SURVEY AND SUMMARY

Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution

Hadas Zur[1] and Tamir Tuller[1,2,*]

[1]Department of Biomedical Engineering, the Engineering Faculty, Tel Aviv University, Tel-Aviv 69978, Israel and [2]The Sagol School of Neuroscience, Tel Aviv University, Tel-Aviv 69978, Israel

ABSTRACT

mRNA translation is the fundamental process of decoding the information encoded in mRNA molecules by the ribosome for the synthesis of proteins. The centrality of this process in various biomedical disciplines such as cell biology, evolution and biotechnology, encouraged the development of dozens of mathematical and computational models of translation in recent years. These models aimed at capturing various biophysical aspects of the process. The objective of this review is to survey these models, focusing on those based and/or validated on real large-scale genomic data. We consider aspects such as the complexity of the models, the biophysical aspects they regard and the predictions they may provide. Furthermore, we survey the central systems biology discoveries reported on their basis. This review demonstrates the fundamental advantages of employing computational biophysical translation models in general, and discusses the relative advantages of the different approaches and the challenges in the field.

INTRODUCTION

Proteins are principal actors in all intra-cellular activities. The protein coding potential inscribed in a species' DNA is converted into proteins through the fundamental cellular process of gene expression, the two major cellular biophysical stages of which are the transcription of the gene to mRNA molecules by the RNA polymerase, and the translation of the mRNA molecules to proteins by the ribosome.

Surprisingly, generally there is only a limited degree of correlation between the levels of mRNAs and their encoded proteins (1–4), emphasizing the significant effect of post-transcriptional regulation on protein levels.

Furthermore, it was shown that translation and its efficiency have substantial effect on the organismal fitness, as it is the intracellular process with the highest energy consumption (5–7). Notably mRNA translation is a central intracellular process which can affect, be related, or interleaved with all central biological phenomena including protein folding, mRNA degradation, ribosomal collisions, abortion and allocation, transcript evolution, tumorigenesis, amongst others (7–23).

In recent years, various large scale techniques for estimating variables related to mRNA translation have been developed (1,3–4,24–41). Thus, in the recent and following years computational models of translation on the basis of high-throughput, genome-wide datasets, are expected to be instrumental in deciphering this process (1–2,7,12,19,42–70). Specifically, a central challenge in the field is to infer computational/mathematical *biophysical* models that consider various aspects of translation based on large scale data.

Indeed in recent years, dozens of studies aimed at modelling various aspects of the translation process. These models differ (and sometimes overlap) in their resolution, complexity, running times, assumptions, and more. Additionally, some of these models have been used for uncovering various fundamental biological phenomena.

The objective of this review is to report the advances in the field of computational large scale biophysical modelling of translation which are based on biological data. We first briefly review large scale biological experiments that can be used for inferring such models, and define what a biophysical model is. The main aim is to cover aspects such as the complexity of the models, their running time, the biological phenomena they uncover, and open challenges in the field. We believe that this is the first comprehensive review on this topic; thus, it is expected to be helpful to researchers studying translation via a quantitative manner.

It is important to mention that while there are some thorough previous reviews related to the field or to per-

*To whom correspondence should be addressed. Tel: +972 3 6405836; Fax: +972 3 6405836; Email: tamirtul@post.tau.ac.il

tinent topics (see, for example, (71,72)), here we focus on various aspects that haven't been comprehensively covered/examined before:

First, we review only models that consider biophysical aspects of the process. Second, we focus on studies that connect the model(s) to real large scale biological measurements and/or provide predictions related to translation of real genes. Third, we present some specific examples of the models' parameter estimation. Fourth, we consider/discuss the complexity of the models and their running times (when possible), and compare the different models based on these measures. Finally, we review systems biology results suggested based on these models.

We believe that the review will be helpful for a very diverse audience. For example, computational/synthetic biologists reading the review will mainly learn about the open questions in the field, the advantages related to each model, and the data that can be used with these models. Systems biologists / genomics researchers will mainly learn about the type of questions that can be answered via these models. Experimentalists (e.g. people studying RNA, gene regulation, and molecular biology) are expected to gain some basic knowledge regarding TASEP based models, and what they can afford them.

Large scale experimental data related to translation

Data related to translation is a crucial factor in accelerating the development of predictive modeling of this process, and for improving its understanding. Thus, we begin with a brief description of the experimental technologies in the field.

Specifically, these data are used for inferring various free parameters of the models and evaluating their performances. Today, it is rather easy to perform large scale measurements of mRNA levels (73) while, for technical reasons, the technologies for performing large scale measurements of protein abundance lag behind. For example, the GEO database includes hundreds of thousands of large scale measurements of mRNA levels, whilst there are only a few such large scale measurements of protein abundance (1,3,29–30) (there are at least two orders of magnitude less measurements of post-transcriptional gene expression stages/aspects, for example, protein levels, translation rates, degradation rates of proteins and mRNA molecules, etc). Thus, researchers are frequently 'forced' to use mRNA levels, the rather rough proxy of protein abundance, instead of the protein abundance itself. Concurrently, technologies to measure various aspects related to translation are now rapidly emerging, transforming our understanding of this process (1,3,24–40). Among others, these include proteomic techniques to measure absolute protein levels (1,3,29–30,40,74), tRNA levels (7,26,39,75–76), mRNA folding (33), elongation dynamics (27), and fractionation of mRNA-ribosome complexes (polysomes) to estimate not only the number of ribosomes on each mRNA (24,77–78), but even the positions of individual ribosomes on each mRNA (25,34,79–91) (using an experimental technique called Ribo-Seq), translation initiation positions (92,93), mRNA degradation rates (36,37), protein degradation rates (94,95), and co-translational protein folding (96,97).

