

# Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization

Namiko Mitarai<sup>1,2</sup>, Kim Sneppen<sup>1</sup> and Steen Pedersen<sup>3\*</sup>

<sup>1</sup>Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Department of Physics, Kyushu University, Fukuoka, Japan

<sup>3</sup>Department of Biology, University of Copenhagen, Ole Maaløes vej 5, DK-2200 Copenhagen N, Denmark

Received 8 May 2008;  
received in revised form  
19 June 2008;  
accepted 23 June 2008  
Available online  
1 July 2008

Individual mRNAs are translated by multiple ribosomes that initiate translation with an interval of a few seconds. The ribosome speed is codon dependent, and ribosome queuing has been suggested to explain specific data for translation of some mRNAs *in vivo*. By modeling the stochastic translation process as a traffic problem, we here analyze conditions and consequences of collisions and queuing. The model allowed us to determine the on-rate (0.8 to 1.1 initiations/s) and the time (1 s) the preceding ribosome occludes initiation for *Escherichia coli lacZ* mRNA *in vivo*. We find that ribosome collisions and queues are inevitable consequences of a stochastic translation mechanism that reduce the translation efficiency substantially on natural mRNAs. The cells minimize collisions by having its mRNAs being unstable and by a highly selected codon usage in the start of the mRNA. The cost of mRNA breakdown is offset by the concomitant increase in translation efficiency.

© 2008 Elsevier Ltd. All rights reserved.

Edited by M. Gottesman

Keywords: modelling; ribosome collisions; codon usage; translation efficiency

## Introduction

The degeneracy of the genetic code opens for additional layers of coding information of importance for protein expression and regulation in response to a changing environment. One of these layers is associated to the fact that different codons are translated with widely different rates, thereby providing regulatory flexibility at the translational level.<sup>1</sup> Furthermore, it has been found that subsequent ribosomes are loaded onto the mRNA sufficiently fast to make them interfere with each other and sometimes even form extended queues causing a substantial delay for producing proteins from older mRNAs.<sup>2</sup> As a consequence, the translation process itself influences the cellular expression of proteins. To quantify the importance of codon-to-codon variation in translation rate, one needs to develop a mathematical model and to calibrate it to available data on mRNA translation kinetics.

Previously, we developed a method to measure the time it takes for ribosomes to synthesize peptides in living cells on individual mRNA species<sup>1</sup> and refined it to have an SEM (standard error of the mean) of about 0.5 s when measuring translation times of mRNA encoding abundant proteins.<sup>3</sup> We found that mRNAs with the ribosomal protein codon usage<sup>4</sup> were translated approximately 35% faster than mRNAs with a more unbiased codon usage as, for instance, the *lacI* mRNA.<sup>1</sup> This rate difference was determined by the codon usage and not influenced significantly by potential hairpin structures in the slowly translated mRNAs.<sup>3</sup> A subsequent study showed that two synonymous codons read by the same tRNA species were translated with a threefold difference in rate, implying that the slow translation of these codons was not due to limits of tRNA abundance. Furthermore, inserting the sequence CGA<sub>3</sub>(GAG)<sub>24</sub>CCG<sub>3</sub> two times in the *lacZ* mRNA provoked queue formation† and reduced the

\*Corresponding author. E-mail address: steenp@bio.ku.dk.

Abbreviations used: RBS, ribosome-binding site; RNAP, RNA polymerase.

†We define collision as the physical hindrance to translation by a preceding ribosome and queuing as mean extensive collisions where more than two ribosomes are involved.

expression slightly, probably because of transcriptional polarity, whereas the same sequence inserted only once gave no detectable ribosome queues and no reduced expression.<sup>2</sup>

To analyze the interplay between the stochastic movement of ribosomes and queue formation, we determined key parameters associated to translation of the *lacZ* mRNA by introducing a model that reproduces the previous experiments.<sup>2</sup> The model is an extension of the stochastic hopping processes, considered for uniform translation rates<sup>5,6</sup> and extended to codon-dependent translation rates.<sup>7-9</sup> Also, Ringquist *et al.* investigated a number of variables that determined the strength of the *lacZ* ribosome-binding site (RBS), including the strength of the SD (Shine-Dalgarno sequence complementary to the 3' end of 16S rRNA) interaction, the distance between the SD and the initiation codon, and the nature of the initiation codon.<sup>10</sup> They concluded that their data were consistent with a simple kinetic model in which a variety of rate constants contribute to the process of translation initiation. Our work attempts to specify these rates by reproducing experimental data on *lacZ* translation with and without slow inserts in a model. We now can analyze how codon usage and mRNA stability may be used to influence speed, expression yield, and metabolic cost associated to protein production.

## Results

### Description of the model of translation kinetics

First, we model the basic dynamics of translation. Figure 1 depicts ribosomes that translate an mRNA. Translation starts by binding a 30S ribosomal subunit to codon 1 with the rate  $K_s$ , provided that the binding site is accessible, which requires that no ribosome is within the occluding distance,  $d$ , from the binding site. The ribosome stays at codon 1 for a time,  $\tau$ , to assemble the translating 70S ribosome.

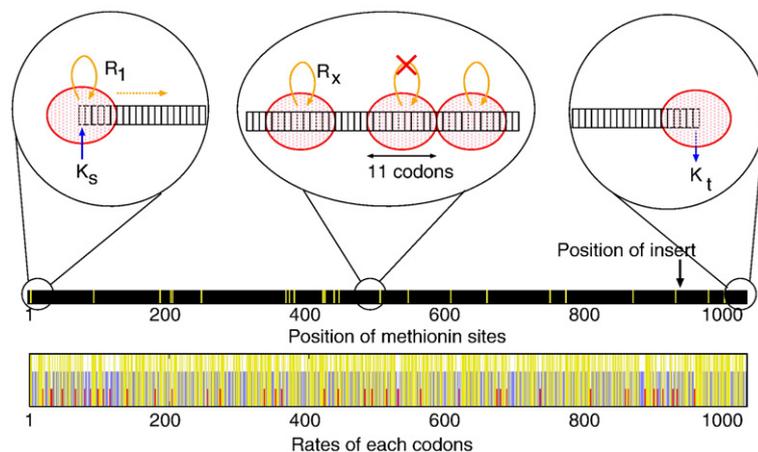
Then, for each time step  $dt$  after  $\tau$ , all ribosomes have a probability  $R_x dt$  to move one codon forward, where  $R_x$  is the translation rate for the codon at position  $x$ . Possible movements are accepted only if the distance to the preceding ribosome is larger than  $d$ . When the ribosome reaches the end of the mRNA, the peptide is released with a rate  $K_t$ .

