

## Gene cloning: studies of nutritional regulation of gene expression

BY JOHN E. HESKETH AND KRISTINE PARTRIDGE

*Division of Biochemical Sciences, Rowett Research Institute,  
Bucksburn, Aberdeen AB2 9SB*

The precise variables of cell function are a result of the interaction between environmental factors such as nutrition and the genetic information present in the DNA. Thus, the modulation of gene expression by dietary components determines aspects of cell function, both normal and pathological: many nutritional and dietary effects on metabolism and tissue function will depend ultimately on the ability of specific nutrients to modulate the expression of genes and, thus, the extent of synthesis of specific proteins. Molecular biology techniques and gene cloning are important for nutritionists as they provide the tools and experimental approaches to study of the interaction between nutrition and gene expression.

Cloning is defined as the isolation of identical host cells containing a recombinant DNA molecule. In most cases the host is a bacterium and the recombinant DNA molecule consists of a bacterial plasmid or bacteriophage DNA linked to the DNA of interest: the plasmid or phage is used as a vector to carry the DNA of interest into the host and to allow the generation of a large number of copies of the DNA of interest (by both multiplication of the vector and the host). Crucial to the cloning process are the availability of suitable vectors, the techniques of DNA purification and the ability to cut (restrict) and join (ligate) pieces of DNA. A thorough overview of gene cloning is given by Brown (1990).

### VECTORS AND ENZYMES

Any discussion of gene cloning needs to be prefaced by a description of the various vectors and enzymes available to the molecular biologist. All the specific techniques, methods, and reagents are listed in detail in laboratory manuals such as Sambrook *et al.* (1989).

The vectors used in cloning are DNA molecules that fall into three categories: (1) plasmids: double-stranded, circular, genetic units which are carried by bacteria and which replicate using host enzymes; (2) lambda-based vectors: also double-stranded DNA but linear, having a more complex manner of replication (lytic or lysogenic); (3) filamentous bacteriophages: viruses that infect bacteria and contain a single-stranded circular DNA molecule. Each of these types of vectors have specific uses, for example M13, a filamentous bacteriophage, is ideal for DNA sequencing and site-directed mutagenesis. Specific plasmids are used for prokaryotic or eukaryotic expression of foreign proteins. The vector of choice will depend on the proposed application.

As far as enzymes go, restriction enzymes are the workhorses of molecular biology. These enzymes recognize short specific sequences of DNA and cut them in a site-specific manner. The recognition sequence can be four to eight nucleotides in length, or longer, and is usually palindromic. The cleavage site varies for different enzymes, as well as the cleavage pattern. Some enzymes cut the DNA so as to leave protruding single-stranded overhang, 'sticky ends', while others leave blunt ends. The specificity and number of

restriction enzymes mean that there are many restriction sites within a gene but that one enzyme may only cut once. This means that restriction digestion will produce identifiable fragments. Other enzymes include the DNA and RNA polymerases used for generating DNA and RNA *in vitro*, DNA ligase (*EC* 6.5.1.1-2) for joining DNA fragments, and DNA kinase and phosphatase for altering the phosphorylation state of the DNA.

#### CLONING A GENE

In essence, cloning involves three steps: producing DNA from the source of interest, ligation of the DNA fragments into a vector and selecting the fragment containing the DNA or gene of interest. This selection is crucial to cloning and usually depends on an oligonucleotide probe which has been designed from the protein sequence. This probe is a short stretch of nucleotides with a sequence complementary to the DNA sequence of a gene of interest. This approach takes advantage of the fact that single-stranded nucleic acid molecules will form base pairs with complementary single-stranded molecules. The probe will 'hybridize' to the RNA or DNA of the gene of interest and, thus, identify that gene. Various methods are employed to design this probe, such as aligning the gene sequence from related organisms and identifying conserved regions, or determining the N-terminal sequence of the protein and back-translating to get a putative DNA sequence. Various considerations are important in designing the probe such as primer dimer and secondary structure formation, and annealing temperature; there are computer programs to aid in primer design. Once synthesized the probe is labelled with either  $^{32}\text{P}$  or a non-isotopic tag such as digoxigenin that allows identification later on an autoradiograph or autoluminograph.

