

GENE 1524

## Enhancement of translational efficiency by the *Escherichia coli atpE* translational initiation region: its fusion with two human genes

(Recombinant DNA; expression vectors; gene cloning;  $\beta$ -interferon; interleukin 2; Shine–Dalgarno sequence; ribosome-binding site; phage  $\lambda$  promoters)

John E.G. McCarthy<sup>a,\*</sup>, Walter Sebald<sup>a</sup>, Gerhard Gross<sup>b</sup> and Reiner Lammers<sup>b</sup>

Departments of <sup>a</sup> Cytogenetics and <sup>b</sup> Genetics. GBF – Gesellschaft für Biotechnologische Forschung mbH., Mascheroder Weg 1, D-3300 Braunschweig (F.R.G.) Tel. 0531-6181-0

(Received October 1st, 1985)

(Revision received November 11th, 1985)

(Accepted November 19th, 1985)

---

### SUMMARY

The cDNA sequences encoding mature human interleukin 2 (IL2) and  $\beta$ -interferon (INF $\beta$ ), respectively, were fused with various translational initiation regions and inserted into two different types of expression vector. The relative levels of expression of the two genes and the functional stability of their respective mRNAs were examined in vivo in *Escherichia coli* hosts. The addition of the 30-bp sequence, found immediately upstream of the *E. coli atpE* gene Shine–Dalgarno (SD) sequence, to the translational initiation regions of IL2 and INF $\beta$  increased the expression of both these genes by a factor of 6–10. Thus this sequence, which naturally acts within the *E. coli atp* operon to enhance the translational initiation frequency of the *atpE* gene, can increase the expression of other genes in *E. coli*. It may exemplify a specific type of recognition signal for the *E. coli* translational apparatus.

---

### INTRODUCTION

Studies on the mechanisms determining the translational efficiency of genes expressed in *E. coli* have demonstrated that alterations ranging from single base substitutions to deletions or additions of many bases in the translational initiation region can signifi-

cantly influence the rate of protein synthesis (see, e.g., Roberts et al., 1979; Scherer et al., 1980; Gold et al., 1981; Hall et al., 1982; Jay et al., 1982; Kastelein et al., 1983; Wood et al., 1984; Hui et al., 1984; Schottel et al., 1984; Coleman et al., 1985; McCarthy et al., 1985). Indeed, there is evidence that translational initiation is an important control point in the overall process of gene expression in *E. coli* (Kastelein et al., 1983; McCarthy et al., 1985). However, a coherent picture of the respective roles played by primary and secondary structure in determining the efficiency of translational initiation is still lacking.

Previous work indicated that a sequence found upstream of the SD of the *atpE* gene is important for enhancing the efficiency of translational initiation

---

\* To whom correspondence and reprint requests should be addressed.

Abbreviations: bp, base pair(s); INF $\beta$ ,  $\beta$ -interferon; IL2, interleukin 2; nt, nucleotide(s);  $p_L$  and  $p_R$ , major leftward and rightward bacteriophage  $\lambda$  promoters, respectively; PA, polyacrylamide; RBS, ribosome-binding site (translational initiation region); SD, Shine–Dalgarno sequence (see INTRODUCTION); SDS, sodium dodecyl sulfate.

(McCarthy et al., 1985). Similar sequences are found in the translational initiation regions of a number of other highly expressed genes. Further investigation of this phenomenon is expected to shed light on the process of translational initiation and the manner in which its efficiency is controlled. In the present study it is shown that the intercistronic sequence found upstream of the *atpE* gene is capable of greatly improving translational efficiency when inserted upstream of other genes, even those foreign to *E. coli*. It is thereby demonstrated that the effect upon translational efficiency is a general, rather than a gene-specific, effect.

In this paper SD is taken to mean a sequence of up to 9 nt complementary to the 3' end of the *E. coli* 16S rRNA (Shine and Dalgarno, 1974), which constitutes therefore a part of the total translational initiation region (RBS).

## MATERIALS AND METHODS

### (a) Bacterial strains and expression vectors

*E. coli* strains DH1 (Hanahan, 1983) and MCG1 (McCarthy et al., 1985) were used for transformation and expression studies.

The *tac* fusion promoter vector pDR540 (Russell and Bennett, 1982) was used for cloning and expression studies (see McCarthy et al., 1985). The double bacteriophage  $\lambda$  promoter ( $p_R p_L$ ) vector pJLf201 was constructed by inserting the bacteriophage  $\lambda$  *cIts857* gene and  $p_R$  promoter upstream from the  $p_L$  promoter of pJLf101 (McCarthy et al., 1985). The unique *SalI* site between the  $p_L$  promoter and the *fd* bacteriophage terminator was used for cloning.

