Expression of eukaryotic genes in *E. coli*

T. J. R. HARRIS

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

GENETIC ENGINEERING 4
ISBN 0-12-270304-9

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

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

Table 1 General strategies for the expression of cloned genes in *E. coli*.

Control level	Strategy
Gene	Synthesise DNA *in vitro* by chemical methods, with optimised codon assignments or obtain cDNA clone to specific mRNA. Chemical DNA synthesis probably required for tailoring genes into expression vector.
Transcription (Initiation and termination)	Clone gene adjacent to strong *E. coli* promoter which is controllable so that transcription can be induced (derepressed) when required. Use a multicopy plasmid to increase gene dosage. Include termination signal after gene to prevent transcriptional read-through.
Translation (Initiation)	Fuse gene in correct translational reading frame to an *E. coli* gene already in the vector, so that normal rbs is maintained. Possible to use both long and short NH_2-terminal fusions. Alternatively, place new gene with its own AUG adjacent to an rbs. The sequence of the SD sequence and distance from the initiating AUG may modulate translation. Accessibility (secondary structure) around SD-AUG may be important. Codon usage can be overcome by using chemically synthesized genes. Not clear if codon bias actually affects the translation of cloned genes. Include stop codon(s) in chemically synthesised genes.
Protein (Secretion and stability)	Use signal sequences to control secretion? Synthesis of precursor proteins followed by their processing ensures removal of NH_2-terminal initiating methionine. Factors affecting folding of foreign proteins and their degradation in *E. coli* are not well defined. Synthesis of long fusion proteins may result in increased stability.

protein is known it is now a relatively straightforward task to design and synthesize chemically, a DNA sequence that will code for the protein without the problem of intervening sequences and with optimized codon assignments. A gene of 514 bp coding for leukocyte Le(α) interferon, a protein of 166 amino acids is the longest DNA sequence that has been synthesized so far (Edge *et al.*, 1981). Although there is no theoretical limit to the size of gene that can be synthesized, practical problems arise for much larger proteins. If the gene is too big for a chemical synthesis, then double stranded DNA copies of mRNA populations can be generated, cloned into a plasmid vector and the clone containing the sequence coding for the protein

of interest selected from the clone bank by hybridization techniques.

Transcription of these genes is controlled by inserting the DNA adjacent to a strong prokaryotic promoter in a cloning vector. Four promoters have been used most widely for this purpose, the *lac* promoter from the *E. coli lac* operon; the *trp* promoter from the *E. coli trp* operon; the strong leftward promoter of phage λ (P_L) and the constitutive and weaker β-lactamase promoter present in the plasmid vector pBR322. The expression vectors themselves are usually derived from high copy number plasmids so that there is increased expression owing to gene dosage (Gelfand *et al.*, 1978; O'Farrell *et al.*, 1978). Termination of transcription can be ensured by placing a termination site after the cloned gene (e.g. Nakamura and Inouye, 1982) although whether this is necessary for the maintenance of high levels of transcription is not yet clear. The consequences of uninterrupted transcription around a small circular plasmid DNA molecule are unknown. It is presumably detrimental since most expression vectors have other genes present (e.g. an antibiotic resistance gene) which are transcribed in the opposite direction from a different promoter and it is known that the transcription of genes in λ phage carrying the *trp* promoter is adversely affected if the *trp* promoter is in an orientation where transcription occurs towards transcripts arising from the P_L promoter (Hopkins *et al.*, 1976).

Translational barriers have been overcome to some extent by two procedures. The foreign gene is either fused (in the correct translational reading frame) to a prokaryotic gene so that the existing rbs is used to initiate translation, or the new gene, with its own initiation codon, is placed adjacent to a naturally occurring *E. coli* rbs (Backman *et al.*, 1980) or a synthetic one (Jay *et al.*, 1981). Since all structural genes, whether eukaryotic or prokaryotic, end with one or more of the three termination codons it is not usually necessary to make special arrangements for translational termination when using a cloned cDNA sequence. However, a termination codon must be included when synthetic DNA is used.

Protein modification and stability are much less easy to control, largely because the structural features governing protein stability in *E. coli* are not well understood. It has been shown that eukaryotic signal sequences are recognised by *E. coli* and that NH_2-terminal fusions of eukaryotic polypeptides to *E. coli* signal sequences results in secretion of the protein to the periplasmic space, with concomitant cleavage of the signal sequence (Talmadge *et al.*, 1980; 1981). There is also some evidence that short "foreign" polypeptides are unstable in *E. coli* (Itakura *et al.*, 1977; Goeddel *et al.*, 1979a). This has been

overcome by fusing the peptide to a larger *E. coli* protein from which the peptide is then cleaved.

IV Expression of DNA from lower eukaryotes

Following the observation that DNA from *S. aureus* could be expressed in *E. coli* (Chang and Cohen, 1974) it was shown that eukaryotic DNA could also be transcribed (Morrow *et al.*, 1974; Chang *et al.*, 1975; Kedes *et al.*, 1975). It was not clear from these experiments, however, whether the normal transcriptional start and stop signals were being recognised. The fundamental question of whether a fungal gene could be transcribed and translated to produce a functional protein in *E. coli* was answered to some extent by the finding that fragments of yeast DNA cloned into phage λ, or the plasmid vector Col E1 could complement auxotrophic mutants of *E. coli* (e.g. His *B* and Leu *B*) (Struhl *et al.*, 1976; Ratzkin and Carbon, 1977; Struhl and Davis, 1977). Similarly segments of *Neurospora crassa* DNA containing the gene for dehydroquinase were successfully expressed in *E. coli* in a pBR322 replicon (Vapnek *et al.*, 1977). Several other yeast genes have now been expressed in this way (e.g. Trp 1, Trp 5 and Arg 4). The functional expression of yeast DNA in *E. coli* not only demonstrated that eukaryotic DNA could be transcribed and translated, paving the way for the experiments described below, but also provided a powerful method for isolating yeast genes. Some of these genes have subsequently been used to provide selection markers in yeast—*E. coli* shuttle vectors (Beggs, 1982; Hinnen and Meyhack, 1982).

The *lac* promoter

The *lac* operon is subject to two types of control. In the absence of lactose (or other inducer) the operon is kept switched off by *lac* repressor (the *lac* i gene product) binding to the operator. Positive regulation is also exerted through the catabolite gene activator protein (CAP). In the absence of glucose, CAP forms a complex with cyclic AMP and this complex stimulates transcription by binding next to the promoter. The operon is derepressed by the presence of lactose, or by the addition of the non-metabolizable inducer IPTG (isopropylthiogalactoside) which binds to the repressor and removes it from the operator.

Plasmid vectors containing parts of the *lac* operon have been constructed by several workers. Polisky *et al.* (1976) cloned an *Eco*RI fragment from λ p *lac* 5 DNA (a transducing phage containing part of the *lac* operon) into a Col E1-derived plasmid to obtain a vector with the *lac* promoter and operator and most of the β-galactosidase gene. Plasmids containing a small "portable" *lac* promoter fragment have also been made. In these constructions a 203 bp *Hae*III fragment of *lac* transducing phage DNA, containing the *lac* promoter and operator and first eight codons of β-galactosidase, was blunt end ligated into *Eco*RI-cut and "filled in" pBR322 DNA. The portability derives from the fact that *Eco*RI sites are reformed at the junctions allowing the promoter fragment to be removed by *Eco*RI digestion (Backman and Ptashne, 1976; Itakura *et al.*, 1977). Colonies harbouring plasmids which carried the *lac* promoter-operator were identified by their constitutive synthesis of β-galactosidase, rendering them blue on agar plates containing X gal (5 chloro-4 bromo 3 indolyl-D galactoside). This is because multiple copies of the operator titrate out all the *lac* repressor resulting in derepression of the chromosomal β-galactosidase gene. Both λ p *lac* 5 DNA and λ h80 *lac* UV5 C_I857 DNA, which contains the CAP site mutation L8 and the up promoter mutation UV5 (making the promoter insensitive to catabolite repression), have been used as a source of *lac* DNA for these constructions (Backman *et al.*, 1976; Itakura *et al.*, 1977; see also Fuller, 1982). Further derivative plasmids containing a 95 bp *Alu*I fragment of *lac* DNA, including the UV5 promoter (minus the CAP binding site), the repressor binding site and most of the rbs, just excluding the ATG of β-galactosidase, have also been constructed for the expression of non-fusion proteins (Fuller, 1982).

A The somatostatin experiment

The first report of the designed expression of a eukaryotic gene in *E. coli* was the production of the small peptide hormone somatostatin (Itakura *et al.*, 1977). Somatostatin was used as a model system because the hormone was a small polypeptide of known amino acid sequence for which sensitive radioimmune and biological assays existed. The experiments illustrate a number of features of methods which are now used to obtain expression of cloned genes. They also demonstrated, although not for the first time, that chemically synthesized DNA was functional in a biological system. In addition, the production of the protein as a fusion polypeptide and its subsequent cleavage into the native hormone at methionine residues by cyanogen bromide (CNBr), has been used quite extensively for other proteins. This overall strategy is depicted in Fig. 1.

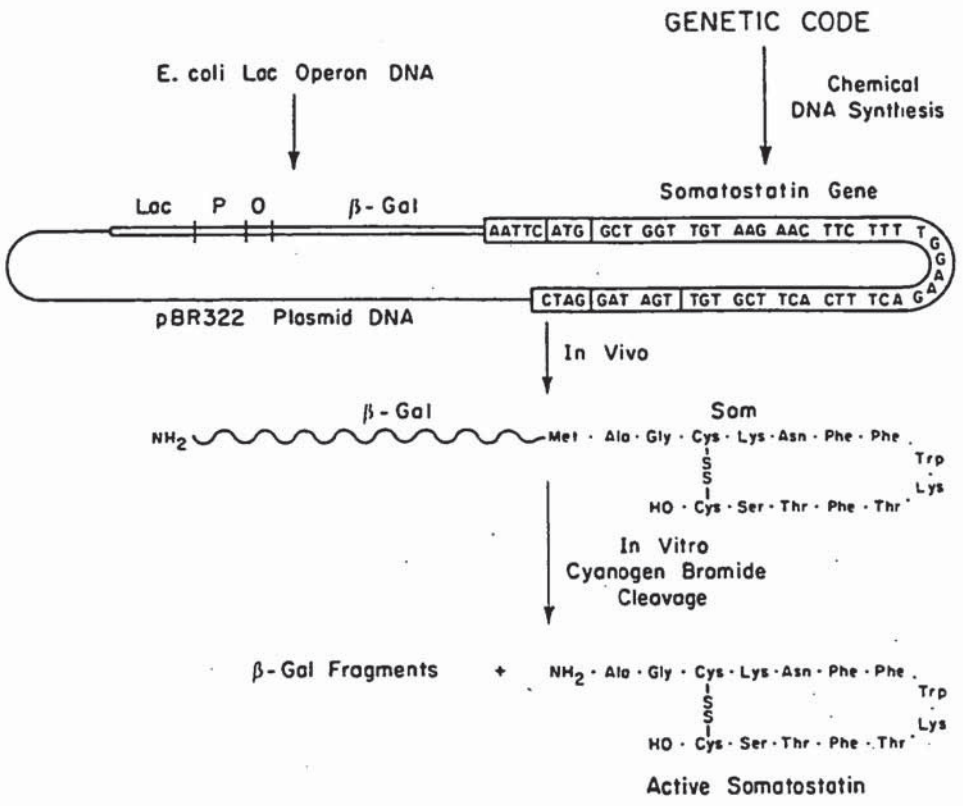


Figure 1 Strategy for the expression of the chemically synthesized somatostatin gene as a β-galactosidase fusion from the *lac* promoter. The active hormone can be cleaved from the hybrid protein by CNBr treatment. (Reproduced from Itakura *et al.*, 1977, copyright by the American Association for the Advancement of Science, with permission.)

In the first set of experiments the chemically synthesized somatostatin gene with synthetic *Eco*RI and *Bam*HI cohesive ends was cloned into a vector containing the wild type *Hae*III *lac* promoter fragment. The DNA sequence indicated that the plasmid should have produced a polypeptide containing the first seven amino acids of β-galactosidase fused to somatostatin. However, no somatostatin was detected in bacterial extracts by radioimmunoassay. As it was found that somatostatin was not stable when added to *E. coli* extracts, the failure to find somatostatin was thought to be due to proteolytic digestion (Itakura *et al.*, 1977). The approach adopted to try to stabilise the somatostatin was to produce it as part of a longer polypeptide from which it could be cleaved by CNBr. This was done by linking the somatostatin gene to the *Eco*RI fragment of λ p *lac* 5 DNA which carries the *lac* promoter and a large proportion of the β-galactosidase gene (Polisky *et al.*, 1976). The translation reading

frame of β-galactosidase was maintained in somatostatin after fusion at the *Eco*RI junction. In these constructions only one orientation of the *Eco*RI *lac* fragment maintained the correct reading frame in somatostatin and when several independent clones were examined, about half produced detectable somatostatin after CNBr cleavage. No immunoreactive protein was detected before cleavage since the antiserum used in the assay required a free NH_2-terminal alanine residue (Itakura *et al.*, 1977).

B Expression of insulin in *E. coli*

The somatostatin work established the feasibility of the synthetic gene fusion approach for the expression of small polypeptides in *E. coli*. It was possible to follow an almost identical strategy to obtain expression of human insulin, as neither the 20 amino acid A chain nor the 30 amino acid B chain of insulin contained methionine and methods were available for the *in vitro* joining of the two chains. Thus, an A chain gene and a B chain gene were chemically synthesized each with *Bam*HI and *Eco*RI cohesive ends (Crea *et al.*, 1978) and cloned separately into pBR322. The B chain gene was synthesized with a *Hind*III site in the middle so that the two halves could be cloned separately and the sequence verified (Goeddel *et al.*, 1979a). Expression was achieved by transcription from the same *lac* promoter as used for the successful somatostatin constructions and insulin A or B-β-galactosidase fusion proteins were produced (Goeddel *et al.*, 1979a). The hybrid proteins represented about 20% of total cell protein, which was about ten-fold higher than the level of expression obtained with somatostatin. The hybrid proteins were insoluble and were found in the first low speed pellet after breaking the cells with a French press.

To obtain A and B peptides suitable for reconstitution into native insulin, the hybrid proteins had to be solubilised, the β-galactosidase portion removed and the peptides S-sulphonated. This was achieved by dissolving the hybrid proteins in 6 M guanidinium chloride followed by dialysis. The precipitate was dissolved in 70% formic acid, the protein cleaved with CNBr and S-sulphonated derivatives of the peptide mixture obtained, using sodium dithionate and sodium sulphite at pH 9. Insulin activity was readily detected by radioimmunoassay after re-constitution. Further studies on the peptides (e.g. chromatographic behaviour) and amino acid compositions established, without doubt, that the bacteria were producing authentic insulin A and B chains (Goeddel *et al.*, 1979a). Insulin, prepared from bacteria containing these constructions by a scaled up and modified process, has now been shown to be active

when injected into human volunteers (Clark *et al.*, 1982) and to interact with insulin receptors in the same way as native human insulin (Keefer *et al.*, 1981).

An alternative approach involves the synthesis of a gene coding for proinsulin, the natural precursor to insulin. Proinsulin is synthesized initially as a preproinsulin molecule consisting of an NH_2-terminal signal sequence, followed by the B chain, a linking C peptide and the COOH-terminal A chain. Enzymatic removal of the signal peptide during secretion generates proinsulin and processing at two trypsin sensitive sites (Arg-Arg, Lys-Arg) allows the removal of the C peptide and the generation of active insulin. The three dimensional structure of insulin indicates that a peptide much shorter than the 35 amino acid connecting C peptide should be sufficient to connect the B and A chains and still allow proper folding of the modified proinsulin. Genes coding for human proinsulin and "mini C" derivatives of proinsulin, where the C peptide is replaced by a six amino acid linker retaining the proteolytic cleavage sites, have been constructed by chemical synthesis (Sung *et al.*, 1979; Wetzel *et al.*, 1981a; Brousseau *et al.*, 1982).

The mini C construction was cloned for expression as a β-galactosidase fusion protein (Wetzel *et al.*, 1981a) and a product with a proinsulin-like structure (as determined by radioimmunoassay and HPLC) was detected after CNBr cleavage and S-sulphonation. The usefulness of this route to insulin production is still not clear however, as there are no data on the behaviour of mini C derivatives in enzymatic proinsulin processing systems and there are already preproinsulin expression constructions available derived from cDNA (see β-lactamase section). However, the modular approach to the chemical synthesis of proinsulin adopted by Brousseau *et al.* (1982) does have the advantage that the shortening and changing of parts of the C peptide or alteration of the codons can be approached rationally by the incorporation of different oligonucleotide blocks during synthesis, obviating the need to synthesise an entire coding sequence each time a specific modification is made.