We also model the finite lifetime of the mRNA by simulating an ensemble of mRNAs produced at random times and subsequently degraded with the rate that reproduces an average of  $p$  translations per mRNA. When mRNA degradation is initiated, translation initiation is stopped and ribosomes that were already initiated continue translation until terminating at the 3' end as suggested by previous experiments.<sup>11,12</sup>

The detailed description of the simulation procedure is given in Methods. The java applet of the model can be found at <http://cmol.nbi.dk/models/RibosomeTraffic/RibosomeTraffic.html>.

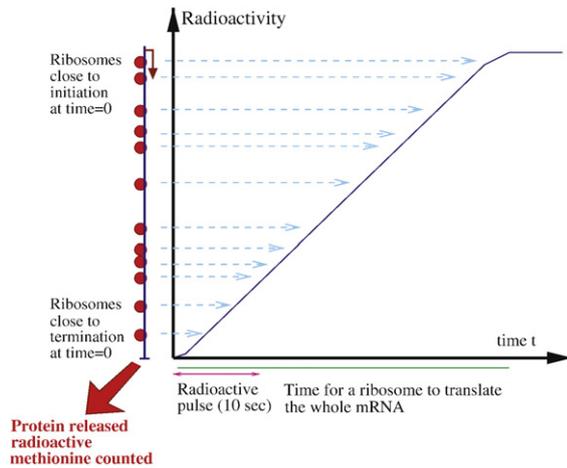
### Description of the experiment

Using this model of translation dynamics, we reproduce the *in vivo* experiment for incorporation of radioactivity in a model protein.<sup>2</sup> The experiment is performed as follows: At time  $t=0$ , the ensemble of mRNAs being translated in steady state is exposed to a 10-s pulse of radioactive <sup>35</sup>S-methionine. The pulse is stopped by adding a large excess of unlabeled methionine to the culture. After this chase, samples are taken at 10-s intervals and the radioactivity in the completed protein is determined. Figure 2 shows a schematic incorporation curve. The sketch of the mRNA along the ordinate axis illustrates that shortly after the <sup>35</sup>S-Met pulse and chase, only ribosomes that were close to terminating at the time of the pulse will contribute radioactivity to the finished peptide. As time progresses and more samples are taken, more and more radioactivity will accumulate in the finished peptide. When the first ribosome on the mRNA at the time of the pulse has terminated



**Fig. 1.** Visualization of the modeled translation of *lacZ* mRNA: Ribosomes move from codon to codon on an mRNA, with translation rates that depend on the codon. Each ribosome occupies a space of 11 subsequent codons, and subsequent ribosomes are initiated with a rate given by  $K_s$ . The time of forming the 70S initiation complex is usually set to 0.2 s. The black middle bar indicates the positions of methionine residues in yellow. The lower panel shows the distribution of the fast translated A codons (yellow), the middle rate B

codons (blue), and the slowly translated C codons (red) in the 1027 codons of the parental *lacZ* gene in pMAS23. When a ribosome reaches the stop codon, it is terminated with the nonlimiting rate constant of 2/s.



**Fig. 2.** Illustration of an idealized experiment where radioactivity is incorporated into full-length protein. The sketch on the ordinate axis indicates that the radioactivity in the early time points originated from the ribosomes close to the 3' end on the mRNA at time of the radioactive pulse.

translation, the incorporation of radioactivity will stop. This defines the translation time as the time it takes to bring the most N-terminal methionine into the full-length protein.

To minimize differences among the experiments, we used a mixture of two cultures, one with the parental plasmid pMAS23 and one encoding *lacZ* with the insert to be compared.<sup>2</sup> The direct comparison in the same experiment was possible because the two full-length proteins were resolved by gel electrophoresis. Furthermore, a double-isotope labeling technique enabled correction for a variable recovery during sample preparation. Normally, 18 data points were collected in an experiment. From these incorporation curves and the position of methionine codons in the mRNA, indicated in Fig. 1, the average translation time of these mRNAs was determined with an SEM of approximately 0.5 s. Comparing mRNA, with and without an insert, we determined the time for translating the insert.<sup>2</sup>

To reproduce this experiment in our model analysis, we prepare the steady-state ensemble of mRNAs, count the incorporation of the radioactive methionine for the ribosomes that pass each methionine codon during the pulse, and calculate how many radioactive methionine residues are included in the finished peptide at each time. This gives the incorporation curve from the simulation, and it is compared with the experimental incorporation curves. Further aspects of the experimental and model methods are discussed in Methods.

### Input parameters to the model

The rate,  $K_s$ , of binding the 30S ribosome to the mRNA can be set in the simulation. For the resulting *lacZ*, mRNA initiation has been estimated to occur about once per 3 s.<sup>2,11</sup> In our analyses, we test on-

rates from 1 per 0.8 s to 1 per 3 s, which will result in initiation rates from approximately 1 per 2 s to 1 per 4 s because it takes approximately 0.2 s to initiate and 1 s to translate the first 11 codons in the *lacZ* mRNA (see later). The rate,  $K_t$ , to release the peptide at the end of the mRNA is set to 2/s, which is faster than the initiation rate and does not provoke ribosome queuing. Thus, this finite termination rate does not influence the translation dynamics in the wild-type situation we consider here.

The occluding distance  $d$  can be estimated to 11 codons because (a) 30S ribosomes protect 30–35 nucleotides against RNase degradation,<sup>13</sup> (b) the 33 nucleotides from -20 to +13 are nonrandom and involved in 30S ribosome binding,<sup>14</sup> and (c) 30S and 70S ribosomes occupy the same space on the mRNA.<sup>14</sup>

An important ingredient in our model is the fact that each mRNA is only used for a limited period. The *lacZ* mRNA was estimated to be translated an average of 40 times<sup>11</sup> and the *trp* mRNA, 20 times.<sup>15</sup> The estimate of 40 proteins per *lacZ* mRNA is perhaps too high: As discussed by Kennell and Riezman, the number for *lacZ* mRNA is critically dependent on the estimate of the enzyme-specific activity of purified  $\beta$ -galactosidase.<sup>11</sup> If the 400,000 enzyme units per milligram of protein<sup>11</sup> is replaced by the 800,000 units per milligram of protein found in a slightly different buffer,<sup>16</sup> the estimated number of protein per mRNA would drop from 40 to 20. The average number for translating the *lacZ* mRNA is therefore likely to be closer to 30 and maybe even as low as 20, which is close to the overall mRNA average translation number in minimal medium.<sup>17</sup> Thus, in our model analysis, we normally set the average translation number to 30 unless otherwise noted.