The search for the gene can take one (or more) of the three approaches shown schematically in Fig. 1, or might involve new techniques such as polymerase chain reaction cloning as discussed by Chesters (1996). The choice of approach depends on several factors; the Southern blotting method may be faster, library construction can be time-consuming. However, a complementary DNA (cDNA) or genomic library may be available from other researchers. Each method has its own advantages and disadvantages and it may actually be necessary to use more than one approach to isolate the gene.

The construction of a cDNA library (Fig. 1) involves isolating the total mRNA from the tissue of interest. The mRNA is translated into DNA by using reverse transcriptase (*EC* 2.7.7.49), an RNA-dependent DNA polymerase. The result is an RNA-DNA hybrid molecule from which it is necessary to remove the RNA with RNase H (*EC* 3.1.26.4). The second strand of DNA is synthesized by DNA polymerase I (a DNA-dependent DNA polymerase; *EC* 2.7.7.7). This process produces copies of the genes expressed in a particular tissue. The DNA molecules are 'packaged' or ligated into vectors and transformed into *Escherichia coli*. Positive colonies are detected by using the labelled probe in a process called colony hybridization.

Construction of a genomic library takes much the same course, with the exception that total genomic DNA is isolated, partially digested with a restriction enzyme, then ligated into a vector. Since large fragments are common, the vector is usually lambda or a cosmid (a special plasmid that can accommodate exceptionally large fragments of DNA, up to 52 kb). Again, the vectors are transformed into *E. coli* and screened by colony hybridization for positive colonies.

The Southern approach is somewhat different from screening libraries. Genomic DNA

