HANDOUTS

BIO732: Gene Manipulation and Genetic Engineering

Virtual University of Pakistan

Recommended Books

i). Principles of Gene Manipulation (Old and Primrose, 1994; Old et al., 2001)
ii). Principles of Gene Manipulation and Genomics (Primrose and Twyman, 2006)
iii). An Introduction to Genetic Engineering (Nicholl, 2002)
iv). Lehninger Principles of Biochemistry (Nelson and Cox, 2005)

Topic No. 1

Introduction to Gene manipulation and genetic engineering

"Gene manipulation is also called as genetic engineering and it refers to the alteration of genes of an organism. It involves manually adding new DNA to an organism to introduce novel traits. A variety of microorganisms, plants and animals has been generated with novel traits by using this technique. Overall, in this course, we will learn about the basic techniques that are used for the isolation of target genes, their analysis by gel electrophoresis and PCR amplification. It addition to that, cutting and joining of target DNA molecules or vectors is also an important component of this discipline. Transformation of target organism and expression of foreign genes will also be focused. We will also learn different techniques to modify microbes, plants and animals. Finally, we also discuss the applications of genetic engineering to get beneficial end products or to cure different types of diseases.

Topic 2

What is genetic engineering?

Genetic engineering is the process of manually adding new DNA to an organism. The goal is to add one or more new traits that are not already found in that organism. However, there are several other terms that can be used to describe the technology, including gene manipulation, gene cloning, recombinant DNA technology, genetic modification etc. In most western countries there is a precise legal definition of gene manipulation

"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 are capable of continued propagation"

Genetic engineering has application in both basic and applied research. In basic research, it is used to study mechanisms of gene replication and expression in prokaryotes, eukaryotes and their viruses. In applied research, it is used to develop microbial cultures that produce valuable metabolic products. It is also important to generate transgenic organisms, in environmental biotechnology and to treat genetic disorders (gene therapy).

Topic 3

Genes and genome

A genome is an organism's complete set of DNA, including all genes. Genome contains all of the information needed to build that organism and allow it to grow and develop. Instructions in our genome are made up of DNA-unique chemical code. Both DNA and RNA contain two major purine bases i.e. adenine (A) and Guanine (G) and two major pyrimidines. In DNA and RNA, one of the pyrimidine-cytosine (C) is common but second pyrimidine in DNA is thymine (T) and in RNA it is uracil (U).

Gene on the other hand, is the nucleic acid sequence that is necessary for the synthesis of a functional gene product i.e. polypeptide or RNA. A cell typically has many thousands of genes and DNA molecules not surprisingly tend to be very large. The storage and transmission of biological information are the only know functions of DNA.

Topic 4

The flow of genetic information

The flow of genetic information in a living system is called central dogma of molecular biology (Fig. 1). Genetic information is encoded in DNA that is replicated before cell division. Genetic information flows from DNA to RNA and then to protein. This flow of information is dependent on the genetic code (a sequence of triplet bases), which defines the relation between the sequence of bases in DNA or mRNA and the sequence of amino acids in a protein. The code is nearly the same in all organisms: a sequence of three bases, called a codon, specifies an amino acid. Codons in mRNA are read sequentially by respective tRNA molecules, which serve as adaptors in protein synthesis. Protein synthesis takes place on ribosomes, which are complex assemblies of rRNAs and more than 50 kinds of proteins.

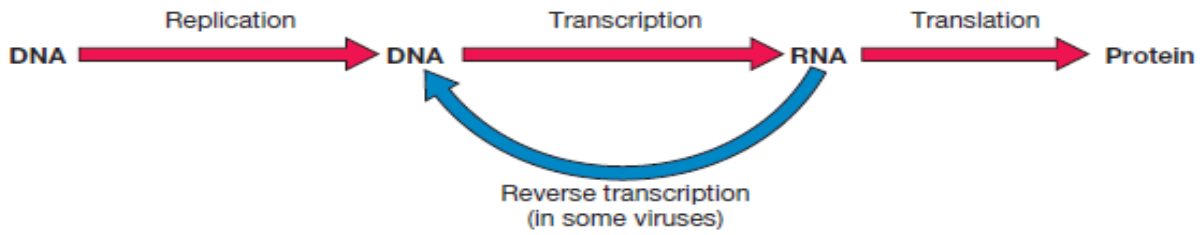

Figure 1. Flow of genetic information in living system.

Topic 5

Molecular cloning

Molecular cloning is at the base of most genetic engineering procedures. The basic strategy of molecular cloning is to move the desired gene from a large complex genome to a small, simple one. Fortunately, our knowledge of DNA chemistry and enzymology allows us to break and join DNA molecules *in vitro*. This process is known as *in vitro* recombination. Restriction enzymes, DNA ligase and synthetic DNA are important tools used for *in vitro* recombination. Molecular cloning can be divided into several steps as elaborated in figure 2.

i). Isolation and fragmentation of source DNA. This can be total genomic DNA from an organism of interest, DNA synthesized from an RNA template by reverse transcriptase, DNA synthesized by the polymerase chain reaction (PCR) and even DNA synthesized *in vitro*. If genomic DNA is the source, it is generally cut with restriction enzymes to give a mixture of fragments.

ii). Joining of DNA fragments to a cloning vector with DNA ligase. The small, independently replicating genetic elements used to replicate genes are known as cloning vectors. Cloning vectors are generally designed to allow recombination of foreign DNA at a restriction site that cut the vector in a way that does not affect its replication. If the source DNA and the vector are cut with the same restriction enzyme, joining can be mediated by annealing of the single stranded regions called “sticky ends”.

iii). Introduction and maintenance of a host organisms. The recombinant DNA molecule made in a test tube is introduced into a host organism, for example, by DNA transformation, where it can replicate. Transfer of the DNA into the host usually yields a mixture of clones. Some cells contain the desired cloned gene, whereas, other cells contain other clones generated by joining the source DNA to the vector. Such a mixture is known as a DNA library or a gene library because many different clones can be purified from the mixture, each containing different cloned DNA segments from the source organism.

iv). Detection and purification of the desired clone. Often one of the most difficult tasks is finding the right clone in a mixture that may contain thousands of others.

v). Production of large numbers of cells or bacteriophage containing the desired clone for isolation and study of the cloned DNA.

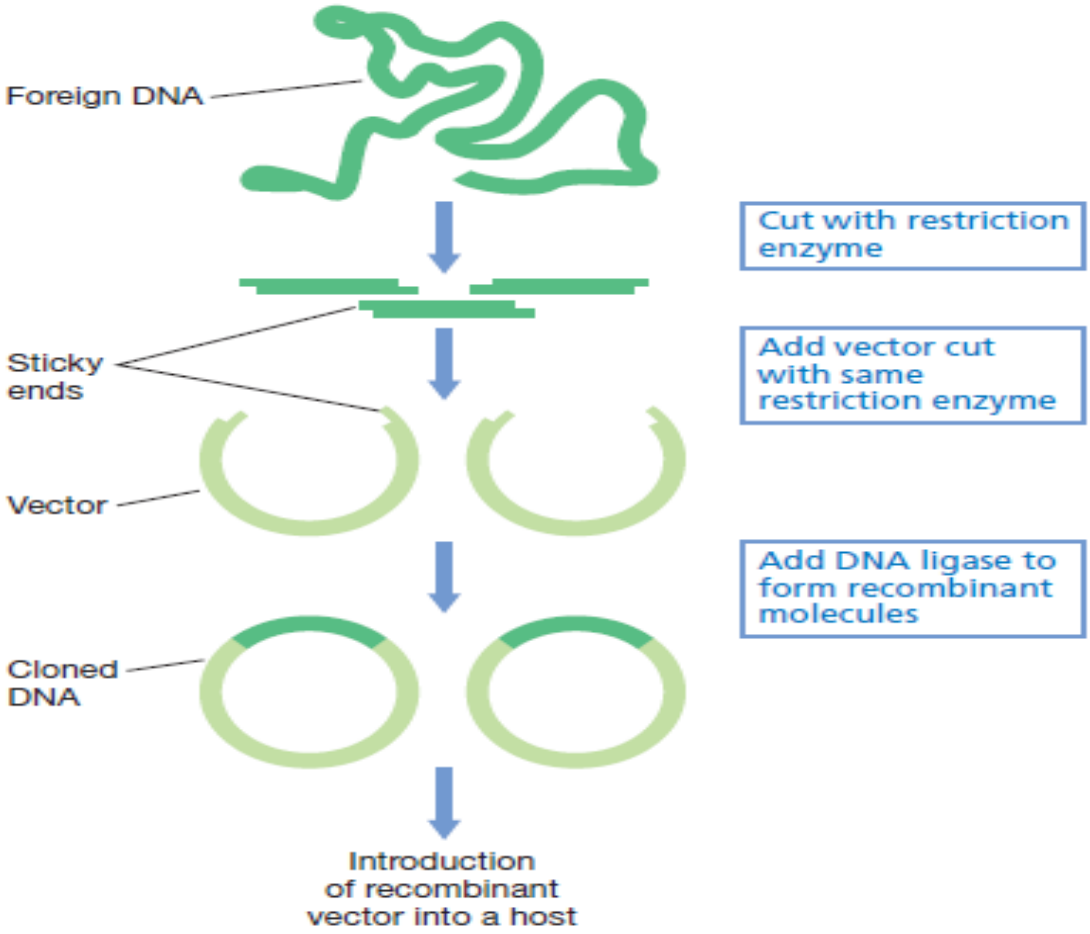

Figure 2. Major steps in gene cloning

Topic 6

Basic techniques: Introduction

The initial impetus for gene manipulation *in vitro* came about in the early 1970s with the simultaneous development of techniques for:

i). genetic transformation of *Escherichia coli*;

ii). cutting and joining DNA molecules;

iii). monitoring the cutting and joining reactions.

Before the advent of these modern gene-manipulation techniques there had been many early attempts at transforming bacterial or eukaryotic cells with foreign DNA. But, in general, little progress could be made. The reasons for this are as follows. Let us assume that the exogenous

DNA is taken up by the recipient cells. There are then two basic difficulties. First, where detection of uptake is dependent on gene expression, failure could be due to lack of accurate transcription or translation. Secondly, and more importantly, the exogenous DNA may not be maintained in the transformed cells. If the exogenous DNA is integrated into the host genome, there is no problem. The exact mechanism whereby this integration occurs is not clear and it is usually a rare event. However this occurs, the result is that the foreign DNA sequence becomes incorporated into the host cell's genetic material and will subsequently be propagated as part of that genome. If, however, the exogenous DNA fails to be integrated, it will probably be lost during subsequent multiplication of the host cells. The reason for this is simple. In order to be replicated, DNA molecules must contain an origin of replication, and in bacteria and viruses there is usually only one per genome. Such molecules are called replicons. Fragments of DNA are not replicons and in the absence of replication will be diluted out of their host cells. It should be noted that, even if a DNA molecule contains an origin of replication, this may not function in a foreign host cell. There is an additional, subsequent problem. If the early experiments were to proceed, a method was required for assessing the fate of the donor DNA. In particular, in circumstances where the foreign DNA was maintained because it had become integrated in the host DNA, a method was required for mapping the foreign DNA and the surrounding host sequences.

Topic 7

The solutions: Basic Techniques

If fragments of DNA are not replicated, the obvious solution is to attach them to a suitable replicon that is also called as vectors or cloning vehicles. Small plasmids and bacteriophages are the most suitable vectors for they are replicons in their own right; their maintenance does not necessarily require integration into the host genome. Moreover, their DNA can be readily isolated in an intact form. Initially, plasmids and phage vectors were used that were mainly suitable for *E. coli*. An important consequence follows from the use of a vector to carry the foreign DNA: simple methods become available for purifying the vector molecule, complete with its foreign DNA insert, from transformed host cells. Thus not only does the vector provide the replicon function, but it also permits the easy bulk preparation of the foreign DNA sequence, free from host-cell DNA. Composite molecules in which foreign DNA has been inserted into a vector molecule are sometimes called DNA chimeras. The construction of such composite or

artificial recombinant molecules has also been termed genetic engineering or gene manipulation because of the potential for creating novel genetic combinations by biochemical means.

Topic 8

Isolation of genomic DNA from *E. coli*

The rapidly growing field of molecular biology and biotechnology has a tremendous need for quick, simple, robust and high-throughput procedures for extraction of DNA from diverse sources. The process of genomic DNA isolation and purification has evolved considerably within the last decade. The new demands of high-throughput facilities have resulted in the development of new technologies for easier and faster DNA processing than ever before. The chemical methods for DNA extraction generally rely on the use of phenol–chloroform but the process is toxic, time-consuming, multi-step and utilizes organic solvent extraction, alcohol precipitation as well as centrifugation. Now, many different methods are available for isolating genomic DNA, and a number of biotech companies sell reagent kits. For genomic DNA isolation, bacterial cells are cultivated in liquid medium overnight. Then cell pellet is separated by centrifugation at 15000 *g* for 2 min. The cells are treated with lysis solution which contains the anionic detergent sodium dodecyl sulphate (SDS) to disrupt membranes and denature proteins. After that, cell lysate is treated with RNase and protein precipitation solutions. Afterwards, DNA can be suspended in solution of isopropenol and then centrifuged. Finally, DNA threads can be made visible by dissolving the DNA pellet in 70% ethanol. DNA pellet can be separated by centrifugation and DNA finally can be dissolved in DNA rehydration solution.

Topic 9

Agarose gel electrophoresis

Agarose is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb. Polyacrlamide is preferred for smaller DNA fragments. The progress of the first experiments on cutting and joining of DNA molecules was monitored by velocity sedimentation in sucrose gradients. However, this has been entirely superseded by gel electrophoresis. Gel electrophoresis is not only used as an analytical method, it is routinely used preparatively for the purification of specific DNA fragments. The gel is composed of polyacrylamide or agarose. The migration of the DNA molecules through the pores of the matrix must play an important role in molecular weight separations since the electrophoretic mobility of DNA in free solution is independent of molecular weight. An agarose gel is a complex network

of polymeric molecules whose average pore size depends on the buffer composition and the type and concentration of agarose used. The larger the pore sizes of the gel, the greater the ball of DNA which can pass through and hence, the larger the molecules which can be separated. Once the globular volume of the DNA molecule exceeds the pore size, the DNA molecule can only pass through by reptation. This occurs with molecules about 20 kb in size and it is difficult to separate molecules larger than this without recourse to pulsed electrical fields. In pulsed-field gel electrophoresis (PFGE) molecules as large as 10 Mb can be separated in agarose gels. This is achieved by causing the DNA to periodically alter its direction of migration by regular changes in the orientation of the electric field with respect to the gel.

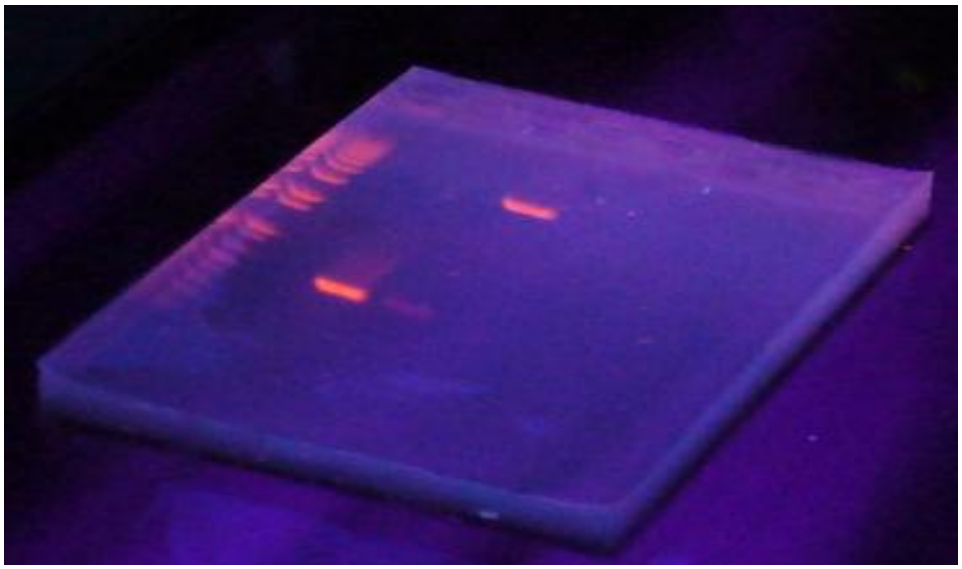

Figure 3. Visualization of gel under ultra violet light

Topic 10

Nucleic acid blotting

The hybridization of nucleic acids on membranes is a widely used technique in gene manipulation. Nucleic acid labeling and hybridization on membranes have formed the basis for a range of experimental techniques central to recent advances in our understanding of the organization and expression of the genetic material (Fig. 4). These techniques may be applied in the isolation and quantification of specific nucleic acid sequences and in the study of their organization, intracellular localization, expression and regulation. A variety of specific

applications includes the diagnosis of infectious and inherited disease. Blotting describes the immobilization of sample nucleic acids on to a solid support, generally nylon or nitrocellulose membranes. The blotted nucleic acids are then used as 'targets' in subsequent hybridization experiments. If the probe and target are 100% identical in sequence, then high stringency hybridization can be carried out. The stringency is determined by the hybridization temperature and the salt concentration in the hybridization buffer. Stringency can be regarded as the specificity with which a particular target sequence is detected by hybridization probe. Probe is a fragment of DNA or RNA which is used to detect the sequences of target DNA.

The main blotting procedures are blotting of nucleic acids from gels, dot and slot blotting, colony and plaque blotting. Overall, nucleic acid hybridization is of three types,

i). Southern blotting

ii). Northern blotting

iii). Western blotting

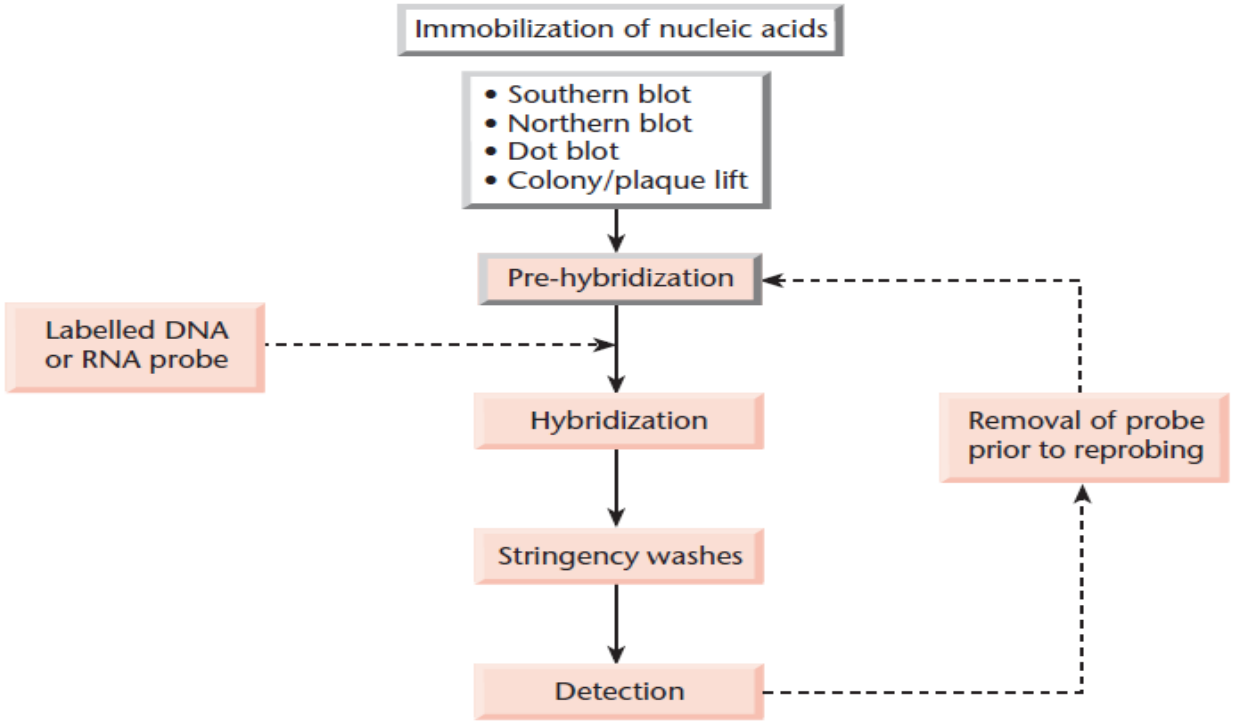

Figure 4. Flow diagram of nucleic acid blotting and hybridization

Topic 11

Autoradiography

It is a technique using X-ray film to visualize fragments of molecules that have been radioactively labeled. For example, it can be used to analyze the length and number of DNA fragments separated by gel electrophoresis. Autoradiography has a large number of practical applications in the biological, chemical, and physical sciences, because it provides both qualitative and quantitative information (eg., images and amounts present). It may be used to image large, small, and microscopic specimens, including sectioned whole organisms, organs, tissues, cellular structures, and nucleic acids that contain some radiolabelled compound.

Topic 12

Southern blotting

The original method of blotting was developed by Southern (1975) for detecting fragments in an agarose gel that are complementary to a given RNA or DNA sequence. In this procedure, referred to as Southern blotting, the agarose gel is mounted on a filter-paper wick which dips into a reservoir containing transfer buffer. The hybridization membrane is sandwiched between the gel and a stack of paper towels or other absorbent material, which serves to draw the transfer buffer through the gel by capillary action. The DNA molecules are carried out of the gel by the buffer flow and immobilized on the membrane (Fig. 5). Initially, the membrane material used was nitrocellulose. The main drawback with this membrane is its fragile nature. Supported nylon membranes have since been developed which have greater binding capacity for nucleic acids in addition to high tensile strength.

For efficient Southern blotting, gel pretreatment is important. Large DNA fragments (> 10 kb) require a longer transfer time than short fragments. To allow uniform transfer of a wide range of DNA fragment sizes, the electrophoresed DNA is exposed to a short depurination treatment (0.25 mol/l HCl) followed by alkali. This shortens the DNA fragments by alkaline hydrolysis at depurinated sites. It also denatures the fragments prior to transfer, ensuring that they are in the single-stranded state and accessible for probing. Finally, the gel is equilibrated in neutralizing solution prior to blotting.

After transfer, the nucleic acid needs to be fixed to the membrane and a number of methods are available. Oven baking at 80°C is the recommended method for nitrocellulose membranes and this can also be used with nylon membranes. Following the fixation step, the membrane is placed

in a solution of labelled (radioactive or non-radioactive) RNA, single-stranded DNA or oligodeoxynucleotide probe which is complementary in sequence to the blot transferred DNA band or bands to be detected. After the hybridization reaction has been carried out, the membrane is washed to remove unbound radioactivity and regions of hybridization are detected autoradiographically by placing the membrane in contact with X-ray film.

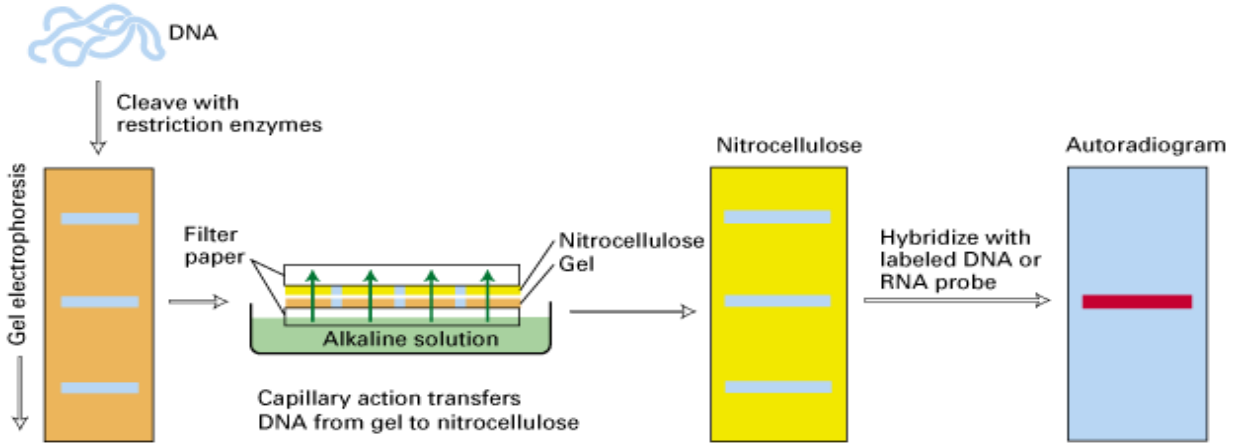

Figure 5. Southern blot hybridization

Topic 13

Northern blotting

A northern blot is a method used to detect specific RNA molecules among a mixture of RNA. Northern blotting can be used to analyze a sample of RNA from a particular tissue or cell type in order to measure the RNA expression of particular genes. Southern's technique has been of enormous value, but initially, it was thought that it could not be applied directly to the blot-transfer of RNAs separated by gel electrophoresis, since RNA was found not to bind to nitrocellulose. Alwine and coworkers (1979) therefore, devised a procedure in which RNA bands are blot-transferred from the gel on to chemically reactive paper, where they are bound covalently. The reactive paper is prepared by diazotization of aminobenzyloxymethyl paper (creating diazobenzyloxymethyl (DBM) paper), which itself can be prepared from Whatman 540 paper by a series of uncomplicated reactions. Subsequently, it was found that RNA bands can indeed be blotted on to nitrocellulose membranes under appropriate conditions and suitable nylon membranes have been developed.

Topic 14

Western blotting

It involves the transfer of electrophoresed protein bands from a gel on to a membrane. The bounded proteins are then available for analysis. Most commonly, antibodies are used to detect specific antigens. The term ‘western’ blotting refers to a procedure which does not directly involve nucleic acids, but which is of importance in gene manipulation. It involves the transfer of electrophoresed protein bands from a polyacrylamide gel on to nitrocellulose or nylon membranes, to which they bind strongly. The bound proteins are then available for analysis by a variety of specific protein–ligand interactions. Most commonly, antibodies are used to detect specific antigens. Lectins have been used to identify glycoproteins. In these cases, the probe may itself be labelled with radioactivity, or some other ‘tag’ may be employed. Often, however, the probe is unlabelled and is itself detected in a ‘sandwich’ reaction, using a second molecule which is labelled, for instance a species-specific second antibody, or protein A of *Staphylococcus aureus* (which binds to certain subclasses of IgG antibodies), or streptavidin (which binds to antibody probes that have been biotinylated). These second molecules may be labelled in a variety of ways with radioactive, enzyme or fluorescent tags. An advantage of the sandwich approach is that a single preparation of labelled second molecule can be employed as a general detector for different probes.

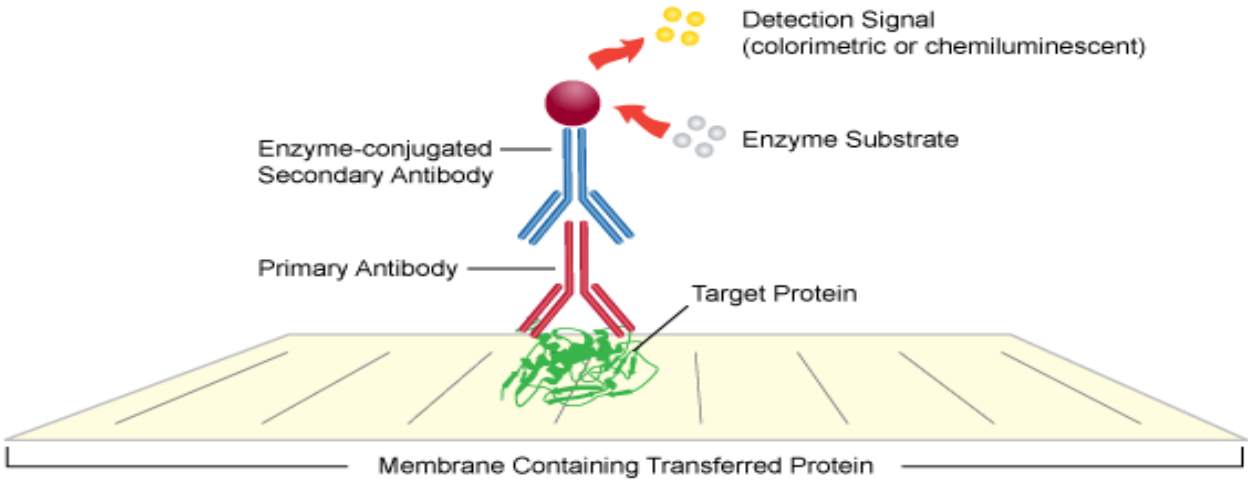

Figure 6. Detection in Western blotting

Topic 15

Transformation of *E. coli*

Transformation is the process of getting the recombinant vector from a reaction mixture or vector solution into *E. coli* cells. To enable the cells to take up circular vector DNA they have to be made competent. Early attempts to achieve transformation of *E. coli* were unsuccessful and it was generally believed that *E. coli* was refractory to transformation. However, later on it was found that treatment with $CaC1_2$ allowed *E. coli* cells to take up DNA from bacteriophage λ. A few years later, it was showed that $CaC1_2$ treated *E. coli* cells are also effective recipients for plasmid DNA. Almost any strain of *E. coli* can be transformed with plasmid DNA, albeit with varying efficiency. Since transformation of *E. coli* is an essential step in many cloning experiments, it is desirable that it be as efficient as possible.

A very simple, moderately efficient transformation procedure for use with *E. coli* involves re-suspending log-phase cells in ice-cold 50 mmol/l calcium chloride at about 10^{10} cells/ml and keeping them on ice for about 30 min. Plasmid DNA (0. 1 μg) is then added to a small aliquot (0.2 ml) of these now competent cells, and the incubation on ice continued for a further 30 min, followed by a heat shock of 2 min at 42°C. The cells are then usually transferred to nutrient medium and incubated for some time to allow phenotypic properties conferred by the plasmid to be expressed, e.g. antibiotic resistance commonly used as a selectable marker for plasmid containing cells.

