Appl Microbiol Biotechnol (1993) 39:324-328

Applied and Microbiology Biotechnology
© Springer-Verlag 1993

Optimization of the synthesis of porcine somatotropin in *Escherichia coli*

H. Wang[1], **D. J. O'Mahony**[1], **D. J. McConnell**[1], **S.-Z. Qi**[2]

[1] Department of Genetics, Trinity College, Dublin 2, Ireland
[2] Laboratory of Animal Biochemistry, Beijing Agricultural University, Beijing 100094, China

Received: 12 August 1992/Accepted: 22 December 1992

Abstract. We report on the influence of choice of promoter and RNA polymerase, 5′-untranslated regions and ribosome binding sites, codon usage, leader peptide coding sequences and poly A tail in the 3′-untranslated region on the synthesis of porcine somatotropin (PST) in *Escherichia coli.* A total of 12 different constructs were tested in this study for the production of porcine somatotropin (PST) in *E. coli.* Several factors have significant effects on PST synthesis. In the presence of a strong promoter and a strong ribosome binding site, the next most important factor seems to be the combination of sequences at the 5′-end of the mRNA including both the 5′-untranslated region and the start of the coding sequence. Codon usage in the 5′-coding sequence per se is not important in determining the level of PST synthesis where high level expression is achieved from a strong ribosome binding site. However, where low level synthesis of recombinant PST (rPST) is achieved, codon usage in the 5′-coding sequence is important in determining the level of PST synthesis. Leader sequences dramatically reduce the level of PST synthesis. The presence of a poly A tail in the 3′-untranslated region has no significant effect on PST synthesis.

Introduction

Somatotropin is synthesized in the pituitary gland as a precursor protein (fPST) where it is processed by removal of the N-terminal leader peptide and release of the biologically active hormone into the bloodstream. Porcine somatotropin (PST) plays an important role in the linear growth of pigs, especially during the pre-adult period. Administration of recombinant porcine somatotropin (rPST) to pigs has been shown to alter nutrient partitioning and results in a number of commercially beneficial effects including increased feed efficiency, increased growth rates, improved carcass quality and earlier marketing.

We have previously described the cloning, sequence analysis and polymorphism of full-length cDNAs coding for PST and the use of the T7 RNA polymerase and the T7 lysozyme system to express rPST in *E. coli*, achieving levels of up to 40% of total cell protein (O'Mahony et al. 1989, 1990). Moreover we have shown that purified preparations of rPST produced by our system are biologically active, capable of restoring bone growth following administration to hypophysectomized rats (O'Mahony et al. 1990). In our interest to optimize the expression and purification of rPST in *E. coli* we have studied a number of factors that are believed to influence efficient heterologous gene expression in *E. coli.* We report on the influence of promoters (and transcribing RNA polymerase), 5′-untranslated regions and ribosome binding sites, codon usage in the 5′-region of the *pst* gene, leader peptide coding sequences and poly A tail in the 3′-untranslated region on the synthesis of rPST.

* *Present address:* Centre for Cardiovascular Science, Department of Medicine and Experimental Therapeutics, Mater Hospital, 41 Eccles Street, Dublin 7, Ireland

Correspondence to: D. J. O'Mahony

Materials and methods

Materials. All DNA-modifying enzymes, hybridization filters and radionucleotides were used according to the supplier's recommendations. A Geneamp polymerase chain reaction (PCR) kit was purchased from Perkin Elmer/Cetus. Synthetic oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer.

Strains and plasmids. The *E. coli* hosts AR68 and BL21(DE3) have been previously described (Shatzman and Rosenberg 1987; Rosenberg et al. 1987, respectively). The inducible expression vectors pJLA503, pET3a and pJW2 have also previously been described (Schauder et al. 1987; Rosenberg et al. 1987; Wang et al. 1990, respectively) and the relevant features of the plasmid important for inducible gene expression, including the sequence of the 5′-untranslated region or translational initiation region (TIR), are summarized (Table 1).

