Journal of Industrial Microbiology, 1 (1987) 277–282
Elsevier
SIM 00044

Review

Factors affecting the expression of foreign proteins in *Escherichia coli*

Bernard R. Glick and Gordon K. Whitney
Department of Biology, University of Waterloo, Waterloo, Ontario, Canada

Received 1 July 1986
Revised 20 October 1986
Accepted 4 November 1986

Key words: Recombinant DNA; Gene expression; Genetic engineering; Biotechnology

SUMMARY

A variety of factors affect the expression of foreign proteins in *Escherichia coli.* These include: promoter strength, efficiency of ribosome binding, stability of the foreign protein in *E. coli,* location of the foreign protein in *E. coli,* the codons used to encode the foreign protein, the metabolic state of the cell, and the location, stability and copy number of the foreign gene. This paper contains a critical review of these factors with the idea that a detailed understanding of them is the key to the development of strategies for the efficient large-scale production of foreign proteins in *E. coli.*

In the past few years recombinant DNA technology has enabled scientists to produce a large number of diverse proteins, in microorganisms, that were previously either unavailable, relatively expensive or difficult to obtain in quantity. While the expression of foreign genes has been reported in a variety of microorganisms and cell lines, most of this work utilizes *E. coli* for the cloning and expression of foreign genes [6]. Despite the fact that the use of organisms other than *E. coli* as recipients of foreign DNA continues to increase, in the short term most of the commercial protein products produced by recombinant DNA technology will be cloned and expressed in *E. coli.* Therefore, this review describes strategies and approaches that have been developed to optimize the expression of foreign proteins in *E. coli.*

The expression of foreign genes in *E. coli* is influenced by a range of factors including: the "strength" of the transcriptional promoter, the "strength" of the ribosome binding site (which regulates the frequency of translation of the messenger RNA), the stability of the cloned protein in *E. coli,* the location of the cloned protein within the *E. coli* cell, the codons utilized in the foreign gene compared to the normal pattern of codon usage in *E. coli,* the metabolic state of the cell, the number of copies of the foreign gene, and the stability of the foreign gene [6,11].

Many plasmid and bacteriophage vectors have been developed in which the cloned gene is situated immediately downstream from a strong transcriptional promotor [6,11]. Use of these vectors requires

Correspondence: Dr. B.R. Glick, Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

0169-4146/87/$03.50 © 1987 Society for Industrial Microbiology

that the promoter not be constitutive (i.e., always turned on) but, rather, be turned on at a specific stage in the growth of the transformed *E. coli* cells. This is often accomplished by the addition of a specific metabolite or by a shift in the temperature of the growth medium. Regulation of promoter activity ensures that the expression of a foreign gene does not interfere with normal cellular gene functions and is not deleterious to the cell. Failure to regulate the expression of strong promoters often results in the loss of the plasmid carrying the strong promoter or the constitutive expression of the strong promoter which may be lethal to the cell.

The most widely used strong promoters are from the *E. coli trp* and *lac* operons, the *tac* promoter (an in vitro construct including elements from both the *trp* and *lac* promoters), and the leftward, or pL, promoter of bacteriophage lambda [9,10,30,44]. The *lac* promoter is negatively regulated by the *lac* repressor protein. Derepression of the *lac* promoter is achieved by the addition of isopropyl-β-D-thiogalactoside (IPTG), a gratuitous inducer which prevents the *lac* repressor from binding to the *lac* operator. In practice, an "up" mutation of the *lac* promoter, *lac* UV5, is often used in place of the wild type *lac* promoter as a component of high expression vectors [12].

The *trp* promoter is negatively regulated, by a complex between tryptophan and the *trp* repressor protein. Derepression of the *trp* promoter is achieved by the addition of 3-indoacrylic acid (IAA). The number of copies of the *lac* or *trp* repressor protein relative to the number of copies of the *lac* or *trp* promoter determines the extent of "background" transcription from these promoters, (i.e., transcription in the absence of added inducer). If the *lac* or *trp* promoter is on a multi-copy plasmid and the repressor protein is present as a single copy, usually as part of the *E. coli* chromosome, expression may be constitutive due to an insufficient amount of repressor protein [3], placing an unacceptably high metabolic load on the *E. coli* cells.

The *tac* promoter [9] is an in vitro construct composed of the -10 region from the *lac* promoter and the -35 region from the *trp* promoter. The *tac* promoter is repressed by the *lac* repressor and is derepressed by the addition of IPTG. It has been reported that the *tac* promoter is up to 10-fold "stronger" than the *lac* UV5 promoter [9], although promoter strength is not readily quantified [34] so that comparisons of this sort, especially between strong promoters, may not be meaningful.

