Selection, Screening and Analysis of Recombinants:

- Various techniques can be used to identify cloned genes.
- Success in any cloning experiment depends on being able to identify the desired gene sequence among the many different recombinants that may be produced.
- Given that a large genomic library may contain a million or more cloned sequences, which are not readily distinguishable from each other by simple analytical methods.
- It is clear that identification of the target gene is potentially the most difficult part of the cloning process.
- Fortunately there are several selection / identification methods that can be used to overcome most of the problems that arise.
- There are two terms that require definition before we proceed, these being selection and screening.
- Selection: is where some sort of pressure, like presence of an antibiotic, is applied during the growth of host cells containing recombinant DNA.
- The cells with the desired characteristic are therefore selected by their ability to survive.
- This approach ranges in sophistication, from simple selection for the presence of a vector, up to direct selection of cloned genes by complementation of defined mutations.
Screening: is a procedure by which a population of viable cells is subjected to some sort of analysis that enable the desired sequences to be identified.

Because only a small portion of the large number of bacterial colonies or bacteriophage plaques being screened will contain the DNA sequence(s) of interest, screening requires methods that are highly sensitive and specific.

In practice, both selection and screening methods may be required in any single experiment, and may even be used at the same time if the procedure is designed carefully.

Genetic Selection and Screening Methods:

- Genetic selection and screening methods rely on the expression or non-expression of certain traits.
- Usually, these traits are encoded by the vector or perhaps by the desired clone sequence if a direct selection method is available.
- One of the simplest genetic selection methods involves the use of antibiotics to select for the presence of vector molecules. For example:
  - The plasmid pBR 322 contains genes for ampicillin resistance (Ap<sup>r</sup>) and tetracycline resistance (Tc<sup>r</sup>).

Fig 5.1 Nicholl 3<sup>rd</sup> Ed
Fig. 5.1  Map of plasmid pBR322.
Important regions indicated are the genes for ampicillin and tetracycline resistance (Ap$^r$ and Tc$^r$) and the origin of replication (ori). Some unique restriction sites are given. The hatched region shows the two fragments that were removed from pBR322 to generate pAT153.
Thus the presence of the plasmid in cells can be detected by plating/growing potential transformants on an agar medium/selective medium that contain either or both of these antibiotics.

Only cells that have taken up the plasmid i.e. transformed cells will be resistant and these cells will therefore grow in the presence of the antibiotic.

The technique can also be used to identify mammalian cells containing vectors with selectable markers.

**Fig 7.6 Peters**

**Use of Chromogenic Substrates:**

- The most popular system uses the compound X-gal (5-bromo-4-chloro-3-indolyl-β-D galactopyranoside), which is a colorless substrate for β-galactosidase (*LacZ*).

- The enzyme is normally synthesized by *E. coli* cells when lactose becomes available.

- However, induction can also occur if a lactose analogue such as IPTG (isopropyl thiogalactoside) is used.

- This has the advantage of being an inducer without being substrate for β-galactosidase.
Figure 7.6
Selective medium is used to identify a subset of cells in a heterogeneous population. In this illustration $tet^+$ host cells are transformed with a $tet^+$ plasmid. The resulting population of treated host cells contains both successful transformants and cells that did not take up and retain the $tet^+$ plasmid. All treated cells are grown in a growth medium containing tetracycline. This medium selects for successfully transformed host cells by preventing the growth of unsuccessful transformants.
- On cleavage of X-gal, a blue colored product is formed, thus the expression of the lacZ gene can be detected easily.

- This system can be used as a screening method for cells or plaques or as a system for the detection of tissue specific gene expression in transgenics.

**Fig 8.2 Peters**

**Fig 8.3 Peters**

**Fig 8.1 Nicholl 3rd Ed**

- The X-gal system can be used where a functional β-galactosidase gene is present in the host/vector system.

- This can occur in two ways:
  - Firstly, an intact galactosidase gene (lacZ) may be present in the vector.

**Fig 5.8 Nicholl 3rd Ed**

- Host cells that are Lac- are used for propagation of phage so that the Lac+ phenotype will only arise when the vector is present.
Figure 8.2
During digestion, the milk sugar lactose is broken down into glucose and galactose by the enzyme beta-galactosidase.
Figure 8.3
The addition of lactose to bacterial cultures results in the almost immediate production of beta-galactosidase and other enzymes involved in lactose metabolism.
Fig. 8.1 Structure of X-gal and cleavage by β-galactosidase. The colourless compound X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) is cleaved by β-galactosidase to give galactose and an indoxyl derivative. This derivative is in turn oxidised in air to generate the dibromo-dichloro derivative, which is blue.
Fig. 5.8 Bacteriophage λ insertion vectors λgt10 and Charon 16A. The cl and lacZ genes in λgt10 and Charon 16A, respectively, are shaded. Within these genes there is an EcoRI site for cloning into. The lengths of the left and right arms (LA and RA, in kb) are given. The size of the wild-type λ genome is marked on the scale bar as λ+. Redrawn from Winnacker (1987), From Genes to Clones, VCH. Reproduced with permission.
A second approach is to employ the $\alpha$-complementation system, in which part of the \textit{lacZ} gene is carried by the vector and the remaining part is carried by the host cell.

