Microbial Biotechnology (2009) 2(3), 379–389 doi:10.1111/j.1751-7915.2009.00107.x

The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA

Laila Berg,[1] Rahmi Lale,[1] Ingrid Bakke,[1] Nigel Burroughs[2] and Svein Valla[1]*
[1]*Department of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.*
[2]*Mathematics Institute and Warwick Systems Biology Centre, University of Warwick, Coventry. CV4 7AL, UK.*

Summary

Secondary structures and the short Shine–Dalgarno sequence in the 5′-untranslated region of bacterial mRNAs (UTR) are known to affect gene expression at the level of translation. Here we report the use of random combinatorial DNA sequence libraries to study UTR function, applying the strong, σ^{32}/σ^{38}-dependent, and positively regulated *Pm* promoter as a model. All mutations in the libraries are located at least 8 bp downstream of the transcriptional start site. The libraries were screened using the ampicillin-resistance gene (*bla*) as reporter, allowing easy identification of UTR mutants that display high levels of expression (up to 20-fold increase relative to the wild-type at the protein level). Studies of the two UTR mutants identified by a modified screening procedure showed that their expression is stimulated to a similar extent at both the transcript and protein product levels. For one such mutant a model analysis of the transcription kinetics showed significant evidence of a difference in the transcription rate (about 18-fold higher than the wild type), while there was no evidence of a difference in transcript stability. The two UTR sequences also stimulated expression from a constitutive σ^{70}-dependent promoter (*P1/$P_{anti\text{-}tet}$*), demonstrating that the UTR at the DNA or RNA level has a hitherto unrecognized role in transcription.

Introduction

The 5′-untranslated region of bacterial mRNAs (UTR) is known to be of critical importance for mRNA translation, and efficient translation is very important for obtaining high levels of recombinant gene expression in both many basic research projects and in industrial protein production. The feature in the UTR that is believed to be most important is the purine rich Shine–Dalgarno (SD) sequence, which by X-ray crystallographic studies of the mRNA bound to the ribosome has been confirmed to interact with the 3′ end of the 16S rRNA (anti-SD) in the 30S ribosomal subunit (Shine and Dalgarno, 1974; Steitz and Jakes, 1975; McCarthy and Brimacombe, 1994; Makrides, 1996; Yusupova *et al*., 2001; Laursen *et al*., 2005). In addition, the ribosomal protein S1 (Roberts and Raboniwitz, 1989; Farwell *et al*., 1992; Subramanian, 1992; Tzareva *et al*., 1994), essential for *in vivo* translation in *Escherichia coli* (Sørensen *et al*., 1998), has been shown to bind to the 11 nucleotides directly upstream of the SD sequence (Sengupta *et al*., 2001).

Secondary structures can prevent initiation of translation from both correct and spurious initiation sites in the mRNA (de Smit and van Duin, 1990), usually when the SD sequence and/or the spacer between the SD sequence and translational start codon is inaccessible to the ribosomes (Spanjaard *et al*., 1989; de Smit and van Duin, 1994a,b). Secondary structures in the UTR can also prevent endonucleases from binding and initiating degradation of the mRNA (Grunberg-Manago, 1999; Rauhut and Klug, 1999; Deana and Belasco, 2005).

Mutations near the transcriptional start site has been shown to influence start site selection, and thereby the expressional level (Walker and Osuna, 2002; Lewis and Adhya, 2004). *In vitro* experiments have demonstrated that RNA polymerase may produce many short (2 to ~12 nucleotides in length) abortive transcripts during initiation because of unsuccessful promoter escape, and mutations near the transcriptional start site has shown to influence the abortive to productive transcription initiation ratio (Hsu, 2002). Abortive transcripts may be an *in vitro* artefact (Kubori and Shimamoto, 1996), but the RNA polymerase has to undergo promoter escape *in vivo* to proceed from the initiation to the elongation phase. During this transformation the RNA exit channel needs to be freed, and based on X-ray crystallographic structures and

Received 10 March 2008; accepted 5 October 2008. *For correspondence. E-mail: svein.valla@biotech.ntnu.no; Tel. (+47) 73598694; Fax (+47) 73591283.

© 2009 The Authors
Journal compilation © 2009 Society for Applied Microbiology and Blackwell Publishing Ltd

biophysical data this involves displacement of the $\sigma_{3.2}$ loop (transcript maximally 12 nucleotides) and release of the contact between σ_4 and the RNA polymerase β-flap (transcript maximally 15 nucleotides) (Murakami and Darst, 2003; Paget and Helmann, 2003).

Based on the information presented above there are many ways by which the UTR may be envisioned to affect expression, and this indicated to us that by creating large combinatorial mutant libraries of UTR nucleotide sequences, its range of potential influence on gene expression might be analysed more efficiently than by site-directed mutagenesis and error-prone PCR, often used in such studies (Komarova *et al.*, 2002; 2005; Huang *et al.*, 2006; Park *et al.*, 2007). Here we have specifically chosen to study its potential to stimulate the level of expression, because this is obviously very important in biotechnology.

The inducible and σ^{32}/σ^{38}-dependent *Pm* promoter (Marqués *et al.*, 1999) and its cognate UTR together with its positive regulator gene *xylS* have previously been integrated into minimal replicon broad-host-range expression vectors based on the RK2 plasmid (Blatny *et al.*, 1997a,b). To facilitate the screening for high level expressing mutants the penicillin resistance gene (*bla*) was used as a reporter. The advantage of this system is that it allows very strong selection for enhanced expression levels by simply plating cells from the library on agar medium supplied with a gradient of increasing levels of ampicillin (Winther-Larsen *et al.*, 2000). The studies led to identification of UTR DNA sequences that strongly stimulate expression at the protein level (up to 20-fold), but more surprisingly, also by enhancing the rate of transcript formation and accumulation. We believe that these results are of great significance, partly because they have important consequences for the basic understanding of gene expression, and partly because the wild-type *Pm/xylS* system was previously shown to be useful for production of recombinant proteins at industrial levels (Sletta *et al.*, 2004; 2007).

Results

Expression from the Pm *promoter can be strongly stimulated by mutating the DNA region corresponding to its cognate UTR*

Initially, a mutated *Pm* UTR library (LI) was constructed in the corresponding part of the *Pm/xylS* expression cassette (*bla* used as reporter gene) by using randomly mutated synthetic oligonucleotides. The mutations were introduced throughout almost the entire UTR sequence and mutants displaying a 10-fold increase in ampicillin-resistance levels compared with the wild type, could easily be isolated from this library. DNA sequence analysis showed that they carried three to six mutations in the UTR DNA sequence, none of which affected the putative SD sequence. Four examples of such mutants are indicated in the legend to Fig. 1. Most of the identified sequences had mutations near the transcriptional start site, but as this was not always the case we wanted to specifically investigate how sequences in the internal parts of the UTR affect gene expression. We therefore prepared a new mutated *Pm* UTR library (LII) in plasmid pIB11 (Fig. 1), based on an oligonucleotide mixture in which only bases more than 7 bp downstream of the transcriptional start site were subjected to random sequence changes. Surprisingly, clones with equally high ampicillin-resistance levels as in LI could be identified also from this new library. Twelve such mutants were selected for further studies, first by sequencing their UTR DNA sequences (LII-1 to -12), followed by new synthesis of the corresponding oligonucleotides (both strands). These were then exchanged with the corresponding part from the wild-type pIB11, reconfirmed by DNA sequencing, and phenotype analysed with respect to ampicillin resistance of the hosts, confirming identity with the original mutants isolated. This means that the phenotypes of the mutants are caused solely by the UTR mutations.