Most of the large scale measurements related to mRNA translation are based on endogenous genes.

The main disadvantage of analyzing endogenous genes is the fact that these genes are shaped by evolution, thus, hindering the ability to understand the causality of the observed correlation between various variables related to translation efficiency (40,98–100). For example, ribosomal proteins have certain amino acid contents which relate to their function, and are known to be highly expressed. As each amino acid is encoded by specific codons they induce/affect various properties of the transcript (e.g. codon usage bias, GC content, mRNA folding). Thus, by analyzing such endogenous genes we may wrongly deduce that these transcript features affect translation efficiency. Furthermore as the different determinants of translation efficiency are correlated (e.g. highly expressed genes have both more efficient codons and more efficient initiation signals) and are under selection, it is difficult to evaluate the contribution of each of them to the translation speed (100).

An additional experimental strategy for studying the variables affecting protein levels and their regulation is via synthetic biology and heterologous gene expression. This approach includes manipulating the nucleotides of a transcript or parts of it, and/or measuring them out of their original endogenous setting, to understand the way translation aspects are encoded in the transcript. Recently, new datasets of protein abundance measurements of heterologous genes have emerged (40,99,101–104). Each of these libraries includes many versions of proteins; each such version includes an amino acid sequence identical to the original gene but with a different set of codons (i.e. synonymous modifications). Thus, by analyzing such datasets it is possible to overcome the two aforementioned problems (40,46,102). The heterologous genes datasets, on the other hand, may have two major problems of their own. First, in order to derive general conclusions about translation, many such datasets of heterologous genes based on numerous amino acid sequences should be generated (significantly more than those existing today). Second, in many cases heterologous genes are expressed in extreme and unnatural regimens, complicating the generalization of their analyses results to endogenous genes (6). Thus, these datasets instigate cautious analysis as well.

Together, these advances provide an enhanced account of basic cellular processes, with extensive implications for our understanding of the regulatory signals encoded in the transcript (e.g. the nature of codon biases), biophysical properties of transcripts, dynamics of translation, and transcript evolution.

However, it is important to mention that today there are no direct large scale in vivo measurements related to many fundamental aspects of translation. For example, we cannot directly measure initiation rates, codon decoding rates, ribosome abortion rates, etc. Thus, one challenge in the field is the inference of such variables from the available data mentioned above. In addition, most of the relevant experimental datasets include (sometimes extreme) non trivial biases (see, for example, (105)); thus, another fundamental challenge, interleaved with the objective of developing biophysical translation models, is the accurate filtering/handling of these biases.

The biophysical vs. machine learning approaches

As aforementioned, we survey the computational biophysical approaches for modeling mRNA translation based on these data. At this stage we would like to briefly mention that there are two major approaches for modelling/predicting translation, the machine learning approach and the biophysical approach. The biophysical approach is usually based on predictive simulations that are inspired by biophysical understanding of the studied processes. These simulations include biophysical phenomena such as diffusion, interactions between particles, folding of macro-molecules, etc. The machine learning approach, on the other hand, is based on statistical predictive inference of relations between sequence features and gene expression aspects, and it does not necessarily require prior knowledge of the biophysical gene expression mechanisms.

Machine learning is the more 'traditional' approach for practically predicting large scale gene translation, protein levels, or related variables (1–3,45,102,106–107). These models weigh the value of all the different variables to generate a prediction of, for example, protein levels. The main advantage of this approach is that it is relatively simple, no prior knowledge related to the biophysics of translation is required, and no knowledge related to the direction/causality of the relations is needed. This approach can often yield very significant predictions and improve our understanding regarding the transcript features relevant to translation efficiency.

The biophysical approach on the other hand, which captures the dynamics and physical nature of the process, is frequently used in ('simple') simulations without analyzing and modeling large scale biological data (108–111). The major advantage of the biophysical approach is the fact that it enables a better *understanding* of the biophysics of translation and its different regulatory stages than the machine-learning approach. Though theoretical physical models and simulations related to translation have been suggested nearly five decades ago (109,112), only recently have such approaches been implemented on real large-scale genomic data. Translation or protein synthesis consists of a complex system of chemical reactions, which ultimately results in the decoding of the mRNA and the production of a protein. The complexity of this reaction system makes it difficult to quantitatively connect its input parameters (such as translation factors, ribosome concentrations, tRNA concentrations, mRNA codon composition or energy availability) to output parameters (such as protein synthesis rates or ribosome densities) (71).