C Synthesis of other hormones as β-galactosidase fusions

The strategy of using the *lac* promoter/operator and β-galactosidase NH_2-terminal fusions has been adopted for several other proteins including other hormones (see Table 2). For example the neuropeptide β-endorphin, a 30 amino acid endogenous opioid has been expressed in this way (Shine *et al.*, 1980). In these experiments a cDNA clone to the precursor peptide of mouse corticotropin (ACTH) and β-lipotropin (LPH) was used as a source of cDNA coding for

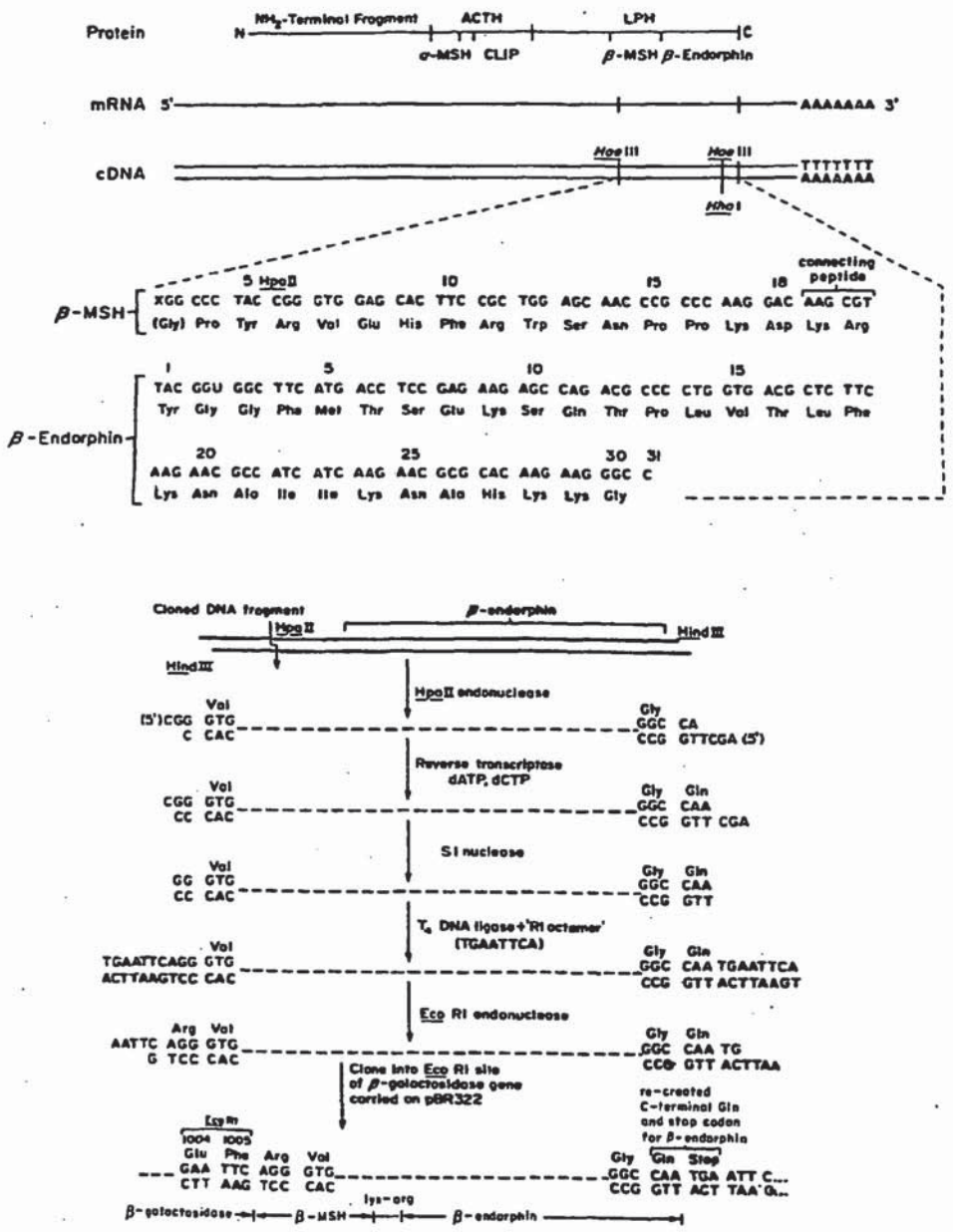


Figure 2 (a) Nucleotide sequence of the cloned fragment of cDNA to ACTH β-endorphin mRNA (b) β-galactosidase fusion construction used for the expression of the hybrid β-endorphin protein. Reprinted by permission from Shine *et al.*, *Nature* 285, 456–461. Copyright © 1980, Macmillan Journals Ltd.

β-endorphin (Roberts *et al.*, 1979a). The cDNA fragment contained all the coding sequence except the C-terminal glutamine. For the expression constructions it was necessary to recreate the C-terminal codon, insert a stop codon and link the gene in phase to β-galactosidase. This process is illustrated in Fig. 2. The cloned

*Hind*III fragment containing the β-endorphin coding sequence was cleaved with *Hpa*II at one end and the cohesive ends filled in using reverse transcriptase in the presence of dATP and dCTP, so that only partial filling in occurred. This step regenerated the C-terminal glutamine. The remaining overhanging ends were then removed with S1 nuclease and the fragment blunt end ligated to a chemically synthesised linker containing a stop codon and an *Eco*RI site (Fig. 2). Subsequent *Eco*RI digestion generated a fragment which could be linked in phase to β-galactosidase. As previously found for the insulin constructions, the hybrid β-endorphin/β-galactosidase hybrid protein was insoluble but represented a substantial proportion of total protein in the pellet obtained after disruption and centrifugation (Shine *et al.*, 1980).

Since β-endorphin contains a methionine residue at amino acid 5, CNBr cleavage could not be used to cleave the hybrid protein. An alternative strategy was developed based on the fact that an arginine residue, which is a site for trypsin cleavage, is present in β-melanocyte stimulating hormone, the peptide preceding β-endorphin, but not in β-endorphin itself and that the lysine residues in β-endorphin can be protected from trypsin attack by citraconylation (which is reversible). Thus, after dissolving the precipitated hybrid protein and treatment with citraconic anhydride at pH 9, the modified β-endorphin was cleaved from the hybrid protein by trypsin digestion. Various immunological and biological criteria showed that authentic active murine β-endorphin had been synthesized (Shine *et al.*, 1980). Since the murine protein differs from the human protein in only two positions (tyr for his at position 27, glu for gln at position 31) it is possible that a clone expressing human β-endorphin could be made by altering the codons in these two places, using oligonucleotide site directed mutagenesis (Smith and Gillam, 1981). It seems unlikely however, that a trypsin cleavage protocol of this kind could be economically or efficiently used on a large scale.

D Expression of ovalbumin

The *lac* promoter has also been used to express an oval bumin cDNA clone. Charnay *et al.* (1979) cloned the promoter fragment back into λ p *lac* 5-1 giving a molecule with two *lac* regulatory regions. In one orientation intramolecular recombination occurred generating a phage with only one *Eco*RI site downstream from the *lac* promoter, after the 8th codon of β-galactosidase (λ p *lac* UV5, Charnay *et al.*, 1979). The *lac* control region from this phage was transferred back to pBR322 to create a plasmid (pOMPO) with an *Eco*RI site into which was blunt end ligated a *Hha*I fragment of an ovalbumin cDNA

clone (Mercereau-Puijalon *et al.*, 1978). In the correct orientation, the translation frame was maintained to generate an ovalbumin fusion protein where the NH_2-terminal five amino acids of ovalbumin were replaced by the NH_2-terminal eight amino acids of β-galactosidase. An ovalbumin-like protein was detected by radioimmunoassay and by polyacrylamide gel electrophoresis of immunoprecipitates of ^{35}S methionine labelled bacterial extracts. In a bacterial strain which overproduces *lac* repressor (repressing the multiple copies of the operator) the synthesis of the hybrid protein was stimulated over 50-fold by the addition of IPTG, showing that the synthesis was under *lac* control. Similar results were reported by Fraser and Bruce (1978), although in their construction the fusion protein contained an additional 18 amino acids at the NH_2-terminus, and was apparently secreted into the periplasmic space. Baty *et al.* (1981) have confirmed that ovalbumin made as a short β-galactosidase fusion protein in *E. coli* is transported to the periplasmic space and have shown further that a derivative of the protein lacking the NH_2-terminal 126 amino acids remains in the cytoplasm.

These were the first reports of the synthesis of a large ($> 40\,000$ molecular weight) eukaryotic protein in *E. coli*. Although there did not appear to be any overriding stability problems for either the plasmid containing the gene, or for the protein, only about 10% of the theoretical yield of ovalbumin (based on native β-galactosidase synthesis) was produced. This could have been due to inefficient transcription or translation or to proteolytic degradation. Another possible reason for the low level of expression was that the codons in the gene derived from ovalbumin mRNA were not optimal for *E. coli* and the availability of certain minor isoaccepting tRNAs could therefore have been limiting (Mercereau-Puijalon *et al.*, 1978).

The production of ovalbumin in *E. coli* was obviously dependent on maintaining the correct translation frame across the β-galactosidase-ovalbumin junction. Charnay *et al.* (1978) constructed a set of plasmid and λ phage vectors allowing fusion of cloned genes in each of the three translational phases. This was achieved by treating the *Eco*RI *lac* UV5 promoter fragment with S1 nuclease and ligating a synthetic octanucleotide linker, containing an *Eco*RI site, to the blunt ends. Cutting with *Eco*RI generated a fragment with two additional base pairs before the cohesive end. Repeating the procedure generated a fragment with a further two base pairs. Considering the original vector to give translation in frame 1, then the two new constructs will give translation frame 2 and frame 3 respectively when fused to the same *Eco*RI fragment (Charnay *et al.*, 1978; 1979).

E Expression of native proteins

In all the constructions examined so far the proteins have been produced from fusions with parts of β-galactosidase. Backman and Ptashne (1978) extended their studies with the λ repressor by making a construction where the λ C_1 gene was fused to the β-galactosidase rbs, rather than within the coding region, forming a hybrid β-galactosidase-C_1 gene SD sequence 8 bases from the AUG of the λ repressor. This construction led to the synthesis of about 30 000 molecules of λ repressor per cell; other fusions which placed the *lac* promoter further away from the C_1 gene reduced the synthesis of repressor quite considerably. Much larger amounts of protein were produced when *lac* rbs fusions were made to the λ *cro* gene by the same group (Roberts *et al.*, 1979b; Lauer *et al.*, 1981). This has allowed a more systematic analysis to be carried out of the effect on expression of changing the distance between the promoter and the gene (Roberts *et al.*, 1979b).

A series of plasmids were constructed in which the λ *cro* gene was placed at varying distances downstream from the *lac* promoter and SD sequence. The level of *cro* protein produced by clones containing different plasmids was then correlated with the nucleotide sequence across the *lac-cro* junction. There was considerable variation in the amount of *cro* produced by the various fusions (Fig. 3). Strains containing plasmids pTR213 or pTR214, produced about 200 000 monomers of *cro* per cell (1.6% total cell protein) calculated to be the number expected from a fully induced *lac* promoter, whereas pTR199 with almost the same number of nucleotides between the *lac* SD and the *cro* AUG directed the synthesis of only about 1/10th as much *cro* protein. Since the same promoter was being used in all the plasmids it was concluded that some post-transcriptional process was responsible for the differences in level of *cro* synthesis. It was unlikely to be caused by plasmid copy number variability nor by the regulatory effects of the *cro* protein itself, since the operators at which *cro* acts were not present (Johnson *et al.*, 1981). An explanation for the observations of Roberts *et al.* (1979b) has been advanced based on secondary structure models for the 5′ terminus of these different *cro* mRNAs (Iserentant and Fiers, 1980). The involvement of the AUG and the SD sequence in secondary structure is thought to reduce the amount of protein produced, presumably because interaction with the 30S ribosomal subunit is inhibited. Sequences which produce mRNAs with secondary structure that leave the AUG accessible (e.g. pTR213) produced large amounts of protein. Involvement of the SD sequence in base paired regions apparently caused far smaller reductions in *cro* protein

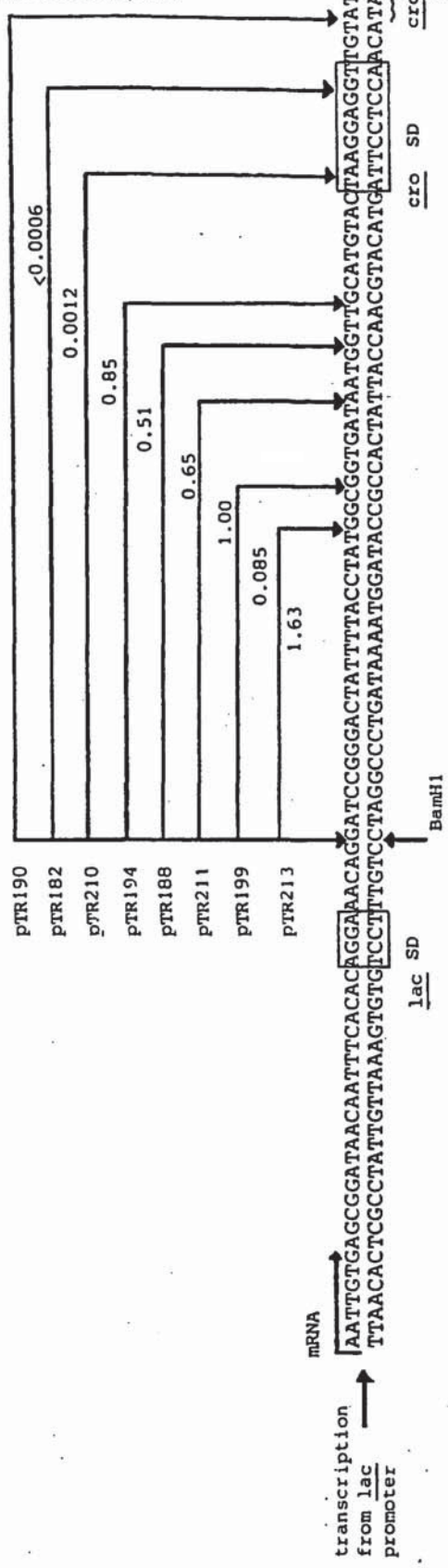


Figure 3 Correlation of nucleotide sequence deletion limits and *cro* protein production for various hybrid rbs constructions between the *lac* SD, the *cro* SD and the *cro* initiation codon. Adapted from Roberts *et al.*, 1979b, by permission.

synthesis than loss of an accessible AUG (Iserentant and Fiers, 1980).

Secondary structure in mRNA is dictated not only by the distance between the SD and AUG but also by the sequence between them and the sequence coding for the remaining N-terminal amino acids. It is not therefore, a trivial task to predict the sequence and optimal distance between the promoter and the AUG to obtain maximal expression of a particular gene. With this fact in mind, Guarente *et al.* (1980) have devised a general method for maximizing the expression of cloned genes in the absence of assays for the gene products. In this method a fragment of DNA bearing the NH_2-terminal region of a gene is fused to a DNA fragment coding for the enzymatically active COOH-terminal fragment of β-galactosidase in an analogous manner to that used to examine the control of prokaryotic promoters (Casadaban *et al.*, 1980; see also Bassford *et al.*, 1978). The portable *lac* promoter is then placed at varying distances in front of the fusion. The constructions which lead to efficient expression of the fused gene are recognised by the amount of β-galactosidase expressed. The gene is then reconstituted in the plasmids directing high levels of β-galactosidase, where an optimal promoter-SD-ATG distance should have been established (Fig. 4). The method was tested with β-globin cDNA and has since been used to obtain expression of a human fibroblast interferon (HFIF) cDNA gene (Guarente *et al.*, 1980; Taniguchi *et al.*, 1980). Plasmids were obtained directing the synthesis of HFIF and the precursor protein pre-HFIF. It was interesting that in each of these constructions the distance separating the SD sequence of the *lac* promoter and the ATG was found to be precisely that normally found between the ATG of β-galactosidase and the *lac* SD sequence. The levels of protein produced were not high, possibly owing to proteolytic degradation (Taniguchi *et al.*, 1980). Other vectors based on the *lac* promoter allowing fusion of DNA coding for enzymatically active β-galactosidase to that coding for amino-terminal fragments of exogenous proteins have also been constructed (Casadaban *et al.*, 1980).

F Expression of human growth hormone

The approach used for the expression of human growth hormone (HGH) as a native rather than a fused polypeptide demonstrates the combination of chemical synthesis and cDNA cloning (Goeddel *et al.*, 1979b). The specific strategy was based on the known restriction map of HGH cDNA. Treatment of cloned HGH cDNA with *Hae*III generated a 551 bp fragment coding for amino acids

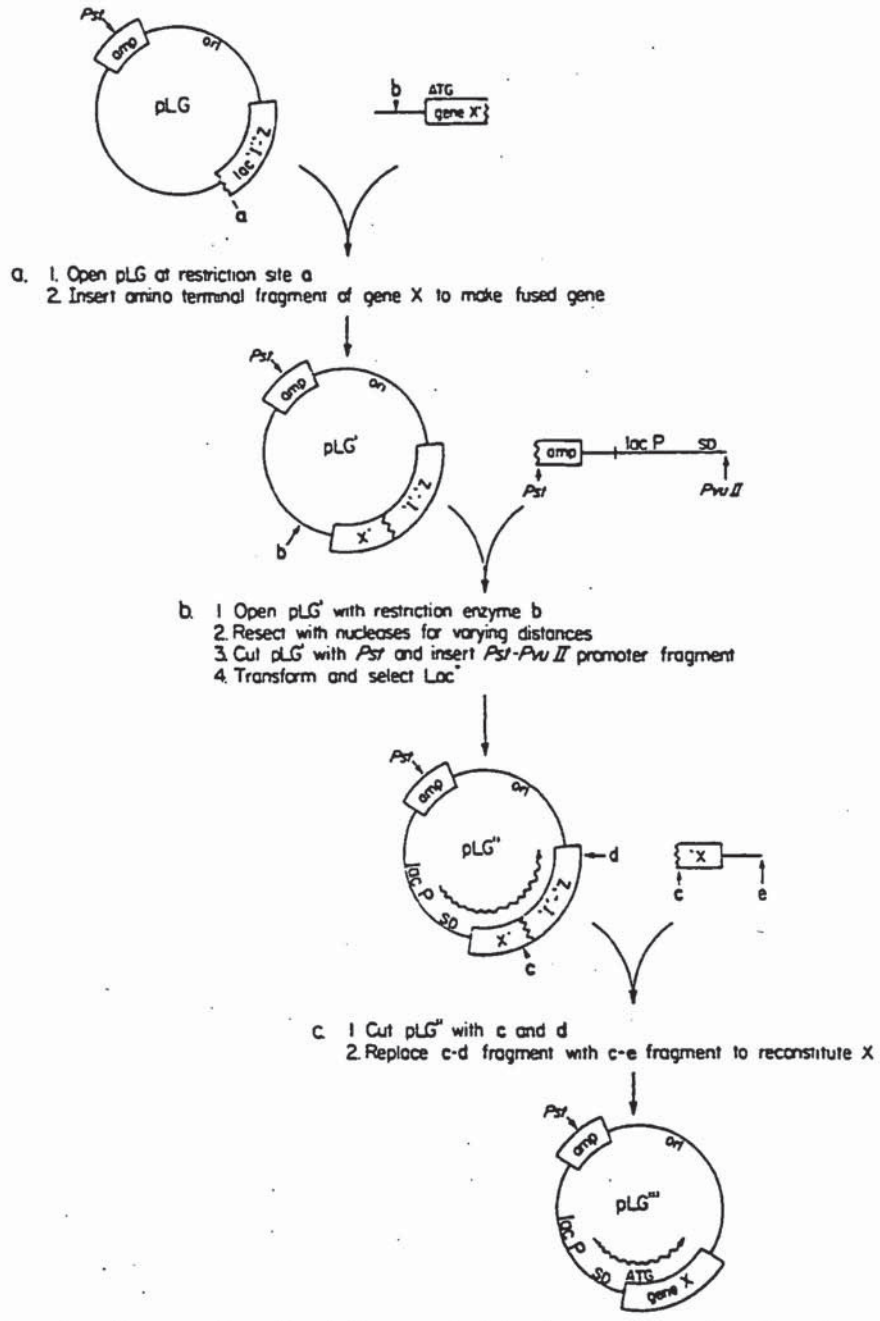


Figure 4 A general method for maximizing expression of a cloned gene in *E. coli*. Reproduced from Guarente *et al.*, 1980, *Cell* 20, 543—553, by permission from MIT press.

