

# *Gene Mapping Techniques*

## **OBJECTIVES**

*By the end of this session the student should be able to:*

- *Define genetic linkage and recombinant frequency*
- *State how genetic distance may be estimated*
- *State how restriction enzymes can be used for isolating genes*
- *Define Restriction Fragment Length Polymorphisms (RFLPs)*
- *Identify the main applications of RFLPs in gene mapping and carrier detection*
- *State the principles used in identifying a specific gene in the genome by hybridisation with a specific gene probe.*

## GENETIC LINKAGE

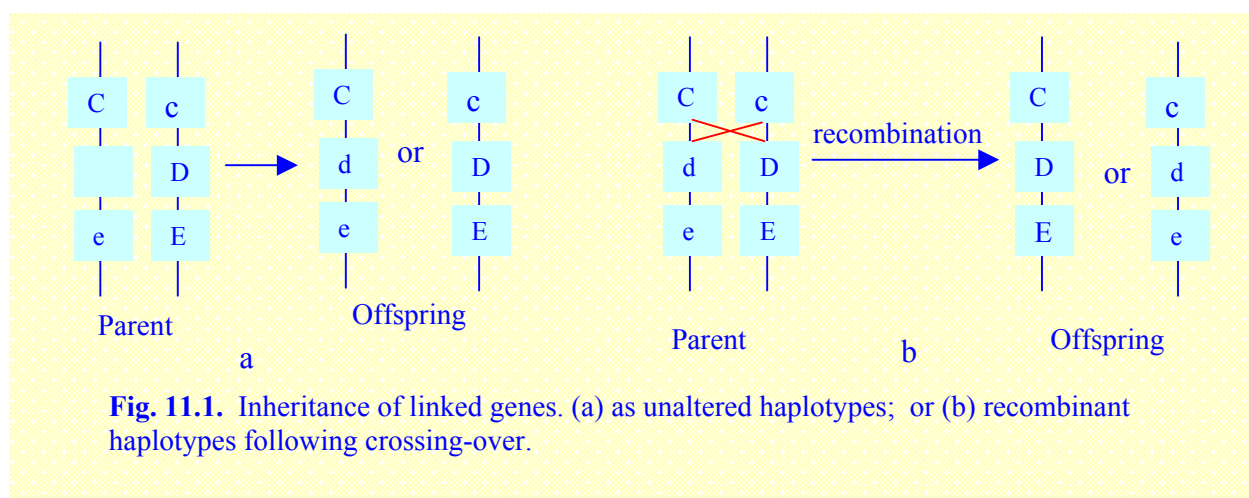
One of the conclusions drawn from Mendel's original experiments was the "law of independent assortment" which states that genes are transmitted from parents to offspring independently of one another. If a person has blood group A (e.g. genotype AO) and brown eyes (e.g. genotype Bb, where B is the allele for brown and b is the allele for blue eyes), the AO alleles are transmitted to the offspring independently of the Bb alleles.

However, not all genes are inherited independently of one another. Genes that are located on the same chromosome and are described as **linked genes**. If each chromosome were to be transmitted from parent to offspring as a whole and unaltered structure, it would be expected that all the genes located on the same chromosome would be transmitted together as a block and not independently of one another as proposed in Mendel's law. However, linked genes are not always transmitted *en bloc* because of the phenomenon of **recombination**. One of the fundamental events that occur in meiosis is **crossing over** in which homologous chromosomes exchange segments causing a reshuffling of genes. If genes are far apart on the same chromosome, it is likely that recombination occurs. Conversely, if they are very close together, they are more likely to be transmitted as a block.