Topic 16

Transformation with other organisms

E. coli often remains the host organism of choice for cloning experiments. However, many other hosts are now used, and with them transformation may still be a critical step. In the case of Gram-positive bacteria, the two most important groups of organisms are *Bacillus* spp. and actinomycetes. That *B. subtilis* is naturally competent for transformation has been known for a long time and hence the genetics of this organism are fairly advanced. For this reason, *B. subtilis* is a particularly attractive alternative prokaryotic cloning host.

With animal cells there is no great problem as only the membrane has to be crossed. In the case of yeast, protoplasts are required. With higher plants, one strategy that has been adopted is either to package the DNA in a plant virus or to use a bacterial plant pathogen as the donor. It has also been shown that protoplasts prepared from plant cells are competent for transformation. A

further remarkable approach that has been demonstrated with plants and animals is the use of microprojectiles shot from a gun. Animal cells, and protoplasts of yeast, plant and bacterial cells are susceptible to transformation by liposomes that were prepared from cationic lipid. Small unilamellar (single-bilayer) vesicles are produced. DNA in solution spontaneously and efficiently complexes with these liposomes. The positively charged liposomes not only complex with DNA, but also bind to cultured animal cells and are efficient in transforming them, probably by fusion with the plasma membrane. The use of liposomes as a transformation or transfection system is called lipofection.

Topic 17

Polymerase chain reaction (PCR)

The PCR involves two oligonucleotide primers, 17 to 30 nucleotides in length, which flank the DNA sequence that is to be amplified. The primers hybridize to opposite strands of the DNA after it has been denatured, and are orientated so that DNA synthesis by the polymerase proceeds through the region between the two primers. The extension reactions create two double stranded target regions, each of which can again be denatured ready for a second cycle of hybridization and extension. The third cycle produces two double stranded molecules that comprise precisely the target region in double-stranded form. By repeated cycles of heat denaturation, primer hybridization and extension, there follows a rapid exponential accumulation of the specific target fragment of DNA. After 22 cycles, an amplification of about 106 fold is expected. Many variations of PCR have been developed with applications throughout gene technology. PCR has revolutionized prenatal diagnosis by allowing tests to be performed using small samples of fetal tissue. In forensic science, the enormous sensitivity of PCR-based procedures is exploited in DNA profiling.

Topic 18

Cutting and joining DNA molecules

Before 1970, there was no method of fragmentation for DNA at specific points. All the available methods for cleaving DNA were non-specific. The available endonucleases had little site specificity and chemical methods produced very small fragments of DNA. The only method where any degree of control could be exercised was the use of mechanical shearing. The long, thin threads which constitute duplex DNA molecules are sufficiently rigid to be very easily broken by shear forces in solution. Intense sonication with ultrasound can reduce the length to

about 300 nucleotide pairs. More controlled shearing can be achieved by high speed stirring in a blender. Typically, high-molecular-weight DNA is sheared to a population of molecules with a mean size of about 8 kb by stirring at 1500 rev/min for 30 min. Breakage occurs essentially at random with respect to DNA sequence. The termini consist of short, single-stranded regions which may have to be taken into account in subsequent joining procedures.

During the 1960s, phage biologists elucidated the biochemical basis of the phenomenon of host restriction and modification. The culmination of this work was the purification of the restriction endonuclease of *Escherichia coli* K12 by Meselson and Yuan (1968). Since this endonuclease cuts unmodified DNA into large discrete fragments, it was reasoned that it must recognize a target sequence. This in turn raised the prospect of controlled manipulation of DNA. Unfortunately, the K12 endonuclease turned out to be perverse in its properties. While the enzyme does bind to a defined recognition sequence, cleavage occurs at a 'random' site several kilobases away. The much sought-after breakthrough finally came in 1970 with the discovery in *Haemophilus influenzae* of an enzyme that behaves more simply. That is, the enzyme recognizes a particular target sequence in a duplex DNA molecule and breaks the polynucleotide chain within that sequence to give rise to discrete fragments of defined length and sequence.

Topic 19

Host-controlled restriction/ modification

Restriction systems allow bacteria to monitor the origin of incoming DNA and to destroy it if it is recognized as foreign. Restriction endonucleases recognize specific sequences in the incoming DNA and cleave the DNA into fragments, either at specific sites or more randomly. When the incoming DNA is a bacteriophage genome, the effect is to reduce the efficiency of plating, i.e. to reduce the number of plaques formed in plating tests. The phenomena of restriction and modification were well illustrated and studied by the behaviour of phage λ on two *E. coli* host strains. If a stock preparation of phage λ, for example, is made by growth upon *E. coli* strain 'C' and this stock is then titred 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 lower. The phage are said to be restricted by the second host strain (*E. coli* K). 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'. Thus, the efficiency with which phage λ plates upon a particular host strain depends upon the strain on which it was last propagated. This non-heritable change conferred upon the phage by the second host strain (*E. coli* K) that allows it to be replated on that strain without further restriction is called modification.

The restricted phages adsorb to restrictive hosts and inject their DNA normally. When the phage are labelled with ^{32}P, it is apparent that their DNA is degraded soon after injection and the endonuclease that is primarily responsible for this degradation is called a restriction endonuclease or restriction enzyme. The restrictive host must, of course, protect its own DNA from the potentially lethal effects of the restriction endonuclease and so its DNA must be appropriately modified. Modification involves methylation of certain bases at a very limited number of sequences within DNA, which constitute the recognition sequences for the restriction endonuclease.

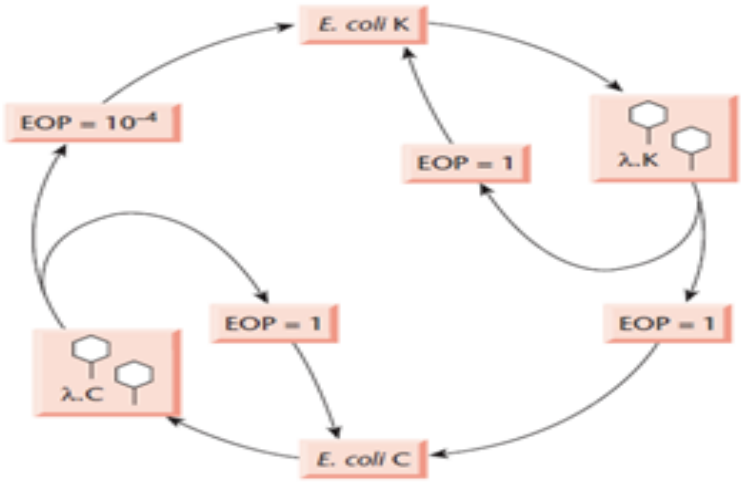

Figure 7. Host-controlled restriction and modification of phage ʎ

Topic 20

Types of restriction and modification

At least four different kinds of R-M system are known: type I, type II, type III and type IIs. The type I systems were the first to be characterized and a typical example is that from *E. coli* K12.

The active enzyme consists of two restriction subunits, two modification (methylation) subunits and one recognition subunit. The methylation and cutting reactions both require ATP and *S*-adenosylmethionine as cofactors. The recognition sequences are quite long with no recognizable features such as symmetry. The enzyme also cuts unmodified DNA at some distance from the recognition sequence. However, because the methylation reaction is performed by the same enzyme which mediates cleavage, the target DNA may be modified before it is cut. These features mean that type I systems are of little value for gene manipulation.

Most of the useful R-M systems are of type II. They have a number of advantages over type I and III systems. First, restriction and modification are mediated by separate enzymes so it is possible to cleave DNA in the absence of modification. Secondly, the restriction activities do not require cofactors such as ATP or *S*-adenosylmethionine, making them easier to use. Most important of all, type II enzymes recognize a defined, usually symmetrical, sequence and cut within it. Many of them also make a staggered break in the DNA and the usefulness of this will become apparent. Although type IIs systems have similar cofactors and macromolecular structure to those of type II systems, the fact that restriction occurs at a distance from the recognition site limits their usefulness.

Topic 21

Nomenclature

The discovery of a large number of restriction and modification systems called for a uniform system of nomenclature. Smith and Nathans (1973) introduced a suitable simplified version of this that included the following key features (Table 1).

i). The species name of the host organism is identified by the first letter of the genus name and the first two letters of the specific epithet to generate a three letter abbreviation. This abbreviation is always written in italics, e.g. *E. coli* = Eco; *H. influenzae* = Hin

ii). Where a particular strain has been the source then this is identified; e.g. *Eco*K

iii). When a particular host strain has several different R-M systems, these are identified by roman numerals. For instance, in case of *H. influenzae* different restriction enzymes can be represented as *HindI, HindII, HindIII*

Table 1. Examples of restriction endonuclease nomenclature

Enzyme s	Enzyme source	Recognition sequence
SmaI	*Serratia marcescens*, 1st enzyme	CCCGGG
HaeIII	*Haemophilus aegyptius*, 3rd enzyme	GGCC
HindII	*H. influenzae*, strain d, 2nd enzyme	GTPyPuAC
HindIII	*H. influenzae*, strain d, 3rd enzyme	AAGCTT
HamHI	*Bacillus amyloliquefaciens*, strain H, 1st enzyme	GGATCC

Topic 22

Target sites

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

5′-GAA|T TC-3′

3′-CT T|AAG-5′

Axis of symmetry

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

5′-G/AA* T T C-3′

3′-C TT A*A/G-5′

For convenience it is usual practice to simplify the description of recognition sequences by showing only one strand of DNA, that which runs in the 5′ to 3′ direction. Thus the *Eco*RI recognition sequence would be shown as G/AATTC (Fig. 8). From the information shown above we can see that *Eco*RI makes single-stranded breaks four bases apart in the opposite strands of its target sequence so generating fragments with protruding 5′ termini:

5′-G 5′-AATTC-3′

3′-CTTAA-5′ G-5′

These DNA fragments can associate by hydrogen bonding between overlapping 5′ termini, or the fragments can circularize by intramolecular reaction. For this reason the fragments are said to have sticky or cohesive ends. In principle, DNA fragments from diverse sources can be joined by means of the cohesive ends and, as we shall see later, the nicks in the molecules can be sealed to form an intact artificially recombinant DNA molecule. Not all type II enzymes cleave their target sites like *Eco*RI. Some, such as *Pst*I (CTGCA/G), produce fragments bearing 3′ overhangs, while others, such as *Sma*I (CCC/GGG), produce blunt or flush ends.

To date, over 10 000 microbes from around the world have been screened for restriction enzymes. From these, over 3000 enzymes have been found representing approximately 200 different sequence specificities. Occasionally enzymes with novel DNA sequence specificities are found but most prove to have the same specificity as enzymes already known. Restriction enzymes with the same sequence specificity and cut site are known as isoschizomers. Enzymes that recognize the same sequence but cleave at different points, for example *Sma*I (CCC/GGG) and *Xma*I C/CCGGG), are sometimes known as neoschizomers.

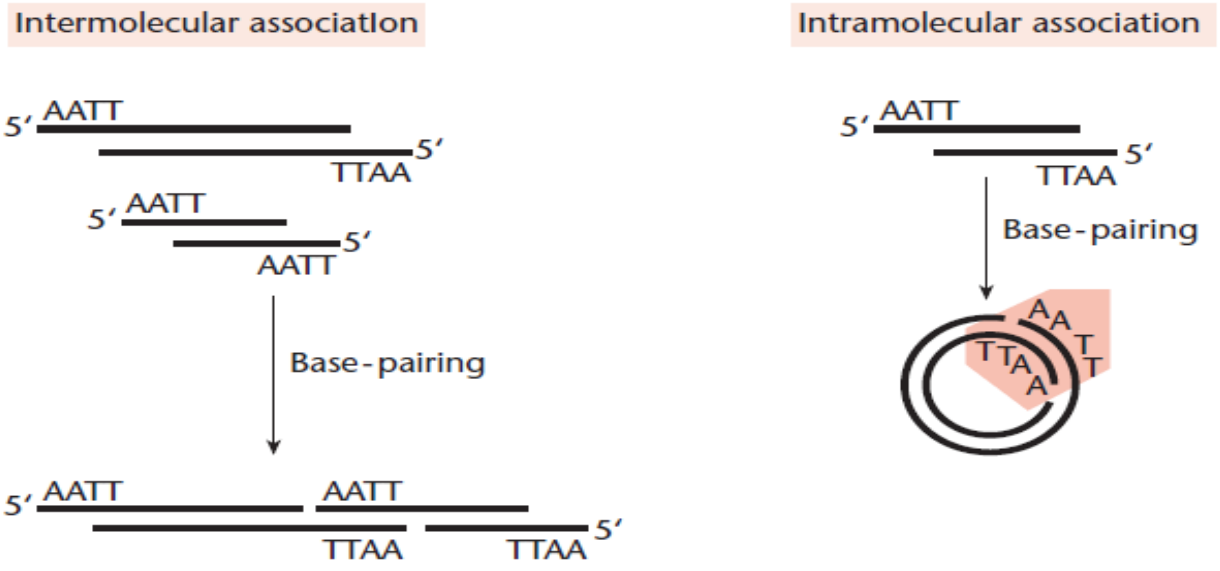

Figure 8. Cohesive fragments of DNA produced by digestion with EcoR1

Topic 23

Number and Size of restriction fragments

The number and size of the fragments generated by a restriction enzyme depend on the frequency of occurrence of the target site in the DNA to be cut. Assuming a DNA molecule with a 50% G+C content and a random distribution of the four bases, a four base recognition site occurs every 44 (256) bp. Similarly, a six-base recognition site occurs every 46 (4096) bp and an eight-base recognition sequence every 48 (65 536) bp. In practice, there is not a random distribution of the four bases and many organisms can be AT- or GC-rich, e.g. the nuclear genome of mammals is 40% G+C and the dinucleotide CG is fivefold less common than statistically expected. Similarly, CCG and CGG are the rarest trinucleotides in most A+T-rich bacterial genomes and CTAG is the rarest tetranucleotide in G+C-rich bacterial genomes. Thus different restriction endonucleases with six-base recognition sites can produce average fragment sizes significantly different from the expected 4096 bp.

Table 2. Average fragment size (bp) produced by different enzymes

Enzyme	Target	*Arabidopsis*	*E. coli*	Human
*Apa*I	GGGCCC	25000	15000	2000
*Bam*HI	GGATCC	6000	5000	5000
*Spe*I	ACTAGT	8000	60000	10000

Topic 24

Summary of restriction endonucleases

Restriction system allows bacteria to monitor the origin of incoming or foreign DNA continuously. If it recognizes as foreign, then it can by degraded. This cutting ability of the restriction enzymes can be exploited to fragment DNA for gene manipulation purposes. Microbes contain four types of restriction or modification (R-M) system that vary in their specificity. It included type I, II, III or IIs. Type I systems were the first to be characterized from *E. coli* K12. Type I systems are of little value for gene manipulation as they lack specificity. Most of the useful R-M system is Type II that recognizes a defined sequence and cut within it. Type III enzymes have symmetrical recognition sequences but otherwise resemble type I systems and are of little value. Type IIs systems have similar cofactors and structure to type II but

restriction occurs at a distance from recognition site that limits their usefulness. The number and size of fragments generated by restriction enzyme depend on the frequency of occurrence of the target site in the DNA to be cut.

Topic 25

DNA modifying enzymes

Restriction enzymes and DNA ligase provide the cutting and joining functions that are essential for the production of recombinant DNA molecules. Other enzymes used in genetic engineering may be loosely termed DNA modifying enzymes, with the term used here to include degradation, synthesis and alteration of DNA. Nuclease enzymes degrade nucleic acids by breaking the phosphodiester bond that holds the nucleotides together. Restriction enzymes are good examples of endonucleases, which cut within a DNA strand. A second group of nucleases, which degrade DNA from the termini of the molecule are known as exonucleases.

Polymerase enzymes synthesise copies of nucleic acid molecules and are used in many genetic engineering procedures. When describing a polymerase enzyme the terms ‘DNA-dependent’ or ‘RNA-dependent’ may be used to indicate the type of nucleic acid template that the enzyme uses. Thus a DNA dependent DNA polymerase copies DNA into DNA, an RNA-dependent DNA polymerase copies RNA into DNA, and a DNA-dependent RNA polymerase transcribes DNA into RNA. These enzymes synthesise nucleic acids by joining together nucleotides whose bases are complementary to the template strand bases. The synthesis proceeds in a 5’→3’ direction, as each subsequent nucleotide addition requires a free 3’-OH group for the formation of the phosphodiester bond. This requirement also means that a short double stranded region with an exposed 3-OH (a primer) is necessary for synthesis to begin.

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

DNA ligase is an important cellular enzyme, as its function is to repair broken phosphodiester bonds that may occur at random or as a consequence of DNA replication or recombination. In genetic engineering it is used to seal discontinuities in the sugar–phosphate chains that arise when recombinant DNA is made by joining DNA molecules from different sources. It can

therefore be thought of as molecular glue, which is used to stick pieces of DNA together. This function is crucial to the success of many experiments and DNA ligase is therefore a key enzyme in genetic engineering. The enzyme used most often in experiments is T4 DNA ligase, which is purified from *E. coli* cells infected with bacteriophage T4. Although the enzyme is most efficient when sealing gaps in fragments that are held together by cohesive ends, it will also join blunt-ended DNA molecules together under appropriate conditions.

Topic 26

Method of joining DNA fragments

Mainly three methods are used for joining DNA *in vitro:*

i). Joining covalently annealed cohesive ends by DNA ligase. DNA ligase facilitates the joining of DNA strands together by catalyzing the formation of a phosphodiester bond.

ii). Joining blunt-ended fragments by DNA ligase from phage T4 infected *E. coli*. T4 DNA ligase was purified from *E. coli* cells infected with bacteriophage T4. This enzyme is most efficient when sealing gaps in fragments that are held together by cohesive ends. However, it will also join blunt-ended DNA molecules together under appropriate conditions.

iii). The third utilizes the enzyme terminal deoxynucleotidyltransferase to synthesize homopolymeric 3' single stranded tails at the ends of fragments.

Topic 27

DNA ligase to create covalent recombinant DNA

E. coli and phage T4 encode an enzyme, DNA ligase, which seals single-stranded nicks between adjacent nucleotides in a duplex DNA chain. Although the reactions catalysed by the enzymes of *E. coli* and T4 infected *E. coli* are very similar, they differ in their cofactor requirements. The T4 enzyme requires ATP, while the *E. coli* enzyme requires NAD+. In each case the cofactor is split and forms an enzyme-MP 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. When termini created by a restriction endonuclease that creates cohesive ends associate, the joint has nicks a few base pairs apart in opposite strands. DNA ligase can then repair these nicks to form an intact duplex. This reaction, performed *in vitro* with purified DNA ligase, is fundamental to many gene-manipulation procedures.

The optimum temperature for ligation of nicked DNA is 37°C, but at this temperature the hydrogen bonded join between the sticky ends is unstable. *Eco*RI generated termini associate through only four AT base pairs and these are not sufficient to resist thermal disruption at such a high temperature. 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 experimentally to be in the range 4-15°C.

Topic 28

Alkaline phosphatase

Alkaline phosphatase is an enzyme responsible for removing phosphate groups from many types of molecules including DNA. Treatment of linearized plasmid vector DNA with alkaline phosphatase to remove 5′ terminal phosphate groups will prevent recircularization and plasmid dimer formation. The ligation reaction can be performed so as to favor the formation of recombinants. First, the population of recombinants can be increased by performing the reaction at a high DNA concentration; in dilute solutions *circularization* of linear fragments is relatively favored because of the reduced frequency of intermolecular reactions. Secondly, by treating linearized plasmid vector DNA with alkaline phosphatase to remove 5′-terminal phosphate groups, both recircularization and plasmid dimer formation are prevented. In this case, circularization of the vector can occur only by insertion of nonphosphatase- treated foreign DNA which provides one 5′-terminal phosphate at each join. One nick at each join remains unligated, but, after transformation of host bacteria, cellular repair mechanisms reconstitute the intact duplex.

Topic 29

Blunt end ligation via linker molecules

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process which has been used extensively to create artificial recombinants. A modification of this procedure depends upon the ability of T4 DNA ligase to join blunt-ended DNA molecules. The *E. coli* DNA ligase will not catalyze blunt ligation except under special reaction conditions of macromolecular crowding. Blunt ligation is most usefully applied to joining blunt-ended fragments via *linker* molecules. Scheller and co-workers (1977) synthesized self-complementary decameric oligonucleotides, which contain sites for one or more restriction endonucleases. The

molecule can be ligated to both ends of the foreign DNA to be cloned, and then treated with restriction endonuclease to produce a sticky-ended fragment, 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 enzyme target sites at each end of the foreign DNA and so enables the foreign DNA to be excised and recovered after cloning and amplification in the host bacterium.

Topic 30

Adaptors

Chemically synthesized adaptor molecules which have a preformed cohesive end can be used to insert blunt ended foreign DNA into vector. Adaptors and linkers are short double stranded DNA molecules that permit different cleavage sites to be interconnected. Under some circumstances, restriction enzyme used to generate the cohesive ends in the linker will also cut the foreign DNA at internal sites. In this situation, the foreign DNA will be cloned as two or more subfragments. One solution to this problem is to choose another restriction enzyme, but there may not be a suitable choice if the foreign DNA is large and has sites for several restriction enzymes. Another solution is to methylate internal restriction sites with the appropriate modification methylase. Alternatively, a general solution to the problem is provided by chemically synthesized adaptor molecules which have a preformed cohesive end. Consider a blunt-ended foreign DNA containing an internal *Bam*HI site, which is to be cloned in a *Bam*HI-cut vector. The *Bam*HI adaptor molecule has one blunt end bearing a 5′ phosphate group and a *Bam*HI cohesive end which is not phosphorylated. The adaptor can be ligated to the foreign DNA ends. The foreign DNA plus added adaptors is then phosphorylated at the 5′ termini and ligated into the *Bam*HI site of the vector. If the foreign DNA were to be recovered from the recombinant with *Bam*HI, it would be obtained in two fragments. However, the adaptor is designed to contain two other restriction sites (*Sma*I, *Hpa*II), which may enable the foreign DNA to be recovered intact. The only difference between an adaptor and a linker is that the former has cohesive ends and the latter has blunt ends. A wide range of adaptors is available commercially.

Topic 31

Homopolymer tailing

A general method for joining DNA molecules makes use of the annealing of complementary homopolymer sequences. Thus, by adding oligo(dA) sequences to the 3′ ends of one population of DNA molecules and oligo(dT) blocks to the 3′ ends of another population, the two types of molecule can anneal to form mixed dimeric circles. An enzyme purified from calf thymus, terminal deoxynucleotidyltransferase provides the means by which the homopolymeric extensions can be synthesized, for if presented with a single deoxynucleotide triphosphate it will repeatedly add nucleotides to the 3′ OH termini of a population of DNA molecules. DNA with exposed 3′ OH groups, such as arise from pretreatment with phage λ exonuclease or restriction with an enzyme such as *Pst*I, is a very good substrate for the transferase.

Topic 32

Cloning of cDNA by Homopolymer tailing

Eukaryotic mRNA can be cloned in vector after converting it to cDNA by using an enzyme reverse transcriptase. Like true DNA polymerases, reverse transcriptase can only synthesize a new DNA strand if provided with a growing point in the form of a preexisting primer which is base paired with the template and bears a 3′-OH group. Fortunately, most eukaryotic mRNA occur naturally in a polyadenylated form with up to 200 adenylate residues at their 3′-termini and hence, we can provide a primer simply by hybridizing a shor oligo (dT) molecule with this poly (A) sequence. The primer is then suitably located for synthesis of a complete cDNA by reverse transcriptase in the presence of all four deoxynucleodie triphoshates. The immediate product of the reaction is an RNA-DNA hybrid. The RNA strand can then be destroyed by alkaline hydrolysis, to which DNA is resistant, leaving a single stranded cDNA which can be converted into double stranded form in a second DNA polymerase reaction. After synthesis of cDNA that is tailed with oligo (dC), it can be annealed with the pBR322 vector which has been cut open with *Pst*I and tailed with oligo (dG). It will be seen that these homopolymers have been chosen so that *Pst*I target sites are reconstructed in the recombinant molecule, thus providing a simple means for excising the inserted sequences after amplification.

Topic 33

Plasmid biology

Plasmids are circular DNA molecules that replicate separately from the host chromosome. Naturally occurring bacterial plasmids range in size from 5,000 to 400,000 bp. They can be introduced into bacterial cells by a process called transformation. Bacterial cells generally of *E.*

coli and plasmid DNA are incubated together at 0 ^{o}C in a calcium chloride solution, then subjected to a shock by rapidly raising the temperature to 37 to 43 ^{o}C. For reasons not well understood, some of the cells treated in this way take up the plasmid DNA. Some species of bacteria are naturally competent for DNA uptake and do not require the calcium chloride treatment. In an alternative method, cells incubated with the plasmid DNA are subjected to a high-voltage pulse. This approach, called electroporation, transiently renders the bacterial membrane permeable to large molecules.

Topic 34

Interconversion of plasmid DNA

Plasmids are replicons which are stably inherited in an extrachromosomal state. Most plasmids exist as double-stranded circular DNA molecules. If both strands of DNA are intact circles the molecules are described as covalently closed circles or CCC DNA. If only one strand is intact, then the molecules are described as open circles or OC DNA. When isolated from cells, covalently closed circles often have a deficiency of turns in the double helix, such that they have a supercoiled configuration. The enzymatic interconversion of supercoiled, relaxed CCC DNA and OC DNA is shown in figure. Not all plasmids exist as circular molecules. Linear plasmids have been found in a variety of bacteria, e.g. *Streptomyces* sp. and *Borrelia* sp. To prevent nuclease digestion, the ends of linear plasmids need to be protected and two general mechanisms have evolved. Either there are repeated sequences ending in a terminal DNA hairpin loop (*Borrelia*) or the ends are protected by covalent attachment of a protein (*Streptomyces*).

Topic 35

Supercoiling of DNA

DNA bands can be visualized by soaking the gel in a solution of ethidium bromide. Due to the differences in their structural configurations, supercoiled and OC DNA separate upon electrophoresis in agarose gels. Addition of an intercalating agent, such as ethidium bromide, to supercoiled DNA causes the plasmid to unwind. If excess ethidium bromide is added, the plasmid will rewind in the opposite direction. Use of this fact is made in the isolation of plasmid DNA.

Topic 36

Phenotypic traits exhibited by plasmids

Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 daltons to greater than 200×10^6, and are generally dispensable. Some of the phenotypes which these plasmids confer on their host cells are listed in Table. Phyenotypic traits are mostly related to antibiotic resistance or production, heavy metal resistance, degradation of aromatic compounds and haemolysin production. Most of these traits; especially, antibiotic resistance can be used in gene cloning experiments to select transformants. Plasmids to which phenotypic traits have not yet been ascribed are called *cryptic* plasmids.

Topic 37

Properties of conjugative and non-conjugative plasmids

Plasmids can be classified into one of two major type i.e. conjugative or non-conjugative plasmids. It depends whether or not they carry a set of transfer genes, called the *tra* genes that mediate the process of bacterial conjugation. Plasmids can also be categorized on the basis of their being maintained as multiple copies per cell called relaxed plasmids or as a limited number of copies per cell i.e. stringent plasmids. Generally, conjugative plasmids are of relatively high molecular weight and are present as one to three copies per chromosome. On the other side, non-conjugative plasmids are of low molecular weight and present as multiple copies per chromosome. An exception is the conjugative plasmid R6K, which has a molecular weight of 25 × 106 daltons and is maintained as a relaxed plasmid.

Topic 38

Host range of plasmids

Plasmids encode only a few of the proteins required for their own replication. All the other proteins required for replication, e.g. DNA polymerases, DNA ligase, helicases, etc., are provided by the host cell. Those replication proteins that are plasmid-encoded are located very close to the *ori* (origin of replication) sequences at which they act. Thus, only a small region surrounding the *ori* site is required for replication. Other parts of the plasmid can be deleted and foreign sequences can be added to the plasmid and replication will still occur. This feature of plasmids has greatly simplified the construction of versatile cloning vectors. The host range of a plasmid is determined by its *ori* region. Plasmids whose *ori* region is derived from plasmid Col E1 have a restricted host range: they only replicate in enteric bacteria, such as *E. coli,*

Salmonella, etc. Other plasmids have a broad host range and these include RP4 and RSF1010. Plasmids of the RP4 type will replicate in most Gram-negative bacteria, to which they are readily transmitted by conjugation. Such promiscuous plasmids offer the potential of readily transferring cloned DNA molecules into a wide range of genetic backgrounds. Plasmids like RSF1010 are not conjugative but can be transformed into a wide range of Gram-negative and Gram-positive bacteria, where they are stably maintained. Many of the plasmids isolated from *Staphylococcus aureus* also have a broad host range and can replicate in many other Gram-positive bacteria. Plasmids with a broad host range encode most, if not all, of the proteins required for replication.

Topic 39

Partioning and segregative stability of plasmids

The loss of plasmids due to defective partitioning is called segregative instability. Naturally occurring plasmids are stably maintained because they contain a partitioning function, *par*, which ensures that they are stably maintained at each cell division. Such *par* regions are essential for stability of low copy-number plasmids. The higher-copy-number plasmid Col E1 also contains a *par* region but this is deleted in many Col E1-derived cloning vectors, e.g. pBR322. Although the copy number of vectors such as pBR322 is usually high, plasmid-free cells arise under nutrient limitation or other stress conditions. The *par* region from a plasmid such as pSC101 can be cloned into pBR322: thereby, stabilizing the plasmid. DNA superhelicity is involved in the partitioning mechanism. pSC101 derivatives lacking the *par* locus show decreased overall superhelical density as compared with wild-type pSC101. Partition-defective mutants of pSC101 and similar mutants of unrelated plasmids are stabilized in *Eseherichia coli* by *top*A mutations, which increase negative DNA supercoiling. Conversely, DNA gyrase inhibitors and mutations in DNA gyrase increase the rate of loss of *par*-defective pSC101 derivatives.