Cells containing pIB11 with the wild-type UTR DNA sequence tolerated about 0.01 (uninduced) and 1 mg ml^{-1} (induced) of ampicillin, while the corresponding numbers for hosts carrying pIB11 with the UTR DNA sequences LII-1 to -12 were 0.15–0.25 and 9.5–11 mg ml^{-1} respectively. The identified mutated sequences carried two to five changes relative to the wild type, but some positions remained unchanged, including the GGAGT in or close to the putative SD sequence (Fig. 1). Certain nucleotide positions were substituted in many of the identified sequences, such as the A to C transversion at position +10 (9 out of 12) relative to the transcriptional start site. Interestingly, a majority of the mutations were A to C and T to C substitutions, whereas no G's had been introduced.

Gene expression can be stimulated both at the transcript and protein product levels by mutating the DNA region corresponding to the UTR

The very strong increases in ampicillin resistances displayed by the mutants containing the LII-1 to -12 sequences were assumed to be a direct result of enhancement of β-lactamase production. An obvious explanation could be that the transcripts produced were more efficiently translated than the corresponding wild-type transcripts. To investigate this we selected six of these mutants (carrying LII-7 to -12) for further studies, and first quantified the β-lactamase activities of the corresponding host cells. The level of stimulation under induced conditions relative to the wild-type was as high as up to about 20-fold (LII-11; Fig. 2A), an unexpectedly high

© 2009 The Authors

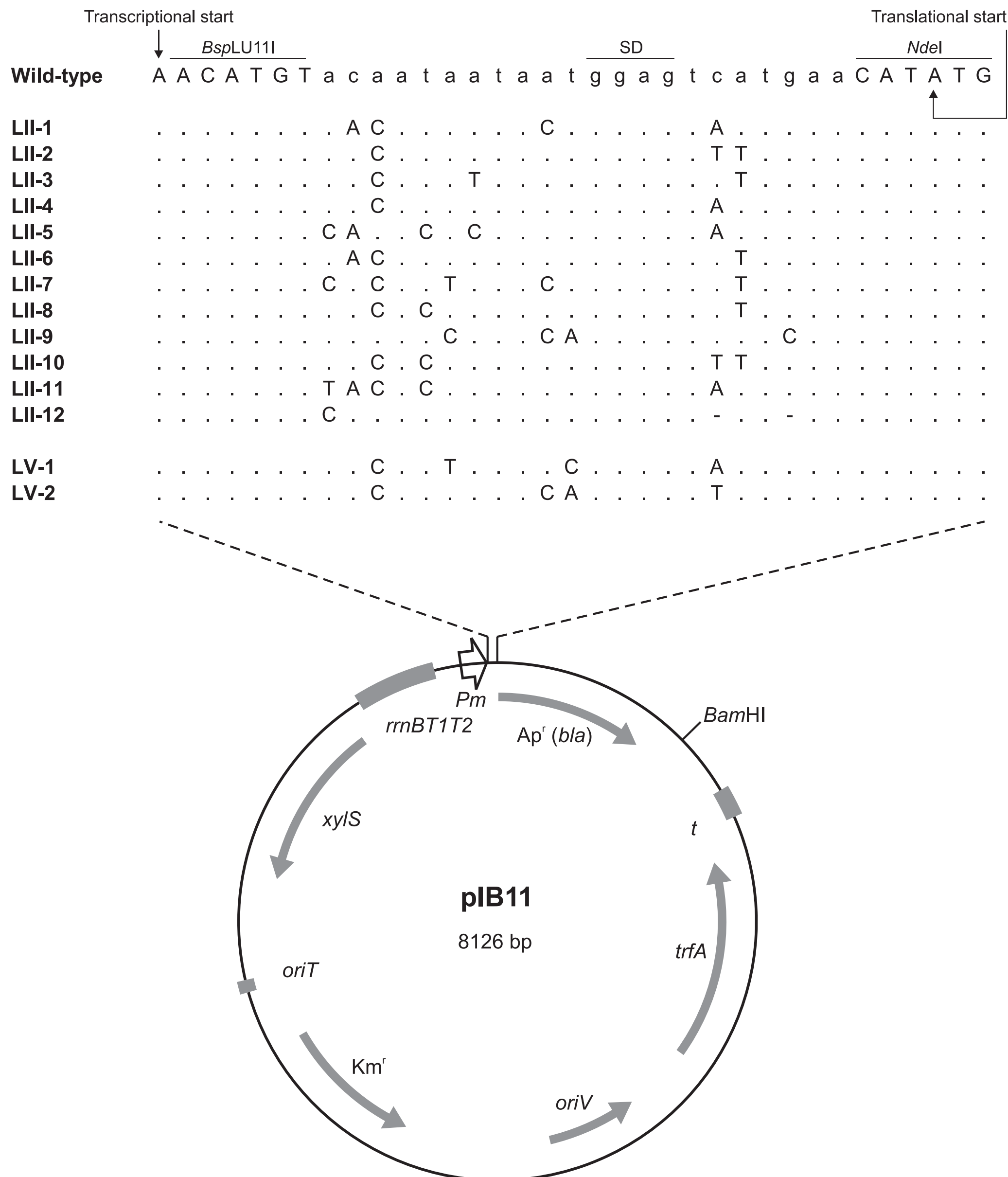

Fig. 1. Map of plasmid pIB11 and mutations giving rise to stimulated expression of *bla*. Restriction sites shown are unique. *Pm*, positively regulated promoter; *xylS*, gene encoding *Pm* activator; Ap^r (*bla*), ampicillin-resistance gene encoding β-lactamase; Km^r, kanamycin-resistance gene; *trfA*, gene encoding the essential replication protein; *oriV*, origin of vegetative replication; *oriT*, origin of transfer; *t*, bidirectional transcriptional terminator; *rrnBT1T2*, bidirectional transcriptional terminator. Details for the DNA sequences corresponding to the *Pm* 5′-untranslated transcript region (UTR) are displayed above the plasmid map. SD is the putative Shine–Dalgarno sequence. Nucleotides in lowercase were randomly mutagenized. Deletion mutations are indicated by short horizontal lines. Transcriptional and translational start sites are indicated with an arrow. The following base substitutions were identified in four selected examples from the screening of the LI library (nucleotide numbers are indicated in the 5′ to 3′ direction). Sequence 1: A2C, C3A, A14T. Sequence 2: C3A, A10C, T12C, A14C. Sequence 3: A2T, C3A, G6C, T12C, A14T. Sequence 4: A2T, A8G, C9T, T12C, A14T, T15A. Note that A10C is also found in nine of the 12 LII sequences and T12C is found in four. The LI library was constructed in plasmid pLB1, which differs from pIB11 in the UTR region in that A4 has been changed to T, and T5 to A.

number taken into consideration that the wild-type system has been shown to be capable of industrial level production of recombinant proteins (see *Introduction*).