24—191 of HGH (Fig. 5). An adaptor fragment encoding an initiation codon and amino acids 1—24 (including the *Hae* III site found in the cDNA) was chemically synthesized. The two fragments were cloned separately, the cDNA by C tailing into the *Pst* I site of pBR322 and

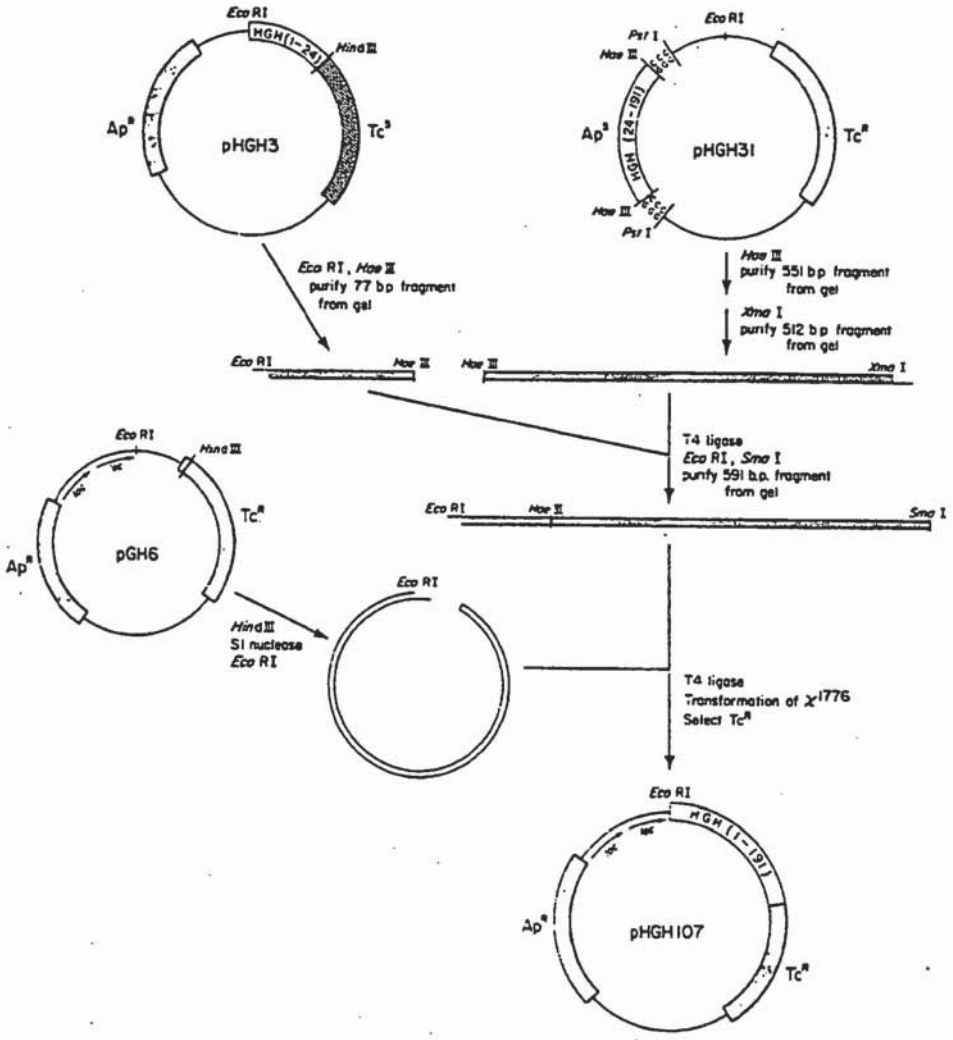


Figure 5 Construction of a plasmid for the expression of human growth hormone in *E. coli* (Goeddel *et al.*, 1979b). Reprinted by permission from *Nature* 281, 544–548. Copyright © 1979 Macmillan Journals.

the adaptor fragment into pBR322 as an *Eco*RI-*Hind*III fragment. Two fragments were isolated from these vectors (Fig. 5), a 77 bp *Hae*III/*Eco*RI fragment from pHGH3 and a 512 bp *Hae*III-*Xma*I fragment from pHGH31. The two fragments were ligated, treated with *Eco*RI and *Sma*I and the 591 bp. DNA coding for HGH isolated and inserted into an expression vector containing two copies of the *lac* promoter (Fig. 5). One of the resulting plasmids (pHGH 107) was found by nucleotide sequence analysis to have 11 base pairs separating the *lac* SD sequence and the ATG for HGH. This was reduced to seven base pairs (the naturally occurring distance for *lac* SD and β-galactosidase) by *Eco*RI digestion, S1 nuclease treatment and blunt end ligation to generate a new plasmid (pHGH107.1) (Goeddel *et al.*,

1979b). Unexpectedly, this new construction produced less immunoreactive HGH than pHGH107 with the longer ATG-SD distance. Growth hormone was readily detectable in polyacrylamide gels of the proteins of extracts of *E. coli* RV 308 (a nutritionally wild type K12 strain) carrying pHGH107.

In contrast to the insulin chain and the β-endorphin fusion polypeptides, HGH is produced as a soluble protein. Partial purification was achieved by ammonium sulphate precipitation and Sephacryl S-200 gel filtration. Preparations of HGH of high purity have now been obtained from *E. coli* carrying these plasmids and this HGH has been shown to have the same specific activity as natural HGH derived from human pituitary glands (Olson *et al.*, 1981; Hintz *et al.*, 1982; Rosenfeld *et al.*, 1982). Very precise authenticity studies have also been carried out on the *E. coli* product; HPLC of the protein and its constituent tryptic peptides has shown that bacterial HGH has the same amino acid sequence and disulphide bond arrangement as the natural hormone apart from the presence of an extra NH_2-terminal methionine (Kohr *et al.*, 1982). The possible side effects of the presence of this methionine (e.g. antigenicity) is a current cause for concern. As this will be a universal problem for bacterially derived products which do not naturally retain an initiating methionine residue, it may be necessary to devise methods for its specific removal.

VI The phage λ P_L promoter

The observation that the *trp* operon genes of *E. coli* could be expressed in phage λ derivatives under the control of the leftward promoter P_L (Moir and Brammar, 1976) suggested that this promoter could be used to drive expression of cloned DNA. In the normal phage infection cycle the P_L promoter controls early leftward transcription of the DNA through gene N to *int* (for review, see Szybalski and Szybalski, 1979). P_L is a strong promoter which is subject to various forms of control. Most importantly the promoter is subject to repression by the C_I gene product and later in infection by the *cro* protein. There is also the *Nut* L sequence downstream from the promoter which allows the N gene product, in association with RNA polymerase, to overcome transcription termination further downstream.

It is possible to obtain expression from P_L either by inserting DNA into the phage itself or by cloning the promoter into a suitable plasmid vector. There have been several reports of the use of hybrid λ phages for expressing prokaryotic genes. In most of these studies the genes have been cloned with their own promoter and oversynthesis

occurs primarily due to the increased copy number attained by phage DNA replication (Panasenko *et al.*, 1977; Hopkins *et al.*, 1976; Kelley *et al.*, 1977). Studies with the *E. coli* DNA polymerase gene (pol A) cloned in a λ phage derivative under P_L control have illustrated some of the problems concerning the use of the P_L promoter within a phage DNA molecule. For example the *cro* gene product, which is required for efficient DNA replication, represses transcription from P_L preventing sustained transcription (Murray and Kelley, 1979). Both P_L and other promoters such as the late rightwards promoter (P'_R) have, however, been used to obtain effective expression of tryptophan synthetase (Moir and Brammar, 1976) and T4 DNA ligase (Murray *et al.*, 1979).

Very few studies have been done concerning the expression of eukaryotic genes from the P_L promoter within phage λ. Kourilsky *et al.* (1977) reported that β-globin cDNA is expressed when cloned into the early region of the phage and low levels of interferon have been detected in *E. coli* infected with a Charon 4A recombinant phage containing a fragment of the human genome (Mory *et al.*, 1981). In addition, the wheat chloroplast gene for the large subunit of ribulose bisphosphate carboxylase has been expressed from P_L when cloned into a phage λ derivative (Gatenby *et al.*, 1981).

Most of the studies have been done using cloned λ promoters because the problem of control of expression of the promoters (by using mutations in other phage genes) can be largely avoided if the promoter itself is cloned in a plasmid vector. One of the potential advantages of the P_L promoter is that, in contrast to the *lac* promoter, there is sufficient λ repressor produced from a single copy of the C_I gene to repress transcription from multiple plasmid-borne copies of P_L. By using a temperature sensitive C_I gene (C_I857), it is possible to control transcription from P_L such that at 28°C it is entirely switched off but at 42°C it is fully induced. (The C_I857 gene, maintained in the host chromosome or in a compatible plasmid, produces a repressor which is inactivated at 42°C.) As an example, Bernard *et al.* (1979) showed that heat induction of plasmids containing the *trp* A gene downstream from P_L, controlled by C_I857, produced up to 6.6% of soluble cell protein as tryptophan synthetase.

Remaut *et al.* (1981) have also described plasmid expression vectors based on λ P_L. All these vectors incorporated a 247 bp DNA fragment from the phage containing the operator-promoter region of P_L and 114 nucleotides of the P_L transcript, excluding the initiation site for N protein, but including the *Nut* L site. Unique restriction sites for *Eco*RI, *Bam* HI and *Hind* III were present further downstream (Fig. 6). P_L activity was controlled by maintaining the plasmids in partial λ lysogens containing a chromosomal C_I857 gene but no *cro*

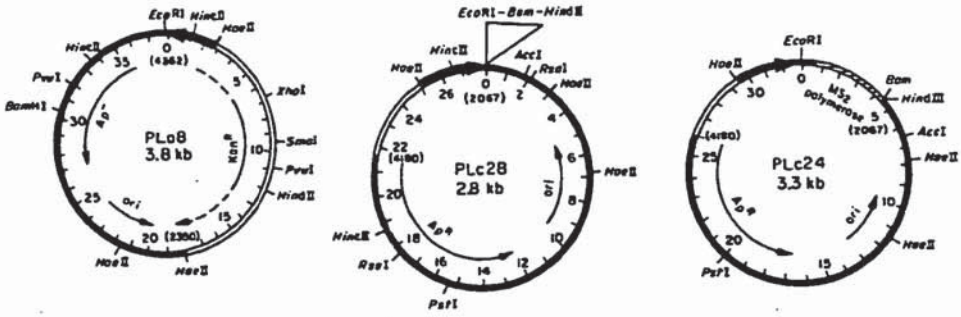


Figure 6 Restriction and genetic maps of representative P_L vectors. The position of the O_LP_L region present on the vectors is indicated by a heavy arrow showing the direction of transcription. The heavy solid lines represent sequences derived from pBR322. Ap^R indicates the region coding for β-lactamase. The direction of translation is shown by an arrow. Kan^R indicates the region carrying the resistance to kanamycin. Ori indicates the region of the origin of replication. The direction of replication is shown by an arrow. Only some particularly relevant restriction sites are shown. The small *Hae*II fragment constituting part of the origin of replication originated from Col E1 in the case of PLa plasmids and from pBR322 in the case of PLc plasmids. Taken from Remaut *et al.* (1981) by permission.

gene (K-12ΔHIΔtrp). In one of the strains (M5219) a functional N gene was also present in the chromosome providing an additional potential control in *trans* by antitermination.

In strains with P_L plasmids containing the β-lactamase gene from pBR322 or the *trp* A gene of *E. coli*, synthesis of high levels of β-lactamase and tryptophan synthetase (up to 10% of total cell protein) could be demonstrated, particularly when the promoter was cloned into the pBR322 replicon and transcribing in the clockwise direction in M5219 (see Fig. 6) (Remaut *et al.*, 1981). This was a somewhat higher level of *trp* A than that reported by Bernard *et al.* (1979) where N protein was expressed from the plasmid rather than from the chromosome.

Large amounts of the λ C_{II} protein, a phage regulatory protein toxic to *E. coli*, have also been produced by vectors containing the P_L promoter (Shimatake and Rosenberg, 1981). The C_{II} gene was cloned into a *Hpa*I site located 321 bp downstream from the P_L transcription start site. In the correct orientation a λ lysogen could be transformed at high efficiency, whereas cells not making λ repressor could not be transformed. In strains containing a chromosomal C_I857 gene, the plasmid directed the synthesis of about 4% total cell protein as the C_{II} gene product on temperature shift induction.

A Expression of eukaryotic genes from P_L plasmids

Vectors based on P_L have been used to express a variety of eukaryotic genes. Derynck *et al.* (1980) cloned two HFIF cDNA genes into the vector containing the β-lactamase gene (pPLa8, Fig. 6) to produce a plasmid (pPLa HFIF-67-12Δ19) containing the HFIF cDNA sequence fused in phase to the β-lactamase gene. The predicted hybrid protein contained 83 amino acids of β-lactamase followed by the methionine of the signal sequence of HFIF. In a second construction (pPLc HFIF-67-8) the acceptor plasmid was pPLc24 (Fig. 6) which contains the P_L promoter followed by an *Eco*RI-*Bam*HI fragment containing the rbs and part of the MS2 phage RNA polymerase gene. In this case, the expected fusion protein consisted of the N-terminal 98 amino acids of MS2 polymerase, 27 amino acids coded for by the linking sequences, followed by the methionine of the signal sequence of HFIF. Antiviral activity could be detected in cleared lysates of bacterial strains containing both constructions after induction at 42°C. Lysis with a mixture of SDS, β-mercaptoethanol and urea resulted in a 10—20 fold increase in activity (HFIF is stable under these conditions), suggesting that some non-specific protein aggregation was occurring. Although the levels of synthesis were low, the solubilized polypeptides had physicochemical, biological and immunological properties resembling HFIF. Polyacrylamide gel electrophoresis suggested, in addition, that some post-translational cleavage of the fusion proteins had occurred generating native HFIF (Derynck *et al.*, 1980).

The small t antigen (tAg) of SV40 has also been expressed in these vectors (Derom *et al.*, 1982). This protein was chosen largely because the role of the tAg in SV40 transformation was equivocal and insufficient protein was available from SV40 transformed cells for detailed biochemical study. In these constructions a DNA segment functioning as an rbs was inserted between P_L and the tAg gene so that a native rather than a fusion protein was produced. The best construction expressed 2.5% of *de novo* protein synthesis as authentic tAg. This level is better than the level of tAg produced in *E. coli* from vectors containing the tAg gene fused to the *lac* promoter, and utilizing a hybrid rbs (Roberts *et al.*, 1979c; Thummel *et al.*, 1981).

In addition to the 19 000 molecular weight small t protein several of the plasmids expressed a related 14 500 molecular weight polypeptide (Derom *et al.*, 1982). Shorter forms of small t were also found in bacteria containing *lac*-small t fusions (Roberts *et al.*, 1979c; Thummel *et al.*, 1981). Tryptic peptide maps indicated that the shortened polypeptides lacked N-terminal peptides suggesting that they arose by initiation at an internal AUG (Thummel *et al.*, 1981;

Derom *et al.*, 1982). In agreement with this conclusion secondary structure maps of the mRNA transcribed from plasmids producing the short tAg indicated the presence of a freely accessible AUG (Gheysen *et al.*, 1982). Systematic alteration of the distance and/or nucleotide sequence between the SD sequence and the AUG of tAg in these plasmids has provided additional evidence that efficient initiation of translation requires an accessible AUG and SD sequence (Gheysen *et al.*, 1982; see also Iserentant and Fiers, 1980).

VII The *trp* promoter

A Construction of vectors

Hershfield *et al.* (1974) showed that high levels of expression of the five genes of the *trp* operon occurred when the operon was transferred to a plasmid vector (for a review of the more molecular aspects of the *trp* operon, see Yanofsky *et al.*, 1981). With these plasmids, the *trp* enzymes accounted for about 20—25% of total cellular protein after induction with 3-indolylacrylic acid (IAA). Moreover, in contrast to the *lac* system, in the absence of inducer, a single copy of the *trp* repressor gene produced sufficient protein to keep the operon fully repressed, despite the increased operator copy number. These observations led to the construction of expression vectors where genes are inserted under the control of the *trp* operon regulatory elements (Hallewell and Emtage, 1980). A 5.4 kb *Hind* III fragment of *E. coli* containing the *trp* promoter, operator, leader and attenuator, all of the *trp* E gene and part of the *trp* D gene was cloned into the *Hind* III site of pBR322 to form p *trp* ED3. The *Hind* III site near the single *Eco*RI site of p *trp* ED3 was subsequently removed using exonuclease III and nuclease S1 and a deleted plasmid isolated (p *trp* ED5-1, Fig. 7) having a *Hind* III site at the end of the *trp* D gene suitable for insertion of cloned genes. Polyacrylamide gel electrophoresis of the proteins made in cells containing p *trp* ED5-1, 3 h after induction with IAA, indicated that about 30% of total cell protein was anthranilate synthetase. This was more protein than was made by similar *lac* based vectors containing the *lac* z gene and furthermore the synthesis was inducible rather than constitutive. In addition, it was shown that the small protein derived from the deleted *trp* D gene (15% of full length) was stable, probably because it was bound to the *trp* E protein (Hallewell and Emtage, 1980).