Topic 40

Incompatibility of plasmids

Plasmid incompatibility is the inability of two different plasmids to coexist in the same cell in the absence of selection pressure. The term incompatibility can only be used when it is certain that entry of the second plasmid has taken place and that DNA restriction is not involved. Groups of

plasmids which are mutually incompatible are considered to belong to the same incompatibility group. Over 30 incompatibility groups have been defined in *E. coli* and 13 for plasmids of *S. aureus.* Plasmids will be incompatible if they have the same mechanism of replication control.

Topic 41

Purification of plasmid DNA

One of the basic requirements for cloning in plasmids is the purification of the plasmid DNA. Although a wide range of plasmid DNAs are now routinely purified, the methods used are not without their problems. Undoubtedly the trickiest stage is the lysis of the host cells; both incomplete lysis and total dissolution of the cells result in greatly reduced recoveries of plasmid DNA. Many methods are available for isolating pure plasmid DNA from cleared lysates. For instance, 'classical' method involves isopycnic centrifugation of cleared lysates in a solution of CsCl containing ethidium bromide (EtBr). EtBr binds by intercalating between the DNA base pairs, and in so doing causes the DNA to unwind. A CCC DNA molecule, such as a plasmid, has no free ends and can only unwind to a limited extent, thus limiting the amount of EtBr bound. A linear DNA molecule, such as fragmented chromosomal DNA, has no such topological constraints and can therefore bind more of the EtBr molecules. Because the density of the DNA-EtBr complex decreases as more EtBr is bound, and because more EtBr can be bound to a linear molecule than to a covalent circle, the covalent circle has a higher density at saturating concentrations of EtBr. Thus covalent circles (i.e. plasmids) can be separated from linear chromosomal DNA.

Currently the most popular method of extracting and purifying plasmid DNA is that of Birnboim and Doly (1979). This method makes use of the observation that there is a narrow range of pH (12.0–12.5) within which denaturation of linear DNA, but not covalently closed circular DNA, occurs. Plasmid containing cells are treated with lysozyme to weaken the cell wall and then lysed with sodium hydroxide and sodium dodecyl sulphate (SDS). Chromosomal DNA remains in a high-molecular-weight form but is denatured. Upon neutralization with acidic sodium acetate, the chromosomal DNA renatures and aggregates to form an insoluble network. Simultaneously, the high concentration of sodium acetate causes precipitation of protein-SDS complexes and of high molecular-weight RNA. Provided the pH of the alkaline denaturation step has been carefully controlled, the CCC plasmid DNA molecules will remain in a native state and in

solution, while the contaminating macromolecules co-precipitate. The precipitate can be removed by centrifugation and the plasmid concentrated by ethanol precipitation. If necessary, the plasmid DNA can be purified further by gel filtration.

Topic 42

Desirable properties of plasmids

Plasmids replicate independently of the host chromosome. In addition to carrying genes required for their own replication, most plasmids are natural vectors because they often carry other genes that confer important properties on their hosts. Plasmids have very useful properties as cloning vectors. These properties include:

i). Small size, which makes the DNA easy to isolate and manipulate

ii). Circular DNA, which makes the DNA more stable during chemical isolation

iii). Independent origin of replication so plasmid replication in the cell proceeds independently from direct chromosomal control

iv). Multiple copy number, so they can be presented in the cell in several or numerous copies, making amplification of DNA possible

v). The presence of selectable markers such as antibiotic resistance genes, making detection and selection of plasmid-containing clones easier.

Topic 43

Natural plasmids as cloning vehicles

The term natural is used loosely to describe plasmids which were not constructed *in vitro* for the sole purpose of cloning. Col E1 is a naturally occurring plasmid which specifies the production of a bacteriocin, colicin E1. By necessity this plasmid also carries a gene which confers on host cells immunity to colicin E1. RSF 2124 is a derivative of Col E1 which carries a transposon specifying ampicillin resistance. For long enough, the origin of pSC101 was not clear. It is now know to be a natural plasmid of *Salmonell panama*. For cloning in pSC101, the plasmid DNA and foreign DNA are digested with *EcoR1*, mixed and treated with DNA ligase. The ligated molecules are hen used to transform a suitable recipient to tetracycline resistance.

Topic 44

Use of PSC101 for cloning

Chang and Cohen (1974) considered *S. aureus* plasmid p1258 as being particularly appropriate for experiments involving interspecies genome construction. It carries several different genetic determinants that were potentially detectable in *E. coli*. Moreover, Agarose gel electrophoresis indicated that this plasmid is cleaved by the *EcoR1* restriction endonuclease into four easily identifiable fragments. Molecular chimeras containing DNA derived from both *Staphylococcus* and *E. coli* were constructed by ligation of a mixture of *EcoR1* cleaved pSC101 and p1258 DNA. After that, chimeric DNA were used to transform a restrictionless strain of *E. coli*. *E. coli* transformants the expressed the ampicillin resistance determinant carried by the *Staphylococcus* plasmid were selected and screened for tetracycline resistance.

Topic 45

PBR322, a purpose built cloning vehicle

pBR322 is an example of *in vitro* constructed cloning vehicle. It contains the Ap^R and Tc^R genes of RSF2124 and pSC101 respectively, combined with replication element of pMB1. Plasmid pBR322 has been completely sequenced. The most useful aspect of the DNA sequence is that it totally characterizes pBR322 in terms of its restriction sites, such that the exact length of every fragment can be calculated. These fragments can serve as DNA markers for sizing any other DNA fragment in the range of several base pairs up to the entire length of the plasmid. There are over 40 enzymes with unique cleavage sites on the pBR322 genome. The target sites of 11 of these enzymes lie within the *Tc*R gene, and there are sites for a further two (*Cla*I and *Hin*dIII) within the promoter of that gene. There are unique sites for six enzymes within the *Ap*R gene. Thus, cloning in pBR322 with the aid of any one of those 19 enzymes will result in insertional inactivation of either the ApR or the TcR markers. However, cloning in the other unique sites does not permit the easy selection of recombinants, because neither of the antibiotic resistance determinants is inactivated. Transformation of typical bacterial cells with purified DNA becomes less successful as plasmid size increases. It is difficult to clone DNA segments longer than about 15,000 bp when plasmids are used as the vector.

Topic 46

Cloning with PBR 322

Researchers have developed many different plasmid vectors suitable for cloning by modifying naturally occurring plasmids. The *E. coli* plasmid pBR322 offers a good example of the features

useful in a cloning vector. pBR322 has an origin of replication, ori, a sequence where replication is initiated by cellular enzymes. This sequence is required to propagate the plasmid and maintain it at a level of 10 to 20 copies per cell. The plasmid contains two genes that confer resistance to different antibiotics as mentioned above. This allows the identification of cells that contain the intact plasmid or a recombinant version of the plasmid. Diagrammatic illustration of the use of pBR322 as cloning vector for *E. coli* is gen below.

Topic 47

Improved vectors derived from PBR322

With the passage of time, a variety of derivatives of pBR322 have been constructed to fulfil special purpose cloning needs. Early work on the improvement of pBR322 was centered on the insertion of additional unique restriction sites and selectable markers. For example, pBR325 encodes chloramphenicol resistance in addition to ampicillin and tetracycline resistance. It also has a unique *Eco*RI site in the *Cm*R gene. Initially, each new vector was constructed in a series of steps analogous to those used in the generation of pBR322. Then the construction of improved vectors was simplified by the use of polylinkers or multiple cloning sites (MCS), as exemplified by the pUC vectors. An MCS is a short DNA sequence, 2.8 kb in the case of pUC19, carrying sites for many different restriction endonucleases. An MCS increases the number of potential cloning strategies available by extending the range of enzymes that can be used to generate a restriction fragment suitable for cloning. By combining them within an MCS, the sites are made contiguous, so that any two sites within it can be cleaved simultaneously without excising vector sequences. Another derivative of pBR322 which has been extensively used as a cloning vehicle is pAT153.

Topic 48

Runaway plasmid vectors

One reason for cloning a gene on a multicopy plasmid is to increase greatly its expression and hence, facilitate purification of the protein it encodes. However, some genes cannot be cloned on high copy number vectors because excess gene product is lethal to the cell. The use of low copy number vectors avoids cell killing but this may be self defeating since expression of the cloned gene will be reduced. A solution to this problem is to use runaway plasmid vectors. At 30 ^{o}C, the plasmid vector is present in a moderate number of copies per cell. Above 35 ^{o}C, all control of

plasmid replication is lost and the number of plasmid copies per cell increase continuously. Cell growth and protein synthesis continue at the normal rates for 2-3 hours at the higher temperature. During this period, products from genes on the plasmid are overproduced. Eventually, inhibition of cell growth occurs and cells lose viability. At this stage plasmid DNA may account for 50% of the DNA in the cell.

Topic 49

Bacteriophage ʎ as cloning vector

Bacteriophage ʎ has a very efficient mechanism for delivering its 48,502 bp of DNA into a bacterium. It can be used as a vector to clone somewhat larger DNA segments as shown in figure. Two key features contribute to its utility. First, one-third of the ʎ genome is nonessential and can be replaced with foreign DNA. Second, DNA is packaged into infectious phage particles only if it is between 40,000 and 53,000 bp long, a constraint that can be used to ensure packaging of recombinant DNA only. Investigators have developed bacteriophage ʎ vectors that can be readily cleaved into three pieces, two of which contain essential genes but which together are only about 30,000 bp long. The third piece, "filler" DNA, is discarded when the vector is to be used for cloning, and additional DNA is inserted between the two essential segments to generate ligated DNA molecules long enough to produce viable phage particles. In effect, the packaging mechanism selects for recombinant viral DNAs. Bacteriophage ʎ vectors permit the cloning of DNA fragments of up to 23,000 bp. Once the bacteriophage ʎ fragments are ligated to foreign DNA fragments of suitable size, the resulting recombinant DNAs can be packaged into phage particles by adding them to crude bacterial cell extracts that contain all the proteins needed to assemble a complete phage. This is called *in vitro* packaging. All viable phage particles will contain a foreign DNA fragment. The subsequent transmission of the recombinant DNA into *E. coli* cells is highly efficient.

Topic 50

Replication of phage ʎ DNA in lytic and lysogenic cycles

The lytic cycle results in the destruction of the infected bacterial cells. Viral DNA exists as separate molecule within the bacterial cell and replicates independently of the host DNA. Lysogeny is characterized by the integration of the ʎ DNA into the host bacterium genome. In the lytic cycle, λ transcription occurs in three temporal stages: early, middle and late. Basically, early gene transcription establishes the lytic cycle in competition with lysogeny. Middle gene

products replicate and recombine the DNA and late gene products package this DNA into mature phage particles. Following infection of a sensitive host, early transcription proceeds from major promoters situated immediately to the left (PL) and right (PR) of the repressor gene (*cI*). This transcription is subject to repression by the product of the *cI* gene and in a lysogen this repression is the basis of immunity to superinfecting λ. Early in infection, transcripts from PL and PR stop at termination sites tL and tR1. The site tR2 stops any transcripts that escape beyond tR1. Lambda switches from early to middle stage transcription by anti-termination. The *N* gene product, expressed from PL, directs this switch. It interacts with RNA polymerase and, antagonizing the action of host termination protein ρ, permits it to ignore the stop signals so that PL and PR transcripts extend into genes such as *red*, *O* and *P* necessary for the middle stage. The early and middle transcripts and patterns of expression therefore overlap. The *cro* product, when sufficient has accumulated, prevents transcription from PL and PR. The gene *Q* is expressed from the distal portion of the extended PR transcript and is responsible for the middle-to-late switch. This also operates by anti-termination. The *Q* product specifically anti-terminates the short PR transcript, extending it into the late genes, across the cohered *cos* region, so that many mature phage particles are ultimately produced.

Topic 51

Modified phages

Wild-type λ DNA contains several target sites for most of the commonly used restriction endonucleases and so is not itself suitable as a vector. Derivatives of the wild-type phage have therefore been produced that either have a single target site at which foreign DNA can be inserted (insertional vectors) or have a pair of sites defining a fragment that can be removed (stuffer) and replaced by foreign DNA (replacement vectors). Since phage λ can accommodate only about 5% more than its normal complement of DNA, vector derivatives are constructed with deletions to increase the space within the genome. The shortest λ DNA molecules that produce plaques of nearly normal size are 25% deleted. Apparently, if too much non-essential DNA is deleted from the genome, it cannot be packaged into phage particles efficiently. In one set of modified lambda phages, called Charon phages, unwanted restriction enzyme sites have been removed by mutation. Foreign DNA can be inserted into variants that have only a single restriction site, such as Charon 16. By contrast, in variants with two sites, such as Charon 4A,

foreign DNA can replace a specific segment of the lambda DNA. Such replacement vectors are especially useful in cloning large DNA fragments. When Charon 4A is used as a replacement vector, the two small interior fragments are cut out and discarded. Charon 4A and 16 are both derivatives of lambda with various substitutions and deletions in the nonessential region. Each has the *lacZ* gene, encoding the enzyme β-galactosidase, which permits detection of phage containing inserted clones. Whereas the wild-type lambda genome is 48.5 kbp, that for Charon 4A is 45.4 kbp and for Charon 16 is 41.7 kbp. The arrowheads in figure maps of each phage indicate the sites recognized by the restriction enzyme *Eco*RI.

Topic 52

Steps in cloning with ʎ

Cloning with lambda replacement vectors involves the following steps:

i). Isolation of vector DNA from phage particles and cutting it into two fragments with the appropriate restriction enzyme.

ii). Connecting the two lambda fragments to fragments of foreign DNA using DNA ligase. Recombinant molecules are formed of a length suitable for packaging into phage particles.

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

iv). Infecting Escherichia coli cells and isolating phage clones by picking plaques on a host strain.

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

Selection of recombinants is less of a problem with lambda replacement vectors (such as Charon 4A) than with plasmids because (1) the efficiency of transfer of recombinant DNA into the cell by lambda is very high and (2) lambda fragments that have not received new DNA are too small to be incorporated into phage particles. Thus, every viable phage virion should contain cloned DNA. Both Charon vectors are also engineered to contain reporter genes, such as the gene for β-galactosidase. When the vectors replicate in a lactose-negative (Lac-) strain of *E. coli*, β-galactosidase is synthesized from the phage gene, and the presence of lactose-positive (Lac+) plaques can be detected by using a color indicator agar. However, if a foreign gene is inserted

into the β-galactosidase gene, the Lac+ character is lost. Such Lac- plaques can be readily detected as colorless plaques among a background of colored plaques.

Topic 53

Packaging phage-ʎ DNA in vitro

Mainly, we have considered only one way of introducing manipulated phage DNA into the host bacterium, i.e. by transfection of competent bacteria. Using freshly prepared λ DNA that has not been subjected to any gene-manipulation procedures, transfection will result typically in about 10^5 plaques/μg of DNA. In a gene-manipulation experiment in which the vector DNA is restricted and then ligated with foreign DNA, this figure is reduced to about 10^4–10^3 plaques/μg of vector DNA. Even with perfectly efficient nucleic acid biochemistry, some of this reduction is inevitable. It is a consequence of the random association of fragments in the ligation reaction, which produces molecules with a variety of fragment combinations, many of which are inviable. Yet, in some contexts, 10^6 or more recombinants are required. The scale of such experiments can be kept within a reasonable limit by packaging the recombinant DNA into mature phage particles *in vitro*. Placing the recombinant DNA in a phage coat allows it to be introduced into the host bacteria by the normal processes of phage infection, i.e. phage adsorption followed by DNA injection. Depending upon the details of the experimental design, packaging *in vitro* yields about 10^6 plaques/μg of vector DNA after the ligation reaction.

Topic 54

Vectors for DNA sequencing: bacteriophage M13

M13 is a filamentous bacteriophage with single-stranded DNA that replicates without killing its host. Mature particles of M13 are released from host cells without lysing by a budding process, and infected cultures can provide continuous sources of phage DNA. Most of the genome of phage M13 contains genes essential for virus replication. However, a small region called the intergenic sequence can be used as a cloning site. Variable lengths of foreign DNA, up to about 5 kbp, can be cloned without affecting phage viability. As the genome gets larger, the virion simply grows longer. Phage M13mp18 is a derivative of M13 in which the intergenic region has been modified to facilitate cloning as shown in figure. One useful modification is the insertion of a functional fragment of lacZ, the *Escherichia coli* gene that encodes β-galactosidase. Cells

infected with M13mp18 can be detected easily by their color on indicator plates. This is achieved using the artificial substrate, Xgal, which is cleaved by β-galactosidase to yield a blue color. The lacZ gene has itself been modified to contain a 54-bp polylinker, which contains several restriction enzyme cut sites absent from the original M13 genome. The polylinker is inserted into the beginning of the coding portion of the lacZ gene, but does not affect enzyme activity. However, inserting additional cloned DNA into the polylinker inactivates the gene. Phages with such DNA inserts yield colorless plaques (no β-galactosidase activity), and can therefore be easily identified.

Topic 55

Cosmid vectors

Plasmids have been constructed which contain a fragment of ʎ DNA including '*cos'* site. These plasmids have been termed cosmids and can be used as gene-cloning vectors in conjunction with the *in vitro* packaging system. Cosmids are constructed from plasmids by ligating the lambda *cos* region to the plasmid DNA. Foreign DNA is then ligated into the vector. The modified plasmid, plus cloned DNA, can then be packaged into lambda virions *in vitro* as described previously and the phage particles used to transduce *E. coli*. One major advantage of cosmids is that they can be used to clone large fragments of DNA, with inserts as large as 50 kbp accepted by the system. With big inserts, fewer clones are needed to cover a whole genome. Using cosmids also avoids the necessity of having to transform *E. coli*, which is especially inefficient with larger plasmids. Cosmids also permit storage of the DNA in phage particles instead of as plasmids. Phage particles are more stable than plasmids, so the recombinant DNA can be kept for long periods of time.

Topic 56

Modified scheme for cloning in cosmid vectors

Cosmids provide an efficient means of cloning large pieces of foreign DNA. Because of their capacity for large fragments of DNA, cosmids are particularly attractive vectors for constructing libraries of eukaryotic genome fragments. Partial digestion with a restriction endonuclease provides suitably large fragments. However, there is a potential problem associated with the use of partial digests in this way. This is due to the possibility of two or more genome fragments

joining together in the ligation reaction, hence creating a clone containing fragments that were not initially adjacent in the genome. This would give an incorrect picture of their chromosomal organization. The problem can be overcome by size fractionation of the partial digest. Even with sized foreign DNA, in practice cosmid clones may be produced that contain non-contiguous DNA fragments ligated to form a single insert. Such difficulties have been overcome in a cosmid cloning procedure devised by Ish-Horowicz and Burke (1981). By appropriate treatment of the cosmid vector pJB8, left-hand and right-hand vector ends are purified which are incapable of self ligation but which accept dephosphorylated foreign DNA. Thus the method eliminates the need to size the foreign DNA fragments and prevents formation of clones containing short foreign DNA or multiple vector sequences. An alternative solution to these problems has been devised by Bates and Swift (1983) who have constructed cosmid c2XB. This cosmid carries a *Bam*HI insertion site and two *cos* sites separated by a blunt-end restriction site. The creation of these blunt ends, which ligate only very inefficiently under the conditions used, effectively prevents vector self-ligation in the ligation reaction.

Topic 57

Phasmid vectors

Phasmid vectors are combination of plasmid and ʎ phage sequences. It consists of a plasmid vector carrying a ʎ attachment (*ʎatt*) site. They contain functional origins of replication of the plasmids and of ʎ and may be propagated as a plasmid or as a phage in appropriate *E. coli* strains. Phage particles are easy to store, they have an effectively infinite shelf life, and screening phage plaques by molecular hybridization often gives cleaner results than screening bacterial colonies. Alternatively, a phasmid may be used as a phage cloning vector from which subsequently a recombinant plasmid may be released. A highly developed and novel phasmid vector, ʎ ZAP with components of ʎ , M13, T3 and T7 phages.

Topic 58

Bacterial artificial chromosomes (BACs)

Bacterial artificial chromosomes are simply plasmids designed for the cloning of very long segments that may ranges from 100,000 to 300,000 bp of DNA. They generally include selectable markers such as resistance to the antibiotic chloramphenicol (Cm^R), as well as a very

stable origin of replication (ori) that maintains the plasmid at one or two copies per cell. DNA fragments of several hundred thousand base pairs are cloned into the BAC vector. The large circular DNAs are then introduced into host bacteria by electroporation. These procedures use host bacteria with mutations that compromise the structure of their cell wall, permitting the uptake of the large DNA molecules.

Topic 59

Yeast artificial chromosomes (YACs)

E. coli cells are by no means the only hosts for genetic engineering. Yeasts are particularly convenient eukaryotic organisms for this work. As with *E. coli*, yeast genetics is a well-developed discipline. The genome of the most commonly used yeast, *Saccharomyces cerevisiae*, contains only 14×10^6 bp (a simple genome by eukaryotic standards, less than four times the size of the *E. coli* chromosome), and its entire sequence is known. Yeast is also very easy to maintain and grow on a large scale in the laboratory. Plasmid vectors have been constructed for yeast, employing the same principles that govern the use of *E. coli* vectors described above. Convenient methods are now available for moving DNA into and out of yeast cells, facilitating the study of many aspects of eukaryotic cell biochemistry. Some recombinant plasmids incorporate multiple replication origins and other elements that allow them to be used in more than one species (for example, yeast or *E. coli*). Plasmids that can be propagated in cells of two or more different species are called shuttle vectors.

Topic 60

Shuttle and expression vectors

Vectors that can replicate and are stably maintained in two or more unrelated host organisms are called shuttle vectors. Genes carried by a shuttle vector can thus be moved between unrelated organisms. Shuttle vectors have been developed that replicate in both *Escherichia coli* and *Bacillus subtilis*, *E. coli* and yeast, and *E. coli* and mammalian cells, as well as in many other pairs of organisms. The importance of a shuttle vector is that DNA cloned in one organism can be replicated in a second host without modifying the vector in any way to do so. Many shuttle vectors have been designed to move genes between *E. coli* and yeast. Bacterial plasmid vectors

were the starting point and were modified to function in yeast as well. Because bacterial origins of replication do not function in eukaryotes, it is necessary to provide a yeast replication origin.

Organisms have complex regulatory systems, and cloned genes are often expressed poorly or not at all in a foreign host cell. This problem is tackled by using expression vectors that are designed to allow the experimenter to control the expression of cloned genes. Generally, the objective is to obtain high levels of expression, especially in biotechnological applications. However, when dealing with potentially toxic gene products, a low but strictly controlled level may be appropriate. Expression vectors contain regulatory sequences that allow manipulation of gene expression. Usually the control is transcriptional because for high levels of expression it is essential to produce high levels of mRNA. In practice, high levels of transcription require strong promoters that bind RNA polymerase efficiently. However, the native promoter of a cloned gene may work poorly in the new host. For example, promoters from eukaryotes or even from other bacteria function poorly or not at all in *E. coli*. Indeed, even some *E. coli* promoters function at low levels in *E. coli* because their sequences match the promoter consensus poorly and bind RNA polymerase inefficiently. For this reason, expression vectors must contain a promoter that functions efficiently in the host and one that is correctly positioned to drive transcription of the cloned gene. Promoters from E. coli that are used in expression vectors include lac (the lac operon promoter), trp (the trp operon promoter), tac and trc (synthetic hybrids of the trp and lac promoters), and lambda PL (the leftward lambda promoter. These are all strong promoters in *E. coli* and in addition they can be specifically regulated.

Topic 61

Comparison of different cloning vectors-Summary

From the above discussion, it can be concluded that different vectors can carry variable sizes of the foreign DNA into the host organisms. A detailed comparison of all the vectors type that included plasmids, cosmids, phasmids, BACs and YACs is given in table below".

Topic 62

"CLONING STRATEGIES: Introduction

Previously we discussed different techniques for cutting and joining of DNA molecules and their ligation into a suitable vector system. Any cell-based cloning procedure has four essential parts that included:

i). A method for generating the DNA fragment for cloning

ii). A reaction that inserts that fragment into the chosen cloning vector

iii). A means for introducing that recombinant vector into a host cell wherein it is replicated

iv). A method for selecting recipient cells that have acquired the recombinant.

There are two major strategies for isolating sequences from complex sources such as genomic DNA. The first, a cell-based cloning strategy, is to divide the source DNA into manageable fragments and clone everything. Such a collection of clones, representative of the entire starting population, is known as a library. We must then screen the library to identify our clone of interest, using a procedure that discriminates between the desired clone and all the others. The second strategy is to selectively amplify the target sequence directly from the source DNA using PCR, and then clone this individual fragment.

Topic 63

Genomic DNA libraries

Suppose that we wish to clone a single-copy gene from the human genome. We could simply digest total human DNA with a restriction endonuclease, such as *Eco*RI, insert the fragments into a suitable phage-λ vector and then attempt to isolate the desired clone. How many recombinants would we have to screen in order to isolate the right one? Assuming *Eco*RI gives, on average, fragments of about 4 kb, and given that the size of the human haploid genome is 2.8×10^6 kb, we can see that over 7×10^5 independent recombinants must be prepared and screened in order to have a reasonable chance of including the desired sequence. In other words, we have to obtain a very large number of recombinants, which together contain a complete collection of all of the DNA sequences in the entire human genome, a human *genomic library*.

There are problems with the above approach. For instance, the gene may be cut internally one or more times by *Eco*RI so that it is not obtained as a single fragment. This is likely if the gene is large. Also, it may be desirable to obtain extensive regions flanking the gene or whole gene

clusters. Fragments averaging about 4 kb are likely to be inconveniently short. Alternatively, the gene may be contained on an *Eco*RI fragment that is larger than the vector can accept. In this case the appropriate gene would not be cloned at all. These problems can be overcome by cloning random DNA fragments of a large size (for λ replacement vectors, approximately 20 kb). Since the DNA is randomly fragmented, there will be no systematic exclusion of any sequence.

Topic 64

ʎ EMBL vectors for genomic library construction

For creation of a genomic DNA library using the phage-λ vector EMBL3A, high-molecular weight genomic DNA is partially digested with *Sau*3AI. The fragments are treated with phosphatase to remove their 5′ phosphate groups. The vector is digested with *Bam*HI and *Eco*RI, which cut within the polylinker sites. The tiny *Bam*HI/*Eco*RI polylinker fragments are discarded in the isopropanol precipitation, or alternatively the vector arms may be purified by preparative agarose gel electrophoresis. The vector arms are then ligated with the partially digested genomic DNA. The phosphatase treatment prevents the genomic DNA fragments from ligating together. Non-recombinant vector cannot re-form because the small polylinker fragments have been discarded. The only packageable molecules are recombinant phages. These are obtained as plaques on a P2 lysogen of *sup*+ *E. coli*. The Spi− selection ensures recovery of phage lacking *red* and *gam* genes. A *sup*+ host is necessary because, in this example, the vector carries amber mutations in genes *A* and *B*. These mutations increase biological containment, and can be applied to selection procedures, such as recombinational selection, or tagging DNA with a *sup*+ gene. Ultimately, the foreign DNA can be excised from the vector by virtue of the *Sal*I sites in the polylinker.

Topic 65

Genomic libraries in high capacity vectors

In place of phage-λ derivatives, a number of higher capacity cloning vectors such as cosmids, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs) and yeast artificial chromosomes (YACs) are available for the construction of genomic libraries. The advantage of such vectors is that the average insert size is much larger than for λ replacement vectors. Thus, the number of recombinants that need to be screened to identify a particular gene

of interest is correspondingly lower, large genes are more likely to be contained within a single clone and fewer steps are needed for a chromosome walk. Generally, to construct genomic libraries, the partial restriction digest conditions are optimized for larger fragment sizes, and size fractionation must be performed by specialized electrophoresis methods that can separate fragments over 30 kb in length. Pulsed field gel electrophoresis (PFGE) and field-inversion gel electrophoresis (FIGE) have been devised for this purpose. High molecular weight donor DNA fragments can also be prepared using restriction enzymes that cut very rarely.

Topic 66

PCR as an alternative to genomic DNA cloning

PCR with specific primers could be used to isolate genes directly from genomic DNA, obviating the need for the production of genomic libraries. However, a serious limitation is that standard PCR conditions are suitable only for the amplification of short products. The maximum product size that can be obtained is about 5 kb, although the typical size is more likely to be 1–2 kb. This reflects the poor processivity of PCR enzymes such as *Taq* polymerase, and their lack of proofreading activity. Both of these deficiencies increase the likelihood of the enzyme detaching from the template; especially, if the template is long. The extreme reaction conditions required for the PCR are also thought to cause damage to bases and generate nicks in DNA strands, which increase the probability of premature termination on long templates.

Topic 67

Properties of cDNA

cDNA is prepared by reverse-transcribing cellular RNA. Cloned eukaryotic cDNAs have their own special uses, which derive from the fact that they lack introns and other non-coding sequences present in the corresponding genomic DNA. Introns are rare in bacteria but occur in most genes of higher eukaryotes. They can be situated within the coding sequence itself, where they then interrupt the collinear relationship of the gene and its encoded polypeptide, or they may occur in the 5′ or 3′ untranslated regions. In any event, they are copied by RNA polymerase when the gene is transcribed. The primary transcript goes through a series of processing events in the nucleus before appearing in the cytoplasm as mature mRNA. These events include the removal of intron sequences by a process called splicing. In mammals, some genes contain

numerous large introns that represent the vast majority of the sequence. For example, the human dystrophin gene contains 79 introns, representing over 99% of the sequence. The gene is nearly 2.5 Mb in length and yet the corresponding cDNA is only just over 11 kb. Thus, one advantage of cDNA cloning is that in many cases the size of the cDNA clone is significantly lower than that of the corresponding genomic clone.

