to patents on processes, in the biotechnological sector including pharmaceutical sector would be granted. This will necessitate, the industry to innovate and invent and not do mere reverse engineering. R&D will have to be taken up in the right earnest. Researchers have to apply their creative skills to producing new inventions. The biotechnology industry will have to focus its resources and efforts on value-added innovation and development of new products.

The modern biotechnology sector is highly scientific and involves skill and innovation. Of all the forms of biotechnology, genetic engineering holds the greatest potential for solving the problems of industry. Under the present condition of TRIPs agreement, the following areas in biotechnology can be targeted for making IPRs.

- Development of patentable techniques for the mass production of ornamentals, tree species and medicinal plants through tissue culture.
- Exploitation of microorganisms for producing patentable processes, products and life – forms (e.g *Spirulina*).
- Identification specific gene sequences exploiting the wealth of biodiversity in India, before some one patents.
- Synthesis of Genetically Modified Organisms (GMO) including transgenic crops to suite the needs of Indian subcontinent.

India has a vast R&D infrastructure that includes the extensive network of national labs, academic institutes and private R&D labs. It is essential to utilize the expertise of these institutions and make them contribute to the development of industry and the national economy. The mettle of Indian scientists is recognized the world over, and many research laboratories in advanced countries have benefited from our national resources that could not be effectively tapped, so far. Further, research in India is far more cost-effective than it is in the west. Therefore, we could well become a source for scientific innovation. The key lies in identifying the right opportunities, matching them with our strengths and relentlessly pursuing the path of making India a preferred supplier of intellectual property in the area of biotechnology research & development and global leader in the sector. It is possible to leverage this factor and find ways of converting it into a source of competitive advantage. What is needed is a more focused attention.

GLOSSARY ON GENETIC ENGINEERING

β-Galactosidase An enzyme encoded by the *lacZ* gene. Splits lactose into glucose and galactose

3' extension A short single-stranded nucleotide sequence on the 3'-hydroxyl end of a double-stranded DNA molecule. Also called 3' protruding end, 3' sticky end, 3' overhang.

5' extension A short single-stranded nucleotide sequence on the 5'-hydroxyl end of a double-stranded DNA molecule. Also called 5' protruding end, 5' sticky end, 5' overhang.

5'-phosphate end The phosphate group that is attached to the 5'carbon atom of the sugar (ribose or deoxyribose) of the terminal nucleotide of a nucleic acid molecule.

Abundance class Refers to the relative abundance of different mRNA molecules in a cell at any given time.

Activation Enhancing the rate of transcription

Activator (1) A substance or physical agent that stimulates transcription of a specific gene or operon. (2) a protein that binds to an operator and enhances the rate of transcription. Also called activator protein

Activator site A DNA sequence to which an activator protein binds. Also called activating site.

ADA Adenosine deaminase, deficiency results in SCIDS (q.v)

Adaptor (1) A synthetic double-stranded oligonucleotide that is blunt ended at one end and at the other has a nucleotide extension that can base pair with a cohesive end created by cleavage of a DNA molecule with a specific type II restriction endonuclease. After blunt-end ligation of the adaptor to the ends of target DNA molecule, the construct can be cloned into a vector by using the cohesive ends of the adaptor. (2) A synthetic single-stranded oligonucleotide that, after self-hybridization, produces a molecule with cohesive ends and an internal restriction endonuclease site. When the adaptor is inserted into a cloning vector by means of the cohesive ends, the internal sequence provides a new restriction endonuclease site.

Adenine (A) Nitrogenous base found in DNA and RNA.

Adeno-associated virus Virus used in gene therapy delivery methods

Adenovirus Virus that can infect through nasal passages, used in gene therapy delivery methods.

Aerobe A microorganism that requires oxygen for growth.

Aetiology Of disease; relating to the causes of the disease

Agarose Jelly-like matrix, extracted from seaweed, used as a support in the separation of nucleic acids by gel electrophoresis.

Agrobacterium tumefaciens Bacterium that infect plants and causes crown gall disease (q.v.). Carries a plasmid (the Ti plasmid) used for gene manipulation in plants.

Airlift fermentor A cylindrical fermentation vessel in which the cells are mixed by air that is introduced at the base of the vessel and rises through the column of culture medium. The cell suspension circulates around the column as a consequence of the gradient of air bubbles in different parts of the reactor.

Alginic acid A polysaccharide polymer produced by different seaweeds and bacteria that is comprised of β -D-mannuronate and α -L-guluronate.

Alkaline phosphatase An enzyme that removes 5' phosphate group from the ends of DNA molecules, leaving 5' hydroxyl groups.

Allele Alternative form of a gene. Alleles of a specific gene occupy the same location on homologous chromosomes.

Allele-specific oligonucleotide Oligonucleotide with a sequence that can be matched precisely to a particular allele by using stringent hybridization conditions.

Allelic frequency The ratio of the occurrence of one particular allele at a locus to the occurrence of all of the alleles of the locus in a large number of individuals of a population.

Alpha-peptide Part of the β -galactosidase protein, encoded by the lacZ' gene fragment

Alternative splicing Cell-specific removal of an exon(s) during processing of a primary transcript that leads to a functional mRNA.

Amino acid A building block of a protein.

Aminoacyl site The portion of a ribosome where the anticodon-codon interaction takes place during translation. Also called an A site.

Aminoacyl-tRNA A charged tRNA; a tRNA with its specific amino acid attached to its 3' end.

Ampicillin (Ap) A semisynthetic β -lactam antibiotic

Amplicon A herpes simplex virus type 1 plasmid vector. Also called amplicon plasmid.

Amylolytic Agents that are capable of breaking down starch into sugars.

Anaerobe A microorganism that grows in the absence of oxygen.

Aneuploidy Variation in chromosome number where single chromosome are affected, thus the chromosome complement is not an exact multiple of the haploid chromosome number.

Animal model Usually a transgenic mouse in which a disease state has been engineered. See also *knockout mouse*, *knockin mouse*, *oncomouse*.

Annealing The process of heating (denaturing step) and slowly cooling (renaturing step) double-stranded DNA to allow the formation of hybrid DNA or DNA-RNA molecules.

Antibiosis The prevention of growth or development of an organism by a substance or another organism.

Antibiotic A biological substance that is produced by one organism and that can inhibit the growth of or kill another organism.

Antibody An immunoglobulin that specifically recognises and binds to an antigenic determinant on an antigen

Anticodon A set of three contiguous nucleotides in a tRNA molecule that are complementary to a set of three contiguous nucleotides (codon) in an mRNA.

Antigen A molecule that is bound by an antibody. Also used to describe molecules that can induce an immune response, although these are more properly described as immunogens.

Antigenic determinant See epitope

Antiparallel orientation The arrangement of the two strands of a duplex DNA molecule, which are oriented in opposite directions so that the 5'-phosphate end of one strand is aligned with the 3'-hydroxyl end of the complementary strand.

Antisense RNA An RNA sequence that is complementary to all or part of a functional RNA. Produced from a gene sequence inserted in the opposite orientation, so that the transcript is complementary to the normal mRNA and can therefore bind to it and prevent translation.

Arabidopsis thaliana Small plant favoured as a research organism for plant molecular biologists

Arbitrarily primed PCR PCR using random primers, useful in the technique of RAPD analysis

ARS Autonomously replicating sequence. Any cloned DNA sequence that initiates and supports extrachromosomal replication of a DNA molecule in a host cell; often used in yeast cells. Also called autonomous(ly) replicating segment, autonomously replicating sequence.

ASO See allele-specific oligonucleotide

Attenuated vaccine a virulent organism that has been modified to produce a less virulent form but nevertheless retains the ability to elicit antibodies against the virulent form.

Autoradiograph Image produced on X-ray film in response to the emission of radioactive particles

Autoradiography A technique that captures the image formed in a photographic emulsion as a result of the emission of either light or radioactivity from a labeled component that is placed next to unexposed film.

Autosome A chromosome that is not a sex chromosome

Auxotroph A cell that requires nutritional supplements for growth

BAC See Bacterial artificial chromosome.

Bacillus thuringiensis Bacterium used in crop protection, and in the generation of Bt plants that are resistant to insect attack. The bacterium produces a toxin that affects the insect

Bacmid a shuttle vector based on the AcMNPV genome that can be propagated in both *E. coli* and insect cells.

Bacterial alkaline phosphatase (BAP) See *alkaline phosphatase*

Bacterial artificial chromosome A vector system based on the *Escherichia coli* F factor plasmid that is used for cloning large (100 to 300 kb) DNA inserts.

Bacteriophage A virus that infects the phage. Also called phage.

Baculovirus A particular type of virus that infects insect cells, producing large inclusions in the infected cells.

Bal 31 nuclease An exonuclease that degrades both strands of a DNA molecule at the same time.

Base pair substitution Permanent replacement in chromosomal DNA of a nucleotide pair with another nucleotide pair.

Base pair The term *base pair* represents the complementary nucleotides; in DNA adenine (A) is hydrogen bonded with the base thymine (T) and guanine (G) is hydrogen bonded with cytosine (C). A thousand base pairs is often called a kilobase (kb).

Batch fermentation A process in which cells or microorganisms are grown for a limited time. At the beginning of the fermentation, an inoculum is introduced into fresh medium, and no addition or removal of medium is done for the duration of the process.

