

Tuning of Recombinant Protein Expression in *Escherichia coli* by Manipulating Transcription, Translation Initiation Rates, and Incorporation of Noncanonical Amino Acids

Orr Schlesinger,^{†,‡,§} Yonatan Chemla,^{†,‡,§} Mathias Heltberg,^{‡,§} Eden Ozer,[†] Ryan Marshall,[§] Vincent Noireaux,[§] Mogens Høgh Jensen,^{*,‡} and Lital Alfonta^{*,†}

[†]Department of Life Sciences and Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 8410501, Israel

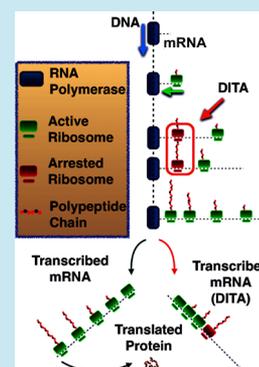
[‡]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

[§]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, United States

Supporting Information

ABSTRACT: Protein synthesis in cells has been thoroughly investigated and characterized over the past 60 years. However, some fundamental issues remain unresolved, including the reasons for genetic code redundancy and codon bias. In this study, we changed the kinetics of the *Escherichia coli* transcription and translation processes by mutating the promoter and ribosome binding domains and by using genetic code expansion. The results expose a counterintuitive phenomenon, whereby an increase in the initiation rates of transcription and translation lead to a decrease in protein expression. This effect can be rescued by introducing slow translating codons into the beginning of the gene, by shortening gene length or by reducing initiation rates. On the basis of the results, we developed a biophysical model, which suggests that the density of co-transcriptional-translation plays a role in bacterial protein synthesis. These findings indicate how cells use codon bias to tune translation speed and protein synthesis.

KEYWORDS: protein translation initiation, transcription initiation, genetic code expansion, rates of translation, codon bias



Protein synthesis, one of the most important and complex functions of living cells, is controlled by several mechanisms. Every stage in the process, from DNA transcription to protein folding dynamics, is tightly regulated to ensure that proteins are produced in required amounts, at the correct times and with minimal waste of energy and resources.¹ In bacteria, the transcription of DNA to mRNA and the subsequent translation into a polypeptide chain are coupled in time and space.^{2,3} The two processes occur simultaneously, which creates a high molecular density area populated with all the components required for protein synthesis. For the dynamics of transcription and translation, such molecular crowding in the cytoplasm plays an important role by stabilizing protein–protein interactions and by controlling the diffusion rates of the components involved in protein synthesis.^{4,5} The molecular densities of RNA polymerases on DNA and of ribosomes on mRNA are known to depend on the transcription and translation initiation rates, which, in turn, are determined by the strengths of the promoter and of the ribosome binding site (RBS). For example, it was shown that the use of a strong RBS with a high initiation rate to overexpress proteins can lead to ribosome collisions and queuing along individual mRNA strands. These queues can generate interference between adjacent translating ribosomes, significantly lowering the yields and efficiency of protein expression.^{6,7} The nature of possible

interactions that may occur between ribosomes on adjacent mRNA strands, however, is not clear.

The kinetics of translation also depend on the codon usage of the encoded gene, which is manifested by its effects on the elongation rate of the growing polypeptide chain.^{8,9} Exploited across species to control translation rates and the ribosome queues along mRNA strands, codon bias is used to optimize protein synthesis and folding. Depending on the elongation rates they dictate, codons can be divided into different rate classes. Slower codons are found to be more favorably encoded for in the first 30–50 codons of the mRNA, thus resulting in ribosome crowding near the translation initiation site. Downstream codons, however, are found to be optimized for fast elongation rates.^{10–12} These findings give rise to several questions: Why is translation that occurs close to the translation initiation site slow? Is this slow translation rate related to the density of the molecular environment in the vicinity of the cotranscriptional–translation event?

Although normally used for applicative purposes,¹³ genetic code expansion through stop codon suppression constitutes an effective, basic research tool to shed light on these questions. One approach of genetic code expansion, the incorporation of

Received: January 17, 2017

Published: February 23, 2017

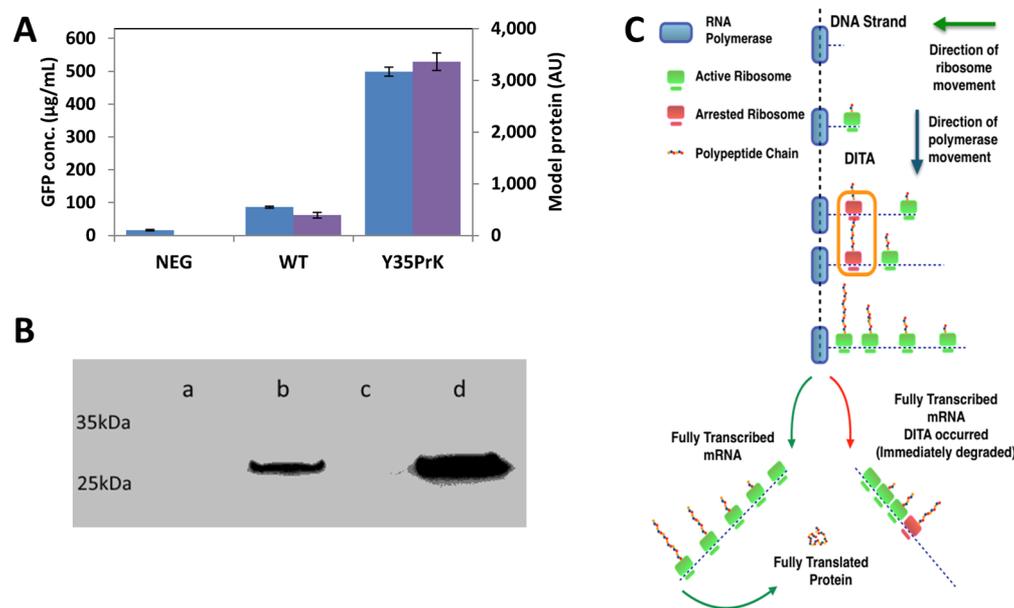


Figure 1. GFP expression using the P70a-UTR1 system. (A) Comparison of the experimental results (blue bars) with the modeled protein quantities (purple bars). (B) Western blot analysis using anti-GFP antibody of GFP expression in C321.ΔA.exp: w/o plasmids (a) pBEST-p70a-UTR1-GFP WT (b) and Y35PrK mutant in the absence and presence of PrK in the growth medium (lanes c and d, respectively). (C) Schematic presentation of the hypothesized DITA phenomenon.

noncanonical amino acids (ncAAs) into proteins, typically exploits the UAG nonsense (stop) codon, essentially transforming it into a sense codon that encodes for the incorporation of an ncAA. This recoding is facilitated by introducing into a host organism an orthogonal translation system (OTS) that comprises of an orthogonal Archaeal *o*-tRNA with an anticodon corresponding to the UAG stop codon and an orthogonal amino-acyl-tRNA synthetase (*o*-aaRS) that selectively recognizes the ncAA of choice and aminoacylates its cognate tRNA_{CUA}.¹⁴ The affinity of the *o*-tRNA to the tertiary complex of the ribosome A-site during translation, is significantly smaller than that of the native bacterial tRNA.^{15,16} This smaller affinity can be exploited to alter ribosomal traffic on the mRNA by decreasing the speed of translation along the mRNA. This approach could be efficiently realized when the OTS and the native release factor (*i.e.*, RF1) are not in direct competition for the UAG codon. That competition can be eliminated by recoding all TAG stop codons in the bacterial genome to TAA and by knocking out the RF1 gene.¹⁷

Here we use the OR2-OR1-pr-UTR1 (P70a-UTR1) expression system, based on a modified lambda PR promoter and the T7 bacteriophage RBS,¹⁸ to perform genetic code expansion. This system has the highest transcription and translation initiation rates reported for an *E. coli* element, and so far, it has been used exclusively *in vitro*. Its high initiation rates promote large and unusual ribosome crowding along the transcribing mRNA. We therefore hypothesized that in the crowded environment of a polysome, a growing polypeptide chain may interact with neighboring translational components inside the polysome in a manner that can significantly retard the process. Indeed, it was previously shown that the nascent polypeptide can regulate the translation process in the ribosome by interacting with the polypeptide exit tunnel in the ribosome.¹⁹ Such interaction may cause ribosome stalling,²⁰ translation arrest²¹ and even accelerated mRNA degradation.²² We exploited both the incorporation of ncAAs using UAG stop

codon suppression, synonymous mutations in the gene and the modular tuning of the P70a-UTR1 expression system to model and control ribosomal traffic, thus optimizing recombinant protein expression.

RESULTS

WT GFP Exhibits Smaller Expression Levels Compared to GFP with an ncAA Incorporated at Position 35.

Compared to its *in vitro* expression, the *in vivo* expression of the WT GFP (WT GFP stands for a protein without incorporated ncAAs) using the strongest *E. coli* promoter so far reported (P70a-UTR1)¹⁸ was unexpectedly weak (Figures 1A, 1B lane b). This outcome was observed not only when using a genomically recoded *E. coli* strain (C321Δ*prf1*) (GRO),¹⁷ but also with two other *E. coli* strains (*i.e.*, BL21(DE3) and DH5α). However, expression in the GRO strain of the same protein, in which a tyrosine residue at position 35 has been replaced with a nonsense stop codon (UAG), led to large and unexpected quantities of mutant GFP with the ncAA Propargyl-L-Lysine (PrK) incorporated into position 35 (Figures 1A, 1B lane d). Correct ncAA incorporation was verified by mass spectrometry (LC-ESI-MS) as well as by MS/MS analysis of peptide fragments (Figure S1).

To understand these initial observations, we first ruled out the possibility that inclusion bodies or secondary mRNA structures were the source of the divergence between the WT GFP and 35TAG GFP quantities. Cryo-electron microscopy (Cryo-EM) imaging of GFP revealed neither inclusion bodies nor any marked difference in bacterial shape compared to Cryo-EM images of bacteria without the GFP expression plasmid (Figures S2A, S2B). Moreover, there was no difference in the mRNA structure encoding for the WT GFP and the mutant GFP (Figures S2C, S2D). We used two different plasmid vectors to express the mutant Y35TAG GFP: one encoding for the mutated protein and one encoding the Pyrrolysine orthogonal translation system (Pyl-OTS), which is the machinery for the ncAA incorporation. To exclude the

possibility that the pEVOL Pyl-OTS plasmid contributed to the unusual overexpression of the mutated reporter protein, we demonstrated that pEVOL Pyl-OTS has no particular effect on the expression of WT GFP (Figure S2E). Taken together, these observations motivated our search for a more fundamental explanation related to the coupling of bacterial transcription and translation kinetics.

Density Induced Translation Arrest Model Predictions Corresponds to Counterintuitive Protein Expression Patterns. Herein, we propose a model to predict protein and mRNA levels that is based on a set of biochemical parameters combined with several assumptions. Model parameters: an increase in the RNA polymerase (RNAP) initiation rate (*i.e.*, promoter “strength”) leads to a decrease in the average distance between transcribing RNAP and *vice versa*.²³ The deterministic average distance between RNAPs, $\langle D \rangle$, is governed by eq 1 (the equation and its solutions are presented in Figures S3A, S3B):

$$\langle D \rangle = D_{\text{pol}} + \frac{R_0}{R_\alpha} \quad (1)$$

where R_α is the RNAP initiation rate and R_0 is the RNAP elongation rate anywhere on the gene, while D_{pol} is the size of the RNAP, which defines the minimal distance between polymerases. The use of the Gillespie stochastic algorithm imposed a distribution of RNAP velocities around the simplified elongation rate: R_0 . This creates a stochastic distribution of the distances between RNAPs and even creates queues of adjacent RNAPs. As the average distance between RNAPs decreases, the density of mRNAs being synthesized along the DNA strand increases and the average distance between adjacent mRNAs decreases. We named this promoter dependent mRNA density along DNA “transcriptional density” (Figure S3E).

The initiation rate of translation depends on the properties of the ribosome binding site (RBS). As the ribosomal translation initiation rate increases, the average distance between the ribosomes translating the same mRNA template becomes shorter. The average distance, $\langle d \rangle$, is governed by ribosome size d_{rib} , the initiation rate r_α and the elongation rate for each codon i given by r'_i . Considering that the time for each step is given by $\langle t \rangle = \frac{1}{r'_i}$ the average distance (in number of codons) can be simplified and expressed by eq 2 (the equation and its solutions are presented in Figures S3C, S3D):

$$\langle d \rangle = d_{\text{rib}} + N + t_n \quad (2)$$

where

$$t_n = \frac{\frac{1}{r_\alpha} - \sum_{i=12}^N \frac{1}{r'_i}}{r_{N+1}}$$

and N fulfills

$$\sum_{i=12}^N \frac{1}{r'_i} < \frac{1}{r_\alpha} < \sum_{i=12}^{N+1} \frac{1}{r'_i}$$

Note that r'_i and r_α units are [s^{-1}], whereas r_i units are [codons $\times \text{s}^{-1}$].

Another factor to include in the model is the different elongation rates of each codon in the mRNA sequence.^{24–26} On the basis of a previous model developed by Mitarai *et al.*, the entire set of bacterial codons was divided into three groups

based on translation rate—fast (A), medium (B) and slow (C)—which correspond to elongation rates of 35, 8, and 4.5 codons per second, respectively²⁵ (Figure S3F). To these canonical codons we added the new noncanonical UAG codon (in the GRO strain its only translated by the o-tRNA). The UAG codon was assigned a new translation rate category, group (D), which had a significantly lower elongation rate of 0.04 codons per second. The rate was estimated from *in vitro* experiments¹⁵ and even though this value has some uncertainty to it, it is at present our best estimate. Moreover, the model based simulated results are quite robust to large perturbations around this estimate. For example, the main observation being that by using an early TAG mutation, significantly more protein is being produced compared to WT GFP. These yields are still achieved for values of a UAG rate ranging between 0.01–0.2 codons/s. Like the case of the RNAP stochastic velocity, the ribosome also moves in a stochastic-probabilistic manner. This means that in addition to the 4 rate groups the actual ribosome velocity is governed by rate distributions for each codon around the group mean. Finally, we included “translational density”, defined as the density of ribosomes along an mRNA. The length of the growing nascent polypeptide is directly proportional to the position of the ribosome along the mRNA relative to the translation initiation site.

In bacteria, transcription and translation are coupled, *i.e.*, as soon as the RBS on the transcribed mRNA emerges from the RNAP, the ribosome binds the RBS and translation begins.² The close proximity of the two processes in time and space means that there may be interactions between them. Accordingly, we hypothesized that highly crowded conditions will promote spatial ribosome density, thus inducing translation arrest in a process that we termed “Density Induced Translation Arrest” (DITA). We propose that in cases in which the promoter and RBS initiation rates are large enough to create regions with high molecular density and in which the nascent polypeptide is long enough, the probability for DITA events increases. In the case of a DITA event, all the ribosomes upstream of the arrested ribosome stall, promoting translation termination and thus reducing the number of full-length proteins produced from crowded mRNA strands (Figure 1C).

Next, we characterized our system’s model parameters (described in detail in the methods section and listed in Table S2). The GFP gene was mapped and the codons were assigned to one of the four codon rate groups (A–D). The ribosome elongation rate is governed by each codon during translation. The average RNAP transcription rate was assumed to be constant.^{2,27} Lastly, the length of the growing nascent polypeptide could not be determined *a priori* since its folding dynamics and interactions with the ribosome are unknown to us. For this reason, we chose the simplest possible approach and we added an empirical constant of proportionality, λ , which governs the length of the polypeptide protruding from its parent ribosome (Figure S3H). This approach allowed us to predict, for a given gene, which transcription-translation instances will generate a full-length protein and, as a result, the protein production rate.