During initiation, the three initiation factors and the charged initiator tRNA may bind the 30S subunit before or after this binds to the mRNA.<sup>3</sup> After forming the 30S mRNA complex, the 50S subunit binds and translation can start. The time for these reactions is likely to be short, but because our method determines the time for bringing the methionine encoded by codon 3 in the gene into the completed peptide, any time spent in the initiation process does not influence the incorporation curve (see Methods). The model operates with a time  $\tau$ , where the ribosome is bound to the mRNA without translating to 0.2 s, but this time might as well be included in the time between initiations.

As the last input to the simulation, we assign one of two sets of translation times to the individual codons. In the first set, all codons are translated at 13.25 codons/s, which gives the observed translation time for the *lacZ* mRNA<sup>3</sup> (see later). In the second set, we assign one of three values (A, B, and C) for each codon-specific translation rate. As explained and justified in the legend to Table 1, the A rate is approximately threefold higher than the B rate, which, again, is approximately threefold higher than the C rate. The A, B, and C rates correspond to high, medium, and small codon adaptation index

values for *Escherichia coli*, respectively.<sup>18</sup> The A, B, and C rates are modeling parameters and varied in our curve-fitting procedure (see later). In the discussion, we analyze the robustness of the conclusions with respect to the assignment of model parameters.

The experimental data used in our modeling are shown in Fig. 3a. The incorporated radioactivity is normalized to the 23 methionine residues in the *lacZ* gene. They are obtained as the average of the data of Sørensen and Pedersen<sup>2</sup> and an independent unpublished measurement. The parental plasmid pMAS23 and the mRNAs with inserts containing many GAA or GAG codons are briefly described in Methods.

The radioactivity incorporated into full-length protein increases monotonously with time in these experiments. As shown in Fig. 3a, the incorporation curve for the pMAS24GAG mRNA with the short insert is displaced with a constant time from the parental pMAS23 curve early in the experiment. This interval is the time for translating the insert, and the early displacement directly reflects that the insert is placed close to the 3' end of the mRNA. For the mRNA with the tandem insert, the incorporation curve is progressively delayed relative to the two

other curves. This more than additive effect of the long insert is explained in the simplest manner by extended queuing from the long, slow insert, where ribosomes that were most 5' on the mRNA at the time of the radioactive pulse not only have to translate the insert but also have to wait for the ribosomes in front to pass the insert. This observation of a queue for the long insert in pMAS48GAG without a noticeable queue for the short insert in pMAS24GAG turned out to constrain the model parameters describing the translation process and give a glimpse into the noise associated to the translating ribosome.

**Results from the modeling**

Figure 3 shows several examples of the model simulation, representing the best fit to the experimental data as well as a number of variants that illustrate the most important parameters. The best fit was selected from combinations of the following values of A, B, and C rates: A: 20, 25, 30, and 35 codons/s; B: 6, 7, 8, and 9 codons/s; and C: 3, 3.5, 4, and 4.5 codons/s.

To choose the parameter sets that best fit all the data, we calculate the variance between each simulated curve and the 13 experimental data points from the rising part of the experimental curve for pMAS23, pMAS24GAG, and pMAS48GAG. Then, we chose the parameter sets that minimize the variance for the data sets.

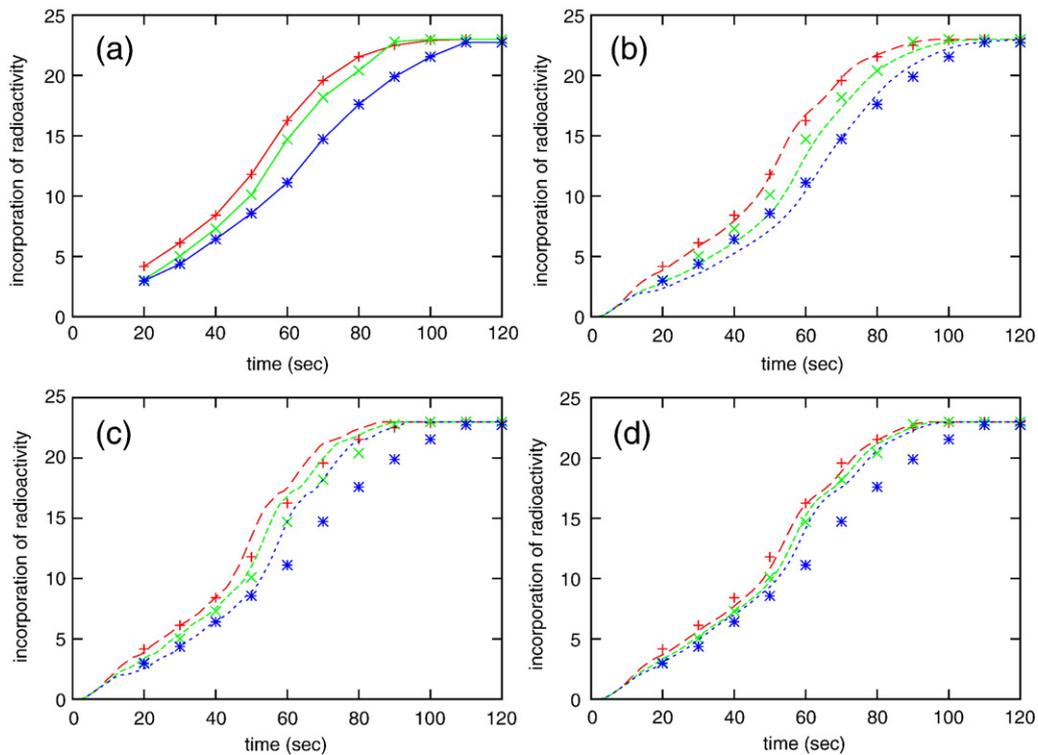
Figure 3b shows the modeled curves with the very best fit: the A, B, and C rates being 35, 8, and 4.5/s, respectively, and the on-rate being 0.9/s. Table 2 gives several sets of A, B, and C rates that almost equally well reproduce the experimental data. In Fig. 3c, the simulation with the same parameter as Fig. 3b but with deterministic translation is shown, where a ribosome waits exactly for the duration 1/R<sub>x</sub> before it moves to the next codon (details are given in Methods). In Fig. 3c, all curves rise faster than Fig. 3b, and the queue formation in pMAS48GAG is not reproduced. The stochastic translation is needed in order to reproduce the divergence of the latter part of the pMAS24GAG and pMAS48GAG curves. Finally, Fig. 3d demonstrates the necessity of having the A, B, and C rates different. The figure shows a fully stochastic simulation with the on-rate of 0.9/s as Fig. 3b, but the A, B, and C rates are all 13.25/s. This can reproduce the data for pMAS23 but not for the mRNAs with inserts.