Bifunctional vector See Shuttle vector.

Binary vector system A two-plasmid system in *Agrobacterium* for transferring a T-DNA region that carries cloned genes into plant cells. The virulence genes are on one plasmid, and the engineered T-DNA region is on the other plasmid.

Bioaccumulation Concentration of a chemical agent (e.g., DDT) in increasing quantities in the organisms of a food chain.

Biodegradation The breakdown of a compound to its chemical constituents by living organisms.

Bioinformatics The emerging discipline of collating and analysing biological information especially genome sequence information

Biolistic Refers to a method of introducing DNA into cells by bombarding them with microprojectiles, which carry the DNA

Biolistics Delivery of DNA to plant and animal cells and organelles by means of DNA-coated pellets that are fired under pressure at high speed. Also called microprojectile bombardment.

Biomass (1) The cell mass produced by a population of living organisms. (2) The organic mass that can be used either as a source of energy or for its chemical components.

Biomass concentration The amount of biological material in a specific volume.

Biopolymer Any large polymeric molecule (protein, nucleic acid, polysaccharide, lipid) produced by a living organism.

Bioreactor A vessel in which cells, cell extracts, or enzymes carry out a biological reaction. Often refers to a growth chamber (fermenter, fermentation vessel) for cells or microorganisms.

Bioremediation a process that uses living organisms to remove contaminants, pollutants, or unwanted substances from soil or water.

Blunt end The end of a DNA duplex molecule in which neither strand extends beyond the other. Also called flush end.

Blunt-end cut To cleave phosphodiester bonds in the backbone of duplex DNA between the corresponding nucleotide pairs on opposite strands. This cleavage process produces no nucleotide extensions on either strand. Also called flush-end cut.

Blunt-end ligation Joining (ligation) of the nucleotides that are at the ends of two DNA duplex molecules, neither of which has an extension.

Bovine somatotropin (BST) Bovine growth hormone, produced as rBST for use in dairy cattle to increase milk production.

Box A short DNA sequence that plays a role in regulating, facilitating, enhancing or silencing transcription.

Broad-host-range plasmid A plasmid that can replicate in a number of different bacterial species.
Bt plants Plants which carry the toxin-producing gene from *Bacillus thuringiensis* as a means to protect the plant from insect attack.

C terminus The last amino acid of a protein. Sometimes denotes the final amino acids of a protein. Also called carboxy(l) terminus, carboxy(l)-terminal end.

CAAT box A sequence located approximately 75 base pairs upstream from eukaryotic transcription start sites. This sequence is one of those that enhance binding of RNA polymerase

Caenorhabditis elegans A nematode worm as a model organism in developmental and molecular studies.

Calf intestinal Phosphatase (CIP) See *alkaline phosphatase*

Candidate gene A coding sequence that has some characteristics, including chromosome location and function, that make it likely that it could be responsible for a specific genetic disorder.

Candidate gene cloning A strategy for isolating a disease gene that is based on an informed guess about the possible gene product.

Cap A chemical modification that is added to the 5' end of eukaryotic mRNA molecule during post-transcriptional processing of the primary transcript

Capsid A structure that is composed of the coat protein(s) of a virus and is external to the viral nucleic acids. The capsid often determines the shape of the virus.

Carrier In genetics, an individual that has one mutant allele and one normal allele of a gene and whose phenotype is normal.

Cassette A combination of DNA elements that performs a specific function and is maintained as a clonable unit.

cDNA DNA that is made by copying mRNA using the enzyme reverse transcriptase.

cDNA clone A double-stranded DNA molecule that is carried in a vector and was synthesized in vitro from an mRNA sequence by using reverse transcriptase and DNA polymerase.

cDNA library A collection of cDNA clones that were generated in vitro from the mRNA sequence by using reverse transcriptase and DNA polymerase.

Centimorgan a unit of genetic distance between two gene loci, abbreviated cM. Named after T. H. Morgan (1866-1945), who used the fruit fly *Drosophila melanogaster* to establish the principles of genetic linkage and gene mapping. In humans 1 cM is approximately 10^6 base pairs of DNA.

Central dogma Statement regarding the unidirectional transfer of information from DNA to NA to protein

CFTR gene (protein) Cystic fibrosis transmembrane conductance regulator, the gene and protein involved in defective ion transport that causes cystic fibrosis.

Charged tRNA A transfer RNA molecule that is coupled to its specific amino acid. Also aminoacylated tRNA, amino acyl tRNA.

Chimaera An organism (usually transgenic) composed of cells with different genotypes.

Chromosomal integration site A chromosomal location where foreign DNA can be integrated, often without impairing any essential function in the host genomes.

Chromosome A DNA molecule carrying a set of genes. There may be a single chromosome, as in bacteria, or multiple chromosomes, as in eukaryotic organisms

Chromosome jumping Technique used to isolate non-contiguous regions of DNA by 'jumping' across gaps that may appear as a consequence of uncloned regions of DNA in a gene library.

Chromosome walking A technique that identifies overlapping cloned sequences that are about 40 kb long that form one continuous segment of a chromosome.

Chymosin (chymase) Enzyme used in cheese production, available as recombinant product.

Cis-acting element A DNA sequence that exerts its effect only when on the same DNA molecule as the sequence it acts on. For example, the CAAT box (q.v.) is a cis-acting element for transcription in eukaryotes.

Cistron A sequence of bases in DNA that specifies one polypeptide

Clone (1) A colony of identical organisms; often used to describe a cell carrying a recombinant DNA fragment. (2) Used as a verb to describe the generation of recombinants. (3) A complex organism (e.g. sheep) generated from a totipotent cell nucleus by nuclear transfer into an enucleated ovum.

Clone bank See cDNA library, genomic library

Cloning Incorporating a DNA molecule into a chromosomal site or a cloning vector.

Cloning site A location on a cloning vector into which DNA can be inserted.

Cloning vector A DNA molecule that can carry inserted DNA and be perpetuated in a host cell. Also called cloning vehicle, vector or vehicle.

Cloning vehicle See Cloning vector

Coding triplet A set of three contiguous nucleotides of the nontranscribed DNA strand of the coding region of structural gene that is complementary to a transcribed triplet.

Codon A set of three nucleotides in mRNA that specifies a tRNA carrying a specific amino acid which is incorporated into a polypeptide chain during protein synthesis.

Codon optimization An experimental strategy in which codons within a cloned gene that are not the ones generally used by the host cell translation system are changed to the preferred codons without changing the amino acids of the synthesized protein.

Codon usage The mean frequency of occurrence of each codon determined from a large sample of structural genes of an organism.

Cohesive ends Complementary single-strand excisions on the ends of duplex DNA molecules. Also called sticky ends. See also cos ends.

Cointegrate vector system A two-plasmid system for transferring cloned genes to plant cells. The cloning vector has a T-DNA segment that contains cloned genes. After introduction into *Agrobacterium*, the cloning vector DNA undergoes homologous recombination with a resident disarmed Ti plasmid to form a single plasmid carrying the genetic information for transferring the genetically engineered T-DNA region to plant cells.

Colony hybridization A technique that uses a nucleic acid probe to identify a bacterial colony with a vector carrying a specific cloned gene(s).

Combinatorial library During the ligation reaction with cDNAs of light and heavy antibody chains into a bacteriophage λ vector, many novel combinations consisting of one heavy and one light chain coding region are formed. The library comprises these combinations, each in a separate vector.

Competence The ability of bacterial cells to take up (usually plasmid) DNA molecules.

Competent Refers to bacterial cells that are able to take up exogenous DNA.

Competitor RT-PCR Technique used to quantify the amount of PCR product by spiking samples with known amounts of a competitor sequence.

Complementarity (1) One of a pair of nucleotide bases that form hydrogen bonds with each other. Adenine (A) pairs with thymine (T) (or with Uracil [U] in RNA), and guanine (G) pairs with cytosine (C). (2) One of a pair of segments or strands of nucleic acid that will hybridize (join by hydrogen bonding) with each other.

Complementary base pairs In double-stranded DNA, adenine forms hydrogen bonds with thymine and cytosine forms hydrogen bonds with guanine. In double-stranded regions of RNA molecules and in both RNA-RNA and DNA-RNA strand interactions, adenine forms hydrogen bonds with uracil and cytosine forms hydrogen bonds with guanine.

Complementary DNA See cDNA

Complementary homopolymeric tailing The process of adding complementary nucleotide extensions to different DNA molecules – e.g., dG (deoxyguanosine) to the 3'-hydroxyl ends of one DNA molecule and dC (deoxycytidine) to the 3'-hydroxyl ends of another DNA molecule – to facilitate, after mixing, the joining of the two DNA molecules by base pairing between the complementary extensions. Also called dG-dC tailing, dA-dT tailing.

Complementation Process by which genes on different DNA molecules interact. Usually a protein product is involved, as this is a diffusible molecule that can exert its effect away from the DNA itself. For example, a *lacZ⁺* gene on a plasmid can complement a mutant (*lacZ*) gene on the chromosome by enabling the synthesis of β-galactosidase.

Complete linkage Two or more adjacent gene loci on the same chromosome that are always inherited together. Also called absolute linkage.

Concatemer A DNA molecule composed of a number of individual pieces joined together via cohesive ends (q.v.).

Congenital Present at birth, usually used to describe genetically derived abnormalities.