The smaller, vector encoded peptide fragments known as the $\alpha$-peptide and the region coding for this is designated \textit{lacZ}'.

Host cells are therefore designated $\textit{lacZ}'$.

Blue colonies or plaques will only be produced when the host and vector fragments complement each other to produce functional $\beta$-galactosidase.

**Fig 5.2 Nicholl 3rd Ed**

**Fig 5.10 Nicholl 3rd Ed**

- Insertional Inactivation:
  - The presence of cloned DNA fragments can be detected if the insert interrupts the coding sequence of a gene.
  - This approach is known as insertional inactivation and can be used with any suitable genetic system.
  - Three systems will be described to illustrate the use of the technique.
Fig. 5.2 Map of plasmid pUC18. (a) The physical map, with the positions of the origin of replication (ori) and the ampicillin resistance gene (Ap') indicated. The lacI gene (lac repressor), multiple cloning site (MCS) or polylinker, and the lacZ' gene (α-peptide fragment of β-galactosidase) are also shown. (b) The polylinker region. This has multiple restriction sites immediately downstream from the lac promoter (P_{lac}). The in-frame insert used to create the MCS is hatched. Plasmid pUC19 is identical with pUC18 apart from the orientation of the polylinker region, which is reversed.
Fig. 5.10 Map of the filamentous phage vector M13mp18. The double-stranded replicative form is shown. The polylinker region (MCS) is the same as that found in plasmid pUC18 (Fig. 5.2). Genes in the REP region encode proteins that are important for DNA replication. The CAP and MOR regions contain genes that specify functions associated with capsid formation and phage morphogenesis, respectively. The vector M13mp19 is identical except for the orientation of the polylinker region.
i. Antibiotic resistance can be used as an insertional inactivation system if DNA fragments are cloned into a restriction site within an antibiotic resistance gene

**Fig. 5.1 Nicholl 3rd Ed**

ii. X-gall system can also be used if a DNA fragment is cloned into a functional β-galactosidase gene

**Fig 5.8 Nicholl 3rd Ed**

This approach can also be used with the α-complementation system

**Fig 8.2 Nicholl 3rd Ed**

iii. Plaque morphology can also be used for certain λ vector such as λgt10, which contain the cI gene

This gene encodes the cI repressor, which is responsible for the formation of lysogens

Plaques derived from cI+ vectors will be slightly turbid, due to survival of some cells that have become lysogens
Fig. 5.1  Map of plasmid pBR322.

Important regions indicated are the genes for ampicillin and tetracycline resistance ($Ape^r$ and $Tce^r$) and the origin of replication ($ori$). Some unique restriction sites are given. The hatched region shows the two fragments that were removed from pBR322 to generate pAT153.
**Fig. 5.8** Bacteriophage \( \lambda \) insertion vectors \( \lambda_{gt10} \) and Charon 16A. The \( cl \) and \( lacZ \) genes in \( \lambda_{gt10} \) and Charon 16A, respectively, are shaded. Within these genes there is an EcoRI site for cloning into. The lengths of the left and right arms (LA and RA, in kb) are given. The size of the wild-type \( \lambda \) genome is marked on the scale bar as \( \lambda^+ \). Redrawn from Winnacker (1987), *From Genes to Clones*, VCH. Reproduced with permission.
Fig. 8.2 Insertional inactivation in the α-complementation system. (a) The chromosome (CS) has a defective lacZ gene that does not encode the N-terminal α-peptide of β-galactosidase (specified by the lacZ'− gene fragment). Thus, the product of the chromosomal lacZ region is an enzyme lacking the α-peptide (α−, hatched). If a non-recombinant pUC plasmid or M13 phage is present in the cell, the lacZ' gene fragment encodes the α-peptide, which enables functional β-galactosidase to be produced. In the presence of X-gal, blue colonies or plaques will appear. If a DNA fragment is cloned into the vector, as shown in (b), the lacZ' gene is inactivated and no complementation occurs; thus, colonies or plaques will not appear blue.
If the *ci* gene is inactivated by cloning a fragment into a restriction site within the gene, the plaques are clear and can be distinguished from the turbid non-recombinants.