Finally, Fig. 4 shows a space-time plot given by the simulation. Here, each line depicts the movement of a ribosome down the mRNA. Where the lines contact each other, this indicates collision between ribosomes. Figure 4 illustrates that the short insert only rarely provokes ribosome queuing when translating mRNA an average 30 times, whereas the tandem insert regularly generates extensive queuing. Figure 4d shows the extensive queuing arising if the pMAS48GAG mRNA had been stable. The space-time plots also showed that collisions normally are resolved quickly whereas

**Table 1.** The standard genetic code table with our tentative assignment of the codon-specific translation rates for *E. coli* growing in glycerol minimal medium

TTT	B	TCT	A	TAT	B	TGT	B
TTC	A	TCC	A	TAC	A	TGC	A
TTA	B	TCA	B	TAA	–	TGA	–
TTG	B	TCG	B	TAG	–	TGG	A
CTT	B	CCT	C	CAT	B	CGT	A
CTC	B	CCC	C	CAC	A	CGC	A
CTA	B	CCA	C	CAA	B	CGA	C (4.2)
CTG	A	CCG	B (7.0)	CAG	A	CGG	C
ATT	A	ACT	A	AAT	B	AGT	B
ATC	A	ACC	A	AAC	A	AGC	B
ATA	C	ACA	B	AAA	A	AGA	C
ATG	A	ACG	B	AAG	B	AGG	C
GTT	A	GCT	A	GAT	B	GGT	A
GTC	B	GCC	B	GAC	A	GGC	A
GTA	A	GCA	A	GAA	A (35.0)	GGA	C
GTG	A	GCG	B	GAG	B (7.5)	GGG	B

To have only three values for the codon-specific translation rates is a simplification and the values we use are tentative, but based on the following reasoning: Four average rates were measured previously: one fast (GAA 22/s), two medium (GAG 6.4/s and CCG 5.8/s), and one slow (CGA 4.2/s).<sup>2</sup> From this study, one can extrapolate several likely conclusions: (1) for each amino acid, the codon that is predominantly used in the genes encoding ribosomal proteins is assigned the A rate. One exception is the preferred CCG codon that was measured to have a B rate. (2) Other codons translated by the same tRNA are translated threefold slower and is assigned a B rate. The B rate is also assumed for codons that are used with medium frequency in the ribosomal-type genes. (3) In some cases, as, for instance, ACT and ACC, the frequency of use in the genes for abundant proteins is high and similar; hence, both are assigned the A rate. (4) The C rate is assigned to the remaining three nonpreferred proline codons, expected to be slower than CCG, and to the rarely used codons ATA, CGA, CCG, AGA, AGG, and GGA. The values for the A, B, and C rates can be optimized independently in the applet. The rates for translating the codons CCG, GAA, GAG, and CGA approximately corrected for collisions are shown in parentheses.



**Fig. 3.** (a) The experimental data from Sørensen and Pedersen<sup>2</sup>, showing incorporation into  $\beta$ -galactosidase encoded by pMAS23 (+, red), pMAS24GAG (x, green), and pMAS48GAG (\*, blue). (b) Best fit to the data. Dotted lines are the simulated curves for the stochastic model and have the same color code as in (a). (c) Deterministic simulation with parameters as in (b). (d) Stochastic simulation with uniform codon rates of  $A = B = C = 13.25$  codons/s. For (b), (c), and (d), the data points are as in (a).

queues, once formed, often takes a long time to untangle.

The higher density of ribosomes in the beginning of the mRNA is seen in all panels. This is because the codons we have designated B and C are predominantly found here.<sup>19</sup> We propose two arguments for the relatively high abundance of the rare and slow codons in the beginning of genes: (1) The occlusion time will respond to changes in substrate level if the rare codons, for instance, GAG, are unsaturated whereas the fast translated codon GAA is saturated (GAA and GAG are read by the same tRNA but are translated with A and B rates, respectively). This allows negative feedback to fine-tune the initiation rate. (2) Slowly translated codons in the beginning of the mRNA will increase the average ribosome distance in the later part of the gene and thereby minimize queues there. The codon distribution

found<sup>19</sup> can therefore be thought of as a way to increase the ribosome efficiency.

To show that the enrichment for “B and C” codons in the mRNA is highly selected for, we determined the frequency of synonymous codon changes in the early part of five highly expressed genes in species closely related to *E. coli* and compared it to the frequency of such changes in the distal part of the same genes. This analysis showed that the selection pressure for a certain codon usage is higher in the beginning of genes compared to the distal parts (Table 3). Treating the frequencies for synonymous codon changes in each end of these genes as coupled variables in a *t* test gave a less than 0.2% probability that this difference in the frequency of synonymous codon replacements is a coincidence.

### Advantage of unstable mRNAs for translation efficiency

As ribosomes are loaded on the mRNA with relatively short time intervals, the stochastic translation relatively fast causes ribosome–ribosome collisions. As each collision delays the lagging ribosome, there is a self-perpetuating tendency for a pileup of ribosomes on slow parts of the mRNA. Even when there are too few collisions to cause long queues, we find that collisions will occur, and will occur more frequently as the mRNA becomes older. Our model allows us to calculate the average translation time of

**Table 2.** Parameter sets that reproduce data almost equally well as the best parameter set:  $A = 35$ ,  $B = 8$ , and  $C = 4.5$  codons/s,  $p = 30$ , and an on-rate of 0.9/s

A rate (/s)	B rate (/s)	C rate (/s)	On-rate (/s)
35	9	3.5	1.1 to 0.9
35	9	3	0.8
35	8	4.5	1.1 to 0.8
35	8	4	1.0 to 0.8
30	9	4	1.1 to 1.0
30	9	3.5	1.0 to 0.8

