Clinical gene analysis and manipulation: overview and applications in pathology

F K Bedford, J A Jankowski

Introduction

"There are no such things as applied sciences, only the applications of science"

Louis Pasteur, 1872

The science of molecular cell biology has advanced dramatically over the last century and in the last decade this has had important implications for medical research particularly for oncology, microbiology, immunology, and inherited genetic diseases. Globally, the oncological sciences have the most resources devoted to them including specialist scientific journals, research funding, national institutions, and number of scientists. The increasing use of molecular cell biological techniques has led to considerable diversity in methodological practice and interpretation of genetic analysis. It is particularly important, therefore, that the calibre of clinical scientific endeavour keeps pace with the scope and potential of new developments in basic science. Furthermore, the rapid and exponential increase in knowledge and potential clinical applications of the new molecular techniques requires that clinical scientists are abreast of many developments so that the strengths and weaknesses of techniques can be fundamentally understood.

The diversity of backgrounds of clinical scientists makes it difficult for new researchers to learn comprehensive and fundamental aspects of molecular cell techniques from their colleagues. In many academic centres training is either rudimentary or learnt by "trial and error" making research in this area unnecessarily expensive in time and resources. Researchers must gain a profound understanding of molecular experiments to appreciate sound laboratory practice and gain confidence to tackle other appropriate techniques.

This leader is directed at those wishing to perform basic semiquantitative experiments such as northern or Southern blots and also to give an insight to those wishing to undertake more specialised genetic manipulation such as gene cloning and novel gene identification.

Louis Pasteur, a founding father of contemporary medicine, recognised that in order to achieve excellence in science there are no easy options. In this regard all we can offer is a brief overview of the simple applications of molecular biology to medicine and pathology, so that it may enthuse others to begin work in this area.

History of molecular genetics

Chromosomes were first described by W Sutton in 1903 but their significance was not ap-

preciated until 1910 when T Morgan suggested the possibility of separate genes which controlled the cellular phenotype. Progress was slow until O T Avery, C M Macleod and M McCarty in 1944 identified DNA as the unique molecule of chromosomes. It was in 1956 when F Crick, M Monod, J Watson, and M Wilkins identified the double helix structure of DNA that modern molecular genetics began. The first samples of artificially engineered DNA (recombinant DNA) were created in 1962 and in 1970 M Mandel successfully introduced foreign DNA species into bacteria (transformation) to create organisms with new genotypes. At this time, however, the methods of DNA manipulation were crude and lacked sensitivity and easy reproducibility. The identification of restriction enzymes which selectively cut sections of DNA, by A Arber, H Smith and D Nathans in 1972, enabled the genome to be studied in greater detail using the new technique of prokaryotic cloning (*Escherichia coli* pSC101) devised by H Boyer and S Cohen in 1973. This permitted selection and propagation of chosen DNA fragments to be amplified in a bacterial plasmic vector. In particular, the technique of DNA sequencing first described in 1972 by W Gilbert enabled the exact order and DNA primary structure to be studied. Unfortunately, these techniques were so revolutionary that genetic analysis was still only accessible to a small minority of highly specialised laboratories. In the mid 1970s, however, Southern gel analysis enabled analysis of genomic DNA to be performed and northern gel analysis permitted analysis of the highly labile RNA molecules for the first time. These techniques were easily learnt and are still the basic tools of molecular study in laboratories. Finally, modern molecular biology entered a new era when K Mullis and R Saiki devised the polymerase chain reaction (PCR) in 1982. This technique is perhaps now the most extensively used experimental procedure in molecular biology laboratories.

The subsequent section will highlight three fundamental molecular biology techniques: gel analysis, PCR based techniques and molecular cloning and sequence analysis.

Modern molecular biology

In general terms molecular techniques can be used to assess gene expression at the RNA or protein level. In most cases this is achieved by gel electrophoresis and the information gained is both quantitative (expressed as gene enhancement over control samples) and qualitative (expressed as the appropriateness of the molecular weight of the gene products).

Department of
Reproductive Biology,
University of
California, San
Francisco, CA, USA
F K Bedford

Imperial Cancer
Research Fund
Laboratories, c/o
Department of
Histopathology, Royal
Postgraduate Medical
School, London
J A Jankowski

Correspondence to:
Dr J A Jankowski, Assistant
Professor in Medicine,
Gastrointestinal Unit,
Department of Medicine, VA
Medical Centre, University
of California, 4150 Clement
Street (151M2), San
Francisco, CA 94121, USA.