Interestingly, corresponding measurements of transcript amounts by quantitative real-time PCR (qRT-PCR) showed that these levels were also very significantly stimulated (a factor of about 7 for LII-11; Fig. 2A), although not to the same extent as at the protein product level. Possible explanations for these observations might be that the UTR mutations were causing the transcripts to be less susceptible to RNA degradation, that more efficient translation leads to protection against

© 2009 The Authors

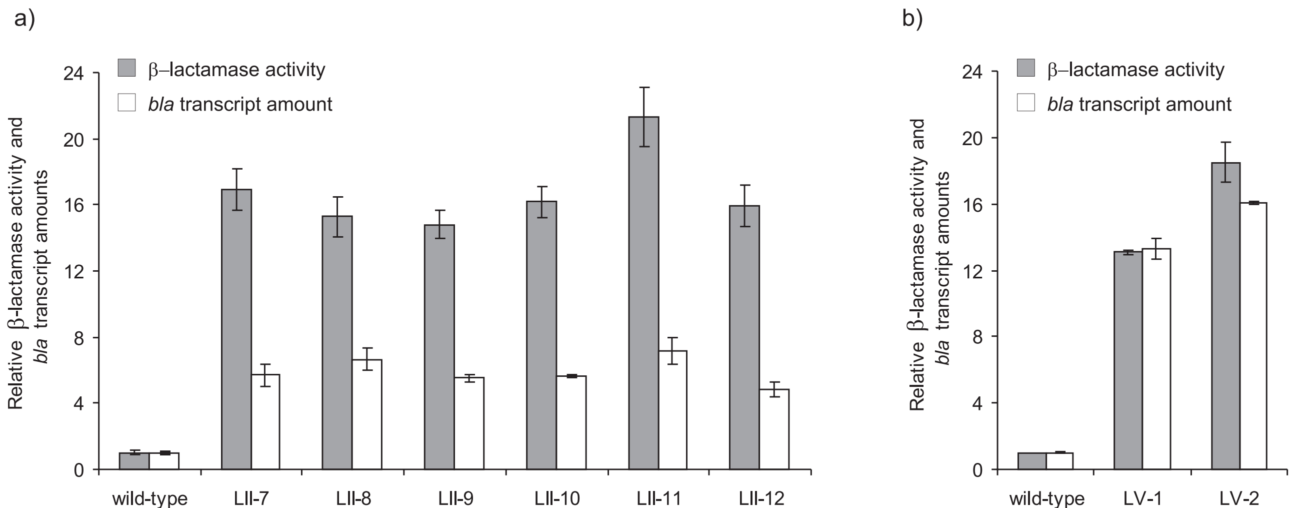

Fig. 2. β-Lactamase activity (grey) and *bla* transcript (white) amounts for pIB11 with the UTR DNA sequences LII-7 to LII-12 (A) and LV-1 and LV-2 (B) relative to the wild type, expressed in *E. coli* DH5α. The values are the average of at least two biological recurrences, and the error bars show the deviation between them. Wild-type values are arbitrarily set to 1.

decay, or that more transcripts were formed by RNA polymerase.

UTR DNA sequences which strongly stimulate the rate of transcript formation independent of the promoter could be identified

The screenings reported here were parts of a bigger effort involving different reporter genes and screening protocols, and results from these studies indicated that the strong stimulation of the amounts of transcripts present was difficult to explain as being the result of enhanced stability alone (data not shown). Furthermore, a change in the transcriptional start site was found not to be essential, as primer extension analysis demonstrated that the start site for *Pm* combined with LII-11 is identical to that of *Pm* combined with the wild-type UTR (data not shown). We therefore decided to directly address the possibility that more transcripts might have been formed. In the screenings reported here selection is directly in favour of stimulated protein product formation. Mutations that putatively stimulate transcript formation might not necessarily be optimal for translation, and thus less optimal for maximal protein product formation. In screening protocols with alternative reporter genes (data not shown) we identified UTR DNA sequences leading to an increase in protein production which correlated with the increase in transcript amounts. Two such UTR DNA sequences (LV-1 and LV-2) are shown in Fig. 1. After insertion of these two sequences in the pIB11 context, it was found that the LV-1 and LV-2 mutations caused an about equal and strong stimulation (about 13- and 17-fold respectively) of the *bla* gene expression at both the transcript and protein product levels (Fig. 2B). This observation might mean that the process of transcript formation has become stimulated by the mutations, leaving the efficiency of translation per transcript more or less unaffected. Alternatively, the mutations have resulted in the production of a more stable transcript, so that the observed accumulation is the result of reduced decay rates.

To distinguish between the hypotheses described above we constructed a transcriptional fusion between the *bla* gene and a sequence encoding a tRNA (*tRNA*^{Arg5}) such that both genes become expressed as one single transcript from plasmid pRL11 (Fig. 3). The rationale behind this is that the *tRNA*^{Arg5} part is essentially stable, so that any stimulation of *bla* transcription would also lead to an increase in the *tRNA*^{Arg5} formation (Lopez *et al*., 1994). In contrast, if the increase in *bla* transcript accumulation is caused by enhanced mRNA stability one would expect essentially no increase in the amounts of *tRNA*^{Arg5}. To test this we used plasmid constructs that were identical, except that the UTR DNA sequence was the wild-type (pRL11), LV-1 or LV-2. It was again, as expected, found that the ampicillin-resistance levels of the corresponding host cells were much higher for cells containing the mutated UTR DNA sequences compared with those containing the wild-type sequence. Quantitative real-time PCR was then used to compare the amounts of *bla* and *tRNA*^{Arg5} transcripts (Fig. 4). The amounts of *bla* transcript were strongly increased (about 11-fold) in the cells containing the LV-1 or LV-2 sequences, relative to those containing the wild-type sequence, as expected. More interestingly, the amounts of *tRNA*^{Arg5} had also increased about eightfold. Based on the assumption that the *tRNA*^{Arg5} part is essentially stable, this appears to

© 2009 The Authors
Journal compilation © 2009 Society for Applied Microbiology and Blackwell Publishing Ltd,