**Fig 5.8 Nicholl 3rd Ed**

- **Screening:**
  - **Table 6.2 Nicholl 3rd Ed**
  - **DNA Hybridization:**
    - The power of nucleic acid hybridization lies in the fact that complementary sequences will bind to each other with a very high degree of fidelity.
    - In practice this depends on the degree of homology between the hybridizing sequences with the help of a probe.
    - There are three possible sources of probes:
      - **Cloned DNA** from a closely related organism, conditions of hybridization reaction can be adjusted to permit considerable mismatch between the probe and target DNA to compensate for the natural difference between two sequences.
      - **cDNA** probe from abundant mRNA.
Fig. 5.8  Bacteriophage λ insertion vectors λgt10 and Charon 16A. The cl and lacZ genes in λgt10 and Charon 16A, respectively, are shaded. Within these genes there is an EcoRI site for cloning into. The lengths of the left and right arms (LA and RA, in kb) are given. The size of the wild-type λ genome is marked on the scale bar as λ⁺. Redrawn from Winnacker (1987), From Genes to Clones, VCH. Reproduced with permission.
<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome size (kb)</th>
<th>20 kb inserts</th>
<th>45 kb inserts</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> (bacterium)</td>
<td>$4.0 \times 10^3$</td>
<td>$6.0 \times 10^2$</td>
<td>$2.7 \times 10^2$</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> (yeast)</td>
<td>$1.4 \times 10^4$</td>
<td>$2.1 \times 10^3$</td>
<td>$9.3 \times 10^2$</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em> (simple higher plant)</td>
<td>$7.0 \times 10^4$</td>
<td>$1.1 \times 10^4$</td>
<td>$4.7 \times 10^3$</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em> (fruit fly)</td>
<td>$1.7 \times 10^5$</td>
<td>$2.5 \times 10^4$</td>
<td>$1.1 \times 10^4$</td>
</tr>
<tr>
<td><em>Stronglyocentrotus purpuratus</em> (sea urchin)</td>
<td>$8.6 \times 10^5$</td>
<td>$1.3 \times 10^5$</td>
<td>$5.7 \times 10^4$</td>
</tr>
<tr>
<td><em>Homo sapiens</em> (human)</td>
<td>$3.0 \times 10^6$</td>
<td>$4.5 \times 10^5$</td>
<td>$2.0 \times 10^5$</td>
</tr>
<tr>
<td><em>Triticum aestivum</em> (hexaploid wheat)</td>
<td>$1.7 \times 10^7$</td>
<td>$2.5 \times 10^6$</td>
<td>$1.1 \times 10^6$</td>
</tr>
</tbody>
</table>

Note: The number of clones ($N$) required for a probability ($P$) of 95% that a given sequence is represented in a genomic library is shown for a range of different organisms. Approximate genome sizes of the organisms are given (haploid genome size, if appropriate). Two values of $N$ are shown, for 20 kb inserts ($\lambda$ replacement vector size) and 45 kb inserts (cosmid vectors). The values should be considered as minimum estimates, as strictly speaking the calculation assumes (1) that the genome size is known accurately, (2) that the DNA is fragmented in a totally random manner for cloning, (3) that each recombinant DNA molecule will give rise to a single clone, (4) that the efficiency of cloning is the same for all fragments, and (5) that diploid organisms are homozygous for all loci. These assumptions are usually not all valid for a given experiment.
Chemically synthesized probe - nucleotide sequence of a synthetic probe is based on probable nucleotide sequence that is deduced form the known amino acid sequence of the protein encoded by the target gene.

Fig 15.21 Klug 5th Ed

- Colony Hybridization/Plaque Hybridization:
  - Fig 27.9 Zubay
  - Fig 7.18 Lodish 3rd Ed
  - Fig 8.5 Nicholl 3rd Ed
  - Fig 28.71 Voet

- Immunological Assay:
  - Can be done for Expression Library
  - Fig 7.21 Lodish 3rd Ed
  - Fig 4.18 Glick 3rd Ed

- Protein Activity:
  - Can be done for Expression Library
  - Ligand – Receptor
  - Fig 20.9 Lodish 3rd Ed
FIGURE 15.21 The process of reverse translation. The coding DNA sequence can be deduced from the amino acid sequence of a small portion of a protein. In this example, two amino acids (trp and met) have only one codon each. Three others (lys, his, and tyr) are encoded by two codons. The sixth amino acid (gly) is encoded by codons that include all base combinations at the third position. For this protein fragment, eight coding sequences encompass all the possible combinations. The exact coding sequence of the gene must be one of the eight. Using a radioactive mixture of these sequences as probes, a cloned library can be screened to isolate the structural gene for the complete protein.
Figure 27.9

Colony hybridization procedure used to identify bacterial clones harboring a plasmid containing a specific DNA. Step 1: Replica-plate the colonies containing plasmids onto nitrocellulose paper. Step 2: Lyse cells with NaOH and fix denatured DNA to paper. Step 3: Hybridize to $^{32}$P-labeled DNA carrying the desired sequence and autoradiograph the product. Locations of desired DNA should be emphasized in the autoradiograph. Clones carrying desired plasmids (circled) may then be isolated from a corresponding agar replica plate carrying untreated colonies.
Individual phage plaques