An example of genetic linkage is given by the Rh blood group system in which the three loci are closely linked i.e. they are situated very close to one another on the same chromosome. At each locus there are three possible genotypes:

C locus:	C C	C c	c c
D locus:	D D	D d	d d
E locus:	E E	E e	e e

Any combination of genotypes can occur. The particular combination present in a given individual is called a **haplotype**. Let us assume that the haplotype of a particular parent is Cc, dD, eE as shown in Fig.11.1. Moreover, in this particular case, the alleles C,d, e are on one chromosome while the alleles c,D,E are on the other chromosome. If these genes were transmitted as two unbroken chains, we would expect that the offspring would inherit either Cde or cDE as in the parent. If the offspring show a different combination e.g. CDE, or cde this would indicate that recombination would have taken place as a result of crossing over between the C and the D loci.



**Fig. 11.1.** Inheritance of linked genes. (a) as unaltered haplotypes; or (b) recombinant haplotypes following crossing-over.

## GENETIC DISTANCE

Crossing over is a random event that may happen anywhere along the chromosome. If we were to consider a long segment of chromosome having loci labelled sequentially from A to Z, it is obvious that genes A and Z are wide apart whereas genes J and K are very close together. It is much more likely that crossing over happens somewhere between A and Z than exactly between J and K. Recombination is much more likely between genes A and Z that are far apart than between genes J and K, which are very close to one another. The frequency of recombination of two genes is proportional to the distance between them. The recombination frequency can be measured in families in which the genotypes of all individuals are known. The frequency with which recombination occurs in the offspring is expressed as a percentage. Genes which are very close together (closely linked) will have a very small recombination frequency (e.g. 1%). A recombination frequency of 1% means that only one out of 100 offspring was the combination of two genes different from that in their parents. In contrast, genes that are very far apart on the same chromosome or those that are on different chromosomes are equally likely to be transmitted together or separately and so would have a recombinant frequency of 50%.

This leads to the conclusion that **the frequency of recombination is directly proportional to the distance between two genes** - the smaller the distance, the smaller is the frequency of recombination. Thomas Hunt Morgan, a leading geneticist of the early twentieth century, first presented this theory after careful and painstaking experiments observing the frequencies of inheritance of combinations of characteristics in the fruit fly *Drosophila*. In recognition of his outstanding work, the unit for measuring genetic distance has been designated as the **centi-Morgan** which is defined as **the distance between two genes in which recombination occurs with a frequency of 1%.**

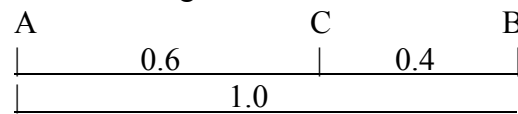
Two genes are linked if they show a recombinant frequency of less than 50%. However, if linked genes are far apart on a chromosome, it is possible that crossing-over occurs one or more times within that distance. This introduces errors in the estimation of gene distance. Hence **measures of genetic distance using recombination frequencies are accurate only if the genes are closely linked i.e. if the gene distance is small.** The unit of gene distance is also called a map unit. One map unit is equal to one centi-Morgan

## CONSTRUCTING A GENE MAP

Studies of genetic linkage and recombination frequencies have been used to create gene maps. This has been done extensively in *Drosophila*. This organism is very suitable for such study because it has very prominent, easily studied characteristics, has a short life cycle and produces hundreds of offspring. The following example illustrates how this is done.

In an experiment investigating two characteristics A and B it was found that their recombinant frequency was 1.0%. In further experiments it was found that the recombinant frequency for characteristics A and C was 0.6% and the recombinant frequency between B and C was 0.4%.

A genetic map of the three genes responsible for these characteristics may be constructed as follows. The values are in centimorgans.