Accepted for publication
23 September 1994

Figure 1 Western blot of E-cadherin immunoreactivity in colorectal tissue in vivo and in vitro. The differential cell lines "col 6" and "col 30" had more E-cadherin. The cancers (ca) express abundant E-cadherin protein compared with normal tissue (n) or mildly dysplastic polyps (md) but this is usually mutated and dysfunctional.

Genomic DNA can be studied by coarse methods which enable approximate sizes of genes or their products to be determined by gel blotting. Fine analysis of the genetic structure, however, can be performed using gene sequencing. Analysis of genes and their products is only the initial step because these genes can be manipulated in vitro to decrease, increase or alter their biological function. In vivo studies also include the generation of transgenic animals and this is already a routine procedure in many laboratories. The potential applications to humans are huge and already over 30 separate clinical treatments involving gene therapy have been reported. This leader will, however, be restricted to the discussion of gene analysis and preliminary genetic manipulation.

Techniques and applications

QUANTITATIVE GEL ANALYSIS

The techniques of western blotting (protein), northern blotting (RNA) and Southern blotting (DNA) have many similarities. The molecules must each be extracted from tissue in the presence of inhibitors which prevent breakdown of DNA, RNA and protein by DNases, RNases and proteases, respectively (molecular extraction). Subsequently, the molecules are electrophoresed on agarose or polyacrylamide gels to separate the molecules by size (gel fractionation). These gels are rather fragile and therefore the molecules must be transferred (blotted) to rigid permanent supports such as nitrocellulose or nylon membranes. These membranes are subsequently probed by either antibodies in the case of western blots or nucleic acid probes in northern and Southern blots (hybridisation).¹

The extraction step is probably the most important step and may frequently be unsatisfactory. As a result it is usually preferable to assess molecular integrity using Cam 5-2, GAPDH or β -actin probes for protein, RNA and DNA, respectively.² Figure 1 shows a western blot of human colorectal cell lines expressing different quantities of the cell adhesion molecule E-cadherin. The immunoreactive fragments are all at the correct molecular weight whereas specimens of intestinal metaplasia of Barrett's oesophagus and gastric metaplasia may have truncated immunoreactive fragments (fig 2). Figure 3 shows a northern blot of the growth factor cripto RNA expression in human colorectal tissue, indicating that all specimens have not only equal quantities of both cripto and the loading control β -actin but also that the sizes of all of the RNA species are of the correct molecular weight.

Figure 2 Western blot of E-cadherin immunoreactivity in gastric tissue. Lanes B and C represent tissue extracted from normal or metaplastic gastric mucosa and express E-cadherin 118 kDa protein; lane D represents a gastric carcinoma expressing truncated E-cadherin 31 kDa protein.

SPECIALISED PCR TECHNIQUES

PCR is theoretically a very simple technique in which specific DNA sequences flanked by

Figure 3 Northern blot of the growth factor *cripto* in colorectal tissue. The mRNA species in all lanes is of the same base pair size, which is correct for the mature *cripto* 2 kilobase RNA. In addition, all lanes show similar densities of autoradiographic signal confirming that this RNA species is not dramatically altered in this cell line. The β -actin bands below confirm equal loading of the RNA.

predesigned primers are amplified. The process of amplification involves three steps which are performed in cycles: denaturation, to separate the DNA strands; annealing, to permit the primers to stick to the single stranded DNA; and polymerisation to form another new complementary strand resulting in double stranded DNA. This process doubles the amount of DNA at each cycle and therefore a few strands

of a chosen RNA (complimentary DNA (cDNA)) or genomic DNA can be amplified one million times. The discovery of the heat stable enzyme Taq polymerase and subsequent homologues enabled this cycle to be performed in the one sample buffer changing only the temperature.

Typical reaction conditions are as follows: step one, 94°C for five minutes (to denature the DNA); step two, 94°C for one minute (another denaturing step)(increase the temperature to 95–97°C if the cDNA is G–C rich); step three, 56°C for two minutes (to enable the primer to anneal to the DNA) (the temperature can be increased to 62°C to achieve more specific products); step four, 74°C for two minutes (to enable the new DNA strand to extend) (increase time especially when original DNA stand is over 2 kilobases in length or when DNA is present at a low concentration); and cycle steps two to four 30 to 40 times. During the final step the DNA is heated to 74°C for four minutes to ensure all the new DNA is synthesised. The reaction products can be easily visualised in a 1% agarose gel containing < 1 μ g/ml ethidium bromide.