Topic 68

cDNA libraries

cDNA library is prepared by reverse-transcribing a population of mRNAs and then screened for particular clones. An important concept is that the cDNA library is representative of the RNA population from which it was derived. Most early cDNA libraries were constructed using plasmid vectors, and were difficult to store and maintain for long periods. They were largely replaced by phage λ libraries, which can be stored indefinitely and can also be prepared to much higher titres. λgt10 and λgt11 were the standard vectors for cDNA cloning until about 1990. Both λgt10 and λgt11 are insertion vectors, and they can accept approximately 7.6 kb and 7.2 kb of foreign DNA, respectively. In each case, the foreign DNA is introduced at a unique *Eco*RI cloning site. λgt10 is used to make libraries that are screened by hybridization. The *Eco*RI site interrupts the phage *cI* gene, allowing selection on the basis of plaque morphology. λgt11 contains an *E. coli lacZ* gene driven by the *lac* promoter. If inserted in the correct orientation and reading frame, cDNA sequences cloned in this vector can be expressed as β-galactosidase fusion proteins, and can be detected by immunological screening.

Topic 69

Preparation of cDNA for library construction

The synthesis of double-stranded cDNA suitable for insertion into a cloning vector involves three major steps:

i). First strand DNA synthesis on the mRNA template, carried out with a reverse transcriptase

ii). Removal of the RNA template

iii). Second strand DNA synthesis using the first DNA strand as a template, carried out with a DNA-dependent DNA polymerase, such as *E. coli* DNA polymerase I. All DNA polymerases, whether they use RNA or DNA as the template, require a primer to initiate strand synthesis

Topic 70

Improved methods for cDNA cloning

For cDNA expression libraries, it is advantageous if the cDNA can be inserted into the vector in the correct orientation. This can be achieved by self-priming method by adding linker molecule to double stranded cDNA. A serious disadvantage of the hairpin method is that cleavage with S1 nuclease results in the loss of a certain amount of sequence at the 5′ end of the clone. This strategy has therefore been superseded by other methods in which the second strand is primed in a separate reaction. One of the simplest strategies was devised by Land and co-workers in 1981. After first strand synthesis, which is primed with an oligo-dT primer as usual, the cDNA is tailed with a string of cytidine residues using the enzyme terminal transferase. This artificial oligo-dC tail is then used as an annealing site for a synthetic oligo-dG primer, allowing synthesis of the second strand.

Topic 71

PCR as an alternative for cDNA cloning

Reverse transcription followed by the PCR (RT-PCR) leads to the amplification of RNA sequences in cDNA form. No modification to the basic PCR strategy is required, except that the template for PCR amplification is generated in the same reaction tube in a prior reverse-transcription reaction. Using gene-specific primers, RT-PCR is a sensitive means for detecting, quantifying and cloning specific cDNA molecules. Reverse transcription is carried out using a specific 3′ primer that generates the first cDNA strand, and then PCR amplification is initiated following the addition of a 5′ primer to the reaction mix. Due to the speed with which RT-PCR can be carried out, it is an attractive approach for obtaining a specific cDNA sequence for cloning.

Topic 72

SCREENING STRATEGIES: Selection for the presence of vector

The identification of specific clone from the DNA library can be carried out by genetic methods, sequence dependent screening or screening expression libraries. The former applies to any type of library, genomic or cDNA, and can involve either nucleic acid hybridization or the PCR. In

each case, the design of the probe or primers can be used to home in on one specific clone or a group of structurally related clones. Screening the product of a clone applies only to expression libraries, i.e. libraries where the DNA fragment is expressed to yield a protein. In this case, the clone can be identified because its product is recognized by an antibody or a ligand of some nature. Genetic selection is a very powerful tool since it can be applied to large populations. All useful vector molecules carry a selectable genetic marker or property. Plasmid and cosmid vectors carry drug resistance or nutritional markers and in the case of phage vectors plaque formation is itself the selected property. Genetic selection for the presence of the vector is a prerequisite stage in obtaining the recombinant population.

Topic 73

Screening by hybridization

Nucleic acid hybridization is the most commonly used method of library screening because it is rapid, it can be applied to very large numbers of clones. Grunstein and Hogness (1975) developed a screening procedure to detect DNA sequences in transformed colonies by hybridization *in situ* with radioactive RNA probes. Their procedure can rapidly determine which colony among thousands contains the target sequence. The colonies to be screened are first replica plated on to 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 bearing the colonies is removed and treated with alkali so that the bacterial colonies are lysed and the DNA they contain is denatured. The filter is then treated with proteinase 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 the filter at 80°C. The labelled RNA is hybridized to this DNA and the result of this hybridization is monitored by autoradiography. A colony whose DNA print gives a positive autoradiographic result can then be picked from the reference plate.

Topic 74

Benton and Davis’ plaque lift procedure

Benton and Davis (1977) devised a method called plaque lift, in which the nitrocellulose filter is

applied to the upper surface of agar plates, making direct contact between plaques and filter. The plaques contain phage particles, as well as a considerable amount of unpackaged recombinant DNA. Both phage and unpackaged DNA bind to the filter and can be denatured, fixed and hybridized. This method has the advantage that several identical DNA prints can easily be made from a single-phage plate. This allows the screening to be performed in duplicate, and hence with increased reliability, and also allows a single set of recombinants to be screened with two or more probes. This procedure is probably the most widely applied method of library screening, successfully applied in thousands of laboratories to the isolation of recombinant phage by nucleic acid hybridization.

Topic 75

Probe design

A great advantage of hybridization for library screening is that it is extremely versatile. Conditions can be used in which hybridization is very stringent, so that only sequences identical to the probe are identified. This is necessary, for example, to identify genomic clones corresponding to a specific cDNA or to identify overlapping clones in a chromosome walk. Alternatively, less stringent conditions can be used to identify both identical and related sequences. This is appropriate where a probe from one species is being used to isolate a homologous clone from another species. Probes corresponding to a conserved functional domain of a gene may also cross hybridize with several different clones in the same species at lower stringency. This can be exploited to identify members of a gene family. The identification of the vertebrate *Hox* genes provides an example in which cross-species hybridization was used to identify a family of related clones. In this case a DNA sequence was identified that was conserved between the *Drosophila* developmental genes *fushi tarazu* and *Antennapedia*.

Topic 76

Chromosome walking

One of the advantages of making genomic libraries from random DNA fragments is that the resulting fragments will overlap. It allows genes to be cloned by chromosome walking. The principle of chromosome walking is that overlapping clones will hybridize to each other, allowing them to be assembled into a contiguous sequence. This can be used to isolate genes

whose function is unknown but whose genetic location is known, a technique known as positional cloning. To begin a chromosome walk, it is necessary to have in hand a genomic clone that is known to lie very close to the suspected location of the target gene. In humans, for example, this could be a restriction fragment length polymorphism that has been genetically mapped to the same region. This clone is then used to screen a genomic library by hybridization, which should reveal any overlapping clones. These overlapping clones are then isolated, labelled and used in a second round of screening to identify further overlapping clones. The process is repeated to build up a contiguous map.

Topic 77

Chromosome jumping

A different strategy called chromosome jumping has been used for human DNA. This involves the circularization of very large genomic fragments generated by digestion with endonucleases, such as *Not*I, which cut at very rare target sites. This is followed by sub-cloning of the region covering the closure of the fragment, thus bringing together sequences that were located a considerable distance apart. In this way a jumping library is constructed, which can be used for long-distance chromosome walks.

Topic 78

Screening by PCR

PCR is widely used to isolate specific DNA sequences from uncloned genomic DNA or cDNA. In addition to that, it is also a useful technique for library screening. As a screening method, PCR has the same versatility as hybridization, and the same limitations. It is possible to identify any clone by PCR but only if there is sufficient information about its sequence to make suitable primers. To isolate a specific clone, PCR is carried out with gene-specific primers that flank a unique sequence in the target. There are also several applications where the use of degenerate primers is favorable. A degenerate primer is a mixture of primers, all of similar sequence but with variations at one or more positions. This is analogous to the use of degenerate oligonucleotides as hybridization probes, and the primers are synthesized in the same way. A common circumstance requiring the use of degenerate primers is when the primer sequences have to be deduced from amino acid sequences.

Topic 79

Expression cloning

If a DNA library is established using expression vectors, each individual clone can be expressed to yield a polypeptide. Expression libraries are useful because they allow a range of alternative techniques to be employed, each of which exploits some structural or functional property of the gene product. This can be important in cases where the DNA sequence of the target clone is completely unknown and there is no strategy available to design a suitable probe or set of primers. For higher eukaryotes, all expression libraries are cDNA libraries, since these lack introns and the clones are in most cases of a reasonable size. It should be noted that bacterial expression libraries and many yeast expression libraries are usually genomic, since there are few introns in bacteria. Efficient expression libraries can be generated by cloning randomly sheared genomic DNA or partially digested DNA, and therefore all genes are represented at the same frequency.

Topic 80

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 most versatile expression cloning strategies, because it can be applied to any protein for which an antibody is available. Additionally, there is also no need for that protein to be functional. The molecular target for recognition is generally an epitope, a short sequence of amino acids that folds into a particular three-dimensional conformation on the surface of the protein. Epitopes can fold independently of the rest of the protein and therefore often form even when the polypeptide chain is incomplete or when expressed as a fusion with another protein.

Topic 81

Immunochemical screening of λgt11

Plasmid libraries can be useful for expression screening. However, it is much more convenient to use bacteriophage-λ insertion vectors, because these have a higher capacity and the efficiency of *in vitro* packaging allows large numbers of recombinants to be prepared and screened.

Immunological screening with phage-λ cDNA libraries were introduced by Young and Davies (1983) using the expression vector λgt11. It generates fusion proteins with β-galactosidase under the control of the *lac* promoter. In the original technique, screening was carried out using colonies of induced lysogenic bacteria, which required the production of replica plates as mentioned earlier. A simplification of the method is possible by directly screening plaques of recombinant phage. In this procedure, the library is plated out at moderately high density (up to 5×10^4 plaques/9 cm^2 plate), with *E. coli* strain Y1090 as the host. This *E. coli* strain overproduces the *lac* repressor and ensures that no expression of cloned takes place until the inducer isopropyl-β-D-thiogalactoside (IPTG) is presented to the infected cells. Y1090 is also deficient in the *lon* protease, hence increasing the stability of recombinant fusion proteins. Fusion proteins expressed in plaques are absorbed on to a nitrocellulose membrane overlay, and this membrane is processed for antibody screening. When a positive signal is identified on the membrane, the positive plaque can be picked from the original agar plate and the recombinant phage can be isolated.

Topic 82

South-Western and North-Western blotting

As mentioned above, fusion proteins expressed in plaques produced by recombinant λgt11 or λZAP vectors may be detected by immunochemical screening. A closely related approach has been used for the screening and isolation of clones expressing sequence specific DNA-binding proteins. As above, a plaque lift is carried out to transfer a print of the library on to nitrocellulose membranes. However, the screening is carried out, without using an antibody, by incubating the membranes with a radiolabelled double stranded DNA oligonucleotide probe, containing the recognition sequence for the target DNA-binding protein. This technique is called south-western screening, because it combines the principles of Southern and western blots. It has been particularly successful in the isolation of clones expressing cDNA sequences corresponding to certain mammalian transcription factors

Topic 83

Screening by functional complementation

Functional complementation is the process by which a particular DNA sequence compensates for a missing function in a mutant cell, and thus restores the wild-type phenotype. This can be a very powerful method of expression cloning, because, if the mutant cells are non-viable under particular growth conditions, cells carrying the clone of interest can be positively selected, allowing the corresponding clones to be isolated. Ratzkin and Carbon (1977) provide an early example of how certain eukaryotic genes can be cloned on the basis of their ability to complement auxotrophic mutations in *E. coli*. These investigators inserted fragments of yeast DNA, obtained by mechanical shearing, into the plasmid ColEl, using a homopolymer-tailing procedure. They transformed *E. coli his*B mutants, which are unable to synthesize histidine, with the recombinant plasmids and plated the bacteria on minimal medium. In this way, they selected for complementation of the mutation and isolated clones carrying expressed yeast *his* gene.

Topic 84

Requirement for expression in *E. coli*

Synthesis of functional protein depends upon transcription of appropriate gene, efficient translation of mRNA and in many cases, post-translational processing and compartmentalization of nascent polypeptide. A failure to perform correctly any one of these processes can result in the failure of a given gene to be expressed. Transcription of the cloned insert requires the presence of a promoter recognized by the host RNA polymerase and ideally, a transcription terminator at the 3' end of the gene. Efficient translation requires that mRNA bears a ribosome binding site.

Topic 85

Secretion of proteins

Gram-negative bacteria such as *E. coli* have a complex wall-membrane structure comprising an inner, cytoplasmic membrane separated from an outer membrane by a cell wall and periplasmic space. Secreted proteins may be released into the periplasm or integrated into or transported across the outer membrane. In E. coli, it has been established that protein export through the inner membrane to the periplasm or to the outer membrane is achieved by a universal mechanism known as the general export pathway (GEP). This involves the *sec* gene products. Proteins which

enter the GEP are synthesized in the cytoplasm with a signal sequence at the N-terminus that is cleaved by signal peptidase during transport.

Topic 86

Protein trafficking

The bacterial inner membrane, periplasmic space, and outer membrane all contain proteins not found in the cytoplasm. There are other proteins exported completely out of the cell into the surrounding medium or into eukaryotic host cells. All of these proteins are first synthesized in the cytoplasm but somehow find their way out. How does this happen? Proteins destined to be integral membrane proteins are tagged with a very hydrophobic N-terminal signal sequence that anchors the protein to the membrane. N-terminal signal sequences range from 15 to 30 amino acids and include 11 hydrophobic amino acids preceded by a short stretch of hydrophilic residues. In contrast to integral membrane proteins, proteins predestined for the periplasm or outer membrane have their signal sequences cleaved following export. The roles of the signal sequence are to mediate binding of nascent polypeptides to the membrane and confer a conformation on the precursor that renders it soluble in the membrane. Signal sequences alone, however, are not enough to export proteins. There are actually seven general mechanisms of protein export that manage the remarkable feat of inserting proteins into membranes and passing hydrophilic proteins through hydrophobic membrane barriers. It is important to note that each system traffics different sets of exported proteins.

Topic 87

Stability of foreign proteins in *E. coli*

Various strategies have been developed to cope with the instability of foreign proteins in *E. coli*
In the case of somatostatin, degradation was prevented by producing a fused protein consisting of somatostatin and β-galactosidase. Subsequently, somatostatin can be cleaved with cyanogens bromide. An alternate approach is to use certain mutants of E. coli which have a reduced complement of intracellular proteases. One particularly useful mutant lacks the lon protease, a DNA binding protein with ATP-dependent proteolylitic activity.

Topic 88

Constructing the optimal promoter

A large number of promoters for E. coli have been analyzed and the most recent compilation gives the sequence of 263 of them. Clearly, we would like to use this information to develop the most efficient promoter possible. Comparison of many promoter has led to the formulation of a consensus sequence which consists of the -35 regions (5'-TTGACA) and the -10 regions or Pribnow box (5'-TATAAT), the transcription start point being assigned position +1. In trying to identify the strongest promoter it is essential to have a measure of the relative efficiencies of the different candidates and a suitable system has been devised. Expression from a strong promoter can result in 20-40% of total cell protein being the cloned gene product. During the exponential growth of the host organism, this level of expression puts such an energetic strain on the cell that it leads to the rapid selection of mutant which produces lower amounts of protein or no protein at all. Thus, it is important that only controllable promoters are used. For example, expression from the *trp* promoter can be induced by tryptophan starvation. Similarly, expression of the *lac* promoter can be induced with IPTG.

Topic 89

Optimizing translation initiation

DNA sequence analysis has yielded considerable information about those factors which affect translational initiation. Central to all of them is the Shine-Dalgarno (S-D) sequence or ribosome binding site (rbs). The degree of Complementarity of this sequence with the 16S rRNA can affect the rate of translation. Perfect Complementarity occurs with the sequence 5'-GGAGG-3' and a single bas change within this region can reduce translation tenfold. The composition of the triplet immediately preceding the AUG start codon also affects the efficiency of translation.

Topic 90

Stability of mRNA and codon choice

Provided the various elements of the translational machinery are present in excess, the rate of synthesis of a particular protein will depend on the steady state level of mRNA in the cell. This level is a reflection of the balance between synthesis and degradation. By using strong promoters, the rate of synthesis is maximized. But similar effect could be achieved by reducing

the degradation of messenger. In addition, a strong positive correlation has been found between tRNA abundance and codon choice.

Topic 91

The effect of plasmid copy number

A major rate limiting step in protein synthesis is the binding of ribosomes to mRNA molecules. Since, the number of ribosomes in a cell far exceeds any one class of messenger; one way of increasing the expression of a cloned gene is to increase the number of the corresponding transcripts. Two factors affect the rate of transcription. First, the strength of the promoters as mentioned earlier. Second, the number of gene copies. The easiest way of increasing the gene dosage is to clone the gene of interest on a high copy number plasmid e.g. pBR322.

Topic 92

Plasmid stability

Having maximized the expression of a particular gene it is important to consider what effects this will have on the bacterium harbouring the recombinant plasmid. Increase in the levels of expression of recombinant genes lead to reductions in cell growth rates and may result in morphological changes such as filamentation and increased fragility. If mutant arises which has either lost the recombinant plasmid, or has undergone structural rearrangement so that the recombinant gene is no longer expressed or has a reduced plasmid copy number, then this will have a faster growth rate and may quickly take over and become predominant in the culture.

The loss of plasmids due to defective partioning is called segregative instability. Naturally occurring plasmids are stably maintained because they contain a partioning function, *par*, which ensures that they are accurately segregated at each cell division.

Topic 93

Structural instability

Structure instability of plasmids may arise by deletion, insertion or rearrangements of DNA. Some of the earliest reports of deletions were in chimaeric plasmids which can replicate in both *E. coli* and *B. subtilis*. Spontaneous deletions have now been observed in a wide range of

plasmid, virus and chromosomal DNAs. A common feature of these deletions is the involvement of homologous recombinations between short direct repeats.

Topic 94

Host cell physiology can affect the level of expression

Gene expression can also be controlled by a less tangible factor i.e. the physiology of the host cell. Factors which will be important include the choice of nutrients and the way in which they are supplied to the culture and environmental parameters such as temperature and dissolved oxygen. So far there has been no systematic study of the effect of different growth conditions on the synthesis of foreign proteins in *E. coli*.

Topic 95

DNA sequencing: benefits and applications

DNA sequencing is a fundamental requirement for modern gene manipulation. Knowledge of the sequence of a DNA region may be an end in its own right, perhaps in understanding an inherited human disorder. More importantly, sequence information is a prerequisite for planning any substantial manipulation of the DNA; for example, a computer search of the sequence for all known restriction endonuclease target sites will provide a complete and precise restriction map. Techniques for large scale DNA sequencing became available in the late 1970s. The Maxam and Gilbert technique, which relies on base-specific chemistry, was popular for a time. But chain-termination techniques soon gained popularity. There have been many modifications of the chain terminator principle and automated sequencing has been developed from it.

Topic 96

Maxam-Gilbert method

This method for DNA sequencing makes use of chemical reagents to bring about base-specific cleavage of the DNA. For large scale sequencing, it is now less favored than enzymatic ‘dideoxy’ method. In this method, starting point is a defined DNA restriction fragment. The DNA strand to be sequenced must be radioactively labelled at one end with a 32P-phosphate group. Chemical reagents have been characterized which alter one or two bases in DNA. These are base-specific reactions; for example, dimethylsulphate methylates guanine. An altered base

can then be removed from the sugar phosphate back bone of DNA. The strand is cleaved with piperidine at the sugar residue lacking the base. This cleave is dependent upon the previous step. When each of the base-specific reagents is used in a limited reaction with end-labelled DNA, a nested set of end-labelled fragments of different lengths is generated. These fragments can be separated on gel to elucidate the sequence of a particular fragment.

Topic 97

Chain Termination or dideoxy procedure

The chain-terminator or dideoxy procedure for DNA sequencing capitalizes on two properties of DNA polymerases. First, their ability to synthesize faithfully a complementary copy of a single-stranded DNA template. Second, their ability to use 2′, 3′-dideoxynucleotides as substrates. Once the analogue is incorporated at the growing point of the DNA chain, the 3′ end lacks a hydroxyl group and is no longer a substrate for chain elongation. Thus, the growing DNA chain is terminated, i.e. dideoxynucleotides act as chain terminators. In practice, the Klenow fragment of DNA polymerase is used because this lacks the 5′ → 3′ exonuclease activity associated with the intact enzyme. Initiation of DNA synthesis requires a primer and this is usually a chemically synthesized oligonucleotide which is annealed close to the sequence being analysed. DNA synthesis is carried out in the presence of the four deoxynucleoside triphosphates, one or more of which is labelled with 32P, and in four separate incubation mixes containing a low concentration of one each of the four dideoxynucleoside triphosphate analogues. Therefore, in each reaction there is a population of partially synthesized radioactive DNA molecules, each having a common 5′ end, but each varying in length to a base-specific 3′ end. After a suitable incubation period, the DNA in each mixture is denatured and electrophoresed in a sequencing gel.

Topic 98

Modifications of chain termination sequencing

The sharpness of the autoradiographic images can be improved by replacing the ^{32}P-radiolabel with the much lower-energy ^{33}P or ^{35}S. In the case of ^{35}S, this is achieved by including an α-^{35}S-deoxynucleoside triphosphate in the sequencing reaction. This modified nucleotide is accepted by DNA polymerase and incorporated into the growing DNA chain. Non-isotopic detection methods have also been developed with chemiluminescent, chromogenic or fluorogenic reporter

systems. Although the sensitivity of these methods is not as great as with radiolabels, it is adequate for many purposes. Other technical improvements to Sanger's original method have been made by replacing the Klenow fragment of *Escherichia coli* DNA polymerase I. Natural or modified forms of the phage T7 DNA polymerase ('Sequenase') have found favour, as has the DNA polymerase of the thermophilic bacterium *Thermus aquaticus* (Taq DNA polymerase). The T7 DNA polymerase is more processive than Klenow polymerase, i.e. it is capable of polymerizing a longer run of nucleotides before releasing them from the template. Also, its incorporation of dideoxynucleotides is less affected by local nucleotide sequences and so the sequencing ladders comprise a series of bands with more even intensities.

Topic 99

Automated DNA sequencing

In manual sequencing, the DNA fragments are radiolabelled in four chain-termination reactions, separated on the sequencing gel in four lanes and detected by autoradiography. This approach is not well suited to automation. To automate the process, it is desirable to acquire sequence data in real time by detecting the DNA bands within the gel during the electrophoretic separation. However, this is not trivial, as there are only about 10^{-15}–10^{-16} moles of DNA per band. The solution to the detection problem is to use fluorescence methods. In practice, the fluorescent tags are attached to the chain-terminating nucleotides. Each of the four dideoxynucleotides carries a spectrally different fluorophore. The tag is incorporated into the DNA molecule by the DNA polymerase and accomplishes two operations in one step: it terminates synthesis and it attaches the fluorophore to the end of the molecule. Alternatively, fluorescent primers can be used with non-labelled dideoxynucleotides. By using four different fluorescent dyes, it is possible to electrophorese all four chain-terminating reactions together in one lane of a sequencing gel. The DNA bands are detected by their fluorescence as they electrophorese past a detector. If the detector is made to scan horizontally across the base of a slab gel, many separate sequences can be scanned, one sequence per lane. Because the different fluorophores affect the mobility of fragments to different extents, sophisticated software is incorporated into the scanning step to ensure that bands are read in the correct order. A simpler method is to use only one fluorophore and to run the different chain-terminating reactions in different lanes.

Topic 100

Sequencing accuracy

As part of a programme to sequence a 96 kb stretch of mouse DNA, Wilson *et al.* (1992) analysed 288 sequences containing part of the vector DNA. By comparing raw sequence data with known vector sequences, it was possible to calculate the error frequency. Sequences that were read beyond 400 bp contained an average of 3.2% error, while those less than 400 bp had 2.8% error. At least one-third of the errors were due to ambiguities in sequence reading. In those sequences longer than 400 bp that were read, most errors occurred late in the sequence and were often present as extra bases in a run of two or more of the same nucleotide. The remainder of the errors were due to secondary structure in the template DNA. However, because the complete sequence was analysed with an average 5.9-fold redundancy and most of it on both strands, the final error frequency is estimated to be less than 0.1%.

Topic 101

DNA sequencing databases

Since the current DNA sequencing technology was developed, a large amount of DNA sequence data has accumulated. These data are maintained in three databases: the National Center for Biotechnology Information in the USA, the DNA Databank of Japan and the European Bioinformatic Institute in the UK. Each of these three groups collects a portion of the total sequence data reported worldwide and all new and updated database entries are exchanged between the groups on a daily basis. In addition, several specialized genome databases exist, including seven for bacterial genomes: four for *E. coli*, two for *B. subtilis* and one at the Institute for Genome Research, an organization responsible for the complete sequencing of a number of genomes.

Topic 102

SITE DIRECTED MUTAGENESIS

Mutagenesis is a process to change the genetic information of an organism. It can occur naturally, or as a result of exposure to mutagens or induced experimentally in the laboratory. In some cases, the goal of protein engineering is to generate a molecule with an improvement in some operating parameter, but it is not known what amino acid changes to make. In this

situation, a random mutagenesis strategy provides a route to the desired protein. However, methods based on gene manipulation differ from traditional mutagenesis in that the mutations are restricted to the gene of interest or a small portion of it. Genetic engineering also provides a number of simple methods of generating chimeric proteins where each domain is derived from a different protein. Three different methods of site-directed mutagenesis have been devised: cassette mutagenesis, primer extension and procedures based on the PCR.

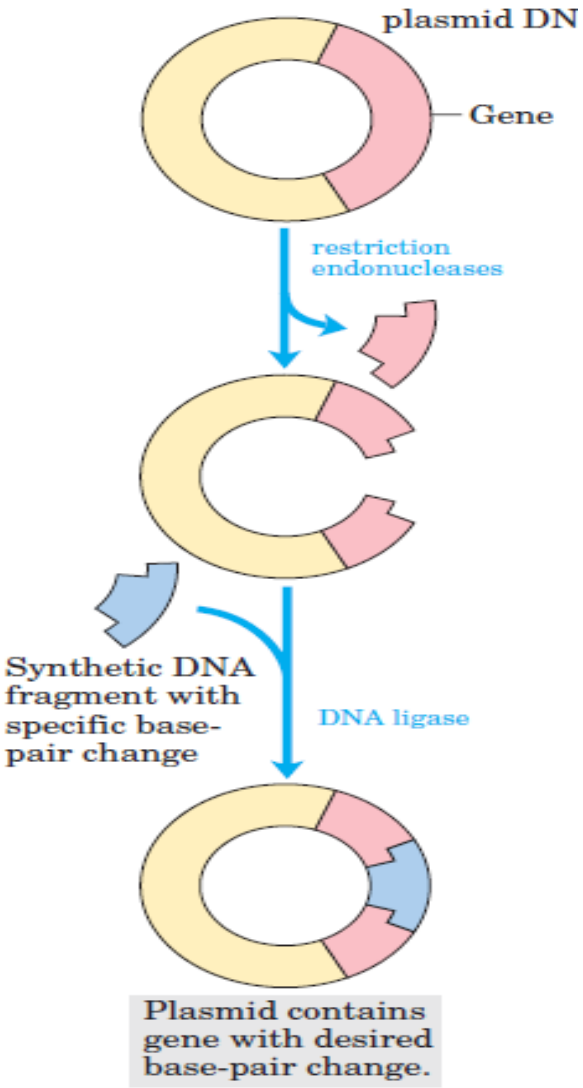

Figure 13. Site directed mutagenesis.

Topic 103

Cassette mutagenesis

In cassette mutagenesis, a synthetic DNA fragment containing the desired mutant sequence is used to replace the corresponding sequence in the wild-type gene. This method was originally used to generate improved variants of the enzyme subtilisin. It is a simple method for which the efficiency of mutagenesis is close to 100%. The disadvantages are the requirement for unique restriction sites flanking the region of interest and the limitation on the realistic number of different oligonucleotide replacements that can be synthesized.

Topic 104

Primer extension: the single primer method

The simplest method of site-directed mutagenesis is the single-primer method. The method involves priming *in vitro* DNA synthesis with a chemically synthesized oligonucleotide (7-20 bp) that carries a base mismatch with the complementary sequence. Method requires that the DNA to be mutated is available in single-stranded form, and cloning the gene in M13-based vectors makes this easy. However, DNA cloned in a plasmid and obtained in duplex form can also be converted to a partially single-stranded molecule that is suitable. The synthetic oligonucleotide primes DNA synthesis and is itself incorporated into the resulting heteroduplex molecule. After transformation of the host *E. coli*, this heteroduplex gives rise to homoduplexes whose sequences are either that of the original wild-type DNA or that containing the mutated base. The frequency with which mutated clones arise, compared with wild-type clones, may be low. In order to pick out mutants, the clones can be screened by nucleic acid hybridization with ^{32}P-labelled oligonucleotide as probe. Under suitable conditions of stringency, i.e. temperature and cation concentration, a positive signal will be obtained only with mutant clones. This allows ready detection of the desired mutant.

Topic 105

PCR method for site directed mutagenesis

Early work on the development of the PCR method of DNA amplification also showed its potential for mutagenesis. Single bases mismatched between the amplification primer and the template become incorporated into the template sequence as a result of amplification. Higuchi *et al.* (1988) have described a variation of the basic method which enables a mutation in a PCR-produced DNA fragment to be introduced anywhere along its length. Two primary PCR

reactions produce two overlapping DNA fragments, both bearing the same mutation in the overlap region. The overlap in sequence allows the fragments to hybridize. One of the two possible hybrids is extended by DNA polymerase to produce a duplex fragment. The other hybrid has recessed 5′ ends and, since it is not a substrate for the polymerase, is effectively lost from the reaction mixture. As with conventional primer extension mutagenesis, deletions and insertions can also be created. The method of Higuchi *et al.* (1988) requires four primers and three PCRs (a pair of PCRs to amplify the overlapping segments and a third PCR to fuse the two segments). Sarkar and Sommer (1990) have described a simpler method, which utilizes three oligonucleotide primers to perform two rounds of PCR. In this method, the product of the first PCR is used as a megaprimer for the second PCR. The advantage of a PCR-based mutagenic protocol is that the desired mutation is obtained with 100% efficiency.