**Table 3.** Comparison of the conservation of synonymous codons in the 5' and 3' ends of mRNAs

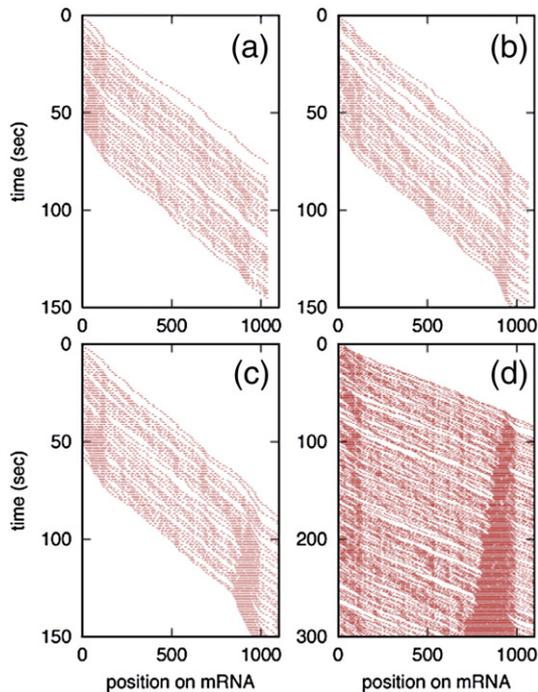
Percentage of synonymous codons replaced in segment	<i>rpo C 5'</i>	<i>rpoC 3'</i>	<i>tufB 5'</i>	<i>tufB 3'</i>	<i>tsf 5'</i>	<i>tsf 3'</i>	<i>rpsL 5'</i>	<i>rpsL 3'</i>	<i>rpsG 5'</i>	<i>rpsG 3'</i>
<i>E. coli 536</i>	0	2	0	0	0	0	0	4	0	0
<i>Shigella dysenteriae</i>	0	0	0	0	0	2	0	0	0	0
<i>Salmonella paratyphi</i>	0	21	7	11	11	23	0	10	3	6
<i>Erwinia carotovora</i>	10	31	5	17	15	23	20	31	15	13
<i>Yersinia pestis</i>	13	48	26	42	27	29	29	26	26	21

The sequences encoding five randomly chosen, highly expressed genes; two ribosomal proteins; two elongation factors; and one RNAP subunit were obtained from The Institute of Genomic Research or Genbank databases. The codons in the 5' end and the 3' end in each gene (40 and 60 codons, respectively) were compared to the corresponding *E. coli* MG1655 sequence. The table lists the percentage of synonymous codon replacements in these gene segments. The database has one *Erwinia*, five *Escherichia*, four *Salmonella*, six *Shigella*, and six *Yersinia* species or variants. The codons used within each of these groups were almost identical and only one example is listed for each family. *E. coli* 536 was the only one in the *E. coli* set that differed from MG1655. Codon changes to nonsynonymous codons are not listed.

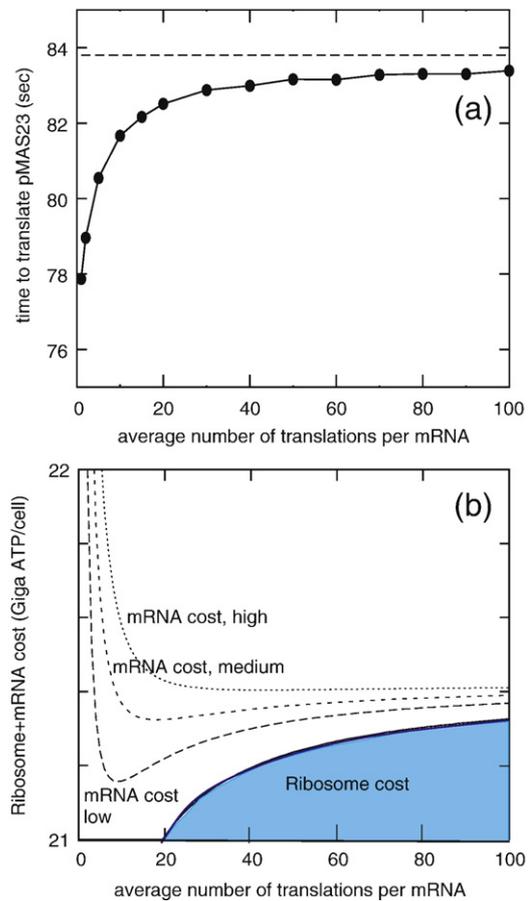
the mRNA as a function of the average number of translations before it is degraded. Figure 5a shows how the *lacZ* mRNA translation time is predicted to increase with the average number of translations of the mRNA. Already at 30 translations per mRNA, the stochastic effects have increased the average translation time by 5% compared to an mRNA that was only translated once. Figure 5a also shows that a stable mRNA would be translated 1.7% slower than the present *lacZ* mRNA translated on average 30 times.

This suggested to us that bacteria might have evolved unstable mRNAs as yet another way to

untangle and minimize ribosome queues. We now compare the cost associated to mRNA production and recycling against the cost of making the ribosomes that would compensate for the ones that waste their time by collisions and the additional



**Fig. 4.** Space-time plot of ribosome traffic as simulated with parameters from Fig. 3b. Each red dotted line refers to the movement of a single ribosome along the DNA. When ribosomes slow down, for example, when queues are developing, one sees increased density of ribosomes. Translation of (a) pMAS23, (b) pMAS24GAG, and (c) and (d) pMAS48GAG. In (a), (b), and (c), we translate mRNAs exactly 30 times. In (d), we show how a queue develops on a stable mRNA with the long, slow insert.



**Fig. 5.** (a) Average translation time and (b) metabolic cost of translation as a function of average number of translations,  $p$ , per mRNA. The metabolic cost is counted by ATP molecules needed to adjust the transcription and translation machinery such that overall protein production is fixed to the same value independent of  $p$  (see Methods). The solid curve is the cost of all ribosomes as a function of  $p$ . The three dashed curves represent the total cost, including an mRNA processing cost of 140, 210, and 280 million ATPs per cell, respectively, at  $p=30$ .

RNA polymerases (RNAPs) needed if mRNAs were unstable. In Fig. 5b, we show that total metabolic cost depends on the average translation number,  $p$ , reaching a minimum in the range  $p=10-30$ . The calculations are detailed in Methods. Obviously, this optimum depends on the ribosome initiation frequency and shifts towards more stable mRNAs for lower initiation rates. As have been realized for a long time,<sup>20</sup> unstable mRNAs also enable the cell to have a timely and efficient regulation in response to changing growth conditions. Our study shows that unstable mRNAs, in addition, will be cost-efficient because of the saved translation time. These calculations are performed on data from cells growing in glycerol minimal medium. At faster growth rates, the density of ribosomes on the mRNA and the on-rates for initiation are expected to be higher<sup>17</sup> and the advantage of having unstable mRNAs may therefore increase with more rapid growth.