The vector pJW3 was constructed by replacing the gene *10* TIR region from pET3a with the corresponding *atpE* TIR region from pJLA503 (Table 1). Plasmid pUCPST3 contains a full-length

Table 1. Summary of plasmid constructs analyzed for the synthesis of recombinant porcine somatotropin (rPST)

(a)

Vector	Prom.	TIR	Sequence of translation initiation region (TIR)
pJLA503	P_RP_L	*atpE*	AACACTACTACGTTTTAACTG-AAACAAACT*GGAG*ACTCATATG
pJW2	P_RP_L	s10	ACCACAACGGTTTCCCTCTA-GAAATAATTTGTTAACTTTAA-G*AAGGAG*ATATACATATG
pJW3	$\phi10$	*atpE*	AACACTACTACGTTTTAACT-GAAACAAACT*GGAG*ACTCA-TATG
pET3a	$\phi10$	s10	ACCACAACGGTTTCCCTCTA-GAAATAATTTGTTAACTTTAA-G*AAGGAG*ATATACATATG

(b)

Plasmid construct	Vector	PST[a] gene	Poly A$^+$ tail
pJW101	pJW2	*pst**	+
pJW111	pJW2	*pst**	−
pJW301	pJW2	*pst*	+
pJW311	pJW2	*pst*	−
pJW201	pJW2	*fpst*	+
pJL101	pJLA503	*pst**	+
pJL111	pJLA503	*pst**	−
pJL301	pJLA503	*pst*	+
pJL311	pJLA503	*pst*	−
pJL201	pJLA503	*fpst*	+
pEW101	pJW3	*pst**	+
pDH101	pJLA503	*pst**/$\phi10$ *pst**	+

(a) Features of the parent vectors that were used in this study, including the identities of the promoter (Prom.), the 5′-untranslated region (TIR) and the sequence of the TIR and the *Nde*I cloning site (CATATG), which contains the initiation codon ATG, are shown. The Shine-Dalgarno sequence is *underlined*. (b) Features of the plasmids constructed for expression of rPST, including the identity of the basic expression vectors, the identity of the porcine somatotropin (*pst*) gene and the presence or absence of poly A$^+$ tail on the *pst* genes are given

[a] *fPST* refers to the full-length cDNA coding for the precursor form of PST. *pst* and *pst** refer to the wild-type (ATG.TTC.CCA.GCC.ATG.CCC.TTG.TCC.AGC.CTA.TTT.GCC.AAC.GCC.GTG) and codon-optimized (ATG.TTC.CCG.GCT.ATG.CCG.CTG.TCT.TCT.CTG.TTC.GCT.AAC.GCT.GTG) PST cDNAs, respectively, coding for mature processed PST

cDNA (*fpst* gene) coding for PST and has been previously described (O'Mahony et al. 1989).

Construction of PST expression vectors. All DNA manipulations, including plasmid constructions and PCR amplification of DNA fragments were performed according to established methodology, essentially as described by Sambrook et al. (1989). Features of the various plasmids constructed to analyse PST expression in this study are summarized in Table 1. Each *pst* plasmid construct used for expression studies was derived by subcloning the relevant *pst* gene into the *Nde*I site of pJLA503, pJW2 or pJW3.

The *pst** cDNA is a derivative of the wild-type *pst* cDNA in which the first 14 codons of the mature PST coding sequence have been altered to reflect the codon bias typical of highly expressed genes in *E. coli* (O'Mahony et al. 1990). In all cases confirmation of each construct across ligated boundaries was performed by DNA sequence analysis.

Analysis of inducible synthesis of PST. Thermo-inducible expression of the *pst*, *pst** and *fpst* genes on the relevant plasmid constructs containing the λP_RP_L promoter (Table 1) was performed in *E. coli* AR68 as host whereas isopropyl-β-D-thiogalactopyranoside (IPTG)-induced, T7 RNA-polymerase-dependent, synthesis of PST was carried out in the host *E. coli* BL21(DE3) as described previously (Wang et al. 1990; O'Mahony et al. 1990, respectively); analysis of cell lysates by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described (Laemmli 1970). Quantification of the level of PST synthesis directed by each construct was performed by densitometric analysis of SDS-polyacrylamide gels.

Results and discussion

We have analysed the synthesis of recombinant PST in *E. coli* under the control of the tandem, thermo-inducible promoters P_RP_L from phage λ (Schauder et al. 1987) in the plasmid series pJL-PST and pJW-PST. In these constructs the temperature-sensitive cI^{857} repressor is also plasmid encoded. In addition, we have analysed the synthesis of rPST in *E. coli* under the control of the bacteriophage T7 gene *10* promoter on the plasmid series pET-PST in the host BL21(DE3), in which the chromosomal-encoded T7 RNA polymerase is under the control of the *lacUV5* promoter (Rosenberg et al. 1987). In conjunction with the choice of controlling promoter used, we have also analysed and compared how variations in codon usage, 5′-untranslated sequences, and how the presence or absence of leader peptide coding sequences or poly A tail in the 3′-untranslated region influenced the levels of rPST synthesis. In total, 12 different constructs were tested for the production of rPST. The level of PST synthesis by each construct, estimated by densitometric scanning, is represented as a percentage of total cell protein throughout.