Topic 106

POLYMERASE CHAIN REACTION (PCR): Basic PCR reaction

PCR is a technology in molecular biology used to amplify a single or a few copies of a target DNA across several orders of magnitude, generating millions of copies of a particular DNA sequence. PCR was developed by Kary Mullis in 1983. It is now a common and often indispensable technique for research and clinical laboratories. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers i.e. short DNA fragments that contain sequences complementary to the target region along with a DNA polymerase, which the method is named after, are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations.

Topic 107

PCR principles and procedure

PCR amplifies a specific target region of DNA that may range from a few to several kilo base pairs. PCR cycles may consist of several steps of denaturation, annealing and extension. Basic reaction requires several components that included:

i). DNA template that contains the DNA region to amplify

ii). Two *primers* that are complementary to the 3' ends of each of the strand of the DNA target.

iii). Taq polymerase as the DNA polymerase cannot attach to a DNA strand. It is heat resistant enzyme and can withstand denaturation process at high temperature

iv). Deoxynucleoside triphosphates (dNTPs) i.e. the building-blocks from which the DNA polymerase synthesizes a new DNA strand.

v). Buffer solution, providing a suitable chemical environment for optimum activity and stability of the DNA polymerase

vi). Bivalent cations, magnesium or manganese ions and monovalent cation i.e. potassium ions.

Topic 108

DNA polymerases

In the original description of the PCR method, Klenow DNA polymerase was used and because of the heat denaturation step, fresh enzyme had to be added during each cycle. A breakthrough came with the introduction of Taq DNA polymerase from the thermophilic bacterium *Thermus aquaticus*. The Taq DNA polymerase is resistant to high temperatures and so does not need to be replenished during the PCR. Furthermore, by enabling the extension reaction to be performed at higher temperatures, the specificity of the prier annealing is not compromised. As a consequence of employing the heat-resistant enzyme, the PCR could be automated very simply by placing the assembled reaction in a heating block with suitable thermal cycling pregame. The Taq DNA polymerase lacks a 3'-5' proofreading exonuclease activity. This may contribute to errors during PCR amplification due to misincorporation of nucleotides. Partly to overcome this problem, other thermostable DNA polymerases with improved fidelity have been used.

Topic 109

Primers

Primer is a short strand of oligonucleotide that serves as a starting point for DNA synthesis. The specificity of the PCR depends crucially upon the primers. The following factors are important in choosing effective primers.

i). Primers should be 17 to 30 bp in length.

ii). A GC content of about 50% is ideal. With low GC content, it is desirable to choose a long primer so as to avoid a low melting temperature.

iii). Sequences with long runs of a single nucleotide should be avoided.

iv). Primers with significant secondary structure are undesirable.

v). There should be no complementarity between the two primers.

Topic 110

Degenerate primers

A degenerate primer is actually a mixture of primers, all of similar sequence but with variations at one or more positions. A common circumstance requiring the use of degenerate primers occurs when the primer sequences have to be deduced from amino acid sequences. Degenerate primers may also be employed to search for novel members of a known family of genes. When a degenerate primer is designed on the basis of an amino acid sequence, the degeneracy of the genetic code must be considered for the design of degenerate oligonucleotide probes. Selection of amino acids with low codon degeneracy is desirable.

Topic 111

Types of PCR

In recent years, modifications or variants have been developed from the basic PCR method to improve performance and specificity, and to achieve the amplification of other molecules of interest in research such as RNA. Some of these variants are

i). Multiplex PCR which simultaneously amplified several DNA sequences usually exonic sequences.

ii). Nested PCR increases the specificity of the amplified product for a second PCR with new primers that hybridize within the amplified fragment in the first PCR.

iii). Semiquantitative PCR which allows an approximation to the relative amount of nucleic acids present in a sample.

iv). RT-PCR which generates amplification of RNA by synthesis of cDNA that is then amplified by PCR.

v). Real time PCR which performs absolute or relative quantification of nucleic acid copies obtained by PCR.

The principles of each of the above techniques are described further.

Topic 112

Competitor RT-PCR

This is a commonly used variant of the basic PCR known as reverse transcriptase PCR (RT-PCR). It can be useful in determining low levels of gene expression by analysing the PCR product of a cDNA prepared from the mRNA transcript. In theory a single mRNA molecule can be amplified, but this is unlikely to be achieved in practice. The process involves copying the mRNA using reverse transcriptase as in a standard cDNA synthesis. Oligo(dT)-primed synthesis is often used to generate the first strand cDNA. PCR primers are then used as normal, although the first few cycles may be biased in favour of copying the cDNA single stranded product until enough copies of the second strand have been generated to allow exponential amplification from both primers.

One use of RT-PCR is in determining the amount of mRNA in a sample (competitor RT-PCR). A differing but known amount of competitor RNA is added to a series of reactions, and the target and competitor amplified using the same primer pair. If the target and competitor products are of different sizes, they can be separated on a gel and the amount of target estimated by comparing with the amount of competitor product. When the two bands are of equal intensity, the amount of target sequence in the original sample is the same as the amount of competitor added.

Topic 113

Real time quantitative PCR (qPCR)

There are many applications of the PCR where it would be advantageous to be able to quantify the amount of starting material. Theoretically, there is a quantitative relationship between the amounts of starting material i.e. target sequence and the amount of PCR product at any given cycle. In practice, replicate reactions yield different amounts of product, making quantitation unreliable. Initially, ethidium bromide was used to quantify PCR products as they accumulate. Amplification produces increasing amounts of double stranded DNA, which binds ethidium bromide, resulting in an increase in fluorescence. The principal drawback to the use of ethidium bromide is that both specific and non-specific products generate a signal. This can be overcome by the use of probe-based methods for assaying product accumulation. The probes used are oligonucleotides with a reporter fluorescent dye attached at the 5′ end and a quencher dye at the

3′ end. While the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter dye. If the target sequence is present, the probe anneals downstream from one of the primer sites. As the primer is extended, the probe is cleaved by the 5′ nuclease activity of the *Taq* polymerase. This cleavage of the probe separates the reporter dye from the quencher dye, thereby increasing the reporter-dye signal. Cleavage removes the probe from the target strand, allowing primer extension to continue to the end of the template strand. Additional reporter-dye molecules are cleaved from their respective probes with each cycle, effecting an increase in fluorescence intensity proportional to the amount of amplicon produced.

Topic 114

Nested PCR

This PCR increases the sensitivity due to small amounts of the target are detected by using two sets of primers, involving a double process of amplification. The first set of primers allows a first amplification. The product of this PCR is subjected to a second PCR using the second set of primers. These primers used in the second PCR are specific to an internal amplified sequence in the first PCR. Therefore, specificity of the first PCR product is verified with the second one. The disadvantage of this technique is the probability of contamination during transfer from the first amplified product into the tube in which the second amplification will be performed. Contamination can be controlled using primers designed to anneal at different temperatures.

Topic 115

Inverse PCR

Often, a stretch of DNA sequence is known, but the desired target sequence lies outside this region. This causes problems with primer design, as there may be no way of determining a suitable primer sequence for the unknown region. Inverse PCR (IPCR) involves isolating a restriction fragment that contains the known sequence plus flanking sequences. By circularising the fragment, and then cutting inside the known sequence, the fragment is essentially inverted. Primers can then be synthesised using the known sequence data and used to amplify the fragment, which will contain the flanking regions. Primers that face away from each other (with respect to direction of product synthesis) in the original known sequence are required, so that on

circularisation they are in the correct orientation. The technique can also be used with sets of primers for nested PCR.

Topic 116

Multiplex PCR

Multiplex PCR is an adaptation of PCR which allows simultaneous amplification of many sequences. This technique is used for diagnosis of different diseases in the same sample. Multiplex PCR can detect different pathogens in a single sample. Also it can be used to identify exonic and intronic sequences in specific genes and determination of gene dosage. This is achieved when in a single tube include sets of specific primers for different targets. In this PCR is important the design of primers because they must be characterized by adherence to specific DNA sequences at similar temperatures. However, it may require several trials to achieve the standardization of the procedure.

Topic 117

RAPD

The aim of PCR is to generate defined fragments from highly specific primers. However, there are some techniques based on low-stringency annealing of primers. Most widely used of these is RAPD-PCR. This stands for random amplification of polymorphic DNA. The technique is also known as arbitrarily primed PCR (AP-PCR). It is a useful method of genomic fingerprinting and involves using a single primer at low stringency. The primer binds to many sites in the genome, and fragments are amplified from these. The stringency of primer binding can be increased after a few cycles, which effectively means that only the ‘best mismatches’ are fully amplified. By careful design the protocol can yield genome-specific band patterns that can be used for comparative analysis. However, there can be problems associated with the reproducibility of the technique, as it is difficult to obtain similar levels of primer binding in different experiments. This is largely due to the mismatched binding of primers at low stringency that is the basis of the technique.

Topic 118

RFLP

Restriction fragment length polymorphism (RFLP) is a technique that exploits variations in homologous DNA sequences. It refers to a difference between samples of homologous DNA molecules from differing locations of restriction enzyme sites. In RFLP analysis, the DNA sample is broken into pieces by restriction enzymes and the resulting restriction fragments are separated according to their lengths by gel electrophoresis. Although now largely obsolete due to the rise of inexpensive DNA sequencing technologies, RFLP analysis was the first DNA profiling technique inexpensive enough to see widespread application. RFLP analysis was an important tool in genome mapping, localization of genes for genetic disorders, determination of risk for disease and paternity testing.

Topic 119

AFLP-PCR

AFLPs-Amplified fragment length polymorphisms are differences in restriction fragments length caused by SNPs or INDELS. AFLP is a PCR based tool used in genetic engineering. Developed in the early 1990s, AFLP uses restriction enzymes to digest genomic DNA, followed by ligation of adaptors to the sticky ends of the restriction fragments. A subset of the restriction fragments is then selected to be amplified. This selection is achieved by using primers complementary to the adaptor sequence, the restriction site sequence and a few nucleotides inside the restriction site fragments. The amplified fragments are separated and visualized on denaturing polyacrylamide gels, either through autoradiography or fluorescence methodologies.

Topic 120

Applications of PCR

During the past 30 years, molecular techniques have been under development, however these have had a rapid and tremendous progress in recent year. Among molecular techniques, PCR and its different variations are highlighted as the most commonly used in laboratories and research institutes. Thus, these have contributed to identification and characterization of several organisms and understanding of physiopathology of diverse diseases in human, animal and plant. Also these have provided clues for future research directions in specific topics with impact in public health such as genetics and biochemistry of antimicrobial resistance. The following

describes some applications of PCR and its variants in studies in human medicine, forensic sciences, and agricultural science and environment.

Topic 121

PCR-Medicine

Molecular techniques; especially, PCR have a major impact on medicine. The versatility of molecular techniques has allowed advances and changes in all fields of medicine. The following is an overview of the main impacts generated for molecular biology in medical sciences. Clinical microbiology has been transformed with the use of molecular technology because it has generated a benefit to the patient affected by infectious diseases. Molecular biology has allowed the development of clinical microbiology because it has been possible to identify microorganisms that are difficult to culture, that have many requirements of laboratory or dangerous for laboratory personnel. The usefulness of PCR in identification of microorganisms has led to the selection and quality assurance of blood that blood banks are using for patients with different pathologies. The incorporation of molecular techniques has been of great importance in the identification and characterization of many viruses, including influenza, which through a rapid, sensitive, and effective molecular diagnosis has allowed inclusion of early treatment to benefit patients and control of a high impact infection.

Topic 122

PCR-Forensic sciences

In forensic pathology, classic morphology remains as a basic procedure to investigate deaths, but recent advances in molecular biology have provided a very useful tool to research systemic changes involved in the pathophysiological process of death that cannot be detected by morphology. In addition, genetic basis of diseases with sudden death can also be investigated with molecular methods. Practical application of RNA analysis has not been accepted for post-mortem research, due to rapid decomposition after death. However, recent studies using variants of conventional PCR (qPCR and RT-PCR) have suggested that relative quantification of RNA transcripts can be applied in molecular pathology to research deaths (molecular autopsy). In a broad sense, forensic molecular pathology involves application of molecular biology in medical science to investigate the genetic basis of pathophysiology of diseases that lead to death.

Therefore, molecular tools support and reinforce the morphological and physiological evidence in research of unexplained death. Molecular methods are used in forensic science to establish the filiations of a person (paternity testing) or to obtain evidence from minimal samples of saliva, semen or other tissue debris.

Topic 123

PCR-Agricultural sciences and environment

Applications of molecular techniques in research in agricultural sciences and environment have been very numerous and varied. It is possible that one of the most important contributions of the applications of some molecular techniques such as PCR has been the identification and characterization of multiple infectious agents that have great impact on human and animal health. Currently, the genome of most domestic animals and major infectious agents that affect animals is known through the use of molecular tools, facilitating the study of mutations associated with disease. Molecular techniques such as conventional PCR or qPCR have also facilitated research in detection of pathogens in plants, animals, and the environment; understanding of their epidemiology; and, development of new diagnostic tests, treatments or vaccines. Conventional PCR or PCR based methods are being applied to identification and characterization of specific pathogens of animals.

Topic 124

PCR-Molecular Paleontology

Molecular paleontology refers to the recovery and analysis of DNA and protein from ancient human, animal and plant remains. The field of molecular paleontology has yielded important insights into evolutionary events, the discovery and characterization of extinct species. By applying molecular analytical techniques to DNA in fossils, one can quantify the level of relatedness between any two organisms for which DNA has been recovered. Advancements in the field of molecular paleontology have allowed scientists to pursue evolutionary questions on a genetic level rather than relying on phenotypic variation alone. Using various biotechnological techniques such as DNA isolation, amplification, and sequencing scientists have been able to gain expanded new insights into the divergence and evolutionary history of countless organisms.”

Topic 125

Mapping and Sequencing Genome

"Genome maps pinpoint the location of specific features on the chromosomes of an organism. They are essential tools in identifying genes responsible for diseases or traits, for comparing the genomes of different species, and for complete genome sequencing. A genome map, like any other map, defines the relative positions of features that are of interest, or which can serve as reference points for navigation. The features that are located on a genome map are collectively referred to as markers, and can include both genes and noncoding sequences. Many methods exist for making genome maps, differing in the types of markers that can be mapped and in their accuracy. Only one technique (fluorescence in situ hybridization) allows the position of markers in the genome to be observed directly. In all other methods, a less direct approach is necessary. A common theme in these indirect methods is that some means is first used to isolate a portion of the genome from the remainder (for example, by cloning in bacterial cells), and to identify the markers present in that portion. The markers found can then be assumed to be physically linked in the genome (for example, all markers found on a single cloned fragment must lie consecutively). By analysing many such samples of the genome, the order and spacing of all markers can be inferred.

Topic 126

Markers for genome mapping

Almost any identifiable feature of the genome can serve as a marker in mapping (Fig. 14). The markers most commonly used are simply short, unique fragments of known deoxyribonucleic acid (DNA) sequence, which can be amplified and detected in a DNA sample using PCR. Such sequence fragments are collectively called 'sequence tagged sites' or STSs. Often, the STS will be some arbitrarily chosen piece of DNA and has no particular biological significance. However, an STS can originate from a coding sequence (for example, by sequencing a clone from a complementary DNA (cDNA) library), in which case it is an 'expressed sequence tag' or EST. A third category of STS consists of a tandem repetition of a simple, short motif (for example, CACACACA...) flanked by a unique sequence. Such 'microsatellites' can be detected by PCR using their unique flanking sequence but often display length polymorphism from one individual to the next, as the number of repeat elements varies. Such polymorphism is of key importance in genetic linkage mapping. Cloned DNA fragments can be mapped, and therefore act as markers,

even when little or nothing is known of their sequence. Techniques for mapping such cloned fragments include physical mapping and fluorescence in situ hybridization. If the anonymous cloned fragment contains a restriction site whose location differs from one allele to the next (a restriction fragment length polymorphism or RFLP), then genetic linkage mapping can be used.

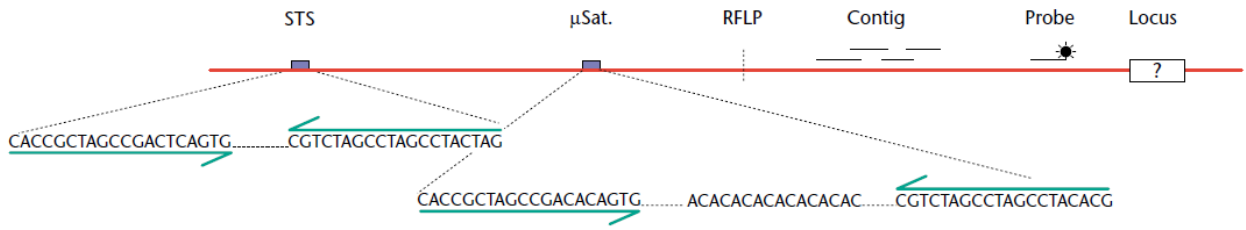

Figure 14. Markers used in genome mapping.

Topic 127

Genetic linkage mapping

Genetic linkage maps illustrate the order of genes or genetic markers on a chromosome and the relative distances between those genes. In organisms that reproduce sexually, meiosis breaks the parental chromosomes at random, recombines the fragments and segregates the shuffled chromosomes into gametes, and thence to offspring. If two markers, A and B, lie close together on a chromosome, then it is unlikely that a meiotic break will occur between them. Hence, A and B will seldom recombine, and will usually segregate together (cosegregate) to the same gametes and hence to the same offspring. In contrast, if markers B and Z lie at opposite ends of a chromosome, it is much more likely that a meiotic break will fall between them, whereupon they may segregate independently. Markers lying on different chromosomes will also segregate independently. Hence, the distance between any two markers is reflected by their recombination frequency: closely linked markers recombine rarely (or cosegregate often) while distant markers recombine often, and hence cosegregate no more than expected by chance – 50% of the time. This provides a way to estimate the distances between pairs of markers and, if all pairwise distances are known, a map can be made.

Topic 128

Physical mapping

Physical mapping involves finding a contiguous series (or ‘contig’) of cloned DNA fragments which contain overlapping portions of the genome. The overlaps define the positions of the clones relative to one another. If at least some of the clones contain markers that have been independently mapped by other means, then the position of the entire contig in the genome is also known. The starting point for physical mapping is a library of cloned genomic fragments, normally prepared by either random mechanical breakage or partial restriction digestion of genomic DNA. The fragments are usually cloned in bacterial hosts (normally *Escherichia coli*), using bacteriophage, cosmid, plasmid or other vector systems. For physical mapping of large genomes, it is desirable to use clones containing large inserts, such as P1 artificial chromosomes (PACs) or bacterial artificial chromosomes (BACs) which can carry inserts in excess of 100 kb. Even larger fragments (over 1Mb) can be cloned in yeast using yeast artificial chromosome (YAC) vectors; however, such clones are often unstable, undergoing deletions or internal rearrangements. Once the library is established, overlapping clones can be identified using several different approaches. In STS content mapping, the library is screened to identify all clones that contain a specific STS marker. Screening can be done either by PCR using the appropriate primers or by arraying the clones on a nylon membrane, lysing them to expose their DNA, and probing the membrane with labelled DNA corresponding to the STS in question. A second common approach to physical mapping is by the use of restriction fingerprinting. In this case, each member of the clone library is digested in turn with one or more restriction enzymes, and the sizes of the resulting fragments are measured by gel electrophoresis.

Topic 129

Physical versus linkage maps

Genetic-linkage maps illustrate the order of genes on a chromosome and the relative distances between those genes. Originally, these maps were made by tracing the inheritance of multiple traits, such as hair color and eye color, through several generations. Genetic-linkage mapping is possible because of a normal biological process called crossing over, which occurs during meiosis-a type of cell division for making sperm and egg cells. During one stage of meiosis, chromosomes line up in pairs along the center of a cell, where they sometimes "stick" to each other and exchange equivalent pieces of themselves. This sticking and exchanging is called crossing over, and is a relatively common event.

Physical maps, by contrast, always give the physical, DNA-base-pair distances from one landmark to another. In the late 1970s, scientists developed new and efficient ways of cutting the genome up into smaller pieces in order to study it. Around the same time they made the first physical maps by using the overlapping DNA sequences at the ends of the genome pieces to help them keep track of where the pieces came from. In other words, a physical map was simply an ordered set of DNA pieces. Today, genome scientists use landmarks known as STSs to help them find their way around the genome. Each STS, or "sequence-tagged site," is a unique DNA sequence-one that is found in only one place in the genome and is a few hundred base pairs long. Some STSs are parts of genes, but an STS can come from anywhere in the genome as long as it is unique.

Topic 130

Use of RLFPs in physical maps

RFLP-Restriction fragment length polymorphism exploits variations in homologous DNA sequences. Botstein et al. (1980) were the first to recognize that DNA probes that target RFLPs can be used for mapping. To generate an RFLP map the probes must be highly informative. This means that the locus must not only be polymorphic, it must be very polymorphic. If enough individuals are studied, any randomly selected probe will eventually discover a polymorphism. However, a polymorphism in which one allele exists in 99.9% of the population and the other in 0.1% is of little utility because it seldom will be informative. Thus, as a general rule, the RFLPs used to construct the genetic map should have two, or perhaps three, alleles with equivalent frequencies.

Topic 131

STS in physical maps

STS is a short region of DNA about 200-300 bases long whose exact sequence is found nowhere else in the genome. Two or more clones containing the same STS must overlap and the overlap must include the STS. Any clone that can be sequenced may be used as an STS provided it contains a unique sequence. A better method to develop STS markers is to create a chromosome-specific library in phage M13. Random M13 clones are selected and 200-400 bases sequenced. The sequence data generated are compared with all known repeated sequences to help identify regions likely to be unique. Two PCR primer sequences are selected from the unique regions which are separated by 100-300 bp and whose melting temperatures are similar. Once identified,

the primers are synthesized and used to PCR amplify genomic DNA from the target organism and the amplification products analyzed by agarose gel electrophoresis. A functional STS marker will amplify a single target region of the genome and produce a single band on an electrophoretic gel at a position corresponding to the size of the target region. Alternatively, an STS marker can be used as a hybridization probe.

Topic 132

SNPs as physical markers

Single nucleotide polymorphisms (SNPs) are single base-pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population. SNPs probably are the most important sequence markers for physical mapping of genomes. Different methods to detect SNPs have been developed that based on enzymatic, electrophoretic, solid phase or chromatographic analysis. In highly outbred populations, such as humans, polymorphisms are considered to be SNPs only if the least abundant allele has a frequency of 1% or more. This is to distinguish SNPs from very rare mutations. In practice, the term SNP is typically used more loosely than required by the above definition. For example, single base variants in cDNAs (cSNPs) are usually classed as SNPs because most of these will reflect underlying genomic DNA variants although they could result from RNA editing. Single base-pair insertions or deletions (indels) also are considered to be SNPs by some workers. A special subset of SNPs is one where the base change alters the sensitivity of a sequence to cleavage by a restriction endonuclease. These are known as restriction fragment length polymorphisms (RFLPs) or "snip-SNPs". SNPs probably are the most important sequence markers for physical mapping of genomes.

Topic 133

Polymorphic DNA detection in the absence of sequence information

Polymorphic DNA can be detected by amplification in the absence of the target DNA sequence information used to generate STSs. One method is based on the amplification of genomic DNA with single primers of arbitrary nucleotide sequence. The nucleotide sequence of each primer was chosen within the constraints that the primer was nine or 10 nucleotides in length, between 50 and 80% G + C in composition and contained no palindromes. Not all the sequences amplified in this way are polymorphic but those that are (randomly amplified polymorphic DNA, RAPD) are easily identified. RAPDs are widely used by plant molecular biologists to construct

maps because they provide very large numbers of markers and are very easy to detect by agarose gel electrophoresis.

Topic 134

AFLPs detection in the absence of sequence information

Amplified fragment length polymorphism (AFLP) is a diagnostic fingerprinting technique that detects genomic restriction fragments and in that respect resembles the RFLP technique. The major difference is that PCR amplification rather than Southern blotting is used for the detection of restriction fragments. The resemblance to the RFLP technique was the basis for choosing the name AFLP. However, the name AFLP should not be used as an acronym because the technique detects presence or absence of restriction fragments and not length differences. The AFLP approach is particularly powerful because it requires no previous sequence characterization of the target genome. For this reason it has been widely adopted by plant geneticists. It also has been used with bacterial and viral genomes. It has not proved useful in mapping animal genomes because it is dependent on the presence of high rates of substitutional variation in the DNA; RFLPs are much more common in plant genomes compared to animal genomes.

Topic 135

Fluorescence in situ hybridization (FISH)

FISH is a cytogenetic method that used fluorescent probes that bind to only those parts of the chromosomes with a high degree of sequence complementarity. The first requirement for conventional FISH is a population of cells in which high proportions are in metaphase, the stage in the cell cycle where the chromosomes are condensed as distinct bodies. This is typically achieved by culturing lymphocytes or fibroblasts in the presence of a mitogen such as phytohaemagglutinin, then treating them with colcemid to block the cell cycle at metaphase. Treatment with bromodeoxyuridine (which is incorporated into the replicating chromosomes) and ethidium bromide (which intercalates into the DNA) improves the morphology of the chromosomes for FISH. The cells are swollen in a hypotonic salt solution and stabilized with a fixative. Drops of the cell suspension are placed on glass slides, to which the chromosomes adhere. Their DNA is denatured, and is then ready to be hybridized to the probe. The resulting images clearly reveal the location of the bound probe, typically as two adjacent spots corresponding to the two chromatids. The characteristic banding pattern of the chromosomes gives the absolute location of the probe in the genome. By using different combinations of

haptens and fluorophores, two or more probes may be hybridized and detected simultaneously. Such multicolour FISH allows the order and spacing of the probes to be determined more accurately than is possible with independent, single-probe experiments.

Topic 136

Radiation hybrid (RH) mapping

RH mapping used X-ray breakage of chromosomes to determine the distances between DNA markers as well as their order on the chromosome. RH mapping was developed originally to facilitate the mapping of the human genome. In this method, a high dose of X-rays is used to break the human chromosome into fragments and these fragments are recovered in rodent cells using somatic cell hybrids. The rodent–human hybrid clones are isolated and examined for the presence or absence of specific human DNA markers. The farther apart two markers are on the chromosome, the more likely a given dose of radiation will break the chromosome between them, placing the markers on two separate chromosomal fragments. By estimating the frequency of breakage, and thus the distance, between markers it is possible to determine their order in a manner analogous to conventional meiotic mapping. Radiation hybrid mapping has a number of advantages over conventional genetic (meiotic) mapping. First, chromosome breakage is random and there are no hot-spots, interference, or gender-specific differences as seen with recombination. Secondly, a much higher resolution can be achieved, e.g. 100-500 kb in radiation mapping as opposed to 1-3 Mb in genetic mapping, and the resolution can be varied by varying the radiation dose. Finally, it is not necessary to use polymorphic markers; monomorphic markers such as STSs can be used as well.

Topic 137

Happing mapping

A number of operational issues have restricted the utility of RH mapping. First, it takes considerable effort to generate an RH panel, even for mammalian genomes, and complications arise because the biological activity of donor fragments in the host cell lines biases their segregation. Second, the presence of a host genome in the hybrids precludes the use of generic markers such as AFLPs and RAPDs. Finally, the technique is not applicable to plants. HAPPY mapping represents an easy and generic alternative to RH mapping and has been used to map the human genome, animal genomes and the *Arabidopsis* genome. HAPPY mapping involves breaking genomic DNA randomly by irradiation or shearing followed by an optional size

fractionation step. Markers then are segregated by diluting the resulting fragments to give aliquots containing one haploid genome equivalent; hence, the term "HAPPY" mapping is used. Markers are detected using the PCR and linked markers tend to be found together in an aliquot.

Topic 138

Map integration

It is essential that the different mapping methods are integrated as each has its own strengths and weaknesses. The size and complexity of the genome being analyzed can greatly influence the methodologies employed in map construction. Nor is it uncommon for different research groups to use different mapping methods even when working on the same genome. Ultimately, the maps generated by the different methods need to be integrated. This is happening. For example, the BAC clones used to construct the human cytogenetic map have STSs that reference the radiation hybrid and linkage maps.

Topic 139

Sequencing genome

Nucleic acid sequencing is a method for determining the exact order of nucleotides present in a given DNA or RNA molecule. In the past decade, the use of nucleic acid sequencing has increased exponentially as the ability to sequence has become accessible to research and clinical labs all over the world. The first major foray into DNA sequencing was the Human Genome Project, a $3 billion, 13-year-long endeavor, completed in 2003. The Human Genome Project was accomplished with first-generation sequencing, known as Sanger sequencing. Sanger sequencing (the chain-termination method), developed in 1975 by Edward Sanger, was considered the gold standard for nucleic acid sequencing for the subsequent two and a half decades. Since completion of the first human genome sequence, demand for cheaper and faster sequencing methods has increased greatly. This demand has driven the development of second-generation sequencing methods, or next generation sequencing (NGS).

Topic 140

High throughput sequencing

Recent advances in DNA sequencing have made it possible to sequence data very rapidly and at a substantially lower cost. High throughput DNA sequencing mainly describes a number of different modern sequencing technologies to sequence genome. Genomes range in size from millions of base pairs to thousands of millions. Given that a single Sanger sequencing reaction

allows 500-600 bases to be sequenced, it is clear that automation is essential. The theoretical sequencing capacity of an automated DNA sequencer is easy to calculate. For a four-dye slab gel system, the capacity is the number of sequencing reactions that can be loaded on each gel, times the number of bases read from each sample, times the number of gels that can be run at once, times the number of days this can be carried out per year. For a 24-channel sequencer the capacity calculated in this way is 2.7 million bases per year. To meet the demands of large-scale sequencing, 96-channel instruments have become commonplace and at least one 384-channel instrument has been developed By switching from slab gels to capillary systems, the electrophoresis run time is greatly reduced and nine runs can be achieved per 24 h period. Various other improvements to the biochemistry of the sequencing reactions and the chemistry of the gel matrix mean that the read length can be extended from the usual 500-600 bases to 800 bases. As a consequence it now is possible to generate 1-6 million bases of sequence per machine per month and Amersham Biosciences claim over 1 million bases in 8 hours or 2.8 million bases in 24 hours for their MegaBACE 4000 instrument.