The cDNA gene sequences encoding mature human INF $\beta$  and IL2, respectively, were fused with various RBS regions to give the constructs shown in Fig. 1. Synthetic SD adapters were used for the construction of  $\beta$ SD1 (see Gross et al., 1985) and ILSD2. A 70-bp *Bam*HI-*Nla*III fragment of the *atp* operon derived from pH116(c) (McCarthy et al., 1985) containing the complete *atpE* RBS was used in the fusions  $\beta$ *atp*1 and *ILatp*2. ILSD3 and *ILatp*3 were constructed using a synthetic oligodeoxynucleotide bearing a derivative of the *atpE* RBS, where a *Bam*HI site had been inserted immediately 5' of the SD (see Fig. 1; J.E.G.McC., H. Blöcker and R. Frank, in preparation). ILSD3 was derived from

*ILatp*3 by removing the 38 nt immediately upstream from the SD site which lies between the indicated *Xho*I and *Bam*HI sites. *Bam*HI and *Sal*I linkers were used for the cloning of the cDNA sequences together with the RBS regions into pDR540 and pJLf201. The sequences were all checked by re-cloning into M13 vectors and sequencing by the dideoxynucleotide procedure (Sanger et al., 1977).

### (b) In vivo pulse labelling

Pulse labelling with [<sup>35</sup>S]L-methionine was performed 20 min after IPTG (at 37°C) or thermal (28 → 42°C) induction (see McCarthy et al., 1985). The same procedure was followed using <sup>3</sup>H-labelled amino acids (Amersham; L-leucine, L-lysine, L-phenylalanine, L-proline, L-tyrosine), except that the labelling pulse was increased to 2 min, after which time the corresponding unlabelled amino acids were added to give a final concentration of each amino acid of 1 mM.

The decay in the rate of incorporation of [<sup>35</sup>S]L-methionine into cells after the addition of 100  $\mu$ g rifampicin per ml was followed by taking 0.8-ml samples of the cultures and pulse labelling for 40 s with a 40-s chase.

Quantitation of radioactive incorporation into specific bands was performed by excising them from dried gels and scintillation counting.

### (c) mRNA local secondary structure

Predictions of mRNA secondary structure were made with the help of a programme kindly supplied by Papanicolaou et al. (1984).

## RESULTS

### (a) Plasmid constructs for the expression of the INF $\beta$ and IL2 cDNA sequences

The cDNA sequences encoding mature INF $\beta$  and mature IL2 (Taniguchi et al., 1980; 1983) were combined with different RBS regions (Fig. 1) and cloned into the expression vectors pDR540 (Russell and Bennett, 1982) and pJLf201 (described in this paper). Cloning into the unique *Bam*HI site of pDR540 allowed transcription from the *tac* promoter to run through the cloned sequences and then through the *galK* gene, which lies downstream from the cloning

```

βSD1                               GUCGACGGAUCCCUAGGGAGAUUCUUCUUCCAUGAGC
βatp1                               UAAUUUACCAACACUACUACGUUUUAACUGAAACAAACUGGAGACUGUUUUGAGC
ILSD2                               GUCGACGGAUCCAGCACAAUAGUAAGGAUUAUAGCCAUGGCA
ILatp2                              UAAUUUACCAACACUACUACGUUUUAACUGAAACAAACUGGAGACUGUCAUGGCA
ILSD3                               GUCGACGGAUCCUGGAGACUGCCAUGGCA
ILatp3                              CUCGAGUAAUUUACCAACACUACUACGUUUUAACUGAAACAAAGGAUCCUGGAGACUGCCAUGGCA

```

Fig. 1. RBS regions of the constructs. The sequences shown start with the linkers (or in the cases of ILSD3 and ILatp3 the sites at the beginning of the oligodeoxynucleotide) used for cloning into the expression vectors followed by the translational initiation regions of the six constructs up to and including the first and second codons of the respective structural genes. The SD's, structural gene start codons and, in the cases of *βatp1*, *ILatp2* and *ILatp3*, the stop codon which normally belongs to the *atpB* gene (see McCarthy et al., 1985), are underlined.

site. Sequences cloned using the unique *SalI* site of pJLf201 were transcribed from the  $p_{RP_L}$  promoters.