B Expression of fusion proteins

Plasmid p*trp*ED5-1 has been used to make a hybrid protein consisting of the NH_2-terminal part of *trp* D fused to HGH. This was

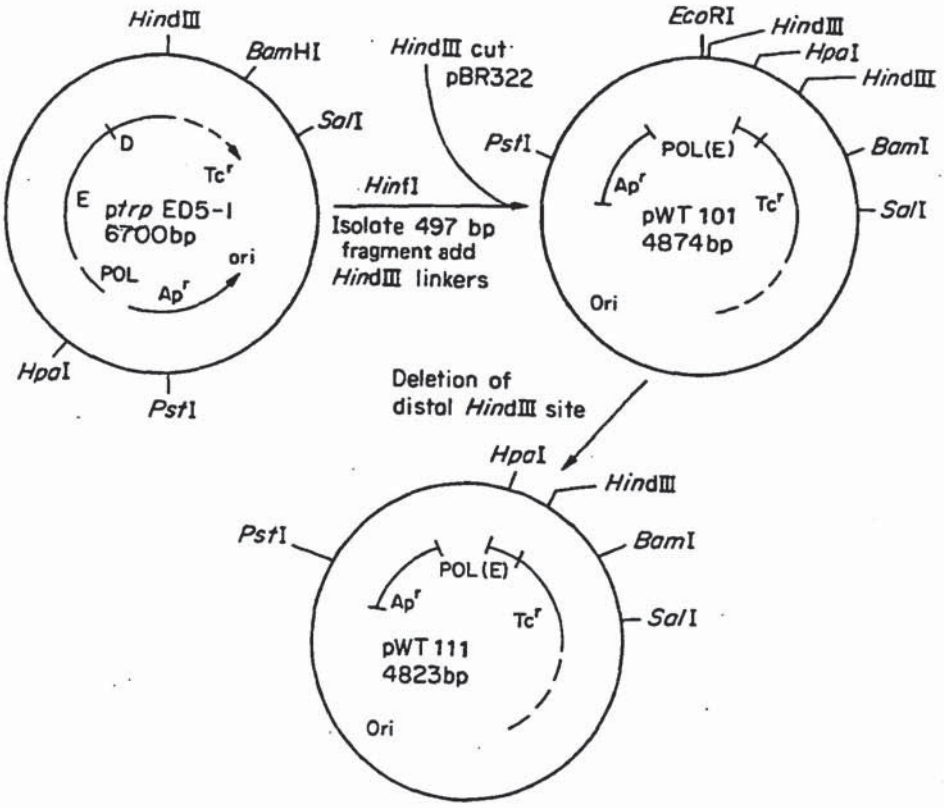


Figure 7 Construction of the pWT series of *trp* promoter expression vectors. Adapted from Hallewell and Emtage 1980, and Tacon *et al.*, 1980.

done by inserting HGH cDNA into the *Hind* III site of the plasmid using *Hind* III linkers maintaining the frame of translation (Martial *et al.*, 1979). About 3% of total cell protein was identified as the hybrid protein by polyacrylamide gel electrophoresis. It is interesting to note that whereas in the original plasmid equal amounts of *trp* E and *trp* D were produced, in the HGH construction the amount of *trp* E made was far greater than the *trp* D fusion protein. This indicates that some feature of the gene other than its transcription or the structure of the rbs affects the overall level of expression. Derivatives of p*trp*ED5-1 have been made by cloning a *Hinf* I fragment, containing the *trp* operator/promoter, into *Hind* III-digested pBR322 using *Hind* III linkers (Tacon *et al.*, 1980). The resulting plasmid (pWT 111, Fig. 7) contained the *trp* regulatory regions plus the first seven codons of *trp* E, upstream of the tetracycline resistance gene of pBR322. Insertion of "foreign" DNA into the *Hind* III site in the correct frame for translation allowed the synthesis of N-terminal *trp* E fusions. Further derivatives, allowing expression of a fusion protein in the other two translation frames,

were also constructed using a technique similar to the one used to phase β-galactosidase fusions (Charnay *et al.*, 1979). The tetracycline resistance genes of all these plasmids were placed under *trp* control by the insertion of the *trp* promoter/operator at the *Hind*III site, allowing transcription through any inserted sequence to be monitored by tetracycline resistance.

One of these derivatives (pWT121) has been used to express the haemagglutinin gene (HA) of fowl plague influenza virus (Emtage *et al.*, 1980). Nucleotide sequencing indicated that the frame of translation of *trp* E would be maintained if the HA gene (obtained by cDNA cloning) was inserted into the *Hind*III site of pWT121. The predicted fusion protein consisted of 7 amino acids of *trp* E, 6 amino acids from the linker, 6 phenylalanine residues from the $(T)_{19}$ "tail" on the cloned HA DNA, 7 amino acids from the normally untranslated 5′ end of the HA gene, 558 amino acids of the HA and its signal sequence plus 5 amino acids from the C terminal *Hind*III linker, giving a total of 589 amino acids (Emtage *et al.*, 1980). A protein with HA immunoreactivity was detected in *E. coli* containing the plasmid but the size of the immunoprecipitated polypeptide on polyacrylamide gels was smaller than the predicted 69 000 molecular weight indicating that some processing had occurred (Emtage *et al.*, 1980). The levels of expression were considerably lower than expected compared to the levels of *trp* E produced by p *trp* ED5-1 (Hallewell and Emtage, 1979) but similar to the levels of ovalbumin-like proteins produced as β-galactosidase fusions from the *lac* promoter (Mercereau-Puijalon *et al.*, 1978; Fraser and Bruce, 1978). A possible explanation for the low level of expression was that the eukaryotic signal sequence contained in the HA gene was somehow inhibitory or toxic to *E. coli*. Rose and Shafferman (1981) showed that analogous *trp* expression plasmids containing the complete glycoprotein gene (G) of vesicular stomatitis virus (VSV) (obtained from cDNA copies of virus mRNA) including the signal peptide, fused in phase to the first seven amino acids of *trp* E, were lethal in cells lacking *trp* repressor (*trp* R^-) but could be transformed into *trp* R^+ cells. A protein of the size expected for a *trp* E-G fusion protein was detected in ^{35}S methionine pulse-labelled *trp* R^+ mini cells (where only protein synthesis from plasmid DNA is examined). Deletion of some of the hydrophobic residues at the COOH-terminus of the glycoprotein gene did not overcome the lethality effect in *trp* R^- cells, whereas expression of a gene lacking ten hydrophobic amino acids from the NH_2-terminal signal sequence was detected (Rose and Shafferman, 1981). In neither instance, however, was the level of expression high enough for a new protein to be detected as a stained or labelled band in polyacrylamide

gels of total cell protein. Heiland and Gething (1981) similarly obtained low HA activity in *E. coli* from plasmids containing the *lac* promoter directing transcription of short β-galactosidase-HA fusions with a deleted HA signal sequence.

Higher levels of expression of HA (5–7% total cell protein) have been achieved by using longer *lac* promoter β-galactosidase fusions. Fragments of HA, lacking the N-terminal signal sequence were fused to a long NH_2-terminal β-galactosidase coding sequence giving rise to very large (> 130 000 molecular weight) insoluble fusion proteins (Davis *et al.*, 1981). Some expression was also detected (by polyacrylamide gel electrophoresis) when mature HA was fused to a *trp* LE gene (see below) transcribed from the *trp* promoter, but constructions designed to obtain direct expression of mature HA from either the *lac* or *trp* promoter were not successful (Davis *et al.*, 1981).

Other animal virus proteins have been expressed in *E. coli* as *trp* E fusions. Kleid *et al.* (1981a) constructed a vector containing the *trp* promoter/operator, designed to direct the synthesis of the immunogenic capsid protein VPI (also called VP3) of foot and mouth disease virus (FMDV) linked to a hybrid protein, consisting of the NH_2-terminus of the *trp* leader peptide fused to the last third of the *trp* E protein gene (ΔLE 1413). The *trp* LE fragment of 190 amino acids is a particularly useful protein for fusions because it is insoluble and resistant to proteolytic cleavage. Moreover, the deletion (ΔLE 1413) removes the attenuator site so that any mRNA secondary structure effects leading to attenuation should be avoided. (For a detailed discussion of attenuation in the *trp* operon, see Yanofsky, 1981.) The expression vector (pFM1) was made by annealing a fragment of the VPI gene (a *Pst*I-*Pvu*II fragment containing codons 8-211) to the *Eco*RI-*Bam*HI fragment of pBR322 and ligating these to a *trp* promoter/operator fragment containing the *trp* LE hybrid coding sequence. The *trp* fragment contained an *Eco*RI site at the end of the *trp* E gene, so an *Eco*RI-*Pst*I linker was used to link the sequences and maintain the frame of translation (Kleid *et al.*, 1981a). (The details of the construction of this *trp* LE fragment are described in Kleid *et al.*, 1981b.)

A *trp* LE-VPI fusion protein which has a molecular weight of about 44 000 was obtained in the pellet from a detergent lysate of bacteria harbouring the plasmid after low speed centrifugation. Relatively large amounts of the protein (170 mg from 800 ml of culture, or about 17% of total cell protein) were produced. After purification by gel electrophoresis the protein was mixed with adjuvant and used to vaccinate both swine and cattle. As might be expected, the fusion protein was no better as an immunogen than

VPI isolated from the virus particle (Kleid *et al.*, 1981). Nevertheless, the results are of importance because a very high level of expression was obtained and the hybrid protein could be purified sufficiently for it to be used as an immunogen. This will enable studies of antigen presentation and the effect of adjuvants on the immunogenicity of the protein to be assessed properly with reasonable amounts of pure material. Kupper *et al.* (1981) have also described the expression of a VPI fusion protein, using cDNA made to a different FMD virus serotype, coupled to a P_L promoter based expression vector (pPLc24, see Fig. 6). VPI was produced with 99 amino acids of MS2 polymerase at the NH_2-terminus and 13 amino acids from the vector at the COOH-terminus. This construction did not lead to the levels of fusion protein obtained by Kleid *et al.* (1981a).

C Expression of interferon

Plasmids have been constructed containing a leukocyte interferon gene (LeIF-A) downstream of the *trp* promoter and *trp* LE fragment (Goeddel *et al.*, 1980a). Although the precursor form of the protein containing the signal sequence (pre LeIF-A) was fused to the *trp* E gene in phase, the size of the expressed protein was consistent with translation having initiated at the signal peptide AUG codon. Not surprisingly, considerably higher amounts of interferon (480 000 u/l) were obtained in this way, than the 20 000 u/l reported for a pre LeIF D gene cloned directly into the *Pst*I site of pBR322 (Nagata *et al.*, 1980a, see below). The designed expression of native, mature LeIF has also been accomplished using the *trp* promoter. The procedure followed was a variation of the method used to express human growth hormone (Goeddel *et al.*, 1979b). As shown in Fig. 8, the *Sau*3A site between codons 1 and 2 of LeIF-A cDNA was used to produce a fragment (*Sau*3A-*Ava*II) to which a synthetic DNA sequence, coding for an initiating ATG plus a codon for the N-terminal cysteine of mature LeIF and an *Eco*RI cohesive end, was ligated. This reconstituted N-terminal coding sequence was then annealed to the rest of the coding sequence and the gene inserted into a plasmid containing the *trp* promoter/operator and leader peptide rbs upstream of an *Eco*RI site. High levels of expression of LeIF were obtained as measured by the levels of antiviral activity present in bacteria containing this plasmid (2.5×10^8 u/l; 600 µg/l). The protein was soluble and had all the expected characteristics of mature leukocyte interferon (e.g. it was stable at pH 2 and the activity was neutralized by antihuman leukocyte interferon antibodies). Moreover, the protein could be shown to protect monkeys against potentially lethal encephalomyocarditis (EMC) virus infection

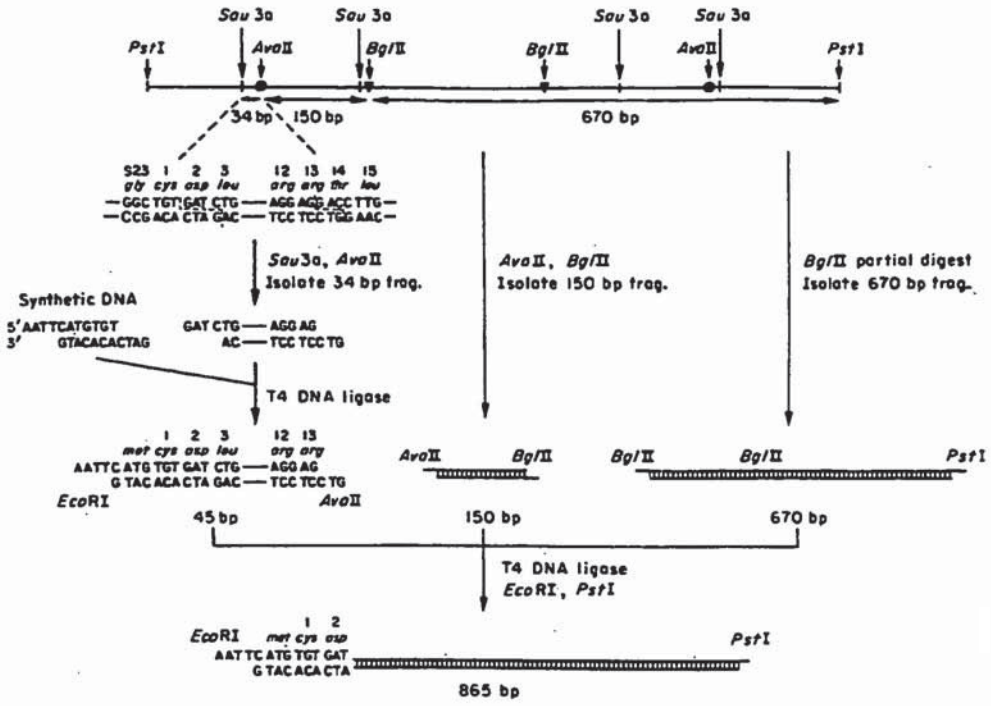
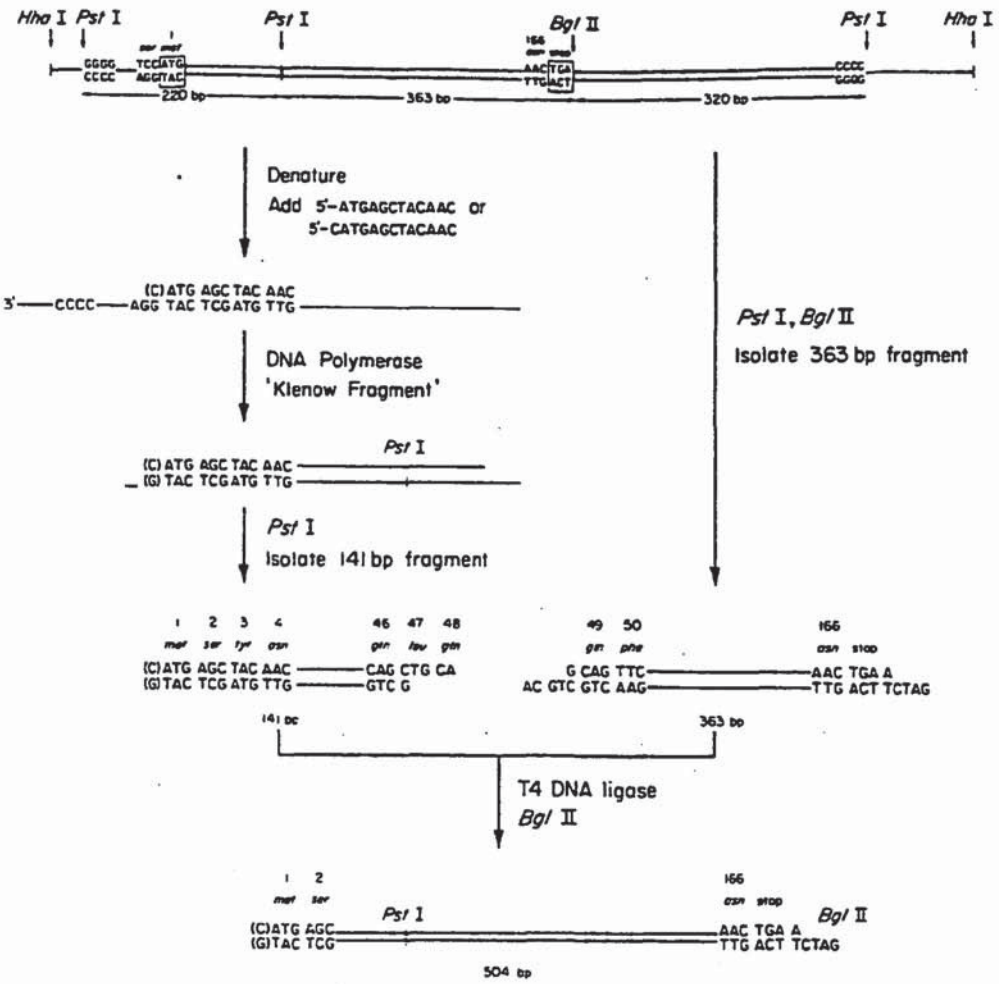


Figure 8 Construction of a gene coding for mature human LeIF-A for expression in a *trp* promoter based vector. From Goeddel *et al.*, 1980a. Reprinted by permission from *Nature*, 287, 411—416. Copyright 1980, Macmillan Journals Ltd.

even in relatively impure form (Goeddel *et al.*, 1980a). This protein has now been purified from *E. coli* and some of its structural properties examined (Wetzel, 1981; Wetzel *et al.*, 1981b).

Human fibroblast interferon has also been synthesised in *E. coli* as a native, mature protein lacking the signal sequence, using both the *trp* and the *lac* promoter (Goeddel *et al.*, 1980b). The construction of these vectors was more complicated than the similar LeIF vectors described above, owing to the lack of conveniently placed restriction sites. The signal peptide region of pFIF3, a plasmid containing HFIF cDNA, was removed using a modification of the primer repair method described for removing the signal peptide from the influenza HA gene (Davis *et al.*, 1981; see also Kleid *et al.*, 1981b). A HhaI fragment of DNA containing the entire HFIF cDNA sequence was denatured (instead of making DNA with long single stranded 3′ extensions with λ exonuclease -- Davis *et al.*, 1981) and a primer, containing an ATG and the coding sequence for the first four amino acids of mature HFIF, annealed. The Klenow fragment of DNA polymerase was then used for repair synthesis. At the same time the 3′—5′ exonuclease activity of the enzyme removed



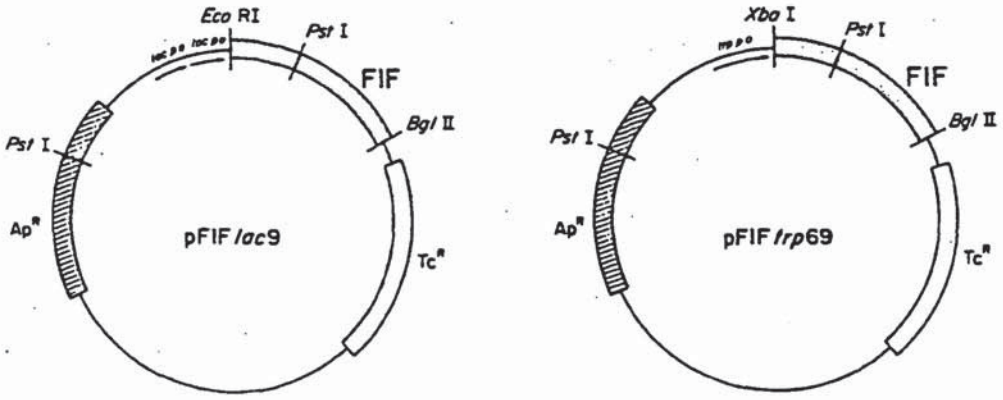


Figure 9 Assembly of the gene coding for HFIF and construction of the *lac* and *trp* based vectors used for expression of the mature protein. Taken from Goeddel *et al.*, 1980b. (Reprinted from *Nucleic Acids Res.* with permission.)

protruding 3′ ends leaving a blunt end where the primer was annealed (Fig. 9). This 5′ fragment was then obtained by restriction enzyme digestion and gel electrophoresis and the coding sequence reestablished by ligation of the fragment to other suitable restriction fragments (Fig. 9). The reconstituted gene was inserted into expression plasmids containing either the *trp* promoter and leader rbs or the *lac* promoter and β-galactosidase rbs (Goeddel *et al.*, 1980b). Bacteria containing the *trp* promoter plasmid apparently produced more HFIF than bacteria containing the *lac* promoter; moreover plasmids containing three *trp* promoters in series before the HFIF gene produced 4—5 times as much HFIF as those plasmids with only one *trp* promoter. Nevertheless, the levels of HFIF produced from one *trp* promoter were 10-fold less than the analogous LeIF constructions (cf. Geoddel *et al.*, 1980a).