Conjugation The unidirectional transfer of DNA from one bacterium to another, involving cell to cell contact.

Consensus sequence A sequence that is found in most examples of a particular genetic element, and which shows a high degree of conservation. An example is the CAAT box (q.v.).

Contig A set of overlapping contiguous clones that cover a chromosome region or a whole chromosome.

Continuous fermentation A process in which cells or microorganisms are maintained in culture in the exponential growth phase by the continuous addition of a fresh medium that is exactly balanced by the removal of cell suspension from the bioreactor.

Copy number (1) the number of plasmid molecules in a bacterial cell. (2) The number of copies of a gene in the genome of an organism.

Corepressor A low-molecular-weight compound that combines with an inactive repressor protein to form a complex that binds to an operator region and prevents transcription.

cos site The region generated when the cohesive ends of λ DNA join together

Cosegregation When two genetic conditions appear to be inherited together.

Cosmid A vector that uses the cos end sequences of bacteriophage λ and in vitro bacteriophage packaging to form, after injection of the vector into a host cell, plasmid that can carry as much as 45 kb of insert DNA.

Cosuppression The transformation of a plant with a gene in the sense orientation that the plant already possesses. This results in the down regulation of both the endogenous and introduced genes. Also called sense suppression.

Cotransfection The introduction of two different DNA molecules into a eukaryotic cell. In baculovirus expression systems, the procedure by which the baculovirus and the transfer vector are simultaneously introduced into insect cells in culture.

Coupling The phase state in which either two dominant or two recessive versions of two different genes occur on the same chromosome. Also called *cis* configuration. See also repulsion.

CpG islands Cluster of GC-rich regions that precede many transcribed vertebrate genes.

Crossover (1) The site of recombination. A single crossover represents one reciprocal breakage-and-reunion event. A double crossover requires two simultaneous reciprocal breakage-and-reunion events. (2) The reciprocal exchange of DNA between two chromosomes or DNA molecules by a breakage-and-reunion process. Also called recombination, recombination event.

Crown gall disease Plant disease caused by the Ti plasmid of *Agrobacterium tumefaciens*, in which a 'crown gall' of tissue is produced after infection.

Cryptic site A functional macromolecular sequence in an unlikely location. Also used, in some instances, to denote a macromolecular sequence whose function is unknown.

Cyanogen bromide Chemical used to cleave a fusion protein product from the N-terminal vector-encoded sequence after synthesis

Cystic fibrosis Disease affecting lungs and other tissues, caused by ion transport defects in the CFTR gene (q.v.).

Cytosine (C) Nitrogenous base found in DNA and RNA

Deletion Change to the genetic material caused by removal of part of the sequence of bases in DNA

Denatured DNA Duplex DNA that has been converted to single strands by breaking the hydrogen bonds of complementary nucleotide pairs.

Deoxynucleoside triphosphate (dNTP) Triphosphorylated ('high energy') precursor required for synthesis of DNA, where N refers to one of the four bases (A, G, T or C).

Deoxyribonuclease (DNase) An nuclease enzyme that hydrolyses (degrades) single- and double-stranded DNA

Deoxyribonuclease I See DNase I.

Deoxyribonucleic acid (DNA) A condensation heteropolymer composed of nucleotides. DNA is the primary genetic material in all organisms apart from some RNA viruses. Usually double-stranded.

Deoxyribose The 5-carbon sugar component of DNA.

Deoxyribozyme A DNA molecule that has catalytic activity.

Dideoxynucleoside triphosphate (ddNTP) A modified form of dNTP used as a chain terminator in DNA sequencing.

Diploid A cell or organism that has a set of all pairs of its chromosomes.

Directed mutagenesis The process of generating nucleotide changes in cloned genes by any one of several procedures, including site-specific and random mutagenesis. Also called *in vitro* mutagenesis.

Disarmed vector A vector in which some characteristic (e.g. conjugation) has been disabled.

DMD See *Duchenne muscular dystrophy*

DNA chip A DNA microarray used in the analysis of gene structure and expression. Consists of oligonucleotide sequences immobilised on a 'chip' array.

DNA codon A set of three contiguous deoxyribonucleotide pairs of the coding region of a structural gene where the bases of one strand are transcribed into a codon.

DNA construct A cloning vector with a DNA insert.

DNA delivery system Generic term for any procedure that facilitates the uptake of DNA by a recipient cell.

DNA Deoxyribonucleic acid; the genetic material of living things.

DNA fingerprint a set of DNA fragments that are characteristic for a particular source of DNA such as an insert of a clone. In some cases, restriction endonuclease DNA fragments are visualized by hybridization after gel electrophoresis. In other instances, the polymerase chain reaction (PCR) is used to generate a distinctive pattern of DNA bands that are evident after gel electrophoresis.

DNA fingerprinting A comparative diagnostic technique that characterizes the DNA of an organism or a sample.

DNA foot printing Method of identifying regions of DNA to which regulatory proteins will bind

DNA hybridization The pairing of two DNA molecules, often from different sources, by hydrogen bonding between complementary nucleotides. This technique is frequently used to detect the presence of a specific nucleotide sequence in a DNA sample.

DNA ligase Enzyme used for joining DNA molecules by the formation of a phosphodiester bond between a 5' phosphate and a 3' OH group.

DNA microarray See *DNA chip*

DNA polymerase an enzyme that links an incoming deoxyribonucleotide, which is determined by complementarity to a deoxyribonucleotide in a template DNA strand, with a phosphodiester bond to the 3'hydroxyl group of the last incorporated nucleotide of the growing strand during replication.

DNA polymerase An enzyme that synthesises a copy of a DNA template

DNA probe A segment of DNA that is labeled (tagged) so that, after DNA hybridization reaction, any base pairing between the probe and a complementary base sequence in a DNA sample can be detected.

DNA profiling Term used to describe the various methods for analysing DNA to establish identity of an individual.

DNA transformation See Transfection, Transformation

DNA typing See Genotyping

DNase I An enzyme that degrades DNA. It is used to remove DNA from RNA preparations and from cell free extracts. Also called deoxyribonuclease I.

Domain A segment of a protein that has a discrete function or conformation. At the protein level, a domain can be as small as a few amino acid residues and as large as half of the entire protein.

Dominant (1) When a heterozygous and homozygous genotype determines the same phenotype. (2) An allele that produces the same phenotype whether the genotype is heterozygous or homozygous. Also called dominance.

Dominant gene One of a pair of alleles that is sufficient to produce a phenotype in a heterozygote.

Dot-blot Technique in which small spots, or 'dots', of nucleic acid are immobilized on a nitrocellulose or nylon membrane for hybridization.

Downstream (1) In molecular biology, the stretch of nucleotides of DNA that lie in the 3' direction of the site of initiation of transcription, which is designated as +1. Downstream nucleotides are marked with plus signs, e.g., +2, +10. Also refers to the 3'side of a particular gene or sequence of nucleotides. (2) In chemical engineering, those phases of a manufacturing process that follow the biotransformation stage. Refer to recovery and purification of the product of a fermentation process. Also called downstream processing.

Downstream processing Refers to the procedures used to purify products (usually proteins) after they have been expressed in bacterial, fungal or mammalian cells.

Drosophila melanogaster Fruit fly used as a model organism in genetic, developmental and molecular studies.

Duchenne muscular dystrophy X-linked (q.v.) muscle-wasting disease caused by defects in the gene for the protein dystrophin (q.v.).

Duplex DNA Double stranded DNA.

Dystrophin Large protein linking the cytoskeleton to the muscle cell membrane, defects in which cause muscular dystrophy.

Electrophoresis A technique that separates molecules (often DNA, RNA, or protein) on the basis of relative migration in a strong electric field.

Electroporation Electrical treatment of cells that induce transient pores, through which DNA is taken into the cell.

Electrotransfer Transfer of a macromolecule by an electric field from a gel to a membrane.

ELISA See Enzyme-linked immunosorbent assay.

Embryo splitting Technique used to clone organisms by separating cells in the early embryo, which then go on to direct development and produce identical copies of the organism.

End labelling Adding a radioactive molecule onto the end(s) of a polynucleotide.

Endonuclease An enzyme that cuts within a nucleic acid molecule, as opposed to an exonuclease (q.v.), which digests DNA from one or both ends.

Endotoxin A component of the cell wall of gram-negative bacteria that elicits, in humans, an inflammatory response and fever.

Enhancer A DNA sequence that increases the transcription of a eukaryotic gene when they are both on the same DNA molecule. Also called enhancer element, enhancer sequence.

Enzyme A protein that catalyses a specific reaction

Enzyme replacement therapy Therapeutic procedure in which a defective enzyme function is restored by replacing the enzyme itself. Cf. *gene therapy*.

Enzyme-linked immunosorbent assay A technique for detecting specific molecules in a mixed sample. An antibody (primary) is bound to the target molecule; another antibody (secondary), which binds to the primary antibody, is added later. The secondary antibody has attached to it an enzyme that can convert a colorless substrate into a colored product. If the target molecule is not present in the sample, washing steps will remove both antibodies and no colored product will be produced. Also called ELISA.

Epigenesis Theory of development that regards the process as an interactive series of steps, in which the various signals and control events interact to regulate development.

Epitope A specific chemical domain on an antigen that is recognized by an antibody. Each epitope on a molecule such as a protein elicits the synthesis of a different antibody. Also called an antigenic determinant.