Topic 141

Shotgun sequencing

Shotgun sequencing is used for sequencing long DNA strands. In shotgun sequencing, DNA is broken up randomly into numerous small segments, which are sequenced using the chain termination method to obtain reads. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

Topic 142

Clone-by-clone sequencing

Sanger sequencing has an accurate read length of 500-800 bases. One approach to sequencing a genome would be a shotgun approach in which a random selection of sequencing reads would be collected from a larger target DNA sequence. With sufficient oversampling (coverage) it should be possible to infer the complete genome sequence by piecing together the individual sequence reads. In practice, two problems will arise. First, the presence of dispersed repeated sequences will confound the sequence assembly. Second, assembly of sequences into contigs will be possible but there will be gaps between contigs that will need closing by other means. The numbers of these problems should increase linearly with the size of the genome to be analyzed.

Consequently, it was assumed that targets the size of cosmids represented the limit of the shotgun approach. Thus, whole genomes would be sequenced by first developing a set of overlapping cosmids that had been ordered by physical mapping. This is the "clone-by-clone" or "map-based" approach and it was used successfully to generate the complete genome sequence of *Saccharomyces cerevisiae* and the nematode *Caenorhabditis elegans*.

Topic 143

Comparative Genomics: Orthologs and paralogs

In order to compare genome organization in different organisms it is necessary to distinguish between orthologs and paralogs. Orthologs are homologous genes in different organisms that encode proteins with the same function and which have evolved by direct vertical descent. Paralogs are homologous genes within an organism encoding proteins with related but non-identical functions. Implicit in these definitions is that orthologs evolve simply by the gradual accumulation of mutations, whereas paralogs arise by gene duplication followed by mutation accumulation.

Topic 144

Comparative genomics of bacteria

By mid-2004, the website of the National Center for Biotechnology Information listed 173 bacteria (19 Archaea and 154 Eubacteria) whose genomes had been sequenced. Simple analysis of the sequence data reveals two features of note. First, the genome sizes vary from 0.49 Mb (*Nanoarchaeum equitans*) to 9.1 Mb (*Bradyrhizobium japonicum* and two species of *Streptomyces*), i.e. a more than 18-fold difference. Secondly, the gene density is remarkably similar across all species and is about 1 gene per kilobase of DNA. This means that large prokaryotic genomes contain many more genes than smaller ones. By contrast, the human genome contains only twice as many genes as *Drosophila*. So how can we account for the size diversity of prokaryotes? When the different genomes are arranged in size order some interesting features emerge. First, the archaebacteria exhibit a very much smaller range of genome sizes. This could be an artifact of the small number of genomes examined but more probably reflects the fact that most of them occupy a specialized environment and have little need for metabolic diversity. The exception is *Methanosarcina acetivorans*. This bacterium is known to thrive in a broad range of environments and at 5.8 Mb has the largest archaeal genome. Second, the smallest eubacterial genomes are found in those organisms that normally are found associated with

animals or humans, e.g. mycoplasmas, rickettsias, chlamydiae, etc. Those organisms that can occupy a greater number of niches have a larger genome size. Not surprisingly, there is a good correlation between genome size and metabolic and functional diversity as demonstrated by the size of the genomes of *Bacillus* and *Streptomyces* (formation of spores, antibiotic synthesis), rhizobia (symbiotic nitrogen fixation), and *Pseudomonas* (degradation of a wide range of aromatic compounds).

Topic 145

Comparative genomics of organelles

The principal function of the mitochondrion is the generation of ATP via oxidative phosphorylation. At least 21 genes encode proteins critical for oxidative phosphorylation and one would expect all of these genes to be located in the mtDNA. Similarly, an mtDNA location would be expected for the genes encoding the 14 ribosomal proteins that are required to translate mtRNA. However, sequence data indicate that many mitochondrial genomes lack a number of key genes and the missing genes can be found in the nucleus. Functional transfer of mitochondrial genes to the nucleus has stopped in animals, hence their consistency in size. Part of the reason for this is that further transfer is blocked by changes in the mitochondrial genetic code. However, this gene transfer continues to occur in plants and protists because there is no genetic code barrier to transfer.

Topic 146

Comparative genomics of Eukaryotes

In determining the minimal genome we are seeking to answer a number of different questions. What is the minimal size of the genome of a free-living unicellular eukaryote and how does it compare with the minimal bacterial genome? That is, what are the fundamental genetic differences between a eukaryotic and a prokaryotic cell? Next, what additional genetic information does it require for multicellular coordination? In animals, what are the minimum sizes for a vertebrate genome and a mammalian genome? Finally, what is the minimum size of genome for a flowering plant? Given that many eukaryotic genomes contain large amounts of noncoding DNA these questions have to be answered by considering both genome size and the number of proteins that are encoded. The smallest eukaryotic genome that has been sequenced is that of the obligate intracellular parasite *Encephalitozoon cuniculi*. This has a genome size of only 2.9 Mb although its close relative *E. intestinalis* may have a genome that is even smaller

(2.3 Mb). Genome compaction in these organisms is achieved by a reduction in the length of intergenic spacers and a shortness of most putative proteins relative to their orthologs in other eukaryotes. Even so, *E. cuniculi* has approximately 2000 ORFs, which is 7–8 times the number in the minimal bacterial genome. The genome of the yeast *Schizosaccharomyces pombe* has about 4800 ORFs but is unlikely to represent the minimal free-living eukaryotic genome unless the *E. cuniculi* genome has lost many more essential genes than those metabolic and biosynthetic ones already recognized. The multicellular fungus *Neurospora crassa* has approximately 10,000 ORFs, about 25% fewer than the fruit fly *Drosophila melanogaster*. Many of these genes do not have homologs in either *Saccharomyces cerevisiae* or *S. pombe* but exactly how many of them are essential for multicellular existence remains to be seen.

Topic 147

DNA microarrays

Microarrays are used to study the expression of many genes at once. It involves the immobilization of thousands of gene sequences on glass slide that are then detected with complementary base pairing between the sample and the gene sequences on the chip. DNA segments from known genes, a few dozen to hundreds of nucleotides long, are amplified by PCR and placed on a solid surface, using robotic devices that accurately deposit nanoliter quantities of DNA solution. Many thousands of such spots are deposited in a predesigned array on a surface area of just a few square centimeters. An alternative strategy is to synthesize DNA directly on the solid surface, using photolithography. Once the chip is constructed, it can be probed with mRNAs or cDNAs from a particular cell type or cell culture to identify the genes being expressed in those cells. Two major types of DNA array are used in expression analysis.

i). Spotted DNA arrays

ii). Printed oligonucleotide chips

Topic 148

Spotted DNA arrays

Spotted DNA arrays are produced by printing DNA samples on treated microscope slides. It is made by transferring or spotting DNA clones or PCR product individually onto a solid support where they are immobilized. The technology arose directly from conventional hybridization analysis, and the first high-density cDNA arrays, now described as *macroarrays*, were essentially the same as the gridded reference libraries. Cloned cDNAs stored in a matrix format

in microtiter plates were transferred to nitrocellulose or nylon membranes in a precise grid pattern, allowing rapid identification of the clones corresponding to positive hybridization signals. For expression analysis, complete libraries could be hybridized with complex probes, generating a "fingerprint" specific to a particular RNA source. Early examples of the use of macroarrays for expression analysis include studies of differential gene expression in the mouse thymus and human muscle.

Topic 149

Oligonucleotide chips

The alternative to a spotted DNA array is a high density prefabricated oligonucleotide chip. These are similar to DNA arrays in that they consist of gridded DNA targets that are interrogated by hybridization. However, while DNA arrays consist of double stranded clones or PCR products that may be up to several hundred base pairs in length, oligo chips contain single-stranded targets ranging from 25-70 nt. Dual hybridization is not used for expression profiling on oligo chips. Instead, probes for chip hybridization are made from cleaved, biotinylated cRNA (RNA that has been transcribed *in vitro* from cDNA). Comparative expression analysis is carried out by hybridization of alternative cRNA samples to identical chips, followed by comparison of signal intensities. Oligo chips can be made in the same way as spotted DNA arrays, by robotically transferring chemically synthesized oligonucleotides from microtiter dishes to a solid support, where they are immobilized. However, the maximum array density is increased almost tenfold if the oligos are printed directly onto the glass surface.

Topic 150

Microbial gene expression analysis

Spotted arrays were manufactured containing PCR-amplified open reading frames representing most of the 6200 genes in the *S. cerevisiae* genome. These investigators analyzed the transcriptional profile of yeast cells shifted from fermentation (anaerobic) to aerobic metabolism, and as they were subjected to a variety of environmental manipulations, including heat shock. In each case, about 5% of the interrogated genes showed highly significant changes in expression induced by the experimental conditions, when unstimulated yeast cells were used as a source of control RNA. Genome wide expression profiling with arrays has also been carried out for a number of complex biological processes in yeast, such as sporulation, the cell cycle, and response to glucose. These studies have allowed tentative functions to be assigned to a number of

previously uncharacterized genes, based on their informative expression patterns. For example, in the study by Spellman and colleagues, 800 cell cycle-regulated genes were identified, about 400 of which were inducible by cyclins. Genes have also been identified whose expression is dependent on the ploidy (number of chromosome sets) of the cell. Furthermore, transcriptional profiling of yeast cells exposed to drugs has allowed novel drug targets to be identified.

Topic 151

Profiling in human diseases

Arrays have been widely used to investigate transcriptional profiles associated with human disease, and to identify novel disease markers and potential new drug targets. Many investigators have used arrays to profile transcriptional changes associated with cancer. In one of the earliest studies, cDNA arrays were used to investigate the ability of human chromosome 6 to suppress the tumorigenic phenotype of the melanoma cell line UACC-903. A number of novel tumor-suppressor genes were identified. Spotted cDNA arrays have also been used to investigate global gene expression in rheumatoid arthritis and inflammatory bowel disease, insulin resistance and asthma. In some cases, the investigation of global gene expression profiles has led to the discovery of novel links between biological processes.

Topic 152

Phage display

Phage display is the technology that allows expression of exogenous polypeptides on the surface of phage particles. The concept is simple in principle: a library of phage particles expressing a wide diversity of peptides is used to select those that bind the desired target. The filamentous phage M13 is the most commonly used vector to create random peptide display libraries. Several methods including recombinant techniques have been developed to increase the diversity of the library. On the other extreme, libraries with various biases can be created for specific purposes. For instance, when the sequence of the peptide that binds the target is known, its affinity and selectivity can be increased by screening libraries created with limited mutagenesis of the peptide. Phage libraries are screened for binding to synthetic or native targets. The initial screening of library by basic biopanning has been extended to column chromatography including negative screening and competition between selected phage clones to identify high affinity ligands with greater target specificity.

Topic 153

Screening phage display libraries

The most common screening method is based on enriching the phage clones with binding affinity for the target by a process called biopanning. Biopanning involves following steps: 1) target immobilization: The target molecules can be immobilized by passive adsorption to a modified 96-well polystyrene microtiter plates. The unbound target is washed off and the remaining sites in the well are blocked with unrelated proteins or nonionic detergents, 2) Phage binding: The phage display random peptide library is added to the target coated well in a solution that allows stability of the target and minimal non-specific binding of the phage. It is important to start the first round with a large and highly diverse library for a better chance of isolating peptides of interest, 3) Removing unbound phage: The first round of biopanning requires higher yield of the fittest phage clones over the background and hence the washes are less stringent. The stringency of selection can be increased by more extensive washes in subsequent rounds to isolate phage with higher affinity, 4) Phage elution: The specific elution of the target bound phage can be carried out in a solution containing either free target or a competing ligand. Due to stability of the filamentous phage, extremes of pH, denaturants, ionic strength, limited proteolysis or sonication can be used for non-specific elution of the target bound phage. The eluted phage is amplified and the biopanning process is repeated three to six times following which the phage clones are analyzed by DNA sequencing to identify the target binding peptides.

Topic 154

Applications of phage display

Over 80% of cellular proteins may work in complexes with other proteins and their protein–protein interactions are regulated by several mechanisms. Phage display has been used in numerous studies of protein-protein interactions. Its use with combinatorial mutagenesis provides a rapid method to identify residues contributing energetically to binding at protein-protein interfaces. Phage display random peptide libraries have been used in identifying novel interacting partners of proteins. Phage display has been used in enzymology to determine the substrate specificity and to develop modulators of both the active and allosteric sites of the enzyme. The method can be used to display mutants of enzymes to study their mechanisms of action. Since filamentous phage is resistant to broad range of proteases, it has been used in identification of substrates of various proteases Phage display of antibody fragments has been used successfully in generating target specific antibodies which can be useful in multiple

applications including proteomics, specific drug delivery and in analysis of intracellular processes. Phage display is a cheap and rapid method to map epitope of the antigen that is involved in specific interaction with the antibody. The identification of epitopes is essential in diagnostics, immunotherapy and vaccine development. Phage display peptide libraries can help identify critical residues within a continuous epitope that are involved in antibody binding.

Topic 155

Knock out and ins

A gene knock out is a genetic technique in which one of an organism's gene is made inoperative that is simply called as knocked out. Knockout is accomplished through a variety of techniques. It can be started in the test tube with a plasmid, a bacterial artificial chromosome or other DNA construct, and proceeding to cell culture. Individual cells are genetically transfected with the DNA construct. Often the goal is to create a transgenic animal that has the altered gene. If successful, embryonic stem cells are genetically transformed and inserted into early embryos. Resulting animals with the genetic change in their germline cells can then often pass the gene knockout to future generations.

On the other hand, a knock-in refers to a gene manipulation method that involves the insertion of a protein coding cDNA sequence at a particular locus in an organism's chromosome. A common use of knock-in technology is for the creation of disease models. It is a technique by which scientific investigators may study the function of the regulatory machinery i.e. promoters that governs the expression of the natural gene being replaced. This is accomplished by observing the new phenotype of the organism in question. The BACs and YACs are used in this case so that large fragments can be transferred.

Topic 156

siRNA technology

Small interfering RNA (siRNA) is the most commonly used RNA interference (RNAi) tool for inducing short-term silencing of protein coding genes. siRNA is a synthetic RNA duplex designed to specifically target a particular mRNA for degradation. The underlying molecular mechanism of gene silencing provides us with short interfering RNAs (siRNAs) which allows to target any gene with high specificity and efficiency. siRNAs can now be obtained in various ways allowing for numerous *in vitro* and *in vivo* applications. Successful knock-downs of disease-related genes indicate that siRNAs open the door for novel therapeutic procedures.

Topic 157

Applications of siRNA

"siRNA is the method of choice to target specific genes for silencing and has provided immense potential as therapeutic tools. The increasing knowledge of this fascinating mechanism of gene silencing provides us with nice tools to analyse gene function in a variety of experimental systems *in vitro* and *in vivo*. Initially, long dsRNAs were used in plants, *Drosophila* and other species. Applications in mammals were not feasable since long dsRNAs (>30 nt) induce a sequence-nonspecific interferon response which in turn results in global inhibition of mRNA translation. It has been shown that chemically synthesized siRNAs when transfected into mammalian cells are able to efficiently silence endogenous genes. Since, a large number of groups used this technique and targeted an ever increasing number of genes in a variety of mammalian cell lines. The transient reduction in mRNA levels can be detected in less than a day and lasts for several days depending on the turnover rate of the RNA or the protein".

Topic 158

Proteomics

The entire complement of proteins synthesized by a given cell or organism has been termed the proteome. Proteomics is the study of the expression of genetic information at the level of proteome and a number of methods are used to study proteins. Proteins can be studied in terms of sequence and abundance like mRNAs but also relevant are three-dimensional structure, modification, localization, biochemical and physiological function, and interactions with other proteins and other molecules. Protein expression analysis is more challenging than mRNA profiling because proteins cannot be amplified like nucleic acids.

Topic 159

Two dimensional (2-D) electrophoresis

The method involved first-dimension isoelectric focusing (i.e. separation of proteins according to charge) followed by second-dimension fractionation according to molecular mass. In the first application of this technology, 1000 proteins from the bacterium *Escherichia coli* were resolved on a single gel. The principle of isoelectric focusing (IEF) is that electrophoresis is carried out in a pH gradient, allowing each protein to migrate to its isoelectric point, i.e. the point at which its pI value is equivalent to the surrounding pH and its net charge is zero. Size fractionation is achieved by equilibrating the isoelectric focusing gel in a solution of the detergent sodium

dodecylsulfate (SDS), which binds non-specifically to all proteins and confers a uniform negative charge. The focused proteins are then separated in the perpendicular plane to the first separation. In the original study, *E. coli* proteins were separated by isoelectric focusing in a tube gel, i.e. a gel cast in a thin plastic tube. When the IEF run was complete, the tube was cracked open and the proteins exposed to SDS by immersion of the gel in an SDS solution. The tube gel was then attached to an SDS-PAGE slab gel, i.e. a flat gel cast between two plates, and the proteins were separated by size. The basic procedure for 2DGE has changed little since this time although the rather cumbersome tube gels have been largely replaced by strip gels, which are easier to handle and give more reproducible separations. After fractionation, the protein gel is stained. There is a wide choice of agents capable of staining proteins non-specifically, including silver nitrate and dyes such as Coomassie brilliant blue, which are commonly used on western blots.

Topic 160

Mass spectrometry

Mass spectrometry (MS) involves the ionization of target molecules in a vacuum, and accurate measurement of the mass of the resulting ions. A mass spectrometer has three component parts: an ionizer, which converts the anylate into gas phase ions; a mass analyzer which separates the ions according to their mass/charge ratio (m/z); and an ion detector. Generally, large molecules such as proteins and nucleic acids are broken up and degraded by the ionization procedure, but more recently, sensitive instruments that are capable of soft ionization, i.e. the ionization of large molecules without significant degradation, have been developed. This allows accurate mass measurements of whole proteins and peptide fragments, data that can be used to search protein databases to identify particular proteins.

Topic 161

Protein microarrays

2D-electrophoresis is an open system for proteome analysis, rather like direct sequence sampling is an open system for transcriptome analysis. The advantage of an open system is that potentially all proteins can be detected, but the disadvantage is that they also have to be characterized, which relies on downstream annotation by mass spectrometry. DNA arrays are closed systems in transcriptome analysis, i.e. the data obtained are constrained by the number and nature of

sequences immobilized on the array. However, it is not necessary to characterize any of the features on the array by sequencing because the sequences are already known. Similarly, protein arrays are emerging as a useful closed system for proteome analysis. These are miniature devices in which proteins, or molecules that recognize proteins, are arrayed on the surface.

In concept, protein arrays are no different to DNA arrays, but they suffer from several practical limitations. First, the manufacture of DNA arrays is simplified by the availability of methods, such as the polymerase chain reaction, for amplifying any nucleic acid sequence. No amplification procedure exists for proteins. Second, all DNA sequences are made of the same four nucleotides and hence behave similarly in terms of their chemical properties. The principles of molecular recognition (hybridization between complementary base pairs) apply to all sequences. For this reason, hybridization reactions can be carried out in highly parallel formats using a single complex probe. Conversely, proteins are made of 22 amino acids specified by the genetic code plus many others generated by post-translational modification, so they have diverse chemical properties.

Topic 162

Cloning in gram negative other than *E. coli*

For majority of the experiments, it is convenient to use *E. coli* as a recipient for genes cloned from eukaryotes or other prokaryotes. Transformation is easy and there is available a wide range of easy-to-use vectors with specialist properties, e.g. regulatable high-level gene expression. However, use of *E. coli* is not always practicable because it lacks some auxiliary biochemical pathways that are essential for the phenotypic expression of certain functions, e.g. degradation of aromatic compounds, antibiotic synthesis, pathogenicity, sporulation, etc. In such circumstances, the genes have to be cloned back into species similar to those whence they were derived. There are three prerequisites for cloning genes in a new host. First, there needs to be a method for introducing the DNA of interest into the potential recipient. The methods available include transformation, conjugation and electroporation. Secondly, the introduced DNA needs to be maintained in the new host. Finally, the uptake and maintenance of the cloned genes will only be detected if they are expressed. Thus the inability to detect a cloned gene in a new bacterial host could be due to failure to introduce the gene, to maintain it or to express it, or to a combination of these factors. Another cause of failure could be restriction. For example, the frequency of transformation of *Pseudomonas putida* with plasmid RSF1010 is 10^5 transformants/μg DNA, but

only if the plasmid is prepared from another *P. putida* strain. Otherwise, no transformants are obtained.

Topic 163

Cloning in gram +ve bacteria: B. subtilis

The development of *B. subtilis* vectors began with the observation that plasmids from *S. aureus* can be transformed into *B. subtilis*, where they replicate and express antibiotic resistance normally. None of the natural *S. aureus* plasmids carries more than one selectable marker; therefore, improved vectors have been constructed by gene manipulation, e.g. pHV11 is pC194 carrying the *Tc*R gene of pT127. In general, these plasmids are stable in *B. subtilis*, but segregative stability is greatly reduced following insertion of exogenous DNA. Reasoning that stable host-vector systems in *B. subtilis* are more likely if endogenous plasmids are used; therefore, cryptic *Bacillus* plasmid pTA1060 as a vector were developed for this purpose. Because of the difficulties experienced in direct cloning in *B. subtilis*, hybrid plasmids were constructed which can replicate in both *E. coli* and *B. subtilis*. Originally most of these were constructed as fusions between pBR322 and pC194 or pUB110. With such plasmids, *E. coli* can be used as an efficient intermediate host for cloning. Plasmid preparations extracted from *E. coli* clones are subsequently used to transform competent *B. subtilis* cells. Such preparations contain sufficient amounts of multimeric plasmid molecules to be efficient in *B. subtilis* competent cell transformation.

Topic 164

Cloning in Streptomyces

Cloning in *Streptomyces* has attracted a lot of interest because of the large number of antibiotics that are made by members of this genus. Although *Streptomyces coelicolor* is the model species for genetic studies, many other species are the subject of intensive study and methods developed for one species may not work particularly well in another. Streptomycete DNA has a G+C content of 70–75% and this affects the frequency of restriction sites. As might be expected, AT-rich recognition sites are rare and this can be useful if large-sized fragments of DNA are wanted. For the construction of gene libraries, the most commonly used enzymes are ones with a high GC content in their recognition sequence, e.g. *Bam*H1 (G′GATCC), *Bgl*II (A′GATCT), and *Bcl*I (T′GATCA). In *Streptomyces*, promoters may be several hundred base pairs upstream of the start of the gene and so can be lost during gene cloning. Also, many *Streptomyces* promoters are

complex and may include tandem sites for recognition by different sigma factors. Streptomycetes are good at expressing genes (promoters, ribosome binding sites, etc.) from low-G+C organisms, but *Streptomyces* genes are usually difficult to express in *E. coli* because most promoters do not function, and translation may be inefficient unless the initial amino acid codons are changed to lower-G+C alternatives.

Topic 165

Cloning in Archea

Organisms are divided into three domains: *Bacteria* (eubacteria), *Archaea* (archaebacteria), and *Eucarya* (eukaryotes). The Archaea comprises at least three major groups of prokaryotic organisms with unusual phenotypes when compared with their eubacterial counterparts. For example, many of them thrive in extreme environments such as ones with very high temperatures, high pH, or high salt concentrations. Others are strictly anaerobic methanogens that live in the rumen of herbivores. Because many Archaea exhibit interesting physiological properties, coupled with a growing number of complete genome sequences, there has been considerable interest in developing suitable gene-cloning procedures. As mentioned earlier, there are three basic requirements for gene cloning in any organism: a means of introducing DNA into recipients; suitable plasmid vectors; and selectable markers. All three requirements have been met but progress lags far behind that made with the major eubacterial groups. In most Archaea investigated so far, electroporation is the only method of getting DNA into cells and the observed efficiencies are low (~10^2 transformants/μg). Several species appear to be naturally transformable but, again, efficiencies are as low as with electroporation. Much higher efficiencies (e.g. 10^7 transformants/g) have been observed with some methanogens using polyethylene glycol-mediated transformation or liposome-mediated transformation.

Topic 166

Cloning in Saccharomyces and other Fungi

In the late 1970s the organism of choice for recombinant DNA technology was the yeast *Saccharomyces cerevisiae*. At that time the primary purpose of cloning was to understand what particular genes do *in vivo* and the concomitant development of DNA sequencing methodology facilitated the identification of the different elements that control gene expression in fungi. A secondary reason for cloning in yeast was to understand those cellular functions unique to eukaryotes such as mitosis, meiosis, signal transduction, obligate cellular differentiation, etc.

Today, just over a quarter of a century later, there are different reasons for cloning in *S. cerevisiae* and other fungi. In 1996 the sequencing of the entire 12 Mb genome of *S. cerevisiae* was completed and most, if not all, of the genes have been identified. The key biological questions now are what products do each of these genes encode, how do the different gene products interact, and under what circumstances is each gene expressed? Answering these questions still requires gene manipulation but the emphasis has switched from analysis of individual genes to that of the whole genome. That is, there has been a switch from the reductive approach to a holistic one. A second reason for the current interest in gene manipulation in fungi is the overproduction of proteins of commercial value. Fungi offer a number of advantages, such as the ability to glycosylate protein, the absence of pyrogenic toxins, and in the case of the methylotrophic yeast *Pichia pastoris*, the ability to get very high yields of recombinant proteins.
Yeasts and other fungi are widely used in the production of food and beverages and recombinant DNA technology can be used to enhance their desirable properties. The third reason for the current interest in the application of recombinant DNA technology in yeast is the ability to clone very large pieces of DNA.

Topic 167

Transformation of fungi with exogenous DNA

Fungi are not naturally transformable and artificial means have to be used for introducing foreign DNA. One method involves the use of spheroplasts (i.e. wall-less cells) and was first developed for *S. cerevisiae*. In this method, the cell wall is removed enzymically and the resulting spheroplasts are fused with ethylene glycol in the presence of DNA and $CaCl_2$. The spheroplasts are then allowed to generate new cell walls in a stabilizing medium containing 3% agar. Electroporation provides a simpler and more convenient alternative to the use of spheroplasts. Cells transformed by electroporation can be selected on the surface of solid media, thus facilitating subsequent manipulation. Both the spheroplast technique and electroporation have been applied to a wide range of yeasts and filamentous fungi. DNA can also be introduced into yeasts and filamentous fungi by conjugation.

Topic 168

Vectors for use in S. cerevisiae

If the heterologous DNA introduced into fungi is to be maintained in an extrachromosomal state then plasmid vectors are required which are capable of replicating in the fungal host. Four types

of plasmid vector have been developed: yeast episomal plasmids (YEps), yeast replicating plasmids (YRps), yeast centromere plasmids (YCps), and yeast artificial chromosomes (YACs). All of them have features in common. First, they all contain unique target sites for a number of restriction endonucleases. Secondly, they can all replicate in *E. coli*, often at high copy number. This is important, because for many experiments it is necessary to amplify the vector DNA in *E. coli* before transformation of the ultimate yeast recipient. Finally, they all employ markers that can be selected readily in yeast and which will often complement the corresponding mutations in *E. coli* as well. The four most widely used markers are *His*3, *Leu*2, *Trp*1, and *Ura*3.

Topic 169

Properties of different yeast vectors

The analysis of yeast DNA sequences has been facilitated by the ease with which DNA from eukaryotes can be cloned in *E. coli* using the vectors. Such cloned sequences can be obtained easily in large amounts and can be altered *in vivo* by bacterial genetic techniques and *in vitro* by specific enzyme modifications. To determine the effects of these experimentally induced changes on the function and expression of yeast genes they need to be taken out of bacteria and returned to the yeast cell. The availability of different kinds of vectors with different properties enables yeast geneticists to do this with relative ease and to perform subsequent manipulations in yeast itself. Thus cloned genes can be used in conventional genetic analysis by means of recombination using YIp vectors or linearized YRp vectors. Complementation can be carried out using YEp, YRp, YCp, or YAC vectors.

Topic 170

Promoter system for yeast

Yeast promoters are more complex than bacterial promoters. First, several consensus sequences are found at the transcription-initiation site. Two of these sequences, TC(G/A)A and PuPuPyPuPu, account for more than half of the known yeast initiation sites. These sequences are not found at transcription-initiation sites in higher eukaryotes, which implies a mechanistic difference in their transcription machinery compared with yeast. The second motif in the yeast promoter is the TATA box. This is an AT-rich region with the canonical sequence TATAT/AAT/A, located 60–120 nucleotides before the initiation site. Functionally, it can be considered equivalent to the Pribnow box of *E. coli* promoters. The third and fourth structural elements are upstream activating sequences (UASs) and upstream repressing sequences (URSs).

These are found in genes whose transcription is regulated. Binding of positive-control proteins to UASs increases the rate of transcription and deletion of the UASs abolishes transcription. An important structural feature of UASs is the presence of one or more regions of dyad symmetry. Binding of negative-control proteins to URSs reduces the transcription rate of those genes that need to be negatively regulated.

Topic 171

Multipurpose vectors for use in yeast

Specialist yeast vectors have been developed which incorporate the useful features found in the corresponding *E. coli* vectors, e.g. an f1 origin to permit sequencing of inserts, production of cloned gene product as a purification fusion, etc. In yeast, proteins destined for the cell surface or for export from the cell are synthesized on and translocated into the endoplasmic reticulum. From there they are transported to the Golgi body for processing and packaging into secretory vesicles. Fusion of the secretory vesicles with the plasma membrane then occurs constitutively or in response to an external signal. Of the proteins naturally synthesized and secreted by yeast, only a few end up in the growth medium, e.g. the mating pheromone α factor and the killer toxin. The remainder, such as invertase and acid phosphatase, cross the plasma membrane but remain within the periplasmic space or become associated with the cell wall. Polypeptides destined for secretion have a hydrophobic amino-terminal extension, which is responsible for translocation to the endoplasmic reticulum. The extension is usually composed of about 20 amino acids and is cleaved from the mature protein within the endoplasmic reticulum. Such signal sequences precede the mature yeast invertase and acid phosphatase sequences.