Increased levels of synthesis of both LeIF and HFIF have since been obtained by manipulation of the nucleotide sequence between the SD sequence of the *trp* leader and the ATG of the interferon coding sequences (Shepard *et al.*, 1982). The optimal spacing for expression of both types of interferon was nine nucleotides, despite the differences in nucleotide sequence between the SD sequence and the AUG for each transcript. Once again, it is likely that the increase in expression is due to the mRNA having a secondary structure which facilitates ribosome binding and initiation (Shepard *et al.*, 1982; Iserentant and Fiers, 1980; Gheysen *et al.*, 1982). This provides a further indication that the efficiency of translation may be one of the major limiting steps in the production of proteins from cloned eukaryotic genes.

The availability of cloned cDNA copies to several of the different types of leukocyte interferon genes which make up the multigene family (Nagata *et al.*, 1980b, Goeddel *et al.*, 1981) has facilitated the construction of vectors for the expression of other leukocyte interferon genes and has enabled the activities of the expressed proteins to be compared to natural human leukocyte interferon (e.g. see Stewart *et al.*, 1980). Yelverton *et al.* (1981) synthesized LeIF-B by inserting the cDNA for this protein in place of the LeIF-A sequence in the plasmid pLeIFA-25 (Goeddel *et al.*, 1980a, see above). The protein produced had markedly different antiviral specificity to LeIF-A, in agreement with other studies using LeIF molecules obtained from different cDNA clones (Streuli *et al.*, 1980). The fact that the LeIF gene family is unusual in not having introns interrupting the coding sequence (Nagata *et al.*, 1980b) extends the possibilities as far as expression of these genes is concerned, since DNA from genomic clones could be used for constructions, instead of relying on the availability of cDNA clones

(e.g. Mory *et al.*, 1981). The observed difference in the biological properties of these *E. coli* derived leukocyte interferons has led to the construction of hybrid genes, consisting of the coding sequence for the NH_2-terminus of one type of interferon (e.g. LeIF A or D) coupled to that coding for the COOH-terminus of another, by recombination *in vitro* through common restriction enzyme sites within their coding regions (Streuli *et al.*, 1981). These hybrid genes have been expressed as short β-galactosidase fusion proteins from the *lac* promoter or as native polypeptides under *trp* promoter control, as described above. Polypeptides with biological characteristics different to either of the parent molecules were obtained (Streuli *et al.*, 1981; Weck *et al.*, 1981).

A potentially greater number of different interferon genes could be constructed synthetically. The availability of a chemically synthesized interferon gene, made from oligonucleotide blocks (Edge *et al.*, 1980) opens up the possibility of altering the coding region specifically by substituting different oligonucleotide blocks. The expression of this chemically synthesised LeIF gene in *E. coli* from the *lac* promoter has been reported recently (De Maeyer *et al.*, 1982). Human immune interferon (interferon γ) has also been cloned, sequenced and the protein expressed to a low level in *E. coli* (Gray *et al.*, 1982), so it should now be possible to compare the specific activity and biological properties of the three different types of human interferon synthesized by *E. coli*.

VIII The β-lactamase promoter

A Synthesis of fusion proteins

An alternative system that has been used to obtain low level expression of hybrid proteins is the β-lactamase (ampicillin resistance) gene of pBR322. The sequence of this gene is known (Sutcliffe, 1979); it has a single *Pst*I site located between codons 181 and 182. The expression studies arose as a result of the fact that the *Pst*I site has been used extensively as the insertion site for cloning cDNA by homopolymer tailing with terminal transferase. Not only does cloning into this site inactivate β-lactamase but the G-C homopolymer tails reconstitute *Pst*I sites at each end of the insert, making removal of the insert potentially easy. (For a detailed description of cDNA cloning see Craig and Hall, 1983.)

Villa-Komaroff *et al.* (1978) cloned cDNA transcripts of rat preproinsulin mRNA into the *Pst*I site of pBR322 by this method. Since G tails of different lengths were added to the plasmid by terminal transferase and the insertion took place within the coding region of

β-lactamase some of the insertions were expected to maintain the reading frame across the G-C junction and give rise to β-lactamase fusion proteins. Using a solid phase radioimmunoassay designed to detect bacterial colonies expressing antigens (Broome and Gilbert, 1978), one of the clones containing insulin sequences was found to produce a β-lactamase-insulin hybrid protein. Furthermore, the hybrid protein was secreted into the periplasmic space indicating that the signal sequence at the NH_2-terminus of β-lactamase was directing secretion of the hybrid protein. Nucleotide sequencing showed that 18 G residues separated the β-lactamase and the rat proinsulin sequence, indicating that a run of six glycines fused the alanine at position 182 of β-lactamase to the glutamine at position 4 of rat proinsulin (Villa-Komaroff *et al.*, 1978).

Rat pregrowth hormone has also been expressed as a β-lactamase hybrid protein by fusion of the β-lactamase gene to the conveniently placed *Pst*I site at codon −24 of the growth hormone signal sequence (Seeburg *et al.*, 1978). Although expression of the hybrid protein, consisting of 181 amino acids of β-lactamase linked to 214 amino acids of rat pregrowth hormone, could be detected in mini-cells, there was little if any growth hormone-related protein detected in the periplasmic space, in contrast to the findings with proinsulin (Villa-Komaroff *et al.* 1978).

B Secretion of native proteins using β-lactamase fusions

The successful transport of a β-lactamase-insulin hybrid to the periplasmic space prompted further experiments to characterise the requirements for the secretion. A series of plasmids (pKT) were made in which the *Pst*I site in the β-lactamase gene was moved so that it was in or near the coding region for the signal sequence of β-lactamase. This was done by deletion mutagenesis using *Bal*31 exonuclease and *Pst*I sites recreated at the boundaries by ligation of *Pst*I linkers. The genes for rat proinsulin and preproinsulin were cloned into the *Pst*I site of these different plasmids, resulting in the formation of different hybrid β-lactamase and insulin signal sequences (Talmadge *et al.*, 1980a). The level of insulin accumulating in the periplasmic space of *E. coli* containing the different plasmids was then measured. It was concluded that possession of a signal sequence was essential for secretion and that this sequence could be either prokaryotic or eukaryotic (or both) in origin. In addition it was shown that for constructions where the entire rat preproinsulin sequence was fused to all or part of the bacterial signal sequence, proinsulin could be detected in the periplasmic space, indicating that the bacterial signal peptidase was recognising the eukaryotic signal

sequence and processing preproinsulin to proinsulin (Talmadge *et al.*, 1980b). Similar results have been reported for human preproinsulin attached to a hybrid signal sequence (Chan *et al.*, 1981).

More recently, Talmadge *et al.* (1981) have shown that if the rat preproinsulin sequence was fused to the first eight amino acids of β-galactosidase under *lac* promoter control, then the expressed protein was processed at the eukaryotic signal sequence as before, and proinsulin found in the periplasmic space. This finding provides additional evidence that it is the signal sequence that is responsible for the secretion, rather than any bacterially derived peptide remaining at the NH_2-terminus. It also demonstrated that the signal sequence need not be precisely at the NH_2-terminus; in these β-galactosidase fusions, 18 amino acids preceded the preproinsulin sequence (Talmadge *et al.*, 1981). One of the potential advantages of making DNA constructions that ensure the secretion of the protein in this way is that periplasmic proteins may be less likely to be degraded than those remaining in the cytoplasm. In this regard, Talmadge and Gilbert (1982) have shown that the rat proinsulin remaining in the cytoplasm produced from constructions not having a signal sequence is more rapidly degraded than proinsulin that is secreted. Whether this will be true for other eukaryotic proteins remains to be seen but it should be noted that similar experiments with other eukaryotic genes have not been reported.

C Synthesis of other native proteins

Mouse dihydrofolate reductase (DHFR) cDNA has also been inserted into the *Pst*I site of pBR322 by G-C tailing and enzymatically active protein detected (Chang *et al.*, 1978). Phenotypic selection for expression of the eukaryotic sequence (i.e. resistance to trimethoprim) allowed the isolation of bacterial cells making a soluble protein that had the enzymatic and immunological properties and molecular size (22 000 molecular weight) of the native mouse enzyme. Nucleotide sequence analysis across the *Pst*I site of the plasmid-cDNA junctions showed that the DHFR cDNA inserts were not in the correct reading frame for expression to be occurring as a fusion polypeptide, implying that initiation of translation was occurring at the DHFR AUG in a polycistronic mRNA. Moreover, some expression was detected in clones with inserts in the opposite orientation to that required for a fusion polypeptide (Chang *et al.*, 1978). DHFR cDNA clones with different sequences before the ATG induced different levels of trimethoprim resistance, indicating that the amount of DHFR made was being controlled at the translational level (Chang *et al.*, 1978). It was postulated that the *Pst*I-poly(G)

sequence at the vector/cDNA junction provided a sufficient homology with the SD sequence for it to direct ribosomes to initiate translation at a nearby AUG. Further evidence for this idea has come from a more detailed correlation of the nucleotide sequences at the vector-insert junction with the levels of enzyme made, for several of the expressing clones (Chang *et al.*, 1980). It was apparent that the optimal distance between the beginning of the artificial SD sequence and the AUG was 12—14 nucleotides — somewhat longer than the optimum found for other genes (e.g. interferon, see Shepard *et al.*, 1982). This difference may be attributable to mRNA secondary structure and the degree of homology between the SD sequence and the 3′-end of 16S rRNA (Chang *et al.*, 1980; Iserentant and Fiers, 1980). In this connection Jay *et al.* (1981) have joined a chemically synthesised rbs sequence with maximum homology to the 3′-end of 16S rRNA, to the SV40 tAg coding sequence and placed this hybrid gene into the *Pst*I site in the β-lactamase gene. The synthetic rbs present in the mRNA (transcribed from the β-lactamase promoter) was apparently recognised because authentic SV40 tAg could b detected in bacterial extracts by immunoprecipitation (Jay *et al.*, 1981). Manipulation of the rbs sequence will now allow a more systematic study of the effect of rbs-16S rRNA homology on the translation of cloned eukaryotic genes.

Phenotypic selection has also been used to obtain plasmids expressing the herpes simplex virus (HSV) thymidine kinase (TK) gene (Garapin *et al.*, 1981). A fragment of the virus DNA containing the gene was fused to each of the three "phased" *lac* promoter β-galactosidase fusion vectors (Charnay *et al.*, 1978) and the plasmids used to transform TK^- *E. coli* to TK^+. When the TK gene was in the right orientation for transcription from the *lac* promoter functional thymidine kinase was found with all three vectors, irrespective of the translational reading frame. Analysis of the nucleotide sequence preceding the start sites identified putative SD sequences, suggesting that initiation of translation was occurring as described for DHFR (Chang *et al.*, 1980). Native thymidine kinase has also been detected in bacteria harbouring plasmids containing the HSV TK gene positioned downstream of the tetracycline promoter of pBR322 (Kit *et al.*, 1981). The plasmids constructed by Nagata *et al.* (1980a) producing LeIF as a β-lactamase fusion using the pKT vectors also appeared to give rise to non-fusion polypeptides, since their production was similarly independent of the translational reading frame predicted by the vector into which the cDNA was inserted.

Reinitiation of translation has also been invoked to explain the production of hepatitis virus core antigen (HBcAg) from restriction fragments of hepatitis virus DNA cloned into the *Pst*I site of pBR322

by G-C tailing (Burrell *et al.*, 1979; Pasek *et al.*, 1979). Although all the inserts were in the correct orientation for expression of β-lactamase fusions, none of them were in the right translation frame for the core antigen to be made as a β-lactamase fusion polypeptide, indicating that initiation was occurring from an ATG in HBcAg. Sufficient core antigen was produced in this way for it to be used to raise a specific antiserum (Pasek *et al.*, 1979). Higher levels of HBcAg have now been expressed in *E. coli* from both the *trp* and *lac* promoters as native and fusion polypeptides (Edman *et al.*, 1981; Stahl *et al.*, 1982). For the expression of the native polypeptide a vector was constructed that contained the *trp* promoter and leader rbs abutted to the HBcAg coding sequence, giving an SD to ATG distance of 13—16 base pairs (Edman *et al.*, 1980). Under conditions of full derepression, 10% of newly synthesized protein was HBcAg. The HBcAg contains 25 arginine codons, 17 of which are coded for by a codon rarely used in *E. coli* (Grantham *et al.*, 1980) so the high level of synthesis obtained suggests that, for this gene at least, the concentration of iso-accepting tRNA species does not inhibit translation. Gel exclusion chromatography and electron microscopy of bacterial HBcAg produced by *lac* promoter constructions (Stahl *et al.*, 1982) has shown that the antigen is in an aggregated form similar in appearance to virus cores seen in extracts of HBV infected liver cells (Cohen and Richmond, 1982). The aggregated protein seems to recognise HBcAg antibodies well enough for it to be used as a diagnostic reagent (Stahl *et al.*, 1982).

Small amounts of another hepatitis B virus antigen, the surface antigen (HBsAg) have also been synthesized in *E. coli* from virus DNA cloned into the *Pst*I site of pBR322 (Pasek *et al.*, 1979; Mackay *et al.*, 1981). Attempts to make larger amounts of this protein as a native polypeptide from the *trp* promoter have not been successful, although some expression of a protein lacking the HBsAg signal sequence but fused to the NH_2-terminus of the signal sequence of β-lactamase transcribed from the *trp* promoter, has been reported (Edman *et al.*, 1981). Very small amounts of long β-galactosidase-HBsAg fusion proteins have also been detected in *E. coli* infected with a recombinant λ phage carrying the HBsAg gene and the *lac* promoter (λ p *lac* 5-1 UV5, Charnay *et al.*, 1980).

IX Conclusions and future prospects

Table 2 presents an up to date summary of the higher eukaryotic proteins that have been expressed in *E. coli*. The levels of expression are indicated where possible and the nature of the construction

Table 2 Expression of eukaryotic DNA in *E. coli.*

Protein (molecular weight)	Source of DNA	Promoter	Construction	Nature of product/cleavage protocol	Cell fraction (before or after lysis)	Level of expression (% total cell protein or mol/cell)	Reference
Somatostatin (c. 1500)	Chemical synthesis	*lac*	β-gal fusion	Hybrid protein; CNBr cleavage	Pellet	Low but immuno-detectable c 0.05%	Itakura *et al.*, 1977
Human insulin A = (c. 2000) B = (c. 3000)	Chemical synthesis of separate A and B chains	*lac*	β-gal fusion	Hybrid A and B CNBr cleavage. Re-constitution to give insulin	Pellet	20% = 10^5 mol/cell; 10 mg/24 g wet cells	Goeddel *et al.*, 1979a
Human proinsulin (mini C) (c. 6500)	Chemical synthesis	*lac*	β-gal fusion	Hybrid protein; CNBr cleavage	Pellet	12 mg/322 g wet cells	Wetzel *et al.*, 1981a
Rat proinsulin (c. 12 000)	cDNA	*lac*	β-gal fusion	Proinsulin	Periplasmic space	Low	Talmadge *et al.*, 1981
Thymosin α_1 (3100)	Chemical synthesis	*lac*	β-gal fusion	Hybrid protein; Nα desacetyl thymosin α_1 formed after CNBr cleavage	Pellet	10^5 mol/cell 40 mg/100 g wet cells	Wetzel *et al.*, 1980
Leu-enkephalin (c. 600)	Chemical synthesis	*lac*	β-gal fusion	Hybrid protein; CNBr gives native hormone	Pellet	2×10^5 mol/cell 5 mg/100 g wet cells	Shemyakin *et al.*, 1980
α neo-endorphin (c. 1200)	Chemical synthesis	*lac*	β-gal fusion	Hybrid protein; CNBr gives native hormone	Pellet	5×10^5 mol/cell 4 mg/10.9 g wet cells	Tanaka *et al.*, 1982
β-endorphin (c. 3200)	ACTH/ βLPH cDNA	*lac*	β-gal fusion	Hybrid protein; Native hormone prepared by citraconylation and trypsin treatment	Pellet	5%	Shine *et al.*, 1980

Rous Sarcoma virus (RSV) protein kinase (src) (60 000)	cDNA	*lac* (× 2)	8 amino acid β-gal fusion	Hybrid protein; NH_2 terminal non src amino acids	Supernatant	0.3%	Gilmer *et al.*, 1982; Gilmer and Erikson, 1981
HSV thymidine kinase (TK) (43 000)	Virus DNA clone	*lac*	Short β-gal fusion	Native enzyme? (40—42 000)	Supernatant	0.2—0.3%	Garapin *et al.*, 1980
Chicken ovalbumin (45 000)	cDNA	*lac*	Short β-gal fusion	β-gal-ovalbumin	Pellet (some in periplasmic space)	1—1.5% 5×10^4 mol/cell	Fraser and Bruce, 1978; Mercereau-Puijalon *et al.*, 1978; Baty *et al.*, 1981
Rabbit β globin (c. 9000)	cDNA	*lac*	rbs fusion	Met-β globin	Pellet?	1.5×10^4 mol/cell	Guarente *et al.*, 1980
Human growth hormone (HGH) (24 000)	Chemical synthesis and cDNA	*lac* (× 2) and *lac-trp* (*tac*)	Constructed rbs fusion, SD-AUG is 7—11 bp	Met HGH	Supernatant (soluble protein	1.8×10^5 mol/cell 2.4 μg/ml	Goeddel *et al.*, 1979b de Boer *et al.*, 1982
Human fibroblast interferon (HFIF) (23 000 — Pre-HFIF)	cDNA	*lac*	rbs fusion	Pre-HFIF Met-HFIF	Pellet (protein aggregates)	Low (50 mol/cell)	Taniguchi *et al.*, 1980
(20 000 — HFIF)	cDNA	*lac*	rbs fusion	Met-HFIF	Supernatant	2.2×10^3 mol/cell	Goeddel *et al.*, 1980b
Human leukocyte interferon (LeIF-α) (20 000)	Chemical synthesis	*lac*	rbs fusion	Met-LeIF	Supernatant	c. 10^7 u/ml	De Maeyer *et al.*, 1982
LeIF-α_2	cDNA	*lac* (in M13 mp7)	Short β-gal fusion	Native LeIF-α_2 19 500	Supernatant	10^9 u/l = 5 mg/l	Slocombe *et al.*, 1982