Analyses were made of the possible local mRNA secondary structures which could theoretically be formed in the region -50 to +50 nt relative to the start codons of the cloned genes. No relationship between the potential secondary structures and the observed differences in potency of the RBS regions could be discerned.

#### (b) The *atpE* sequence enhances expression in vivo

The potency of the different RBS regions was examined using strains of *E. coli* (DH1 or MCG1) containing the described constructs. To measure the relative levels of expression of the cloned genes in vivo, whole cells were solubilized after induction of the expression vectors, and samples thereof were loaded onto SDS-PA gels. The relative rates of specific protein synthesis in vivo during short time spans after induction were estimated on the basis of pulse labelling (Fig. 2; Table I). The relative amounts of specific proteins synthesized over a longer time period were estimated by means of Coomassie staining (Table I). Moreover, IFN $\beta$  activity in extracts was also measured to give a further, independent indication of the relative amounts of this protein expressed under the direction of those constructs bearing the IFN $\beta$  gene (see also Table I).

The presence of the *atpE* sequence pattern upstream from the SD improves the expression of IFN $\beta$  and IL2 by a factor of at least six relative to constructs bearing (three different) SD's alone. The construct  $\beta$ SD1 is very similar to that described by Gross et al. (1985) except that two extra bases (UC) have been inserted between the SD and the start

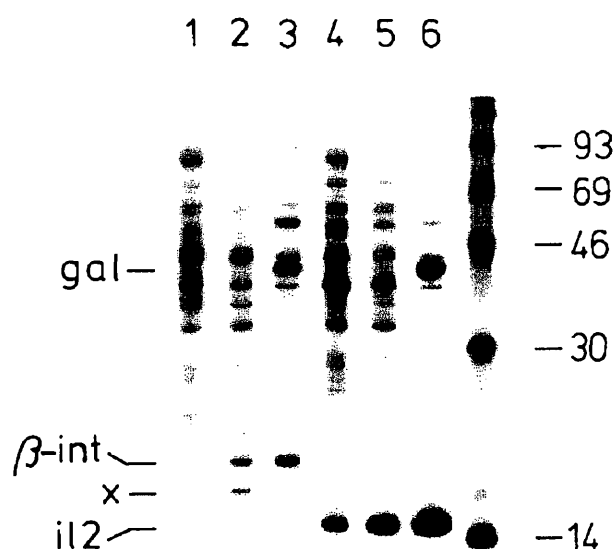


Fig. 2. In vivo pulse labelling of strains containing the described recombinant plasmids. A fluorograph of a SDS-15% PA gel loaded with the sonicated and solubilized cells after pulse-labelling with  $^3\text{H}$ -labelled amino acids. Experiments were performed with *E. coli* strain DH1, transformed with derivatives of pJLf201 bearing the constructs  $\beta$ SD1 (lane 1), *βatp1* (lane 2), ILSD2 (lane 4), *ILatp2* (lane 5) and MCG1, transformed with derivatives of pDR540 bearing *βatp1* (lane 3) and *ILatp2* (lane 6). The same amount of cellular material was loaded in each lane. The symbols on the left margin indicate the positions of galactokinase (gal), IFN $\beta$  ( $\beta$ -int), IL2 (il2) and also of a breakdown (or premature termination) product of IFN $\beta$  (x) which appeared in significant amounts only during the early stages of heat induction. The sizes (kDa) of protein markers are indicated on the right margin.

TABLE 1

Efficiencies of in vivo  $\text{INF}\beta$  and IL2 synthesis obtained with various RBS constructs

Construct <sup>a</sup>	Pulse-chase experiments <sup>b</sup>		Coomassie staining <sup>c</sup>		$\text{INF}\beta$ activity (I.U.) <sup>d</sup>	
	pJLf201 (%)	pDR540 (%)	pJLf201 (%)	pDR540 (%)	pJLf201	pDR540
$\beta\text{SD1}$	1.5	0.5	1.0	n.m.	$10^6$ – $10^7$	n.m.
$\beta\text{atp1}$	9.0 <sup>e</sup>	6.0	12	1.0	$2 \times 10^9$ – $10^{10}$	$10^6$
ILSD2	5.6	6.1	4.5	4.9		
ILatp2	36	36	32	35		
ILSD3	3.0	n.d.	3.3	n.d.		
ILatp3	30	n.d.	32	n.d.		