Effect of the 5′-untranslated region on PST synthesis

We have analysed the effect of the 5′-untranslated region derived from the *E. coli atpE* gene (Schauder et al. 1987) or from the phage T7 gene *10* gene (Rosenberg et al. 1987) on PST synthesis using the plasmid series derivated from pJLA503 and pJW2, respectively, in which the phage λP_RP_L promoters directed expression. The levels of rPST synthesis directed by plasmids pJL101, pJL111, pJL311 and pJL301 were low (Fig. 1, lanes 9–12, respectively; Table 2) with rPST at best accounting for up to 6% of total cell protein as directed by the plasmids pJL101 and pJL111. In contrast, PST synthesis directed by the plasmids pJW101, pJW111, pJW311 and pJW301 was relatively higher with rPST accounting for up to 30% of total cell protein in general in the case of each plasmid in this series (Fig. 1, lanes 3–6, respectively; Table 2).

We have previously reported that PST synthesis directed by the plasmid pET3a/PST, which contained the T7 gene *10* promoter and 5′-untranslated sequence, in

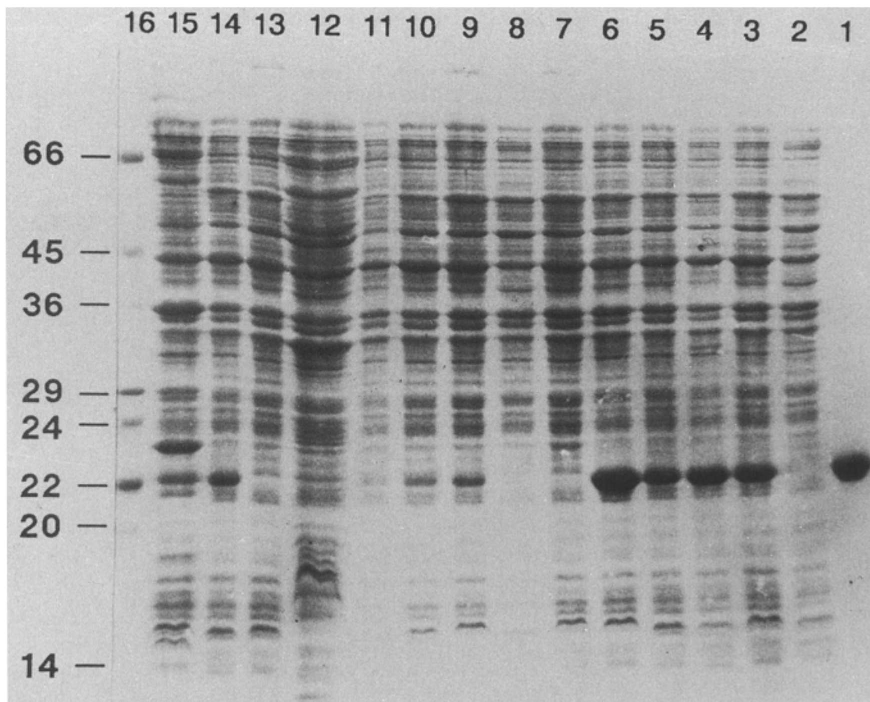

Fig. 1. Analysis of expression of recombinant porcine somatotropin (rPST) by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein products synthesized by various PST expression vectors were analysed by SDS-PAGE on a 12.5% polyacrylamide gel. All cultures were in the host AR68 except construct pEW101, which was induced in BL21(DE3). The symbols (−) and (+) refer to uninduced and induced cultures, respectively. *Lane 1,* pituitary purified PST; *lane 2,* pJW101 (−); *lane 3,* pJW101 (+); *lane 4,* pJW111 (+); *lane 5,* pJW311 (+); *lane 6,* pJW301 (+); *lane 7,* pJW201 (+); *lane 8,* pJL101 (−); *lane 9,* pJL101 (+); *lane 10,* pJL111 (+); *lane 11,* pJL311 (+); *lane 12,* pJL301 (+); *lane 13,* pJL201 (+); *lane 14,* pDH101 (+); *lane 15,* pEW101 and pLysS (+); *lane 16,* molecular mass standards: 14 kDa, 20 kDa, 22 kDa (purified pituitary) (pPST), 24 kDa, 29 kDa, 36 kDa, 45 kDa and 66 kDa