Epitope tag An affinity tag that is recognized by an antibody.

Escherichia coli The most commonly used bacterium in molecular biology.

Ethidium bromide A molecule that binds to DNA and fluoresces when viewed under ultraviolet light. Used as a stain for DNA.

Eukaryotes Organisms, including animals, plants, fungi, and some algae, that have (1) Chromosomes enclosed within a membrane bounded nucleus and (2) functional organelles, such as mitochondria and chloroplasts, in the cytoplasm of their cells.

Exogenous DNA DNA that has been derived from a source organism and has been cloned into a vector and introduced into a host cell. Also referred to as foreign or heterologous DNA.

Exon A segment of a gene that is transcribed as part of the primary transcript and is retained, after processing, with other exons to form a functional mRNA molecule.

Exon trapping A protocol for the isolation of exons from genomic clones. Genomic DNA (1 to 6 kb) is cloned into an intron that is flanked by two exons (artificial gene) in a vector. When an exon is present in the cloned DNA, reverse transcriptase polymerase chain reaction (RT-PCR) shows that the mRNA from the artificial gene is longer than expected. The cloned DNA can be amplified, subcloned, and characterized. Also called exon entrapment, exon amplification.

Exonuclease An enzyme that digests a nucleic acid molecule from one or both ends.

Expressed sequence tag A sequence tagged site derived from a cDNA clone. Also called EST, expressed sequence tagged site, eSTS.

Expression library A population of different DNA molecules cloned into one kind of expression vector.

Expression Transcription and translation of a gene.

Expressivity The degree to which a particular genotype generates its effect in the phenotype. Cf. *Penetrance*.

Extension A single-stranded DNA region consisting of one or more nucleotides at the end of a strand of duplex DNA. Also called protruding end, stick end, overhang, cohesive end.

Extrachromosomal DNA A replicatable DNA element that is not a part of a chromosome.

Ex-vivo Outside the body. Usually used to describe gene therapy procedure in which the manipulations are performed outside the body, and the altered cells returned after processing. Cf. *in vivo, in vitro*

False negative A negative assay result that should have been positive.

False positive A positive assay result that should have been negative.

Fermentation (1) In chemical engineering, the growth of cells or microorganisms in specialized vessels (fermenters, bioreactors). (2) In biochemistry, the breakdown of carbon compounds by cells or organisms to ATP without using molecular oxygen.

Fermenter See Bioreactor.

Flavr Savr (sic) Transgenic tomato in which polygalacturonase (q.v.) synthesis is restricted using antisense technology. Despite the novel science, the Flavr Savr was not a commercial success.

Foldback DNA Class of DNA which has palindromic or inverted repeat regions that re-anneal rapidly when duplex DNA is denatured.

Foreign DNA A DNA molecule that is incorporated into either a cloning vector or a chromosomal site.

Four cutter A type II restriction endonuclease that binds and cleaves DNA at sites that contain four nucleotide pairs.

Frameshift mutation In chromosomal DNA, an insertion or deletion of base pairs that changes the reading frame of a gene.

Functional gene cloning A strategy for isolating a gene that depends on information about its gene product.

Fusion protein A hybrid recombinant protein that contains vector-encoded amino acid residues at the N terminus

Fusion protein The product of two or more coding sequences from different genes have been cloned together and that, after translation, form a single polypeptide sequence. Also called hybride protein, chimeric protein.

Gamete Refers to the haploid male (sperm) and female (egg) cells that fuse to produce the diploid zygote (q.v.) during sexual reproduction

Gel electrophoresis Technique for separating nucleic acid molecules on the basis of their movement through a gel matrix under the influence of an electric field. See agarose and polyacrylamide

Gel retardation Method of determining protein-binding sites on DNA fragments on the basis of their reduced mobility, relative to unbound DNA, in gel electrophoresis experiments.

Gene A segment of nucleic acid that encodes a functional protein or RNA. The unit of inheritance.

Gene bank A population of organisms, each of which carries a DNA molecule that was inserted into a cloning vector. Ideally, all of the cloned DNA molecules represent the entire genome of another organism. Also called gene library, clone bank, library. This term is sometimes also used to denote all of the vector molecules, each carrying a piece of the chromosomal DNA of an organism, prior to the insertion of these molecules into a population of host cells.

Gene cloning Insertion of a gene into a DNA vector (often a plasmid) to form a new DNA molecule that can be perpetuated in a host cell. Also called recombinant DNA technology, genetic engineering, gene splicing, gene transplantation, molecular cloning, cloning.

Gene map The linear array of genes on a chromosome.

Gene protecting technology Range of techniques used to ensure that particular commercially derived recombinant constructs cannot be used without some sort of control or process, usually supplied by the company marketing the recombinant. Also known as genetic use restriction technology and genetic trait control technology.

Gene The unit inheritance, located on a chromosome. In molecular terms, usually taken to mean a region of DNA that encodes one function. Broadly, therefore, one gene encodes one protein.

Gene therapy The use of cloned genes in the treatment of genetically derived malfunctions. May be delivered *in vivo* or *ex vivo*. May be offered as gene addition or gene replacement versions.

Genetic code The triplet codons that determine the types of amino acid that are inserted into a polypeptide during translation. There are 61 codons for 20 amino acids (plus three stop codons), and the code is therefore referred to as degenerate.

Genetic complementation When two genomes or DNA molecules that are in the same cell produce a function that neither genome or DNA molecule can supply on its own. Also called complementation.

Genetic fingerprinting A method which uses radioactive probes to identify bands derived from hypervariable regions of DNA (q.v.). The band pattern is unique for an individual, and can be used to establish identity or family relationships.

Genetic immunization Delivery of a cloned gene that encodes an antigen to a host organism. After the cloned gene is expressed, it elicits an antibody response that protects the organism from infection by virus, bacterium, or other disease-causing organism.

Genetic map The linear array of genes on a chromosome based on recombination frequencies. Also called linkage map.

Genetic mapping Determining the linear order of marker sites along a chromosome. Also called mapping.

Genetic marker A phenotypic characteristic that can be ascribed to a particular gene.

Genetic polymorphism When two or more alleles of a locus in a population of individuals occur at a frequency of 1% or greater. Often, in the appropriate context, it is called polymorphism.

Genetic trait control technology Version of gene protection technology (q.v.), sometimes called 'traitor technology'.

Genetic use restriction technology (GURT) See *gene protection technology*

Genetically modified organism (GMO) An organism in which a genetic change has been engineered. Usually used to describe transgenic plants and animals

Genome (1) The entire complement of genetic material of an organism, virus, or organelle. (2) The haploid set of chromosomes (DNA) of a eukaryotic organism.

Genomic library A collection of clones which together represent the entire genome of an organism.

Genomics The study and development of genetic and physical maps, large-scale DNA sequencing, gene discovery, and computer-based systems for managing and analyzing genomic data.

Genotype (1) The genetic constitution of an organism. (2) The alleles at a genetic locus.

Genotyping The determination of the alleles of a chromosome of an individual. Also called DNA typing, haplotyping.

Germ line Gamete producing (reproductive) cells that give rise to eggs and sperm

GMO See *genetically modified organism*

Gram-negative organism Any prokaryotic organism that does not retain the first stain (crystal violet) used in the Gram technique. It does retain the second stain (safranin O) and therefore has a pink color when viewed under a light microscope. Retention of the stain is due to the structure of the cell wall.

Gram-positive organism Any prokaryotic organism that retains the first stain (crystal violet) used in the Gram technique, which gives a purple-black color when viewed under a light microscope. Retention of the stain is due to the structure of the cell wall.

Grandfather method Knowledge of the X-linked alleles of a father is used to determine the genetic phase of a daughter whose sons can be examined directly as recombinant or nonrecombinant types for the genetic mapping of the X-linked gene loci.

Guanine (G) Nitrogenous base found in DNA and RNA

GURT See *genetic use restriction technology, gene protection technology*.

Haploid Having one set of chromosomes. C.f. *diploid*

Haplotype The alleles of the loci of a chromosome. Derived by combining *haplo* from haploid and *type* from genotype.

Haplotyping See Genotyping.

Helper plasmid A plasmid that provides a function(s) to another plasmid in the same cell. Some helper plasmids are used to mobilize nonconjugative plasmids from a donor cell into a recipient cell.

Helper virus A virus that provides a function(s) to another virus or viral genome in the same cell.

Heterologous Refers to gene sequences that are not identical, but show variable degrees of similarity

Heteropolymer A polymer composed of different types of monomer. Most protein and nucleic acid molecules are heteropolymers

Heterozygote An individual that has different alleles at the same locus in its two homologous chromosomes.

High-resolution map Closely spaced sites throughout a genetic or physical map.

Hogness box See *TATA box*.

Homodimer A protein with two identical polypeptide chains.

Homologous chromosomes Chromosomes that normally synapse during zygotene.

Homologus (1) Refers to paired chromosomes in diploid organisms (2) Used to strictly describe DNA sequences that are identical; however, the percentage homology between related sequences sometimes quoted.

Homomer A protein with two or more identical protein chains. Also called homomeric polypeptide, homomeric protein.

Homopolymer A polymer composed of only one type of monomer, such as polyphenylalanine (protein) or polyadenine (nucleic acid)

Homopolymeric tailing See Tailing.