<sup>a</sup> See Fig. 1, and MATERIALS AND METHODS, sections a and b.<sup>b</sup> Incorporation of [<sup>35</sup>S]L-methionine presented as a percentage of the total incorporation into cellular protein during the pulse. These data indicate the relative rates of specific protein synthesis during the designated time period. Data are presented for each gene/translational initiation region combination inserted into pJLf201 and pDR540.<sup>c</sup> The amount of Coomassie-stained specific protein band on SDS-PA gels as a percentage of total Coomassie-stained cellular protein (estimated using an LKB 2202 Ultrosan laser densitometer). The gels were loaded with solubilized extracts from cells which had been induced for 4 h.<sup>d</sup> Expressed as international units of  $\text{INF}\beta$  activity with an NIH standard as reference.<sup>e</sup> The sum of the  $\text{INF}\beta$  band and the 16-kDa breakdown (premature termination) product (see Fig. 2) was equivalent to  $16\%$ , n.d., not determined; n.m., not measurable.

codon of  $\text{INF}\beta$ . The two constructs directed the same level of expression of  $\text{INF}\beta$ . A ratio of about 10 is also apparent when comparing the expression levels of ILatp3 and ILSD3, where both bear the same (*atpE*) SD and following sequence 3' to it. This latter result is directly analogous to the data reported previously for the *atp* operon (McCarthy et al., 1985).

Induction of the expression of IL2 and  $\text{INF}\beta$  directed by the constructs  $\beta\text{atp1}$ , ILatp2 and ILatp3 (in contrast to induction of  $\beta\text{SD1}$ , ILSD2 and ILSD3) was accompanied by a marked inhibition of bacterial growth, particularly in the case of strains producing  $\text{INF}\beta$ . This observation, taken together with the competitive inhibitory effect of expression of the cloned foreign genes upon the expression of other cellular genes (Fig. 2), and also the level of expression attained (Devos et al., 1983; Remaut et al., 1983), indicates that IL2 and  $\text{INF}\beta$  were both expressed at very near the physiological limits of over-production for these two genes in *E. coli*.

The influence of the *atpE*-type RBS region upon expression was similar using both types of expression vector. The nature of the vector did, however, influence to some extent the characteristics of expression. For example, the maximal expression of

$\text{INF}\beta$  was appreciably higher with pJLf201 than with pDR540, whereas the expression of IL2 was generally at least as high with pDR540 as with pJLf201. Another difference is that a high percentage of the expressed  $\text{INF}\beta$  appeared as an approx. 16-kDa breakdown (or premature termination) product shortly after induction when pJLf201 was the vector rather than pDR540 (Fig. 2). This effect may be associated with the initial shock of heat induction, since the total synthesis of  $\text{INF}\beta$  over a longer time period was greater with pJLf201 than with pDR540. Expression of the *galK* gene serves as a useful marker for transcriptional activity in pDR540. The fact that the expression of this gene is less in those strains bearing  $\beta\text{atp1}$ , ILatp2 and ILatp3 constructs indicates that the observed effects upon expression cannot be due to any significant extent to variations in transcriptional efficiency (see McCarthy et al., 1985).

### (c) The functional half-lives of mRNA transcripts of the constructed plasmids.

Comparative estimates of the functional mRNA half-lives (Fig. 3) indicate that the alterations in the translational initiation regions had little effect upon the stability of the mRNA transcripts encoding IL2

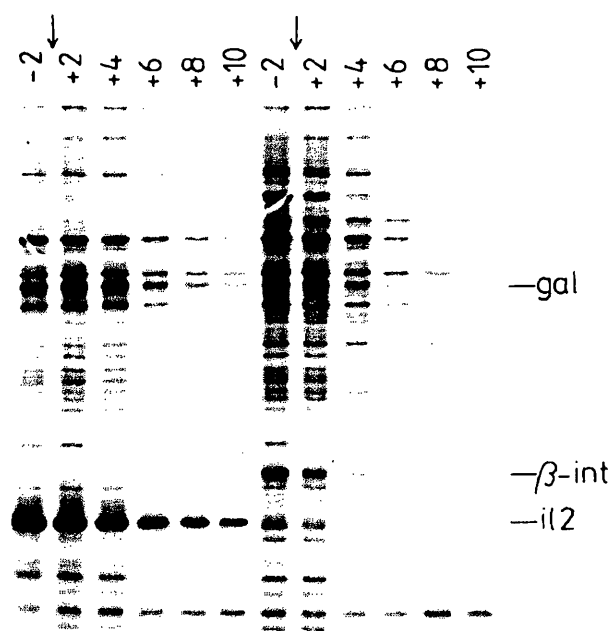


Fig. 3. Functional half-lives of IFN $\beta$  and IL2 transcripts in vivo. Cells were pulse-labelled with [ $^{35}$ S]-methionine before and after the addition of rifampicin. The *E. coli* strain MCG1 had been transformed with derivatives of pDR540 bearing either ILatp2 (the six lanes on the left hand side) or  $\beta$ atp1. The numbers above the lanes refer to the times (in min) of sampling relative to the addition of rifampicin (time 0). For symbols on the right margin see legend to Fig. 2.