Researchers frequently use PCR to make comparisons of the quantity of gene expression between several clinical samples. It is imperative, therefore, to amplify a control single copy gene in addition to the test gene to ensure normalisation (equal loading) of the starting DNA prior to amplification. Figure 4 shows amplification of the tyrosine kinase receptor of the *c-erbB3* gene in non-dysplastic specimens of Barrett's mucosa. The dysplastic specimens do not express identifiable amounts of this gene but express the control GAPDH sequences ensuring validity of the reaction.³ We have also applied the use of this extremely sensitive technique to the analysis of cytological specimens of Barrett's mucosa and it can be seen that

Figure 4 Reverse transcriptase – PCR of the *c-erbB3* gene in Barrett's mucosa. The four lanes containing either cancers (Ca) or dysplastic mucosa (Dys) have dramatically less *c-erbB3* than the normal tissue (n).

Figure 5 Reverse transcriptase – PCR of the *c-erbB3* gene in oesophageal cytology. Serial brushings were taken from the distal oesophagus at 37 to 40 cm during a routine endoscopy. After taking the cytology, pinch biopsy specimens were analysed to assess the morphology of each 1 cm strip of mucosa. At 37 and 38 cm, the tissue consisted of intestinal metaplasia with dysplasia, whereas at 39 cm the tissue did not have dysplasia. The *c-erbB3* expression seems closely associated with cellular phenotype.

Trouble-shooting for failed PCR experiments	
Problem	Solution
No bands	<ul style="list-style-type: none">– Repeat and increase DNA concentration– Decrease annealing temperature– Increase magnesium concentration– Check primers, etc.– Check enzyme (amplitaq good, although vent or pfu has proof-reading action also)
Background bands	<ul style="list-style-type: none">– “Hot start” (increase denaturation temperature)– Decrease magnesium concentration (normal 15 mmol, range 5–25 mmol)– Increase annealing temperature
Smear	<ul style="list-style-type: none">– Decrease starting DNA by a factor of 2–10

consecutive 1 cm sweeps of the distal oesophagus reveals different *c-erbB3* expression. In particular, 39 cm corresponded to an area of Barrett’s gastric-type metaplasia with no dysplasia whereas at 37 and 38 cm the oesophagus had intestinal-type mucosa and dysplasia (fig 5).

Three types of problems can occur during PCR reactions: failure to amplify any bands, multiple bands or a DNA smear. The possible solutions to these problems are listed in the table.

Further modifications to the basic PCR protocol have resulted in several different applications including mutational analysis by single stranded conformational polymorphisms (SSCP) and novel gene identification by Alu

Figure 6 Single stranded conformational polymorphisms in colorectal tissue. Both alleles of each E-cadherin (E-CAD) gene can be seen in all samples (two parallel arrows) but in lane 1 an extra band (arrow) is noted corresponding to a conformational polymorphism (mutation in this case).

Figure 7 Degenerate PCR in colorectal tissue. Degenerate PCR primers were produced so as to amplify other members of the cadherin family containing a 200 base pair (bp) fragment which is highly conserved between species. The annealing temperature of the degenerate primer was increased (left to right) by 1°C. At 54°C, the products are mainly of an incorrect size (<190 base pairs). As the annealing temperature increases, bands of >200 base pairs also appear in addition to the correct 200 base pair fragment. At 57°C, only the correct product is amplified.

PCR or degenerate oligonucleotide primed (DOP) PCR. SSCP can be identified by running the labelled PCR products on a non-denaturing polyacrylamide gel. Aberrant DNA strands will run at different speeds in the gel, resulting in new bands when compared with tissue without mutations (fig 6). DOP PCR is the method whereby oligonucleotides (primers) which have homology within a gene family can be used to amplify all members of related molecules including those undiscovered previously.⁴ The PCR reaction conditions are important in deciding how specific the homology of the resulting amplified products is compared with the chosen gene family. Figure 7 shows the effect of serial decreases of 1°C of the annealing temperature on the final reaction products. The resulting PCR fragments can then be subcloned and individual colonies further analysed by sequencing (vide infra).