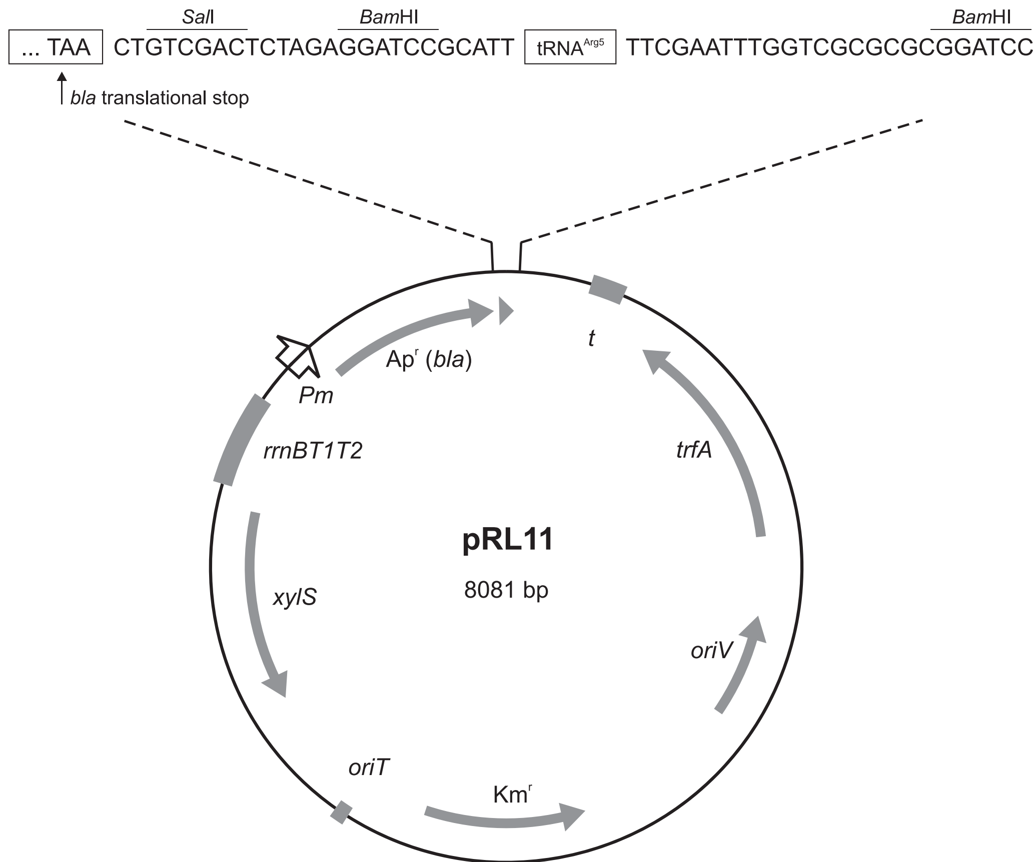

Fig. 3. Map of plasmid pRL11. The DNA sequence in the region of the fusion between *bla* and *tRNA^Arg5* genes is displayed above the plasmid map. The *tRNA^Arg5* DNA sequence is reported elsewhere (Lopez *et al*., 1994). For further details see the legend to Fig. 1.

confirm that the UTR mutations have resulted in a large increase in the amounts of transcript formed by the RNA polymerase.

As understanding of the parameters controlling the levels of gene expression is obviously of very general importance in biology and biotechnology, we wanted to confirm the conclusion above by an alternative and independent type of experiment. To achieve this we studied the kinetics of transcript induction from time zero, immediately before the inducer is added, and up to the point where transcript accumulation had ceased. If the rate of *bla* transcript formation is increased in the cells containing the mutated UTR sequence, one would expect the kinetics of accumulation to be faster in the mutant strain than in the wild type from just after induction and continuously up to an equilibrium state where no further accumulation per unit of cell mass would take place. In contrast, if the observed transcript amounts were the result of enhanced transcript stability, one would expect the rates of accumulation to be similar at early stages after induction, while accumulation would continue for much longer time for the strain containing the mutated UTR sequence than for the strain with the wild-type sequence. We found that qRT-

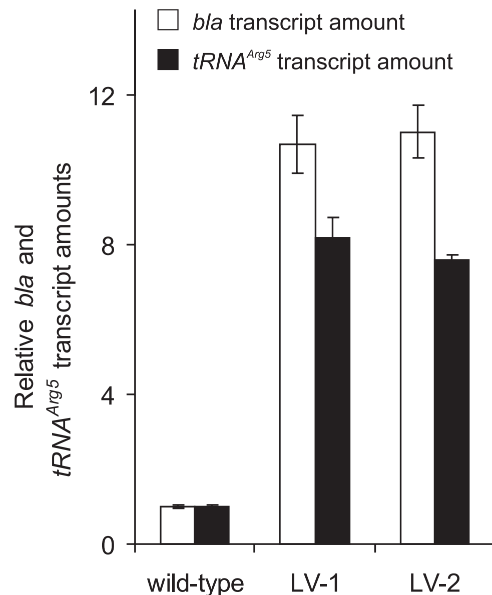

Fig. 4. *bla* (white) and *tRNA^Arg5* (black) transcript amounts for pRL11 with UTR DNA sequence LV-1 and LV-2 relative to the wild type, expressed in *E. coli* DH5α. The error bars show the deviation between three technical recurrences. Wild-type values are arbitrarily set to 1.