and IFN $\beta$ . For example, the estimated functional half-life for the construct ILatp2 in pDR540 was 2.8 min, whereas that of ILSD2 in the same vector was 2.4 min.

#### DISCUSSION

The RBS upstream of the *E. coli atpE* gene promotes highly efficient translational initiation of the human IL2 and IFN $\beta$  genes. The segment of 30 bp or less upstream from the *atpE* SD plays a primary role in determining this high level of translation, just as was shown previously for the *atpE* gene itself (McCarthy et al., 1985). The maximal levels of synthesis of the two human genes in *E. coli* obtained in the present work using the *atpE* sequence are at least as high as those previously reported by others (Devos et al., 1983; Remaut et al., 1983).

The present data have several implications regarding the regulation of translational efficiency in *E. coli*. They demonstrate the function of a naturally

occurring sequence pattern (exemplified by the *atpE* RBS) as an element (additional to the SD) which can act generally to enhance the translational initiation efficiency of genes. This pattern comprises a U-rich sequence followed by an interrupted A-rich sequence; the *atpE* form of it is UUUUAACUGAAACAAA. It is likely to lie close to, or within, the region of mRNA bound by a ribosome as it initiates translation. Additional nt may also belong to the complete pattern. The enhancing effect on translational yield is apparently not due to changes in mRNA stability or the rate of transcription. The *atpE*-type pattern has been observed in the RBS regions of other genes that are highly expressed in *E. coli* (e.g., ribosomal and bacteriophage  $\lambda$  genes; McCarthy et al., 1985). Moreover, at least the U-rich part of the sequence is evident in the RBS regions of an even wider range of genes (especially among bacteriophages; see, e.g., Jay et al., 1982). Further relevant experimental evidence can be found in the study of Mott et al. (1985). They showed that a much higher efficiency of translation of the *E. coli rho* gene was attained when the  $\lambda$  cII RBS region, which closely resembles the *atpE* pattern, replaced the normal *rho* one.

There is no obvious relationship between the levels of expression associated with the described constructs and expected (or predictable) local mRNA secondary structure in the RBS region. Therefore, the observed effects of the *atpE* RBS are likely to be due to its primary structure acting as a recognition signal for parts of the translation apparatus rather than to the influence of this region upon local secondary structure.

The present study underlines the importance of reaction steps involved in initiation in determining the overall efficiency of translation. The two human genes, especially IFN $\beta$ , have codon usages typical of many genes that are normally weakly expressed in *E. coli* (e.g., Grosjean and Fiers, 1982). Yet these genes can be highly expressed in *E. coli* provided that they are fused with efficient RBS regions. Thus codon usage cannot play a primary role in determining the efficiency of translation of the two genes, at least up to the maximal levels of expression described here. The factors which might determine the level of expression of IFN $\beta$  in *E. coli* will be discussed elsewhere (G.G. and J.E.G. McC., in preparation).

In conclusion, both experimental data and sequence comparisons (McCarthy et al., 1985) prompt formulation of the hypothesis that certain sequence patterns (such as the *atpE* type) function generally to enhance the efficiency of translational initiation. They occur as a component of RBS regions which promote a high rate of protein synthesis. It remains to be established to what extent the mode of presentation (in relation to primary and secondary mRNA structures) as well as the precise structure of the *atpE*-type pattern itself determines the strength of the enhancer sequence's influence upon translational initiation efficiency.

#### ACKNOWLEDGEMENTS

We are grateful to Prof. K. von Meyenburg and Dr. B.B. Jørgensen for the donation of a plasmid bearing the bacteriophage  $\lambda$  *cIts857* gene together with the  $p_R$  promoter.