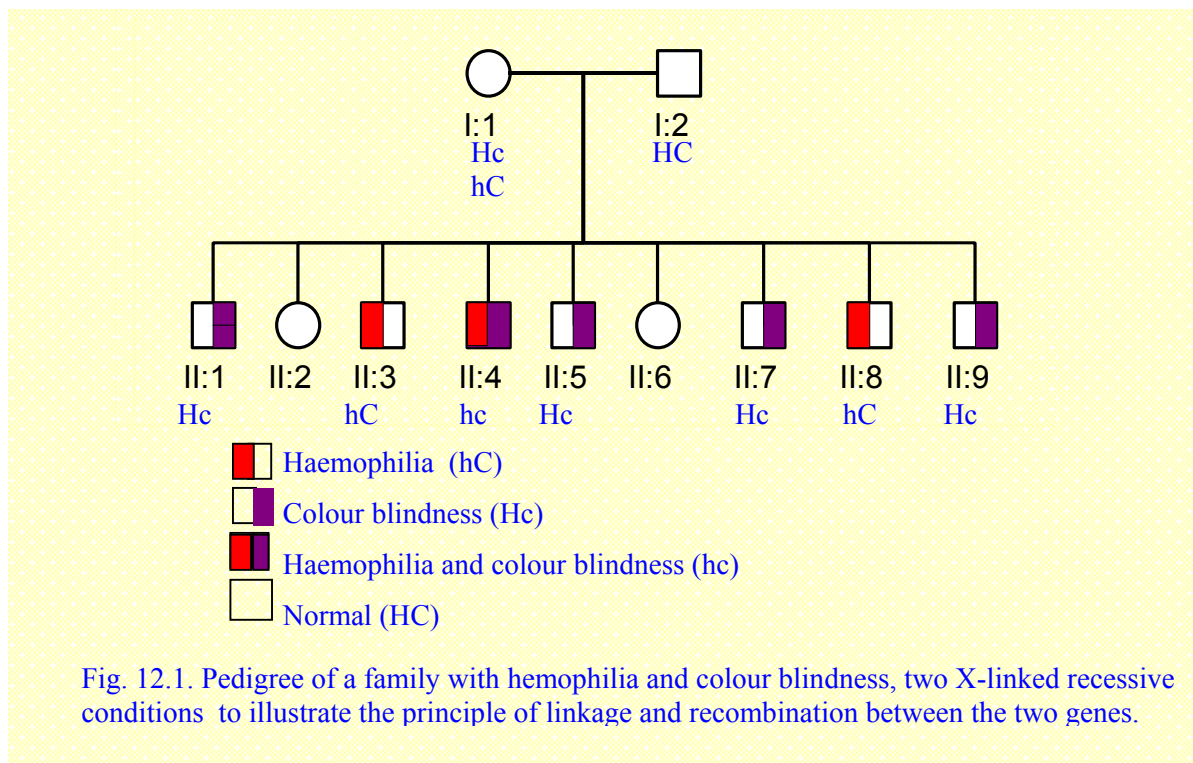


Further recombinant studies can be performed to estimate the gene distances between gene C and other genes D and E, and then between B and other genes, E, F, G and so on. In this way a larger and more detailed map is gradually constructed. This approach has been used to construct an extensive and detailed gene map of *Drosophila*.

## CONSTRUCTING A HUMAN GENE MAP

In humans it is not possible to perform recombinant experiments. Most disease-causing genetic traits are rare, the generation time in humans extends over decades and the number of offspring per family is very small. In the past genetic linkage and recombination frequencies were determined for as few diseases using accumulated information derived from studies on families in which there were two observable traits. For example, it was found from studies of families in which individuals had red-green colour blindness and haemophilia that the genes for these two conditions were linked. Furthermore, observations on recombinant frequencies have shown that the distance between these two genes is about 8 centi-Morgans.

**Problem:** Fig.12.1 shows an informative pedigree of a family with haemophilia and colour blindness. The gene for haemophilia is indicated as 'h' and that for colour blindness as 'c' while the normal corresponding genes are 'H' and 'C' respectively. Both these genes are X-linked and so males are hemizygous.



Note that six of the seven boys had either colour blindness or haemophilia. Boys inherit their X chromosome from their mother. This implies that, in this family, one X chromosome has the genes for haemophilia while the other has the gene for colour blindness. The six boys inherited either one or the other. One of the 7 boys had both haemophilia and colour blindness suggesting that there must have been a crossover between the two maternal X chromosomes in this case. Therefore, the recombinant frequency in this family was  $\frac{1}{7} = 0.143$ . Of course the small number gives a rather inaccurate estimate. Data from a number of families can be pooled together to give more accurate figures.