The simplest comprehensive (whole-cell) biophysical model of translation should consider the following aspects:

1. A (possibly) different speed/time for each codon, which is related to the local biophysical properties of the mRNA (e.g. its folding and the encoded amino acids), and their interaction with trans factors and/or the availability of such trans factors (e.g. tRNA levels).
2. Initiation rates (which are affected by the properties of the mRNAs and initiation factors, and global factors such as the concentrations of ribosomes and translation factors).

3. The fact that more than one ribosome can translate an mRNA at a certain time point.
4. Excluded volume interactions between ribosomes and possible traffic-jams.
5. The stochastic nature of the process.
6. The movement has a directionality (in the case of translation it is totally asymmetric).
7. The fact that different mRNA molecules compete for the same pool of ribosomes (and other translation factors); i.e. limited resources.

All the points above are known to be fundamental aspects of the canonical translation mechanism in all domains of life (5). The list above was chosen based on the majority of the previous studies in the field (as reviewed in this manuscript).

Of course, there are various (some less canonical) aspects of translation that are not mentioned above (e.g. ribosome recycling) that appear in some of the reviewed computational models. Importantly there are fundamental translation aspects that have not been considered in biophysical translation models (e.g. re-initiation after translating a uORF, frame shifts, numerous alternative initiation start sites, to name a few).

Currently there is no computational predictive model that considers all the aspects above; however, as described below, there are models that consider many of these aspects. In the following sections we review the models suggested in the field sorted according to their complexity (see also, Figures 1 and 2).

Models that are based on averaging estimations of codon decoding rates

The simplest biophysical models of translation assume a 'speed' (or decoding rate) for each codon and average these speeds over the codons of a gene. The estimated speed of each codon can be inferred based on various gene expression variables such as tRNA levels, ribosome densities, and/or mRNA levels. Thus, these models consider only aspect 1 (but in some cases, as described below, they indirectly also consider aspect 7).

As a typical example of these models we would like to describe in further detail the tRNA Adaptation Index (tAI, (48,113), see Figures 1A and 2). While there were previous models that aimed at estimating the translation rate of a gene based on averaging weights related to the optimality of each codon (114–116) and based on their frequencies; the tAI is the first that aimed at considering basic aspects related to the *biophysics* of translation.

The tAI assumes that the relative concentrations of the tRNA molecules that recognize a codon have a significant effect on the codon translation rate, and it gauges the availability of the different tRNA molecules for each codon along an mRNA. Though this is clearly a very rough approximation of the reality and there are many additional relevant factors, the tAI has been shown to have high performance levels, and is employed as a main feature in many of the more sophisticated predictors (2,46–47).