#### REFERENCES

- Coleman, J., Inouye, M. and Nakamura, K.: Mutations upstream of the ribosome binding site affect translational efficiency. *J. Mol. Biol.* 181 (1985) 139–143.
- Devos, R., Plaetinck, G., Cheroutre, H., Simons, G., Degraeve, W., Tavernier, J., Remaut, E. and Fiers, W.: Molecular cloning of human interleukin 2 cDNA and its expression in *E. coli*. *Nucl. Acids Res.* 11 (1983) 4307–4323.
- Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G.: Translational initiation in prokaryotes. *Annu. Rev. Microbiol.* 35 (1981) 365–403.
- Grosjean, H. and Fiers, W.: Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 18 (1982) 199–209.
- Gross, G., Mayr, U., Stüber, D., Westphal, W., Zacher, M., Frank, R. and Blöcker, H.: Stabilisation of human interferon  $\beta$  synthesis in *Escherichia coli* by zinc ions. *Biochim. Biophys. Acta* 825 (1985) 207–213.
- Hall, M.N., Gabay, J., Débarbouillé, M. and Schwartz, M.: A role for mRNA secondary structure in the control of translation initiation. *Nature* 295 (1982) 616–618.
- Hanahan, D.: Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166 (1983) 557–580.
- Hui, A., Hayflick, J., Dinkelspiel, K. and De Boer, H.A.: Mutagenesis of the three bases preceding the start codon of the  $\beta$ -galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO J.* 3 (1984) 623–629.
- Jay, E., Seth, A.K., Rommens, J., Sood, A. and Jay, G.: Gene expression: chemical synthesis of *E. coli* ribosome binding sites and their use in directing the expression of mammalian proteins in bacteria. *Nucl. Acids Res.* 10 (1982) 6319–6329.
- Kastelein, R.A., Berkhout, B., Overbeek, G.P. and Van Duin, J.: Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein. *Gene* 23 (1983) 245–254.
- McCarthy, J.E.G., Schairer, H.U. and Sebald, W.: Translational initiation frequency of *atp* genes from *Escherichia coli*: identification of an intercistronic sequence that enhances translation. *EMBO J.* 4 (1985) 519–526.
- Mott, J.E., Grant, R.A., Ho, J. and Platt, J.: Maximizing gene expression from plasmid vectors containing the  $\lambda$   $p_L$  promoter: strategies for overproducing transcription termination factor p. *Proc. Natl. Acad. Sci. USA* 82 (1985) 88–92.
- Papanicolaou, C., Gouy, M. and Ninio, J.: An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules. *Nucl. Acids Res.* 12 (1984) 31–44.
- Remaut, E., Stanssens, P. and Fiers, W.: Inducible high level synthesis of mature human fibroblast interferon in *Escherichia coli*. *Nucl. Acids Res.* 11 (1983) 4677–4688.
- Roberts, T.M., Kacich, R. and Ptashne, M.: A general method for maximizing the expression of a cloned gene. *Proc. Natl. Acad. Sci. USA* 76 (1979) 760–764.
- Russell, D.R. and Bennett, G.N.: Construction and analysis of in vivo activity of *E. coli* promoter hybrids and promoter mutants that alter the -35 to -10 spacing. *Gene* 20 (1982) 231–243.
- Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 (1977) 5463–5467.
- Scherer, G.F.E., Waskinshaw, M.D., Arnott, S. and Morrè, D.J.: The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both the leader and protein coding segments. *Nucl. Acids Res.* (1980) 3895–3907.
- Schottel, J.L., Sninsky, J.J. and Cohen, S.N.: Effects of alterations in the translation control region on bacterial gene expression: use of *cat* gene constructs transcribed from the *lac* promoter as a model system. *Gene* 28 (1984) 177–193.
- Shine, J. and Dalgarno, L.: The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* 71 (1974) 1342–1346.
- Taniguchi, T., Guarente, L., Roberts, T., Kimelman, D., Douhan, J. and Ptashne, M.: Expression of the human fibroblast interferon gene in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 77 (1980) 5230–5233.
- Taniguchi, T., Matsui, H., Fujita, T., Takoaka, C., Kashima, N., Yoshimoto, R. and Hamuro, J.: Structure and expression of a cloned cDNA for human interleukin-2. *Nature* 302 (1983) 305–310.
- Wood, C.R., Boss, M.A., Patel, T.P. and Spencer Emtage, J.: The influence of messenger RNA secondary structure on expression of an immunoglobulin heavy chain in *Escherichia coli*. *Nucl. Acids Res.* 12 (1984) 3937–3950.

Communicated by T. Bickle.