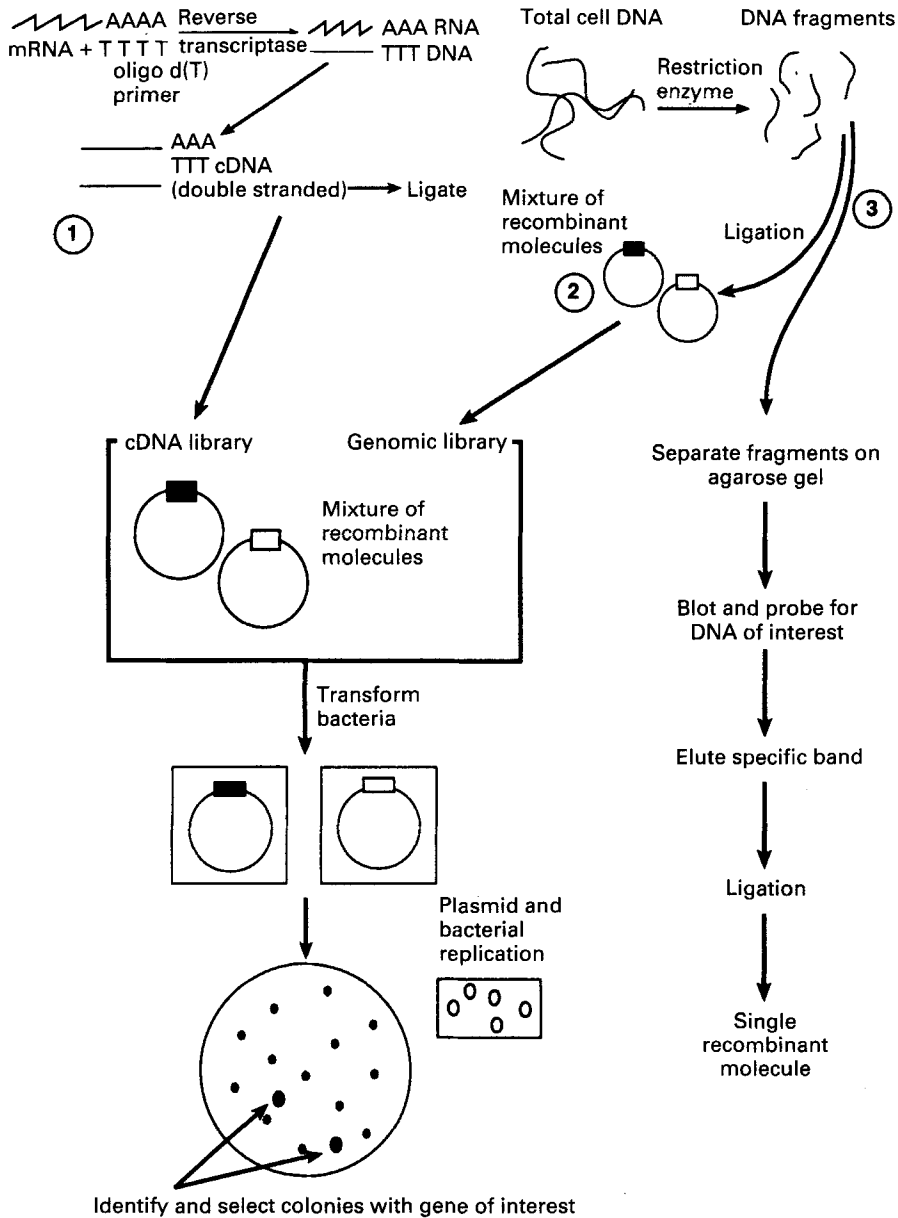


Fig. 1. Approaches to gene cloning. Three alternative methods of cloning are shown schematically: (1) cDNA library, (2) Genomic library and (3) Southern blotting.

is isolated and digested with one or a combination of two restriction enzymes. The fragments are separated by size by electrophoresis through an agarose gel. The DNA is transferred to nitrocellulose, then probed with the labelled oligonucleotide, and a DNA fragment identified by size. This fragment is isolated and then ligated to a vector. It should be kept in mind that in all these approaches the full gene may not be isolated at

first, and it may take several rounds of probe design and isolating gene fragments before the full gene is cloned.

#### SUBCLONING

The precise experimental objectives of a particular project may require that the DNA of interest is present in a vector of particular characteristics and so, having obtained the DNA of interest, it may be necessary or desirable to put the whole, or part, of the DNA of interest into a different vector (subcloning).

Particularly important vectors are those which possess the capability to express the inserted DNA in mammalian cells (mammalian expression vectors) and those which contain DNA polymerase promoters so that they can generate labelled RNA copies of the DNA. The expression vectors are used in transfection and transgenic experiments and the production of cells and animals which will continue to express the gene over many generations. The production of labelled RNA allows the production of riboprobes, which are useful for *in situ* hybridization, and labelled RNA for studies of RNA-binding proteins.

Specific parts of a cDNA or gene can be cut out by restriction enzymes and then ligated into a different cDNA to produce a chimaeric construct. The expression in mammalian cells of either cDNA with such deletions or of chimaeric cDNA allows the study of different regulatory domains (e.g. promoter, 3' untranslated region (UTR)).

#### GENE CLONING IS ESSENTIAL FOR THE STUDY OF REGULATION OF GENE EXPRESSION

There is considerable potential for nutritional factors to influence the synthesis of protein from the gene. As shown in Fig. 2, there are at least four stages at which nutrition can regulate gene expression. First, there can be regulation of gene transcription through control of the promoter, as has been shown to occur in the case of Zn regulating the thymidine kinase (*EC 2.7.1.21*) gene (Chesters *et al.* 1995).

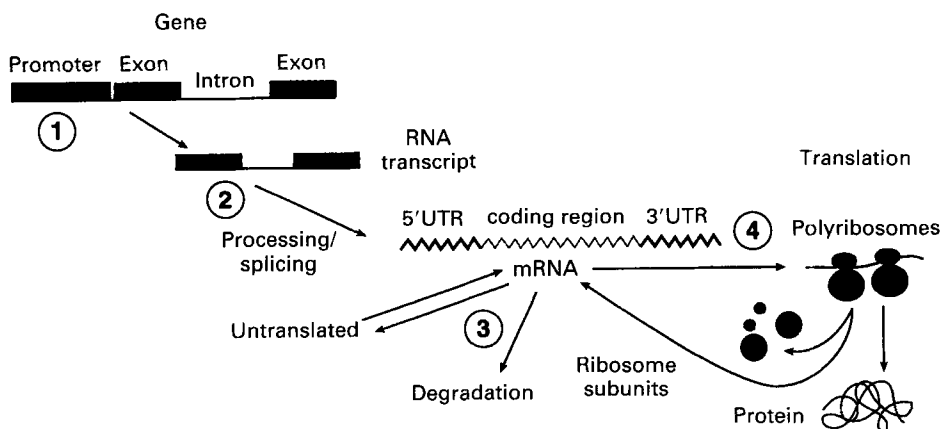


Fig. 2. Regulation of gene expression. Potential control points are identified: (1) gene transcription, (2) splicing and RNA processing, (3) mRNA degradation and (4) mRNA translation. UTR, untranslated region.