Homozygote An individual that has identical alleles at the same locus in its two homologous chromosomes.

Homozygous dominant A genotype with two dominant alleles at a gene locus.

Homozygous recessive A genotype with two recessive alleles at a gene locus.

Homozygous Refers to a diploid organism (cell or nucleus) which has identical alleles at a particular locus

Host A cell used to propagate recombinant DNA molecules.

Human artificial chromosome a chromosome that is assembled from telomere, centromere, and human genomic DNA sequences. Also called HAC, microchromosome.

Human genome project An international research effort dedicated to developing high-resolution human genetic and physical maps and, eventually, the complete genomic DNA sequences of humans and other organisms. Also called HGP.

Human minisatellite DNA Human DNA that is noncoding and generally G+C rich and contains tandem repeats of short (9- to 40- base-pair) stretches of DNA.

Hybrid – arrest translation Techniques used to identify the protein product of a cloned gene, which translation of its mRNA is prevented by the formation of a DNA mRNA hybrid

Hybrid gene The combination of two genes or the parts of two genes in the correct reading frame that encodes a single protein that has amino acid sequences from both genes.

Hybrid selection A protocol for determining which genomic clones hybridize to a cDNA or mRNA molecule.

Hybridization The joining together of artificially separated nucleic acid molecules *via* hydrogen bonding between complementary bases.

Hybridoma The product of the fusion of a myeloma cell with an antibody-producing lymphocyte. This cell combination (hybridoma) can continue to divide in cell culture and secrete a single type of antibody.

Hybrid-release translation Technique in which a particular mRNA is selected by hybridization with its homologous clones DNA sequence, and is then translated to give a protein product that can be identified.

Hyperchromic effect Change in absorbance of nucleic acids, depending on the relative amounts of single-stranded and double-stranded forms. Used as a measurement in denaturation / renaturation studies.

Hypervariable region (HVR) A region in a genome that is composed of a variable number of repeated sequences and its diagnostic for the individual. See *genetic fingerprinting*.

Hypervariable region the parts of both the heavy and light chains of an antibody molecule that enable it to bind to a specific site on an antigen.

Hypervariable segment A region of a protein that varies considerably between strains or individuals. Ice-minus bacteria Bacteria engineered to disrupt the normal ice-forming process, used to protect plants from frost damage

IGF-1 See *insulin-like growth factor*

Immunoassay A protocol that uses antibody specificity to detect the presence of a particular compound in a biological sample.

Immunogen A substance that induces an antibody response. Also called an antigen.

Immunogloblin See Antibody.

Immunosuppression A substance, an agent, or a condition that prevents or diminishes the immune response.

Immunotherapeutic procedure The use of an antibody or a fusion protein containing the antigen binding site of an antibody to treat a disease and enhance the well-being of a patient.

Immunotoxin A fusion protein that has separate domains with antibody and toxin activity. The antibody portion of the molecule facilitates binding to a target molecule or cell, and the toxin inactivates the target molecule or kills the cell.

In vitro Literally 'in glass', meaning in the test-tube, rather than in the cell or organism.

In vitro translation Protein synthesis that is directed by either purified DNA with bacterial extracts or mRNA with wheat germ or rabbit reticulocyte extracts that provide ribosomes, tRNAs, and protein synthesis factors. The reaction mixture is often supplemented with ATP, GTP, and amino acids.

In vivo Literally 'in life', meaning the natural situation, within a cell or organism

Initiation The start of the biosynthesis of a polymeric macromolecule.

Initiator element See Response element

Initiator tRNA The fMet-tRNA_{F^{Met}} in prokaryotes and Met-tRNA_{M^{et}} in eukaryotes that starts translation.

Initiaiton codon The codon AUG, which specifies the first amino acid (methionine[N-formylmethionine in prokaryotes]) of a protein. Also called initiator, translational start codon, translational initiation signal.

Initiation complex The fMet-tRNA_f^{Met}-mRNA-small ribosomal subunit – large ribosomal subunit combination in prokaryotes or the Met-tRNA^{Met}-mRNA-small ribosomal subunit-large subunit combination in eukaryotes that is ready for the elongation phase of the translation.

Insertion vector A bacteriophage vector that has a single cloning site into which DNA is inserted.

Insulin-like growth factor (IGF-1) Polypeptide hormone, synthesis of which is stimulated by growth hormone. Implicated in some concerns about the safety of using recombinant bovine growth hormone in cattle to increase milk yields.

Integrating vector A vector that is designed to integrate cloned DNA into the host cell chromosomal site.

Integration Insertion of a DNA molecule (usually by homologous recombination) into a chromosomal site.

Integration-excision (I/E) region The portion of bacteriophage λ DNA that enables bacteriophage λ DNA to be inserted into a specific site in the *Escherichia coli* chromosome and excised from this site.

Internal ribosomal entry site A nontranslated sequence following a coding region of a polycistronic RNA that binds to a small ribosomal subunit and forms an intiation-of-translation complex.

Intervening sequence Region in a eukaryotic gene that is not expressed via the processed mRNA

Intron A segment of a gene that is transcribed but is then excised from the primary transcript during processing to a functional RNA molecule. Also called intervening sequence.

Inverted repeat A short sequence of DNA that is repeated, usually at the ends of a longer sequence, in a reverse orientation.

IPTG iso-Propyl-thiogalactoside, a gratuitous inducer which de-represses transcription of the *lac* operon.

Jumping library See Chromosome jumping

Kilobase (kb) 1000 base pairs or base-pairs, used as a unit for measuring or specifying the length of DNA or RNA molecules

Klenow fragment a product of proteolytic digestion of the DNA polymerase I from *E. coli* that retains both polymerase and 3'-exonuclease activities but not 5'-exonuclease activities.

Knockin mouse A transgenic mouse in which a gene function has been added or 'knocked in'. Used primarily to generate animal models for the study of human disease. Cf. *Knockout mouse*.

Knockout mouse A transgenic mouse in which a gene function has been disrupted or 'knocked out'. Used primarily to generate animal models for the study of human disease, e.g. cystic fibrosis. Cf. *Knockin mouse*

Ligase chain reaction A technique for determining the presence or absence of a specific nucleotide pair within a target gene. Also called LCR.

Ligation Joining of two DNA molecules by the formation of phosphodiester bonds. In vitro, this reaction is usually catalyzed by the enzyme T4 DNA ligase.

Linkage map See Genetic map.

Linkage mapping Genetic mapping (q.v.) technique used to establish the degree of linkage between genes. See also *recombination frequency mapping*.

Linkage The occurrence of two or more genes on the same chromosome.

Linker A synthetic double-stranded oligonucleotide that carries the sequence for one or more restriction endonuclease sites.

Lipase Enzyme that hydrolyses fats (lipids).

Liposome A circular collection of lipid molecules in which the hydrophobic portions of the molecule are facing inward; a lipid vesicle with an aqueous interior that can carry nucleic acids, drugs, or other therapeutic agents.

Locus The site at which a gene is located on a chromosome

Long terminal repeats Similar blocks of genetic information that are found at the ends of the genomes of retroviruses. Also called LTRs.

Lysogeny A condition in which a bacteriophage genome (prophage) survives within a host bacterium either as part of the host chromosome or as part of an extrachromosomal element and does not initiate lytic functions.

Lytic Refers to bacteriophage infection that causes lysis of the host cell.

M13 + strand The single-stranded DNA molecule that is present in infective bacteriophage M13.

Map distance The number of map units or centimorgans between two gene loci.

Map unit A measure of genetic distance between two linked genetic loci; usually 1% recombination corresponds to one map unit.

Maternal inheritance Pattern of inheritance from female cytoplasm. Mitochondrial genes are inherited in this way, as the mitochondria are inherited with the ovum.

Mega (M) SI prefix, 10^6

Messenger RNA (mRNA) the ribonucleic acid molecule transcribed from DNA that carries the codons specifying the sequence of amino acids in a protein

Messenger RNA An RNA molecule carrying the information that, during translation, specifies the amino acid sequence of a protein molecule. Also called mRNA.

Methylation The addition of a methyl group(s) to a macromolecule. For example, the addition of a methyl group to specific cytosine and, occasionally, adenine residues in DNA.

Micro (μ) SI prefix 10^{-6}

Microinjection Introduction of DNA into the nucleus or cytoplasm of a cell by insertion of a microcapillary and direct injection

Microsatellite DNA Types of sequence repeated many times in the genome. Based on dinucleotide repeats, microsatellites are highly variable and can be used in mapping and profiling studies.

Milli (m) SI prefix 10^{-3}

Missense mutation A genetic mutation that changes a codon for one amino acid into a codon specifying another amino acid.

Molecular cloning Alternative term for gene cloning

Molecular ecology Use of molecular biology and recombinant DNA techniques in studying ecological topics

Molecular paleontology Use of molecular techniques to investigate the past, as in DNA profiling from mummified or fossilised samples

Monocistronic Refers to an RNA molecule encoding one function

Monoclonal antibody A single type of antibody that is directed against a specific epitope (antigenic determinant) and is produced by a hybridoma cell line, which is formed by the fusion of a lymphocyte with a myeloma cell. Some myeloma cells synthesize single antibodies naturally. Also called MAb.

Monogenic Trait caused by a singel gene. C.f. *polygenic*

Monomer The unit that makes up a polymer. Nucleotides and amino acids are the monomers for nucleic acids and proteins, respectively.