© 2009 The Authors

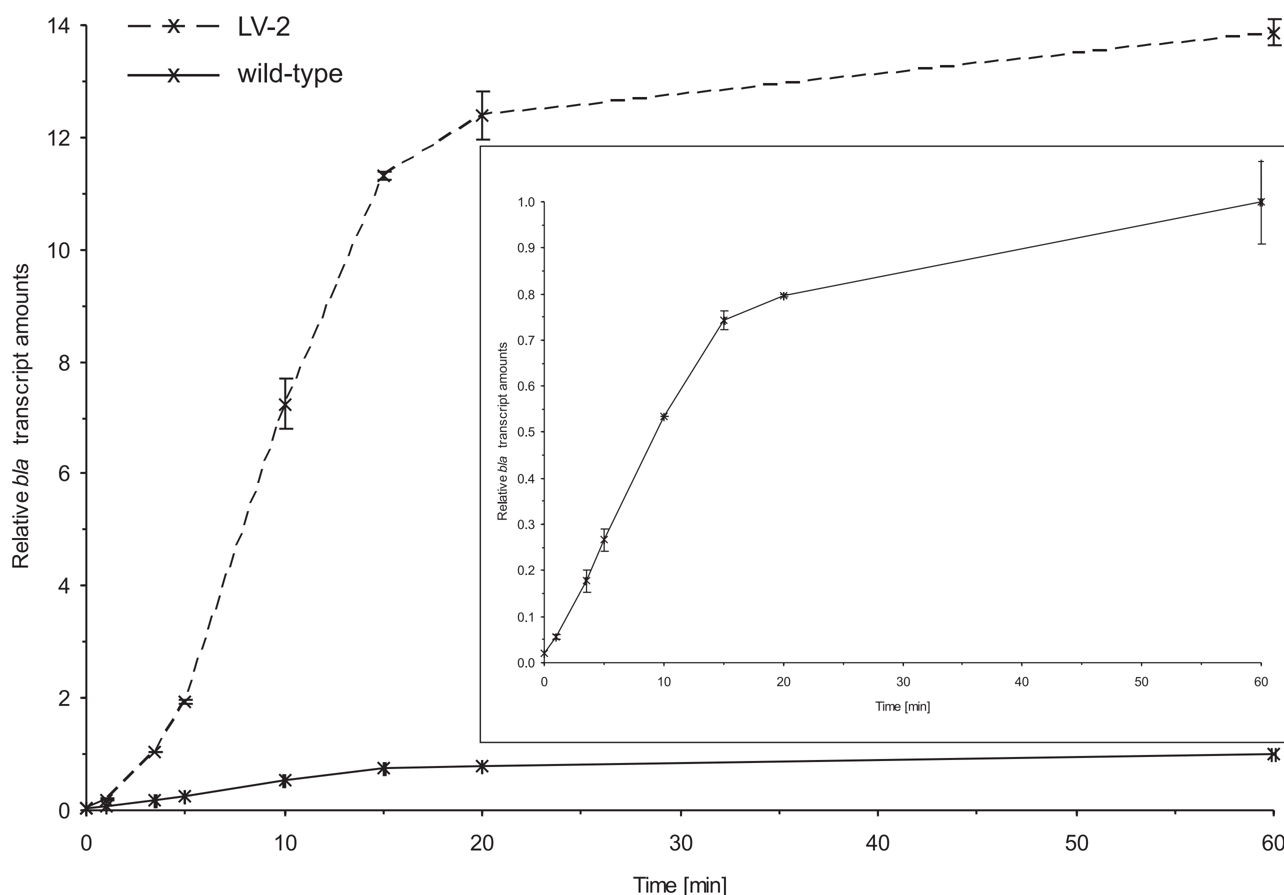

Fig. 5. Kinetics of *bla* transcript accumulation for pIB11 with wild-type (intact line) or LV-2 (dashed line) UTR DNA sequence, expressed in *E. coli* DH5α. Samples were collected at time points 0, 1, 3.5, 5, 10, 15, 20 and 60 min. Inducer was added immediately after the first sample (time zero) was collected. The *bla* transcript amounts are relative to the wild-type at time zero, and the wild-type value at 60 min is arbitrarily set to 1. Error bars show the deviation between two biological recurrences. The relative transcript amounts immediately before onset of induction (time zero) for cells containing the wild-type and LV-2 UTR DNA sequence are 0.02 and 0.04 respectively. The *bla* transcript kinetic curve for the cells containing the wild-type UTR DNA sequence is magnified in the inserted section.

PCR was a very good and reproducible method to quantify the kinetics of transcript accumulation. By comparing the accumulation of the *bla* transcript from cells with plasmid pIB11 (wild-type UTR) and the corresponding plasmid containing the LV-2 UTR, a very clear conclusion could be drawn (Fig. 5). The transcript accumulation rate is significantly higher for the strain with the LV-2 UTR compared with the strain with the corresponding wild-type sequence, with an approximately 14-fold higher final accumulation of transcript, consistent with previous estimates (Fig. 2B). However, the time required to reach the point where accumulation levels off is nearly identical (approximately 20 min). Transcript levels are a balance between transcription rate and decay of individual transcripts, the latter determining the time scale for equilibration after promoter induction. Thus, the similar time scales for transcript level equilibration strongly suggests that transcript stability is similar in these two constructs. This was further analysed by fitting a simple model for transcript accumulation (see *Supporting Information*), where we found no evidence for a difference in transcript stability ($P > 5\%$), average lifetime 10 min, but found a highly significant increase ($P < 1\%$), by at least 18-fold, in the transcription rate in the LV-2 UTR strain. Thus, this kinetics analysis provides strong evidence that the rate of transcript formation has increased, and the results of this experiment are consistent with the *tRNA*^{Arg5} fusion experiment.

Both the wild type and the mutant strain display a short delay in transcript accumulation (from 0 to 1 and about 5 min respectively) (Fig. 5). The explanation for this is not known but may simply be due to the combined time needed for sufficient inducer influx into the cells and full activation of the components involved in *Pm* activation.

Furthermore, to eliminate the possibility that the wild type and the LV-2 mutant have different transcriptional start sites, we decided to determine it experimentally. The outcome of these studies showed that the start site is identical in both constructs, also consistent with earlier findings (Ramos *et al.*, 1987; Domínguez-Cuevas *et al.*,

© 2009 The Authors

2005) as indicated in Fig. 1. The possibility that RNA polymerase selects a new start site in the LV-2 mutant can thus be excluded.

Finally, to test if the transcription stimulation displayed by the two LV DNA sequences is independent of the promoter, we replaced the *Pm* promoter (σ^{32}/σ^{38}-dependent) in plasmid pIB11 with the constitutive σ^{70} promoter *P1* (generating plasmid pLB9). The UTR sequence in pLB9 was then substituted with those of LV-1 and LV-2. Phenotypic characterization of the resulting mutant cells showed that the ampicillin-resistance levels were about 10-fold higher than that of the corresponding wild-type (pLB9). This strongly indicates that the LV-1 and LV-2 sequences can stimulate transcription from very different types of promoters.

Based on the results reported here we conclude that the UTR itself and/or the DNA region corresponding to it are of critical importance for the process of transcription. This conclusion appears to have important consequences for the basic understanding of gene expression in general, and also for recombinant protein production in research and industry.

Discussion

The data presented in this paper illustrate the complexity of understanding the role the UTRs and their corresponding DNA sequences play in determining gene expression at the transcript and protein levels. Taking into consideration the very extensive research that has been ongoing in this field for several decades it may seem surprising that the effects reported here have not been discovered earlier. We believe that this can partly be explained by the more recent methodological improvements allowing more precise transcript level quantifications, and efficient random mutant library constructions combined with a very strong positive screening system (level of host resistance to ampicillin).

The UTR DNA sequences identified in the LII library are characterized by a strong stimulation at both the transcriptional and translational levels (Fig. 2A), while those derived from the LV library stimulate mainly by enhancing the rate of transcription (Figs 2B, 4 and 5). We believe that this difference can be explained by analysing how the screening system works. The LII DNA sequences were identified based on the highest possible level of protein product (β-lactamase). This is not necessarily achieved by maximizing transcription, because the mutations might at the same time also affect translation or transcript stability, potentially negatively. Thus, maximum protein output may be achieved through a compromise between transcriptional and translational stimulation, possibly also including transcript stability. This compromise might be reached by a dominating transcriptional stimulation (as seen for LV-1 and LV-2), a dominating translational stimulation, or a mixture of both (LII mutants).