The pL promoter is also negatively regulated by a temperature-sensitive cI repressor protein. Cells carrying a temperature-sensitive cI repressor are first grown at 28–30°C, where the cI repressor is functional. At this temperature the cI repressor prevents transcription from the pL promoter. When the cell culture reaches the mid-log stage of growth, the temperature may be shifted to 42°C thereby inactivating the cI repressor protein, and active transcription from the pL promoter ensues. The gene for the cI repressor may be on the same plasmid as the pL promoter, on a separate plasmid or F factor, or integrated in the bacterial chromosome, usually as part of an excision-defective bacteriophage lambda lysogen [4,25,26,30,31]. Transcription from the pL promoter is dependent on the amount of cI repressor protein relative to the number of copies of the pL promoter [30,31 and B. Glick, unpublished observations].

It was reported recently that the addition of a short specific DNA sequence (approximately 89 base pairs) to the distal end of cloned genes may stabilize the mRNA transcribed from that gene, thereby increasing gene expression [46]. This "retroregulator" sequence probably becomes incorporated at the 3′ end of the mRNA, protecting it from exonuclease digestion.

Gene expression in procaryotes such as *E. coli* may be regulated at the translational as well as at the transcriptional level. For example, whereas the RNA bacteriophage MS2 contains all four of its coding sequences on a single polycistronic mRNA, the encoded proteins are produced in varying amounts; e.g., 60 molecules of coat protein are produced for every one molecule of maturation protein [24]. While a comprehensive description of the regulation of procaryotic translation initiation remains to be elucidated, there are, nevertheless, some features of mRNA which are known to affect this process. These include the Shine-Dalgarno (S-D) se-

quence at the 5′ end of the mRNA that is thought to help position the mRNA on the ribosome, the distance between the S-D sequence and the initiation codon and the secondary and tertiary structure of the mRNA [42].

In an effort to increase foreign gene expression, several attempts have been made to construct plasmid vectors that optimize the complementarity between the S-D sequences on the mRNA and the 3′ end of the 16 S ribosomal RNA. In addition, varying the distance between the S-D sequence and the translation initiation codon as well as varying the nucleotide sequences 5′ to the S-D sequence region have been shown to affect expression levels [8,21,39,41]. Construction of a series of chimeric plasmids, in which the human fibroblast interferon gene is under the control of the *trp* promoter and of variable distance between the S-D sequence and the initiation codon, revealed that the expression of interferon could vary by a factor of 250 [39]. This difference was attributed to differences in mRNA secondary structure within the ribosome binding site affecting mRNA stability.

In addition to changes in the region of the mRNA 5′ to the initiation codon, it was observed that introduction of additional codons 3′ to the initiation codon, that altered the NH_2-terminal amino acid sequence of cloned bovine growth hormone, resulted in a dramatic (60-fold) increase in the amount of bovine growth hormone synthesized by cells transformed with this plasmid construct [36]. These results indicate that the secondary and/or tertiary structure of the mRNA in the region surrounding the initiation codon is critically important in determining the ultimate level of expression of a protein whose gene has been cloned in *E. coli* and that both translational and transcriptional regulation must be considered in the design of procaryotic expression vectors.

Since the 20 amino acids are encoded by 61 different trinucleotide codons, several trinucleotide codons can encode the information for the insertion of the same amino acid into protein. Organisms show marked differences in codon preference, especially in genes encoding highly expressed proteins [14,16,33]. In fact, it appears that the frequency of codon usage in an organism is a direct reflection of the pool of cognate tRNAs. Highly expressed genes use codons for which there is a large pool of cognate tRNAs while regulatory genes often use codons for which there is only a very small pool of cognate tRNAs [14,16,22,33]. Accordingly, expression of a foreign gene may be limited by the availability of a particular aminoacyl-tRNA. Expression limitations due to codon usage may be avoided when developing a strategy for chemical synthesis of a gene whose subsequent optimal expression is desired.

While *E. coli* proteins are synthesized in the cytoplasm, it is possible to direct a cloned gene product to the cytoplasm, the inner or outer membrane or the periplasmic space. Secretion of a cloned gene product to the periplasmic space often allows for higher levels of expression of the foreign protein that might be degraded by proteases in the cytoplasm [18]. *E. coli* is capable of recognizing and correctly processing eucaryotic (as well as procaryotic) signal sequences so that secretion of eucaryotic proteins like rat proinsulin into the *E. coli* periplasmic space is possible [15,43]. Expression vectors have recently been constructed which place the genes for foreign proteins, not normally secreted, behind a DNA fragment encoding a signal sequence. This results in the foreign protein being efficiently secreted (in large amounts) to the periplasmic space with no evidence for accumulation of the unprocessed form in the cytoplasm [13]. While it is technically feasible to direct the protein products of foreign genes to the inner or outer membrane, high levels of a foreign protein in the membrane may interfere with normal cellular functions and be lethal to the cell.