Monosomic Diploid cells in which one of a homologous pair of chromosomes has been lost. C.f. *trisomy*.

Monozygotic Refers to identical twins, generated from the splitting of a single embryo at an early stage

Mosaic An embryo or organism in which not all the cells carry identical genomes.

Multifactorial Caused by many factors, e.g. genetic trait in which many genes and environmental influences may be involved.

Multigene RNA An RNA transcript of an operon.

Multi-locus probe DNA probe used to identify several bands in a DNA fingerprint or profile. Generates the 'bar code' pattern in a genetic fingerprint

Multiple cloning site (MCS) A short region of DNA in a vector that has recognition sites for several restriction enzymes

Multipotent Cell which can give rise to a range of differentiated cells. C.f. *totipotent, pluripotent*

Mutagenesis Chemical or physical treatment that changes the nucleotides of the DNA of an organism.

Mutagenesis The process of inducing mutations in DNA

Mutant An organism that differs from the wild type because it carries one or more genetic changes in its DNA. Also called a variant.

Mutant complementation See Genetic complementation

Mutation An alteration to the sequence of bases in DNA. May be caused by insertion, deletion or modification of bases.

Mutation detection assay a protocol that identifies the difference of one or a few nucleotide pairs between the same DNA molecules from different sources.

N terminus Amino terminus, defined by the $-NH_2$ group of an amino acid or protein

Nano (n) SI prefix, 10^{-9}

Narrow-host-range plasmid A plasmid that can replicate in one, or at most a few, different bacterial species.

Native protein A recombinant protein that is synthesised from its own N terminus, rather than from an N terminus supplied by the cloning vector.

Negative control A system of regulation of transcription that requires the removal of a repressor from an operator.

Nested fragments A series of nucleic acid fragments that differ from each other (in terms of length) by one or only a few nucleotides

Nick translation Method for labelling DNA with radioactive dNTPs

Nod box A DNA sequence that controls the transcriptional regulation of *Rhizobium* nodulation genes.

Northern blotting Similar to Southern blotting, except that RNA that has been separated by gel electrophoresis is transferred from a gel onto a matrix such as a nitrocellulose or nylon membrane, and the presence of a specific RNA molecule is detected by DNA-RNA hybridization.

Nuclear transfer Method for cloning organisms in which a donor nucleus is taken from a somatic cell and transferred to the recipient ovum.

Nuclease An enzyme that hydrolyses phosphodiester bonds.

Nucleoid Region of a bacterial cell in which the genetic material is located.

Nucleoside A base (purine or pyrimidine) that is covalently linked to a five-carbon (pentose) sugar. When the sugar is ribose, the nucleoside is a ribonucleoside; when the sugar is deoxyribose, the nucleoside is a deoxyribonucleoside.

Nucleotide A nucleoside with one or more phosphate groups linked to the 5'carbon of the pentose sugar. Ribose-containing nucleosides are often called ribonucleoside monophosphate (NMP), ribonucleoside diphosphate (NDP), or ribonucleoside triphosphate (NTP). When the nucleoside contains the sugar deoxyribose, the nucleotides are called deoxyribonucleoside mono-, di-, or triphosphates (dNMP, dNDP, or dNTP).

Nucleus Membrane-bound region in a eukaryotic cell that contains the genetic material.

Oligo(dT)-cellulose Short sequence of deoxythymidine residues linked to a cellulose matrix, used in the purification of eukaryotic mRNA.

Oligolabelling See *primer extension*

Oligomer General term for a short sequence of monomers

Oligonucleotide A short molecule (usually 6 to 100 nucleotides) of single-stranded DNA. Oligonucleotides are sometimes called oligodeoxyribonucleotides or oligomers and are usually synthesized chemically.

Oligonucleotide ligation assay A diagnostic technique for determining the presence or absence of specific nucleotide pair within a target gene which indicates whether a gene is wild type (normal) or mutant (defective). Also called OLA.

Oligonucleotide-directed mutagenesis Process by which a defined alteration is made to DNA using a synthetic oligonucleotide

Oncogene A gene that plays a role in the cell division cycle. Often mutated forms of oncogenes cause a cell to grow in an uncontrolled manner.

Oncomouse Transgenic mouse engineered to be susceptible to cancer

Oocyte Stage in development of the female gamete or ovum (egg). Often the terms oocyte and ovum are used interchangeably.

Open reading frame A sequence of nucleotides in a DNA molecule that encodes a peptide or protein. This term is often used when, after the sequence of a DNA fragment has been determined, the function of the encoded protein is not known. Also called ORF.

Operator The region of DNA that is upstream from a prokaryotic gene(s) and to which a repressor or activator binds.

Operon A cluster of bacterial genes under the control of a single regulatory region.

Organismal cloning The production of an identical copy of an individual organism by techniques such as embryo splitting or nuclear transfer. Used to distinguish the process from molecular cloning (q.v.).

P1 artificial chromosome A plasmid vector system based on bacteriophage P1 that uses electroporation for introducing a vector with a large DNA insert (100 to 300 kb) into *Escherichia coli*. Also called PAC.

PAC See P1 artificial chromosome.

Palindrome A DNA sequence that reads the same on both strands when read in the same (e.g. 5' 3') direction. Examples include many restriction enzyme recognition sites.

PCR See *polymerase chain reaction*

Pedigree analysis Determination of the transmission characteristics of a particular gene by examination of family histories

Penetrance A certain fraction, usually expressed as a percentage, of individuals with a mutant gene(s) at a locus that show the expected abnormal (mutant) phenotype. Also called incomplete penetrance, nonpenetrance, reduced penetrance.

Peptide bond The covalent bond between the free carboxyl group of the α carbon of one amino acid and the free amino acid group of the α carbon of an adjacent amino acid in a peptide or protein.

Peptidyl site The portion of a ribosome where the tRNA with the peptide chain participates in peptide bond formation with the aminoacyl-tRNA during translation. Also called P site.

Peptidyl-tRNA The tRNA that has a growing peptide chain attached to it during translation.

Phage See *bacteriophage*

Phagemid A vector containing plasmid and phage sequences

Pharm animal Transgenic animal used for the production of pharmaceuticals.

Phenocopy A nongenetic condition that closely resembles a genetic disorder.

Phenotype The observable characteristics of an organism, determined both by its genotype (q.v.) and its environment.

Phosphodiester bond The linkage of a phosphate group to the 3' carbon of one nucleotide and the 5' carbon of another nucleotide; the linkage between nucleotides of the same nucleic acid strand.

Physical map A map of the positions of chromosome sites such as restriction endonuclease recognition or sequence tagged sites on a chromosome. The distance between sites is measured in base pairs.

Physical mapping Mapping genes with reference to their physical location of the chromosome. Generates the next level of detail compared to genetic mapping (q.v.).

Physical marker A sequence-based tag that labels a region of the genome. There are several such tags that can be used in mapping studies. Cf. *RFLP*, *STS*

Pico (p) SI prefix, 10^{-12}

Plaque A Cleared area on a bacterial lawn caused by infection by a lytic bacteriophage.

Plasmid A circular extrachromosomal element found naturally in bacteria and some other organisms. Engineered plasmids are used extensively as vectors for cloning.

Ploidy number Refers to the number of sets of chromosomes, e.g. haploid, diploid, triploid, etc.

Pluripotent Cell which can give rise to a range of differentiated cells. Cf. *multipotent*, *totipotent*.

Plystuffer An expendable stuffer fragment in a vector that is composed of many repeated sequences.

Poly (A) tail See Polyadenylation

Polyacrylamide A cross-linked matrix for gel electrophoresis (q.v.) of small fragments of nucleic acids, primarily used for electrophoresis of DNA. Also used for electrophoresis of proteins.

Polyadenylic acid A string of adenine residues. Poly (A) tails are found at the 3' ends of most eukaryotic mRNA molecules.

Polycistronic Refers to an RNA molecule encoding more than one function. Many bacterial operons are expressed via polycistronic mRNAs.

Polycistronic RNA An mRNA that encodes two or more proteins.

Polyclonal antibody A serum sample that contains antibodies that bind to different antigenic determinants of one antigen.

Polygalacturonase Enzyme involved in pectin degradation. Target for antisense control in the Flavr Savr tomato (q.v.).

Polygenic trait A trait determined by the interaction of more than one gene, e.g. eye colour in humans.

Polyhedra Capsid structures in baculoviruses, composed of the protein polyhedrin.

Polylinker See *multiple cloning site*

Polymer A long sequence of monomers

Polymerase An enzyme that synthesises a copy of a nucleic acid

Polymerase chain reaction a technique for amplifying a specific segment of DNA by using a thermostable DNA polymerase, deoxyribonucleotides, and primer sequences in multiple cycles of denaturation, renaturation, and DNA synthesis. Also called PCR.

Polymorphism Refers to the occurrence of many allelic variants of a particular gene or DNA sequence motif. Can be used to identify individuals by genetic mapping and DNA profiling techniques

Polynucleotide A polymer made up of nucleotide monomers

Polynucleotide kinase (PNK) An enzyme that catalyses the transfer of a phosphate group onto a 5' hydroxyl group

Polypeptide A chain of aminoacid residues. Cf. *protein*

Positional cloning Cloning genes for which little information is available apart from their location on the chromosome

Positive control A system of regulation of transcription that requires the addition of a protein activator to an activator site on the DNA.