The results of the Gillespie algorithm simulation agreed with the experimental results for both WT and Y35TAG mutant GFP (Figure 1A). The model suggests that WT GFP expression levels are negligible because of the high probability for DITA occurrences when a strong promoter and RBS, such as P70a and UTR1, respectively, are used. In the case of the Y35TAG GFP mutant, the model suggests that the small-

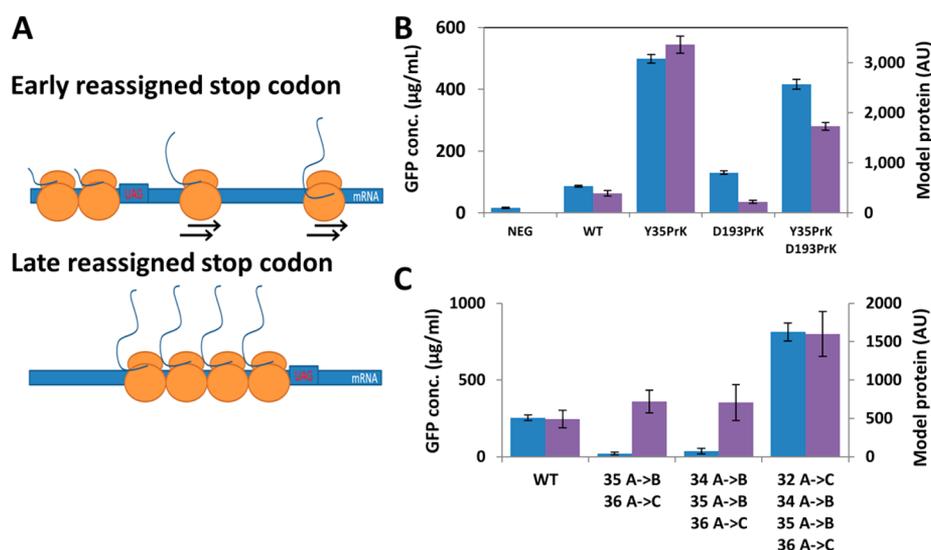


Figure 2. The “rescue effect” caused by nCAA incorporation and the associated attenuation in transcription rate is position dependent as well as “slow translating” codon dependent. (A) Schematic presentation of the ribosome queue caused by UAG codons at various positions along the mRNA. In early introduced UAG codons, the nascent polypeptide is relatively short and the ribosome traffic downstream of the codon is low, reducing the chance of spatial ribosome density. However, with later UAG codons, the nascent peptide is longer and has more chances to interact with other molecules in the polysome due to the ribosomal queue caused by the slow rate of the UAG codon. (B) Comparison of the experimental results (blue bars) with the modeled protein amounts (purple bars). (C) WT GFP expression upon introduction of synonymous mutations around position 35. Experimental results (blue bars) and modeled protein quantities (purple bars) show a significant increase in expression after the introduction of the 4th synonymous mutation.

translation-rate UAG codon (group D) inserted in this position serves as a “traffic light” that reduces ribosomal density downstream. Taken together, the reduction in translational density downstream of the UAG codon and the low probability of a DITA event result in high yields of expressed protein.

Early Reassigned Amber Stop Codon Rescues Protein Expression As Predicted by the Model. In the proposed model, we suggest that a suppressed UAG stop codon functions as a traffic light, thus its position along the mRNA is of importance. Due to its substantially slower nCAA incorporation kinetics compared to those of codons encoding for canonical amino acids, a queue of ribosomes will grow behind the reassigned stop codon. The transient stalling generated by an early UAG codon significantly reduces ribosome occupancy downstream, thereby reducing the chance of a DITA event (Figure 2A). As the translation process continues, the chance that the elongating polypeptide chain will have a DITA grows. Indeed, both our experimental results and our simulations indicated that the earlier the stop codon is introduced, the lower the chance of a DITA event. For cases in which both promoter and RBS are strong, our hypothesis predicts that the closer a UAG codon is positioned to the C-terminal, the smaller will be the protein yields in a manner similar to what is observed for the WT GFP. Indeed, the choice of a late D193TAG site in the simulation resulted in high DITA levels and small protein yields compared to those in the Y35TAG GFP mutant and protein yields equal to that of the WT protein. To test our prediction, we mutated position D193TAG in GFP. The experimental results coincided with those of the simulation, *i.e.*, low protein levels (Figure 2B). Note that D193TAG GFP is a permissive mutation site, as compared *in vitro*^{28,29} to the WT and Y35PrK mutant expression (Figure S4A). The relationship between the position of the UAG codon and the protein expression levels was simulated (Figure S4B) revealing that only the first 37 codons enable rescue of protein

levels. This result is in agreement with our experimental results and with earlier reports by Tuller *et al.* of an early slow translating “ramp” region close to the translation initiation region.^{9,10}

Next, we tested the influence of adding an early UAG codon to a mutant that already contains a late mutation (Y35TAG +D193TAG). The model predicted that the early mutation would decrease the translational density around the later UAG stop codon, thus reducing the probability of DITA and conferring a rescue mechanism on protein levels. The expression levels of the double mutant Y35TAG+D193TAG GFP, its protein expression kinetics and the final yields with the different mutants predicted *in silico* and tested *in vivo* showed high correlation and a clear rescue effect on protein expression (Figure 2B).

Observing these results, we wanted to test whether the rescue effect could be achieved with synonymous (silent) sense codon mutations. Hence, we have tried to mutate the codons around the early Y35 site in the WT GFP gene to slower synonymous codons. As an example; tyrosine 35 was mutated from TAC (group A) to TAT (group B). When tested *in silico*, we predicted that at least four slow translating codon mutations (two A → B mutations and two A → C mutations) should be introduced in order to increase protein yields (Figure 2C, purple bars). We tested our predictions experimentally and only when four slow translating codon mutations were introduced, the WT GFP expression was rescued and showed significant increase in expression levels (Figure 2C, blue bars). When two or three slow codon mutations were introduced, the expression levels of WT GFP were only basal levels in both the simulation and the experiments. These results reconfirm that translation rates are crucial for high yields of recombinant protein expression.

Slower Transcription and Translation Initiation Rates Rescue Protein Expression. The use of weaker variants of

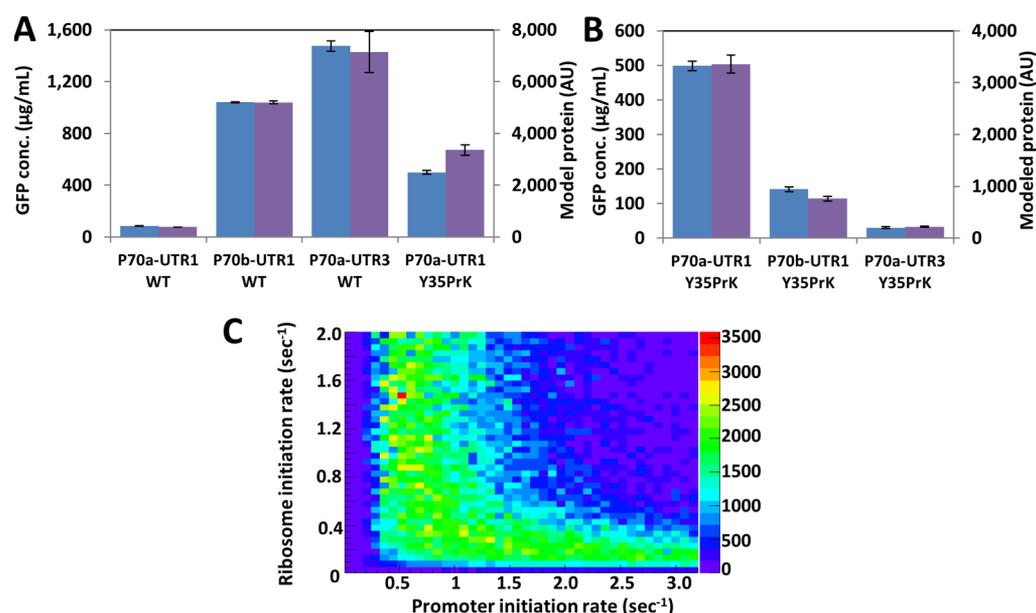


Figure 3. Smaller initial transcription and translation rates decrease the probability of DITA. (A) Comparison of the experimental results (blue bars) with the modeled protein amounts (purple bars) of WT GFP expression under the control of promoter and RBS with variable initiation rates. (B) Comparison of the expression of GFP Y35PrK mutant under the control of different control regions. Experimental and the modeled results, shown in blue and purple bars, respectively. (C) Heat map of the expected amounts of WT GFP protein using different combinations of transcription and translation initiation rates.

both promoter and RBS to increase the average distances between adjacent mRNAs and between translating ribosomes, respectively, should reduce the probability of DITA and increase protein expression. We engineered weaker variant of the P70a promoter and the UTR1 RBS by introducing point mutations into the control regions. *In vitro* transcription and translation experiments showed that the transcription initiation rate of the weaker promoter variant, P70b, was about 20 times smaller than that of the P70a promoter (Figure S5A). *In vitro* tests of the weak RBS variant UTR3 found that its translation initiation rate was 10 times smaller than that of the original UTR1 (Figure S5B). Our use of either a weaker promoter or RBS enabled us to test whether DITA is affected only by transcriptional or translational density or, as our model suggests, that both factors influence the expression density, the chances for DITA and thus, the amount of expressed protein. Intuitively, the use of weaker promoter and RBS regions is expected to result in smaller amounts of synthesized protein. However, as predicted by our hypothesis and model, the counterintuitive trend was observed, according to which the weaker the control region, the higher the protein yields. This finding is true both for the weaker promoter and RBS variant, P70b-UTR1 and P70a-UTR3, respectively (Figure 3A, purple bars). Experimental tests of this prediction showed that the weakened variants yielded up to 20 times more protein than the strong promoter-RBS construct (Figure 3A, blue bars), suggesting that DITA can be mitigated by increasing either the spacing of RNAP on DNA or ribosomal spacing on mRNA. Notably, when the same experiment was performed with the Y35TAG GFP mutant it showed the opposite trend, both experimentally as well as by simulation, where a weaker control region yielded less protein (Figure 3B). Thus, by using a simple set of mutated reporter genes and incorporation of non canonical amino acids, we showed how protein synthesis yields depend, in a counterintuitive manner, on the strengths of the regulatory elements, *i.e.*, promoter and RBS strengths, as well as

on codon usage. Figure 3C is a heat map generated by the model that exemplifies the intricate relationship between promoter initiation rate and ribosomal initiation rate and resulting protein levels, it could be seen from this heat map that there is a certain set of conditions that will afford high protein yields, even for a combination of a very low promoter initiation rate and a high ribosomal initiation rate.

An Analysis of Mutants and Initiation Rate Variants Suggests That DITA Influences mRNA Levels. Under the DITA assumption, we propose that the stalling of translation somewhere along an mRNA causes all upstream ribosomes to stall while all downstream ribosomes complete translation. This hypothesis also suggests that the stretch of mRNA between the DITA site and the 3' end will be more exposed to endonuclease cleavage. For that reason, we predicted that the larger the chances of DITA, the lower the mRNA levels will be, because mRNA is more exposed to endonucleases. Using the model, we determined the amount of mRNA produced by each of the mutants and compared it to the relative quantity (RQ) of GFP mRNA found in mid log phase cultures of the same mutants using qPCR (Figure 4A). A comparison of the qPCR and the modeled results revealed a strong correlation, suggesting that DITA affects both protein and mRNA levels by rapidly degrading not only the mRNA, but also nascent peptides. Since high mRNA levels usually correspond to high protein expression levels, it is essential to optimize protein expression for high levels of mRNA while maintaining the half-life of mRNA by avoiding DITA. This can be accomplished by exploiting the optimal regions, in terms of transcription and translation initiation rates, for maintaining a high level of GFP mRNA and by using regulatory elements that are strong enough but calibrated to prevent DITA under high expression density conditions. The heat map shown in Figure 4B is a result of a simulation of different initiation rates of the promoter and ribosomes and their influence on mRNA levels. It can be seen from the map that as expected ribosomal initiation rates have a

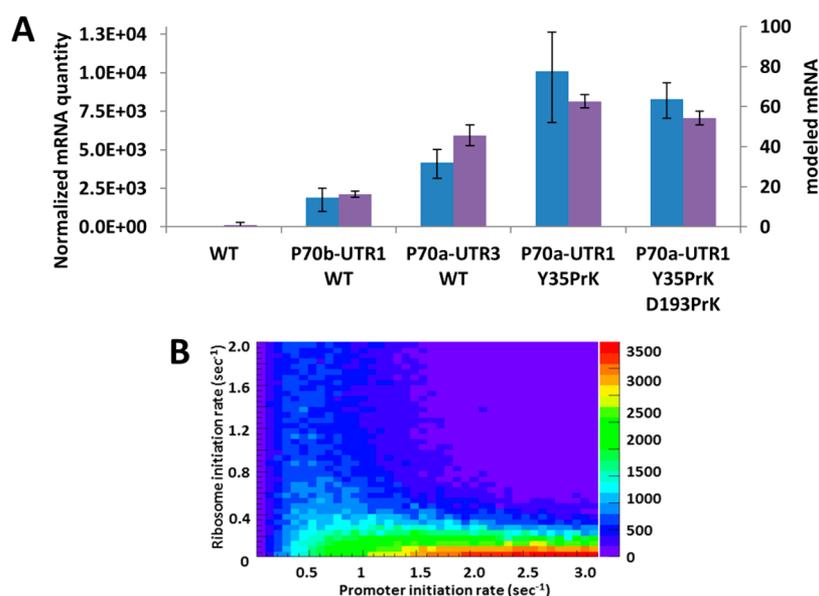


Figure 4. mRNA levels are also affected by DITA. (A) Comparison of the relative quantities of GFP mRNA transcripts found in mid log phase cultures and (blue bars) with the modeled mRNA quantities (purple bars). (B) Heat map of the expected amounts of GFP mRNA transcripts using different combinations of transcription and translation initiation rates.

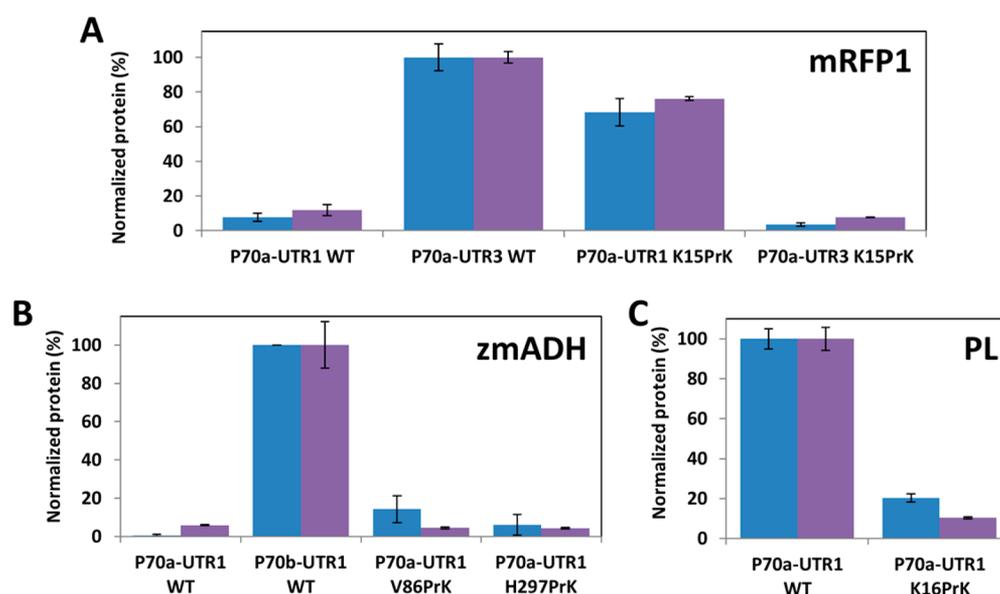


Figure 5. DITA is not limited to GFP and can be seen in other genes expressed using the P70a-UTR1 system and its variants. (A) Experimental (blue bars) and modeled (purple bars) protein levels of a codon optimized WT and K15PrK mRFP1. (B) Experimental (blue bars) and modeled (purple bars) normalized expression levels of zmADH. (C) Experimental (blue bars) and modeled (purple bars) normalized expression levels of WT and K16PrK protein L (PL).

very low influence on mRNA levels, however, our experimental results as well as the model have identified a set of conditions in which mRNA levels are influenced by ribosomal initiation rates. We do not exclude other explanations for the reduction in mRNA levels such as effect on transcription initiation by the density, or a codon bias effect³⁰ possibly mediated by a protein.³¹ However, if this is the case, then it is mutually inclusive to our hypothesis.