## Discussion

We have here modeled the translation process in a living bacterial cell and, thanks to the precise input, succeeded in finding a narrow range for the kinetic parameters that reproduce the experimental data. The fortunate experimental finding that one insert did not provoke ribosome queue formation while the tandem insert did restricted the kinetic parameters. We conclude the following:

- (1) The experimental data for translating the *lacZ* mRNA *in vivo* is reproduced by a local model taking into account only stochasticity and variations in individual codon translocation speed, without any need for effects associated to mRNA secondary or tertiary structures. This does not exclude that special mRNA structures have evolved to inhibit the progress of the ribosome in, for example, the pseudo-knot structures essential for programmed frameshifting in a widespread but rare number of cases.<sup>21</sup>
- (2) The major time-limiting step in translation is stochastic with the broadest possible time distribution, the Poisson distribution. Because the two codons GAA and GAG, read by the same tRNA, have different translation rates<sup>2</sup> and because a change in the modification status of wobble base in the anti-codon loop of this tRNA alters the translation rate dramatically,<sup>22</sup> we find it likely that this rate-limiting step is bringing the cognate ternary complex into the A site of the ribosome.
- (3) Neither a fully deterministic nor a uniform stochastic mechanism can reproduce the experimental data.
- (4) In contrast, values spanning 10-fold between the codon-specific rates as the tentative values in Table 1 are needed to give the stochastic noise that will give ribosome queue formation

on the longest insert but not on the insert with the same sequence but with half the length.

- (5) The on-rate of initiating on the *lacZ* RBS is close to 1/s. Including an occlusion time estimated to 1 s and 0.2 s as the time from binding of the 30S ribosome to the formation of the first peptide bond, the resulting initiation frequency on the *lacZ* mRNA becomes close to 1 per 2.3 s, in reasonable agreement with the previous determinations.<sup>2,11</sup>

Changing the model parameters (the time for forming the 70S translating ribosome in the interval 0.1 to 0.3 s, the occlusion length to 10 or 12 codons, or decreasing the average number of translations per mRNA to 20) does not change the above conclusions or the finding that unstable mRNAs are more economical to translate. However, minor quantitative changes are seen. If, for instance, the average number of translations of each mRNA is decreased to 20, the best fitting on-rate will be larger and close to 1.2/s.

While our model fulfils the overall feature of increased queuing on the older mRNAs, it does not fully describe the extensive queuing in the case of the long insert in pMAS48GAG. To improve the fit, we have explored a number of variations in the model, as well as in the parameters. We have increased the average number of translations per mRNA to 60, varied on-rate rates, and changed codon translation rates. In all cases, the change affects all three constructs proportionally and does not address the difference in queuing for the long compared to the short, slow insert. The only way that we can better reproduce the additional queuing in the long insert is by assuming that the tRNA<sub>Glu</sub> becomes more rate limiting for GAG translation in the case of the long insert. This assumption is not in agreement with the finding that rare codons are translated with a rate that is independent of the expression yield.<sup>3</sup> However, the effect might be small and not noticed in the previous study or the inserts used here use many identical codons. We estimate the use of Glu-tRNA to be increased by approximately 15% under full induction of pMAS48GAG. The drain on glu-tRNA<sub>Glu</sub> is smaller when inducing pMAS48GAA because the tRNA spends less time on the ribosome in this case. Assuming a varying specific effect for GAG translation, the model can now reproduce the observed queuing difference between pMAS24GAG and pMAS48GAG, provided that we set the on-rate to be at the threshold for queuing.

With respect to the assigned values for the codon-specific translation rates given in Table 1, they, in addition to reproducing the data of Fig. 3a, also reproduce the 35% increased rate of translating the *tuf*, *tsf*, *fus*, and *rpsA* mRNA with the ribosomal protein-type codon usage compared to the *lacI* and *bla* mRNAs, which have a broader codon usage.<sup>1</sup> The used A, B, and C rates also reproduce the observed translation times of the five additional inserts in Sørensen and Pedersen<sup>2</sup> and not modeled here.

Our modeling has drawn attention to ribosome collisions and shown them to be frequent and to affect the kinetics of translation significantly. Compared to the measured average rates, the fast rates must especially be corrected upwards due to the effect of collisions because a ribosome on a fast codon is particularly likely to collide with a preceding ribosome when this translates a slow codon. The fastest individual codon translation rate ever measured was 47 codons/s for translating codon GAA in the *trmE* strain, deficient in the modification of the anti-codon loop of tRNA<sub>Glu</sub>.<sup>22</sup> Also adjusting this rate upwards as a consequence of collisions, the estimate of the translation  $k_{\text{cat}}$  value from *in vivo* measurements is now in total agreement with the  $k_{\text{cat}}$  value of 50–100/s estimated from *in vitro* experiments.<sup>23</sup>

With the use of the codon translation rates from Table 1 to estimate the occlusion times for natural RBSs of mRNAs, the analysis of 100 such sites indicated that the occlusion time is likely to range from 0.3 to 1.3 s among different mRNAs. With on-rates of 1 every few seconds, variances in occlusion time will modulate the expression substantially from most if not all RBSs.

Finally, our model allows us to obtain an accurate estimate on ribosome initiation rates and translation speed of individual codons, which is consistent with a *lacZ* mRNA stability that minimizes the overall metabolic cost of producing proteins. We predict that for *lacZ* mRNA, about 5% of ribosome translation time is wasted because of collisions along the mRNA. Although this waste could be minimized by engineering mRNAs with shorter lifetimes, the cost of producing more unstable mRNA will exceed the gain associated to increased ribosome efficiency. If, on the other hand, one lowers the waste associated to ribosome–ribosome collisions by decreasing ribosome initiation rate, this would increase the distance between translating ribosomes. An increase in naked mRNA along the mRNA and also between the first ribosome and the RNAP may, in turn, increase the access for degradation factors such as Rho, and degradation may be initiated more often at random positions along the mRNA. As 0.4% of all *in vivo* translations that are terminated prematurely<sup>24</sup> in itself are a large cost in terms of ribosomes that waste their time on dysfunctional mRNA, the cost of increasing this termination fraction could easily become larger than the 5% cost of ribosome–ribosome collisions. Overall, we therefore find that *E. coli* has optimized both metabolic cost and protein production time by translating each *lacZ* mRNA 30 times with 2.3-s intervals.