Topic 172

Cloning for large DNA fragments

When the first YACs were developed it was found that their segregative stability was determined by their size. If the size of the YAC was less than 20 kb then centromere function was impaired whereas much larger YACs segregated normally. This fact was considered in developing a vector for cloning large DNA molecules. One problem with large DNA molecules is that they are difficult to manipulate in the liquid phase prior to transformation and keeping them intact is very difficult. Thus, many of the early YAC libraries had average insert sizes of only 50–100 kb. By removing small DNA fragments by PFGE fractionation prior to cloning, the average insert size was increased to 350 kb.

Topic 173

Deficiencies and advantages of YACs

There are a number of operational problems associated with the use of YACs. The first of these is that it is estimated that 10–60% of clones in existing libraries represent chimeric DNA sequences: i.e. sequences from different regions of the genome cloned into a single YAC. Chimeras may arise by co-ligation of DNA inserts *in vitro* prior to yeast transformation, or by recombination between two DNA molecules that were introduced into the same yeast cell. It is possible to detect chimeras by *in situ* hybridization of the YAC to metaphase chromosomes: hybridization to two or more chromosomes or to geographically disparate regions of the same chromosome is indicative of a chimera. A second problem with YACs is that many clones are unstable and tend to delete internal regions from their inserts.

Topic 174

GENE TRANSFER TO PLANT CELLS-Introduction

Plants provide human beings with all manner of useful products: food and animal feed, fibers and structural materials, and small molecules that can be used as dyes, scents, and medicines. Plants have been cultivated for these products since the dawn of history, and for the same length of time people have sought to improve plants by breeding them and selecting the better-performing and most useful varieties. The one limitation of this approach is that breeders are restricted to the existing gene pool in each species or sexually compatible group of species. In order to surmount this barrier, it has been necessary to develop technologies for gene transfer to plants. During the 1960s and 1970s several attempts to transfer DNA into plant tissues were reported but stable transformation was never confirmed. The introduction of foreign DNA into a plant followed by stable transmission through the germline was first demonstrated in 1981, when transgenic tobacco plants were generated by transformation using the soil bacterium *Agrobacterium tumefaciens*. In the 25 years following this report, foreign genes have been introduced into well over 100 different plant species either through the use of *A. tumefaciens* or alternative strategies involving direct DNA transfer to plant cells and tissues. In addition, plant viruses have been developed as versatile episomal vectors, allowing high-level transient gene expression. This research has founded agricultural biotechnologies industry in which plants are manipulated to make them resistant to pests and diseases, to improve their tolerance of stress, to

improve their nutritional characteristics and even to act as factories producing therapeutic proteins and industrial enzymes.

Topic 175

Callus culture

Tissue culture is the process whereby small pieces of living tissue (explants) are isolated from an organism and grown aseptically for indefinite periods on a nutrient medium. For successful plant tissue culture it is best to start with an explant rich in undetermined cells because such cells are capable of rapid proliferation. The usual explants are buds, root tips, nodal stem segments or germinating seeds, and these are placed on suitable culture media where they grow into an undifferentiated mass known as a callus (Fig. 15). Since the nutrient media used for plants can also support the growth of microorganisms, the explant is first washed in a disinfectant such as sodium hypochlorite or hydrogen peroxide. Once established, the callus can be propagated indefinitely by subdivision. Usually callus cultures are maintained in the dark because light can induce differentiation of callus cells. For plant cells to develop into a callus it is essential that the nutrient medium contains the correct balance of plant hormones (phytohormones) to maintain the cells in an undifferentiated state. There are five main classes of plant hormones: auxins, cytokinins, gibberellins, abscisic acid, and ethylene. The correct balance of auxins and cytokinins is most important for callus culture, and the exact relative amounts need to be determined empirically for each species and explant type. A low auxin:cytokinin ratio leads to shoot formation whereas a high ratio favors the formation of roots.

Figure 16. Close-up view of a callus culture

Topic 176

Cell-suspension culture

Depending on the species and culture conditions, callus tissue can become hard and compact or soft and easily breakable. The latter is known as friable callus, and when transferred into liquid medium and agitated, the cell mass breaks up to give a suspension of isolated cells, small clusters of cells, and larger aggregates. Such suspensions can be maintained indefinitely by subculture but, by virtue of the presence of aggregates, are extremely heterogeneous. Genetic instability adds to this heterogeneity, so that long-term culture results in the accumulation of mutations (somaclonal variation) which can adversely affect the vitality and fertility of regenerated plants.

If placed in a suitable medium, isolated single cells from suspension cultures are capable of division.

Topic 177

Protoplasts

Protoplasts are cells from which the cellulose walls have been removed. They are very useful for genetic manipulation for two reasons: first, several transformation protocols have been developed that work specifically with protoplasts; and second, because under certain conditions, protoplasts from similar or contrasting cell types can be fused to yield somatic hybrids, a process known as protoplast fusion. Protoplasts can be produced from suspension cultures, callus tissue, or intact tissues, e.g. leaf mesophyll cells, by mechanical disruption or, preferably, by treatment with cellulolytic and pectinolytic enzymes. Pectinase is necessary to break up cell aggregates into individual cells and the cellulase digests away the cell wall. After enzyme treatment, protoplast suspensions are collected by centrifugation, washed in medium without the enzyme, and separated from intact cells and cell debris by flotation on a cushion of sucrose. When plated onto nutrient medium, protoplasts will synthesize new cell walls within 5-10 days and then initiate cell division.

Topic 178

Regeneration of fertile plants

The developmental plasticity of plant cells means that whole fertile plants can often be regenerated from tissue explants, callus, cell suspensions, or protoplasts by placing them on appropriate media. As discussed above, the maintenance of cells in an undifferentiated state requires the correct balance of phytohormones. However, only cytokinin is required for shoot

culture and only auxin for root culture, therefore increasing the level of cytokinins available to the callus induces shoot formation and increasing the auxin level promotes root formation. Ultimately plantlets arise through the development of adventitious roots on shoot buds, or through the development of shoot buds from tissues formed by proliferation at the base of rootlets. The formation of roots and shoots on callus tissue is known as organogenesis. The culture conditions required to achieve organogenesis vary from species to species, and have not been determined for every type of callus. As discussed, the adventitious organogenesis of shoots and roots can also occur directly from organized plant tissues such as stem segments, without first passing through a callus stage.

Topic 179

Major strategies of gene transfer to plants

Gene transfer to plants can be achieved through four types of mechanism-viral transduction, bacterial gene delivery, and chemical and physical direct DNA transfer. Unlike the situation in animals, where bacterial gene transfer is a relatively new development, *Agrobacterium*-mediated transformation is the most widely used transformation method, particularly for dicotyledonous plants. Physical methods are the next most popular, especially particle bombardment for the transformation of monocotyledonous plants such as cereals. Chemical transfection methods are little used, and are compatible only with protoplasts, which behave in many ways analogously to animal cells. Many of the techniques used to transfect animal cells can therefore be applied to plant protoplasts, e.g. calcium phosphate transfection. All three of the above methods can be used for either transient expression or stable transformation. Another major difference between gene-transfer strategies in animal and plant cells is that no known plant viruses integrate their genetic material into the plant genome as part of the natural infection cycle. Therefore, plant viruses are used as episomal vectors rather than for stable transformation. However, while stable transformation cannot be achieved, plant viruses often cause systemic infections resulting in the rapid production of high levels of recombinant protein throughout the plant, and they can be transmitted through normal infection routes, or by grafting infected scions onto virus-free hosts.

Topic 180

Agrobacterium- mediated transformation

Gene transfer from bacteria to plants occurs naturally and is responsible for crown gall disease. This is a plant tumor that can be induced in a wide variety of gymnosperms and dicotyledonous

angiosperms (dicots) by inoculation of wound sites with the Gram negative soil bacterium *A. tumefaciens*. The involvement of bacteria in this disease was established nearly 100 years ago. It was subsequently shown that the crown gall tissue represents true oncogenic transformation, since the undifferentiated callus can be cultivated *in vitro* even if the bacteria are killed with antibiotics, and yet retains its tumorous properties. These properties include the ability to form a tumor when grafted onto a healthy plant, the capacity for unlimited growth as a callus in tissue culture even in the absence of phytohormones necessary for the *in vitro* growth of normal cells, and the synthesis of opines, such as octopine and nopaline, which are unusual amino acid derivatives not found in normal plant tissue. Since the continued presence of *Agrobacterium* is not required to maintain plant cells in their transformed state, it is clear that some "tumor-inducing principle" is transferred from the bacterium to the plant at the wound site. It was noted that virulent strains of *A. tumefaciens* harbor large plasmids (140–235 kbp), and experiments involving the transfer of such plasmids between various octopine- and nopaline-utilizing strains soon established that virulence and the ability to use and induce the synthesis of opines are plasmid-borne traits (Fig. 17). The plasmids therefore became known as tumor-inducing plasmids (Ti-plasmids).

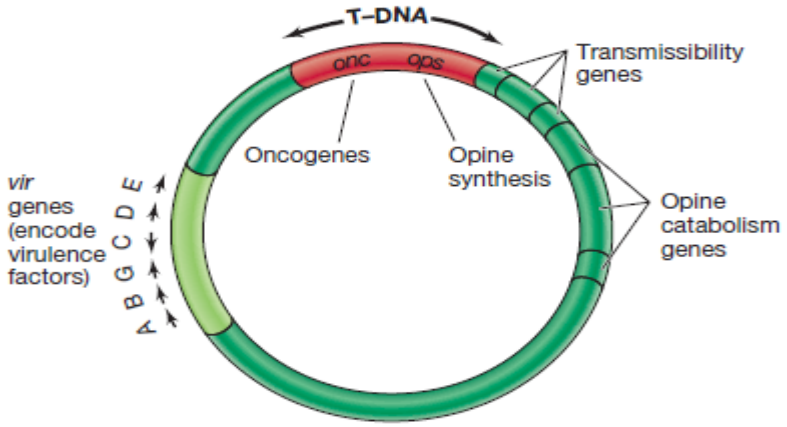

Figure 17. Structure of Ti-plasmid of *Agrobacterium tumefaciens*

Topic 181

T-DNA transfer

Complete Ti-plasmid DNA is not found in plant tumor cells but a small, specific segment of the plasmid, about 23 kbp in size, is found integrated in the plant nuclear DNA at an apparently random site. This DNA segment is called T-DNA (transferred DNA) and carries genes that confer both unregulated growth and the ability to synthesize opines upon the transformed plant tissue. However, these genes are non-essential for transfer and can be replaced with foreign DNA In the Ti-plasmid itself, the T-DNA is flanked by 25-bp imperfect direct repeats known as border sequences, which are conserved between octopine and nopaline plasmids. The border sequences are not transferred intact to the plant genome, but they are involved in the transfer process.

Topic 182

Function of T-DNA genes

Genetic maps of T-DNA have been obtained by studying spontaneous and transposon-induced mutants that affect tumour morphology, generating tumours that are larger than normal or that show ‘shooty’ or ‘rooty’ phenotypes. Although normal tumours can grow on medium lacking auxins and cytokinins, the tumour cells actually contain high levels of these hormones. It has been proposed that the oncogenes carried on the TDNA encoded products involved in phytohormone synthesis and that the abnormal morphologies of T-DNA mutants were due to a disturbance in the balance of plant hormones in the callus. The cloning and functional analysis of T-DNA genes has confirmed that those with ‘shooty’ mutant phenotypes encode enzymes for auxin biosynthesis and those with ‘rooty’ phenotypes are involved in cytokine production. Other genes have been identified as encoding enzymes for opine synthesis, while the function of some genes remains unknown.

Topic 183

Disarmed Ti vectors

Ti plasmid is a natural vector for genetically engineering plant cells because it can transfer its T-DNA from the bacterium to the plant genome. However, wild-type Ti plasmids are not suitable as general gene vectors because the T-DNA contains oncogenes that cause disorganized growth of the recipient plant cells. To be able to regenerate plants efficiently, we must use vectors in which the T-DNA has been *disarmed* by making it nononcogenic. This is most effectively

achieved simply by deleting all of its oncogenes. For example, scientist have substituted pBR322 sequences for almost all of the T-DNA of pTiC58, leaving only the left and right border regions and the *nos* gene. The resulting construct was called pGV3850 (Fig. 18). *Agrobacterium* carrying this plasmid transferred the modified T-DNA to plant cells. As expected, no tumour cells were produced, but the fact that transfer had taken place was evident when the cells were screened for nopaline production and found to be positive. Callus tissue could be cultured from these nopaline-positive cells if suitable phytohormones were provided, and fertile adult plants were regenerated by hormone induction of plantlets. The creation of disarmed T-DNA was an important step forward, but the absence of tumour formation made it necessary to use an alternative method to identify transformed plant cells. In the experiment described above, opine production was exploited as a screenable phenotype, and the *ocs* and *nos* genes have been widely used as screenable markers.

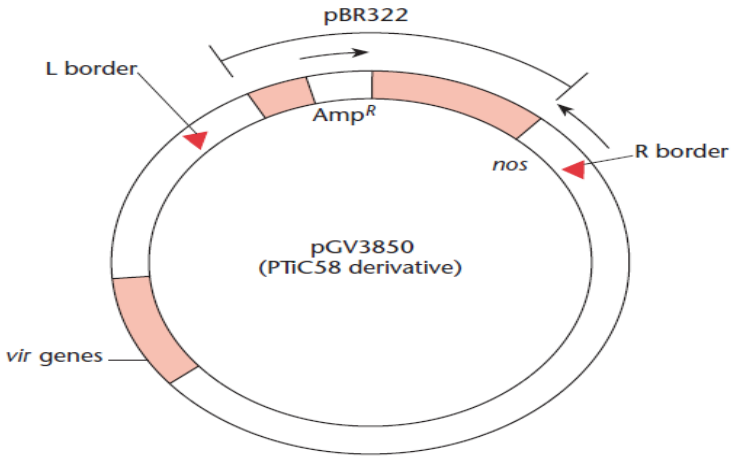

Figure 18. Structure of the Ti-plasmid in which T-DNA has been disarmed

Topic 184

Two plasmid strategy for transformation

A common cloning strategy employs an *Agrobacterium* with two different recombinant plasmids. The first is a Ti plasmid from which the T DNA segment has been removed in the laboratory. The second is an *Agrobacterium–E. coli* shuttle vector in which the 25 bp repeats of the T DNA flank a foreign gene that the researcher wants to introduce into the plant cell, along

with a selectable marker such as resistance to the antibiotic kanamycin. The engineered *Agrobacterium* is used to infect a leaf, but crown galls are not formed because the T DNA genes for the auxin, cytokinin, and opine biosynthetic enzymes are absent from both plasmids. Instead, the *vir* gene products from the altered Ti plasmid direct the transformation of the plant cells by the foreign gene-the gene flanked by the T DNA 25 bp repeats in the second plasmid. The transformed plant cells can be selected by growth on agar plates that contain kanamycin, and addition of growth hormones induces the formation of new plants that contain the foreign gene in every cell.

Topic 185

Leaf-disk transformation by *A. tumefaciens*

Once the principle of selectable, disarmed T-DNA vectors was established, there followed an explosion in the number of experiments involving DNA transfer to plants. In the original report, small discs (a few millimetres in diameter) were punched from leaves, surface-sterilized and inoculated in a medium containing *A. tumefaciens* transformed with the recombinant disarmed T-DNA (as a cointegrate or binary vector). The foreign DNA contained a chimeric *neo* gene conferring resistance to the antibiotic kanamycin. The discs were cultured for 2 days and transferred to medium containing kanamycin, to select for the transferred *neo* gene, and carbenicillin, to kill the *Agrobacterium.* After 2–4 weeks, developing shoots were excised from the callus and transplanted to root-induction medium. Rooted plantlets were subsequently transplanted to soil, about 4–7 weeks after the inoculation step.

Topic 186

Agrobacterium and monocots

Most monocotyledonous plants (monocots) were thought to be outside the host range of *Agrobacterium*, prompting research into alternative transformation methods. During the 1980s, limited evidence accumulated showing that some monocots might be susceptible to *Agrobacterium* infection. However, in most cases there was no convincing evidence for T-DNA integration into the plant genome. In the laboratory, it proved possible to induce tumours in certain monocot species, such as asparagus and yam. In the latter case, an important factor in the success of the experiment was pretreatment of the *Agrobacterium* suspension with wound exudate from potato tubers. It has been argued that *Agrobacterium* infection of monocots is inefficient because wounded monocot tissues do not produce phenolics, such as acetosyringone,

at sufficient levels to induce *vir* gene expression. In the last 10 years, amazing progress has been made in the transformation of cereals using *Agrobacterium*. The first species to be transformed was rice.

Topic 187

A. rhizogenes and Ri plasmids

Agrobacterium rhizogenes causes hairy-root disease in plants and this is induced by root-inducing (Ri) plasmids, which are analogous to the Ti plasmids of *A. tumefaciens*. Ri plasmids are of interest from the point of view of vector development, because opine-producing root tissue induced by Ri plasmids in a variety of dicots can be regenerated into whole plants by manipulation of phytohormones in the culture medium. Ri TDNA is transmitted sexually by these plants and affects a variety of morphological and physiological traits, but does not in general appear deleterious. The Ri plasmids therefore appear to be already equivalent to disarmed Ti plasmids. Transformed roots can also be maintained as hairy root cultures, which have the potential to produce certain valuable secondary metabolites at higher levels than suspension cultures and are much more genetically stable

Topic 188

Protoplasts transformation

Until comparatively recently, the limited host range of *A. tumefaciens* precluded its use for the genetic manipulation of a large number of plant species, including most monocots. At first, the only alternative to *Agrobacterium*-mediated transformation was the introduction of DNA into protoplasts. This process has much in common with the transfection of animal cells. The protoplasts must initially be persuaded to take up DNA from their surroundings, after which the DNA integrates stably into the genome in a proportion of these transfected cells. Gene transfer across the protoplast membrane is promoted by a number of chemicals, of which polyethylene glycol has become the most widely used, due to the availability of simple transformation protocols. Alternatively, DNA uptake may be induced by electroporation, which has also become a favoured technique. As with animal cells, the introduction of a selectable marker gene along with the transgene of interest is required for the identification of stable transformants. This can be achieved using plasmid vectors carrying both the marker and the transgene of interest.

Topic 189

Particle bombardment

An alternative procedure for plant transformation was introduced in 1987, involving the use of a modified shotgun to accelerate small (1-4 μm) metal particles into plant cells at a velocity sufficient to penetrate the cell wall (~250 m/s). In the initial test system, intact onion epidermis was bombarded with tungsten particles coated in tobacco mosaic virus (TMV) RNA. Three days after bombardment, approximately 40% of the onion cells that contained particles also showed evidence of TMV replication. A plasmid containing the *cat* reporter gene driven by the CaMV 35S promoter was then tested to determine whether DNA could be delivered by the same method. Analysis of the epidermal tissue 3 days after bombardment revealed high levels of transient chloramphenicol transacetylase (CAT) activity.

Topic 190

Chloroplast transformation

Chloroplast is also a useful target for genetic manipulation, because thousands of chloroplasts may be present in photosynthetic cells and this can result in levels of transgene expression up to 50 times higher than possible using nuclear transformation. Furthermore, transgenes integrated into chloroplast DNA do not appear to undergo silencing or suffer from position effects that can influence the expression levels of transgenes in the nuclear DNA. Chloroplast transformation also provides a natural containment method for transgenic plants, since the transgene cannot be transmitted through pollen.

Topic 191

In planta transformation

Until recently, gene transfer to plants involved the use of cells or explants as transformation targets and an obligatory tissue-culture step was needed for the regeneration of whole fertile plants. Experiments using the model dicot *Arabidopsis thaliana* have led the way in the development of so-called *in planta* transformation techniques, where the need for tissue culture is minimized or eliminated altogether. Such methods involve the introduction of DNA, either by *Agrobacterium* or by direct transfer, into intact plants. The procedure is carried out at an appropriate time in the plant's life cycle, so that the DNA becomes incorporated into cells that will contribute to the germ line, directly into the germ cells themselves (often at around the time of fertilization) or into the very early plant embryo. Generally, *in planta* transformation methods

have a very low efficiency, so the small size of *Arabidopsis* and its ability to produce over 10 000 seeds per plant is advantageous. This limitation has so far prevented *in planta* techniques from being widely adopted for other plant species.

Topic 192

Plant viruses as vectors

As an alternative to stable transformation using *Agrobacterium* or direct DNA transfer, plant viruses can be employed as gene-transfer and expression vectors. There are several advantages to the use of viruses.

i). viruses are able to adsorb to and introduce their nucleic acid into intact plant cells. However, for many viruses, naked DNA or RNA is also infectious, allowing recombinant vectors to be introduced directly into plants by methods such as leaf rubbing.

ii). infected cells yield large amounts of virus, so recombinant viral vectors have the potential for high-level transgene expression.

iii). viral infections are often systemic. The virus spreads throughout the plant, allowing transgene expression in all cells.

iv). viral infections are rapid, so large amounts of recombinant protein can be produced in a few weeks.

v). Finally, all known plant viruses replicate episomally; therefore the transgenes they carry are not subject to the position effects that often influence the expression of integrated transgenes. Since plant viruses neither integrate into nor pass through the germ line, plants cannot be stably transformed by viral infection and transgenic lines cannot be generated. However, this limitation can also be advantageous in terms of containment".

Topic 193

“Genetic modification of animals

Techniques are available for the introduction of DNA or RNA into hundreds of different cell types in culture, and such experiments can be used to study gene function and regulation, to produce recombinant proteins and to manipulate the endogenous genome. Mammalian cells are the most widely used hosts for gene delivery, since they allow the production of recombinant human proteins with authentic post-translational modifications that are not carried out by bacteria, yeast, or plants. Indeed, mammalian cells are cultured on a commercial scale for the synthesis of many valuable products, including antibodies, hormones, growth factors, cytokines, and vaccines. There has been intense research into the development of efficient vector systems and transformation methods for animal cells, based on viral, bacterial, and synthetic delivery vectors.

Topic 194

Strategies to transform animal cells

Gene transfer to animal cells can be achieved using four broad types of delivery mechanism (Fig. 19). Two of these are described as biological mechanisms because the target cells need to be infected with a biological delivery vector, such as a virus or bacterium, which carries the exogenous genetic material. Delivery using a viral vector is known as transduction, and many different viruses have been adapted as gene-delivery vectors. Delivery using bacterial vectors is a more recent development, which in most cases relies on bacteria which also invade animal cells. In this case, however, the transgene is delivered not as part of the bacterial genome, but on a plasmid which is carried by the bacterium. Bacterial gene delivery is sometimes termed bactofection. The other two delivery mechanisms are described as non-biological because biological delivery vectors are not required. To distinguish such methods from those involving *infection* with a bacterium or virus, the term transfection is used. In *chemical transfection* methods, cells are persuaded to take up DNA from their surroundings when the DNA is presented as a synthetic complex – either a complex with an overall positive charge, allowing it to interact with the negatively charged cell membrane and promote uptake by endocytosis; or a lipophilic complex which fuses with the membrane and deposits the transgene directly into the cytoplasm. In physical transfection methods, naked DNA is introduced directly into the cell by

exploiting a physical forcc e.g. microinjection, particle bombardment, ultrasound, and electroporation.

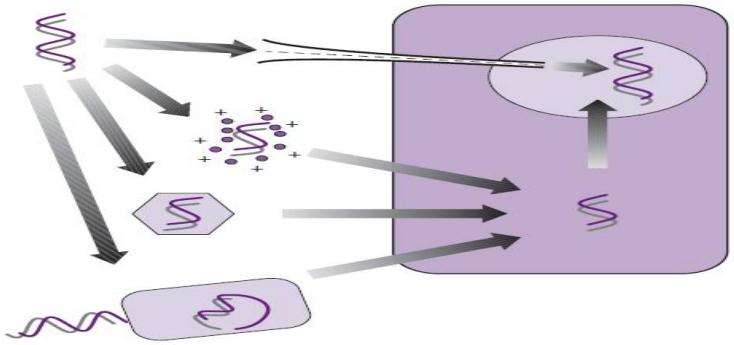

Figure 19. The four principal mechanisms by which DNA can be introduced into animal cells

Topic 195

Chemical transfection techniques-Calcium phosphate method

The calcium phosphate method involves the formation of a co-precipitate which is taken up by endocytosis. In this method, total uncloned genomic DNA was used to transfect human cells deficient for the enzyme hypoxanthine-guanine phosphoribosyltransferase (HPRT). Rare HPRT-positive cells, which had presumably taken up fragments of DNA containing the functional gene, were identified by selection on HAT medium. At this time, the actual mechanism of DNA uptake was not understood. Much later, it was appreciated that successful DNA transfer in such experiments was dependent on the formation of a fine DNA/calcium phosphate co-precipitate, which first settles onto the cells and is then internalized. The precipitate must be formed freshly at the time of transfection. It is thought that small granules of calcium phosphate associated with DNA are taken up by endocytosis. The DNA then escapes and some reaches the nucleus and can be expressed.

Topic 196

Transfection with polyplexis

The calcium phosphate method is applicable to many cell types, but some cell lines are adversely affected by the co-precipitate due to its toxicity and are hence difficult to transfect. Alternative

chemical transfection methods have been developed to address this problem using polycationic compounds that form soluble complexes (polyplexes) through spontaneous electrostatic interactions with DNA. The earliest of these methods utilized DEAE-dextran (diethylaminoethyl dextran), a soluble polycationic carbohydrate. This technique was initially devised to introduce viral RNA and DNA into cells but was later adapted as a method for plasmid DNA transfer. The efficiency of the original procedure was improved later on. Although efficient for the transient transfection of many cell types, DEAE-dextran cannot be used to generate stably transformed cell lines and has thus fallen out of favor. A new generation of polycationic compounds has been developed more recently, which are more versatile and can be used for the stable transformation not only of proliferating cell lines but also of primary cells and post-mitotic cells such as neurons and myocytes.

Topic 197

Transfection with liposomes and lipoplexes

An alternative chemical transfection procedure is to package DNA inside a fusogenic phosopholipid vesicle, which interacts with the target cell membrane and facilitates DNA uptake. In first example, bacterial protoplasts containing plasmids to transfer DNA into mammalian cells were used. Briefly, bacterial cells were transformed with a suitable plasmid vector and then treated with chloramphenicol to amplify the plasmid copy number. Lysozyme was used to remove the cell walls, and the resulting protoplasts were gently centrifuged onto a monolayer of mammalian cells and induced to fuse with them using polyethylene glycol. The procedures are very efficient in terms of the number of transformants obtained, but they are also labor-intensive and so have not been widely adopted as a general transfection method. However, an important advantage is that they are gentle, allowing the transfer of large DNA fragments without shearing. Yeast cells with the cell wall removed (spheroplasts) have therefore been used to introduce YAC DNA into mouse ES cells by this method, for the production of YAC transgenic mice. More widespread use has been made of artificial phospholipid vesicles, which are called liposomes. Initial liposome-based procedures were hampered by the difficulty encountered in encapsulating the DNA, and the transfection efficiency was no better than that of the calcium phosphate method. However, a breakthrough came with the discovery that cationic/neutral lipid mixtures can spontaneously form stable complexes with DNA (lipoplexes) that interact productively with the cell membrane, resulting in DNA uptake by endocytosis

Topic 198

Physical transfection techniques-Electrophoration and ultrasound

Electroporation is a physical transfection technique which involves the generation of transient, nanometer-sized pores in the cell membrane, by exposing cells to a brief pulse of electricity. DNA enters the cell through these pores, and is transported to the nucleus. This technique has been successfully used to introduce plasmid DNA into mouse fibroblasts. The electroporation technique has been adapted to many other cell types. The most critical parameters are the intensity and duration of the electric pulse, and these must be determined empirically for different cell types. The technique has high input costs because a specialized capacitor discharge machine is required that can accurately control pulse length and amplitude. Additionally, larger numbers of cells may be required than for other methods because in many cases, the most efficient electroporation occurs when there is up to 50% cell death. Ultrasound transfection involves the exposure of cells to a rapidly oscillating probe, such as the tip of a sonicator. The transfection mechanism is similar in some ways to electroporation in that the application of ultrasound waves to a dish of cells or a particular tissue results in the formation and collapse of bubbles in the liquid, including the cell membrane, a process known as cavitation. The transient appearance of such cavities allows DNA to cross the membrane into the cytoplasm.

Topic 199

Microinjection: Direct transfer of DNA

Another method involves the direct transfer of DNA into the cell, without a synthetic carrier. One such procedure is microinjection, a technique that is guaranteed to generate successful hits on target cells but that can only be applied to a few cells in any one experiment. This technique has been applied to cultured cells that are recalcitrant to other transfection methods. For *in vivo* applications, conventional needle injection appears to be an efficient way to transfer DNA into target cells. The DNA can be injected directly into tissues such as skin, muscle, or internal organs or it can be injected into the blood.

Topic 200

Selectable markers for animal cells-Endogenous markers

Endogenous selectable markers are already present in the cellular genome, and mutant cell lines are required when they are used. For calcium phosphate transfection method, it was shown that

mouse cells deficient for the enzyme thymidine kinase (TK) could be stably transformed to a wildtype phenotype by transfecting them with the herpes simplex virus (HSV) *Tk* gene. As for the HPRT+ transformants discussed earlier in the chapter, cells positive for TK can be selected on HAT medium. This is because both enzymes are required for nucleotide biosynthesis via the salvage pathway. In mammals, nucleotides are produced via two alternative routes, the *de novo* and salvage pathways. In the *de novo* pathway, nucleotides are synthesized from basic precursors such as sugars and amino acids, while the salvage pathway recycles nucleotides from DNA and RNA. If the *de novo* pathway is blocked, nucleotide synthesis becomes dependent on the salvage pathway, and this can be exploited for the selection of cells carrying functional *Hprt* and *Tk* genes. The drug aminopterin blocks the *de novo* synthesis of both inosine monophosphate (IMP) and thymidine monophosphate (TMP) by inhibiting key enzymes in the *de novo* pathway. Cells exposed to aminopterin can thus survive only if they have functional *Hprt* and *Tk* genes and a source of hypoxanthine and thymidine. *Hprt*+ and *Tk*+ transformants can therefore both be selected using HAT medium, which contains hypoxanthine, aminopterin, and thymidine.