Testing Additional Proteins Supports the Generality of the DITA Phenomenon. To investigate whether the proposed phenomenon is a general mechanism and that it is not specific to GFP, we tested our model on three different genes: red fluorescent protein (mRFP1), *Zymomonas mobilis*

alcohol dehydrogenase II (*zmADH*) and the B1 domain of Protein L (PL), which is a small, 73-amino-acid polypeptide. The mRFP1 gene was chosen because it is a reporter protein as is GFP, however, mRFP1 shares only 26% similarity with the GFP amino acid sequence (sequences are available in the SI section), and it represents an optimized gene in terms of codon usage (it consists almost entirely of rapidly translating codons (A-type codons)). The genes were tested under similar conditions to those used for GFP. The experimental results for mRFP1 were in a good agreement with the model simulations (Figure 5A). This protein has shown the same trends as GFP both in the model and in the experiments. In contrast to mRFP1, *zmADH* is a larger, more complex gene

with lower translation rates owing to its abundance of codons from groups B and C, which attenuate the translation process and result in more complex folding dynamics. The results were once more in a good agreement with the simulations (Figure 5B), but we observed, contrary to model predictions, a partial rescue effect when testing expression levels with a late mutation (V86TAG). Still, a weaker promoter has given higher expression for this enzyme as well, which was observed both in the model prediction and experimentally. This observation is evidence that our model does not account for all factors that influence transcription/translation. Moreover, this finding suggests that cotranslational folding and chaperons may introduce bifurcation points at which nascent polypeptide length is significantly reduced. Thus, the special case of a late mutation can rescue a protein from DITA. Lastly, PL was chosen to test the model prediction that a protein with a relatively short polypeptide chain should have a much lower propensity for a DITA event (Figure S3A–D). Indeed, because it is a small protein, WT PL is efficiently produced at significantly greater levels than the TAG mutated variant (Figure 5C). The results with PL are additional experimental evidence that if the polypeptide is short enough, spatial collisions are less likely to occur although density is very high. The good agreement found between our model simulation and the experimental results for other proteins suggest that our proposed model is applicable not only to GFP.

DISCUSSION

Due to the fact that we have no direct evidence to the occurrence of DITA, we wanted to test our hypothesis by exploring alternative explanations for this phenomenon. Alternative explanations that were excluded by us are: differences in protein stability between a protein with an ncAA and WT protein, differences in plasmid copy numbers, mRNA secondary structure differences, as well as nCAAs interference with fluorescence of the reporting protein, GFP. In order to demonstrate no apparent change in protein stability between WT GFP and Y35TAG GFP, two experiments were conducted: We monitored the stability of the WT and the mutant protein in a crude cell lysate over the course of 24 h, showing that both proteins were stable with no significant change in fluorescence (Figure S6). In the second experiment we used synonymous, slow translating codons that were consecutively mutated around position 35, demonstrating that after the addition of four and above slowly translating codons, protein yields improve significantly to yields that are even higher than that of the protein with incorporated ncAA (Figure 2C), these important results indicate that the same protein with no structural change, but a change in the coding sequence, can be expressed with higher yields when the rate of translation slows down significantly, these results are in agreement with a recent report of Zhong and co-workers.³² These results also show that even with a strong promoter as is being used in this study, no hindrance from plasmid replication is observed. Evidence that attests to the fact that there is no hindrance for plasmid replication due to the existence of a strong promoter are the results with PL, since this protein is very short (*ca.* 70 AA) it is not affected by DITA, and high yields of expression are observed for this protein even with the strong promoter (Figure 5C). mRNA secondary structure could have accounted for the apparent differences in expression profiles between WT GFP and Y35GFP; however, an analysis of the mRNA secondary structure according to an algorithm written by

Mathews and co-workers³³ has shown no difference in mRNA secondary structure. The algorithm calculates mRNA secondary structure by taking into account base pairing, free energy minimization and other thermodynamic considerations. The analysis has shown that the single nucleotide change of C → G (Figure S2C,D) has no implications on mRNA's secondary structure, hence could not explain the discrepancy in expression levels. Moreover, once ribosomes bind mRNA during translation, the secondary structure is rendered almost linear, hence the predicted secondary structure is not relevant any longer and could not account for the observed difference. In order to exclude the possibility that nCAAs may interfere in any way with GFP fluorescence, we have quantified WT and mutant GFP and thus report their quantities rather than their fluorescence.

Additional possibilities were tested as well: ribosome abortion due to ribosome collisions was not excluded it could be an additional hindrance in the system but not an exclusive explanation since we could see elevated expressions of WT GFP also with a weak promoter and a strong RBS (Figure 3A). Another possibility is that due to the strong promoter and RBS there will be an extreme consumption of translation factors (*i.e.*, ribosomes, tRNAs, elongation factors, release factors), this possibility was excluded since it should have been seen for the much slower mutant as well (Y35PrK GFP), with the same strong promoter, multiple mRNAs will require multiple ribosomes too. Lastly, we have considered the plasmid copy number as a possible cause of low protein expression levels as is common with very strong promoters, however, our observations point to very low effect of plasmid copy numbers if any: the fact that the relatively small protein WT PL have shown high yields compared to the mutant protein using the same expression vector as for WT GFP expression, while the WT GFP have shown very small expression levels under the same conditions, contradicts the effect of plasmid copy number as the cause for low protein yields. In addition, for the same plasmid Y35PrK GFP have exhibited very high yields as well, again contradicting the effect of high plasmid copy number. Moreover, the synonymous mutations experiment (shown in Figure 2C), demonstrates very well that after the insertion of four synonymous “slow” translating codons in the beginning of the gene, protein expression levels are recovered, for the same plasmid, yet again demonstrating that plasmid copy number could not be the cause for low protein yields.

CONCLUSIONS

The ability of the model to accurately predict the expression trends of various proteins under different conditions led us to suggest that spatial expression density and DITA have significant effects on protein expression in cells. We note that our model does not take into account co-translational folding and therefore should not be applied to these cases. We would like to stress out that a natural system could not have been evolved to have such strong elements to drive higher protein expression, maybe due to DITA, hence, natural systems have evolved to prevent inefficiency and energy loss. We have used artificial transcription and translation elements as well as a recombinant GFP with a synthetic sequence to demonstrate DITA. These elements were then modified to control DITA levels. In our model the expression density of any gene relies on a combination of four key determinants: translation initiation and termination rates, transcription initiation and termination rates, gene length and codon bias. Herein, we propose an additional hypothesis for the important roles of codon bias and

genetic code redundancy. Although this effect was only observed in this study due to the use of highly efficient transcription and translation control regions, we infer that its effects could have significant, yet not always easy to observe implications, on the expression of all recombinant heterologous proteins. We propose that what is widely known as exogenous expression toxicity due to resource and energy depletion in some cases could be explained by DITA. In addition, we were able to show that by reducing the strength of the regulatory elements, we could lower expression density, resulting in a counterintuitive outcome that significantly improved protein yields. These protein expression dependencies were also observed at the mRNA levels of the various mutants, showing that it affects both cellular protein and mRNA levels, thus affecting the final quantities of protein produced. We showed that DITA occurs for several, highly dissimilar proteins, suggesting that it could be a general mechanism found in all bacteria. Moreover, our findings may also point out the importance of separating transcription and translation processes to increase the production rate of proteins, especially with longer and more complex genes. Obtaining a deep understanding of the transcription and translation processes is of an utmost importance; our findings are a novel step towards the ability to control and modify these processes, which may have a significant impact on protein expression both for fundamental research as well as for biotechnological applications.

■ EXPERIMENTAL PROCEDURES

GFP and mRFP1 Quantification and Purity Assessment. GFP and mRFP1 fluorescence were measured during overnight incubation at 37 °C. ncAA mutants were supplemented with PrK in a final concentration of 2 mM of ncAA. The various mutants were grown in 96 well plates while OD600 and fluorescence were measured every 20 min for up to 20 h. GFP and mRFP1 fluorescence were measured with the respective excitation/emission wavelengths of 488/510 nm and 584/607 nm. GFP mutants were purified using nickel affinity chromatography, and the resulting samples were measured using a commercial Bradford assay (Thermo Scientific, Waltham, MA). Western blot analysis was used to verify the integrity of fluorescence as a measure of protein quantity when comparing the various mutants and to eliminate the possibility of fluorescence reduction due to ncAA incorporation. For Western blot analysis, goat anti-GFP and donkey antigoat (HRP-conjugated) antibodies were used as primary and secondary antibodies (Santa Cruz, CA, USA), respectively.

GFP Purification and Mass Spectrometry. For the LC–MS validation of PrK incorporation, nickel affinity chromatography purification (IMAC) of 6xhis-tagged GFP was performed. Overnight cultures of 100 mL were lysed using BugBuster protein extraction reagent (Novagen, WI, USA) and 6xHis tagged GFP was purified from the crude lysate using His-Bind nickel affinity chromatography resin (Novagen). The protein-containing eluted fraction was concentrated using a 10 kDa cutoff Vivaspin concentrator (Sartorius, Göttingen, Germany) to a final concentration of 2 mg/mL. The resulting concentrated fraction was analyzed by LC–MS (Finnigan Surveyor Autosampler Plus/LCQ Fleet (Thermo Scientific, Waltham, MA), using Chromolith monolithic column (EMD Millipore). The results were analyzed using Xcalibur (Thermo) and Promass (Novatia) software. MS/MS analysis was performed using standard protocols for in-gel trypsin digestion and desalting using ZipTip μ C18 (EMD Millipore). The

desalted peptides were analyzed on an LTQ/Orbitrap mass spectrometer (Thermo). Collision induced dissociation (CID) was used to analyze ions of interest for tandem mass spectrometry.

Zymomonas mobilis Alcohol Dehydrogenase (zmADH) Expression and Quantification. Cultures of C321.ΔA.exp harboring the pBEST-zmADH plasmid with the various mutants were incubated at 37 °C. Cultures intended for ncAA incorporation were also supplemented with PrK at a final concentration of 2 mM. zmADH expression was analyzed by quantifying ADH activity in the samples.³⁴ The results were also semiquantitatively verified by densitometry analysis of a Western blot of the different mutants. Blotting was done using anti His-tag antibodies made in mice (Santa Cruz, CA, USA). The Western blot results were analyzed using ImageJ software.³⁵

B1 Domain of Protein L (PL) Expression and Quantification. The PL gene was subcloned to the pBEST P70a-UTR1 vector. The K16TAG mutant was created using site-directed mutagenesis (primer sequences can be found in the SI section). The two variants were transformed separately into C321ΔPrf1.EXP already harboring the pEVOL-Pyl OTS plasmid. The cultures were incubated overnight at 37 °C in LB media supplemented with 2 mM of propargyl-L-lysine. The OD of the cultures was calibrated and lysis was performed using the protocol supplied with the BugBuster Reagent (Novagen, WI, USA). A sample of each lysate was loaded onto SDS-PAGE (WT sample was diluted by a factor of 10) and then blotted using anti His-Tag antibodies produced in mice (Santa Cruz, CA, USA). The Western blot results were analyzed using imageJ software³⁵ and the conversion to molar concentration was done using a calibration curve.

mRNA Quantification. A GeneJET RNA purification kit (Thermo Scientific, Waltham, MA, USA) was used to extract total RNA from bacterial cultures during midexponential phase. cDNA samples were synthesized from RNA samples using iScript cDNA synthesis kit (Biorad, Hercules, CA, USA). qPCR was performed using KAPA SYBR FAST qPCR Kit (KapaBiosystems, Wilmington, MA, USA) with the recommended relative calibration curve protocol, in the StepOnePlus Real-Time PCR System (Thermo Scientific, Waltham, MA, USA).

Modeling the Expression Density under the DITA Assumption. The system was modeled in a 2D temporal-spatial model and was simulated using a Gillespie algorithm. All the parameters used in the model are detailed in Supporting Information Table S2. The parameters were assessed and determined from literature and from experimentation as described in detail in the supporting experimental procedures section. The computational simulation enabled the assessment of mRNA and protein production kinetics and statistical assessment of the propensity for density induced translation arrest under different parameter regimes. Detailed explanation about the construction of the model including literature sources, and experiments along with the computational simulation code are available in the SI section.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00019.

Additional experimental procedures; Supporting figures, tables and gene sequences (PDF)
Complete code of the model (PDF)
Constants used in the model (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: mhjensen@nbi.ku.dk. Phone: +45-353-25371. Fax: +45-353-25425.

*E-mail: alfontal@bgu.ac.il. Phone: +972-8-6479066. Fax: +972-8-6479576.

ORCID

Orr Schlesinger: 0000-0002-2648-6235

Lital Alfonta: 0000-0002-3805-8625

Author Contributions

#O.S., Y.C. and M.H. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank the following for their valuable contributions: Prof. Gilad Haran for providing us the PL gene, Prof. Yuval Shoham for providing the zmADH gene, Prof. Edward Lemke for providing us with pEVOL-PylRS plasmid, Dr. Namiko Mitarai for fruitful discussions and Prof. Kim Sneppen for invaluable advice. ERC-SG number 260647 (L.A.), The Danish Council for Independent Research (M.H.J. and M.H.) and Office of Naval Research award number N00014-13-1-0074 (V.N.) are gratefully acknowledged.

