

## MOLECULAR BIOLOGY – BASIC ASPECTS.

R.J.A. TRENT, MBBS, BSc. PhD. FRCPA. FRACP, DPhil.

Senior Lecturer, Department of Molecular Biology, Clinical Immunology Research Center, University of Sydney.

### HISTORICAL BACKGROUND

F.H.C. Crick has described "molecular biology" as the study of:



DNA was first isolated by Miescher in 1871 (nucleic acid comes from the acid material "nuclein" obtained from nuclei of pus cells). In the 1950's **Chargoff** and **Wyatt** showed that the ratios of the bases adenine (A) to thymine (T) and guanine (G) to cytosine (C) were close to unity. This, coupled with X-ray crystallographic data from Wilkins and Franklin enabled Watson and Crick to describe the *double-stranded helical nature of DNA*. Subsequently, two major discoveries preceded rapid advances in molecular biology. First, **Temin** and **Baltimore** isolated the enzyme *reverse transcriptase* from certain RNA viruses. Up to this time it was dogma that DNA coded for RNA and the reverse could not happen.

However, reverse transcriptase enabled single-stranded DNA to be prepared from an RNA template. This type of DNA (complementary or *cDNA*) could then be used as a probe to identify gene sequences. Characterisation of enzymes which digest DNA (*restriction endonucleases*) by **Nathans**, **Smith** and **Arber** was a further landmark in molecular biology. Restriction enzymes are found in bacteria and are thought to represent a primitive immune system. Their usefulness comes from the ability to recognise specific DNA sequences 4-7 base pairs in length. Restriction enzymes enabled DNA to be cleaved predictably into smaller units. Thus, total (*genomic*) DNA could be divided into more basic segments (i.e. a restriction fragment containing a gene). Genes were then identified with DNA probes.

### MOLECULAR MACHINERY OF THE CELL

Genetic information is carried in the linear sequence of nucleotide bases in DNA. Base pairing between G-C and A-T underlies the process of DNA replication. The human genome consists of about  $3 \times 10^9$  base pairs of DNA. A *gene* is that linear stretch of DNA

which codes for a single polypeptide chain. Genes consist of coding segments (*exons*) which are represented in the final gene product and non-coding segments (*introns*) which are removed from the final gene product during the *processing* of mRNA. Manufacture of mRNA copies from genes is called *transcription*. The enzyme *RNA polymerase* initiates transcription after binding to a specific DNA sequence termed a *promoter* and continues until the termination signal is reached. The RNA transcript is then processed by *splicing* and *polyadenylation* before transport to the cytoplasm. Each triplet codon is recognised by a specific tRNA molecule carrying the relevant amino acid. Regulation of gene expression may be *transcriptional* e.g. by the action of activators or repressors on *enhancer segments* of DNA which influence the binding of RNA polymerase to its promoters or post-transcriptional e.g. by affecting the processing or stability of mRNA or the translation process.

### GENE MAPPING – METHODOLOGY & APPLICATIONS

The basis of *gene mapping* is characterisation of restriction endonuclease sites associated within and around genes. Since DNA recognition sequences for each restriction enzymes are specific, it is possible to construct a unique *restriction map* pattern for individual genes. Steps involved in this process include:

(i) **DNA preparation.**

Any nucleated cell serves as a source of DNA. For convenience, peripheral blood is often used. DNA and proteins are separated by phenol extraction.

(ii) **Digestion of DNA with restriction enzymes:**

Restricted DNA is fractionated into its various fragments by electrophoresis in agarose gels. Following completion of electrophoresis double-stranded DNA is made single-stranded (ss) by reaction with sodium hydroxide. ss DNA fragments are then transferred to nylon or nitrocellulose membrane filters.

Article reprinted with the kind permission of the author from the Education Supplement of the Combined XXI Congress of the International Society of Haematology and XIX Congress of the International Society of Blood Transfusion, Sydney, May 1986.

## (iii) Hybridisation with DNA probes:

DNA sequences homologous to the gene(s) under investigation are called *probes*. These may be derived from the action of reverse transcriptase on mRNA (*cDNA probes*) or alternatively represent a discrete segment of DNA (*genomic probes*). Probes are labelled with <sup>32</sup>P-nucleotides. Boiling converts probes into ss DNA. *Hybridisation* enables the radiolabelled probes to anneal with corresponding ss DNA sequences. Excess probe which has not annealed is removed by washing. Hybridisation is detected by autoradiography of the membrane filter.

An unaltered gene structure is indicated by restriction enzyme fragment lengths which correspond to those expected in normal. Fragments of different size indicate a gene rearrangement or a DNA polymorphism. More recently, synthetic probes (*oligonucleotides*) have been produced. These are ss nucleotide sequences of approximately 10 bases. Two are usually necessary in diagnostic work. One oligonucleotide has normal base sequences. The second contains the sequences found in the mutant DNA segment. Careful selection of hybridisation and washing conditions enables differential hybridisation of the two oligonucleotides to their respective gene sequences. DNA with normal, heterozygous or homozygous gene patterns may then be identified.