From the above data one can also work out the haplotypes:

Boys with colour blindness have Hc; Boys with haemophilia have hC; Boys with both haemophilia and colour blindness have hc; Normal males HC.

The mother was obviously the carrier for both conditions, a double heterozygote. Her haplotype must have been hC ; H c.

The two daughters (II:2 and II:6) were both normal. They must have inherited one X chromosome from their father (HC); their maternally derived X chromosome could be either Hc or hC. However, they could also have received the recombinant genotypes ch or CH, which would occur with a frequency of 0.143 (14.3%), assuming that this is a reliable estimate.

It will be noted that X-linked traits have been more useful in gene mapping than autosomal traits, because hemizygoty of X-linked genes in males makes it easier to work out the haplotypes. It is much more difficult to find informative pedigrees for autosomal traits.

Although such studies were valuable in early human gene mapping, informative pedigrees are very rare. However, genetic linkage studies using polymorphic markers have been used extensively for gene mapping. These are described below.

## **RESTRICTION FRAGMENT LENGTH POLYMORPHISMS**

Markers have been used extensively in human gene mapping. Earlier on it was noted that a disease-causing gene could be mapped by linkage and recombination studies with other known genes. However, informative families for such studies are rare. Genes can be mapped by linkage studies with polymorphic markers, which are nucleotide sequences identifiable at specific sites along the genome. Numerous markers have been identified throughout the genome using restriction endonucleases and so it is possible to construct maps of disease genes in relation to closely linked markers.

**Restriction endonucleases** are naturally occurring enzymes produced by bacteria as a defence against invasion by viruses. The bacterial endonucleases cut the viral DNA thus restricting its further proliferation. A particular restriction endonuclease recognises a specific nucleotide sequences in DNA and cleave it. For example the restriction enzyme Hind III recognises the following DNA sequence and cuts it open as shown:



As this sequence occurs by chance at several sites along the human genome. If a segment of DNA is exposed to Hind III, this will cut the DNA into several fragments of various sizes. These can be sorted out by electrophoresis according to the length of the fragments. This will form a pattern of fragments, identified by the length of each fragment as shown in Fig. 12.2. In this example, the specific sequence occurs at the sites indicated as A, B, C, D, E and F and exposure to the corresponding restriction enzyme will generate five fragments of size 16, 4, 1, 2 and 8 units respectively.

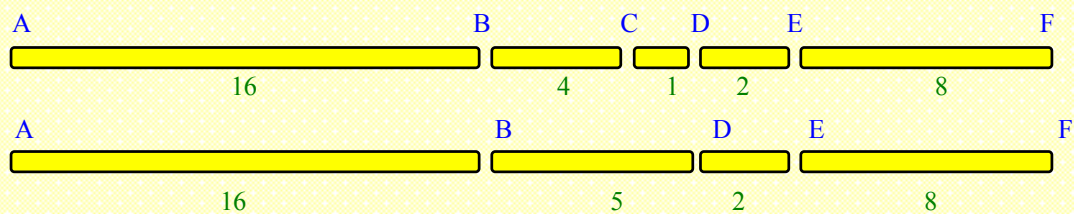


Fig. 12.2. An example of Restriction Fragment Length Polymorphism generated by Hind III. A,B,C,D,E and F indicate the sites where DNA is cleaved. The second individual lacks the restriction site at C and gives a different pattern of fragments from individual 1.

If the sequence were missing at site C, there would be 4 fragments of lengths 16, 5, 2 and 8 units. This variation is referred to as a **restriction fragment length polymorphism (RFLP)**. Using a large number of restriction endonucleases, it is likely that one finds one or more RFLPs close to the gene of interest. Such RFLPs are then used as markers for linkage studies with known genes. Linkage studies have been one of the most important tools for gene mapping.