REFERENCES

- (1) Browning, D. F., and Busby, S. J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65.
- (2) Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508.
- (3) Chen, H.-Z., and Zubay, G. (1983) Prokaryotic coupled transcription–translation. *Methods Enzymol.* 101, 674–690.
- (4) Ellis, R. J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* 26, 597–604.
- (5) Tan, C., Saurabh, S., Bruchez, M. P., Schwartz, R., and Leduc, P. (2013) Molecular crowding shapes gene expression in synthetic cellular nanosystems. *Nat. Nanotechnol.* 8, 602–608.
- (6) Mitarai, N., Sneppen, K., and Pedersen, S. (2008) Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J. Mol. Biol.* 382, 236–245.
- (7) Jin, H., Björnsson, A., and Isaksson, L. A. (2002) Cis control of gene expression in *E. coli* by ribosome queuing at an inefficient translational stop signal. *EMBO J.* 21, 4357–4367.
- (8) Sørensen, M. A., Kurland, C. G., and Pedersen, S. (1989) Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* 207, 365–377.
- (9) Fredrick, K., and Ibba, M. (2010) How the sequence of a gene can tune its translation. *Cell* 141, 227–229.
- (10) Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborse, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354.
- (11) Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., and Barral, Y. (2010) A role for codon order in translation dynamics. *Cell* 141, 355–367.
- (12) Tuller, T., and Zur, H. (2015) Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Res.* 43, 13–28.
- (13) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500.
- (14) Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8, 883–890.
- (15) Wang, J., Kwiatkowski, M., and Forster, A. C. (2016) Kinetics of tRNA Pyl-mediated amber suppression in *Escherichia coli* translation reveals unexpected limiting steps and competing reactions. *Biotechnol. Bioeng.* 113, 1552–1559.
- (16) Fan, C., Xiong, H., Reynolds, N. M., and Söll, D. (2015) Rationally evolving tRNA Pyl for efficient incorporation of non-canonical amino acids. *Nucleic Acids Res.* 43, e156–e156.
- (17) Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., and Isaacs, F. J. (2013) Genomically recoded organisms expand biological functions. *Science* 342, 357–360.
- (18) Shin, J., and Noireaux, V. (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4, 8.
- (19) Mankin, A. S. (2006) Nascent peptide in the “birth canal” of the ribosome. *Trends Biochem. Sci.* 31, 11–13.
- (20) Komar, A. A. (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* 34, 16–24.
- (21) Woolhead, C. A., Johnson, A. E., and Bernstein, H. D. (2006) Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* 22, 587–598.
- (22) Sunohara, T., Jojima, K., Tagami, H., Inada, T., and Aiba, H. (2004) Ribosome stalling during translation elongation induces cleavage of mRNA being translated in *Escherichia coli*. *J. Biol. Chem.* 279, 15368–15375.
- (23) Klumpp, S., and Hwa, T. (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20245–20250.
- (24) Quax, T. E. F., Claassens, N. J., Söll, D., and van der Oost, J. (2015) Codon bias as a means to fine-tune gene expression. *Mol. Cell* 59, 149–161.
- (25) Mitarai, N., and Pedersen, S. (2013) Control of ribosome traffic by position-dependent choice of synonymous codons. *Phys. Biol.* 10, 56011.
- (26) Dana, A., and Tuller, T. (2015) Mean of the typical decoding rates: a new translation efficiency index based on the analysis of ribosome profiling data. *G3: Genes, Genomes, Genet.* 5, 73–80.
- (27) Klumpp, S., and Hwa, T. (2008) Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18159–18164.
- (28) Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V., and Alfonta, L. (2015) Genetically expanded cell-free protein synthesis using endogenous Pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* 112, 1663–1672.
- (29) Ozer, E., Chemla, Y., Schlesinger, O., Aviram, H. Y., Riven, I., Haran, G., and Alfonta, L. (2016) In vitro suppression of two different stop codons. *Biotechnol. Bioeng.* DOI: 10.1002/bit.26226.
- (30) Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K.-H., Su, M., Luff, J. D., Valecha, M., Everett, J. K., Acton, T. B., Xiao, R., Montelione, G. T., Aalberts, D. P., and Hunt, J. F. (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529, 358–363.
- (31) Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., and Collier, J. (2016) The DEAD-Box protein dhh1p couples mrna decay and translation by monitoring codon optimality. *Cell* 167, 122–132.
- (32) Zhong, C., Wei, P., and Zhang, Y.-H. P. (2016) Enhancing functional expression of codon-optimized heterologous enzymes in *Escherichia coli* BL21(DE3) by selective introduction of synonymous rare codons. *Biotechnol. Bioeng.* DOI: 10.1002/bit.26238.

(33) Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 7287–7292.

(34) Amir, L., Carnally, S. A., Rayo, J., Rosenne, S., Melamed Yerushalmi, S., Schlesinger, O., Meijler, M. M. M., and Alfonta, L. (2013) Surface display of a redox enzyme and its site-specific wiring to gold electrodes. *J. Am. Chem. Soc.* *135*, 70–73.

(35) Abramoff, M. D., Magalhães, P. J., and Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics Int.* *11*, 36–42.

Supporting Information

Tuning of recombinant protein expression in *Escherichia coli* by manipulating transcription, translation initiation rates and incorporation of non-canonical amino acids

Orr Schlesinger,^{†,#} Yonatan Chemla,^{†,#} Mathias Heltberg,^{‡,#} Eden Ozer,[†] Ryan Marshall,[§] Vincent Noireaux,[§] Mogens Høgh Jensen,^{*,‡} and Lital Alfonta^{*,†}

[†]Department of Life Sciences and Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, POBox 653, Beer-Sheva 8410501, Israel

[‡]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

[§]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, United States

*Correspondence: mhjensen@nbi.ku.dk (M.H.J), alfontal@bgu.ac.il (L.A)

These authors contributed equally.

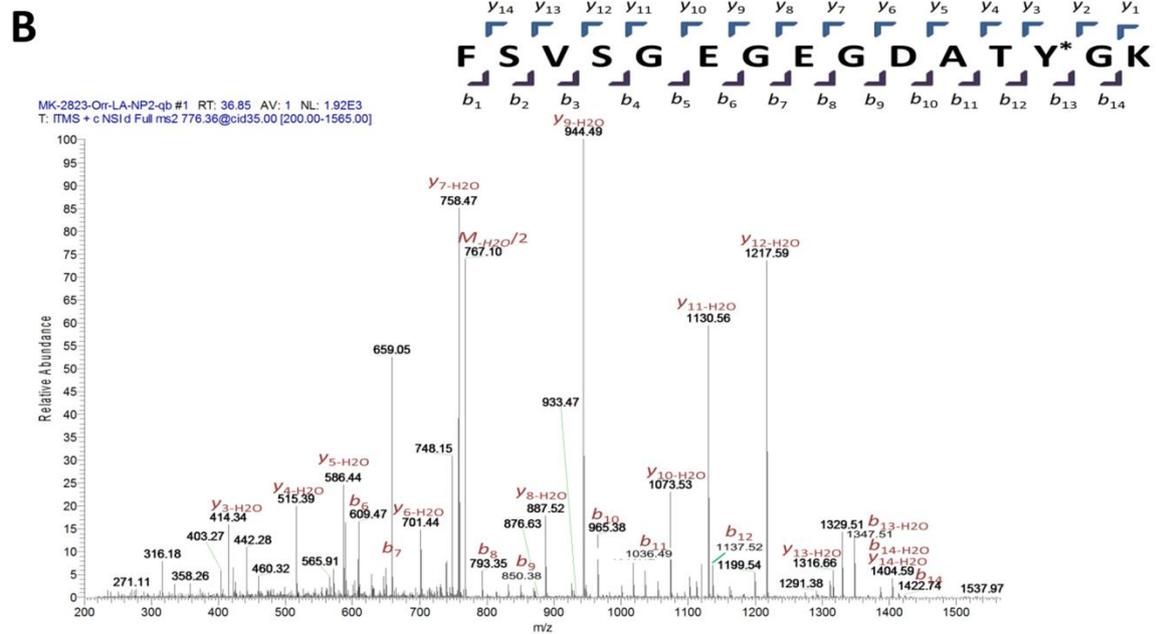
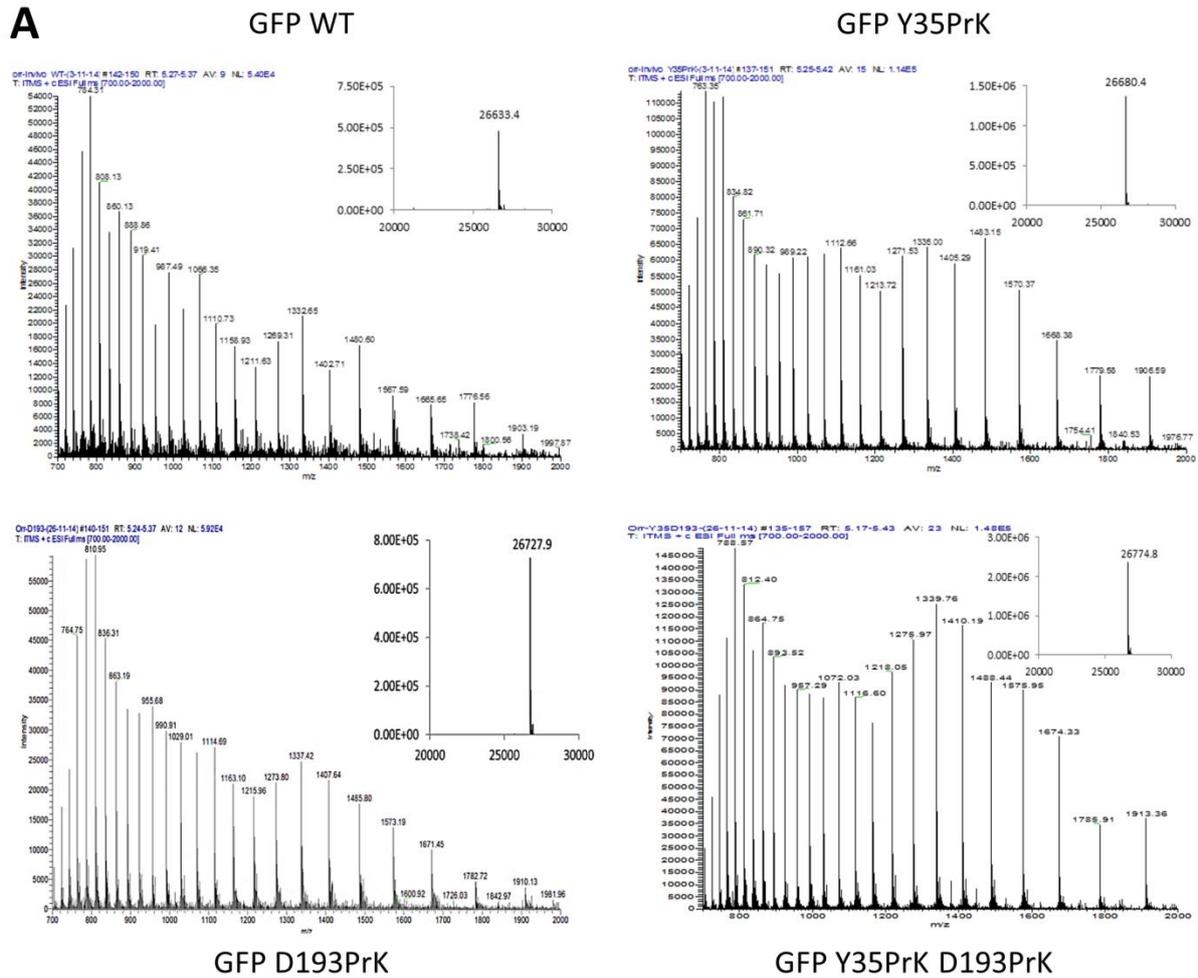


Figure S1. Mass spectrometry analysis of the GFP mutants. (A) LC-MS analysis of GFP mutants. The insets show the deconvoluted mass of each mutant. Calculated masses of GFP WT: 26634Da; Y35PrK: 26680Da; D193PrK: 26729Da; Y35D193PrK: 26776Da. (B) MS/MS spectrum and fragmentation analysis of a peptide confirms the presence of PrK in the 35th position of GFP (denoted as Y* in the fragmentation analysis).

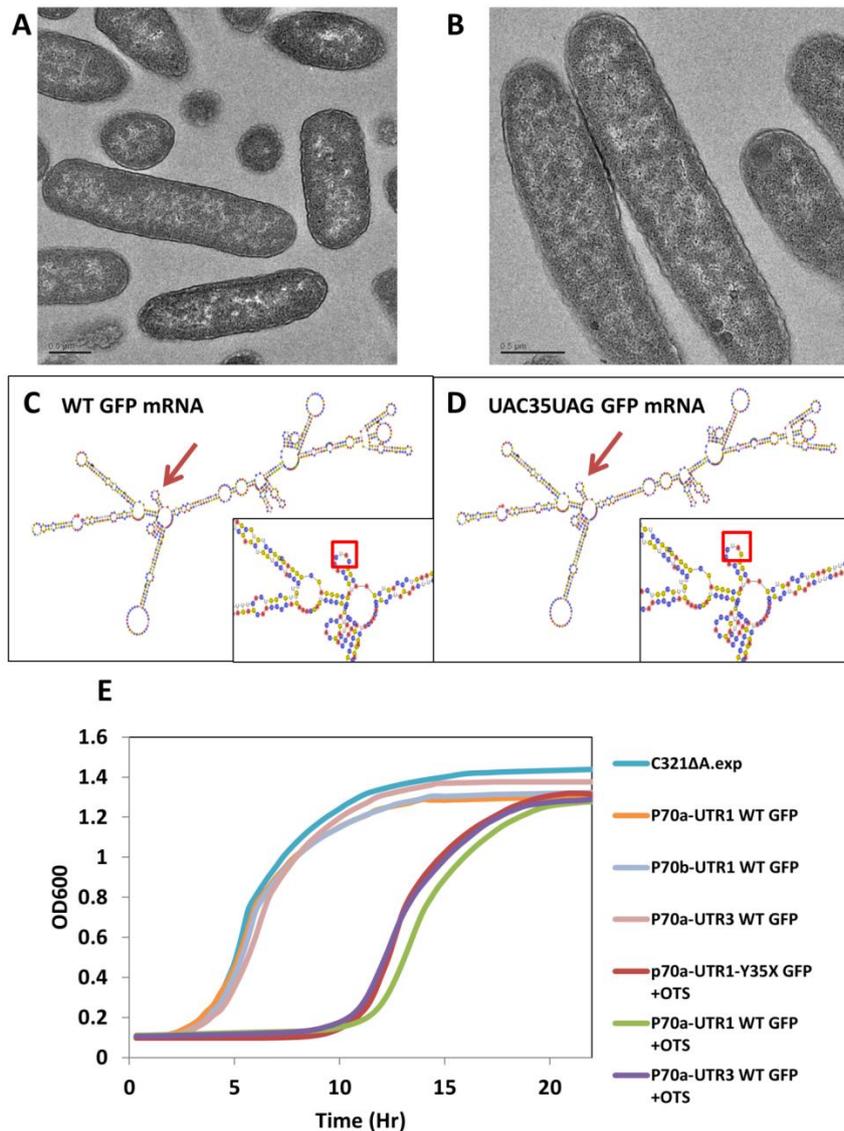


Figure S2. There is no apparent effect of the GFP mutants on bacterial morphology, mRNA secondary structure and growth rate. (A and B) Cryo-TEM images show that there is no significant change in bacterial morphology upon GFP overexpression. (A) TEM images of C321ΔA.exp bacteria. (B) The same bacteria cultured for the expression of WT GFP under P70a-UTR1. (C and D) There is no observable change in predicted mRNA secondary structure upon introduction of the UAC35UAG mutation. (C) mRNA secondary structure for WT GFP and (D) for the UAC35UAG mutant. The inset shows a close-up of the mutation site (marked with a red arrow). The prediction of the secondary structure was performed using an algorithm developed by Turner and coworkers (Mathews and Turner, 2002; Mathews et al., 2004). The algorithm uses known base pairing, free-energy minimization and other thermodynamic parameters to predict the secondary structure of them RNA from the input sequence. this prediction shows no base-pairing around the mutation site nor any significant change in secondary structure around the mutation site. Note that the model is for free RNA molecule, while in our system there are ribosomes RNA binding proteins and elongation factors, straightening and modifying the final structure during translation. (E) Bacterial growth rates remain unaffected for the different mutants during incubation at 37°C. ncAA mutants were supplemented with 2 mM PrK per ncAA mutation. The longer lag phase for the mutants incorporating ncAAs is due to the addition of an additional antibiotic (Chloramphenicol) to the growth medium. a control experiment demonstrates a lag phase even when WT GFP is expressed along with the OTS plasmid. When the two plasmids are present in the presence of the corresponding antibiotics we observed a lag phase similar to that observed with the ncAA incorporated mutants.