## Methods

### Description of the translated mRNAs

The *lacZ* genes used in this study are derived from the wild-type *lacZ* and have the following inserts: pMAS23 has the sequence CGG TCG ACC GAT inserted at codon 927 in *lacZ*; the plasmids pMAS24GAA, pMAS24GAG,

pMAS48GAA, and pMAS48GAG have the sequences CGA(GAA)<sub>8</sub>CCG and CGA(GAG)<sub>8</sub>CCG inserted three and six times, respectively, in the AccI restriction enzyme site in the CGG TCG ACC GAT sequence in pMAS23.<sup>2</sup>

### Remarks on the experimental method

The method determines the time for bringing the radioactivity incorporated into the most N-terminal methionine residue into the full-length protein. The initiating methionine is split in the case of  $\beta$ -galactosidase, which happens before the protein is completed,<sup>25</sup> and the most N-terminal methionine residue is therefore the methionine encoded by codon 3 in *lacZ*. This means that the time the ribosome spends in the initiation process is not included. Similarly, the time the ribosome takes for terminating translation does not influence our measurements, as long as it is not so slow as to provoke queue formation. This is because the bond between the tRNA and the nascent peptide is hydrolyzed during our sample preparation, which includes boiling with SDS. The full-length position on the gel is therefore achieved before the termination factor has released the peptide. This also means that the time for recycling mRNA, ribosomes, initiation and termination factors, and tRNA does not influence the incorporation kinetics. With respect to equilibrating the radioactivity into the pool of methionine and met-tRNA<sub>Met</sub>, this time has been estimated to be shorter than 5 s at 25 °C, that is, shorter than 2 s at the 37 °C used here.<sup>26</sup> This time for pool heating is therefore close to negligible in the 10-s pulse we use. Furthermore, because we use a mixture of the two cultures, the small amount of time for pool heating affects the incorporation curves identically for the strains to be compared in an experiment.

### Simulation methods

#### Stochastic simulation

We performed Monte Carlo simulations with a constant time step  $dt$ . We consider a system that consists of multiple mRNAs, which are constantly reshuffled as new mRNAs are transcribed and old ones are decaying. New mRNAs are transcribed stochastically with a rate  $R_{\text{birth}}$ . Ribosomes can bind to existing mRNAs, as long as the binding site is accessible. At each time step, all ribosomes on all mRNAs proceed to the next codon with a probability  $R_i dt$  under the condition that the distance to the preceding ribosome is larger than the occluding distance  $d$ . When a ribosome reaches the 3' end, it terminates the translation with a rate  $K_t$ . The decay of mRNA with rate  $R_{\text{stop}}$  is simulated by stopping translation initiation of each mRNA with probability  $R_{\text{stop}} dt$ . For each of these stopped mRNAs, the already initiated ribosomes continue translation until terminating at the 3' end. When all the ribosomes finish the translation, the mRNA is removed from the system.

The dynamics of the model is implemented in time steps  $dt$ , which are sufficiently short that even the events with the largest rate occur with probability  $\ll 1$  within  $dt$ . Each time step consists of the following actions, taking care of the population of mRNAs and the movement of ribosome on each mRNA:

- (a) Add a new mRNA to the system with a probability  $R_{\text{birth}} dt$ .

- (b) Perform the following ribosome dynamics (b-1 to b-3) for all the mRNAs in the system:
- (b-1) If the first  $d$  codons are not occupied by ribosomes, a new ribosome can bind to the first codon with probability  $K_s dt$ . If a new ribosome binds to the first codon, the time counter  $w$  for that ribosome is set to zero. Also, if the first  $d$  codons are not occupied by ribosomes, translation initiation is stopped with a probability  $R_{\text{stop}} dt$  and the mRNA starts degrading from the 5' end.
  - (b-2) If there is a ribosome on the codon 1 and  $w$  for that ribosome is smaller than the assigned time ( $\tau = 0.2$  s) it takes to initiate transcription, then there is no movement. For a ribosome at codon 1 with  $w > \tau$ , and for any of the other ribosome at positions  $x < L$ , there is a possibility for movement. For each such ribosome, one chooses a random number  $r$  equally distributed in the interval  $[0:1]$ , and if  $r < R_x dt$ , the ribosome attempts to move one step forward from its current position  $x$ . This movement,  $x \rightarrow x+1$ , is performed if and only if there are no other ribosomes in  $[x, x+d]$ . In case movement is occluded, then the ribosome stays where it is. An eventual ribosome at position  $x=L$  cannot move but will finish translation with probability  $K_t dt$ , and in that case, it will leave the mRNA.
  - (b-3) Add  $dt$  to  $w$ , the duration that a ribosome spent at the first codon.
- (c) mRNAs where translation initiation has been stopped and have no ribosomes are removed from the system.

After procedures (a)–(c), the time  $t$  proceeds by the time step  $dt$ . Procedures (a)–(c) are repeated to see the time evolution of the system.

#### Parameters used in the simulation

The average number of translations per mRNA,  $p$ , is determined by  $R_{\text{birth}}/R_{\text{stop}}$ , which is adjusted such that each mRNA is translated 30 times unless otherwise noted. We use the time step  $dt = 1/300$  s for stochastic simulation, while for deterministic simulation,  $dt$  is set so that  $1/(R_x dt)$  becomes an integer to minimize the error in the waiting time. In most of the simulation, we set  $R_{\text{birth}} = 12$ , which gives about 2000 mRNAs in the system at a given time in the steady state when  $p = 30$ .

#### Deterministic simulation

In the deterministic simulations, the rates  $R_x$  and  $K_t$  in step (b) are not used to give probabilities to proceed to the next codon but each ribosome waits for the duration  $1/R_x$  ( $1/K_t$ ) before it proceeds to the next codon (it terminates the translation). The rest of the simulation has been performed in the same way as the stochastic simulation.

#### Reproduction of the incorporation curves

The incorporation curve is obtained as follows: First, the ensemble of the mRNAs is prepared by performing the simulation for a long enough time so that the system reaches the steady state. Then, the time  $t$  is reset to zero. From  $t = 0$  s to  $t = 10$  s, the pulse of radioactive methionine is added.

When a ribosome passes a methionine site during this pulse, the ribosome carries one radioactive methionine, and the number of radioactive methionines is added up as the ribosome passes more methionine sites during the pulse. When a ribosome finishes the translation, the number of radioactive methionines that the ribosome is carrying is counted as the radioactive signal. The accumulated counts until time  $t$ ,  $C(t)$ , gives the amount of incorporation, which reaches a constant after a long enough time. The  $C(t)$  is normalized so that this constant becomes 23, the number of methionine codons on the mRNA. This normalized incorporation curve  $C(t)$  is represented by broken lines in Fig. 3.

