Expression of eukaryotic genes in
E. coli

T. J. R. HARRIS

Celltech Ltd, 250 Bath Road, Slough SL1 4DY, Berks, UK

128   *T. J. R. Harris*

I   Introduction

In recent years the techniques of *in vitro* DNA recombination followed by transfection of suitable host cells with recombinant vectors (gene cloning) has led to a great increase in our understanding of the structure and function of the genomes of many organisms. In the early stages of this work it became clear that genes which were cloned in this way could be expressed in the new host if the genetic elements controlling expression were suitably arranged. The results of these efforts will find application in two spheres. In the first, new approaches to fundamental studies on the relationship of protein structure to function will be possible. Already, molecules have been produced which are hybrids of the appropriate regions of different interferon molecules and their functions are being examined. This is possible not only because the genes for the proteins can be recombined but because they can then be expressed in *E. coli* in quantities sufficient for purification and biological study (Streuli *et al.*, 1981; Weck *et al.*, 1981). Further extensions of this kind of work can be foreseen where one or a few selected amino acids (e.g. near the active site of an enzyme) are altered by *in vitro* mutagenesis (Shortle *et al.*, 1981; Lathe *et al.*, 1983) and the effect on enzymatic function assayed. Secondly, such is the power of these gene cloning and expression techniques that they are already having a profound impact on the practice of biotechnology and it seems that few areas of this technology will remain unaffected by them. Indeed, the first proteins made by recombinant DNA techniques are now being produced in sufficient quantity for extensive safety and efficacy testing.

Insulin and growth hormone, both conventionally isolated from human endocrine tissue have now been made in *E. coli* and the proteins purified (Goeddel *et al.*, 1979a, 1979b). Considerable effort has been expended on the isolation and expression of both leukocyte (Le or α) and fibroblast (F or β) interferon genes so that the potential of these antiviral compounds can be evaluated properly (see Scott and Tyrrell, 1980). There is also the possibility of producing proteins for use as vaccines against a variety of infectious agents by cloning and expressing the genes coding for the relevant surface immunogens. Notable progress has been made towards a vaccine for foot and mouth disease virus (FMDV) using this approach, where one of the capsid proteins (VPI) produced in *E. coli* has been shown to elicit

neutralizing antibody (Kleid *et al.*, 1981). Genetically engineered vaccines for other viruses such as hepatitis B and rabies virus are also being considered.

Although none of these initial examples of the expression of proteins from recombinant organisms is as yet established as a biotechnological process, the way in which the expression of the recombinant DNA was achieved forms a general paradigm for all future studies. However, at the same time, it is clear that not all the rules governing the expression of cloned genes have been elaborated and those rules that do exist are still largely empirical. In this article the ways in which expression has been achieved are reviewed, some of the problems discussed and some of the probable future systems considered.

II   Gene expression in *E. coli*

E. coli has been used as the host cell for expression of foreign genes mainly because more is known about the control of gene expression in this organism than in any other. It is well established, for example, that the genes involved in a particular metabolic activity tend to be clustered in transcriptional units (operons) with the major control regions (the operator and promoter) located at the beginning of the cluster (for a detailed description of bacterial gene expression, see Miller and Reznikoff, 1980). The operon is transcribed into a polycistronic mRNA from which the polypeptides are then translated. Transcriptional control is exerted over the expression of an operon and varies depending on the function of the genes in the operon (see Miller and Reznikoff, 1980). Since relatively few promoter systems are currently being utilized to express cloned genes, the essential elements of their control mechanisms will be dealt with when considering each system. Expression of a cloned gene requires efficient and specific transcription of the DNA, translation of the mRNA and in some cases post-translational modification of the resulting protein.

A   Transcription

The first step in the initiation of transcription in *E. coli* is the binding of RNA polymerase to a promoter sequence in the DNA. Analysis of the DNA sequence of many promoters in *E. coli* has revealed two regions of homology located about 35 base pairs (bp) upstream from the transcription initiation site (the -35 region) and about 10 bp upstream (the -10 region or Pribnow-Schaller box). The

130　　*T. J. R. Harris*

conserved sequences in the -35 and -10 regions (TTGACA and TATAAT respectively, Rosenberg and Court, 1979; Siebenlist *et al.*, 1980) probably represent those bases most intimately involved in polymerase binding and orientation via sigma factor, so that RNA chain initiation can take place just downstream.