Thaumatin (22 000)	cDNA	*lac*	rbs fusion	Preprothaumatin	—	Low	Edens *et al.*, 1982
Influenza virus HA gene (61 000)	cDNA	*lac*	linker-part HA fusion	HA protein lacking some NH_2-terminal amino acids	—	3×10^3 mol/cell	Heiland and Gething, 1981
			Long β-gal fusion	Large hybrid protein (130 000)	Pellet	5—7% 5—7 x 10^4 mol/cell	Davis *et al.*, 1981
Hepatitis B virus core antigen (HBcAg) (19 000 or 22 000)	Virus DNA clone	*lac*	short β-gal fusion	Native and hybrid HBcAg	Supernatant (small aggregate)	Low	Stahl *et al.*, 1981
Hepatitis B virus surface antigen (HBsAg) (22 600)	Virus DNA clone	*lac* (in phage)	long β-gal fusion	Hybrid protein (138 000)	Supernatant	0.05%	Charnay *et al.*, 1980
Polyoma small t antigen (20 000)	Virus DNA	*lac*	β-gal linker fusion	Hybrid protein (26 000)	Supernatant (soluble)	0.15%	Horwich *et al.*, 1980
SV40 tAg (19 000)	Virus DNA	*lac*	rbs fusion	Native tAg	Pellet; high mol wt complex	1—5 x 10^3 mol/cell 0.8%, 4—8 x 10^4 mol/cell	Roberts *et al.*, 1979c; Thummel *et al.*, 1981
SV40 tAg (19 000)	Virus DNA	P_L^*	rbs fusion	Native protein (19 000 and 14 500 derivative	Pellet	2.5% 7%	Derom *et al.*, 1982
HFIF (20 000)	cDNA	P_L^*	β-lactamase or MS_2 polymerase fusions	Native HFIF by processing?	Pellet	Low	Derynck *et al.*, 1980
Foot and mouth disease virus VP1 (FMDV-VP1) (24 000)	cDNA from virus RNA	P_L^*	MS_2 polymerase fusion	MS_2 pol-VP1 fusion polypeptide	Pellet	10^3 mol/cell	Küpper *et al.*, 1981

HGH (24 000)	cDNA	*trp*	*trp* D fusion	*trp* D-HGH hybrid protein (32 000)	Supernatant	3%	Martial *et al.*, 1979
Influenza HA gene (61 000)	cDNA	*trp*	*trp* E-linker fusion	Hybrid protein (69 000) some processing to (61 000)	Pellet?	0.75%	Emtage *et al.*, 1980
VSV G protein (62 000)	cDNA	*trp*	*trp* E fusion	Hybrid protein (lacking G signal sequence)	Supernatant	1%	Rose and Shafferman, 1981
FMDV-VP1 (24 000)	cDNA	*trp*	*trp* LE fusion	Hybrid protein (44 000)	Pellet	17% 1—2 x 10^6 mol/cell	Kleid *et al.*, 1981a
RSV src (60 000)	cDNA	*trp*	Short HGH-NH_2-terminal fusion	Hybrid protein (60 000)	Supernatant	5% 1.7 x 10^5 mol/cell	McGrath and Levinson, 1982
HBcAg (22 000)	Virus DNA clone	*trp*	*trp* leader rbs fusion	HBcAg	100S complex	10% 5 x 10^5 mol/cell	Edman *et al.*, 1981
HBsAg (22 600)	Virus DNA clone	*trp*	β-lactamase fusion	Hybrid protein (41 000)	—	8.5% 1.7 x 10^5 mol/cell	Edman *et al.*, 1981
LeIF-A (c. 20 000)	cDNA	*trp*	*trp* LE fusion	Pre LeIF-A	Supernatant	1 μg/l	Goeddel *et al.*, 1980a;
			trp leader rbs fusion	Met LeIF-A	Supernatant	600 μg/l	Shepard *et al.*, 1982
Le-1F B (20 000)	cDNA	*trp*	*trp* leader rbs fusion	Met LeIF-B	Supernatant	200 μg/l	Yelverton, *et al.*, 1981
HF1F (20 000)	cDNA	*trp* (x 3)	*trp* leader rbs fusion	Met HF1F	Supernatant	4.5—20 x 10^3 mol/cell	Goeddel *et al.*, 1980b
Human immune interferon (17 000?)	cDNA	*trp*	*trp* leader rbs fusion	Met IF-γ	Supernatant after sonication	Low c. 80 mol/cell	Gray *et al.*, 1982 1982

Human serum albumin (67 000)	cDNA	*trp*	*trp* leader rbs fusion	Met-HSA	—	Low		Lawn *et al.*, 1981
Chymosin (35 000)	cDNA	*trp*	*trp* leader rbs fusion	Prochymosin	Pellet	1—5% 5×10^4 mol/cell		Emtage *et al.*, 1983
Thaumatin (22 000)	cDNA	*trp*	*trp* leader rbs fusion	Preprothaumatin	—	Low		Edens *et al.*, 1982
Rat prepro insulin (12 000)	cDNA	β-lactamase	β-lactamase fusion	Hybrid protein	Periplasmic space	100 mol/cell		Villa-Komaroff *et al.*, 1978
			β-lactamase/ preproinsulin signal sequence fusions	Proinsulin	Periplasmic space	6×10^3 mol/cell		Talmadge *et al.*, 1980b; Talmadge and Gilbert, 1982
Human prepro-insulin (12 000)	cDNA	β-lactamase	β-lactamase fusions	β-lactamase/ preproinsulin hybrid proteins and proinsulin	Periplasmic space	0.2—1%		Chan *et al.*, 1981
Rat prepro-hormone (c. 22 000)	cDNA	β-lactamase	β-lactamase fusion	Hybrid protein (44 000)	—	2.4×10^4 mol/cell		Seeburg *et al.*, 1978
Rat prolactin (22 000)	cDNA	β-lactamase	β-lactamase fusion	Hybrid protein	—	Low		Erwin *et al.*, 1981
Bovine growth hormone (21 000)	cDNA	β-lactamase	β-lactamase fusion	Hybrid protein	—	Low		Keshet *et al.*, 1981
Porcine secretion (c. 3000)	Chemical synthesis	β-lactamase	β-lactamase fusion	Pre β-lactamase and β-lactamase-secretin hybrids (23 000 and 19 000)	Supernatant	2 mol/cell (desamido secretin)		Suzuki *et al.*, 1982

LeIF (20 000)	cDNA	β-lactamase	Short β-lactamase fusions	Met LeIF	Supernatant	1—2 mol/cell	Nagata *et al.*, 1980; Streuli *et al.*, 1980
Urokinase (35 000)	cDNA	β-lactamase	β-lactamase fusion	Enzymatically active proteins (hybrids?) (32 000—150 000)	Supernatant (periplasmic space)	Low	Ratzkin *et al.*, 1981
Mouse DHFR (22 000)	cDNA	β-lactamase	β-lactamase fusion	Met-DHFR	—	Low	Chang *et al.*, 1978; 1980
SV40 tAg (19 000)	Virus DNA clone	β-lactamase	Synthetic rbs fusion within β-lactamase gene	SV40 tAg	—	0.4%	Jay *et al.*, 1981
HBcAg (22 000)	Virus DNA clone	β-lactamase	β-lactamase fusion	HBcAg (native?)	—	Low	Burrell *et al.*, 1979; Pasek *et al.*, 1979
HBsAg (22 600)	Virus DNA clone	β-lactamase	Short β-lactamase fusions	HBsAg	Supernatant	Low	Mackay *et al.*, 1981
HSV TK (43 000)	Virus DNA clone	Tet gene (pBR322)	Tet gene fusion, internal rbs	Enzymatically active protein (40 000)	Supernatant	Low	Kit *et al.*, 1981

*Only constructions using P_L on a plasmid are considered.

(i.e. either rbs or coding sequence fusion) and of the polypeptide produced (either native or hybrid) are noted. The promoter from which the mRNA coding for the protein was transcribed is also indicated.

Several interesting points arise from Table 2. Firstly, there are very few reports of eukaryotic proteins expressed to really high levels in *E. coli*. The best examples are insulin, growth hormone, interferons (various) and the *trp* LE FMDV VPI fusion protein. It is significant that these proteins are the furthest advanced towards a clinical application. For example, insulin, prepared by recombinant DNA techniques is already under extensive clinical trial in the UK (Clark *et al.*, 1982). Similarly, interferon and growth hormone (Hintz *et al.*, 1982) are in clinical trials and the FMDV-VPI fusion is being tested as a vaccine (Kleid *et al.*, 1981a). By far the majority of the proteins in Table 2, however, have not been reported to be expressed to the same high levels. The importance of this limitation, as far as commercial considerations are concerned, is clearly related to the value of the product. For a protein such as SV40 tAg, the levels of expression may well be sufficient to enable the protein to be purified and its function and structure analysed (Thummel *et al.*, 1981) but for other proteins (e.g. chymosin, Emtage *et al.*, 1983) higher expression levels may well be necessary.

A Alternative promoters and constructions

The *lac*, *trp* and P_L promoters fulfill several criteria which make them attractive promoters to use to drive transcription of eukaryotic genes in *E. coli*. They are controlled by repressors which can be removed or inactivated when expression is required, and this controllability may be very important when considering proteins which are toxic to the cell (e.g. hydrophobic proteins). Nevertheless, there are several other strong promoters which have not yet been harnessed to express eukaryotic genes. The promoter for the outer membrane lipoprotein of *E. coli* (the most abundant protein in *E. coli*), a constitutive strong promoter, has been utilized to construct versatile expression vectors. The approach has been to make the *lpp* promoter controllable by inserting a *lac* promoter/operator fragment between it and the 5′-end of the lipoprotein gene. The expression of genes placed downstream of the promoters (the cloning of which is facilitated by a linker containing restriction sites) is then repressed in bacterial strains containing the I^q mutation (which overproduce repressor) but can be induced by IPTG. The 3′-end of the *lpp* gene is maintained in the vector to provide both transcriptional and translational stop signals (Nakamura *et al.*, 1982; Nakamura and Inouye,

1982). The presence of both promoters results in high levels of synthesis of lipoprotein and β-galactosidase when these genes are cloned in the vectors (Nakamura *et al.*, 1982).

A hybrid promoter consisting of the −35 region of the *trp* promoter fused to the *lac* −10 region and the *lac* operator has also been constructed. This promoter (the *tac* promoter) which is regulated by the *lac* repressor (from *lac* I^q) but has the strength of the *trp* promoter, has been used for the controlled expression of HGH (de Boer *et al.*, 1982). One of the problems with strong promoters, however, is that they are difficult to clone, presumably owing to transcriptional blocking of the origin of replication and other plasmid functions. The T-odd phages probably contain the strongest promoters known in terms of RNA polymerase complex formation and rate of RNA chain initiation *in vitro*. Attempts to clone the early promoter of T5 phage were unsuccessful until a vector containing the major terminator of phage *fd* was used (Gentz *et al.*, 1981); similar vectors are being used to clone the early promoter of phage T7. The demonstration by McAllister *et al.* (1981) that T7 RNA polymerase (an early gene product) can utilize T7 late promoters in plasmids, with its usual specificity, suggests a further expression system where the coding sequence for a desired protein is cloned behind a promoter for T7 polymerase and the polymerase itself is supplied to the cell by infection or from a cloned polymerase gene (McAllister *et al.*, 1981).

Another two-plasmid expression system is also being developed. Sninsky *et al.* (1981) have constructed vectors for the temperature regulated expression of cloned genes. In this system a plasmid, containing the *lac* repressor, which fails to replicate at 42°C, coexists with a plasmid that undergoes multicopy "runaway" replication at 42°C (Uhlin *et al.*, 1979) and which contains a cloned gene under the control of the *lac* promoter/operator. Concurrent derepression of the *lac* promoter and copy number amplification occur at elevated temperatures leading to extensive transcription of the cloned gene. As a model system, Sninsky *et al.* (1981) reported the high level expression of chloramphenicol transacetylase (the CAT gene product) at 42°C. Relatively high copy number vectors have also been prepared by cloning fragments of λ DNA into pBR322 (Rao and Rogers, 1978). In addition, the copy number of Col E1 has been increased by deletion mutagenesis (Twigg and Sherratt, 1980) and single base pair substitution (Muesing *et al.*, 1981). Whether these higher copy number vectors or the more complicated bipartite expression systems can be used on a large scale remains to be seen. One of the problems in a scaled up system could well be that of plasmid instability. It may be possible to control stability to some extent (in the absence of a selection, such as antibiotic resistance) by insertion of DNA sequences

which are involved in plasmid partitioning during growth (e.g. the *par* locus, Meacock and Cohen, 1980) into the vectors.

Another interesting expression system that has been developed recently utilizes the high copy number of the RF (replicative form) of the single stranded DNA phage M13. Slocombe *et al.* (1982) achieved high levels of expression of LeIF $\alpha 2$ (5 mg/l) by cloning LeIF cDNA adjacent to the *lac* promoter in M13 mp7 to make an NH_2-terminal-short β-galactosidase fusion protein.

B mRNA structure and stability

Considerable evidence has accumulated that the secondary structure of the mRNA around the rbs and AUG is important in determining its translational efficiency (Iserentant and Fiers, 1980; Shepard *et al.*, 1982; Gheysen *et al.*, 1982) but there have been few studies on the stability of the analogue of eukaryotic mRNA in *E. coli*. The expression of *N. crassa* dehydroquinase is increased in bacterial strains lacking polynucleotide phosphorylase, apparently owing to the fact that this phosphorylase is involved in the turnover of eukaryotic mRNA and its absence increases the half life of these molecules (Hautala *et al.*, 1979). The use of bacterial strains lacking polynucleotide phosphorylase has since been reported for expression of other proteins such as insulin (Talmadge and Gilbert, 1982) but not to any large extent, possibly because not all mRNAs are stabilised in this way.

C Nature of the proteins produced

Another important point to emerge from Table 2 is the nature of the proteins that have been produced from *E. coli*. A number of hormones and other effector proteins and some virus antigens have been expressed but there are only a few examples of eukaryotic proteins whose primary function is enzymatic, e.g. DHFR (Chang *et al.*, 1978; 1980); urokinase (Ratzkin *et al.*, 1981); Rous sarcoma virus *src* protein kinase (Gilmer and Erikson, 1981; McGrath and Levinson, 1982), HSV thymidine kinase (Garapin *et al.*, 1981; Kit *et al.*, 1981) and calf chymosin (Emtage *et al.*, 1982). The lack of examples of enzymes may simply reflect the fact that fewer experiments have been done with genes coding for eukaryotic enzymes than have been done with hormone genes. Alternatively it may be a reflection of the fact that large proteins do not seem to be made (or tolerated) as well in *E. coli* as smaller ones (c. 20 000 molecular weight, see Table 2); enzymes tend to be larger proteins than hormones.

The problem of degradation of "foreign" proteins in *E. coli* may well be difficult to resolve. Nevertheless, the observation that hybrid signal sequences can be used to transport insulin to the periplasmic space (Talmadge *et al.*, 1980a, b) has given the option of making constructions designed to ensure secretion (Talmadge and Gilbert, 1982). Extension of this approach to other proteins will be useful and vectors based on the *lpp* gene (a secreted protein) have been constructed for this purpose (see Nakamura and Inouye, 1982). The use of bacterial strains lacking one or more of the normal complement of proteases in *E. coli* (e.g. Lon$^-$, Gottesman *et al.*, 1981; for review of *E. coli* proteases see Goldberg *et al.*, 1981) may be another way of preventing or slowing down degradation of eukaryotic proteins.

Another aspect to be considered is that the conformation of the protein (determined by the primary amino acid sequence) may actually not be the same when made in *E. coli* as it is when made in its normal location. This might explain why several normally soluble proteins are found to be insoluble when made in *E. coli* (Table 2). It is interesting that the available evidence for HGH and leukocyte interferon which are both produced as soluble proteins in *E. coli* indicates that these small proteins (c. 20 000 molecular weight) have S-S bridges formed correctly (Kohr *et al.*, 1982; Wetzel, 1981). Lack of the proper S-S bridges or incorrect folding due to cross-bridging might lead to insolubility. Further work is clearly needed towards gaining an understanding of the conformation of eukaryotic proteins during their synthesis by *E. coli*.

D Other host-vector systems

Several other host and vector systems are being developed for the expression of eukaryotic DNA. Detailed discussion of them is outside the scope of this review but it is nevertheless worthwhile mentioning some of the alternatives to *E. coli* that are becoming available. There is considerable interest in developing expression systems in *B. subtilis* largely because this organism is a non-pathogenic, Gram positive bacterium which does not produce endotoxins. Moreover strains of *Bacillus* are used widely commercially and several secrete large amounts of extracellular proteins. So far, however, there are few reports of the successful expression of eukaryotic proteins in *B. subtilis*, even though molecular cloning in this organism is now an established technique (see Gryczan, 1982). As an example, Hardy *et al.* (1981) have recently reported the expression of native HBcAg and an FMDV-VPI fusion polypeptide in *B. subtilis* utilizing an erythromycin resistance gene promoter. The obligate methylotroph,

M. methylotrophus is also being used as an alternative host since it can be fermented on a large scale using cheap feedstock (ammonia and methanol). The vector is a composite plasmid (pGSS15) containing the antibiotic resistance genes of pBR322 and the broad host range characteristics of R300B, an Inc Q group plasmid. DHFR (from the β-lactamase promoter), an ovalbumin fusion protein and interferon (both from the *lac* promoter) have been expressed in this system (Hennam *et al.*, 1982; De Maeyer *et al.*, 1982).

Considerable effort is also going into developing cloning and expression systems in other organisms fermentable on a large scale such as *Streptomyces* (Chater *et al.*, 1982) and the yeast *S. cerevisiae* (Hollenberg, 1982). There are reports of the synthesis of cloned prokaryotic enzymes (Hollenberg, 1982) and low levels of ovalbumin in yeast (Mercereau-Puijalon *et al.*, 1980). More recently leukocyte interferon has been expressed to high levels in yeast, using either the promoter for the alcohol dehydrogenase I gene or the phosphoglycerate kinase gene promoter, in vectors based on the 2 μ circle (Hitzeman *et al.*, 1981; Tuite *et al.*, 1982). However, at the same time, it has been reported that rat growth hormone is not made in yeast using a similar expression vector containing the rat growth hormone gene (quoted in Hitzeman *et al.*, 1981).