Second, removal of the intron sequences from the RNA transcript (splicing) can produce different mRNA which in turn can provide different regulatory sequences or different isoforms of the protein. In the case of the insulin receptor, splicing produces two major mRNA species in the rat and the relative proportion of the two species, as well as overall mRNA levels, is influenced by feeding diets high in phytohaemagglutinin lectin (Knott *et al.* 1992).

Third, mRNA levels, which partly determine the extent of protein synthesis, can be regulated not only by changes in gene transcription, but also by alterations in mRNA stability and degradation, as has been suggested for the regulation of the selenoenzymes glutathione peroxidase (*EC* 1.11.1.9) and phospholipid peroxidase (*EC* 1.11.1.12) by dietary Se (Bermano *et al.* 1995). Fourth, protein production can be controlled also by regulation of the translation of mRNA; this has been elegantly demonstrated in the case of regulation of ferritin synthesis by Fe, where the extent of translation is controlled by binding of a protein to untranslated sequences in the mRNA (Klausner *et al.* 1993). In addition, recent evidence shows that certain mRNA are localized to particular cytoplasmic sites within cells, probably to provide local synthesis of proteins close to where they will function (Hesketh, 1994), and potentially there is the possibility that nutritional factors may also regulate the site of synthesis of specific proteins.

Study of all previously mentioned control mechanisms depends critically on the measurement of mRNA. This requires the availability of a complementary nucleic acid sequence which can be labelled in some way and used as a 'probe': these probes can be part of a complementary cDNA clone (cDNA), an RNA sequence derived from the cDNA (riboprobe) or a short (twenty-five to fifty bases) length of DNA synthesized from the known sequence (oligonucleotide). Cloning can not only generate such probes but also, in addition, cloning of the full mRNA or gene sequence allows the introduction of these sequences in a native or modified form into cells (transfection) or animals (transgenics). These two techniques allow a powerful approach to the study of nutrient-gene interaction; the effects of these gene products on metabolism can be studied, nutrient requirements can be modulated, and the regulatory regions of genes and mRNA identified.

*Mechanisms of nutritional control of gene expression:  
modulation of selenoenzyme by dietary selenium*

The genes for the mammalian selenoenzyme cytosolic glutathione peroxidase, phospholipid hydroperoxide glutathione peroxidase and 5'-deiodinase have been cloned. The availability of the appropriate cDNA has allowed the investigation both of the mechanism of Se incorporation (Berry *et al.* 1991) and the effects of dietary Se supply on the levels of the mRNA (Bermano *et al.* 1995). Thus, dietary Se has been found to have differential effects on levels of the three mRNA within a particular tissue (e.g. liver) and also to have different effects in different tissues, for example, between liver, thyroid and heart. The availability of the cloned cDNA also provides the tools for analysis of transcription rates in isolated liver nuclei and rates for the three selenoenzyme genes were found to be similar in nuclei from rats fed on Se-deficient and control diets, suggesting that the differential effect of diet on liver mRNA levels is due to an effect on mRNA stability (Bermano *et al.* 1995).

*Modification of nutritional requirements:  
mammalian cells which make their own threonine*

The amino acid threonine is essential in the mammalian diet because it is not synthesized by these organisms. However, the pathway for its synthesis exists in bacteria and the genes encoding the enzymes responsible have been cloned. Subcloning these genes into a mammalian expression vector and transfection into mammalian fibroblasts has enabled the introduction of all the bacterial enzymes responsible for synthesis of threonine from aspartate into mammalian cells (Rees & Hay, 1995); these cells are capable of synthesizing threonine, demonstrating that the combination of cloning and transfection has achieved a modification of the cells' nutritional requirements.