Even though the design of the screening system can to a great extent explain why the particular phenotypes were selected the underlying molecular mechanisms are far less clear. It is well known that translation is affected by the UTR, and the main focus in the literature has been on the importance of the SD sequence and of secondary structures which may or may not involve parts of the coding sequence. Concerning the translational effects observed with the mutants from the LII library it was interesting to note that the mutations did not affect the putative SD sequence, and in addition we were unable to detect any correlations with putative secondary structures eliminated or established by the mutated UTRs. Furthermore, the assumed preference for A and T/U in UTR to diminish secondary structures (Laursen *et al.*, 2005) does not agree with the high number of nucleotide substitutions in which C had been introduced. Based on these results we therefore believe that the role of the entire UTR sequence context in the control of gene expression levels through translational effects is more complex than previously anticipated, possibly due to subtle interactions with parts of the components in the ribosome. This hypothesis may potentially be studied in more detail by *in vitro* binding and translational studies involving the entire ribosome or parts of it, like the S1 protein (see *Introduction*).

The strong stimulation of transcript levels derived from all the mutated UTR DNA sequences reported here, but most clearly the sequences originating from the LV library were surprising, and illustrate that the transcriptional process can be strongly stimulated by changing bases quite far downstream of the transcriptional start site in the DNA region corresponding to the UTR. The mechanisms underlying this is not clear, but it seems tempting to speculate that the effects could be caused by a more rapid transition from closed to open transcription initiation complex and/or a faster promoter escape by RNA polymerase. The DNA sequence from –11 to +20 occupies the active-channel in both the first kinetically significant intermediate in RNA polymerase open complex formation and in the RNA polymerase open complex itself, based on real-time hydroxyl radical (•OH) and potassium permanganate ($KMnO_4$) footprints (Davis *et al.*, 2007). During the subsequent promoter escape, the RNA polymerase has to undergo major structural changes, including the displacement of the $\sigma_{3.2}$ loop to free the RNA exit channel when the transcript is maximally ~12 nucleotides in length (see *Introduction*). In both of these processes there are contacts between the RNA polymerase and the DNA region corresponding to the UTR. In addition, during the promoter escape the UTR itself might influence the clearing of the RNA exit channel.

© 2009 The Authors

Based on the results reported here we conclude that the UTR and/or the DNA region corresponding to it have a significantly more complex role in affecting gene expression than previously realized, through a hitherto unrecognized effect on transcript formation. In addition, the entire UTR sequence, and not only the SD sequence, seems to play an important role in the translational process. The underlying mechanisms behind these effects are unknown, but it appears likely that the answers will have to be found by gaining deeper insights into the process of promoter escape *in vivo*, and in the complex interactions that take place between the UTR and the ribosome during translation.

Experimental procedures

Bacterial strains, plasmids and growth media

Escherichia coli DH5α (Bethesda Research Laboratories) was used as host, except for in the construction and screening of the mutated *Pm* UTR library, where *E. coli* Gold (Stratagene) was used.

pJT19*bla* (Winther-Larsen *et al.*, 2000) represents the origin for the plasmid constructs used in this work, and is a RK2-based expression vector containing the *Pm/xylS* expression cassette with the ampicillin-resistance gene, *bla*, as the reporter gene for *Pm*. Plasmid pIB11, used for construction of the mutated *Pm* UTR library, is a pJT19*bla* derivative with a XbaI/MunI fragment containing the *rrnBT1T2* (transcriptional terminator) (Santos *et al.*, 2001) cloned into the same sites of pJT19*bla*, an AflIII restriction site introduced upstream of the *Pm* promoter, and a BspLU11I restriction site inserted downstream of the *Pm* transcriptional start site (Fig. 1). pRL11 (Fig. 3) is a pIB11 derivative, and its construction required a large number of steps that are described in *Supporting Information*. The differences between pRL11 and pIB11 consist in an altered DNA sequence directly upstream of the unique BamHI restriction site in pIB11, including the insertion of a SalI site, and the deletion of approximately 150 bp directly downstream of the translational stop codon of the *bla* gene in pIB11. The $tRNA^{Arg5}$ gene is cloned into the BamHI and thus transcriptionally fuses *bla* and $tRNA^{Arg5}$ (Fig. 3). The $tRNA^{Arg5}$ gene was PCR amplified from plasmid pMAMA5-Arg5 (Lopez *et al.*, 1994) by using the primer set 5′-CAGTAATTTCGAGGATCCGCATTGTC-3′ and 5′-AAGG ATCCGCGCGCGACCA-3′. Plasmid pLB9 is a pIB11 derivative with a BglII/BspLU11I fragment containing the *P1* (also known as $P_{anti\text{-}tet}$) promoter (Brosius *et al.*, 1982) cloned into the same sites of pIB11. The *P1* promoter was PCR amplified from plasmid pPK34 (Karunakaran *et al.*, 1998) by using the primer set 5′-AGCCTATGCCTAGATCTTCCAGG GTGACG-3′ and 5′-TATCATCGATAACATGTAATGCGGTA GT-3′.

Cells were grown in L broth (10 g l^{-1} tryptone, 5 g l^{-1} yeast extract, and 5 g l^{-1} NaCl) or on L agar (L broth with 15 g l^{-1} Agar) at 37°C, except for in expression studies, where 30°C was used. Kanamycin (50 μg ml^{-1}) was used as a marker for establishment of the mutated *Pm* UTR libraries and the *Pm/xylS* inducer, *m*-toluate, was always used at a concentration of 2 mM.

Standard DNA manipulations

Transformations of *E. coli* were performed with a modified RbCl protocol (http://www.promega.com) in cloning experiments. WizardPlus SV minipreps DNA purification kit (Promega) or Qiagen plasmid midi kit (Qiagen) were used for plasmid DNA purifications. Enzymatic manipulations were performed as described by the manufacturers. PCR reactions were performed using the Expand high fidelity PCR system kit (Boehringer Mannheim) for cloning purposes, the polymerase DynazymeII (Finnzymes) for generation of templates for DNA sequencing, and Quickchange II site-directed mutagenesis kit (Stratagene) for site-directed mutagenesis. PCR templates were treated with the enzyme mixture ExoSapIt (USB) prior to DNA sequencing. Sequencing reactions were performed with the ABI PRISM BigDye sequencing kit (Applied Biosystems), and analysed using an Applied Biosystems 3130 XL DNA sequencing machine (Applied Biosystems).