Post-translational modification The specific addition of phosphate groups, sugars (glycosylation), or other molecules to a protein after it has been synthesized.

Preformationism Refers to the idea that all development is pre-coded in the zygote, and the development is simply the unfolding of this information. Now considered too simplistic. Cf. *epigenesis*.

Pribnow box Sequence found in prokaryotic promoters that is required for transcription initiation. The consensus sequence (q.v.) is TATAAT.

Primary transcript The initial, and often very large, product of transcription of a eukaryotic gene. Subjected to processing to produce the mature mRNA molecule.

Primer extension Synthesis of a copy of a nucleic acid from a primer. Used in labelling DNA and in determining the start site of transcription.

Probe (1) For diagnostic tests, the agent that is used to detect the presence of a molecule in a sample. (2) A DNA sequence that is used to detect the presence of a complementary sequence by hybridization with a nucleic acid sample.

Proinsulin Precursor of insulin that includes an extra polypeptide sequence that is cleaved to generate the active insulin molecule.

Prokaryotes Organisms, usually bacteria, that have neither a membrane-bound nucleus enclosing their chromosomes nor functional organelles such as mitochondria and chloroplasts.

Promoter A segment of DNA to which RNA polymerase attaches. It usually lies upstream of (5' to) a gene. A promoter sequence aligns the RNA polymerase so that the transcription will initiate at a specific nucleotide.

Promoter DNA sequence(s) lying upstream from a gene, to which RNA polymerase binds

Pronucleus One of the nuclei in a fertilised egg prior to fusion of the gametes

Prophage A repressed or inactive state of a bacteriophage genome that is maintained in a bacterial host cell as part of the chromosomal DNA.

Protease An enzyme that hydrolyzes peptide bond linkages and cleaves proteins into smaller peptides. Also called proteinase or proteolytic enzyme.

Protein A condensation (dehydration) heteropolymer composed of amino acid residues linked together by peptide bonds to give a polypeptide

Proteome Refers to the population of proteins produced by a cell. Cf. *genome, transcriptome*.

Protoplast A cell from which the cell wall has been removed

Prototroph A cell that can grow in an unsupplemented growth medium.

Purine Fusion of a pyrimidine and imidazole ring, e.g., adenine and guanine.

Pyrimidine A heterocyclic ring, e.g., thymine, cytosine and uracil.

Reading frame The pattern of triplet codon sequences in a gene. There are three reading frames, depending on which nucleotide is the start point. Insertion and deletion mutations can disrupt the reading frame and have serious consequences, as often the entire coding sequence becomes nonsense after the point of mutation.

Recessive An allele where the expression is masked in the phenotype in heterozygous individuals. Cf. Dominant

Recognition site See Restriction site

Recombinant DNA A DNA molecule made up of sequences that are not normally joined together

Recombination frequency mapping Method of genetic mapping that uses the number of crossover events that occur during meiosis to estimate the distance between genes. Cf. *physical mapping*

Regulatory gene A gene that exerts its effect by controlling the expression of another gene

Renaturation kinetics Method of analysing the complexity of genomes by studying the patterns obtained when DNA is denatured and allowed to renature

Renaturation The reassociation of two nucleic acid strands after denaturation.

Repetitive sequence A sequence that is repeated a number of times in the genome

Replacement vector A bacteriophage vector in which the cloning sites are arranged in pairs, so that the section of the genome between these sites can be replaced with insert DNA.

Replication Copying the genetic material during the cell cycle. Also refers to the synthesis of new phage DNA during phage multiplication

Replicon A piece of DNA carrying an origin of replication

Reporter gene A gene that encodes a product that can readily be assayed. For example, reporter genes are used to determine whether a particular DNA construct has been successfully introduced into a cell, organ, or tissue.

Repression Inhibition of transcription by preventing RNA polymerase from binding to the transcription initiation site; a repressed gene is “turned off”.

Repulsion The phase state in which a dominant version and a recessive version of two different genes occur on the same chromosome. Also called *transconfiguration*. See also Coupling

Response element A sequence of deoxyribonucleotides of a gene that acts as a binding site for a protein (transcription factor) that regulates transcription. Also called initiator element, signal region.

Restriction endonuclease (type II) An enzyme that recognizes a specific duplex DNA sequence and cleaves phosphodiester bonds on both strands between definite nucleotides.

Restriction enzyme An endonuclease that cuts DNA at sites defined by its recognition sequence

Restriction fragment A piece of DNA produced by digestion with a restriction enzyme

Restriction fragment length polymorphism The occurrence of variations in the lengths of certain DNA fragments that are produced after cleavage with a type II restriction endonuclease. The differences in DNA lengths are due to the presence or absence of a specific restriction

endonuclease recognition site(s) and are detected by DNA hybridization with DNA probes after separation by gel electrophoresis. Also called RFLP.

Restriction map The linear array of restriction endonuclease sites on a DNA molecule.

Restriction mapping Technique used to determine the location of restriction sites in a DNA molecule

Restriction site The sequence of nucleotide pairs in duplex DNA that is recognized by a type II restriction endonuclease. Sometimes called restriction enzyme site, restriction endonuclease site or recognition site.

Retrovirus A virus that has an RNA genome that is copied into DNA during the infection

Reverse transcriptase An RNA dependent DNA polymerase that uses an RNA molecule as a template for the synthesis of a complementary DNA strand.

Reverse transcriptase An RNA-dependent DNA polymerase found in retroviruses, used *in vitro* for the synthesis of cDNA.

Reverse transcriptase-polymerase chain reaction A two-step protocol for synthesizing cDNA molecules. First cDNA strands are synthesized *in vitro* by reverse transcriptase with oligo (dT) as a primer and mRNA as the template. Second, a specific cDNA strand is amplified by the polymerase chain reaction (PCR), with one primer directed to a sequence of the first cDNA and the other to a sequence of the complementary cDNA strand (second strand) that is synthesized during the first PCR cycle. Also called RT-PCR.

Ribonuclease (Rnase) An enzyme that hydrolyses RNA.

Ribonucleic acid (RNA) A condensation heteropolymer composed of ribonucleotides.

Ribose The 5-carbon sugar component of RNA.

Ribosomal RNA The RNA molecules that form part of the large and small ribosomal subunits. Also called rRNA.

Ribosome The subcellular structure that contains both RNA and protein molecules and mediates the translation of mRNA into protein. Ribosomes contain both large and small subunits.

Ribosome-binding site A region on an mRNA molecule that is involved in the binding of ribosomes during translation.

Ribozyme An RNA molecule that has catalytic activity.

RNA polymerase An enzyme that links an incoming ribonucleotide, which is determined by complementarity to a base in a template DNA strand, with a phosphodiester bond to the 3'hydroxyl group of the last incorporated ribonucleotide of the growing RNA strand during transcription.

RNA processing The formation of functional RNA from a primary transcript (q.v.). In mRNA production this involves removal of introns, addition of a 5' cap and polyadenylation.

RNA Ribonucleic acid; a polynucleotide that has ribose as its pentose sugar and uracil as one of its pyrimidines.

S1 nuclease An enzyme that specifically degrades single-stranded DNA.

SCIDS Severe combined immunodeficiency syndrome, a condition that results from a defective enzyme (Adenosine deaminase, q.v.).

Sequence tagged site Refers to a DNA sequence that is unique in the genome and which can be used in mapping studies. Usually identified by PCR amplification

Shine-Dalgarno sequence See *ribosome-binding site*

Short tandem repeat A DNA sequence with a sequential repeating set of two (di-), three (tri-), or four (tetra-) nucleotide pairs. Also called STR.

Short template A DNA strand that is synthesized during the polymerase chain reaction and has a primer sequence at one end and a sequence complementary to the second prime at the other end.

Shuttle vector A plasmid cloning vehicle,, usually a plasmid, that can replicate in two different organisms because it carries two different origins of replication. Also called a bifunctional vector.

Sigma factor An accessory bacterial protein(s) that directs the binding of RNA polymerase to specific promoters.

Signal peptide See Signal sequence

Single nucleotide polymorphism Polymorphic pattern at a single base, essentially the smallest polymorphic unit that can be identified

Single-cell protein Dried mass of a pure sample of a protein-rich microorganism, which may be used either as feed (for animals) or as a food (for humans). Also called SCP.

Single-locus probe Probe used in DNA fingerprinting that identifies a single sequence in the genome. Diploid organisms therefore usually show two bands in a fingerprint, one allelic variant from each parent.

Site-specific mutagenesis A technique to change one or more specific nucleotides in a cloned gene in order to create an altered form of protein with a specific amino acid change(s). Also called oligonucleotide-directed mutagenesis.

Somatic cell Body cell, as opposed to germ-line cell

Somatotropin Growth hormone, see also *bovine somatotropin*

Southern blotting A technique for transferring denatured DNA molecules that have been separated electrophoretically from a gel to a matrix (such as a nitrocellulose or nylon membrane) on which a hybridization assay can be performed.

Specific activity The amount of radioactivity per unit material, e.g. a labelled probe might have a specific activity of 10^6 counts/minute per microgram. Also used to quantify the activity of an enzyme

Splice site The nucleotides at (1) the end of an exon and the beginning of an intron and (2) the end of an intron and the beginning of the next exon that are required for the joining of two exons and removal of an intron during the processing of a primary transcript to a functional mRNA.