This measure is determined by combining the estimated thermodynamic properties of the codon-anticodon interac-
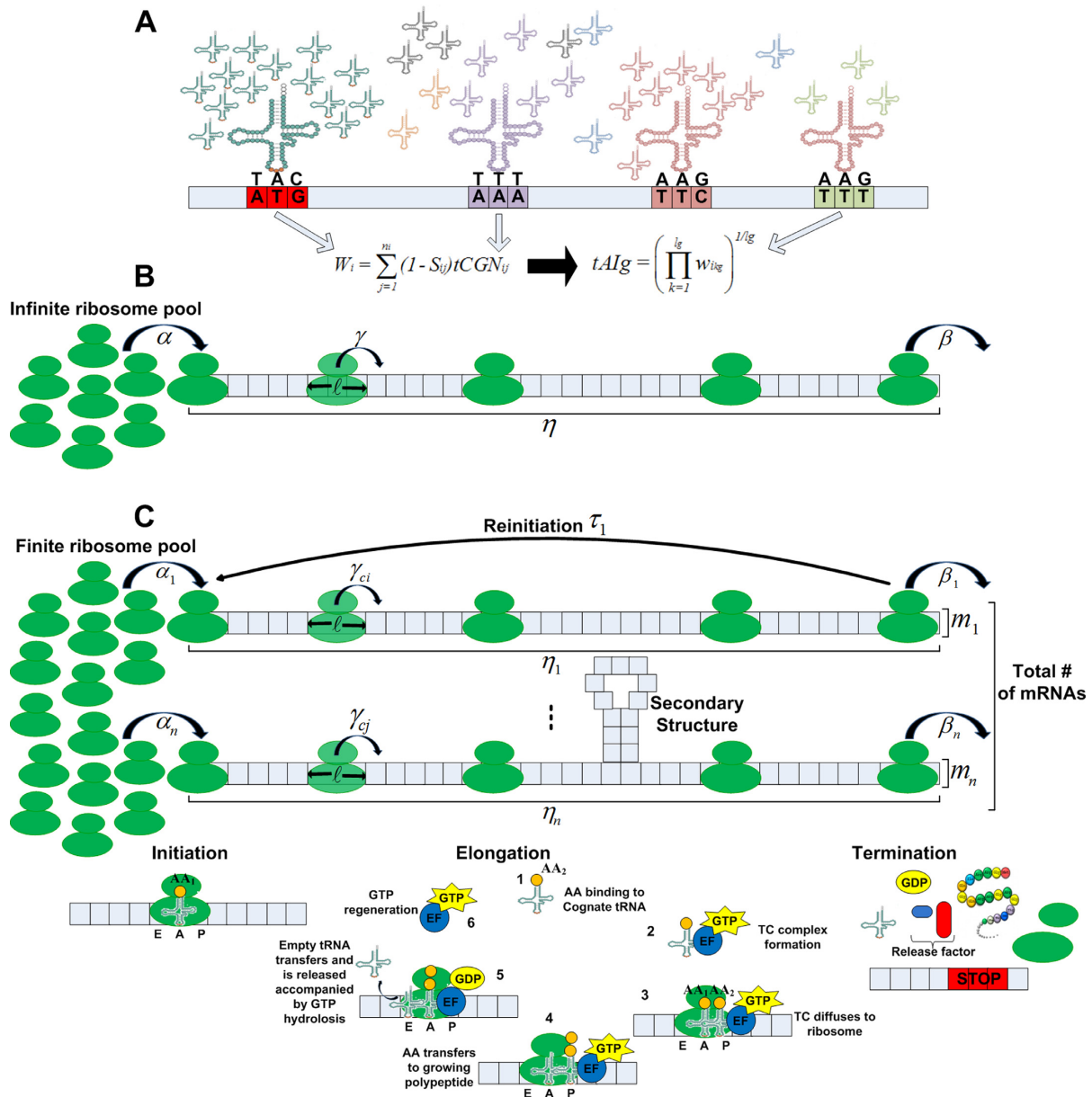
Figure 1. Illustration of the biophysical models reviewed and envisioned. (**A**) The simplest biophysical model, the tRNA adaptation index (tAI), which gauges the availability of the different tRNA molecules for each codon along an mRNA; the output of this model can be an average over all the codons (one number per gene) or a profile of the nominal codon adaptation along the mRNA (a vector per gene). (**B**) The simplest TASEP (totally asymmetric simple exclusion process) model, which considers a single mRNA and its length η, 'infinite' ribosomal pool (i.e. no feedback between the initiation rate and the ribosomes on the mRNAs), ribosome length ℓ, single initiation rate α, single elongation rate γ, single termination rate β. (**C**) An intricate TASEP model including all the canonical aspects necessary towards a whole cell simulation: finite ribosomal pool, ribosome length ℓ, genome-wide mRNAs, their levels m_i, lengths η_i, and secondary structures, transcript specific initiation rates α_i, codon specific elongation rates γ_{cj} (based on numerous factors such as tRNA concentrations), transcript specific termination rates β_i, transcript specific re-initiation rates $\tau_{i,}$. Each of the three main stages of gene translation, namely initiation, elongation, and termination, can be broken down to the multifarious steps comprising them, see, for example, Zhang *et al.* (51).

tion, taking into account that due to wobble interactions, several anti-codons can recognize the same codon, with different efficiency weights. Let n_i be the number of tRNA isoacceptors recognizing codon i. Let $tCGN_{ij}$ be the copy number of the jth tRNA that recognizes the ith codon, and let S_{ij} be a parameter corresponding to the efficiency of the codon-anticodon coupling between codon i and tRNA j. Currently the S_{ij} are inferred by optimizing the correlation between the *tAI* and gene expression measurements.

The absolute adaptiveness, W_i, for each codon i is defined as (48,113):

$$W_i = \sum_{j=1}^{ni} (1 - S_{ij}) tCGN_{ij}$$

From W_i we obtain w_i, which is the *relative* adaptiveness value of codon i, by normalizing the W_i's values (dividing
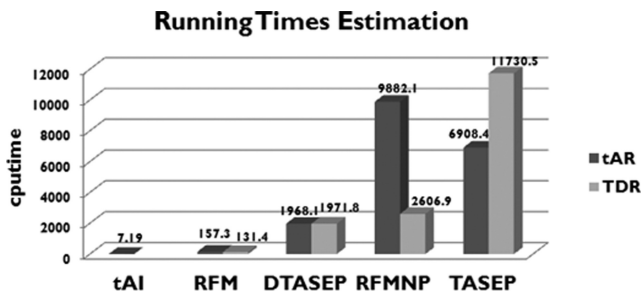
Figure 2. An illustrative comparison of the running times of five representative translation biophysical models, namely, the tAI (48); the RFM (46) which is a deterministic mean field approximation of the TASEP, modeled here with a single initiation rate for comparability; the model of (128) which is a deterministic TASEP model (DTASEP); the RFMNP (129) which is a general dynamical model for large-scale simultaneous mRNA translation and competition for ribosomes based on combining several ribosome flow models (RFMs); and a relatively simple TASEP which has an infinite ribosome reservoir, a single initiation rate, a single termination rate, individual codon rates, calculated across the entire genome, but for each gene separately (in this sense we assume an infinite set of ribosomes). The comparison was performed on 6664 *S. cerevisiae* coding sequences, with aside from the tAI, the individual codon times/rates were calculated according to two estimation methods: (i) What we termed the tAR, tRNA Adaptation Rate, which is estimated based on the tAI logic. (ii) The TDR, Typical Decoding Rate, which is described in (161), and the initiation rate was taken to be 0.8 based on (209). The RFMNP is initiation dependent thus using a single initiation rate for it makes no sense, thus they were estimated to be the measured ribosomal read count divided by the mRNA levels and then normalized to have a median of 0.8. In addition we constrained the free ribosome pool to be 15% (24) of the total number of ribosomes (which is around 200 000 (210)). The running times are of course parameter dependent, and we tried to make them as similar as possible for comparability. We used Matlab 2015b for all simulations.

them by the maximal of all the 61 W_i). The tAI of a gene is the geometric mean of the w_i of its codons.