#### Metabolic costs

The metabolic cost estimate is based on cells growing in a minimal medium with a doubling time of 60 min and an average mRNA half-life of 1.5 min. Under these conditions, a 1-ml culture at a density  $OD_{450}$  of 1 ( $\sim 4 \times 10^8$  cells) contains 0.5  $\mu\text{g}$  mRNA ( $\sim 2.2$  million nucleotides/cell), 1.1  $\mu\text{g}$  of RNAP, and 35  $\mu\text{g}$  of ribosomes.<sup>17,27</sup> When calculating the metabolic cost in terms of number of ATPs per cell, we considered the fact that production of 1 g of cellular material requires approximately 2 g of glucose. One glucose molecule weighs 180 Da and supplies an energy equivalent of 36 ATP molecules, which, individually, is assumed to be enough for recycling between one and two nucleic acid bonds in mRNA. First, we estimate the mRNA processing cost, including both the mRNA recycling cost and the production cost of the fraction of RNAP that is actively engaged in transcribing mRNA. At a number of proteins per mRNA ( $p = 30$ ), the cost of recycling the 2.2 million mRNA nucleotides 30 times within one cell generation is between 70 and 140 million ATPs per cell. The cost of producing the 10% of the RNAP that is actively transcribing mRNA<sup>17</sup> is an additional 70 million ATPs, a number that is uncertain because a fraction of nontranscribing RNAPs may be needed to secure the activity of the active ones. Accordingly, we estimate RNAP processing cost to be between 70 and 140 million ATPs per cell, which results in a total mRNA processing cost of 140–280 million ATPs/cell. When  $p$  is decreased, the cost in mRNA processing changes as  $1/p$ , with the cost = 140–280 million ATPs/cell for  $p = 30$ . Thus, the mRNA translation cost increases up to 30-fold as  $p$  approaches 1. Ribosomes represent a metabolic cost of 21 billion ATPs per wild-type cell, which presumably represents a translation speed that is obtained at  $p = 30$ . Compared to  $p = 30$ , Fig. 5a shows that the translation speed changes from being 5% faster for  $p = 1$  to 1.7% slower for  $p \gg 100$ . The corresponding variation in metabolic cost for producing the needed ribosomes therefore changes between 20 and 21.3 billion ATPs/cell as  $p$  is varied from  $p = 1$  to stable mRNAs. Although the ribosome cost only varies by a few percentages, the fact that ribosomes are expensive in absolute terms makes it possible to save metabolic cost by reducing mRNA stability. This is illustrated in Fig. 5b.

#### Acknowledgements

We thank Tom Silhavy, Princeton University, for comments on the manuscript. The Danish National Research Foundation funded this work. N.M. thanks the Yamada Science Foundation for supporting her stay at the Niels Bohr Institute.

## References

1. Pedersen, S. (1984). *Escherichia coli* ribosomes translate in vivo with variable rate. *EMBO J.* **3**, 2895–2898.
2. Sørensen, M. A. & Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*: the two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280.
3. Sørensen, M. A., Kurland, C. G. & Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* **207**, 365–377.
4. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. & Dennis, P. P. (1979). Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **76**, 1697–1701.
5. MacDonald, C. T., Gibbs, J. H. & Pipkin, A. C. (1968). Kinetics of biopolymerization on nucleic acid templates. *Biopolymers*, **6**, 1–25.
6. Bergmann, J. E. & Lodish, H. F. (1979). A kinetic model of protein synthesis. Application to hemoglobin synthesis and translational control. *J. Biol. Chem.* **254**, 11927–11937.
7. von Heijne, G., Nilsson, L. & Blomberg, C. (1978). Models for mRNA translation: theory versus experiment. *Eur. J. Biochem.* **92**, 397–402.
8. Shaw, L. B., Zia, R. K. P. & Lee, K. H. (2003). Totally asymmetric exclusion process with extended objects: a model for protein synthesis. *Phys. Rev. E*, **68**, 021910.
9. Gilchrist, M. A. & Wagner, A. (2006). A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *J. Theor. Biol.* **239**, 417–434.
10. Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D. & Gold, L. (1992). Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **9**, 1219–1229.
11. Kennell, D. & Riezman, H. (1977). Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J. Mol. Biol.* **114**, 1–21.
12. Jaquet, M. & Kepes, A. (1971). Initiation, elongation and inactivation of lac messenger RNA in *Escherichia coli* studied by measurement of its  $\beta$ -galactosidase synthesizing capacity in vivo. *J. Mol. Biol.* **60**, 453–472.
13. Steitz, J. A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*, **224**, 957–964.
14. Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu. Rev. Biochem.* **57**, 199–233.
15. Baker, R. F. & Yanofsky, C. (1970). Transcription initiation frequency for the tryptophan operon of *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol.* **35**, 467–470.
16. Ullmann, A., Jacob, F. & Monod, J. (1968). On the subunit structure of wild-type versus complemented beta-galactosidase of *Escherichia coli*. *J. Mol. Biol.* **32**, 1–13.
17. Bremer, H. & Dennis, P. P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. (Neidhardt, F. C., Curtiss, I. I., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. & Umberger, H. E., eds), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. vol. 2, pp. 1553–1569, ASM Press, Washington, DC.
18. Sharp, P. M. & Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295.
19. Bulmer, M. (1988). Codon usage and intragenic position. *J. Theor. Biol.* **133**, 67–71.
20. Monod, J., Pappenheimer, A. M., Jr & Cohen-Bazire, G. (1952). The kinetics of the biosynthesis of  $\beta$ -galactosidase in *Escherichia coli* as a function of growth. *Biochim. Biophys. Acta*, **9**, 648–660.
21. Kontos, H., Naphthine, S. & Brierley, I. (2001). Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell. Biol.* **21**, 8657–8670.
22. Krüger, M. K., Pedersen, S., Hagervall, T. G. & Sørensen, M. A. (1998). The modification of the wobble base of tRNA<sup>Glu</sup> modulates the translation rate of glutamic acid codons in vivo. *J. Mol. Biol.* **284**, 621–631.
23. Lovmar, M. & Ehrenberg, M. (2006). Rate, accuracy and cost of ribosomes in bacterial cells. *Biochimie*, **88**, 951–961.
24. Moore, S. D. & Sauer, R. T. (2005). Ribosome rescue: tmRNA tagging activity and capacity in *Escherichia coli*. *Mol. Microbiol.* **58**, 456–466.
25. Engbæk, F., Kjeldgaard, N. O. & Maaløe, O. (1973). Chain growth rate of  $\beta$ -galactosidase during exponential growth and amino acid starvation. *J. Mol. Biol.* **75**, 109–118.
26. Pedersen, S. (1984). *Escherichia coli* individual genes are translated with different rates in vivo. In Gene Expression, A Benzon Symp 19, (Clark, B. F. C. & Petersen, H. U., eds), pp. 101–107, Munksgaard, Copenhagen, Denmark.
27. Maaløe, O. & Kjeldgaard, N. O. (1966). *Macromolecular Synthesis*. Benjamin Press, New York.