Master plate of \( \lambda \)-phage plaques on \( E. \) coli lawn

Place nitrocellulose filter on plate to pick up phages from each plaque

Nitrocellulose filter

Incubate filter in alkaline solution to lyse phages and denature released DNA

Single-stranded phage DNA bound to filter

Hybridize with labeled probe; perform autoradiography

Signal appears over phage DNA that is complementary to probe

FIGURE 7-18 Identification of a specific clone from a \( \lambda \)-phage library by membrane hybridization to a radiolabeled probe. The position of the signal on the autoradiogram identifies the desired plaque on the plate. In practice, in the initial plating of a library the plaques are not allowed to develop to a visible size so that up to 50,000 recombinants can be analyzed on a single plate. Phage particles from the identified region of the plate are isolated and replated at low density so that the plaques are well separated. Then pure isolates can be obtained by repeating the plaque hybridization as shown in the figure.
Fig. 8.5 Screening plaques at high and low densities. A radiolabelled probe was used to screen a genomic library in the \( \lambda \) vector EMBL3. (a) Initial screening was at a high density of plaques, which identified two positive plaques on this plate. The boxed area shows a false positive. (b) The plaques were picked from the positive areas and re-screened at a lower density to enable isolation of individual plaques. Many more positives are obtained because of the high proportion of 'target' plaques in the re-screened sample. Photograph courtesy of Dr M. Stronach.
FIGURE 28-71. Chromosome walking. A DNA segment too large to sequence in one piece is fragmented and cloned. A clone is picked and the DNA insert it contains is sequenced. A small fragment of the insert near one end is subcloned (cloned from a clone) and used as a probe to select a clone containing an overlapping insert, which, in turn, is sequenced. The process is repeated so as to “walk” down the chromosome. Chromosome walking can, of course, extend in both directions.
FIGURE 7-21 Use of λ expression cloning to identify a cloned DNA based on binding of the encoded protein to a specific antibody. The λgt11 vector was engineered to express the E. coli protein β-galactosidase at high levels. The only EcoRI recognition site (red) in this vector lies near the 3’ end of the β-galactosidase gene. If a cDNA (green), or protein-coding fragment of genomic DNA, is inserted into this EcoRI site in the correct orientation and proper reading frame, it will be expressed as a fusion protein in which most of the β-galactosidase sequence is at the N-terminal end and the protein sequence encoded by the inserted DNA is at the C-terminal end. Plaques resulting from infection with recombinant λgt11 contain high concentrations of such fusion proteins. These proteins can be transferred and bound to a replica filter, which then is incubated with a monoclonal primary antibody (blue) that recognizes the protein of interest. Rinsing the filter washes away antibody molecules that are not bound to the specific fusion protein attached to the filter. Bound antibody usually is detected by incubating the filter with a second radiolabeled antibody (dark red) that binds to the primary antibody. Any signals that appear on the autoradiogram are used to locate plaques on the master plate containing the gene of interest. [Adapted from J. D. Watson et al., 1992, Recombinant DNA, 2d ed., Scientific American Books.]
Figure 4.18. Immunological screening of a gene library (colony immunoassay). Cells from the transformation reaction are plated onto solid agar medium under conditions that permit transformed but not nontransformed cells to grow. (1) From the discrete colonies formed on this master plate, a sample from each colony is transferred to a solid matrix such as a nitrocellulose or nylon membrane. (2) The cells on the matrix are lysed, and their proteins are bound to the matrix. (3) The matrix is treated with a primary antibody that binds only to the target protein. (4) Unbound primary antibody is washed away, and the matrix is treated with a secondary antibody that binds only to the primary antibody. (5) Any unbound secondary antibody is washed away, and a colorimetric reaction, which can occur only if the secondary antibody is present, is carried out. (6) A colony on the master plate that corresponds to a positive response on the matrix is identified. Cells from the positive colony on the master plate are subcultured because they may carry the plasmid-insert DNA construct that encodes the protein that binds the primary antibody.
Plasmid expression vector

Double-stranded cDNA from cell normally synthesizing receptor protein

SV40 origin of DNA replication and promoter

Signals for termination and polyadenylation of mRNA

Plasmid DNA

 Signals for DNA replication in bacterial cells

Fig 20.9 Lodish
Fig 20.9 Lodish
(b) Rescreening of positive cDNA pools

Subdivide

Subpool 1

Negative

Subpool 2

Transfect plasmids
Add radioactive ligands
Autoradiography

Negative

Subpool 3

Positive

Subdivide and rescreen until a single clone is obtained

Fig 20.9 Lodish
ii. Enzyme – Substrate
- Can be done for Expression Library
  • 17β- hydroxysteroid dehydrogenase
    $^{14}$C-Androstendione $\rightarrow^{14}$C-Testosterone
  • Separation by TLC