Recently novel indexes/models aimed at improving the estimation of codon decoding rates using similar ideas but in a more accurate manner:

The normalized translation efficiency (nTE) index (68) considers both the 'supply' (the number of tRNA of each type) and 'demand' (i.e. the number of codons of each type in the transcriptome) of tRNA species. This is achieved via normalizing the w_i values of the tAI by the frequency of the corresponding codons in the transcriptome.

Another recent index named the mean of the typical codon decoding rate (MTDR) (117), aimed at estimating the typical translation rate of each codon based on Ribo-Seq (25) (after filtering experimental biases and considering aspects such as ribosomal traffic jams). By definition this index also aims at considering both the 'supply' of tRNA species and their demand, in addition to other aspects of the process, such as amino acid availability and Aminoacyl-tRNA synthesis levels and the possible diffusion rates of different tRNAs, not considered in the other indexes mentioned above.

Both the nTE and the MTDR partially consider aspect 7 in the previous section: competition on limited elongation resources.

It is important to mention that many predictors and models integrated the tAI (or similar measures) into more complicated multivariate models (1–2,47,118–119). Such mod-

els try to predict gene expression aspects based on many transcript features, where the tAI (or similar measures) is only one of the features, but often has been one of the most predictive variables (2,47,118–119).

Finally, one fundamental disadvantage of the models described above is the fact that they attribute a *single* speed to each coding region (which is presumably the mean of all the codons' decoding rates). However, in recent years a natural generalization to these models has been used for studying various biological phenomena (7,68): compute a *vector (or a profile)* of estimated nominal decoding rates (the decoding rates when there are no interactions between ribosomes and traffic jams) which includes a value for each codon instead of an average of all these values (see Figures 1C and 3A).

Rudimentary totally asymmetric simple exclusion process (TASEP) models

The idea that protein synthesis involves unidirectional ribosomal movement along an mRNA molecule, was proposed by Warner and Rich in 1962 (71,120). The statistical properties of ribosome movement were explored initially by Zimmerman and Simha (121,122), and then refined by MacDonald and Gibbs (108,123). In these works mRNAs were considered as lattices on which ribosomes move with specific hopping probabilities, the latter being functions of the intrinsic kinetics of elongation; the ratio of initiation, elongation and termination rates (which determine ribosome density on the messenger RNA); and the probability that progress of a ribosome is unimpeded by preceding ribosomes (which is itself a function of the ribosome density on the message) (71). The statistical approach as developed by MacDonald and Gibbs (108,123) continued to be developed theoretically and eventually became known as the 'Totally Asymmetric Simple Exclusion Process' or TASEP model for translation. The model is called TASEP since it assumes that the movement of the ribosomes is only in one direction, from the 5′ end to the 3′ end ('Totally Asymmetric'), a particle (ribosome) can only hop to the next position (codon) in the lattice ('Simple'), and only if it is empty; i.e. there are excluded volume interactions ('Exclusion').

Characteristics of early versions of the TASEP include assumptions of limitless ribosome-supply (which is related to a very high probability of ribosome attachment to the transcript, which affects the initiation rate), a single, uniform elongation rate-constant along the mRNA, and a coarse-grained description of the elongation process, which is simply regarded as a 'hopping-probability' (71,124–127) (Figures 1B and 2).

In the case of the rudimentary TASEP models, the assumption is that the initiation rate is constant and/or not coupled/related to the number of available ribosomes. In other models, which will be discussed in the next section, the initiation rate is affected by the number of available ribosomes via the fact that the number of ribosomes is finite (increasing the number of ribosomes on one mRNA should cause a decrease in the number of ribosomes on other mRNAs). Specifically, if the number of mRNAs (and ribosomes on the mRNA) is very small/large (respectively) relatively to the total number of mRNAs/ribosomes in the cell the initi-

ation rate of the gene is expected to be practically constant / change (respectively) when there are changes in the codons of the genes and/or the ribosome density along the mR-NAs, as these changes have negligible/significant (respectively) effect on the ribosome pool and other mRNAs.

More comprehensive models of the TASEP or deterministic mean field approximations of it include *different* translation rates for different codons (44,46,53–54) (Figure 1C). It is important to mention that even though the translation process is stochastic some of the models were based on a deterministic approximation of the process (46,51,128–129). In (128), for example, each codon has a deterministic nominal decoding time, ribosomes have excluded volume interactions, and they cover nine 9 codons. These models are usually computationally faster (among others, due the fact that they do not require averaging over many possible trajectories of the model), and can be analysed analytically in a simpler manner.