Construction of the mutated Pm *UTR libraries*

Ampicillin-resistance levels of host cells containing a plasmid-encoded β-lactamase gene (*bla*) is approximately proportional to the copy number of the plasmid (Uhlin and Nordström, 1977), reflecting enhanced expression due to increased gene dosage. Thus, ampicillin resistance can be used to estimate changes in the *Pm* UTR activity. The pJT19*bla* derivative pIB11 was used for constructing the mutated *Pm* UTR library, with the *bla* gene as a reporter of *Pm* activity. A kanamycin-resistance gene allows for plasmid selection, avoiding that *Pm* expression levels (ampicillin resistance) affect the composition of the UTR libraries. Construction of the LII library was performed according to the procedure previously described (Winther-Larsen *et al.*, 2000), with the exception that the oligonucleotides were designed to constitute a double-stranded DNA fragment containing the *Pm* UTR and with BspLU11I- and NdeI-compatible ends when annealed, for subsequent easy cloning into the pIB11 vector. One of the oligonucleotides corresponded to the wild-type *Pm* UTR sequence (5′-TATGTTCATGACTC CATTATTATTGTA-3′), and the oligonucleotide corresponding to the complementary strand was randomly mutagenized by the use of a mixed oligonucleotide solution (5′-CATGT121141141433134214311CA-3′, where the numbers in sequence indicate the doping percentages of the nucleotides: 1 = 70% A, 10% C, G and T; 2 = 70% C, 10% A, G and T; 3 = 70% G, 10% A, C and T; 4 = 70% T, 10% A, C and G). Approximately 25 000 transformants were obtained, mixed and used as a library (library LII). The LV library was constructed in a similar way by using the same oligonucleotide mixture, but the plasmid construct used in the original screening had a different history (see *Supporting Information*).

Screening for mutated UTR sequences leading to stimulation of expression from the Pm *promoter*

Screening for mutated *Pm* UTR sequences resulting in increased ampicillin-resistance levels was performed

© 2009 The Authors

according to the procedure previously described (Winther-Larsen *et al.*, 2000). The sequences of the *Pm* UTRs were determined for the selected high level expression candidates using the sequencing primer 5′-CTATCAAACCGGACA CGTTTATCGTGGTTATGC-3′. PCR products generated with the primers 5′-CTTTCACCAGCGTTTCTGGGTG-3′ and 5′-GATGTAGAAAGGCGCCAAGTC-3′ were used as templates for the sequencing reactions. All mutants reported under *Results* were reproduced by cloning annealed newly synthesized oligonucleotides with the mutations identified in the original isolates. The reproduced mutants were also confirmed by DNA sequencing, as described above.

β-Lactamase assay

Growth of strains and β-lactamase assay were performed according to the procedures previously described (Winther-Larsen *et al.*, 2000). All enzyme activity analyses were repeated at least twice, and measurements were carried out with minimum three technical recurrences.

RNA isolation, cDNA synthesis, and transcript quantification by qRT-PCR

Strains were grown as previously described (Winther-Larsen *et al.*, 2000). Cell cultures were stabilized with RNAprotect (Qiagen) prior to freezing and RNA was isolated from the frozen cell pellets using the RNAqeous kit (Ambion) as described by the manufacturers. The RNA preparations were treated with DNase (DNA-free, Ambion) and cDNA was produced from 3 μg total RNA as template using the First-Strand cDNA synthesis kit (Amersham Biosciences) with random $pd(N)_6$ primers as described by the suppliers. Two-step qRT-PCR with the power SYBR green PCR master mix (Applied Biosystems) in a 7500 Real Time PCR System instrument (Applied Biosystems) was used for quantification of *bla* and $tRNA^{Arg5}$ transcripts. PCR cycles were 95°C for 10 min, followed by 40 cycles of amplification (95°C for 15 s; 60°C for 1 min) except for the *bla* and $tRNA^{Arg5}$ fusion experiment were the PCR cycles were 95°C for 10 min, followed by 40 cycles of amplification (95°C for 15 s; 47°C for 1 min). Results were analysed using 7500 system software v1.3, and data were normalized by the $2^\wedge\Delta\Delta C_T$ method (Livak and Schmittgen, 2001). Primers were designed using the primer express software (Applied Biosystems). Primer pairs used for transcript quantification were 5′-ACGTTTTCCAATGATGAGCACTT-3′ and 5′-TGCCCGGCGTCAACAC-3′ for *bla*, and 5′-GTCC TCTTAGTTAAATGG-3′ and 5′-AGGAATCGAACCTGC-3′ for $tRNA^{Arg5}$. A fragment from the kanamycin-resistance gene (primer pair 5′-TACCTTTGCCATGTTTCAGAAACA-3′ and 5′-AATCAGGTGCGACAATCTATCGA-3′) or the 16S rRNA gene (primer pair 5′-ATTGACGTTACCCGCAGAAGAA-3′ and 5′-GCTTGCACCCTCCGTATTACC-3′) was used as a normalizer. The normalizer genes gave nearly identical results.

Transcriptional start site determination

The transcriptional start site was determined for the pIB11 construct containing the wild-type or LV-2 DNA sequence by generating 5′-RACE-Ready cDNA from 100 μg ml^{-1} of total RNA (isolated as described above) by SMART RACE cDNA amplification kit (Clontech). Takara Bio's PrimeScript (Clonetech) was used as reverse transcriptase. cDNA was purified by QIAquick PCR purification kit (Qiagen), and used as a template in the subsequent PCR step. 5′-RACE PCR reaction was carried out with SMART universal primer mix, the *bla* gene specific primer 5′-CTTTCACCAGCGTTTCTG GGTG-3′, and Expand high fidelity PCR system kit (Boehringer Mannheim). Single PCR product was excised from gel, purified by QIAquick gel extraction kit (Qiagen) and TA cloned by using the PCR cloning kit (Qiagen). Selected clones from the resulting constructs were subjected to sequencing (Eurofins MWG Operon) by using the M13 uni (-21) 5′-TGTAAAACGACGGCCAGT-3′ as sequencing primer.

Acknowledgements

We would like to thank professor Agamemnon J. Carpousis for suggesting the tRNA fusion experiment, and Professor Mark Dreyfus for kindly providing us with important strains, and MWG Biotech AG for help with the design of primers for qRT-PCR quantification of tRNA. This work was supported by the Norwegian Research Council, Grant Number 140533.

References

Blatny, J.M., Brautaset, T., Winther-Larsen, H.C., Haugan, K., and Valla, S. (1997a) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the rk2 replicon. *Appl Environ Microbiol* **63:** 370–379.

Blatny, J.M., Brautaset, T., Winther-Larsen, H.C., Karunakaran, P., and Valla, S. (1997b) Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. *Plasmid* **38:** 35–51.

Brosius, J., Cate, R.L., and Perlmutter, A.P. (1982) Precise location of two promoters for the β-lactamase gene of pBR322. *J Biol Chem* **257:** 9205–9210.

Davis, C.A., Bingman, C.A., Landick, R., Record, M.T., Jr, and Saecker, R.M. (2007) Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci USA* **104:** 7833–7838.

Deana, A., and Belasco, J.G. (2005) Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* **19:** 2526–2533.

Domínguez-Cuevas, P., Marín, P., Ramos, J.L., and Marqués, S. (2005) RNA polymerase holoenzymes can share a single transcription start site for the Pm promoter: Critical nucleotides in the -7 to -18 region are needed to select between RNA polymerase with σ38 or σ32. *J Biol Chem* **280:** 41315–41323.

Farwell, M.A., Roberts, M.W., and Rabinowitz, J.C. (1992) The effect of ribosomal protein S1 from *Escherichia coli* and *Micrococcus luteus* on protein synthesis in vitro by *E. coli* and *Bacillus subtilis*. *Mol Microbiol* **6:** 3375–3383.