Table 2. Summary of the percentage of rPST synthesized by each plasmid construct

Plasmid construct	Expression of rPST (%)
pJW101	30
pJW111	30
pJW301	30
pJW311	30
pJW201	5
pJL101	6
pJL111	6
pJL301	1
pJL311	2
pJL201	0.5
pEW101	9
pDH101	12

Levels of rPST synthesis are given as the percentage of total cell protein as quantified by densitometric scanning of sodium dodecyl sulphate-polyacrylamide gel electrophoresis gels

the presence of plasmid pLysS, which coded for low levels of T7 lysozyme, accounted for up to 40% of total cell protein (O'Mahony et al. 1990). Replacement of the T7 gene *10* 5′-untranslated region with the 5′-untranslated region of the *E. coli atpE* gene in the plasmid pEW101 resulted in a reduction in expression directed by the T7 promoter with PST synthesis now accounting for about 9% of total cell protein (Fig. 1, lane 15; Table 2).

The double-cistron plasmid pDH101 contained tandem copies of the *pst** gene under the control of the $\lambda P_R P_L$ promoters and arranged such that the 5′ and 3′ *pst** gene contained the *atpE* and T7 gene *10* 5′-untranslated sequences, respectively. Analysis of PST synthesis directed by this plasmid showed that plasmid pDH101 directed moderate levels (12%) of rPST synthesis (Fig. 1, lane 14; Table 2) in comparison to pJL101 (6%) (Fig. 1, lane 9; Table 2), which contained a single copy of the *PST* gene.

Taken together the results suggested that replacement of the 5′-untranslated sequence of the phage T7 gene *10* with the corresponding region of the *E. coli atpE* gene resulted in reduced levels of synthesis of rPST directed by the specified promoters. Furthermore, the presence of a poly-A tail in the 3′-untranslated region had no significant effect on the synthesis of rPST (e.g., compare pJW101 and pJW111; pJW301 and pJW311 in Table 2). We also observed that the level of *PST* mRNA is higher from the "pJW" constructs as opposed to the "pJL" constructs and that the mRNA level was higher after 2 h induction as opposed to 4 h induction (data not shown).

There is now a very large body of evidence showing that the 5′-untranslated region or TIR sequence affects the rate of gene expression (reviewed by Gold 1988). Such features of the 5′-untranslated sequence that are believed to contribute to the efficiency of translation initiation and gene expression include the nucleotide sequence of the Shine-Dalgarno (SD) (Morelle and Mayer 1988), the distance and sequence between the SD and initiation codon (Dalbøge et al. 1988; Vize and Wells 1987; de Boer et al. 1983) and the nucleotide sequence 5′ to the SD sequence (Schauder and McCarthy 1989; Olins and Rangwala 1989). The T7 ϕ10 TIR is consistently better than the *atpE* TIR in permitting efficient synthesis of rPST in *E. coli* in this study. The SD sequence of the pJW series, 5′AGAAGGAG, is positioned eight nucleotides from the initiation codon whereas the SD sequence of the pJL series, 5′GGAG, is positioned six nu-

cleotides from the initiation codon (see Table 1). However, an increase in the nucleotides between the SD sequence and ATG by two bases in pJL101 or pJL301 did not lead to any detectable increase in the synthesis of rPST (data not shown). The ability of the mRNA sequence in the vicinity of the translation start site to adopt secondary and/or tertiary structures can also influence the level of expression of the gene in question (Stanssens et al. 1985; Tessier et al. 1984; Lee et al. 1987; Schauder and McCarthy 1989).

Effects of codon usage on PST expression

Following the observation that the usage of synonymous codons is biased in various organisms (Grantham et al. 1980), it was suggested that high frequency codons, which corresponded to high abundancy tRNAs, were important in optimizing translation of mRNAs from highly expressed genes (Ikemura 1981). This led to the controversial idea that low frequency codons would limit the rate of translation (Hoekema et al. 1987; Andersson and Kurland 1990). The weight of evidence now shows that high-frequency codons have been selected in microorganisms because they lead to greater efficiency in the overall process of translation. Low-frequency codons do not usually limit the rate of translation of a particular gene; highly expressed genes contain many rare codons. Our own data are in partial agreement with this conclusion.