MOLECULAR CLONING AND SEQUENCE ANALYSIS

Gene cloning is the formation of multiple clones of identical single genes and enables analysis and manipulation of single genes and their products. In simple terms the segment of DNA of interest (termed insert DNA) is placed between the linearised ends of a circular bacterial plasmid DNA so that specific *E coli* strains may subsequently process and amplify the insert DNA as its own. There are many applications such as structural, expression and mutational analysis of segments of insert DNA, yet this technique is simple and involves four

basis steps: purification of insert DNA, modification (if any required) of insert and vector DNA, transformation of *E coli* bacterial cells, and finally the selection of bacterial colonies containing cloned insert DNA. This technique can be performed with single genes or with entire genomes such as is commonly performed in the construction of genomic or cDNA libraries.⁵

Two main options are available for cloning using either blunt ended or cohesive insert DNA fragments. The latter option is 100 to 1000 times more efficient and is the preferred method but if large quantities of insert DNA are available then the former option may be satisfactory. Once subcloned colonies are grown and purified, the DNA can then be sequenced usually by the dideoxy chain termination technique. This technique permits a modified T4 polymerase to extend four identical aliquots of the insert DNA until a dideoxy GTP, ATP, TTP or CTP base (which blocks further DNA chain extension) is substituted for the normal deoxy base. The end result of subcloning insert DNA from a cancer cell line with an invasive phenotype is shown in fig 8A. The sequencing gel indicates a single base pair transition of E-cadherin gene resulting in a change from leucine to proline.

Summary and future prospects

These techniques can be used to aid diagnosis, prognosis and therapy in many aspects of clinical medicine. It is perhaps in the area of tumour biology where the applications are greatest. In particular the identification of *c-ras* mutations in codon 12 has already been used as a potential screening aid for colorectal cancer when applied to faecal cytology.⁶ Furthermore, identification of mutations in the p53 and *c-ras* genes has also been shown to have prognostic significance in addition to tumour size, stage and morphology.⁷ At present, only 5% of the human genome is characterised and therefore identification of novel genes which have a role in normal development and tumorigenesis is urgently required. In this regard simple techniques such as DOP PCR enable such strategies to be achieved. This technique has been used successfully to identify *HEX*, a novel homeo-domain gene expressed during haematopoiesis.⁸ This strategy has also been employed to identify E-cadherin homologues, which have an important role in colorectal cancer.

As suggested in the introduction gene transfer therapy has already been performed in several clinical areas, most notably in the treatment of inherited enzyme deficiencies. As yet, its application in cancer is still experimental but several strategies are possible such as enhancing immune surveillance and antibody targeting.⁹

Finally, this short leader has several key references which may be of use to those wishing to undertake any of the procedures outlined above.

1 Jankowski J, Jones M. Quantitative gel analysis. In: Jankowski J, Polak J, eds. *Clinical gene analysis and manipulation*. London: Cambridge University Press, 1995 (in press).

Figure 8 Sequencing reaction of E-cadherin mutations in colorectal tissue. In the colorectal cancer in lane 2 there is a transition changing CTC to CCC. This amino acid change results in a marked conformational change in the E-cadherin molecule.

- 2 Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbour Laboratory Press, 1989.
- 3 Innis MA, Gelfand DH, Sninsky JJ, White TJ. *PCR protocols*. San Diego: Academic Press, 1990.
- 4 Bedford F, Jankowski J. Identification of Novel Gene Species. In: Jankowski J, Polak J, eds. *Clinical gene analysis and manipulation*. London: Cambridge University Press, 1995 (in press).
- 5 Boothwell A, Yancopoulos GD, Alt FW. *Methods for cloning and analysis of eukaryotic genes*. Boston: Jones and Barlett, 1990.
- 6 Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, *et al*. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 1992;256:102-5.
- 7 Scott N, Quirke P. Molecular biology of colorectal neoplasia. *Gut* 1993;34:289-92.
- 8 Bedford FK, Ashworth A, Enver T, Wiedeman LM. HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. *Nucleic Acids Res* 1993;21:1245-9.
- 9 Jankowski J, McGarrity TJ. Screening for alimentary carcinoma: outstanding problems and a new strategy for research. *Eur J Cancer Prev* 1993;2:211-14.