Finally, there is increasing interest in using eukaryotic cells to express eukaryotic proteins. This might be considered rather obvious but it is only very recently that effective vectors have become available (mostly based on virus replicons) for eukaryotic cloning (for review, see Rigby, 1982). There is no doubt that these systems will become important for the analysis of the mechanism of eukaryotic gene expression and its control. It will be interesting to see whether the systems are also exploited for expressing cloned genes in the way *E. coli* has been. Some problems such as those of post-translational modification and stability may be diminished by a eukaryotic expression system but this advantage may be offset by, for example, greater difficulty in controlling transcription or lower yields of biomass or of the desired protein. No doubt, a comparison of the expression of cloned eukaryotic genes in *E. coli* with their expression in eukaryotic cells will be the subject of a future review.

X Acknowledgments

I would like to thank Spencer Emtage, Mike Doel, Gwyn Humphreys and particularly Norman Carey for their constructive criticisms of the manuscript and many other colleagues for helpful advice and information. I am also grateful to Margaret Turner and Jackie

Hardwick for typing the manuscript and to Ed Trewhella and Celltech's internal library service for provision of space and computer on-line search time.

XI References

Backman, K. and Ptashne, M. (1978). Maximizing gene expression on a plasmid using recombination *in vitro*. *Cell* **13**, 65—71.

Backman, K., Ptashne, M. and Gilbert, W. (1976). Construction of plasmids carrying the CI gene of bacteriophage λ. *Proc. Natn. Acad. Sci. U.S.A.* **73**, 4174—4178.

Bassford, P., Beckwith, J., Berman, M., Brickman, E., Casadaban, M., Guarente, L., Saint-Girous, I., Sarthy, A., Schwartz, M., Schuman, H. and Silhavy, T. (1978). Genetic fusions of the *lac* operon. A new approach to the study of biological processes. *In* "The Operon" (Miller, J. H. & Reznikoff, W. S., eds), pp. 245—261. Cold Spring Harbor.

Baty, D., Mercereau-Puijalon, O., Perrin, D., Kourilsky, P. L. and Lazdunsky, C. (1981). Secretion into the bacterial periplasmic space of chicken ovalbumin synthesized in *Escherichia coli*. *Gene* **16**, 79—87.

Beggs, J. D. (1982). Gene cloning in yeast. *In* "Genetic Engineering" (Williamson, R., ed). Vol. 2, 175—203. Academic Press, London.

Bernard, H. U., Remaut, E., Hershfield, M. V., Das, H. K., Helinski, D. R., Yanofsky, C. and Franklin, N. (1979). Construction of plasmid cloning vehicles that promote gene expression from the bacteriophage lambda P_L promoter. *Gene* **5**, 59—76.

Bollen, A., Glineur, C., Delcuve, G., Loriau, R. and Herzog, A. (1981). Expression in *Escherichia coli* of urokinase antigenic determinants. *Biochem. Biophys. Res. Comm.* **103**, 391—401.

Breathnach, R. and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**, 349—383.

Broome, S. and Gilbert, W. (1978). Immunological screening method to detect specific translation products. *Proc. Natn. Acad. Sci. U.S.A.* **75**, 2746—2749.

Brousseau, R., Scarpulla, R., Sung, W., Hsiung, H. M., Narang, S. A. and Wu, R. (1982). Synthesis of a human insulin gene V. Enzymatic assembly, cloning and characterization of the human proinsulin DNA. *Gene* **17**, 279—289.

Burrell, C. J., Mackay, P., Greenaway, P. J., Hofschneider, P. H. and Murray, K. (1979). Expression in *E. coli* of hepatitis B virus DNA sequences cloned in plasmid pBR322. *Nature* **279**, 43—47.

Casadaban, M., Chou, J. and Cohen, S. N. (1980). *In vitro* gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. *J. Bact.* **143**, 971—980.

Chan, S. J., Weiss, J., Konrad, K., White, T., Bahl, C., Yu, S. D., Marks, D. and Steiner, D. F. (1981). Biosynthesis and periplasmic segregation of human proinsulin in *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* **78**, 5401—5405.

Chang, A. C. Y. and Cohen, S. N. (1974). Genome construction between bacterial species *in vitro*: Replication and expression of *Staphylococcus*

plasmid genes in *Escherichia coli. Proc. Natn. Acad. Sci. U.S.A.* **71**, 1030—1034.

Chang, A. C. Y., Lansman, R. A., Clayton, D. A. and Cohen, S. N. (1975). Studies of mouse mitochondrial DNA in *Escherichia coli*: Structure and function of the eukaryotic — prokaryotic chimeric plasmids. *Cell* **6**, 231—244.

Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. and Cohen, S. N. (1978). Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature* **275**, 617—624.

Chang, A. C. Y., Erlich, H. A., Gunsalus, R. P., Nunberg, J. H., Kaufman, R. J., Schimke, R. T. and Cohen, S. N. (1980). Initiation of protein synthesis in bacteria at a translational start codon of mammalian cDNA: Effects of the preceding nucleotide sequence. *Proc. Natn. Acad. Sci. U.S.A.* **77**, 1442—1446.

Charnay, P., Perricaudet, M., Galibert, F. and Tiollais, P. (1978). Bacteriophage lambda and plasmid vectors allowing fusion of cloned genes in each of three translational phases. *Nucleic Acids Res.* **5**, 4479—4494.

Charnay, P., Louise, A., Fritsch, A., Perrin, D. and Tiollais, P. (1979). Bacteriophage lambda *Escherichia coli* K12 vector host system for gene cloning and expression under lactose promoter control II. DNA fragment insertio i at the vicinity of the *lac* UV 5 promoter. *Molec. Gen. Genet.* **170**, 171 - 178.

Charnay, P., Gervais, M., Louise, A., Galibert, F. and Tiollais, P. (1980). Biosynthesis of hepatitis B virus surface antigen in *Escherichia coli. Nature* **286**, 893—895.

Chater, K. F., Hopwood, D. A., Kiester, T. and Thompson, C. J. (1982). Gene cloning in *Streptomyces. Curr. Topics Microbiol. Immunol.* **96**, 69—95.

Clarke, A. J. L., Adeniyi-Jones, R. O., Knight, G., Leiper, J. M., Wiles, P. G., Jones, R. H., Keen, H., Maccuish, A. C., Ward, J. D., Watkins, P. J., Cauldwell, J. M., Glynne, A. and Scotton, J. B. (1982). Biosynthetic human insulin in the treatment of diabetes. *Lancet* **2**, 354—357.

Cohen, B. J. and Richmond, J. E. (1982). Electron microscopy of hepatitis B core antigen synthesized in *E. coli. Nature* **296**, 677—678.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Promoter sequences of eukaryotic protein coding genes. *Science* **209**, 1406—1414.

Craig, R. and Hall, D. (1983). *In* "Genetic Engineering" (Williamson, R., ed). Vol. 4, pp. 57—125. Academic Press, London.

Crea, R., Kraszeuski, A., Hirose, T. and Itakura, K. (1978). Chemical synthesis of genes for human insulin. *Proc. Natn. Acad. Sci. U.S.A.* **75**, 5765—5769.

Davis, B. D. and Tai, P.-C. (1980). The mechanism of protein secretion across membranes. *Nature* **283**, 433—438.

Davis, A. R., Nayak, D. P., Cleda, M., Hiti, A. L., Dowbenko, D. and Kleid, D. G. (1981). Expression of antigenic determinants of the haemagglutinin gene of a human influenza virus in *E. coli. Proc. Natn. Acad. Sci. U.S.A.* **78**, 5376-5380.

de Boer, H. A., Comstock, L. J., Yansura, D. G. and Heynecker, H. L. (1982). Construction of a tandem *trp-lac* promoter for efficient and controlled expression of the human growth hormone gene in *Escherichia coli. In* "Promoter Structure and Function" (Rodriquez, R. L. and Chamberlin, M. J., eds) Praegar Publishers, New York.

De Maeyer, E., Skup, D., Prasad, K. S. N., De Maeyer-Guignard, J., Williams, B.,

Sharpe, G., Pioli, D., Hennam, J., Schuch, W. and Atherton, K. T. (1982). Expression of a chemically synthesized human α interferon gene. *Proc. Natn. Acad. Sci. U.S.A.* 79, 4256–4259.

Derom, C., Gheysen, D. and Fiers, W. (1982). High level synthesis in *E. coli* of the SV40 small-t antigen under control of the bacteriophage lambda P_L promoter. *Gene* 17, 45–54.

Derynck, R., Remaut, E., Saman, E., Stanssens, P., De Clercq, E., Content, T. and Fiers, W. (1980). Expression of human fibroblast interferon gene in *E. coli*. *Nature* 287, 193–197.

Edens, L., Heslinga, L., Klok, R., Ledeboer, A. M., Maat, J., Toonen, M. Y., Visser, C. and Verrips, C. T. (1982). Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in *Escherichia coli*. *Gene* 18, 1–12.

Edge, M. D., Greene, A. R., Heathcliffe, G. R., Meacock, P. A., Schuch, W., Scanlon, D. B., Atkinson, T. C., Newton, C. R. and Markham, A. F. (1981). Total synthesis of a human leukocyte interferon gene. *Nature* 191, 756–762.

Edman, J. C., Hallewell, R. A., Valenzuela, P., Goodman H. M. and Rutter, W. J. (1981). Expression of hepatitis B surface and core antigens in *E. coli*. *Nature* 291, 503–506.

Emtage, J. S., Tacon, W. C. A., Catlin, G. H., Jenkins, B., Porter, A. G. and Carey, N. H. (1980). Influenza antigenic determinants are expressed from haemagglutinin genes cloned in *E. coli*. *Nature* 283, 171–174.

Emtage, J. S., Angal, S., Doel, M. T., Harris, T. J. R., Jenkins, B., Lilley, G. and Lowe, P. A. (1983). In preparation.

Erwin, C. R., Maurer, R. A. and Donelson, J. E. (1980). A bacterial cell that synthesises a protein containing the antigenic determinants of rat prolactin. *Nucleic Acids Res.* 8, 2537–3546.

Fraser, T. H. and Bruce, B. J. (1978). Chicken ovalbumin is synthesised and secreted by *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* 75, 5936–5940.

Fuller, F. (1982). A family of cloning vectors containing the *lac* UV5 promoter. *Gene* 19, 43–54.

Garapin, A. C., Colbere-Garapin, F., Cohen-Solal, M., Horodniceanu, F. and Kourilsky, P. (1981). Expression of herpes simplex virus type 1 thymidine kinase gene in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 78, 815–819.

Gatenby, A. A., Castleton, J. A. and Saul, M. W. (1981). Expression in *E. coli* of maize and wheat chloroplast genes for large subunit of ribulose bisphosphate carboxylase. *Nature* 291, 117–121.

Gelfand, D. H., Shepard, H. M., O'Farrell, P. H. and Polisky, B. (1978). Isolation and characterisation of a Col E1 derived plasmid copy-number mutant. *Proc. Natn. Acad. Sci. U.S.A.* 75, 55869–55873.

Gentz, R., Lagner, A., Chang, A. C. Y., Cohen, S. N. and Bujard, H. (1981). Cloning and analysis of strong promoters is made possible by the downstream placement of an RNA termination signal. *Proc. Natn. Acad. Sci. U.S.A.* 78, 4936–4940.

Gheysen, D., Iserentant, D., Derom, C. and Fiers, W. (1982). Systematic alteration of the nucleotide sequence preceding the translation initiation codon and the effects on bacterial expression of the cloned SV40 small-t antigen gene. *Gene* 17, 55–63.

Gilmer, T. M. and Erikson, R. L. (1981). Rous sarcoma virus transforming protein p60 src expressed in *E. coli* functions as a protein kinase. *Nature* 294, 771–773.

Gilmer, T. M., Parsons, J. T. and Erikson, R. L. (1982). Construction of plasmids for expression of Rous sarcoma virus transforming protein p60 src in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 79, 2152–2156.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heynecker, H. C., Yansura, D. G., Crea, R., Hirose, T., Kraszeuski, A., Itakura, K. and Riggs, A. D. (1979a). Expression in *E. coli* of chemically synthesised genes for human insulin. *Proc. Natn. Acad. Sci. U.S.A.* 76, 106–110.

Goeddel, D. V., Heynecker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G., Crea, R. and Seeburg, P. H. (1979b). Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 281, 544–548.

Goeddel, D. V., Yelverton, E., Ullrich, A., Heynecker, H. L., Miozzari, G., Holmes, W., Seeburg, P. H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, N., Sloma, A., Tabar, J. M., Gross, M., Familletti, P. C. and Pestka, S. (1980a). Human leukocyte interferon production in *E. coli* is biologically active. *Nature* 287, 411–416.

Goeddel, D. V., Shepard, H. M., Yelverton, E., Leung, D. and Crea, R. (1980b). Synthesis of human fibroblast interferon by *E. coli*. *Nucleic Acids Res.* 8, 4057–4074.

Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E. and Gray, P. W. (1981). The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature* 290, 20–26.

Goldberg, A. L., Swamy, K. H. S., Chung, C. H. and Larimore, F. S. (1980). Proteases in *Escherichia coli*. *Methods Enzymol.* 80, 680–702.

Gottesman, S., Gottesman, M., Shaw, J. E. and Pearson, M. L. (1981). Protein degradation in *E. coli*: The Lon mutation and bacteriophage lambda N and CII protein stability. *Cell* 24, 225–233.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R., (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 9, 443–r75.

Gray, P. W., Leung, D. W., Pennica, D., Yelverton, E., Wajarian, R., Simonsen, C. C., Derynck, R., Sherwood, P. J., Wallace, D. M., Berger, S. L., Levinson, A. D. and Goeddel, D. V. (1982). Expression of human immune interferon cDNA in *E. coli* and monkey cells. *Nature* 295, 503–508.

Greenblatt, J. (1981). Regulation of transcription termination by the N gene protein of bacteriophage λ. *Cell* 24, 8–9.

Grosjean, H. and Fiers, W. (1982). Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and selective codon usage in efficiently expressed genes. *Gene* 18, 199–209.

Gryczan, T. J. (1981). Molecualr cloning in *Bacillus subtilis*. *In* "The Molecular Biology of the Bacilli" (Dubnau, D., ed), Vol. 1, pp. 307–329. Academic Press, New York.

Guarente, L., Lauer, G., Roberts, T. M. and Ptashne, M. (1980). Improved methods for maximizing expressing of a cloned gene: A bacterium that synthesises rabbit β-globin. *Cell* 20, 543–553.

Hallewell, R. A. and Emtage, S. (1980). Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulated expression of foreign genes. *Gene* 9, 24–47.

Hardy, K., Stahl, S. and Kupper, H. (1981). Production in *B. subtilis* of hepatitis B core antigen and of major antigen of foot and mouth disease virus. *Nature* 293, 481–483.

Hautala, J. A., Bassett, C. L., Giles, N. H. and Kushner, S. R. (1979). Increased

expression of a eukaryotic gene in *E. coli* through stabilization of its messenger RNA. *Proc. Natn. Acad. Sci. U.S.A.* **76**, 5774—5778.

Heiland, I. and Gething, M. J. (1981). Cloned copy of the haemagglutinin gene codes for human influenza antigenic determinants in *E. coli. Nature* **292**, 851—852.

Hennam, J. F., Cunningham, A. E., Sharpe, G. S. and Atherton, K. T. (1982). Expression of eukaryotic coding sequences in *Methylophilus methylotrophus. Nature* **297**, 80—82.

Hershfield, M. V., Boyer, H. W., Yanofsky, C., Lovett, M. and Helinski, D. R. (1974). Plasmid Col E1 as a molecular vehicle for cloning and amplification of DNA. *Proc. Natn. Acad. Sci. U.S.A.* **71**, 3455—3459.

Hinnen, A. and Meyhack, B. (1982). Vectors for cloning in yeast. *Curr. Topics Microbiol. Immunol.* **96**, 101—117.

Hintz, R. L., Wilson, D. M., Finno, J., Rosenfeld, R. G., Bennett, A. and McClellan, B. (1982). Biosynthetic methionyl human growth hormone is biologically active in adult man. *Lancet* **1**, 1276—1279.

Hintzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G. and Hall, B. D. (1981). Expression of a human gene for interferon in yeast. *Nature* **293**, 717—722.

Hollenberg, C. P. (1982). Cloning with 2 μm DNA vectors and the expression of foreign genes in *Saccharomyces cerevisiae. Curr. Topics Microbiol. Immunol.* **96**, 119—144.

Hopkins, A. S., Murray, N. E. and Brammar, W. J. (1976). Characterisation of λ *trp*-transducing bacteriophages made *in vitro. J. Mol Biol.* **107**, 549—569.

Horwich, A., Koop, A. H. and Eckhart, W. (1980). Expression of a gene for the polyoma small t antigen in *Escherichia coli. J. Virol.* **36**, 125—132.

Ikemura, T. (1981). Correlation between the abundance of *E. coli* transfer RNAs and the occurrence of respective codons in its protein genes. *J. Mol. Biol.* **146**, 1—21.

Iserentant, D. and Fiers, W. (1980). Secondary structure of mRNA and efficiency of translation initiation. *Gene* **9**, 1—12.

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H. W. (1977). Expression in *E. coli* of a chemically synthesised gene for the hormone somatostatin. *Science* **198**, 1056—1063.

Jay, G., Khoury, G., Seth, A. K. and Jay, E. (1981). Construction of a general vector for efficient expression of mammalian proteins in bacteria: Uses of a synthetic ribosome binding site. *Proc. Natn. Acad. Sci. U.S.A.* **78**, 5543—5548.

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. and Ptashne, M. (1981). λ repressor and cro—components of an efficient molecular switch. *Nature* **294**, 217—223.

Kedes, L. H., Chang, A. C. Y., Houseman, D. and Cohen, S. N. (1975). Isolation of histone genes from unfractionated sea urchin DNA by sub culture and cloning in *E. coli. Nature* **255**, 533—538.

Keefer, L. M., Piron, M. A. and De Meyts, P. (1981). Human insulin prepared by recombinant DNA techniques and native human insulin interact identically with insulin receptors. *Proc. Natn. Acad. Sci. U.S.A.* **78**, 1391—1395.

Kelley, W. S., Chalmers, K. and Murray, N. E. (1977). Isolation and characterisation of a λ *pol* A transducing phage. *Proc. Natn. Acad. Sci. U.S.A.* **74**, 5362—5366.

Keshet, E., Rosner, A., Bernstein, Y., Gorecki, M. and Aviv, H. (1981). Cloning

of bovine growth hormone gene and its expression in bacteria. *Nucleic Acids Res.* 9, 19–30.

Kit, S., Otsuka, H., Qavi, H. and Kit, M. (1981). Functional expression of the herpes simplex virus thymidine kinase gene in *E. coli* K-12. *Gene* 16, 287–295.