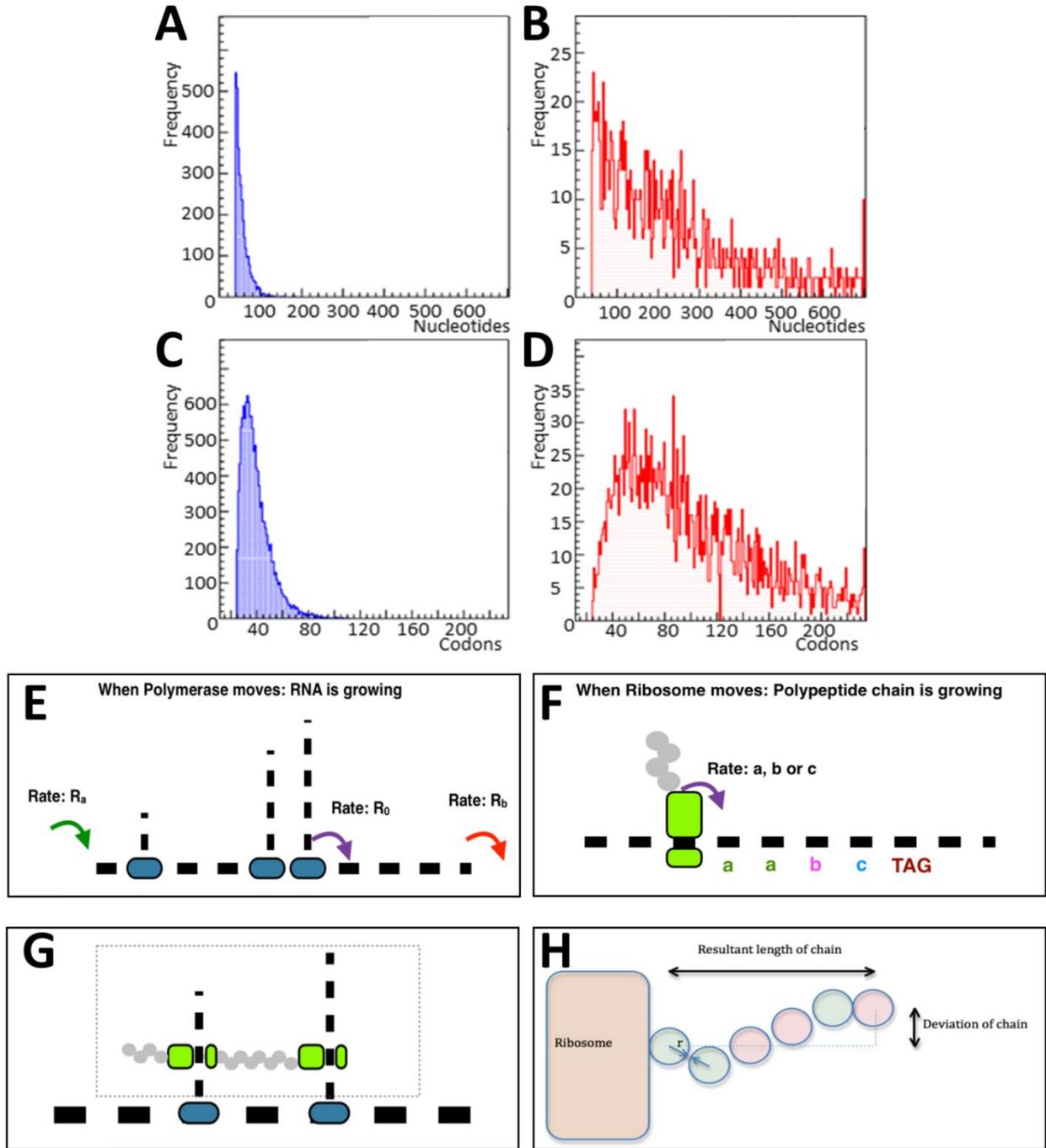


Figure S3. Model description and analytical solutions of polymerase and ribosome distributions under different initiation rates. (A and B) RNA polymerase distributions for (A) P70a (strong) and (B) P70b (weak) promoters. In the following calculations we study the DITA model, but treat every step in a purely deterministic way. These results are compared to the simulations. The time it takes to move at a step with rate r_i is $\frac{1}{r_i}$ instead of a probability distribution. This can be expressed by equation 6:

$$P(x = i, t) = r_i e^{-r_i t} \Rightarrow P(x = i, t) = \begin{cases} 1 & \text{if } t \leq \frac{1}{r_i} \\ 0 & \text{if } t > \frac{1}{r_i} \end{cases} \quad (6)$$

We first want to calculate the distance between the two polymerases (P1 and P2) with size L_p when P2 has just been initiated, i.e., the position $X_2 = 1$. P1 must move L_p nucleotides (36 in the present model), after which the equations for the times, $\frac{1}{R_\alpha} = \frac{N}{R_0} \Rightarrow N_{pi} = \frac{R_0}{R_\alpha}$ where N_{pi} is the initial distance between P1 and P2, are calculated. Inserting numbers for the P70a promoter, $R_0 = 51$ nucleotides/s (polymerase elongation rate) and $R_\alpha = 3/s$ (polymerase elongation rate) gives an average distance: $\langle D \rangle = 36 + N_{pi} = 53.01$ nucleotides. We compare this to the simulations in the model and find $\langle D \rangle = 53.4 \pm 0.27$ nucleotides, which is slightly larger than one standard deviation above, so they are in a good agreement. Inserting numbers $R_0 = 36$ nucleotides/s and $R_\alpha = 0.2/s$ (for the P70b promoter) gives: $\langle D \rangle = 36 + N_{pi} = 291$ nucleotides. We compare this to the simulations in the model and find $\langle D \rangle = 227.58 \pm 4.35$ nucleotides. Here we have a disagreement between the averaged equations and the simulations. This is due to the fact that polymerase queuing plays an important role, thus decreasing the average distance between polymerases. (C and D) Ribosome distribution for (C) UTR1 (strong) and (D) UTR3 (weak) ribosome binding domains (RBS) can likewise be calculated. However, this distance is not very informative, since, in contrast with the case for DNA, the rates vary widely on the mRNA. To correct for that we calculated the average distance by equation 7:

$$\langle d \rangle \approx L_p + \sum_{i=L_p}^n \frac{1}{r_i} + \frac{\frac{1}{r_\alpha} - \sum_{i=L_p}^n \frac{1}{r_i}}{r_n + 1} \quad (7)$$

where r_α is the ribosome initiation rate, r_i is the specific codon elongation rate and r_n is the ribosome termination rate. Inserting the rates, we find that $\langle d_{P70a} \rangle \approx 25$ and $\langle d_{P70b} \rangle \approx 126$. This is only valid if we assume no queue, so simulating the distance between the first and second ribosome, we find that $\langle d_{P70a} \rangle \approx 24.56$ and $\langle d_{P70b} \rangle \approx 110$. While the analytical calculation agrees quite well for the strong RBS, there is a mismatch between these calculations and the simulation for the weak RBS due to ribosomes queuing on the mRNA. The position at which DITA can occur can also be calculated as follows: the length of the polypeptide chain is given as equation 5: $L = W_R + \lambda x_i$, where W_R is the average linear width of the tRNA inside the ribosome and λ is the proportionality constant for the linear width of the average amino acid in the polypeptide chain. This means that DITA can occur (in terms of codons), which is given by equation 8:

$$L_p = W_R + \lambda x_i \Rightarrow x_i = \frac{L_p - W_R}{\lambda} \quad (8)$$

And for the purely averaged model this would be equation 9:

$$x_i = \frac{\langle D \rangle - W_R}{\lambda} \quad (9)$$

In both situations, we have a relation of inverse proportionality. Lastly, the average number of ribosomes on a fully transcribed mRNA molecule can be calculated. This is done while considering that the polymerase will transcribe the $L_{t_{gene}}$ number of codons (length of the gene – size of ribosome [L_r]). For example in the case of the GFP gene $L_{t_{GFP}} = 234 - 12 = 222$. The average number of ribosomes on an actively transcribed mRNA molecule, N_r , is calculated by equation 10:

$$\frac{L_{t_{gene}}}{R_0} = \frac{1}{r_\alpha} + (N_r - 1) \left(\frac{1}{r_\alpha} + \sum_{i=1}^{L_r} \frac{1}{r_i} \right) \Rightarrow N_r = 1 + \frac{\frac{L_{t_{gene}}}{R_0} - \frac{1}{r_\alpha}}{\frac{1}{r_\alpha} + \sum_{i=1}^{L_r} \frac{1}{r_i}} \quad (10)$$

For the P70a promoter we get $\langle N_r \rangle \approx 8.8$ and for the P70b promoter we get $\langle N_r \rangle \approx 1.9$. A comparison with the stochastic simulations ($\langle N_r \rangle \approx 8$ for P70a and $\langle N_r \rangle \approx 1.6$ for P70b) shows that the results are in good agreement.

(E) RNA polymerase moves in a stochastic manner with a mean rate R_0 . The polymerized mRNA molecule's length is proportional to the RNA polymerase along the transcribed gene. (F) In the model, the ribosome moves in a stochastic manner, with a mean time spent on a certain codon according to the 3 speed classes (a, b and c) for canonical amino acids and the slower, d class for ncAA incorporation in response to a UAG codon. (G) DITA may occur when the distance to the polymerase is the same as for the ribosome on one of the neighboring RNA strands and the length of the polypeptide chain is longer than the distance between polymerases, the two ribosomes are "arrested". This is one possibility to represent the expression density and its effect on protein production (H) The proportionality constant lambda (λ) governs the length of the polypeptide protruding from its parent ribosome in the model.

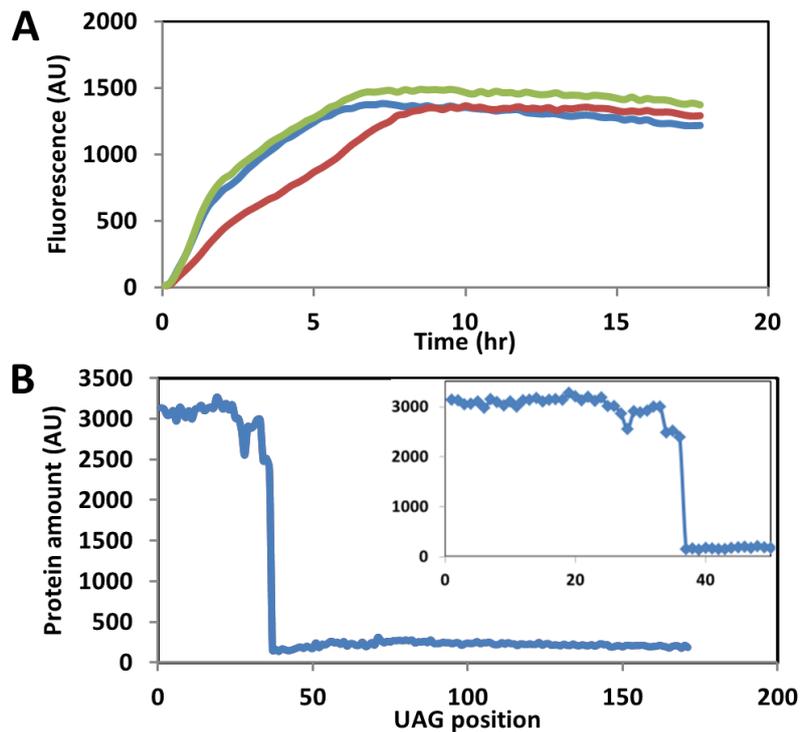


Figure S4. D193 is a permissive mutation for incorporation of ncAAs into GFP and expression can be ‘rescued’ by an earlier ncAA incorporation. (A) In-vitro experiments show that both ncAA incorporation sites, Y35PrK (red curve) and D193PrK (green curve), are permissive and mutants are expressed well compared to GFP WT (blue curve). (Chemla et al., 2015). (B) Model simulation for the ability of a UAG codon to ‘rescue’ later codons and prevent DITA in GFP. Inset shows a magnification of the first 50 codons.

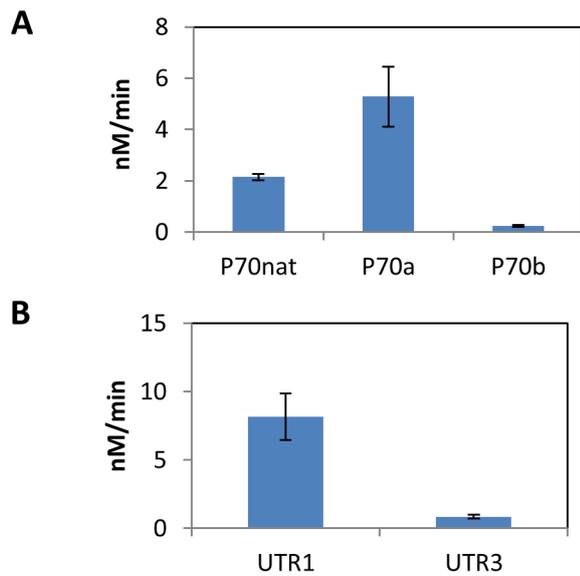


Figure S5. Initiation rates of the different promoter and ribosomal binding site variants. (A) Transcription initiation rates expressed as nM mRNA produced per minute for the native viral promoter (P70nat), the stronger variant P70a, and the weaker variant, P70b, determined by fluorescence (the fluorescent Broccoli RNA aptamer was cloned under each promoter). (B) Translation initiation rates for UTR1 and UTR3 expressed as protein expression rates as determined *in-vitro* by fluorescence (GFP was cloned under each UTR).

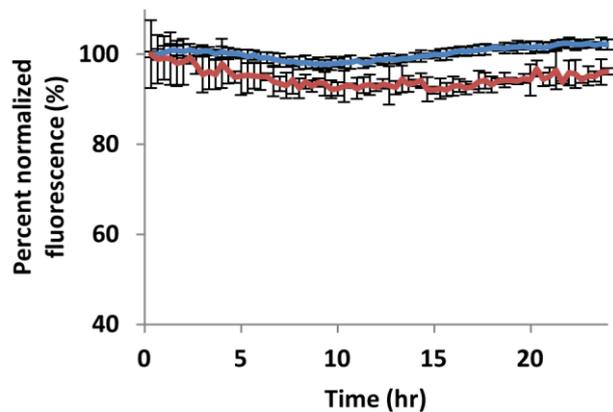


Figure S6. Protein stability is unchanged due to incorporation of ncAA in position 35 The stability and fluorescence of GFP WT (blue curve) and GFP Y35PrK (red curve) in a crude lysate without protease inhibitors over the course of 24 hours.

Experimental procedures

Bacterial strains and plasmids transformation

E. coli DH5 α were used to amplify plasmids, to construct GFP mutants and promoter variants. *E. coli* C321. Δ A.exp (Addgene strain #49018) was used for expression of GFP under the P70a promoter and its variants. *E. coli* BL21(DE3) was used for the expression of GFP mutants under the control of a T7 promoter. All transformations were done by electroporation using standard protocols. All strains not containing any plasmids were grown in Luria-Bertani (LB) broth (10 g/L NaCl, 10 g/L trypton and 5 g/L yeast extract) overnight at 37°C for sequential inoculation of the expression cultures.