*Fig 6 1999 Publication*

➢ Functional Complementation:
- Screening method that depend on the full biological activity of the protein
- This is often termed functional cloning
- It is possible in complete ignorance of the whereabouts of the gene in the genome and requires no prior knowledge of the nucleotide sequence of the clone or amino acid sequence of its product
- As long as the expressed protein is functional and that function can be exploited to screen an expression library, the corresponding clone can be identified
- 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
- If the mutant cells are non-viable under particular growth conditions, cell carrying the clone of interest can be positively selected, allowing the corresponding clones to be isolated
Fig. 6. A, Down-regulation of enzymatic activity of 17βHSD by hCG in rat Leydig cells. Leydig cells were obtained from male rats injected sc with vehicle (control) or 2.5 μg hCG and obtained at different times (1, 4, 12, and 24 h) after treatment. Intact cells (1 × 10⁶ cells) after preincubation with aminoglutethimide (100 μg) were incubated with 2 μM [14C]androstenedione (A) and 5 μM unlabeled A for 60 min in the presence of 5.8 mM glucose. The substrate [14C]A and product [14C]testosterone (T) were separated by TLC, and radioactivity from product and substrate were recorded by autoradiography for visual display (upper panel). TLCs were quantified by phosphorimager analysis (lower panel). Results represent the percent conversion to testosterone from androstenedione relative to the control value. A representative experiment (mean ± SE) of triplicates is shown. Three separate such experiments were performed. B, Comparative profiles of type III 17βHSD and 3βHSD enzyme activities after administration of a single dose of hCG (2.5 μg). Results for 3βHSD are from our previous study (16) and used for comparative purpose in the Discussion.
• Ratkin and Carbon (1977) provided an early example of how certain eukaryotic genes can be cloned on the basis of their ability to complement autotrophic mutations in *E. coli*

• These investigators inserted fragments of yeast DNA, obtained by mechanical shearing, into the plasmid ColE1, using a homo-polymer tailing procedure

• They transformed *E.coli* hisB mutants which were unable to synthesize histidine, with recombinant plasmids and plated the bacteria on minimal medium

• In this way, they selected for complementation of the mutation and isolated clones carrying an expressed yeast *his* gene

• If the function of the gene is highly conserved, it is quite possible to carry out functional cloning of, for example, mammalian proteins in bacteria and yeast

• Thus complementation in yeast has been used to isolate cDNA for a number of mammalian metabolic enzymes and certain highly conserved transcription factors as well as regulator of meiosis in plants

• Functional complementation is also possible in transgenic animals and plants

Fig 14.3 Glick 3rd Ed
Fig 14.8 Glick 3rd Ed
K. pneumoniae cells are treated with a dose of a mutagenic agent that allows ~0.1 to 1% of the cells to survive.

- Some of these mutagenized cells are able to grow on a minimal medium containing a source of fixed nitrogen such as NH$_4$Cl but do not grow in the absence of fixed nitrogen.

- These cells are likely to have a mutation in a nif genes and are designated Nif -
Figure 14.8 Procedure for the isolation of *R. meliloti* nodulation genes. The DNA from wild-type *R. meliloti* is cloned into the broad-host-range cosmid pLAFR1, packaged into bacteriophage λ, and introduced into *E. coli*. The clone bank is then transferred from *E. coli* to a Nod*+* *R. meliloti* strain by conjugation. Alfalfa plants are inoculated with transformed Nod*+* *R. meliloti*. Those plants that develop root nodules have been infected with Nod*+* *R. meliloti* cells that presumably carry a complementing nodulation gene inserted into the cosmid vector. The transformed Nod*+* *R. meliloti* cells can be isolated directly from the root nodules.
In this way Probst et al. (1998) were able to clone the mouse deafness associated gene *Shaker*-2 and from there its human homologue, *DFNB3*

**Fig 6.16 Primrose 6th Ed**

- **Analysis of Cloned Gene:**
  - Once clones have been identified by using one of the mentioned techniques, more detailed characterization of DNA will now be begin
  - There are many ways to tackle this and
  - The choice of approach will depend on
  - What is already know about the gene in question? And
  - What are the ultimate aims of the experiment?
- **Characterization Based on mRNA Translation *in vitro***:
  - In some cases the identification of a particular clone may require confirmation
  - This is particularly true when the plus/minus method of screening has been used, as the results of such a process are usually somewhat ambiguous
Fig. 6.16 Functional complementation in transgenic mice to isolate the Shaker-2 gene. Homozygous shaker-2 fertilized mouse eggs were injected with BAC clones derived from the Shaker-2 candidate region of a wild-type mouse. Progeny were screened for restoration of the wild-type phenotype, thus identifying the BAC clone corresponding to the Shaker-2 gene. This clone is then sequenced and used to isolate and map the corresponding human disease gene DFNB3.
If the desired sequence codes for a protein, and the protein has been characterized, it is possible to identify the protein product by two methods based on translation of mRNA *in vitro*

i. Hybrid arrest translation (HART)

ii. Hybrid release translation (HRT)