Gene mapping has been applied extensively to the detection of genetic disease. The thalassaemia syndromes provide useful models for these disorders. For example,  $\alpha$  thalassaemias in the majority of cases result from *gene deletions*. Therefore, altered restriction enzyme mapping patterns associated with the  $\alpha$  globin genes become diagnostic. In contrast are *non-deletional disorders* such as the  $\beta$  thalassaemias. The majority of molecular defects in  $\beta$  thalassaemia involve intragenic point-mutations. In these situations restriction enzymes patterns are usually normal and conventional gene mapping is unhelpful for diagnosis. However, gene mapping may still be useful through an indirect approach which relies on linkage between normal or mutant genes and DNA polymorphisms. Information obtained from gene mapping may also be applied to carrier identification (e.g. the haemophilias) and antenatal diagnosis.

## CLONING AND SEQUENCING DNA

Foreign DNA may be inserted into *vectors* such as phage, plasmids or cosmids. Vectors

are then inserted (*transfected*) into *host* bacteria (usually *E. coli*). Vectors may contain markers such as antibiotic resistance genes to enable selective growth of transfected bacteria. Replication of the transfected bacteria and the vector within the host ensures production of large amounts of the foreign (*cloned*) DNA. The vector can be purified easily after bacterial lysis and foreign DNA, now in large amounts, is recovered by digestion with restriction enzymes. The genetic code is deciphered by *sequencing*. In the commonly used Sanger method, a series of transcripts of the DNA is made. Each of these terminates with a *dideoxynucleotide (dd)*. The transcripts are radiolabelled and separated according to length by gel electrophoresis. The lengths of these transcripts indicate the intervals between which a given base occurs since, for example, there will be fixed length transcript ending in ddC for each dG in the sequence being copied and read. If the procedure is repeated using each of the remaining dideoxynucleotides (ddG, ddT, ddA) the entire sequence can be deduced.

## MAKING A DNA "LIBRARY"

A genomic library consists of the entire genome of the cell cut into small pieces each of which is cloned into bacteria as described above. Each bacterial colony in the library represents one such piece of the genome. cDNA is made from mRNA template using reverse transcriptase. When cloned into bacteria each bacterial colony will represent one species of cDNA, or one species of mRNA in the cell of origin. Although the cell contains only about 12,000 different mRNA's, approximately 170,000 colonies are required to provide a 99% chance of representing every species in the library. Libraries may be *screened* by seeking a bacterial colony which contains DNA complementary to a labelled probe e.g. a short stretch of synthetic DNA (oligonucleotide). This is conveniently performed by lysing the bacterial colonies on nitrocellulose, to which DNA binds tightly, and hybridising a solution of the radiolabelled probe to the filter.

## GENE EXPRESSION

One additional application of recombinant DNA technology has been the ability to isolate a gene sequence, insert it into appropriate cells and manipulate the genetic environment in order to obtain protein production from the

gene. Already therapeutic agents such as insulin, growth hormone, interferon, plasminogen activator, hepatitis B vaccine have been prepared in this way. Whilst the genetic code is universal, a number of problems must be overcome before a functional protein can be synthesised by genetic engineering. They include:

- (i) Transcription of a cloned gene insert:  
This requires the presence of a promoter recognised by the host (e.g. *E. coli*) RNA polymerase. The number of gene copies present in the vector (e.g. plasmid) may also be significant.
- (ii) Translation of mRNA:  
To be efficient, translation requires mRNA to have a ribosomal binding site. A commonly used host system is *E. coli*. In this case the translation start codon (AUG, GUG) and a nearby DNA sequence comprise the ribosomal binding site.

- (iii) Post-translational processing:

A common post-translational modification of protein involves removal of *signal* sequences. The latter are involved in passage of protein through the cell membrane and so will determine if expression of protein is *intracellular*. The post-translation event of glycosylation required by some proteins for activity does not occur in the *prokaryote* system described above. Some post-translational problems need to be overcome by use of *eukaryote* expression systems. For example, vectors such as SV40, retroviral or caccinia viruses which are *integrated* into cultured mammalian cell lines may be used. Viral DNA in this instance becomes incorporated into host cell chromosomes.

The principles and problems inherent in gene expression may be illustrated by reference to preparation of insulin as a therapeutic agent and the ultimate goal of gene therapy through insertion of genetically engineered DNA sequences into patients with genetic disorders.

#### SUGGESTED READING

1. *DNA in Medicine*. 10 articles in Lancet starting in the 13 October 1984 issue. Brief, broad reviews in a wide range of topics.
2. *DNA structure, transcription, translation*. Scientific American. October 1985.
3. *Genetic Engineering vols 1-4 cd. R. Williamson*, Academic Press. Comprehensive reviews in a number of areas.