Transcription termination is also controlled by signals in the DNA sequence, characteristically a GC rich region having a two-fold symmetry before the termination site, followed by an AT rich sequence at the site of termination (Rosenberg and Court, 1979). Several protein factors are also involved in the control of termination, most notably the rho factor. Anti-termination proteins such as the N gene product of phage λ can also be involved in specialised systems (Greenblatt *et al.*, 1981).

B　Translation

Efficient translation of mRNA in prokaryotic cells requires the presence of a ribosome binding site (rbs). For most *E. coli* mRNAs the rbs consists of two components, the initiation codon AUG and, lying 3–12 bases upstream, a sequence of 3–9 bases called the Shine-Dalgarno (SD) sequence complementary to the 3' end of the 16S rRNA (Shine and Dalgarno, 1975). It is believed that hybridization to this region is involved in the attachment of the ribosomal 30S subunit to the mRNA (Steitz, 1979). The SD sequence is not identical in all mRNAs but a semi-conserved consensus sequence has been identified just as for promoter sequences. It is possible that differences in SD sequences form part of a translational control system. In addition, ribosome binding is probably modulated by the secondary structure at the 5' end of the RNA since more efficient translation occurs if the AUG and SD sequence are freely accessible to 30S ribosomal subunits (Iserentant and Fiers, 1980). Termination of translation usually occurs whenever one of the three stop codons is encountered in the mRNA by a ribosome complex, provided that an aminoacylated suppressor tRNA is not present.

C　Post-translational modification

There are a variety of modifications that bacterial proteins can undergo following translation. The formyl group on the NH_2-terminal methionine is hydrolysed and one or more NH_2-terminal residues may be removed. Many secreted proteins are synthesized as large precursors with additional hydrophobic NH_2-terminal signal sequences that are cleaved off by a membrane bound enzyme (for review, see Davis and Tai, 1980). However, glycosylation and

phosphorylation, which are common modifications of proteins in eukaryotic cells do not occur to any great extent in *E. coli*.

III   Problems encountered in the expression of eukaryotic DNA in *E. coli*

Successful expression of a eukaryotic gene in *E. coli* requires that the cellular machinery is organised so that the level of expression of the cloned gene is as good or better than the resident genes. Probably the most important difference between eukaryotic genes (at least from higher organisms) and prokaryotic genes is the presence of intervening sequences (introns) which interrupt the coding sequences. Normally these sequences are spliced out of the initial RNA transcript, producing cytoplasmic mRNA suitable for translation. There are no introns in prokaryotic genes and consequently no splicing enzymes present, so in general genomic DNA cannot be used as a source of genes for expression in bacterial cells. A second problem is that transcriptional signals in eukaryotes are different from those in prokaryotes (Corden *et al.*, 1980; Breathnach and Chambon, 1981) and are not usually recognised by bacterial RNA polymerase. This difference again emphasizes the fact that eukaryotic genomic DNA is not a suitable gene source for construction of expression vectors. Thirdly, the structure of eukaryotic mRNA is different to bacterial mRNA. Eukaryotic mRNA is polyadenylated at the 3′ end and normally capped at the 5′ end, features which may affect mRNA stability and ribosome binding (Breathnach and Chambon, 1981). Furthermore eukaryotic mRNA does not seem to have an equivalent of the SD sequence present in prokaryotic mRNA (Kozak, 1981).

An additional problem is that of codon usage. The codons used in mRNA coding for highly expressed prokaryotic genes are not random; there is a marked preference for particular codons for some amino acids (Grantham *et al.*, 1981; see Grosjean and Fiers, 1982). This preference appears to correlate with the abundance of different tRNA species (Ikemura, 1981). As codon selection preferences are different for eukaryotic genes it is possible that the levels of certain tRNAs will affect translational efficiency of these genes in a prokaryotic system. Finally, it is known that many eukaryotic proteins are subject to a number of post-translational modifications which may affect either activity or stability. Most of these modifications do not occur in prokaryotes.

A number of strategies have been developed to try to overcome these difficulties (Table 1). Once the amino acid sequence of a

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

fastcase®
Smarter legal research.