*Transfection of chimaeric constructs to study control of translation and mRNA stability by 3' untranslated regions*

There is increasing evidence for the importance of regulation of mRNA translation in the overall control of gene expression (McCarthy & Kollmus, 1995), and this regulation seems to involve signals within the untranslated sequences of the mRNA. Untranslated sequences are also important in the regulation of mRNA stability, insertion of Se into selenoenzymes (Berry *et al.* 1991) and in determining the site of mRNA translation within the cell (Hesketh *et al.* 1994).

Cloning is crucial to these studies of UTR function: first, it allows the generation of gene constructs in which the UTR are either modified, exchanged with those of other genes or linked to a sequence which allows detection of the transcript (reporter); second, subcloning into a mammalian expression vector allows these sequences to be expressed in cultured cells. For example, cutting out the c-myc 3'UTR with restriction enzymes and ligating it to the coding sequences of the  $\beta$ -globin gene in an appropriate vector produces a chimaeric construct of  $\beta$ -globin coding sequences linked to the c-myc 3'UTR. Transfection of fibroblasts with this construct results in expression of a modified gene and production of a chimaeric mRNA transcript. This technique has shown that the c-myc 3'UTR influences mRNA localization within the cell (Hesketh *et al.* 1994). A similar approach can be used to study mRNA stability or translation and this is presently being used in nutrition research by producing other chimaeric constructs in which 3'UTR regions from metallothionein and selenoenzyme mRNA are linked to reporter sequences; the cloning and subsequent transfection of these constructs will allow investigation of the function of these 3'UTR regions and their role in nutritional regulation of gene expression.

#### REFERENCES

- Bermano, G., Nicol, F., Dyer, J. A., Sunde, R. A., Beckett, G. J., Arthur, J. A. & Hesketh, J. E. (1995). Tissue-specific regulation of selenoenzyme gene expression during selenium deficiency in rats. *Biochemical Journal* **311**, 425–430.
- Berry, M. J., Banu, L., Chen, Y., Mandel, S. J., Kieffer, J. D., Harney, W. & Larsen, P. R. (1991). Recognition of UGA as a selenocysteine codon in type 1 deiodinase requires sequences in the 3' untranslated region. *Nature* **353**, 273–276.
- Brown, T. A. (1990). *Gene Cloning: An Introduction*, 2nd ed. London: Chapman and Hall.
- Chesters, J. K. (1996). Polymerase chain reaction. *Proceedings of the Nutrition Society* **55**, 599–604.
- Chesters, J. K., Boyne, R., Petrie, L. & Lipson, K. E. (1995). Role of the promoter in the sensitivity of human thymidine kinase to lack of zinc. *Biochemical Journal* **308**, 659–664.

- Hesketh, J. E. (1994). Translation and the cytoskeleton: a mechanism for targeted protein synthesis. *Molecular Biology Reports* **19**, 233–243.
- Hesketh, J., Campbell, G., Piechaczyk, M. & Blanchard, J.-M. (1994). Targeting of c-myc and B-globin sequences to cytoskeletal-bound polysomes by c-myc untranslated region. *Biochemical Journal* **298**, 143–148.
- Klausner, R. D., Rouault, T. A. & Harford, J. B. (1993). Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* **72**, 19–28.
- Knott, R. M., Grant, G., Bardocz, S., Pustzai, A., Carvalho, A. F. F. & Hesketh, J. E. (1992). Alterations in the level of insulin receptor and Glut-4 mRNA in skeletal muscle from rats fed a kidney bean (*Phaseolus vulgaris*) diet. *International Journal of Biochemistry* **20**, 897–902.
- McCarthy, J. E. G. & Kollmus, H. (1995). Cytoplasmic mRNA-protein interactions in eukaryotic gene expression. *Trends in Biochemical Sciences* **20**, 191–197.
- Rees, W. D. & Hay, S. M. (1995). The biosynthesis of threonine by mammalian cells: expression of a complete bacterial biosynthetic pathway in a mammalian cell. *Biochemical Journal* **309**, 999–1007.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.