For a simple and intuitive example according to the Ribosome Flow Model (RFM) (46) (Figure 2), a ribosome that occupies the *i*th site moves, with rate λ_i, to the consecutive site provided the latter is not occupied by another ribosome. Transition rates are determined by the codon composition of each site and the tRNA pool of the organism. Briefly, the elongation rate associated with a codon is proportional to the abundance of the tRNA species that recognize it, taking into account the affinity of the interactions between the tRNA species and the codons. Denoting the probability that the *i*th site is occupied at time *t* by $p_i(t)$, it follows that the rate of ribosome flow into/out of the system is given by: $\lambda[1 - p_1(t)]$ and $\lambda_n p_n(t)$ respectively. Hence, the rate of ribosome flow from site *i* to site *i* + 1 is given by: $\lambda_i p_i(t)[1 - p_{i+1}(t)]$. Thus we get the following set of differential equations that describe the process of translation elongation:

$$\begin{cases} \frac{dp_1(t)}{dt} = \lambda[1 - p_1(t)] - \lambda_1 p_1(t)[1 - p_2(t)] \\ \frac{dp_i(t)}{dt} = \lambda_{i-1} p_{i-1}(t)[1 - p_i(t)] - \lambda_i p_i(t)[1 - p_{i+1}(t)] \ 1 < i < n \\ \frac{dp_n(t)}{dt} = \lambda_{n-1} p_{n-1}(t)[1 - p_n(t)] - \lambda_n p_n(t) \end{cases}$$

The steps of the mechano-chemical cycle of individual ribosomes during the elongation stage were not captured explicitly in the simple TASEP models due to the one single 'hopping' parameter that was used to describe the rate of translation of one codon. Moreover, these basic TASEP models neither incorporate any mechanism for specifically selecting the correct amino acid monomer, nor do they allow for the possibility of translational error, etc. Therefore, such TASEP models are too simple to account for the effects of various mechano-chemical processes on the statistical properties of polysomes (58,72,111,130–134). Finally, these models did not consider more complicated and global aspects of translation such as competition between mRNA molecules on ribosomes and translation factors, and between cognate and near-cognate tRNAs.

For a thorough review of TASEP modelling and its development we refer the reader to (72) and (71).

Complex totally asymmetric simple exclusion process (TASEP) models

It is easy to see that each of the different stages of a rudimentary TASEP (described in the previous subsection) actually represents various/many sub-stages and aspects of the translation process. For example, these aspects can include, among others, the different conformational steps of the ribosome during one iteration of elongation (56,59,135–136), the initiation itself includes various sub-steps (5′ end recognition, scanning, start codon recognition, etc) (55,137) that are usually not modeled in a rudimentary TASEP, the unfolding of the mRNA (61,138–139), the diffusion of tRNA molecules (19,46,51,54,61–62,66,69,131,134,140–142), interactions between the nascent peptide and the ribosome (61,143–144), diffusion of elongation factors (42), the addition of an amino acid to the growing polypeptide chain (5), tRNA molecule release (51), re-cycling of elongation factors (58,65,145), ribosome frame-shift (146), interactions between ribosomes and ribosome drop-off (46,53–54,59,69,131,135,142), mRNA decay (70), and many additional aspects (Figure 1C).

Indeed in recent years, various 'complex' TASEP models that consider some of these phenomena have been suggested. Specifically, these models consider not only linear initiation-elongation-termination, but also additional more detailed aspects of the process. For instance, some of these models divided each stage (or elongation iteration) of the translation process into multiple sub-steps, while others considered phenomena such as scanning during the initiation step and/or ribosome recycling (Figure 1C).

For example, Ciandrini *et al.* (54) suggested a TASEP model where each elongation step dynamics includes two sub-steps: (1) the recognition of the cognate tRNA with rate k_i depending on the codon *i*, and (2) the translocation towards the next codon with rate γ. At the last codon, the ribosomes detach and release the protein with a rate β (termination).

Zhang *et al.* (51) developed a more complicated model in which the rate of translation of a single codon in the cell is determined by following four subsequent processes: (i) attachment of the amino acid to its cognate tRNA by the corresponding aaRS; (ii) aa-tRNA forms a ternary complex (TC) with GTP-bound elongation factor (EF-Tu in prokaryotes and eEF1A in eukaryotes); (iii) TC diffuses to a ribosome requesting it and transfers the amino acid to the growing polypeptide chain, whereby the tRNA is released; and (iv) the GDP-bound elongation factor is regenerated into its GTP-bound state by EF-Ts (prokaryotes) or EF-1B/C (eukaryotes) (see Figure 1C). In addition they modelled spatial diffusion of the TC complexes around ribosomes translating repetitive sequences using a Brownian random-walker model.