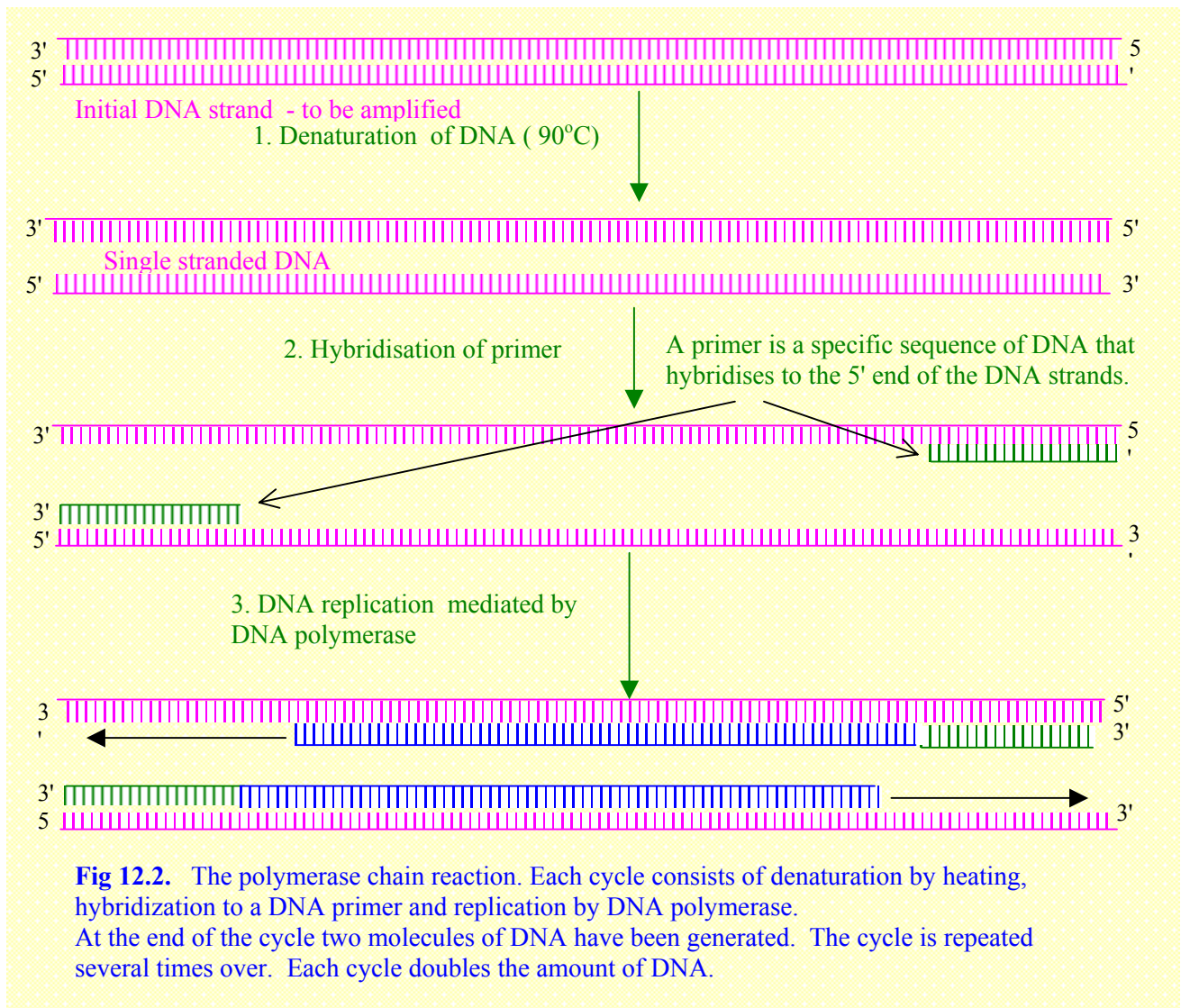
Although the gene causing a particular trait may not be known it is possible to identify markers which are very closely linked to it. Hundreds of markers might have to be screened before the right one is found. The polymorphic marker is then used to identify its location on a chromosome. It can also be identified in several family members and used as a means of tracking the gene. It is presumed that the gene is present where its linked marker is identified. The error of this method is equal to the recombination frequency which would be very small if the genes are very closely linked. If more than one marker is used the accuracy of the procedure is further increased.