Staggered cuts Symmetrically cleaved phosphodiester bonds that lie on both strands of duplex DNA but are not opposite one another.

Structural gene A gene that encodes a protein product.

STS See *sequence tagged site*

Stuffer fragment The section in a replacement vector (q.v.) that is removed and replaced with insert DNA. See *polystuffer*

Subcloning Splicing part of a cloned DNA molecule into a different cloning vector.

Substitution vector See *replacement vector*

Suicide gene A plasmid-borne, inducible sequence that produces a protein that directly or indirectly kills the host cell.

Superbug Jargon for the bacterial strain of *Pseudomonas* developed by Dr. A. Chakrabarty, who combined a hydrocarbon-degrading gene carried on different plasmids into one organism. Although this genetically engineered microorganism is neither “super” nor a “bug”, it represents a landmark example because it showed how genetically modified microbial strains could be used in a novel way and because it was the basis for the precedent-setting legal decision that, in the United States, declared that genetically engineered organisms were patentable.

T4 DNA polymerase end labeling DNA that has been cut with a restriction endonuclease(s) is mixed with T4 DNA polymerase and one labeled deoxyribonucleotide. The 3' exonuclease activity of the T4 DNA polymerase removes deoxyribonucleotides from the 3' ends of the DNA fragments. Immediately after a deoxyribonucleotide that is the same as the deoxyribonucleotide in the reaction mix is cleaved off, the T4 DNA polymerase activity incorporates a labeled deoxyribonucleotide from the reaction mixture. No further incorporation of deoxyribonucleotides occurs because there is only one kind of deoxyribonucleotide in the reaction mixture.

Tandem repeat A repeat composed of an array of sequences repeated contiguously in the same orientation.

Taq polymerase Thermostable DNA polymerase from the thermophilic bacterium *Thermus aquaticus*. Used in the polymerase chain reaction (q.v.).

TATA box The DNA sequence, to which RNA polymerase binds, that lies upstream from the site of initiation of transcription and ensures that transcription starts at a specific nucleotide. Also called a Pribnow box in prokaryotes and a Hogness box in eukaryotes after the researchers who discovered the functions of the TATA box in prokaryotes and eukaryotes, respectively.

T-DNA The segment of a Ti plasmid that is transferred and integrated into chromosomal sites in the nuclei of plant cells.

Temperate Refers to bacteriophages that can undergo lysogenic infection of the host cell.

Terminal transferase An enzyme that adds nucleotide residues to the 3' terminus of an oligo- or poly-nucleotide

Tetracycline (Tc) A commonly used antibiotic

Thermal cycler Heating / cooling system for PCR applications. Enables denaturation, primer binding and extension cycles to be programmed and automated

Thermus aquaticus Thermophilic bacterium from which Taq polymerase (q.v.) is purified. Other bacteria from this genus include *Thermus favus* and *Thermus thermophilus*.

Thymine (T) Nitrogenous base found in DNA only.

T_i-plasmid Plasmid of *Agrobacterium tumefaciens* that causes crown gall disease (q.v.).

Tissue plasminogen activator (TPA) A protease that occurs naturally, and functions in breaking down blood clots. Acts on an inactive precursor (Plasminogen), which is converted to the active form (plasmin). This attacks the clot by breaking up fibrin, the protein involved in clot formation.

Totipotent A cell that can give rise to all cell types in an organism. Totipotency has been demonstrated by cloning carrots from somatic cells, and nuclear transfer experiments in animals.

Traitor technology See *genetic trait control technology*.

Trans-acting element A genetic element that can exert its effect without having to be on the same molecule as a target sequence. Usually such an element encodes a protein product (perhaps an enzyme or a regulatory protein) that can diffuse to the site of action.

Transcription (Tc) The synthesis of RNA from a DNA template

Transcription factor A protein that facilitates RNA synthesis by binding to a specific DNA sequence or another transcription factor that is bound to a specific DNA sequence.

Transcription frequency The fraction of a cell population that takes up foreign DNA; expressed as the number of transformed cells divided by the total number of cells in a population.

Transcription mapping Assigning gene transcripts, in the form of cDNA clones or expressed sequence tags, to specific chromosome regions by fluorescence *in situ* hybridization, hybridization, polymerase chain reaction (PCR), analysis of somatic cell hybrid mapping panels, or other strategies. Also called transcript mapping, transcriptional mapping.

Transcriptional unit The DNA sequence that encodes the RNA molecule, i.e. from the transcription start site to the stop site.

Transcriptome The population of RNA molecules (usually mRNAs) that is expressed by a particular cell type. Cf. *genome, proteome*.

Transduction The transfer of nonviral DNA by a virus to a cell.

Transfection Introduction of purified phage or virus DNA into cells.

Transfer RNA (tRNA) A small RNA (~ 75-85 bases) that carries the anticodon and the amino acid residue required for protein synthesis

Transformant A cell that has been transformed by exogenous DNA

Transformation (1) The uptake and establishment of DNA in a bacterium or yeast cell, in which the introduced DNA often changes the phenotype of the recipient organism. (2) Conversion, by various means, of animal cells in tissue culture from controlled to uncontrolled cell growth.

Transformation efficiency The number of cells that take up foreign DNA as a function of the amount of added DNA; expressed as transformants per microgram of added DNA.

Transgene A gene from one source that has been incorporated into the genome of another organism. Often refers to a gene that has been introduced into a multicellular organism.

Transgenesis The introduction of a gene(s) into animal or plant cells that leads to the transmission of the input gene (transgene) to successive generations.

Transgenic An organism that carries DNA sequences that it would not normally have in its genome

Transgenic animal A fertile animal that carries an introduced gene(s) in its germ line.

Transgenic plant A fertile plant that carries an introduced gene(s) in its germ line.

Translation The process of protein (polypeptide) synthesis in which the amino acid sequence of a protein is determined by mRNA mediated by tRNA molecules, and carried out on ribosomes.

Translocation (1) The movement of peptidyl-tRNA and mRNA from the aminoacyl site to the peptidyl site on the ribosome during the elongation phase of translation; this movement opens the aminoacyl site for the next codon. (2) The transfer of chromosome material from one chromosome to another. (3) The movement of compounds through a plant.

Transposable element A genetic element that carries the information that allows it to integrate at various sites in the genome. Transposable elements are sometimes called 'jumping genes'.

Transposase An enzyme that is encoded by a transposon gene and that facilitates the insertion of the transposon into a new chromosomal site and excision from a site.

Transposon A DNA sequence (mobile genetic element) that can insert randomly into a chromosome, exit the site, and relocate at another chromosomal site. For example, Tn5 is a bacterial transposon that carries the genes for resistance to the antibiotics neomycin and kanamycin and the genetic information for its insertion and excision. Also called transposable element.

Trisomy Aneuploid (q.v.) condition where an extra chromosome is present. Common example is the trisomy-21 condition that causes Down syndrome

Upstream (1) In molecular biology, the stretch of DNA base pairs that lie in the 5' direction from the site of initiation of transcription. Usually, the first transcribed base is designated +1 and the upstream nucleotides are indicated with minus signs, e.g., -1, -10. Also, to the 5'side of a particular gene or sequence of nucleotides. (2) In chemical engineering, those phases of a manufacturing process that precede the biotransformation step. Refers to the preparation of raw materials for a fermentation process. Also called upstream processing.

Uracil (U) Nitrogenous base found in RNA only.

Variable number tandem repeat (VNTR) Repetitive DNA composed of a number of copies of a short sequence, involved in the generation of polymorphic loci that are useful in genetic

fingerprinting. Also known as hypervariable regions. See also *minisatellite* and *microsatellite* DNA.

Vector A DNA molecule that is capable of replication in a host organism, and can act as a carrier molecule for the construction of recombinant DNA

Virulent Refers to bacteriophages that cause lysis of the host cell

Virus An infectious agent that cannot replicate without a host cell.

VNTR See *variable number tandem repeat*.

Western blotting Transfer of electrophoretically separated proteins onto a membrane for probing with antibody.

Wild type A genetic term that denotes the most commonly observed phenotype, or the normal state, in contrast to a mutant.

X linkage When a gene is present on the X chromosome.

Xenobiotic A chemical compound that is not produced by living organisms; a manufactured chemical compound.

Xenogeneic From a different species or individual (an attribute of cells or tissue). Also called Xenogenic.

Xenotransplantation The use of tissues or organs from a non-human source for transplantation

X-gal 5-Bromo-4-chloro-3indolyl- β -D-galactopyranoside: a chromogenic substrate for β -D-galactosidase; on celavage it yields a blue-coloured product

X-linked Pattern of inheritance where the allele is located on the X-chromosome. In humans, this can result in males expressing recessive characters that would normally be masked in an autosomal heterozygote

Yeast artificial chromosome (YAC) A yeast-based vector system for cloning large (>100 kb) DNA inserts. Also called YAC.

Yeast episomal vector A cloning vector for the yeast *Saccharomyces cerevisiae* that uses the 2 μ m plasmid origin of replication and is maintained as an extrachromosomal nuclear DNA molecule.

Zygote Single-celled product of the fusion of a male and a female gamete (q.v.). Develops into an embryo by successive mitotic divisions.