The overproduction of cloned proteins in *E. coli* may result in the formation of an insoluble aggregate (or inclusion body) of the foreign protein within the *E. coli* cytoplasm [7,37]. Depending on the particular protein, solubilization of this aggregate may result in the loss of biological activity. In this instance it may be advantageous to clone the gene into a secretion vector so that the cloned protein does not accumulate in the cytoplasm.

The large-scale production of eucaryotic pro-

teins in *E. coli* is often limited by the instability of these polypeptides within the bacterial host. A common strategy which has been used to overcome this problem is to fuse the gene for the eucaryotic protein to a portion of a bacterial gene [20,27]. With this type of construct, the desired polypeptide constitutes a small portion of the fusion product. An alternate approach to stabilizing a cloned protein is to clone multiple copies of the gene in tandem onto the same plasmid [17,23,38]. This approach has been successfully employed with both the human alpha interferon and the human proinsulin genes [23,38].

After a foreign gene has been cloned into an expression vector, the vector is introduced into competent *E. coli* cells that become a source of the foreign protein. However, plasmids are not always stable, especially in cells grown for many generations in large-scale cultures [28], so that when a process is scaled up it is important that vector stability be addressed.

A two-stage model of plasmid loss has been proposed [2] in which plasmid loss first occurs at the level of the individual cell as a result of defective segregation at cell division, and then at the population level. Since a plasmid-free strain has a faster specific growth rate than a plasmid-containing strain, as a result of the metabolic energy which is expended for plasmid maintenance, the plasmid-free strain will eventually outcompete the plasmid-containing strain [2]. It has been observed that as the expression of a cloned gene product increases, the appearance of plasmid-free segregants also increases, presumably as a result of the increase in metabolic energy required for plasmid maintenance and function. In practice, maximal levels of gene expression may occur when the plasmid copy number is relatively low [1,19].

The most common method of ensuring that a recombinant plasmid is not lost during the growth of the microorganism is the inclusion of antibiotics which select for the presence of plasmids carrying the appropriate antibiotic resistance genes. However, scale-up of this approach may not be economically feasible due to the cost of the added antibiotics [28,35].

A unique strategy for enhancing plasmid stability is to incorporate a gene onto the plasmid which allows the plasmid-containing strain to grow more rapidly than the plasmid-free strain. This approach, while somewhat limited in application, has been successfully employed in the transformation of a methylotrophic bacterium where the introduction of a recombinant plasmid effectively removed growth inhibition of the one carbon substrate [35].

An alternative strategy is to utilize cloning vectors with regulatable promoters so that the foreign protein is not expressed until the promoter is "turned on", thus minimizing the metabolic load placed on the cell. An analogous strategy involves the use of runaway-replication plasmid vectors where plasmid copy number is relatively low at lower temperatures and is increased when the temperature is raised [5,31,45]. The lower plasmid copy number during much of the cell growth cycle reduces the metabolic load on the cell and ensures plasmid stability. At the same time the higher plasmid copy number for a portion of the growth cycle results in high levels of expression of the cloned foreign gene.

A consequence of the use of cloning vectors with regulatable promoters is the development of two-stage fermentors in which the cell growth and gene expression phases are separated [32,40]. Using a two-stage fermentor, cells are grown to a predetermined stage, under conditions in which the plasmid-encoded foreign gene is not expressed, before they are transferred to a second fermentor where expression of the foreign gene is induced. Separation of growth and expression phases of the process reduces the metabolic load and increases plasmid stability, even in the absence of direct selective pressure. This makes it possible for the overall process to be carried out continuously, increasing the productivity and decreasing the capital costs of the process.

Regardless of the nature of the foreign gene or the design of the fermentor, the introduction of an exogenous plasmid into an *E. coli* cell is bound to impose some metabolic load. This may be avoided by integrating the foreign gene into the *E. coli* chromosome through the use of a defective bacterio-

phage lambda lysogen carrying the foreign gene (see, e.g., Ref. 26) or by the direct insertion of a foreign gene into a specific site on the host chromosome [29]. With the direct insertion strategy, a foreign gene is cloned into the middle of an *E. coli* gene on a plasmid and when this construct is used to transform *E. coli* cells, homologous recombination occurs so that the foreign gene and the DNA sequences flanking it become integrated into the chromosomal DNA.