Mitarai *et al.* (69) developed and analyzed a translation model that captures in addition to the basic TASEP features, also the fact that ribosomes recycle. Specifically they added a parameter corresponding to the fact that with a certain probability a ribosome finishing translating the mRNA can immediately start another round of translation (e.g. due to the cyclization of the mRNA molecule). Among others, they used their model to investigate the ef-

fect of sequences of synonymous codons on ribosome traffic, based on codon-dependent rates estimated from experiments. They quantify the ability of codon usage to regulate ribosome traffic by simulating the translation of several natural genes, and demonstrate that those with a high codon usage bias, tend to have the ability to reduce ribosome collisions, hence optimizing the cellular investment in the translation apparatus. Recently, it was also demonstrated how ribosomal recycling can be modelled and analyzed via an RFM based approach (147).

In (61), the authors developed and analyzed a TASEP based model where the nominal codon decoding rate of each codon is affected by three different components: (i) the adaptation of the codon to the tRNA pool; (ii) the strength of the local mRNA folding of the 40 nucleotides downstream of the codon; (iii) the net charge of the 31 amino acids inside the exit tunnel—as the exit tunnel is negatively charged a positive charge is expected to slow down the ribosome.

In summary, the most important advantages of the TASEP (rudimentary and complex) models are related to the fact that they theoretically can capture and provide high resolution predictions of all the important *local* aspects of translation dynamics (aspects 1–6 in the third section). In addition, the fact that the different mRNAs are not 'connected' can enable accelerating and parallelizing the simulations. However, the models cannot capture aspects related to the coupling between the translation components (e.g. tRNAs, mRNAs, ribosomes) (aspect 7 in the third section) such as the 'competition' between mRNAs for ribosomes. It is clear that more complex translation models are preferred if the estimated parameters in the models are accurate and the running time is feasible, but this is obviously not always the case.

Whole cell simulations of translation

Currently there are no models that consider all the fundamental mRNA translation aspects at a cellular level. However, there are previous studies that contributed towards developing such a model. The aim of this section is to review some of these studies, specifically we review models which consider the dynamics of many interacting mRNA molecules. While there are quite a few previous studies that aimed at analyzing many mRNAs in the cell in a single model (51–52,129,148–157), many of them are based on mathematical analyses and simulations that are based on artificial genes and/or a small number of genes. As explained below, even the more comprehensive models suggested recently do not consider all the fundamental relevant translation aspects.

Many studies have demonstrated/suggested the importance of considering the 'competition' of mRNAs on a finite pool of resources (149–158). Specifically, unlike many other models of translation (46,53,65), which treat each mRNA molecule in isolation and assume an inexhaustible supply of free ribosomes that initiate the message at a constant rate, those models consider simultaneously every tRNA, mRNA, and ribosome molecule in the cell, and so capture the indirect effects of one gene's translation on another's. In particular, if many ribosomes are engaged in translating the

mRNAs of one gene, this reduces the pool of free ribosomes and tRNAs available to translate other genes (Figures 1C and 3B).

For example, Siwiak *et al*. (67) developed a TASEP based comprehensive and quantitative model of translation, enabling the estimation of translation parameters which are difficult to measure experimentally. They model the number of transcripts, ribosome density, protein abundance, tRNAs' decoding specificities, average cell volume, mean codon translation time, total transcript translation time, total time required for translation initiation and elongation, translation initiation rate, average number of active ribosomes in a cell, ribosome size/coverage, mRNA half-lives (optionally).

Shah *et al.* (62) developed a continuous-time, discrete-state Markov model of translation in a yeast cell. The model keeps track of every ribosome and tRNA molecule; according to the model, at any time point each ribosome/tRNA molecule is either freely diffusing in the cell or bound to a specific mRNA molecule at a specific codon position. Their model includes many parameters and describes translation dynamics at a single-nucleotide resolution for the entire transcriptome. In combination with ribosomal profiling data (Ribo-Seq), they aimed at using their model to infer the initiation rates of all abundant yeast transcripts, and systematically explore how the codon usage, transcript abundance, and initiation rates of genes jointly determine protein yield and cellular growth rate. The rates of initiation and elongation are based on physical parameters that have been experimentally determined in yeast, including the cell volume, the abundances of ribosomes and tRNAs, and their diffusion constants.

Another whole cell model which includes translation modeling was suggested by Karr *et al*. (156). This model was based on the human pathogen *Mycoplasma genitalium* and aimed at including all of its molecular components and their interactions (specifically translation). It is important to mention, however, that the resolution of translation modelling was not high and most aspects mentioned above (e.g. ribosome interactions and traffic jams during elongation, different codon decoding rates, initiation modelling, etc.) were not included. Similarly, Roberts *et al.* developed a whole cell simulation based on *E. coli* (159); also in this case they assume that a single rate-limiting event affects translation, ignoring the fundamental higher resolution details related to the translation process.

We conclude that currently there are no models that are based on real data that encapsulate all fundamental mRNA translation aspects. For example, the following canonical aspects (mentioned in the previous section) are not considered in (62,67,156,159): the different conformational steps of the ribosome during one iteration of elongation, the different sub-steps of the initiation stage (5′ end recognition, scanning, start codon recognition), the unfolding of the mRNA and interactions with mRNA folding, interactions between the nascent peptide and the ribosome, re-cycling of elongation factors, ribosome frame-shift, ribosome drop-off, to name a few.