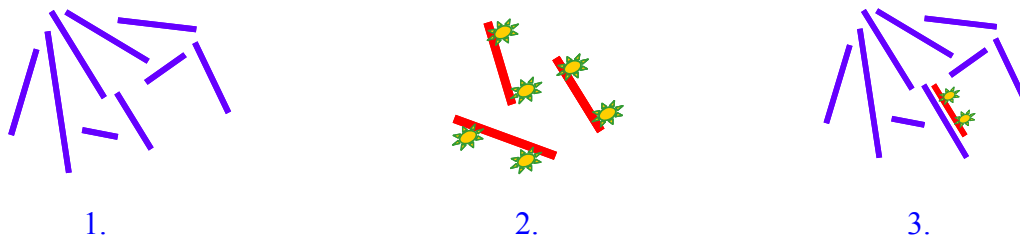
## THE POLYMERASE CHAIN REACTION - PCR

This is one of the most important tools in molecular biology and gene mapping. It is used mainly for amplification of the genes. This is one of the most important tools in modern genetics because it can amplify one gene to make millions of copies. Detection of the gene is thus facilitated. The main steps in the PCR are shown in figure 12.2

At the end of each cycle the DNA has doubled to yield - 2, 4, 8, 16, 32, 64, 128, 256 ....etc. copies of the same gene. After 20 to 25 cycles it would have amplified several million times.



## MOLECULAR HYBRIDIZATION IS USED TO DETECT SPECIFIC GENES.



1. Single-stranded DNA is generated
2. A probe is a known sequence of part of gene to be identified tagged with a radioactive label. Specific probes are synthesised in the laboratory..
3. The probe hybridizes only to the fragment with the corresponding sequence. This is detected by the label , which gives a fluorescent signal.

## FLUORESCENCE IN-SITU HYBRIDISATION (FISH)

This technique provides the most direct method of visualising genes on chromosomes. A DNA probe for a particular gene is a complementary sequence for part of the gene. This is usually labelled with a fluorescent dye. If the DNA in a chromosome preparation is first denatured, the probe hybridises specifically with the corresponding gene on the chromosome. The site of hybridisation is visualised under a fluorescence microscope. The site of the gene appears as a bright spot.

### FISH

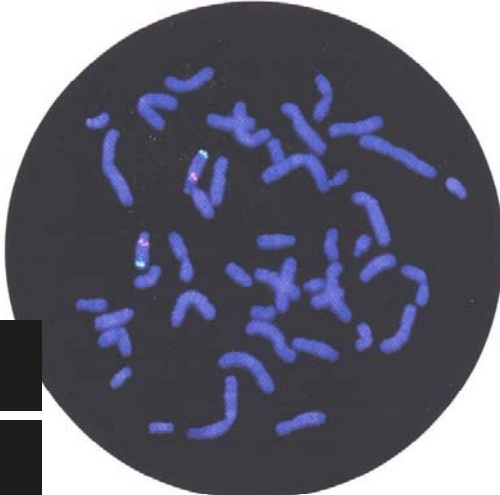
#### Fluorescence In-Situ Hybridisation

For the detection of single genes directly on the chromosomes

- Chromosomes preparations
- Heat denaturation
- Hybridise to specific probe with fluorescent label

Red spot: Gene for Prader-Willi syndrome on chrom 15

Green spot: telomere for chromosome 15 (control)

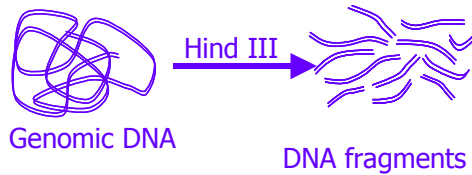


The diagram illustrates the FISH technique. It shows a circular field of view containing numerous chromosomes, which appear as blue, thread-like structures. Two specific spots are highlighted: a red spot and a green spot. The red spot is located on one of the chromosomes, indicating the presence of the gene for Prader-Willi syndrome on chromosome 15. The green spot is located on the same chromosome, indicating the presence of the telomere for chromosome 15, which serves as a control.