Grunberg-Manago, M. (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet* **33:** 193–227.

© 2009 The Authors

Hsu, L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta* **1577:** 191–207.

Huang, X., Zhang, X., Zhou, Y., Zhang, Z., and Cass, A.E.G. (2006) Directed evolution of the 5′-untranslated region of the *phoA* gene in *Escherichia coli* simultaneously yields a stronger promoter and a stronger Shine-Dalgarno sequence. *Biotechnol J* **1:** 1275–1282.

Karunakaran, P., Blatny, J.M., Ertesvåg, H., and Valla, S. (1998) Species-dependent phenotypes of replication-temperature-sensitive *trfA* mutants of plasmid RK2: a codon-neutral base substitution stimulates temperature sensitivity by leading to reduced levels of *trfA* expression. *J Bacteriol* **180:** 3793–3798.

Komarova, A.V., Tchufistova, L.S., Supina, E.V., and Boni, I.V. (2002) Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA* **8:** 1137–1147.

Komarova, A.V., Tchufistova, L.S., Dreyfus, M., and Boni, I.V. (2005) AU-rich sequences within 5′ untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol* **187:** 1344–1349.

Kubori, T., and Shimamoto, N. (1996) A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. *J Mol Biol* **256:** 449–457.

Laursen, B.S., Sørensen, H.P., Mortensen, K.K., and Sperling-Petersen, H.U. (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69:** 101–123.

Lewis, D.E.A., and Adhya, S. (2004) Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol Microbiol* **54:** 692–701.

Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* **25:** 402–408.

Lopez, P.J., Iost, I., and Dreyfus, M. (1994) The use of a tRNA as a transcriptional reporter: the T7 late promoter is extremely efficient in *Escherichia coli* but its transcripts are poorly expressed. *Nucleic Acids Res* **22:** 1186–1193.

McCarthy, J.E.G., and Brimacombe, R. (1994) Procaryotic translation: the interactive pathway leading to initiation. *Trends Genet* **10:** 402–407.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* **60:** 512–538.

Marqués, S., Manzanera, M., González-Pérez, M.M., Gallegos, M.T., and Ramos, J.L. (1999) The XylS-dependent Pm promoter is transcribed with σ32 or σ38 depending on the growth phase. *Mol Microbiol* **31:** 1105–1113.

Murakami, K.S., and Darst, S.A. (2003) Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* **13:** 31–39.

Paget, M.S.B., and Helmann, J.D. (2003) The σ70 family of sigma factors. *Genome Biol* **4:** 203–208.

Park, Y.S., Seo, S.W., Hwang, S., Chu, H.S., Ahn, J., Kim, T., *et al.* (2007) Design of 5′-untranslated region variants for tunable expression in *Escherichia coli*. *Biochem Biophys Res Commun* **356:** 136–141.

Ramos, J.L., Mermod, N., and Timmis, K.N. (1987) Regulatory circuits controlling transcription of TOL plasmid operon encoding meta-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. *Mol Microbiol* **1:** 293–300.

Rauhut, R., and Klug, G. (1999) mRNA degradation in bacteria. *FEMS Microbiol Rev* **23:** 353–370.

Roberts, M.W., and Raboniwitz, J.C. (1989) The effects of *Escherichia coli* ribosomal protein S1 on the translational specificity of bacterial ribosomes. *J Biol Chem* **264:** 2228–2235.

Santos, P.M., Di Bartolo, I., Blatny, J.M., Zennaro, E., and Valla, S. (2001) New broad-host-range promoter probe vectors based on the plasmid RK2 replicon. *FEMS Microbiol Lett* **195:** 91–96.

Sengupta, J., Agrawal, R.K., and Frank, J. (2001) Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proc Natl Acad Sci USA* **98:** 11991–11996.

Shine, J., and Dalgarno, L. (1974) The 3′-terminal sequence of *E. coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* **71:** 1342–1346.

Sletta, H., Nedal, A., Aune, T.E.V., Hellebust, H., Hakvåg, S., Aune, R., *et al.* (2004) Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. *Appl Environ Microbiol* **70:** 7033–7039.

Sletta, H., Tøndervik, A., Hakvåg, S., Aune, T.E.V., Nedal, A., Aune, R., *et al.* (2007) The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl Environ Microbiol* **73:** 906–912.

de Smit, M., and van Duin, J. (1990) Control of prokaryotic translational initiation by mRNA secondary structure. *Prog Nucleic Acids Res Mol Biol* **38:** 1–35.

de Smit, M., and van Duin, J. (1994a) Translational initiation on structured messengers (Another role for the Shine-Dalgarno interaction). *J Mol Biol* **235:** 173–184.

de Smit, M., and van Duin, J. (1994b) Control of translation by mRNA secondary structure in *Escherichia coli* (A quantitative analysis of literature data). *J Mol Biol* **244:** 144–150.

Spanjaard, R.A., van Dijk, M.C.M., Turion, A.J., and van Duin, J. (1989) Expression of the rat interferon-α1 gene in *Escherichia coli* controlled by the secondary structure of the translation-initiation region. *Gene* **80:** 345–351.

Steitz, J.A., and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S RNA and the mRNA during initiation of protein synthesis in *E. coli*. *Proc Natl Acad Sci USA* **72:** 4734–4738.

Subramanian, A.R. (1992) Structure and functions of ribosomal protein S1. *Prog Nucleic Acids Res Mol Biol* **28:** 101–142.

Sørensen, M.A., Fricke, J., and Pedersen, S. (1998) Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in *Escherichia coli in vivo*. *J Mol Biol* **280:** 561–569.

Tzareva, N.V., Makhno, V.I., and Boni, I.V. (1994) Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS Lett* **337:** 189–194.

Uhlin, B., and Nordström, K. (1977) R plasmid gene dosage effects in *Escherichia coli* K-12: copy mutants of the R plasmid R1drd-19. *Plasmid* **1:** 1–7.

Walker, K.A., and Osuna, R. (2002) Factors affecting start site selection at the *Escherichia coli fis* promoter. *J Bacteriol* **184:** 4783–4791.

© 2009 The Authors

Winther-Larsen, H.C., Blatny, J.B., Valand, B., Brautaset, T., and Valla, S. (2000) Pm promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. *Metab Eng* **2:** 92–103.

Yusupova, G.Z., Yusupov, M.M., Cate, J.H., and Noller, H.F. (2001) The path of messenger RNA through the ribosome. *Cell* **106:** 233–241.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Simulation of mRNA concentration after induction at t = 0.

Fig. S2. Time series with transcript levels rescaled between 0.1.

Fig. S3. Fits of the recovery to time points 3.5 min and above.

Fig. S4. Simulation of mRNA concentration after induction at t = 0 with a model for a finite time for the inducer activity.

Table S1. Model fit and parmeter estimates.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2009 The Authors