Using the criteria of Sharp and Li (1986), we have optimized the *pst* coding sequence such that the 5′ most 14 codons of the *pst** gene present on the plasmids pJW101, pJW111, pJL101 and pJL111 reflect the codon usage typical of highly expressed *E. coli* genes (see Table 1). These plasmids are otherwise identical to plasmids pJW301, pJW311, pJL301 and pJL311, respectively, which contain the unaltered codons of *pst*. No differences were observed in the levels of rPST synthesis directed by the plasmids pJW111 and pJW311 whereas pJW301 directed marginally higher level synthesis of PST than pJW101 (Fig. 1, lanes 4, 5; 3, 16, respectively; Table 2) indicating that optimization of codon usage had no significant effects on the efficiency of rPST synthesis directed by the pJW series in *E. coli*.

In contrast, optimization of codon usage does appear to have an effect on the synthesis of rPST by the pJL series with higher levels of rPST synthesized by the *pst** gene on pJL101 and pJL111 (Fig. 1, lanes 9 and 10, respectively; Table 2) in comparison to the unaltered *pst* gene on pJL311 and pJL301 (Fig. 1, lanes 11 and 12, respectively; Table 2).

Whereas the presence of the wild-type codons do not hinder high-level expression as in the case of plasmids pJW311 and pJW301, which direct the synthesis of rPST at levels of 30% of total cell protein, it seems more likely that the comparatively lower level synthesis of rPST by pJL311 and pJL301 results from the combination of the *atpE* TIR and the wild-type codons in the first 14 positions. One possible explanation is that this combination at the 5′-end of the *pst* mRNA allows it to adopt a secondary and/or tertiary structure that inhibits translation initiation. Alternatively, the effect of codon usage on gene expression may only be critical depending on the efficiency of translation initiation promoted by a particular ribosome binding site.

Effects of leader-peptide coding sequences on PST synthesis

In pJW201 and pJL201 we have cloned the full-length cDNA for PST, including its own leader-peptide coding sequence (fPST), under the control of the T7 gene *10* TIR and *atpE* TIR, respectively. We could not detect any intracellular rPST coded by pJL201 either by SDS-PAGE analysis (Fig. 1, lane 13) or Western blotting (data not shown). In contrast, low level expression of rPST directed by pJW201 was confirmed by Western blot and showed the presence of a 23-kDa immunoreactive peptide band that was intermediate in size between that of the mature PST (22 kDa) and full-length PST (24 kDa) (data not shown). It is possible that *E. coli* cleaves fPST within the leader peptide and not at the junction between the leader peptide and mature PST. By subcellular fractionation we have localized the processed PST synthesized by pJW201 to the plasma membrane. The level of rPST synthesized by pJW201 is much less than that synthesized by the other pJW plasmids in *E. coli* AR68 (Table 2). In further studies, we have established that the presence of leader-peptide coding sequences from either the ompA gene of *E. coli* (Ghrayeb et al. 1984), the α-amylase gene from *B. licheniformis* (Wood et al. 1990) or from the *fPST* gene dramatically reduced the synthesis of rPST from the pET3a plasmid (unpublished data).

In summary, synthesis of rPST in a precursor form substantially reduced the overall level of PST synthesis. However, the choice of leader-peptide coding sequence together with the choice of the 5′ TIR is an important determinant of the overall level of expression achieved and in the final processing of the precursor PST. Vernet et al. (1989) have previously shown that the putative signal sequence of the protease papain impairs the synthesis of this protein and β-galactosidase fusion proteins in *E. coli* and have suggested that the synthesis of papain in *E. coli* may require deletion of the signal sequence. Our results would suggest that efficient synthesis of rPST in *E. coli* also necessitates removal of leader-peptide coding sequences.

From these studies we conclude that:

1. The choice of the 5′-untranslated region, including the SD sequence, has a critical effect on the synthesis of rPST in *E. coli*. The phage T7 ϕ10 TIR allows for higher levels of synthesis of rPST protein relative to the *E. coli atpE* TIR.
2. When rPST is synthesized at a high level (30% of total cell protein) from a strong ribosome binding site (T7 ϕ10) then codon usage at the first 14 codons of the *pst* gene does not have any significant effect on the level of synthesis of rPST. However, when rPST is synthesized at a low level (6% of total cell protein) from a weak ri-

bosome binding site (*atpE*) then optimization of the 14 5′-codons of the *pst** gene does influence the level of synthesis of rPST.