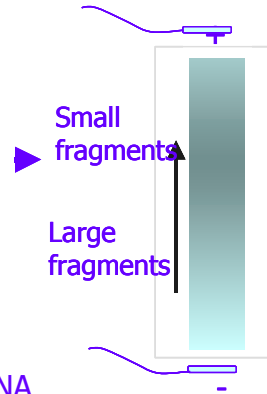


## Southern Blotting - A technique for detecting a particular gene

### 1. Cut DNA into fragments

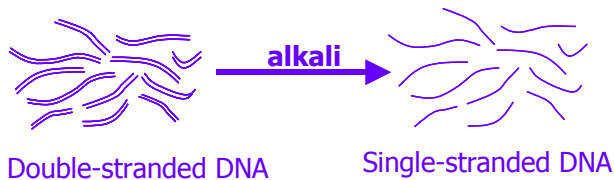


Separation of fragments by electrophoresis on agarose gel  
– bands not visualised. DNA fragments migrate to the positive (+) pole – small fragments move faster.



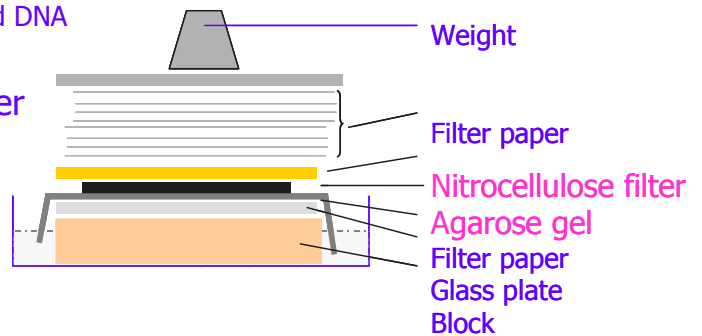
### 2. Denaturation of DNA.

The agarose gel is treated with alkali to denature the DNA



### 3. Blotting the DNA on to filter paper

The single-stranded DNA is transferred from the agarose gel on to a nitrocellulose filter.



## PROBLEMS

1. In a large family of eight children, two boys have haemophilia, three other boys are colour blind while one other boy and two girls are normal. Both parents are normal. If the gene for haemophilia is indicated as 'h', the gene for colour blindness as 'c' and the normal genes as 'H' and 'C' respectively,

- write the genotypes of the boys;
- write the genotypes of the parents;
- in which child has recombination occurred?
- from the available data, estimate the frequency of recombination;
- what are the possible genotypes in the girls assuming that there is no recombination?

2. Using somatic cell hybridisation for localising the gene for the enzyme  $\alpha$ -glucosidase the only cells which expressed this enzyme had the following human chromosomes:

A: 4, 11, 17;    B: 2, 9, 13, 17;    C: 6, 17, 21;    D: 4, 11, 13, 17, 22;    E: 4, 13, 17

On which chromosome is the gene for  $\alpha$ -glucosidase located?

3. Linkage analysis was performed on families with a genetic disease D and three polymorphic markers, N, O, P. The recombinant frequencies were as follows:

D and N 0.1%;	O and N 0.5%;
D and O 0.6%;	O and P 1.0%.
D and P 0.4 %;	

Construct a gene map for the disease locus and the three polymorphic markers, marking out the genetic distances in centi-Morgans. What would be the recombinant frequency between N and P?

4. What is a gene probe? What steps would be required in order to hybridise a specific gene probe to DNA? How can this be achieved ?

\*\*\*\*\*