While the efficient expression of foreign genes in *E. coli* is dependent on a number of factors, it is nevertheless reasonable to expect that most foreign genes may be expressed at high levels in *E. coli* and that this expression will be amenable to scale-up. Although the strategy of gene expression and scale-up is likely to vary, there are more similarities than differences from one gene to the next, resulting in the development of a "systems" approach to the cloning, expression and scale-up of foreign genes in *E. coli.*

REFERENCES

1 Aiba, S., H. Tsunekawa and T. Imanaka. 1982. New approach to tryptophan production by *Escherichia coli:* genetic manipulation of composite plasmids *in vitro*. Appl. Environ. Microbiol. 43: 289–297.

2 Ashby, R.E. and K.A. Stacey. 1984. Stability of a plasmid F Trim in populations of a recombination-deficient strain of *Escherichia coli* in continuous culture. Antonie van Leeuwenhoek 50: 125–134.

3 Backman, K., M. Ptashne and W. Gilbert. 1976. Construction of plasmids carrying the cI gene of bacteriophage lambda. Proc. Natl. Acad. Sci. USA 73: 4174–4178.

4 Bernard, H.-U. and D.R. Helinski. 1979. Use of the lambda phage promoter pL to promote gene expression in hybrid plasmid cloning vehicles. Methods Enzymol. 68: 482–492.

5 Bittner, M. and D. Vapnek. 1981. Versatile cloning vectors derived from the runaway-replication plasmid pKN402. Gene 15: 319–329.

6 Carrier, M.J., M.E. Nugent, W.C.A. Tacon and S.B. Primrose. 1983. High expression of cloned genes in *E. coli* and its consequences. Trends Biotechnol. 1: 109–113.

7 Cheng, Y.E., D.Y. Kwoh, T.J. Kwoh, B.C. Soltvedt and D. Zipser. 1981. Stabilization of a degradable protein by its overexpression in *Escherichia coli.* Gene 14: 121–130.

8 Crowl, R., C. Seamans, P. Lomedico and S. McAndrew. 1985. Versatile expression vectors for high-level synthesis of cloned gene products in *Escherichia coli.* Gene 38: 31–38.

9 De Boer, H.A., L.J. Comstock and M. Vasser. 1983. The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. USA 80: 21–25.

10 Dworkin-Rastl, E., P. Swetly and M.B. Dworkin. 1983. Construction of expression plasmids producing high levels of human leukocyte-type interferon in *Escherichia coli.* Gene 21: 237–248.

11 Friesen, J.D. and G. An. 1983. Expression vehicles used in recombinant DNA technology. Biotechnol. Adv. 1: 205–227.

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

13 Ghrayeb, J., H. Kimura, M. Takahara, H. Hsiung, Y. Masui and M. Inouye. 1984. Secretion cloning vectors in *Escherichia coli.* EMBO J. 3: 2437–2442.

14 Gouy, M. and C. Gautier. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucl. Acids Res. 10: 7055–7074.

15 Gray, G.L., J.S. Baldridge, K.S. McKeown, H.L. Heynecker and C.N. Chang. 1985. Periplasmic production of correctly processed human growth hormone in *Escherichia coli:* natural and bacterial signal sequences are interchangeable. Gene 39: 247–254.

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

17 Hartley, J.L. and T.J. Gregori. 1981. Cloning multiple copies of a DNA segment. Gene 13: 347–353.

18 Hoffman, C.S. and A. Wright. 1985. Fusions of secreted proteins to alkaline phosphatase: An approach for studying protein secretion. Proc. Natl. Acad. Sci USA 82: 5107–5111.

19 Imanaka, T. and S. Aiba. 1981. A perspective on the application of genetic engineering: stability of recombinant plasmid. Ann. N.Y. Acad. Sci. 369: 1–14.

20 Itakura, K., T. Hirose, R. Crea and A.D. Riggs. 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198: 1056–1063.

21 Jay, G., G. Khoury, A.K. Seth and E. Jay. 1981. Construction of a general vector for efficient expression of mammalian proteins in bacteria: Use of a synthetic ribosome binding site. Proc. Natl. Acad. sci USA 78: 5543–5548.

22 Konigsberg, W. and G.N. Godson. 1983. Evidence for use of rare codons in the *dna* G gene and other regulatory genes of *Escherichia coli.* Proc. Natl. Acad. Sci. USA 80: 687–691.

23 Lee, N., J. Cozzitorto, N. Wainwright and D. Testa. 1984. Cloning with tandem gene systems for high level gene expression. Nucl. Acids Res. 12: 6797–6812.