Plasmids and mutant construction

The pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene plasmid #40019) plasmid was constructed by V.N. Mutations in the GFP, promoter and RBS were introduced by site-directed mutagenesis using standard protocols. Primers and reagents used for vector construction and promoter/RBS modifications can be found in the SI section.

Modeling the expression density under the DITA assumption

Gillespie Algorithm

When modeling the system, it was essential to consider the process from a stochastic approach. We used the Gillespie algorithm as described in the original paper from 1977(Gillespie, 1977). We considered seven possible reactions: Initiation of transcription, transcription elongation, termination of transcription, initiation of translation, elongation of translation, termination of translation and mRNA degradation. The time to the next reaction was calculated according to equation (3):

$$t_{next} = -\ln\left(\frac{rand}{S}\right) \quad (3)$$

where $S = \sum_{ribosome} r_i + \sum_{polymerase} R_j + \sum_{mRNA} \Phi_k$.

The distribution of time until the next reaction followed an exponential function determined by the sum of the rates of all possible reactions. The formula above transforms a random number between 0 and 1 into an exponentially dependent parameter. This guarantees stochasticity following the nature of the reactions in a living bacterial cell. Likewise, the subsequent reaction should be stochastic but weighted by the rate Q_i relative to the sum of all rates. Here to choose the subsequent reaction, we picked a random number between 0 and 1 to ensure that the next reaction would satisfy the following condition described by equation (4):

$$Next\ reaction: \frac{Q_i}{S} \leq rand \leq \frac{Q_{i+1}}{S} \quad (4)$$

For the sake of simplicity, we chose to make our model 2D rather than 3D. The main justification for this simplification is that the RNA polymerases transcribing a single gene are assumed to be on the same plane of

rotation as the DNA strand in front of them to facilitate elongation on a single plane (Harada et al., 2001). As translation takes place on a single-stranded RNA molecule that is relaxed by the translating ribosomes and that has only one orientation (as nucleotides are unlikely to flip or rotate), we made an additional simplifying assumption, namely, that all the ribosomes were also on one plane.

Model characteristics:

Movement of Polymerase and Transcription: We consider the DNA strand of the GFP gene with a length of 714 nucleotides. First the RNA-polymerase can attach to the DNA promoter and initiate transcription with a rate R_a given there is no other RNA-Polymerase already attached to the DNA at any of the 36 first nucleotides, since this is the approximated width of an RNA-Polymerase. Next the polymerase moves on the 714 nucleotides that with an R_0 rate. It should be noted that because all positions on the DNA have the same rate, it does not mean that the RNA-Polymerase moves in a uniform velocity. As in a stochastic movement, a constant rate means that the distribution of time the polymerase spends at a given position before moving, is given by an exponential distribution with mean R_0 . Furthermore, the basic assumption that no polymerase can share the same position, means that a queuing structure will be created, and this will lead to regions where some polymerases will move very slowly due to leading polymerases, there by chance, and that have moved a bit slower down the strand. When the polymerase reaches the final nucleotide of the gene, it can terminate transcription and release the mRNA with a rate R_b . Whenever the RNA-polymerase moves one position, it produces mRNA, meaning that the length of the mRNA strand will to be equal to the position of the polymerase (Fig S3E).

Movement of Ribosome and Translation: Whenever the length of the transcribed mRNA is equal to 12 codons, which is the approximated width of the ribosome, a ribosome can attach to the mRNA strand with rate r_a . Once attached the ribosome can move on the mRNA strand with rates that are governed by each codon. For simplicity reasons all the codons has been divided into 3 groups: $a=35\text{codons/s}$, $b=8\text{codons/s}$, $c = 4.5\text{codons/s}$ while the genetically expanded codon - UAG = 0.04 codons/s. The ribosome can only move if there is no ribosome occupying any of the codons in front of it. Whenever the ribosome moves one position, a new amino acid is translated and added to the growing polypeptide chain. This means that the length of the polypeptide chain perpendicular to the direction of movement of the ribosome, can be described by the linear relation, depending on the position of the ribosome.

Length of the polypeptide chain and the proportionality constant λ : The nascent polypeptide folding dynamics were not modeled in this work. Therefore, the length of the growing polypeptide chain could be fitted empirically in a proportional manner to the number of amino acids added to the chain (Fig S3F). The constant of proportionality is denoted λ , and is estimated as 0.1 nm/codon. The length of the polypeptide, L , is governed by equation (5):

$$L = \lambda * x_i + W_R \quad (5) \text{ where } x_i \text{ is the codon position of the ribosome.}$$

Termination of Ribosome Movement and Protein Production Whenever the ribosome reaches the terminal stop codon, it can terminate translation of the mRNA strand with a rate r_a , thereby releasing the polypeptide chain to produce a protein. From the above description, this can only happen after the polymerase has detached

from the DNA strand. Therefore the released mRNA can facilitate protein production, until they are degraded with a rate k_{RNA} .

Density Induced Translation Arrest: From the model description, it can be deduced, that the distance between two ribosomes is given by the distance between polymerases. Thus if two polymerases are close to each other, the distance between the ribosomes attached to the mRNA will be given by the distance between the polymerases, and in this picture, the size of the polypeptide chain can be longer than this distance. Therefore, we assume that if a ribosome moves into a position where the distance to the polymerase is the same as the ribosome on one of the neighboring mRNA strands and the length of the polypeptide chain is longer than the distance between polymerases, the two ribosomes are “arrested”. This arrest, causes the two ribosomes to stop moving and after transcription ends, an mRNA that has arrested ribosomes is immediately degraded (Fig S3G). It’s important to note that we do not argue that this is a physical mechanism of collision but rather these conditions enable a way to model the local expression density and we allow that this is not the only possible mechanism of density induced translation arrest.

Model parameters:

We used the experimental parameters closest to those in the literature, when available. Since all of the *in-vivo* experiments were carried at 37°C, the parameters were chosen from literature describing *E. coli* experiments done at high doubling rates, i.e., two doublings per hour. The RNA polymerase (RNAP) initiation rate, R_{α}^S , for the P70a synthetic promoter was determined using the reported initiation rate for the natural lambda bacteriophage P70nat promoter of 1.1 initiations/second (McClure, 1983). The natural initiation rate was augmented to account for the increase in rate caused by the synthetic mutations (Shin and Noireaux, 2010). The rates of both natural and synthetic promoters were measured *in-vitro* and the synthetic promoter was found to be three times stronger than the natural promoter (Extended data figure E7A). These results led us to approximate the rate of R_{α}^S to be 3 initiations/second. The increased kinetics are in agreement with the promoter-RNAP K_m (Hawley and McClure, 1980) taking into account the number of free polymerases in the cell (Klumpp and Hwa, 2008a). Note that this rate is larger than the average time it takes the polymerase to evacuate the promoter binding site, resulting in saturation of initiation (Klumpp and Hwa, 2008b).

The experimentally measured transcription initiation rate of the mutated weak promoter, P70b, was found to be ca. 20 times lower than that of P70a (Extended data fig. E7A). RNAP elongation rate, R_0 , was set at 51 nt/second (Proshkin et al., 2010; Vogel and Jensen, 1994). Transcription termination, i.e., R_{β} , was set at 3 terminations/second (adjusted from Arndt et. al. (Arndt and Chamberlin, 1988)). Ribosome initiation rate, i.e., r_{α}^S , for UTR1 ribosome binding site was determined using the available data on high rate ribosome binding domains (Brandt et al., 2009; Kennell and Riezman, 1977) and was set at 1.5 initiations/second.

The experimentally measured translation initiation rate for the mutated weak RBS, UTR3, i.e., r_{α}^W , was found to be 10 times lower than that of UTR1, and therefore, it was set at 0.15 (Extended Data Fig. E7B). Ribosomal elongation rates were calculated by dividing the canonical codon elongation rates into three groups (A, B and C). The rates of the groups R_A , R_B , and R_C were set at 35, 8 and 4.5 codons/second, respectively.

This approach was adopted from Mitarai et al., and these groups were found to be a reasonable simplification in good agreement with the experimental results(Mitarai and Pedersen, 2013). The elongation rate of UAG, i.e., group D codons, was assessed as two orders of magnitude lower than group C codons, i.e., $R_D=0.04$ codons/second. It has been widely accepted that in the absence of the cognate release factor 1 (RF1), the elongation rate of the suppressed UAG codon is slow. This inefficiency is a result of the low affinity of the orthogonal tRNA to the ribosomal A site tertiary complex(Wang et al., 2015). Moreover, it was shown that efficiency can be enhanced by improving the orthogonal tRNA (i.e., tRNA_{cua}^{pyl})(Fan et al., 2015).

The ribosome termination rate, i.e., r_β , was set at 1 termination/second (adjusted from Gritsenko et. al. (Gritsenko et al., 2015)). The model takes into account the ribosomal and RNA polymerase footprints, as no initiation event can occur until the RBS or the promoter are vacant. Moreover, the translational and transcriptional densities take this size into account when calculating the occupancy density. The ribosome(Ingolia, 2014) and polymerase(Zaychikov et al., 1995) footprints were set at 12 codons and 36 nucleotides, respectively. The distance from the mRNA strand and the elongating nascent polypeptide, i.e., W_R , was calculated to be 8 codons, the estimated height of a tRNA molecule. The WT mRNA half-life rate, i.e., $t_{1/2}$, was set to be 60 seconds(Liang et al., 1999; Pedersen and Reeh, 1978). The mRNA half-lives of mutants containing 1, 2 or 3 TAG mutations were determined empirically to be 95%, 90% and 80%, respectively, of the WT mRNA half-life(Morse and Yanofsky, 1969). As the ribosome coverage of the UTR3 Y35TAG GFP and UTR3 K26TAG mRFP1 variants is very low, their mRNA half-life times were fitted to be only 10% of the WT mRNA. All the model parameters are listed in supplemental table S2.

Electron microscopy of *E. coli*

Cells transformed with either pBEST-P70a-UTR1-GFP(WT) or pBEST-P70a-UTR1-GFP(Y35TAG) were fixed in 2.5% (v/v) glutaraldehyde in PBS (pH=7.0) and 1% (m/v) osmium tetroxide in PBS (pH=7.0). The cells were sequentially dehydrated in ethanol and embedded in araldite. Ultrathin sections were made using Leica UltraCut UCT ultramicrotome (Leica Microsystems GmbH, Austria) and stained with uranyl acetate followed by lead citrate. The ultrathin sections were imaged using Jeol JEM-1230 working at 120 kV (JEOL, Ltd., Peabody, MA).

***In vitro* translation**

The Pyl-OTS transformed C321 Δ prfA *E. coli* strains were subjected to a 30S cell extract protocol(Chemla et al., 2015). All strains were grown to O.D₆₀₀ 2.0 \pm 0.05 and then lysed according to the 30S cell extract protocol. Next, cell extracts were used for a cell free protein synthesis assay using the GFP harbored on the pBEST plasmid as the fluorescence reporter. The assay was conducted in Nunc 384-well plates with 120 μ L added to each well (Thermo Fisher Scientific, Waltham, MA) and was monitored using time dependent fluorescence measurements using a plate reader (Excitation 485 nm Emission 525 nm). A typical cell-free reaction assay consisted of 10 μ L of reaction mixture containing 33% (by volume) *E. coli* cell extract and 66% of the reaction volume was composed of the reaction buffer, which contained nutrients, metabolites and crowding agents. The

reporter plasmid (pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene #40019)) was added to final concentration of 2 nM. For detailed methodology please see (Chemla et al., 2015).

GFP stability assay

Two overnight cultures of C321.ΔA.exp expressing GFP WT and GFP Y35PrK, respectively, were prepared as previously described. The two cultures were lysed using BugBuster reagent (Novagen) according to manufacturer's protocol **in the absence of protease inhibitors**. The crude lysates were monitored for GFP fluorescence (480/510nm) over the course of 24 hours at 37°C in a 96 well plate (Nunc) using a Synergy HT plate reader (Biotek, Winooski, VT).

Promoter and UTR initiation rate assessment

Promoter initiation rate assessment was done in a 5-μL cell-free reaction that was incubated in a 96-well plate. Reactions contained 40 μM DFHBI-1T dye (specific to the Broccoli RNA aptamer), and linear PCR DNA at either 2 or 10 nM. The DNA contained the Broccoli aptamer (Filonov et al., 2014) downstream of a promoter of interest (P70nat, P70a, P70b). Broccoli fluorescence was calibrated using pure Broccoli produced in an *in vitro* transcription reaction. Fluorescence kinetics were recorded on a Biotek H1m plate reader. The transcription rates were determined as the linear slope of the kinetics. The ratio of the strength of each promoter was measured relative to that of the strong P70a. The plasmid containing P70a-Broccoli at 1 nM was put in reaction and the RNA synthesis rate was measured. For the other three promoters, we used the strength ratio of the linear DNA to correspond to the plasmid RNA synthesis rates. UTR initiation rates were calculated using a 5-μl cell-free reaction that was incubated in a 96-well plate at 29°C. The rate of GFP synthesis was measured for two identical plasmids except for the RBS (P70a-UTR1-GFP and P70a-UTR3-GFP) at 1 nM each.

DNA sequences

GFP DNA sequence

```
ATGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTTCAGCGTGTCGGCGAGGGCGAGGGCGATGCCACCTTACGGCAAGCTGACCCTGAAGTTCATC
TGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCA
GTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCG
TGCAGCTCGCCGACCACTTACCAGCAGAACACCCCATCGGCCGACGGCCCCGTGCTGCTGCCCGA
CAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
GTCTGCTGGAGTTCGTGACCGCCGCCGGGATCTCTAGAGTG CACCACCACCACCATCACGTGTA
A
```

GFP Protein sequence

MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFS
RYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQNTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMLLEFVTAAGISR V **HHHHHH**V

- ncAA incorporation sites are underlined (mutated into TAG nonsense codons).
- C-terminal 6xHis tag is highlighted in red.

mRFP1 DNA sequence

ATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTTTC AAAGTTCGTATGGAAGGTTCCGT
TAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTGCGTCCGTACGAAGGTACCCAGACC
GCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTGCTTGGGACATCCTGTCCCCGCAGTT
CCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCTT
TCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTGTTACCGTTACC
CAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCC
GTCCGACGGTCCGGTTATGCAGAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTAC
CCGGAAGACGGTGCTCTGAAAGGTGAAATCAAATGCGTCTGAAACTGAAAGACGGTGGTCACT
ACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAACCGGTTACGCTGCCGGGTGCTTACAA
AACCGACATCAAAGTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAA
CGTGCTGAAGGTGCTCACTCCACCGGTGCTTAA

- ncAA incorporation sites are underlined (mutated into TAG nonsense codon).

mRFP1 protein sequence

MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFQY
GSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPDGP
VMQKKTMGWEASTERMYPEDGALKGEIKMRLKLDGGHYDAEVKTTYMAKKPVQLPGAYKTDIK
LDITSHNEDYTIVEQYERAEGRHSTGA

Control regions

P70a-UTR1

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAGCTAGCAATA
ATTTTGTTTAACTTTAAGAAGGAGATATA

P70b-UTR1

TGAGCTAACACCGTGCGTGT AGACAATTTTACCTCTGGCGGTGATAATGGTTGCAGCTAGCAATA
ATTTTGTTTAACTTTAAGAAGGAGATATA

P70a-UTR3

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAGCTAGCAATA
ATTTTGTTTAACTTTAAGAA CGAGATATA

- The mutation used to generate the weaker control region from the P70a-UTR1 template is underlined.