3. The presence of leader-peptide coding sequences at the 5′-end of the *PST* gene reduces the synthesis of rPST in *E. coli*.

Acknowledgements. We are especially grateful to Dr. Therese Kinsella for many helpful comments, suggestions, discussions and critical readings of this manuscript. This work was supported by the Board of Trinity College, Dublin, by the National Board for Science and Technology Ireland and by CITIC, China.

References

Andersson SGE, Kurland CG (1990) Codon preferences in free-living microorganisms. Microbiol Rev 54:198–210

Boer HA de, Hui A, Comstock LJ, Wong E, Vasser M (1983) Portable Shine-Dalgarno regions: a system for a systematic study of defined alterations of nucleotide sequences within *E. coli* ribosome binding sites. DNA 2:231–235

Dalbøge H, Carlsen S, Jensen EB, Christensen T, Dahl HM (1988) Expression of recombinant growth hormone in *Escherichia coli*: effect of the region between the Shine-Dalgarno sequence and ATG initiation codon. DNA 7:399–405

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

Gold L (1988) Post-transcriptional regulatory mechanisms in *E. coli*. Annu Rev Biochem 57:199–233

Grantham R, Grutier C, Gouy M (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome typing. Nucleic Acids Res 8:1893–1912

Hoekema A, Kastelein RA, Vasser M, Boer HA de (1987) Codon replacement in the *PGKI* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol Cell Biol 7:2914–2924

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol 151:389–409

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lee N, Zhang S-2, Cozzitorto J, Yang J-S, Testa D (1987) Modification of mRNA secondary structure and alteration of the expression of human interferon α1 in *Escherichia coli*. Gene 58:77–86

Morelle G, Mayer H (1988) Increased synthesis of human parathyroid hormone in *Escherichia coli* through alterations of the 5′ untranslated region. Biochem Biophys Acta 950:459–462

Olins PO, Rangwala SH (1989) A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the *lacZ* gene in *Escherichia coli*. J Biol Chem 264:16973–16976

O'Mahony DJ, Wang H, McConnell DJ, Jia F, Xia L, Qi S (1989) Polymorphism in porcine somatotropin cDNA sequences. Anim Genet 20:313–316

O'Mahony DJ, Wang H, McConnell DJ, Jia F, Zhou SW, Yin D, Qi S (1990) The effect of phage T7 lysozyme on the production of biologically active porcine somatotropin in *Escherichia coli* from a gene transcribed by T7 RNA polymerase. Gene 91:275–279

Rosenberg AH, Lade BN, Chui D, Lin S, Dunn JJ, Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Schauder B, McCarthy JEG (1989) The role of bases upstream of the Shine-Dalgarno region and in the coding sequences in the control of gene expression in *Escherichia coli*: translational and stability of mRNAs *in vivo*. Gene 78:59–72

Schauder B, Blöcker H, Frank R, McCarthy JEG (1987) Inducible expression of vectors incorporating the *Escherichia coli atpE* translational initiation region. Gene 52:279–283

Sharp PM, Li W-M (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res 14:7737–7749

Shatzman A, Rosenberg M (1987) Expression, identification and characterization of recombinant gene products in *Escherichia coli*. Methods Enzymol 152:661–673

Stanssens P, Remaut E, Friers W (1985) Alterations upstream from the Shine-Dalgarno region and their effect on bacterial gene expression. Gene 36:211–223

Tessier LM, Sondermeyer P, Faure T, Dreyer D, Benevente A, Villerual D, Courtney M, Lecorq J-P (1984) The influence of mRNA primary and secondary structure on human IFN-γ gene expression in *E. coli*. Nucleic Acids Res 12:7663–7675

Vernet T, Tessier DC, Laliberté F, Dignard D, Thomas DY (1989) The expression in *Escherichia coli* of a synthetic gene coding for the precursor of papain is prevented by its own putative signal sequence. Gene 77:229–236

Vize PD, Wells JRE (1987) Spacer alterations which increase the expression of porcine growth hormone in *E. coli*. FEBS Lett 213:155–158

Wang H, McConnell DJ, O'Mahony DJ (1990) An efficient temperature-inducible vector incorporating the T7 gene *10* translational initiation leader region. Nucleic Acids Res 18:1070

Wood HE, Dawson MT, Devine KM, McConnell DJ (1990) Characterization of PBSX, a defective prophage of *Bacillus subtilis*. J Bacteriol 172:2667–2674