24 Lewin, B. 1977. Gene Expression: Plasmids and Phages, Vol. 3, pp. 790–824, Wiley-Interscience, Toronto.

25 Mieschendahl, M. and B. Muller-Hill. 1985. F′-Coded, temperature-sensitive lambda cI857 repressor gene for easy construction and regulation of lambda promoter-dependent expression systems. J. Bacteriol. 164: 1366–1369.

26 Murray, N.E., S.A. Bruce and K. Murray. 1979. Molecular

cloning of the T4 DNA ligase gene from bacteriophage T4. J. Mol. Biol. 132: 493–505.
27 Nagai, K. and H.C. Thorgersen. 1984. Generation of B-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli.* Nature 309: 810–812.
28 Pierce, J. and S. Gutteridge. 1985. Large-scale preparation of ribulosebisphosphate carboxylase from a recombinant system in *Escherichia coli* characterized by extreme plasmid instability. Appl. Environ. Microbiol. 49: 1094–1100.
29 Raiboud, O., M. Mock and M. Schwartz. 1984. A technique for integrating any DNA fragment into the chromosome of *Escherichia coli.* Gene 29: 231–241.
30 Remaut, E., P. Stanssens and W. Fiers. 1981. Plasmid vectors for high-efficiency expression controlled by the pL promoter of coliphage lambda. Gene 15: 81–93.
31 Remaut, E., H. Tsao and W. Fiers. 1983. Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. Gene 22: 103–113.
32 Robinson, C.W., B.R. Glick, J.J. Pasternak and M. Moo-Young. 1984. Continuous process for production, recovery and purification of rDNA products. First Meeting of the International Biotechnology Network on Downstream Processing, Toulouse, France, Abstract No. 35.
33 Robinson, M., R. Lilley, S. Little, J.S. Emtage, G. Yarranton, P. Stephens, A. Millican, M. Eaton and G. Humphreys. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli.* Nucl. Acids Res. 12: 6663–6671.
34 Rosenberg, M., A.B. Chepelinsky and K. McKenney. 1983. Studying promoters and terminotors by gene fusion. Science 222: 734–739.
35 Ryder, D.F. and D. DiBiasio. 1984. An operational strategy for unstable recombinant DNA cultures. Biotechnol. Bioeng. 26: 942–947.
36 Schoner, B.E., H.M. Hsiung, R.M. Belagaje, N.G. Mayne and R.G. Schoner. 1984. Role of mRNA translational efficiency in bovine growth hormone expression in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 81: 5403–5407.
37 Schoner, R.G., L.F. Ellis and B.E. Schoner. 1985. Isolation and purification of protein granules from *Escherichia coli* cells overproducing bovine growth hormone. Bio/Technology 3: 151–153.
38 Shen, S.H. 1984. Multiple joined genes prevent product degradation in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 81: 4627–4631.
39 Shepard, H.M., E. Yelverton and D.V. Goeddel. 1982. Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome binding sites. DNA 1: 125–130.
40 Siegel, R. and D.Y. Ryu. 1985. Kinetic study of instability of recombinant plasmid pPLc32 *trpAI* in *E. coli* using two-stage continuous culture system. Biotechnol. Bioeng. 27: 28–33.
41 Stanssens, P., E. Remaut and W. Fiers. 1985. Alterations upstream from the Shine-Dalgarno region and their effect on bacterial gene expression. Gene 36: 211–223.
42 Stormo, G.D., T.D. Schneider and L.M. Gold. 1982. Characterization of translational initiation sites in *E. coli.* Nucl. Acids Res. 10: 2971–2996.
43 Talmadge, K., J. Kaufman and W. Gilbert. 1980. Bacteria mature preproinsulin to proinsulin. Proc. Natl. Acad. Sci. USA 77: 3988–3992.
44 Tanaka, S., T. Oshima, K. Ohsuye, T. Ono, A. Mizono, A. Ueno, H. Nakazato, M. Tsujimoto, N. Higashi and T. Noguchi. 1983. Expression in *Escherichia coli* of a chemically synthesized gene for the human immune interferon. Nucl. Acids Res. 11: 1707–1723.
45 Uhlin, B.E., S. Molin, P. Gustafsson and K. Nordstrom. 1979. Plasmids with temperature-dependent copy number for amplification of cloned genes and their products. Gene 6: 91–106.
46 Wong, H.C. and S. Chang. 1986. Identification of a positive retroregulator that stabilizes mRNAs in bacteria. Proc. Natl. Acad. Sci. USA 83: 3233–3237.