Rate maps used in the model

WT *GFP*:

K16TAG protein L:

A BBBBBBBBAAAABA **TAG(D)** AABABABCABBAAAAACBBAAAAABAAAA
ABABBAABC AAA BAAA BAAA ABBAABAC stop(TAA)

WT ZADH:

A BBBBBBBBABBAAAABAABBAAAABABBACABAAAAAABBAAAABBBA
ABABABBAAAABA AAAAAA ABA AAAAAABABAAAABAAAABBAAAABA
AAAAAABBBAABBABBAAAABA AAAAAACB AABABA AABA ABA ABA
AAAAAABACBACBABAABBAAAABA AAAAAABABA AABA ABA ABA
AAAAABA AABABCABA AAAAAACA A ABBAAAABA AAAAAABA ABA
ABAAA BAAAAA A ABBABA AAAAAABBAC AAAAAABBBA BABA ABA
BBABBBA AAAAAABA AAAAAABBABA AAAAAABBBA AABA BABA
AAAAABBABBBA BAAAAABA AAAAAABA AAAAAACABA ABA ABA
ABBBAAAAAABA BAAAAABA AAAAAABA AAAAAABA ABA ABA ABA stop(TAA)

H297TAG ZADH:

A BBBBBBBBABBAAAABAABBAAAABABBACABAAAAAABBAAAABBBA
ABABABBAAAABA AAAAAA ABA AAAAAABABAAAABAAAABBAAAABA
AAAAAABBBAABBABBAAAABA AAAAAACB AABABA AABA ABA ABA
AAAAAABACBACBABAABBAAAABA AAAAAABABA AABA ABA ABA
AAAAABA AABABCABA AAAAAACA A ABBAAAABA AAAAAABA ABA
ABAAA BAAAAA A ABBABA AAAAAABBAC AAAAAABBBA BABA ABA
BBABBBA AAAAAABA AAAAAABBABA AAAAAABBBA AABA BABA
AAAAAABBABBBA BAAAAABA AAAAAABA AAAAAABA AAAAAACABA ABA ABA
AABABBBA AAAAAABA BABA AAAAAABA AAAAAABA ABA ABA ABA stop(TAA)

Supplemental tables

Table S1. Primers and sequences used in this study

#	Oligo name	Sequence	Comments	
1	Y35X F	GCGATGCCACCTAGGGCAAGCTGACCCTGAA	Y35TAG GFP mutant generation	
2	Y35X R	TTCAGGGTCAGCTTGCCCTAGGTGGCATCGC		
3	D193X F	CCCCGTGCTGCTGCCCTAGAACCACTACCTGAGCA	D193TAG GFP mutant generation	
4	D193X R	TGCTCAGGTAGTGGTTCTAGGGCAGCAGCACGGGG		
5	P70b F	GCTAACACCGTGCGTGTAGACAATTTACCTCTGG	Pr1 promoter variant generation	
6	P70b R	CCAGAGGTAAAATTGTCTACACGCACGGTGTTAGC		
7	GFP UTR3 F	TTGTTTAACTTTAAGAACGAGATATAACCATGGAGCT	UTR3 site directed mutagenesis for GFP	
8	GFP UTR3 R	AGCTCCATGGTATATCTCGTTCTTAAAGTTAAACAA		
9	PL K16TAG F	GGAAGAAGTAACAATCTAGGCTAACCTAATCTTTGC	K16TAG protein L mutant generation	
10	PL K16TAG R	GCAAAGATTAGGTTAGCCTAGATTGTTACTTCTTCC		
11	GFP mRNA quant F	ATGAAGCAGCACGACTTCTT	GFP specific primers for mRNA quantification	
12	GFP mRNA quant R	GTGGCTGTTGTAGTTGTACTC		
13	16S 1369F	CGGTGAATACGTTTCYCGG	16S rRNA endogenous control for mRNA quantification	
14	16S 1492R	GGWTACCTTGTTACGACTT		
15	35 TAC->TAT + 36 GGC-> GGA F	CGAGGGCGATGCCACCTATGGAAAGCTGACCCTGAAG	35 A->B + 36 A->C	Synonymous slow codons mutations
16	35 TAC->TAT + 36 GGC-> GGA R	CTTCAGGGTCAGCTTTCATAGGTGGCATCGCCCTCG		
17	34 ACC->ACA F	GGGCGAGGGCGATGCCACATATGGAAAGCTG	34 A->B	
18	34 ACC->ACA R	CAGCTTTCATATGTGGCATCGCCCTCGCCC		
19	32 GGC->GGA F	CCGGCGAGGGCGAGGGAGATGCCACATATG	32 A->C	
20	32 GGC->GGA R	CATATGTGGCATCTCCCTCGCCCTCGCCGG		

Table S1. DNA oligonucleotides used in this study for vector construction, mutant generation and real-time PCR.

Table S2. Parameters used in the model

Parameters used in this study	Notation	Value
Initiation rate of P70a promoter (strong)	R_a^S	3.0/s
Initiation rate of P70b promoter (weak)	R_a^W	0.2/s
Elongation rate of polymerase	R_0	36 nt/s
Termination rate of polymerase	R_β	3.0/s
Initiation rate of UTR1 (strong)	r_a^S	1.5/s
Initiation rate of UTR3 (weak)	r_a^W	0.15/s
Elongation rate of ribosome A (fast)	r_A	35.0 codons/s
Elongation rate of ribosome B (medium)	r_B	8.0 codons/s
Elongation rate of ribosome C (slow)	r_C	4.5 codons/s
Elongation rate of ribosome D (UAG - very slow)	r_D	0.04 codons/s
Termination rate of ribosome	r_β	1/s
Size of ribosome	L_r	12 codons
Size of polymerase	L_p	51 nt
mRNA half-life time	$t_{1/2}$	60 seconds
Proportionality constant lambda	λ	0.11 nm/codon

Table S2. Constants and parameters used in the model. The sources and experiments resulted in these values are described in the supplemental experimental procedures section.

References

- Arndt, K.M., and Chamberlin, M.J. (1988). Transcription termination in *Escherichia coli*. *J. Mol. Biol.* *202*, 271–285.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The Native 3D Organization of Bacterial Polysomes. *Cell* *136*, 261–271.
- Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V., and Alfanta, L. (2015). Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* *112*, 1663–1672.
- Fan, C., Xiong, H., Reynolds, N.N.M., and Söll, D. (2015). Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* *43*, e156.
- Filonov, G.S., Moon, J.D., Svendsen, N., and Jaffrey, S.R. (2014). Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* *136*, 16299–16308.
- Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* *81*, 2340–2361.
- Gritsenko, A.A., Hulsman, M., Reinders, M.J.T., and de Ridder, D. (2015). Unbiased Quantitative Models of Protein Translation Derived from Ribosome Profiling Data. *PLoS Comput. Biol.* *11*, e1004336.
- Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., and Kinosita, K. (2001). Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase. *Nature* *409*, 113–115.

- Hawley, D.K., and McClure, W.R. (1980). In vitro comparison of initiation properties of bacteriophage lambda wild-type PR and x3 mutant promoters. *Proc. Natl. Acad. Sci. U. S. A.* *77*, 6381–6385.
- Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* *15*, 205–213.
- Kennell, D., and Riezman, H. (1977). Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J. Mol. Biol.* *114*, 1–21.
- Klumpp, S., and Hwa, T. (2008a). Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 20245–20250.
- Klumpp, S., and Hwa, T. (2008b). Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 18159–18164.
- Liang, S.T., Ehrenberg, M., Dennis, P., and Bremer, H. (1999). Decay of rplN and lacZ mRNA in *Escherichia coli*. *J. Mol. Biol.* *288*, 521–538.
- Mathews, D.H., and Turner, D.H. (2002). Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.* *317*, 191–203.
- Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 7287–7292.
- McClure, W. (1983). biochemical analysis of the effect of RNA polymerase concentration on the in vivo control of RNA chain initiation frequency. In *Biochemistry of Metabolic Processes*, D. Lennon, F. Stratman, and R. Zahltan, eds. (Elvesier), pp. 207–217.
- Mitarai, N., and Pedersen, S. (2013). Control of ribosome traffic by position-dependent choice of synonymous codons. *Phys. Biol.* *10*, 056011.
- Morse, D.E., and Yanofsky, C. (1969). Polarity and the Degradation of mRNA. *Nature* *224*, 329–331.
- Pedersen, S., and Reeh, S. (1978). Functional mRNA half lives in *E. coli*. *Mol. Gen. Genet.* *166*, 329–336.
- Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* *328*, 504–508.
- Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* *4*, 8.
- Vogel, U., and Jensen, K.F. (1994). The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.* *176*, 2807–2813.
- Wang, J., Kwiatkowski, M., and Forster, A.C. (2015). Kinetics of tRNA Pyl-mediated amber suppression in *Escherichia coli* translation reveals unexpected limiting steps and competing reactions. *Biotechnol. Bioeng.* *113*, 1552-1559.
- Zaychikov, E., Denissova, L., and Heumann, H. (1995). Translocation of the *Escherichia coli* transcription complex observed in the registers 11 to 20: “jumping” of RNA polymerase and asymmetric expansion and contraction of the “transcription bubble”. *Proc. Natl. Acad. Sci.* *92*, 1739–1743.

Constants that were used in the simulations:

Tuning of recombinant protein expression in *Escherichia coli* by manipulating transcription, translation initiation rates and incorporation of non-canonical amino acids

Orr Schlesinger,^{†,#} Yonatan Chemla,^{†,#} Mathias Heltberg,^{‡,#} Eden Ozer,[†] Ryan Marshall,[§] Vincent Noireaux,[§] Mogens Høgh Jensen,^{*,‡} and Lital Alfonta^{*,†}

[†]Department of Life Sciences and Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, POBox 653, Beer-Sheva 8410501, Israel

[‡]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

[§]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, United States

*Correspondence: mhjensen@nbi.ku.dk (M.H.J.), alfontal@bgu.ac.il (L.A)

These authors contributed equally.

This is an h file that defines the constants that are used in the simulation file. The user should define the length of the protein of interest (in amino acids/codons) in the "#define L" row.

```
#ifndef NEWAPPROACHPARAMETERS_H
#define NEWAPPROACHPARAMETERS_H

#include <math.h>

// Defining the sphere
#define RibMax 300
#define PolMax 300
#define L 234 //INPUT LEGTH OF PROTEIN (IN AMINO ACIDS)

// Defning objects
// Objects regarding Transcription
double raPL; double PrThr;
double ProdTime[200][15];
int reactStat [5];
int Ribos[RibMax][PolMax]; double UpdRib[RibMax][PolMax]; int
NewRibos[RibMax][PolMax]; double NewUpdRib[RibMax][PolMax];
int CollidRib[RibMax][PolMax];int NewCollidRib[RibMax][PolMax];
int Polys[PolMax]; double UpdPol[PolMax];
double RatDna[L]; double RatRna[L];
int Nri[PolMax]; int Nst[PolMax];
int FiCou[PolMax][5];
double AgDec[PolMax];
double SumDe; double SumRn; double SumRb; double SumTot; double
SumStart; double SumDeAgg;
```

```
// General Parameters
double phi;double Time;double NuOn;
double NuOnPr; double NuOff; double NuRiOn; double NuRiOff;
double Tag;double lambda;

// Helping objects
double LifeTimer; double LifeTimes;
int ProUn; int ProCo; int cUn; int cPro;
int DistRibCalc;int DistPolCalc;
int UnCol;int SumCol;int ColMr;
int Avgfmr;int Avgapo;
int apo;int fmr;
int DistPol; int DistRib;
double A; int Astop;
double FiPro; int FimRNA; int DemRNA;
double ColPro; int collcount;
double randT;
double RT; double RT0;
int click; int counter; int coufmr; double checker; int react;
#endif
```

Simulation Code

Tuning of recombinant protein expression in Escherichia coli by manipulating transcription, translation initiation rates and incorporation of non-canonical amino acids

Orr Schlesinger,^{†,#} Yonatan Chemla,^{†,#} Mathias Heltberg,^{‡,#} Eden Ozer,[†] Ryan Marshall,[§] Vincent Noireaux,[§] Mogens Høgh Jensen,^{*,‡} and Lital Alfonta^{*,†}

[†]Department of Life Sciences and Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, POBox 653, Beer-Sheva 8410501, Israel

[‡]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

[§]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, United States

*Correspondence: mhjensen@nbi.ku.dk (M.H.J), alfontal@bgu.ac.il (L.A)

These authors contributed equally.

This is a c⁺⁺ code generated to simulate bacterial transcription and translation in a stochastic way. This file includes all the parameter values used in the manuscript. The parameters can be altered to explore other genes and regulatory elements. The first string list in the code refers to the different GFP conditions examined in this study. The specific string can be chosen using the "int type" variable. Each run of the script explores one of those conditions under a defined parameter regime and repeats it as many times as defined by the TRI < value. The results are presented in a sequence according to the 11 columns mentioned in the code. The time of the reaction before the promoter shuts off can be defined by the "time" variable. The main parameters to be changed by the user are the following:

- 1) Size of the ribosome footprint in codons = DistRib
- 2) Size of the RNAP footprint in nucleotide triplets = DistPol
- 3) Promoter initiation rate (initiation/sec)= NuOnPr
- 4) RBS initiation rate (initiation/sec)= NuRiOn
- 5) Promoter termination rate (initiation/sec) = NuOff
- 6) Ribosome termination rate (initiation/sec) = NuRiOff

7) UAG codon rate of translation (codons/sec) = Tag

8) mRNA half life time (seconds) = phi

9) The probability of a DITA event occur once conditions are met = PrThr

A map of the protein codons rates should be provided in the file folder and inserted in the "std::ifstream readFile" frame.