Kleid, P. G., Yansura, D., Small, B., Dowbenko, D., Moore, D., Grubman, M., McKercher, P., Morgan, D. O., Robertson, B. H. and Bachrach, H. L. (1981a). Cloned viral protein vaccine for foot and mouth disease: Responses in cattle and swine. *Science* 214, 1125–1129.

Kleid, D. G., Yansura, D. G., Heynecker, H. I. and Miozzari, G. F. (1981b). A method of producing a polypeptide product and a plasmidic expression vehicle therefore, a method of creating an expression plasmid, a method of cleaving double stranded DNA and specific plasmids. *UK Patent Application 2073203 A.*

Kohr, W. J., Keck, R. and Harkins, R. N. (1982). Characterization of intact and trypsin-digested biosynthetic human growth hormone by high-pressure liquid chromatography. *Anal. Biochem.* 122, 348–359.

Kourilsky, P., Gros, D., Rougeon, F. and Mach, B. (1977). Transcription of a mammalian sequence under phage λ control. *Nature* 267, 637–639.

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9, 5233–5252.

Kupper, H., Keller, W., Kurz, C., Forss, S., Schaller, H., Franze, R., Strohmaier, K., Marquardt, O., Zaslavsky, V. G. and Hofschneider, P. H. (1981). Cloning of cDNA of major antigen of foot and mouth disease virus and expression in *E. coli. Nature* 289, 555–559.

Lathe, R. F., Lecocq, J. P. and Everett, D. (1983). *In* "Genetic Engineering" (Williamson, R., ed) Vol. 4, pp. 1–56. Academic Press, New York and London.

Lauer, G., Pastrana, R., Sherley, J. and Ptashne, M. (1981). Construction of overproducers of bacteriophage 434 repressor and cro proteins. *J. Mol. Appl. Genet.* 1, 139–147.

Lawn, R., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R., Seeburg, P. H. and Wion, K. L. (1981). The sequence of human serum albumin cDNA and its expression in *E. coli. Nucleic Acids Res.* 9, 6103–6114.

Mackay, P., Pasek, M., Magazin, M., Kovakic, R. T., Allet, B., Stahl, S., Gilbert, W., Schaller, H., Bruce, S. A. and Murray, K. (1981). Production of immunologically active surface antigens of hepatitis B virus by *Escherichia coli. Proc. Natn. Acad. Sci. U.S.A.* 78, 4510–4514.

Martial, J. A., Hallewell, R. A., Baxter, J. D. and Goodman, H. M. (1979). Human growth hormone: Complementary DNA cloning and expression in bacteria. *Science* 205, 602–607.

McAllister, W. T., Morris, C., Rosenberg, A. H. and Studier, F. W. (1981). Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.* 153, 527–544.

McGrath, J. P. and Levinson, A. D. (1982). Bacterial expression of an enzymatically active protein encoded by RSV *src* gene. *Nature* 295, 423–435.

Meacock, P. A. and Cohen, S. N. (1980). Partitioning of bacterial plasmids during cell division: A cis-acting locus that accomplishes stable plasmid inheritance. *Cell* 20, 529–540.

Mercereau-Puijalon, O., Royal, A., Carni, B., Garapin, A., Krust, A., Gannon, F.

and Kourilsky, P. (1978). Synthesis of an ovalbumin-like protein by *E. coli* K12 harbouring a recombinant plasmid. *Nature* 275, 505—510.

Mercereau-Puijalon, O., Lacroute, F. and Kourilsky, P. (1980). Synthesis of a chicken ovalbumin-like protein in the yeast *Saccharomyces cerevisiae*. *Gene* 11, 163—167.

Miller, J. H. and Reznikoff, W. S. (1980). The Operon. Cold Spring Harbor Laboratory.

Moir, A. and Brammer, W. J. (1976). The use of specialised transducing phages in the amplification of enzyme production. *Molec. Gen. Genet.* 149, 87—99.

Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M. and Helling, R. B. (1974). Replication and transcription of eukaryotic DNA in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 71, 1743—1747.

Mory, Y., Chernajovsky, Y., Feinstein, S. I., Chen, L., Nir, U., Weissenbach, J., Malpiece, Y., Tiollais, P., Marks, D., Ladner, M., Colby, C. and Revel, M. (1981). Synthesis of human interferon β in *Escherichia coli* infected by a λ phage recombinant containing a human genomic fragment. *Eur. J. Biochem.* 120, 197—202.

Muesing, M., Tamm, J., Shepard, H. M. and Polisky, B. (1981). A single base pair alteration is responsible for the DNA overproduction phenotype of a plasmid copy number mutant. *Cell* 24, 235—242.

Murray, N. E. and Kelley, W. S. (1979). Characterisation of λ *pol* A transducing phages; effective expression of the *E. coli* pol A gene. *Molec. Gen. Genet.* 175, 77—87.

Murray, N. E., Bruce, S. A. and Murray, K. (1979). Molecular cloning of the DNA ligase gene from bacteriophage T4. *J. Mol. Biol.* 132, 493—505.

Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsodi, J., Boll, W., Cantell, K. and Weissmann, C. (1980a). Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. *Nature* 284, 316—320.

Nagata, S., Mantei, N. and Weissmann, C. (1980b). The structure of one of the eight or more distinct chromosomal genes for human interferon-α. *Nature* 287, 401—408.

Nakamura, K. and Inouye, M. (1982). Construction of versatile expression cloning vehicles using the lipoprotein gene of *E. coli*. *EMBO Journal* 1, 771—775.

Nakamura, K., Masui, Y. and Inouye, M. (1982). Use of a *lac* promoter-operator fragment as a transcriptional control switch for expression of the constitutive *lpp* gene in *Escherichia coli*. *J. Mol. Appl. Genet.* 1, 289—299.

Olson, K. C., Fenno, J., Lin, N., Harkins, R. N., Snider, C., Kohr, W., Ross, M. J., Fodge, D., Prender, G. and Stebbing, N. (1981). Purified human growth hormone from *E. coli* is biologically active. *Nature* 293, 408—411.

O'Farrell, P. H., Polisky, B. and Gelfand, D. H. (1978). Regulated expression by readthrough translation from a plasmid-encoded β-galactosidase. *J. Bact.* 134, 645—654.

Panasenko, S. M., Cameron, J. R., Davis, R. W. and Lehman, I. R. (1977). Five-hundredfold overproduction of DNA ligase after induction of a hybrid λ lysogen constructed *in vitro*. *Science* 196, 188—189.

Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G. and Murray, K. (1979). Hepatitis B virus genes and their expression in *E. coli*. *Nature* 282, 575—579.

Polisky, B., Bishop, R. J. and Gelfand, D. H. (1976). A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria. *Proc. Natn. Acad. Sci. U.S.A.* 73, 3900—3904.

Rao, R. and Rogers, S. G. (1978). A thermoinducible λ phage-Col E1 plasmid chimera for the overproduction of gene products from cloned DNA segments. *Gene* 3, 247–263.

Ratzkin, B. and Carbon, J. (1977). Functional expression of cloned yeast DNA in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 74, 487–491.

Ratzkin, B., Lee, S. G., Schrenk, W. J., Roychoudhury, R., Chen, M., Hamilton, T. A. and Hung, P. P. (1981). Expression in *Escherichia coli* of biologically active enzyme by a DNA sequence coding for the human plasminogen activator urokinase. *Proc. Natn. Acad. Sci. U.S.A.* 78, 3313–3317.

Remaut, E., Stanssens, P. and Fiers, W. (1981). Plasmid vectors for high-efficiency expression controlled by the P_L promoter of coliphage lambda. *Gene* 15, 81–93.

Rigby, P. W. J. (1982). Expression of cloned genes in eukaryotic cells using vector systems derived from viral replicons. *In* "Genetic Engineering" (Williamson, R., ed), Vol. 3, pp. 88–141. Academic Press, London.

Roberts, J. L., Seeburg, P. H., Shine, J., Herbert, E., Baxter, J. D. and Goodman, H. M. (1979a). Corticotropin and β endorphin: Construction and analysis of recombinant DNA complementary to mRNA for the common precursor. *Proc. Natn. Acad. Sci. U.S.A.* 76, 2153–2157.

Roberts, T. M., Kacich, R. and Ptashne, M. (1979b). A general method for maximizing the expression of a cloned gene. *Proc. Natn. Acad. Sci. U.S.A.* 76, 760–764.

Roberts, T. M., Bikel, I., Rogers-Yocum, R., Livingston, D. M. and Ptashne, M. (1979c). Synthesis of simian virus 40 t antigen in *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* 76, 5596–5600.

Rose, J. K. and Shafferman, A. (1981). Conditional expression of the vesicular stomatitis virus glycoprotein gene in *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* 78, 6670–6674.

Rosenberg, M. and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Ann. Rev. Genetics* 13, 319–353.

Rosenfeld, R. G., Aggarwal, B. B., Hintz, R. L. and Dollar, L. A. (1982). Recombinant DNA-derived methionyl human growth hormone is similar in membrane binding properties to human pituitary growth hormone. *Biochem. Biophys. Res. Comm.* 106, 202–209.

Scott, G. M. and Tyrrell, D. A. J. (1980). Interferon: Therapeutic fact or fiction for the '80s? *Brit. Med. J.* 280, 1558–1562.

Seeburg, P. H., Shine, J., Martial, J. A., Ivarie, R. D., Morris, J. A., Ullrich, A., Baxter, J. D. and Goodman, H. M. (1978). Synthesis of growth hormone by bacteria. *Nature* 276, 795–598.

Shemyakin, M. F., Chestukin, A. V., Dolganov, G. M., Khodkova, E. M., Monastyrskaya, G. S. and Sverdlov, E. D. (1980). Leu-enkephalin purification from *E. coli* cells carrying the plasmid with fused synthetic leu-enkephalin gene. *Nucleic Acids Res.* 8, 6163–6173.

Shepard, H. M., Yelverton, E. and Goeddel, D. V. (1982). Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome binding sites. *DNA* 1, 125–131.

Shimatake, H. and Rosenberg, M. (1981). Purified λ regulatory protein CII positively activates promoters for lysogenic development. *Nature* 292, 128–132.

Shine, J. and Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature* 254, 34–38.

Shine, J., Fettes, I., Lan, N. C. Y., Roberts, J. L. and Baxter, J. D. (1980).

Expression of a cloned β-endorphin gene sequence by *Escherichia coli*. *Nature* 285, 456–461.

Shortle, D., Dimaio, D. and Nathans, D. (1981). Directed mutagenesis. *Ann. Rev. Genetics* 15, 265–294.

Siebenlist, U., Simpson, R. B. and Gilbert, W. (1980). *E. coli* polymerase interacts homologously with two different promoters. *Cell* 20, 269–281.

Slocombe, P., Easton, A., Boseley, P. and Burke, D. C. (1982). High level expression of an interferon α2 gene cloned in phage M13 mp7 and subsequent purification with a monoclonal antibody. *Proc. Natn. Acad. Sci. U.S.A.* 79, 5455–5459.

Smith, M. and Gillam, S. (1981). Constructed mutants using synthetic oligodeoxyribonucleotides as site-specific mutagens. *In* "Genetic Engineering" (Setlow, J. K. and Hollaender, A., eds) Vol. 3, pp. 1–32. Plenum Press, New York.

Sninsky, J. J., Uhlin, B. E., Gustafsson, P. and Cohen, S. N. (1981). Construction and characterisation of a novel two-plasmid system for accomplishing temperature-regulated amplified expression of cloned adventitious genes in *Escherichia coli*. *Gene* 16, 275–286.

Stahl, S., Mackay, P., Magazin, M., Bruce, S. A. and Murray, K. (1982). Hepatitis B virus core antigen: Synthesis in *E. coli* and application in diagnosis. *Proc. Natn. Acad. Sci. U.S.A.* 79, 1606–1610.

Steitz, J. A. (1979). Genetic signals and nucleotide sequences in messenger RNA. *In* "Biological Regulation and development 1. Gene Expression". (Goldberger, R. F., ed) pp. 349–399. Plenum Press, New York.

Streuli, M., Nagata, S. and Weissman, C. (1980). At least three human type α interferons: Structure of α_2. *Science* 209, 1343–1347.

Streuli, M., Hall, A., Boll, W., Stewart, W. E., Nagata, S. and Weissmann, C. (1981). Target cell specificity of two species of human interferon-α produced in *E. coli* and of hybrid molecules derived from them. *Proc. Natn. Acad. Sci. U.S.A.* 78, 2848–2852.

Stewart, W. E. II, Sarkar, F. H., Taira, H., Hall, A.. Nagata, S. and Weissmann, C. (1980). Comparisons of several biological and physiochemical properties of human leukocyte interferons produced by human leukocytes and by *E. coli*. *Gene* 11, 181–186.

Struhl, K., Cameron, J. R. and Davis, R. W. (1976). Functional genetic expression of eukaryotic DNA in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 73, 1471–1475.

Struhl, K. and Davis, R. W. (1977). Production of a functional eukaryotic enzyme in *Escherichia coli*: Cloning and expression of the yeast structural gene for imidazole-glycerol phosphate dehydratase (his 3). *Proc. Natn. Acad. Sci. U.S.A.* 74, 5255–5259.

Sung, W. L., Hsiung, H. M., Brousseau, R., Michniewicz, J., Wu, R. and Narang, S. A. (1979). Synthesis of the human insulin gene. Part II. Further improvements in the modified phosphotriester method and the synthesis of seventeen deoxyribooligonucleotide fragments constituting human insulin chains B and mini-C DNA. *Nucleic Acids Res.* 7, 2199–2212.

Sutcliffe, J. G. (1979). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* 43, 77–90.

Suzuki, M., Sumi, S.-I., Hasegawa, A., Nishizawa, T., Miyoshi, K.-I., Wakisaka, S., Miyake, T. and Misoka, F. (1982). Production in *Escherichia coli* of biologically active secretin a gastro-intestinal hormone. *Proc. Natn. Acad. Sci. U.S.A.* 79, 2475–2479.

Szybalski, E. H. and Szybalski, W. (1979). A comprehensive molecular map of bacteriophage λ. *Gene* 7, 217–270.

Tacon, W., Carey, N. H. and Emtage, S. (1980). The construction and characterization of plasmid vectors suitable for the expression of all DNA phases under the control of the *E. coli* tryptophan promoter. *Mol. Gen. Genet.* 177, 427–438.

Talmadge, K. and Gilbert, W. (1982). Cellular location affects protein stability in *Escherichia coli*. *Proc. Natn. Acad. Sci. U.S.A.* 79, 1830–1833.

Talmadge, K., Kaufman, J. and Gilbert, W. (1980a). Bacteria mature preproinsulin to proinsulin. *Proc. Natn. Acad. Sci. U.S.A.* 77, 3988–3992.

Talmadge, K., Stahl, S. and Gilbert, W. (1980b). Eukaryotic signal sequence transports insulin antigen in *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* 77, 3369–3373.

Talmadge, K., Brosius, J. and Gilbert, W. (1981). An internal signal sequence directs secretion and processing of proinsulin in bacteria. *Nature* 294, 176–178.

Tanaka, S., Oshima, T., Ohsue, K., Ono, T., Oikawa, S., Takano, I., Noguchi, T., Kangawa, K., Minamino, N. and Matsuo, H. (1982). Expression in *Escherichia coli* of chemically synthesized gene for a novel opiate peptide α neo-endorphin. *Nucleic Acids Res.* 10, 1741–1754.

Taniguchi, T., Guarente, L., Roberts, T. M., Kimelman, D., Douhan, J. and Ptashne, M. (1980). Expression of the human fibroblast interferon gene in *E. coli*. *Proc. Natn. Acad. Sci. U.S.A.* 77, 5230–5233.

Thummel, C. S., Burgess, T. L. and Tjian, R. (1981). Properties of simian virus 40 small t antigen overproduced in bacteria. *J. Virol.* 37, 683–697.

Tuite, M. F., Dobson, M. J., Roberts, N. A., King, R. M., Burke, D. C., Kingsman, S. M. and Kingsman, A. J. (1982). Regulated high efficiency expression of human interferonalpha in *Saccharomyces cerevisiae*. *EMBO Journal* 1, 603–608.

Twigg, A. and Sherratt, D. (1980). Trans complementable copy number mutants of plasmid Col E1. *Nature* 283, 216–218.

Uhlin, B. E., Molin, S., Gustafsson, P. and Nordstrom, K. (1979). Plasmids with temperature-dependent copy number for amplification of cloned genes and their products. *Gene* 6, 91–106.

Vapnek, D., Hautala, J., Jacobson, J. W., Giles, N. H. and Kushner, S. R. (1977). Expression in *Escherichia coli* K12 of the structural gene for catabolic dehydroquinase of *Neurospora crassa*. *Proc. Natn. Acad. Sci. U.S.A.* 74, 3508–3512.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. and Gilbert, W. (1978). A bacterial clone synthesising proinsulin. *Proc. Natn. Acad. Sci. U.S.A.* 75, 3727–3731.

Weck, P.K., Apperson, S., Stebbing, N., Gray, P. W., Leung, D., Shepard, H. M. and Goeddel, D. V. (1981). Antiviral activities of hybrids of two major human leukocyte interferons. *Nucleic Acids Res.* 9, 6153–6166.

Wetzel, R. (1981). Assignment of the disulphide bonds of leukocyte interferon. *Nature* 289, 606–607.

Wetzel, R., Heyneker, H. L., Goeddel, D. V., Thurman, G. B. and Goldstein, A. L. (1980). Production of biologically active Nα-desacetylthymosin α_1 in *E. coli* through expression of a chemically synthesized gene. *Biochemistry* 19, 6096–6104.

Wetzel, R., Kleid, D. G., Crea, R., Heyneker, H. L., Yansura, D. G., Hirose, T., Kraszeuski, A., Riggs, A. D., Itakura, K. and Goeddel, D. V. (1981a). Expression in *E. coli* of a chemically synthesized gene for a mini C analog of human proinsulin. *Gene* 16, 63–71.

Wetzel, R., Perry, L. J., Estell, D. A., Lin, N., Levine, H. L., Slinker, B., Fields, F., Ross, M. J. and Shively, J. (1981b). Properties of a human alpha-interferon purified from *E. coli* extracts. *J. Interferon Res.* 1, 381–390.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature* 289, 751–758.

Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., Van Cleemput, M. and Wu, A. M. (1981). The complete nucleotide sequence of the tryptophan operon of *E. coli*. *Nucleic Acids Res*, 9, 6647–6668.

Yelverton, E., Leung, D., Weck, P., Gray, P. W. and Goeddel, D. V. (1981). Bacterial synthesis of a novel human leukocyte interferon. *Nucleic Acids Res*, 9, 731–741.