**Fig 8.8 Nicholl 3rd Ed**

**Restriction Mapping:**

- Obtaining a restriction map for cloned fragments is usually essential before additional manipulation is carried out

- This is particularly important where phage or cosmid vectors have been used to clone large pieces of DNA

- If restriction map is available, smaller fragments can be isolated and used for various procedures including
  - Sub-cloning into other vector
  - Preparation of probe for chromosome walking and
  - DNA sequencing

**Table 4.2 Nicholl 3rd Ed**

**Fig 4.4 Nicholl 3rd Ed**
- Components of Translation and radioactive amino acid $^{35}$S Methionine
Table 4.2. Digestion of a 15 kbp DNA fragment with three restriction enzymes

<table>
<thead>
<tr>
<th></th>
<th>BamHI</th>
<th>EcoRI</th>
<th>PstI</th>
<th>BamHI + EcoRI</th>
<th>BamHI + PstI</th>
<th>EcoRI + PstI</th>
<th>BamHI + EcoRI + PstI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>14</td>
<td>12</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: Data shown are lengths (in kbp) of fragments that are produced on digestion of a 15 kbp DNA fragment with the enzymes BamHI, EcoRI, and PstI. Single, double, and triple digests were carried out as indicated. Fragments produced by each digest are listed in order of length.
Fig. 4.4 Restriction mapping. (a) The 15 kb fragment yields two fragments of 14 and 1 kb when cut with BamHI. (b) The EcoRI fragments of 12 and 3 kb can be orientated in two ways with respect to the BamHI site, as shown in (b)i and (b)ii. The BamHI/EcoRI double digest gives fragments of 11, 3, and 1 kb, and therefore the relative positions of the BamHI and EcoRI sites are as shown in (c). Similar reasoning with the orientation of the 8 and 7 kb PstI fragments (d) gives the final map (e).
**Blotting Techniques:**

- Although a clone may have identified and
- Its restriction mapping has been determined
- This information itself does not provide much of insight into the fine structure of cloned fragment and the gene that it contains
- Ultimately the aim may be to obtain gene sequence
- But it not usually not sensible to begin sequencing straight away
- For example if a 20 kb fragment of genomic DNA has been cloned in a λ replacement vector and
- The area of interest is only 2kb in length, much effort would be wasted by sequencing the entire clone
- In many experiments it is therefore essential to determine which parts of the original clone contain the region of interest
- This can be done by using a variety of methods based on blotting nucleic acid molecules on to membranes and hybridization with specific probes
- Such approach is some ways an extension of clone identification by colony or plaque hybridization with the refinement that information about the structure of the clone is obtained

Fig 2.4 Primrose 6th Ed
Fig. 2.4 Overview of nucleic acid blotting and hybridization (reproduced courtesy of Amersham Pharmacia Biotech).
Southern Blotting:

- The first blotting technique was developed by Ed Southern and now called after his name.
- In this technique fragments of DNA, generated by restriction digestion, are subjected to agarose gel electrophoresis.
- The separated fragment are then transferred to nitrocellulose or nylon membrane by blotting technique.

Fig 3.5 Nicholl 3rd Ed
Fig 4.11 Brown 2nd Ed
Fig 8.10 Nicholl 3rd Ed

Northern Blotting:

- Use to detect particular RNA in a mixture of RNAs.
- Transcripts of mRNA and level of expression.
- Presence of gene in various tissues i.e. tissue distribution of mRNA.

Fig 2 1999 Publication

- Amount comparison of a particular mRNA in cells under different conditions.

Fig 7.32 Lodish 3rd Ed / Fig 7.33 Lodish 4th Ed
Fig. 3.5 A typical system used for agarose gel electrophoresis. The gel is just covered with buffer; therefore, the technique is sometimes called submerged agarose gel electrophoresis (SAGE). Nucleic acid samples placed in the gel will migrate towards the positive electrode as indicated by the horizontal arrow.
Figure 4.11  Separation of DNA molecules by agarose gel electrophoresis.