Lastly, inside the loops that describe each of the conditions defined earlier the user can choose and change the location of the UAG codon, if relevant. Please note that as mentioned in the manuscript the mRNA half life time changes slightly with the addition of each UAG codon (our results suggest something in the order of 5% reduction for each UAG codon added).

```
#include "NewApproachParameters.h"
#include <math.h>
#include <stdlib.h>
#include <time.h>
#include <fstream>
#include <ctype.h>
#include <iomanip>
#include <iostream>
#include <sstream>
#include <string>
#include <cmath>
using namespace std;

int main(){
    srand (time(0));
    int tal = 401;

    string list[] = {"WT",
"WeakPromo", "WeakRibo", "1Mut", "2Mut", "3Mut", "1Mut193", "1MutWeakPr
o", "1MutWeakRibo"};
    // This should be used for statistics over entire runs. It contains
11 columns:
    // 1) NuOnPr, 2) lambda 3) Avarage number of finished mRNA
templates, 4) Avarage number of transcribed templates, 5) Avarage
distance between polymerases
    // 6) #Proteins, 7) #Unaggregated Proteins 8) # of Aggregated
proteins, 9) # of produced mRNAs, 10) # of uncollided mRNAs, 11) #
of collided mRNAs
    int type = 8;
    ostringstream filename1;
    filename1 << "deGFP_Final_";
    filename1 << list[type];    filename1 << ".txt";
    std::ofstream TriFile (filename1.str().c_str());
```

```

// This saves the number of proteins at different times
ostream filename2; filename2 << "deGFP_"; filename2 <<
list[type]; filename2 << ".txt";std::ofstream SpeedFile
(filename2.str().c_str());

for (unsigned int TRI = 0; TRI < 3; TRI++){ //////////////// Starts
different trials ////////////////

// Initialize objects and variables ///
counter = 0; FiPro = 0; ColPro = 0; RT = 0; RT0 = 0;
click = 0; checker = 0; coufmr = 0;
SumRb = 0; SumRn = 0; apo = 0; fmr = 0; SumDe = 0; SumStart = 0;
FimRNA = 0; DemRNA = 0; Avgfmr = 0; DistRibCalc = 0; DistPolCalc =
0; Avgapo = 0;
UnCol = 0; ColMr = 0; ProCo = 0; ProUn = 0; RT0 = 0;
LifeTimes = 0.0; cUn = 0; cPro = 0; react = 0; collcount = 0;
int unfmr = 0;
Time = 2000;

////////////////////// THESE ARE THE MAIN
PARAMETERS TO BE CHANGED: ////////////////////////
DistRib = 12; DistPol = 12; lambda = 0.333;
NuOnPr = 3.00; NuRiOn = 1.5; NuRiOff = 1;
NuOff = 3.0; Tag = 0.04; phi = log(2.0)/60.0;
PrThr = 1;
std::ifstream readFile ("WTMAPNumbers.txt"); //This is where you
should add you protein TXT file, wheres every codon speed is stated
in seconds followed by 1 space.A group codons will be written as 35.0
, B group codons as 8.0 and C group as 4.5//
////////////////////// ACTUAL SETTINGS FOR THE
RUN ////////////////////////
double a; int c = 0;
while (readFile >> a){
RatRna[c] = a;
c++;}

if (type == 1){ //insert promoters
NuOnPr = NuOnPr/15;
}
else if (type == 2){
NuRiOn = NuRiOn/10.0;
}
else if (type == 3){
phi = phi/0.95;
RatRna[34] = Tag;
}
else if (type == 4){ // insert mutations
phi = phi/0.90;
RatRna[34] = Tag;
RatRna[192] = Tag;
}
}

```

```

else if (type == 5){
    phi = phi/0.80;
    RatRna[34] = Tag;
    RatRna[178] = Tag;
    RatRna[192] = Tag;
}
else if (type == 6){
    phi = phi/0.95;
    RatRna[192] = Tag;
}
else if (type == 7){
    phi = phi/0.95;
    RatRna[34] = Tag;
    NuOnPr = NuOnPr/14;
}
else if (type == 8){
    phi = phi/0.1;
    NuRiOn = NuRiOn/10.0;
    RatRna[34] = Tag;
}

double avgttime = 1.0/NuRiOn;
avgttime += 1.0/NuRiOff;
for (unsigned int i = 0; i < L; i++){
    avgttime += 1.0/(double)(RatRna[i]);
}
///// initialize the rates for the ploymerase
for (unsigned int k = 0; k < L; k++){
    RatDna[k] = 17.0; }
// Initialize the objects
for (unsigned int i = 0; i < PolMax; i++){
    Polys[i] = -1;
    UpdPol[i] = 0;
    Nri[i] = 0;
    FiCou[i][0] = 0;    FiCou[i][1] = 0;    FiCou[i][2] = 0;
FiCou[i][3] = 0;    FiCou[i][4] = 0;
    for (unsigned int j = 0; j < RibMax; j++){
        Ribos[j][i] = -1;
        CollidRib[j][i] = 0;
        NewCollidRib[j][i] = 0;
        UpdRib[j][i] = 0;}}
for (unsigned int k = 0; k < 5; k++){ reactStat[k] = 0;}

```

```

////////////////////////////////////
STARTING TIME

```

```

////////////////////////////////////
std::cout << "-> Run Number " << TRI << " R_alfa " << NuOnPr <<
" r_alfa " << NuRiOn << " Phi " << phi << "Time to finish mRNA " <<
avgtime << std::endl;

while (RT < Time) {
    counter++;
    if (RT > click*10.0){ //Save to datafile
        ProdTime[click][TRI*3] = RT;          ProdTime[click][TRI*3 + 1]
= FiPro; ProdTime[click][TRI*3 + 2] = unfmr;
        click++;
    }

    // Checking if we can insert a new polymerase
    if (Polys[0] > DistPol || Polys[0] == -1){
        NuOn = NuOnPr;}
    else if (Polys[0] <= DistPol){
        NuOn = 0;}
    if (RT > 1700){
        NuOn = -NuOn*(RT - RT0);
        if (NuOn < 0){NuOn = 0;}
    }

////////////////////////////////////
//
//      randT = rand() / double(RAND_MAX); // Updating time
//
//      SumTot = SumStart + SumRb + SumRn + SumDe + NuOn;
//      RT0 = RT;
//      RT = RT - log(randT)/(SumTot);
//
//      randT = rand() / double(RAND_MAX);
//      A = 0;      Astop = 0;
//      A = A + NuOn/SumTot;
//
////////////////////////////////////
//
//
//      react = 0;
//      /////////////////////////////////// Choosing next reaction
////////////////////////////////////
////////////////////////////////////
//      if (A > randT){ // 1) Insert RNA - polymerase
//      react = 1;
//      reactStat[react-1]++;
//      Astop++;
//      for (unsigned int i = 0; i < apo + fmr; i++){
//          Polys[apo + fmr - i] = Polys[apo + fmr - 1 -i];
//          UpdPol[apo + fmr - i] = UpdPol[apo + fmr - 1 -i];

```



```

    }
  }
}

if (Astop == 0){ // 4) Insert New Ribosome
for (unsigned int i=0; i<apo+fmr; i++){
  if (Polys[i] > DistRib && (Ribos[0][i]==-1 || Ribos[0][i] >
DistRib) ){
    A = A + NuRiOn/SumTot;
  }
  if (A > randT){
    react = 4;
    reactStat[react-1]++;
    for (unsigned int j=0; j<Nri[i]; j++){
      Ribos[Nri[i]-j][i] = Ribos[Nri[i]-1-j][i];
      CollidRib[Nri[i]-j][i] = CollidRib[Nri[i]-1-j][i];
      UpdRib[Nri[i]-j][i] = UpdRib[Nri[i]-1-j][i];
    }
    Ribos[0][i] = 0;
    Nri[i]++;
    Astop++;
    break;
  }
}
}
}

```

```

if (Astop == 0){ // 5) Move Ribosome
for (unsigned int i = 0; i < (apo + fmr); i++){
  if (Astop == 1){
    break;}
  for (unsigned int j1 = 0; j1 < Nri[i]; j1++){
    A = A + UpdRib[j1][i]/SumTot;
    if (A > randT){
      react = 5;
      reactStat[react-1]++;

      Ribos[j1][i]++;
      if (i < apo-1){
        for (unsigned int j2 = 0; j2 < Nri[i+1]; j2++){
          if(( Polys[i+1]-Ribos[j2][i+1] == Polys[i] -
Ribos[j1][i]) && (j1 < Nri[i]) ){
            if ((j2 < Nri[i+1]) &&
((double)(Ribos[j2][i+1])*lambda*0.3 + 8.0 >
(double)(Polys[i+1]-Polys[i]))) {
              raPL = rand() / double(RAND_MAX);
              if (raPL < PrThr){
                collcount++;
                CollidRib[j1][i] = 1;
                CollidRib[j2][i+1] = 1;
              }
            }
          }
        }
      }
    }
  }
}
}

```

```

        }
    }
    }
    Astop++;
    break;
}
}
}
// This ends the choice of events
//

// Debugging
if (Astop == 0){
    RT = Time;
    std::cout << " PROBLEM with Astop STILL ZERO!!" << randT << "
A " << A << std::endl;}
    if (A > 1.000001){
        RT = Time;
        std::cout << " PROBLEM with A TOO LARGE!!!" << randT << " A "
<< A << std::endl;}

// Updating all rates
//
// Updating Polymerases
SumRn = 0;
for (unsigned int i = 0; i < apo; i++){
    SumCol = 0;
    if (Polys[i+1] <= Polys[i] + DistPol && Polys[i+1] != -1 &&
Polys[i+1] != L+1){
        UpdPol[i] = 0;}
    else if (Polys[i+1] >= Polys[i] + DistPol || Polys[i+1] == -1){
        UpdPol[i] = RatDna[Polys[i]];
    }
    if (Polys[i] == L-1){ // Entering last step
        UpdPol[i] = NuOff;
    }
    if (Polys[i] == L){ // Entering category as finished mRNA
        FiCou[i][2] = RT;

        if (FiCou[i][1] == 0){
            UnCol++;
            unfmr++;
        }

        else if (FiCou[i][1] == 1){
            ColMr++;}

        Polys[i] = L+1;
        apo = apo -1;
        FimRNA++;

```

```

    fmr++;
    UpdPol[i] = 0;
    }
    SumRn = SumRn + UpdPol[i];
}

////////// Updating decay rate
SumDe = phi*fmr;    SumRb = 0;    SumStart = 0;

////////// Starting Updating Ribosome rates
for (unsigned int i = 0; i < (apo + fmr); i++){ // looping over
all Polymerases
    if (Polys[i] > DistRib && (Ribos[0][i] == -1 || Ribos[0][i] >
DistRib)){
        SumStart = SumStart + NuRiOn;}
        for (unsigned int j = 0; j < Nri[i]; j++){ //// Looping over
ribosomes at template i
            // Debugging
            if (Ribos[j][i] > Ribos[j+1][i] && Ribos[j+1][i] != -1 &&
Ribos[j][i] != -2 && Ribos[j+1][i] != -2){
                RT = Time;    std::cout << " PROBLEM some greater than
another !" << std::endl;    }
                if (Ribos[j][i] == -1){
                    RT = Time;    std::cout << " PROBLEM Some at -1 !" <<
std::endl; }

                //Actual updates
                if (Ribos[j+1][i] <= Ribos[j][i] + DistRib && Ribos[j+1][i] !=
-1){
                    UpdRib[j][i] = 0;}
                    else if (Ribos[j+1][i] > Ribos[j][i] + DistRib || Ribos[j+1][i]
== -1){
                        UpdRib[j][i] = RatRna[Ribos[j][i]];
                    }
                    if (Ribos[j][i] == Polys[i]-DistRib && Polys[i] != L+1){
                        UpdRib[j][i] = 0;
                    }

                    if (CollidRib[j][i] > 0){
                        UpdRib[j][i] = 0;}

                    if (Ribos[j][i] == L-1){
                        UpdRib[j][i] = NuRiOff;}

                // Producing proteins
                if (Ribos[j][i] == L){
                    if (FiCou[i][1] == 0) {
                        cUn++;
                    }
                    else if (FiCou[i][1]==1){
                        cPro++;
                    }
                }

```

```

else{
    std::cout << "Marker is " << FiCou[i][1] << std::endl;
}

Ribos[j][i] = -1;
UpdRib[j][i] = 0;
Nri[i] = Nri[i]-1;
FiPro++;
FiCou[i][0]++;
}
SumRb = SumRb + UpdRib[j][i];
}
}
////////////////////////////////////// Ends Ribosome rate update
if (RT > 0.25*Time && apo > 1){
    coufmr++;
    Avgfmr = Avgfmr + fmr;
    Avgapo = Avgapo + apo;
    DistPolCalc = DistPolCalc + ((double) (Polys[apo-1]) -
(double) (Polys[0]))/(double) (apo-1);
}

////////////////////////////////////// All rates should now be
updated
SumTot = SumStart + SumRb + SumRn + SumDe + NuOn;
if (SumTot <= 0){
    RT = Time;
}

} //////////////////////////////////////// This ends time
//////////////////////////////////////
// Adding all proteins that are being translated at present.
Should this be done???
for (unsigned int i = apo; i < fmr + apo; i++){
    if (FiCou[i][1] == 0){
        ProUn = ProUn + FiCou[i][0];
    }
    else if (FiCou[i][1] == 1){
        ProCo = ProCo + FiCou[i][0];
    }
}
////////////////////////////////////// Printing after script
//////////////////////////////////////
std::cout << "* Finished Protein: " << FiPro << std::endl;
std::cout << "* Uncilloded mRNA: " << UnCol << " Produced from
" << (double) (ProUn)/(double) (UnCol) << std::endl;
std::cout << "* Number of available mRNA " << fmr << " Actice
polymerases " << apo << endl;
std::cout << "* Finished mRNA " << FimRNA<< " Decayed " << DemRNA
<< " counter " << counter << std::endl;
std::cout << "* Collided mRNAs " << ColMr << " Produced from "

```

```

<< (double) (ProCo) / (double) (ColMr) << std::endl;
    std::cout << "* Avarage <D> " <<
    (double) (DistPolCalc) / (double) (coufmr) << " Avarage number of
polymerases " << (double) (Avgapo) / (double) (coufmr) << endl;
    std::cout << "* Avarage available mRNA " <<
    (double) (Avgfmr) / (double) (coufmr) << " LifeTimes " <<
LifeTimes / FiPro << std::endl;
    std::cout << "* Real Un = " << cUn << std::endl;
    std::cout << "* Real Coll = " << cPro << std::endl;
    std::cout << "* Total number of collisions " << collcount <<
std::endl;
    /// Saves to STAT file
    TriFile << NuOnPr << "\t" << NuRiOn << "\t" << Tag << "\t" << lambda
<< "\t" << (double) (Avgfmr) / (double) (coufmr) << "\t" <<
    (double) (Avgapo) / (double) (coufmr) << "\t" <<
    (double) (DistPolCalc) / (double) (coufmr) << "\t" << FiPro << "\t" <<
ProUn << "\t" << ProCo << "\t" << FimRNA << "\t" << UnCol << "\t" <<
ColMr << "\n";

    } // This ends the different trials

    /// Saves to Production file
    for (unsigned int isav = 0; isav < 200; isav++){
        for (unsigned int tsav = 0; tsav < 10; tsav++){
            SpeedFile << ProdTime[isav][tsav] << "\t";
        }
        SpeedFile << "\n";
    }
} // This ends main

/* Description of objects
FiCou[i][0] -> number of finished proteins from mRNA template i
FiCou[i][1] -> Marker for wheather an aggregation has occured at this
template
FiCou[i][2] -> Gives the time for when the template was finished. To
check the lifetimes of mRNAs.
FiCou[i][3] -> Highest position of aggregation on the template
FiCou[i][4] -> Number of ribosomes in front of the aggregation

CollidRib[i][j] -> Marker of the ribosome of ij'th postion is blocked.
Ribos[j][i] -> the position of the j'th ribosome, on mRNA template
number i.
UpdRib[j][i] -> the rate of reaction for ribosome j on template i.
Polys[i] -> The position of the i'th mRNA polymerase

*/

/*          //Checking and printing all polymerases
std::cout << "Ribosomes " << std::endl;
for (unsigned int k1 = 0; k1 < apo + fmr; k1++){

```

```

for (unsigned int k2 = 0; k2 < Nri[k1]; k2++){
    std::cout << Ribos[k2][k1] << " ";
}
std::cout << " " << std::endl;
}
std::cout << "Collided " << std::endl;
for (unsigned int k1 = 0; k1 < apo + fmr; k1++){
for (unsigned int k2 = 0; k2 < Nri[k1]; k2++){
    std::cout << CollidRib[k2][k1] << " ";
}
std::cout << " " << std::endl;
}
std::cout << "Updates " << std::endl;
for (unsigned int k1 = 0; k1 < apo + fmr; k1++){
    for (unsigned int k2 = 0; k2 < Nri[k1]; k2++){
        std::cout << UpdRib[k2][k1] << " ";
    }
    std::cout << " " << std::endl;
}
std::cout << "Polys " << std::endl;
for (unsigned int k1 = 0; k1 < apo; k1++){
    std::cout << Polys[k1] << " ";}
std::cout << " " << std::endl;
std::cout << "Finished " << std::endl;
for (unsigned int k1 = apo; k1 < apo + fmr; k1++){
    std::cout << Polys[k1] << " ";}
std::cout << " " << std::endl;
std::cout << "Updates " << std::endl;
for (unsigned int k1 = 0; k1 < apo; k1++){
    std::cout << UpdPol[k1] << " ";}
std::cout << " " << std::endl;
*/

```