Topic 201

Bacterial vectors for the transfection of animal cell

Stable transformation by integration can be achieved using any source of DNA. The early gene-transfer experiments were carried out using complex DNA mixtures, e.g. genomic DNA, bacterial plasmids, and phage. Calcium phosphate transfection was used in most of these experiments, and the specific donor DNA was often bulked up with a nonspecific carrier such as cleaved salmon sperm DNA. However, it is generally more beneficial to use a purified source of the donor transgene. The use of plasmid vectors for transfection provides numerous other advantages, depending on the modular elements included on the plasmid backbone. (i) The convenience of bacterial plasmid vectors can be extended to animal cells, in terms of the ease of subcloning, *in vitro* manipulation, and purification of recombinant proteins. (ii) More importantly, modular elements can be included to drive transgene expression, and these can be used with any transgene of interest. The pSV and pRSV plasmids are examples of early expression vectors for use in animal cells, containing transcriptional control sequences from SV40 and Rous sarcoma virus which are functional in a wide range of cell types. The incorporation of these sequences into pBR322 generated convenient expression vectors in which any transgene could be controlled by these promoters when integrated into the genome of a

transfected cell. (iii) The inclusion of a selectable marker gene obviates the need for co-transformation, since the transgene and marker remain linked when they co-integrate into the recipient cell's genome. A range of pSV and pRSV vectors was developed containing alternative selectable marker genes, e.g. pSV2-neo, pSV2-gpt, and pSV2-dhfr. (iv) some plasmid vectors for gene transfer to animal cells are designed to be shuttle vectors, i.e. they contain origins of replication functional in animal cells allowing the vector to be maintained as an episomal replicon.

Topic 202

Viruses as vector for gene transfer

Virus particles have a natural ability to adsorb to the surface of cells and gain entry, and this can be exploited to deliver recombinant DNA into animal cells. Due to the efficiency with which viruses can deliver their nucleic acid into cells, and the high levels of replication and gene expression it is possible to achieve, viruses have been used as vectors not only for gene expression in cultured cells, but also for gene transfer to living animals. Several classes of viral vector have been developed for use in human gene therapy, and at least eight have been used in clinical trials. Still others have been developed as recombinant vaccines.

Topic 203

Methods for the production of transgenic mice

The ability to introduce DNA into the germline of mice is one of the greatest achievements of the twentieth century and has paved the way for the transformation of other mammals. Genetically modified mammals have been used not only to study gene function and regulation, but also as bioreactors producing valuable recombinant proteins, e.g. in their milk. Several methods for germline transformation have been developed, all of which require the removal of fertilized eggs or early embryos from donor mothers, brief culture *in vitro*, and then their return to foster mothers, where development continues to term. Note that these methods have been developed with nuclear transgenesis in mind, but mitochondria have their own genome. Recently, methods have been developed for the production of mitochondrial transgenics.

Topic 204

Pronuclear microinjection

Direct microinjection of DNA was the first strategy used to generate transgenic mice. Simian virus 40 (SV40) DNA was injected into the blastocoele cavities of preimplantation embryos. The

embryos were then implanted into the uteri of foster mothers and allowed to develop. The DNA was taken up by some of the embryonic cells and occasionally contributed to the germline, resulting in transgenic mice containing integrated SV40 DNA in the following generation. Transgenic mice have also been recovered following the injection of viral DNA into the cytoplasm of the fertilized egg. The technique that has become established is the injection of DNA into one of the pronuclei of the egg. Just after fertilization, the small egg nucleus (female pronucleus) and the large sperm nucleus (male pronucleus) are discrete. Since the male pronucleus is larger, this is usually chosen as the target for injection. About 2 pl of DNA solution is transferred into the nucleus through a fine needle, while the egg is held in position with a suction pipette. The injected embryos are cultured *in vitro* to the morula stage and then transferred to pseudopregnant foster mothers. The procedure requires specialized microinjection equipment and considerable dexterity from the handler. The exogenous DNA may integrate immediately or, less commonly, may remain extrachromosomal for one or more cell divisions. Thus the resulting animal may be transgenic or may be mosaic for transgene insertion.

Topic 205

Recombinant retroviruses-germline transformation

Recombinant retroviruses provide a natural mechanism for stably introducing DNA into the genome of animal cells. Retroviruses are able to infect early embryos (as well as ES cells), so recombinant retroviral vectors can be used for germline transformation. An advantage over the microinjection technique is that only a single copy of the retroviral provirus is integrated, and the genomic DNA surrounding the transgenic locus generally remains intact. The infection of preimplantation embryos with a recombinant retrovirus is technically straightforward and, once the infected embryos are implanted in the uterus of a foster mother, can lead to germline transmission of the transgene. However, there are also considerable disadvantages to this method, including the limited amount of foreign DNA that can be carried by the virus, the possible interference of viral regulatory elements with the expression of surrounding genes, and the susceptibility of the virus to *de novo* methylation, resulting in transgene silencing.

Topic 206

Transfection of ES cells

ES cells are derived from the inner cell mass of the mouse blastocyst and thus have the potential to contribute to all tissues of the developing embryo. The ability of ES cells to contribute to the

germline was first demonstrated in 1984 and requires culture conditions that maintain the cells in an undifferentiated state. Since these cells can be serially cultured like any other established cell line, DNA can be introduced by transfection or viral transduction and the transformed cells can be selected using standard markers. In contrast, since there is no convenient way to select for *eggs or embryos* that have taken up foreign DNA, each potential transgenic mouse generated by pronuclear microinjection must be tested by Southern blot hybridization or the PCR to confirm transgene integration.

Topic 207

Gene transfer to *Xenopus*

Gurdon *et al*. (1971) first showed that *Xenopus* oocytes synthesized large amounts of globin after they had been microinjected with rabbit globin mRNA. Since then, the *Xenopus* oocyte expression system has been a valuable tool for expressing a very wide range of proteins from plants and animals. *X. laevis* is an African clawed frog. Oocytes can be obtained in large numbers by removal of the ovary of an adult female. Each fully grown oocyte is a large cell (0.8–1.2 mm diameter) arrested at first meiotic prophase. This large cell has a correspondingly large nucleus (called the *germinal vesicle*), which is located in the darkly pigmented hemisphere of the oocyte. Due to the large size of the oocytes, mRNA-either natural or synthesized by transcription *in vitro*, using phage-T7 RNA polymerase (Melton 1987)-can be readily introduced into the cytoplasm or nucleus by microinjection. This is achieved using a finely drawn glass capillary as the injection needle, held in a simple micromanipulator. DNA can also be injected. The oocyte nucleus contains a store of the three eukaryotic RNA polymerases, enough to furnish the needs of the developing embryo at least until the 60,000-cell stage. The RNA polymerases are available for the transcription of injected exogenous DNA. Using this system, it has therefore been possible to express complementary DNAs (cDNAs) linked to a heat-shock promoter or to mammalian virus promoters. In addition, vaccinia virus vectors can be used for gene expression in the cytoplasm.

Topic 208

Rabbit reticulocytes

Reticulocytes are immature red blood cells. Reticulocytes are a valuable tool for biologists who study protein translation. Reticulocytes are unusual among cells in that they contain all of the

machinery necessary to translate proteins but lack a nucleus. Since a cell's nucleus contains many components that make studying translation difficult, these cells are quite useful. Scientists can collect reticulocytes from animals such as rabbits and extract the mRNA and translation enzymes to study protein translation in a cell-free, in vitro system, allowing greater control over the environment in which proteins are being synthesized.

Topic 209

Frog oocyte system

All modes of *Xenopus* research (embryos, cell-free extracts, and oocytes) are commonly used in direct studies of human disease genes and to study the basic science underlying initiation and progression of cancer. *Xenopus* oocytes for studies of gene expression and channel activity related to human disease. Another strength of *Xenopus* is the ability to rapidly and easily assay the activity of channel and transporter proteins using expression in oocytes. This application has also led to important insights into human disease.

Topic 210

The different ways that recombinant DNA technology has been exploited

Trait modification through germline manipulation is practiced in animals and plants, but the same technology can be applied to humans in a more limited sense to prevent or cure diseases. This field of medicine, known as gene medicine, includes the concepts of DNA vaccines, using gene transfer to kill cancer cells, and introducing DNA into specific cells to repair or compensate for genetic defects. In the converse approach, animal genomes may be deliberately mutated by gene-transfer techniques to recreate the mutations seen in human inherited diseases or develop disease models that can be used to investigate how diseases occur and test novel drugs. A comprehensive survey of all the applications of gene manipulation is mentioned in **figure**.

Topic 211

Commercial production of recombinant therapeutic proteins

Recombinant therapeutic proteins are produced commercially in bacteria, yeast, and mammalian cells. One of the earliest commercial applications of gene manipulation was the production of human therapeutic proteins in bacteria. Not surprisingly, the first such products were recombinant versions of proteins already used as therapeutics: human growth hormone and insulin. Prior to the advent of genetic engineering, human growth hormone was obtained from pituitary glands removed from cadavers. Not only did this limit the supply of the hormone but in

some cases resulted in recipients contracting Creutzfeld–Jacob syndrome. The recombinant approach resulted in unlimited supplies of safe material. This safety aspect has been extended to various blood components, anticoagulants, and growth factors that were originally isolated from blood but now carry the risk of HIV infection. As the methods for cloning genes became more and more sophisticated an increasing number of cytokines, inerferons, and interleukins have been identified and produced in bacteria.

Topic 212

Biopharmaceuticals approved in different countries

Biopharmaceuticals approved in the United States and Europe includes recombinant blood factors, anticoagulants, hormones, interferones, vaccines, monoclonal antibody-based products. The first generation of protein drugs were exact copies of the human molecules but protein engineering now is being used to develop second-generation molecules with improved properties, e.g. by replacing certain amino acids or removing particular protein domains. Many of the proteins are modified in this manner to improve efficacy or longevity when administered by injection. Another trend is the increasing production of novel therapeutics combining parts of different proteins. Nowhere is this more apparent than in the production of antibody-based drugs, which often combine the binding specificity of antibodies with the activity of other proteins such as cytokines and lymphokines to achieve therapeutic effects on particular target cells.

Topic 213

Transgenic animals and plants to produce recombinant 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. In later years, several groups reported the secretion of recombinant proteins in mouse milk. In each case, this was achieved by joining the transgene to a mammary-specific promoter, such as that from the casein gene. The first proteins produced in this way were sheep β-lactoglobulin and human tissue plasminogen activator (tPA). Although proteins can be produced at high concentrations in mouse milk (e.g. 50 ng/ml for tPA) the system is not ideal due to the small volume of milk produced. Therefore, other animals, such as sheep and goats, have been investigated as possible bioreactors.

Topic 214

Therapeutic recombinant proteins expressed in genetically modified plants

Plants have been explored as cheap, safe, and efficient production systems for subunit vaccines, with the added advantage that orally administered vaccines can be ingested by eating the plant, therefore limiting the need for processing and purification. The earliest demonstration was the expression of a surface antigen from the bacterium *S. mutans* in tobacco. This bacterium is the cause of tooth decay, and it was envisaged that stimulation of a mucosal immune response would prevent the bacterium colonizing the teeth and therefore protect against cavities. A number of edible transgenic plants have been generated expressing antigens derived from animal viruses. For example, rabies glycoprotein has been expressed in tomato, hepatitis B virus antigen in lettuce, and cholera antigen in potato. As well as animal virus antigens, autoantigens associated with diabetes have also been produced. Plants have also been infected with recombinant viruses expressing various antigen epitopes on their surfaces.

Topic 215

Metabolic engineering-production of small molecules in bacteria

Metabolic engineering allows the directed production of small molecules in bacteria. When the large-scale production of penicillin began in the 1940s, yields were measured in micrograms per liter of culture. Demand for the antibiotic was outstripping supply and higher yielding strains were badly needed. Since nothing was known about the biosynthetic pathway a program of strain improvement was set in place that involved random mutation and screening. The best strain from each cycle of improvement then became the starting point for the next round of selection. In this way the yield of penicillin was steadily increased until it reached the tens of grams per liter that can be achieved today. As each new antibiotic was discovered the same process of strain improvement was applied. In every case, the biochemical and genetic bases of the beneficial mutations were not known. Only when the details of gene regulation and metabolic pathway regulation had been elucidated could we even begin to understand how antibiotic yields might have been improved.

Topic 216

Metabolic engineering-in plant to produce diverse chemical structures

Plants synthesize an incredibly diverse array of useful chemicals. Most are products of secondary

metabolism, that is, biochemical pathways that are not involved in the synthesis of essential cellular components but more complex molecules that provide additional functions. Examples of these functions are attraction of pollinators and resistance to pests and pathogens. In many cases, these secondary metabolites have specific and potent pharmaceutical properties in humans: well-known examples include caffeine, nicotine, morphine, and cocaine. Plants have long been exploited as a source of pharmaceutical compounds, and a number of species are cultivated specifically for the purpose of extracting drugs and other valuable molecules. We already discussed that how gene transfer to bacteria and yeast can be used to produce novel chemicals, so in theory it would be possible to transfer the necessary components from these useful plants into microbes for large-scale production. However, the secondary metabolic pathways of plants are so extensive and complex, that in most cases such a strategy would prove impossible. Fortunately, advances in plant transformation have made it possible to carry out metabolic engineering in plants themselves, and largescale plant cell cultures can be used in the same manner as microbial cultures for the production of important phytochemicals.

Topic 217

Herbicide resistance in commercial transgenic plants

Herbicides are used to kill weeds, and generally affect processes that are unique to plants, e.g. photosynthesis or amino acid biosynthesis. Both crops and weeds share these processes, and developing herbicides that are selective for weeds is very difficult. An alternative approach is to modify crop plants so that they become resistant to broad spectrum herbicides, i.e. incorporating selectivity into the plant itself rather than relying on the selectivity of the chemical. Two strategies to engineer herbicide resistance have been adopted. In the first, the target molecule in the cell either is rendered insensitive or is overproduced. In the second, a pathway that degrades or detoxifies the herbicide is introduced into the plant.

Topic 218

Production of virus resistance crops

Major crop losses occur every year as a result of viral infections, e.g. tobacco mosaic virus (TMV) causes losses of over $50 million per annum in the tomato industry. There is a useful phenomenon known as cross-protection in which infection of a plant with one strain of virus protects against superinfection with a second, related strain. The mechanism of cross-protection is not fully understood but it is believed that the viral coat protein is important. Transgenic plants

have been developed which express the TMV coat protein and which greatly reduced disease symptoms following virus infection. Since that observation, the principle of heterologous coat protein expression has been extended to many different plants and viruses. In the case of resistance to TMV, the coat protein must be expressed in the epidermis and in the vascular tissue through which the virus spreads systemically. Transgenic squash containing multiple viral coat protein genes and demonstrating resistance to cucumber mosaic virus, watermelon mosaic virus 2, and zucchini yellow mosaic virus was the first virus-resistant transgenic crop to reach commercial production.

Topic 219

Resistance against fungal pathogens

Progress also has been made in developing resistance to fungal pathogens, which are traditionally controlled by appropriate farming practices (e.g. crop rotation) and the application of expensive and environmentally harmful fungicides. A straight forward approach is to engineer plants with antifungal proteins from heterologous species. For instance, it has been shown that expression of bean chitinase can protect tobacco and oilseed rape from post-emergent damping off caused by *Rhizoctonia solani*. Plants synthesize a wide range of so-called "pathogenesis-related proteins" (PR proteins), such as chitinases and glucanases, which are induced by microbial infection and there are now many examples of such proteins expressed in plants to provide protection against fungi – in some cases multiple PR proteins have been expressed with synergistic effects: tobacco, carrot, and tomato have each been engineered to express both chitinase and glucanase simultaneously, and these plants show greater fungal resistance, indicating that the two enzymes can work. Plants also synthesize anti-fungal peptides called defensins, and other anti-fungal proteins that have been shown to confer resistance to pathogens when overexpressed in transgenic plants.

Topic 220

Resistance to bacterial diseases

Bacterial diseases cause significant losses in crop yields, and many different transgenic strategies have been developed to prevent infection or reduce the severity of symptoms. One of the most prevalent bacterial diseases of rice is bacterial blight, which causes losses totaling over $250 million every year in Asia alone. This disease has received a great deal of attention due to the discovery of a resistance gene complex in the related wild species *O. longistaminata*. The trait

was introgressed into cultivated rice line IR-24 and was shown to confer resistance to all known isolates of the blight pathogen *Xanthomonas oryzae* pv. *oryzae* in a few Asian countries. Further investigation of the resistance complex resulted in the isolation of a gene, named *Xa*21, encoding a receptor tyrosine kinase. The transfer of this gene to susceptible rice varieties resulted in plant lines showing strong resistance to a range of isolates of the pathogen.

Topic 221

Bacillus thuringiensis-source of insect resistance genes

Insect pests represent one of the most serious biotic constraints to crop production. For example, more than one-quarter of all the rice grown in the world is lost to insect pests, at an estimated cost of nearly $50 billion. This is despite an annual expenditure of approximately $1.5 billion on insecticides for this crop alone. Insect-resistant plants are therefore desirable not only because of the potential increased yields, but also because the need for insecticides is eliminated and, following on from this, the undesirable accumulation of such chemicals in the environment is avoided. Typical insecticides are non-selective, so they kill harmless and beneficial insects as well as pests. For these reasons, transgenic plants have been generated expressing toxins that are selective for particular insect species. Research is being carried out on a wide range of insecticidal proteins from diverse sources. However, all commercially produced insect-resistant transgenic crops express toxin proteins from the Gram-positive bacterium *Bacillus thuringiensis* (Bt). Unlike other *Bacillus* species, *B. thuringiensis* produces crystals during sporulation, comprising one or a small number of ~130 kDa protoxins called crystal proteins. These proteins are potent and highly specific insecticides. The specificity reflects interactions between the crystal proteins and receptors in the insect midgut. In susceptible species, ingested crystals dissolve in the alkaline conditions of the gut and the protoxins are activated by gut proteases. The active toxins bind to receptors on midgut epithelial cells, become inserted into the plasma membrane, and form pores that lead to cell death through osmotic lysis.

Topic 222

Drought resistant transgenic crops

After pests and diseases, unfavorable environmental conditions (*abiotic stresses*) represent the next major limitation on crop production. One of the most prevalent of these conditions is drought, and the development of transgenic crops with built-in drought resistance could increase the global yield of food by up to 30%. Many plants respond to drought (prolonged dehydration)

and increased salinity by synthesizing small, very soluble molecules such as betaines, sugars, amino acids, and polyamines. These are collectively termed *compatible solutes*, and they increase the osmotic potential within the plant, therefore preventing water loss in the short term and helping to maintain a normal physiological ion balance in the longer term. Compatible solutes are non-toxic even at high concentrations, so one strategy to provide drought resistance is to make such molecules accumulate in transgenic plants. For example, several species have been engineered to produce higher levels of glycine betaine but in most cases the levels achieved have fallen short of the 5–40 μmol/g fresh weight observed in plants that naturally accumulate this molecule under salt stress conditions.

Topic 223

Transgenic plants to cope poor soil quality

About 65% of the world's potential arable land consists of marginal soils, which are characterized by extremes of pH, limited nutrient availability (particularly phosphorus and iron), and high levels of toxic metal ions. Acidic soils, which account for 40% of the arable land, have low levels of available phosphorus and iron but high levels of aluminum. In an acidic environment, both iron and aluminum sequester phosphorus into insoluble or poorly soluble molecules. Aluminum is also toxic in its own right, its major effect being the inhibition of root development. Alkaline soils account for 25% of arable land. The major problem in alkaline soils is the high level of calcium and magnesium ions, which also sequester phosphorus into insoluble and sparingly soluble molecules. Calcium is an essential signaling molecule in plants and high levels of this metal ion can interfere with normal plant growth and metabolism. Despite these problems, many plants have adapted to grow in marginal soils and some tolerant varieties of crop plants have also been produced by mutation and conventional breeding. A common factor among these tolerant plants is the increased exudation of organic acids, such as citrate, malate, and oxalate, from the roots. The production of transgenic crop plants engineered to exude higher levels of organic acids is therefore an attractive strategy to increase the use of marginal soils. Several species have been transformed with bacterial or plant citrate synthase genes to increase organic acid production and induce tolerance of poor soils.

Topic 224

Plant biotechnology to increase food yields

The amount of usable food obtained from a field of plants can vary tremendously, and much effort has been expended in attempts to increase yields by conventional breeding and optimizing farming practices. Genetic engineering provides a wide range of strategies not only for reducing yield losses (by increasing resistance to pests and diseases and providing tolerance to abiotic stress) but also by increasing the intrinsic yield potential of the plants. In terms of yield enhancement, photosynthesis is perhaps the most obvious target for genetic intervention because it determines the rate of carbon fixation, and therefore the overall size of the organic carbon pool. Strategies for increasing photosynthetic activity include the modification of light-harvesting phytochromes and key photosynthetic enzymes. Progress has been made in crop species by attempting to introduce components of the energy-efficient C_4 photosynthetic pathway into C_3 plants, which lose a proportion of their fixed carbon through photorespiration. The key step in C_4 photosynthesis is the conversion of CO_2 into C_4 organic acids by the enzyme phosphoenolpyruvate carboxylase (PEPC) in mesophyll cells. The maize gene encoding PEPC has been transferred into several C_3 crops, including potato and rice in order to increase the overall level of carbon fixation. Transgenic rice plants were also produced expressing pyruvate orthophosphate dikinase (PPDK) and NADP-malic enzyme. Preliminary field trials in China and Korea demonstrated 10-30% and 30-35% yield increases for PEPC and PPDK transgenic rice plants, respectively, which was quite unexpected since only one C4 enzyme was expressed in each case.

Topic 225

Transgenic animals as model of human disease

Mammals have been used as models for human disease for many years, since they can be exploited to carry out detailed analyses of the molecular basis of disease and to test newly developed therapeutics prior to clinical trials in humans. Before the advent of transgenic animal technology, however, models of inherited diseases (i.e. diseases with a genetic basis) were difficult to come by. They could be obtained as spontaneously occurring mutants, suitable mutant animals identified in mutagenesis screens, and susceptible animal strains obtained by selective breeding. Gene manipulation now offers a range of alternative strategies to create *specific* disease models. Some of the earliest transgenic disease models were mice predisposed to

particular forms of cancer because the germ line contained exogenously derived oncogenes. This exemplifies so-called gain-of-function diseases, which are caused by a dominantly acting allele and can be modeled simply by adding that allele to the normal genome, e.g. by microinjection into eggs.

Topic 226

Gene medicine to prevent, treat, or cure disease

Gene medicine refers to the use of the same technology to ameliorate or even permanently cure diseases in humans. Gene medicine has a wide scope and includes the use of DNA vaccines, the targeted killing of disease cells (e.g. cancer cells), the use of oligonucleotides as defects (gene therapy). Gene transfer can be carried out in cultured cells, which are then reintroduced into the patient, or DNA can be transferred to the patient *in vivo*, directly or by using viral vectors. The *ex vivo* approach can be applied only to certain tissues, such as bone marrow, in which the cells are amenable to culture. Gene therapy can be used to treat diseases caused by mutations in the patient's own DNA (inherited disorders, cancers), as well as infectious diseases, and is particularly valuable in cases where no conventional treatment exists, or that treatment is inherently risky.

Topic 227

DNA vaccines

DNA vaccines are expression constructs whose products stimulate the immune system. The immune system generates antibodies in response to the recognition of proteins and other large molecules carried by pathogens. With typical vaccines, the functional component of the vaccine introduced into the host is the protein that elicits the immune response. The introduction of DNA into animals does not generate an immune response against the DNA molecule, but, if that DNA is expressed to yield a protein, that protein can stimulate the immune system. This is the basis of DNA vaccination, as first demonstrated in 1993. DNA vaccines generally comprise a bacterial plasmid carrying a gene encoding the appropriate antigen under the control of a strong promoter that is recognized by the host cell. The advantages of this method include its simplicity, its wide applicability, and the ease with which large quantities of the vaccine can be produced. The DNA may be administered by injection, using liposomes or by particle bombardment.

Topic 228

Gene augmentation therapy

The first human genetic engineering experiment was one of gene marking rather than gene therapy, and was designed to demonstrate that an exogenous gene could be safely transferred into a patient and that this gene could subsequently be detected in cells removed from the patient. Both objectives were met. Tumor-infiltrating lymphocytes (cells that naturally seek out cancer cells and then kill them by secreting proteins such as tumor necrosis factor, TNF) were isolated from patients with advanced cancer. The cells were then genetically marked with a neomycin resistance gene and injected back into the same patient. The first clinical trial using a therapeutic gene transfer procedure involved a four-year-old female patient, Ashanthi DeSilva, suffering from severe combined immune deficiency, resulting from the absence of the enzyme adenosine deaminase (ADA). This disease fitted many of the ideal criteria for gene-therapy experimentation. The disease was life threatening (therefore making the possibility of unknown treatment-related side-effects ethically acceptable) but the corresponding gene had been cloned and the biochemical basis of the disease was understood (Fig 20).

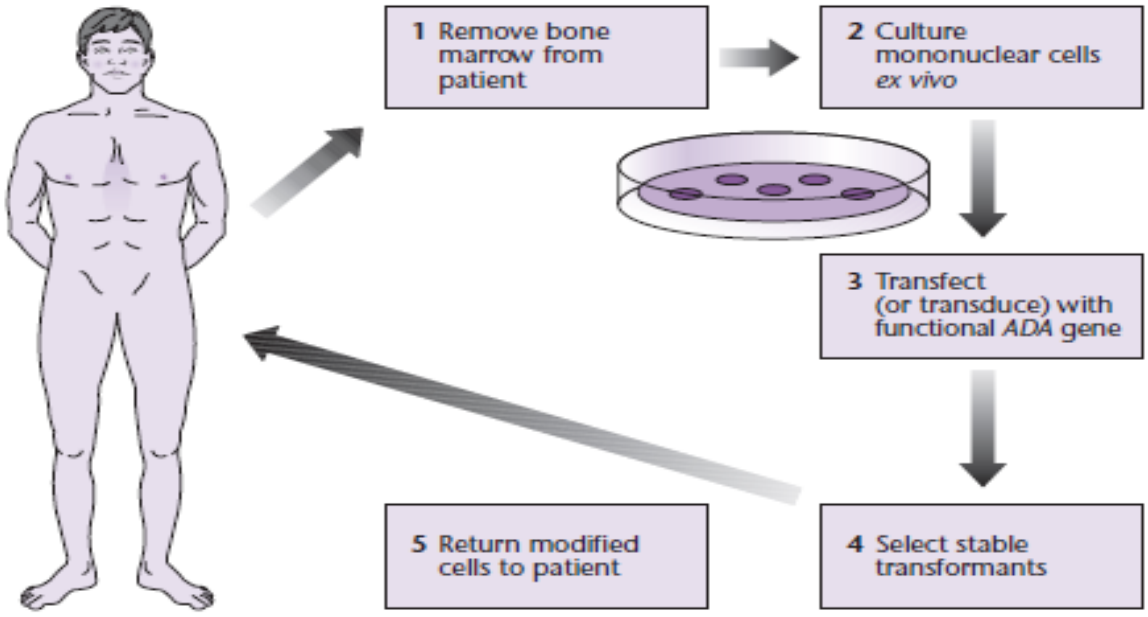

Figure 20. Procedure for *ex vivo* gene therapy

Topic 229

Gene therapy strategies for cancer

Cancer gene therapy was initially an extension of the early gene-marking experiments. The tumor infiltrating leukocytes were transformed with a gene for TNF in addition to the neomycin-resistance gene, with the aim of improving the efficiency with which these cells kill tumors by increasing the amount of TNF they secreted. Although TNF is highly toxic to humans at levels as low as 10 μg/kg body weight, there have been no side-effects from the gene therapy and no apparent organ toxicity from secreted TNF. One alternative strategy is to transform the tumor cells themselves, making them more susceptible to the immune system through the expression of cytokines or a foreign antigen. Another is to transform fibroblasts, which are easier to grow in culture, and then co-inject these together with tumor cells to provoke an immune response against the tumor. A number of such "assisted killing" strategies have been approved for clinical trials.

Topic 230

Ethical issues in manipulation

Advances in the basic science of genetics usually pose few problems from an ethical standpoint. The major concerns are usually separate from the actual experiments – perhaps the use of animals in research, or the potential for transgenic crops to contaminate non-transgenic or wild populations. In medicine, few would argue against the development of new drugs and therapies, where clear benefit is obtained. Perhaps the one area in the medically related applications of genetic research that is difficult is the human genome information. Genetic screening, and thus the possibility of genetic discrimination, is an area of active debate at the moment. The molecular diagnosis of genetically based disease is now well established, and the major ethical dilemmas tend to centre around whether or not a foetus should be aborted if a disease-causing trait is detected.

Transgenic organisms set up several ethical questions. The one thing that has been a little surprising is the reversal of the usual plant/animal debate as far as transgenesis is concerned. Traditionally, animal welfare has been the major source of difficulty between pressure groups, concerned individuals, scientists and regulators. Plants were largely ignored in the ethical debate until the late 1990s, when the public backlash against genetically modified foods began to

influence what biotech companies were doing. Concerns were in two areas – the effect of GMO-derived foods on health, and the effect on the environment. The environmental debate in particular has been driven by many different groups, who claim that an ecological disaster might be waiting to emerge from GM plant technology due to cross-pollination. It is impossible to predict what might happen in such cases, although the protagonists of GM crops claim to have evaluated the risks. The simple answer is that we do not know what the long-term ecological effects might be".