The range of fragment sizes that can be resolved depends on the concentration of agarose in the gel. Electrophoresis has been performed with three different concentrations of agarose. The labels indicate the sizes of bands in the left and right lanes. Photo courtesy of BioWhittaker Molecular Applications.
Fig. 8.10  Southern blotting. A hypothetical 20 kb fragment from a genomic clone is under investigation. A cDNA copy of the mRNA is available for use as a probe. (a) Gel pattern of fragments produced by digestion with various restriction enzymes; (b) autoradiograph resulting from the hybridisation. Lanes 1 and 6 contain \( \lambda \) HindIII markers, sizes as indicated. These have been marked on the autoradiograph for reference. The intact fragment (lane 2) runs as a single band to which the probe hybridises. Lanes 3, 4, and 5 were digested with EcoRI (E), PstI (P), and BamHI (B). Fragment sizes are indicated under each lane in (a). The results of the autoradiography show that the probe hybridises to two bands in the EcoRI and BamHI digests; therefore, the clone must have internal sites for these enzymes. The PstI digest shows hybridisation to the 7 kb fragment only. This might, therefore, be a good candidate for subcloning, as the gene may be located entirely on this fragment.
Fig. 2. Tissue distribution of 17βHSD type III mRNA. A. Polyadenylated RNA (10 μg) from several tissues of adult male and female rat was subjected to Northern blot analysis. The loading of the RNA was monitored by hybridization of β-actin. B. RT-PCR analysis in rat ovary and Leydig cells. RNA samples extracted from rat Leydig cells and from ovaries of rat treated with hormones (see Materials and Methods and diagram in this figure) were reverse transcribed as described in Materials and Methods. RT-PCR transcripts of Leydig cell RNA was used as a positive control.
**FIGURE 7-33** Northern blots of β-globin mRNA in extracts of erythroleukemia cells that are growing but uninduced (UN lane) and in cells that are induced to stop growing and allowed to differentiate for 48 hours or 96 hours. The density of a band is proportional to the amount of mRNA present. The β-globin mRNA is barely detectable in uninduced cells but increases more than 1000-fold by 96 hours after differentiation is induced. [Courtesy of L. Kole.]
✓ Dot Blotting:

• Nucleic acids are not subjected to electrophoresis but are spotted on to the filters then

• Hybridization can be carried out are for Southern and Northern blots

• This is useful in obtaining quantitative data in the study of gene expression.

**Fig 10.3, Nicholl 3rd Ed**

✓ Sequencing:

• This provides much useful information about coding sequences, control regions and other factures such as intervening sequences

✓ Nuclease Protection Assay:

• Another important method for detecting and quantifying specific RNA molecule in a mixture and mapping them employs endonucleases that digest single stranded but not double stranded nucleic acids

**Fig 7.32, Lodish 3rd Ed / Fig 7.34 Lodish 4th Ed**
Fig. 10.3  Dot-blot analysis of mRNA levels. Samples of total RNA from synchronous cell cultures of *Chlamydomonas reinhardtii* grown under batch culture and turbidostat (control) culture conditions were spotted onto a membrane filter. The filter was probed with a radiolabelled cDNA specific for an mRNA that is expressed under conditions of flagellar regeneration. (a) An autoradiograph was prepared after hybridisation. Batch conditions (i) show a periodic increase in transcript levels with a peak at 15 h. Control samples (ii) show constant levels. Data shown in (b) were obtained by counting the amount of radioactivity in each dot. This information can be used to determine the effect of culture conditions on the expression of the flagellar protein. Photograph courtesy of Dr J. Schloss. From Nicholl et al. (1988), *Journal of Cell Science* **89**, 397–403. Copyright (1988) The Company of Biologists Limited. Reproduced with permission.
Figure 7-34 Nuclease-protection method for quantitating specific RNAs in a mixture and mapping them. (a) A radiolabeled, single-stranded DNA probe (blue) is mixed with a mixture of cellular RNAs; the probe hybridizes only to the complementary RNA (red), which is a small fraction of the total RNA sample. In this example, the probe contains a sequence complementary to the entire RNA of interest. Digestion with S1 endonuclease degrades all the unprotected (unhybridized) RNA and DNA sequences, leaving a double-stranded RNA-DNA hybrid equal in length to the RNA. The protected hybrid is detected by gel electrophoresis followed by autoradiography. The density of the resulting bands is proportional to the amount of the hybridized RNA in the original mixture. (b) With a “partial” DNA probe, containing only a portion of the DNA sequence complementary to the RNA, the protected S1-digestion product is shorter than the RNA and equal in length to the complementary region of the probe. (c) In this example of mapping an RNA on the genome, a 1.7-kb RNA was approximately mapped to the region between 22.4 and 24.1 from the left end of the 36-kb adenovirus genome. Four radiolabeled restriction-fragment probes (A, B, C, and J) from this region of the viral DNA were prepared, hybridized with RNA from virus-infected cells, and then treated with S1 endonuclease. An autoradiogram of the S1-digestion products is shown at the right. Probes A and B produced S1-digestion products of 1.7 kb, indicating the RNA sequence maps entirely within these restriction fragments. The results with the partial probes C and J map the RNA sequence relative to the restriction site separating fragments C and J. [Photograph in part (c) from A. J. Berk and P. A. Sharp, 1977. Cell 12:721; copyright M.I.T.]
FIGURE 7-34 Nuclease-protection method for quantitating specific RNAs in a mixture and mapping them. (a) A radiolabeled, single-stranded DNA probe (blue) is mixed with a mixture of cellular RNAs; the probe hybridizes only to the complementary RNA (red), which is a small fraction of the total RNA sample. In this example, the probe contains a sequence complementary to the entire RNA of interest. Digestion with S1 endonuclease degrades all the unprotected (unhybridized) RNA and DNA sequences, leaving a double-stranded RNA-DNA hybrid equal in length to the RNA. The protected hybrid is detected by gel electrophoresis followed by autoradiography. The density of the resulting band is proportional to the amount of the hybridized RNA in the original mixture. (b) With a "partial" DNA probe, containing only a portion of the DNA sequence complementary to the RNA, the protected S1-digestion product is shorter than the RNA and equal in length to the complementary region of the probe. (c) In this example of mapping an RNA on the genome, a 1.7-kb RNA was approximately mapped to the region between 22.4 and 24.1 from the left end of the 36-kb adenovirus genome. Four radiolabeled restriction-fragment probes (A, B, C, and J) from this region of the viral DNA were prepared, hybridized with RNA from virus-infected cells, and then treated with S1 endonuclease. An autoradiogram of the S1-digestion products is shown at the right. Probes A and B produced S1-digestion products of 1.7 kb, indicating the RNA sequence maps entirely within these restriction fragments. The results with the partial probes C and J map the RNA sequence relative to the restriction site separating fragments C and J. [Photograph in part (c) from A. J. Berk and P. A. Sharp, 1977, Cell 12:721; copyright M.I.T.]
Mapping of Transcription Start Sites:

- As discussed earlier that some of the DNA regulatory elements that control transcription of genes into mRNA located near transcription start site.
- Mapping the start site for synthesis of a particular mRNA often helps in identifying the DNA regulatory sequences that control its transcription.
- There are two methods used to map the 5’ end of a particular mRNA on a complementary DNA:
  i. \( S_1 \) Protection
  ii. Primer extension
- The first step in both methods is to identify the general region of a DNA that includes the start site of interest.
- This can be done by Northern blot analysis or nuclease protection using various cloned restriction fragments as probes.

Fig 7.34, Lodish 4th Ed / Fig 3.33 3rd Ed
Concept Map 8, Nicholl 3rd Ed
Fig 7.38, Lodish 3rd Ed
FIGURE 7-34 Two methods for mapping the start site for transcription of a particular gene in a region of DNA of known sequence. (a) Diagram of the DNA fragment containing the gene of interest (light blue) and the corresponding mRNA (red). The end-labeled (red dot) single-stranded DNA fragment used as a probe in the S1 mapping technique and the end-labeled oligonucleotide primer used in the primer-extension technique are shown below the position of their sequence in the DNA. (b) In the S1 mapping technique, the probe is hybridized with the mRNA, and unpaired nucleic acid is then digested with S1 endonuclease (see Figure 7-33). Denaturation leaves a labeled DNA fragment whose length accurately marks the distance of the starting nucleotide of the mRNA from the nucleotide that hybridized with the labeled DNA end. (c) In the primer-extension technique, a short (approximately 20 nucleotides) oligodeoxynucleotide is synthesized and end-labeled. After the primer is hybridized to the mRNA, it is extended by reverse transcriptase until it reaches the first nucleotide of the mRNA. The length of the primer-extension product, determined by gel electrophoresis, measures the distance from the 5’ end of the primer to the 5’ end of the mRNA.
Clone identification

Highly specific method

requires a

which enables

Isolation of specific genes

for further analysis by

Translation of mRNA in vitro

two methods

Hybrid arrest

Hybrid release

Restriction mapping

in conjunction with

Blotting techniques

followed by

DNA sequence analysis

Selecting a clone

which requires

Genetic selection

for direct selection

Presence of vector

Complementation of a defined mutation

Screening a clone bank

two main methods

Nucleic acid hybridisation using a

Replica filter

Labelled probe

which may be

RNA

cDNA

Genomic DNA

Oligonucleotide

Immunological screening

detects proteins from

cDNA expression library

Concept map 8
FROM PROTEIN TO GENE

Isolate protein on the basis of its molecular function (e.g., enzymatic or hormonal activity)

Determine partial amino acid sequence of the protein

Synthesize oligonucleotides that correspond to portions of the amino acid sequence

Use oligonucleotides as probes to select cDNA or genomic clone encoding the protein from library

Sequence isolated gene

FROM GENE TO PROTEIN

Isolate genomic clone corresponding to an altered trait in mutants (e.g., nutritional auxotrophy, inherited disease, developmental defect)

Use genomic DNA to isolate a cDNA for the mRNA encoded by the gene

Sequence the cDNA to deduce amino acid sequence of the encoded protein

Compare deduced amino acid sequence with that of known proteins to gain insight into function of the protein

Use expression vector to produce the encoded protein

FIGURE 7-38 Overview of alternative strategies for studying the roles of specific proteins in complex cellular processes. This chapter highlights the protein-to-gene strategy; Chapter 8 describes the gene-to-protein strategy. In both cases, recombinant DNA technology can link a gene and its encoded protein, so that traditional biochemical and genetic approaches are combined in powerful ways.