Parameter estimation for the biophysical models

In this section, we provide some specific examples of parameter usage and estimation in some of the biophysical models described. There are variables such as protein levels (1,3,29–30,40,74), mRNA levels (160), number of ribosomes on the mRNA (24,77–78), that can be directly measured; however, as described here, other parameters can be estimated only via developing sophisticated approaches, which require a good understanding of the biophysics of translation, molecular evolution, machine learning, and more.

One fundamental parameter in biophysical translation models is the nominal decoding rate of codons. Various methods for estimating this parameter have been suggested in recent years. One approach is based on the tAI or normalized versions of it (2,46–48,61,68,113). The tAI was described above. One of the most challenging issues when computing the tAI is the selection of a meaningful set of S_{ij}'s, which are the selective constraints on the efficiency of the codon–anticodon coupling. Since translation efficiency, also in terms of adaptation to the tRNA pool, is expected to be higher for highly expressed genes, it was suggested to find the set of S_{ij}-values that maximize the correlation between expression levels and tAI values (48). This was first achieved based on microarray data from yeast (48); later, measurements of protein levels were used to optimize these values (7). Recently, it was suggested that these values can be estimated per organism, even when there are no gene expression measurements, based on optimizing the correlation with codon usage bias (113). Another component of the tAI is the tRNA concentration; today there are very few measurements of tRNA levels (7,26,39,75–76); however, it was suggested that tRNA copy numbers correlate with tRNA levels (48), and in many studies they are used as a proxy for tRNA levels. Codon decoding time estimation may be improved via considering the demand vs. supply of codons (68).

Recently, it was suggested that codon decoding rates can be estimated based on ribosome profiling data (161,162). To this end, ribosome profiling (Ribo-Seq) data should be normalized to control for the different initiation rates and mRNA levels of genes; then, based on the histogram of the normalized Ribo-Seq values of each codon the typical decoding rate can be estimated. In (161) a novel filter was suggested that enabled estimating the typical codon decoding rate via filtering various problems/phenomena such as pauses/biases/ribosomal traffic jams related to the Ribo-Seq protocol and the translation process. A newer study (163) has suggested inferring the elongation rates and initiation rates of mRNAs by finding parameters that yield the best fitting of the state variables (profiles of local ribosome densities) of the TASEP to the corresponding Ribo-Seq measurements.

An additional fundamental parameter which cannot be inferred in a trivial manner is the translation initiation rate. Translation initiation estimation is not trivial since it is based on many aspects such as the ribosome's attempt to bind the mRNA, the concentration of initiation factors and the presence of secondary structures in the 5′UTR region, the nucleotide composition surrounding the start codon, the number of START codons in the 5′ UTR and their nu-

cleotide context, and more (5,54–55,92–93,98,137,164). An approach for estimating translation initiation was suggested in (54): they estimate mRNA-specific in vivo translation initiation rates on a genome-wide basis by integrating a computational model of mRNA translation with experimental datasets of ribosome occupancy. They first apply their TASEP based translation model which predicts the translation rate J and the ribosome density ρ on each mRNA as a function of the translation initiation rate α. Then, by utilising genome-wide experimental data of ribosome density ρ_φ they identify the physiological translation initiation rate α_φ as the one which, when used in their simulations, replicates the experimentally observed density:

$$\rho(\alpha_\varphi) = \rho_\varphi$$

This yields a translation initiation rate value α_φ for each mRNA.

Recently, it was suggested that the ribosome drop off rate can be predicted based on the analysis of the decrease in ribosome density along coding regions via ribosome profiling data (165). These data can be used in the future when ribosome drop off rates are integrated into translation models.

Finally, a good source that includes many parameters related to translation (e.g. ribosomes' size, initiation rates, codon decoding rates, diffusion coefficients, etc) based on measurements or estimations is the biological numbers database BioNumbers (166). Parameters from this database can be used in future development of translation models.

It is important to emphasize that some of the translation parameters (e.g. ribosome footprint size) are probably universal or very similar for various organisms; other parameters (e.g. initiation rates and ribosome drop off) can be estimated for various organisms when the related data (e.g. ribosome profiling) is available. However, there are parameters that currently are not available at a genome wide level for multiple organisms (e.g. ribosome recycling rates and non-programmed frame shift rates).

Comparison of different models in terms of running time and complexity

In the previous sections, we described various representative biophysical models of translation. In the Table 1 below we include a comparison of some of the recently suggested models. We focused on the following aspects: scope of the biophysical modelling, computational complexity (expected running time), number of parameters in the model, and properties of the model, etc.

Systems biology studies based on computational biophysical translation models

In this section we briefly review some central biological results and discoveries derived from the translation biophysical models. A central objective is to demonstrate how biophysical models of translation can contribute to biomedical science; specifically, as we explain, they enable understanding fundamental topics related to genome evolution, translation biophysical dynamics, and regulatory processes related to translation (see also Table 2 and Figure 3).

Table 1. The table below summarizes and compares the biophysical models discussed, classifying them according to the features they encapsulate

Short description of the model	Reference	Relative running time	Parameters fit to real data	Additional info. and data	Deterministic/ stochastic
The tRNA adaptation index (tAI)	Dos Reis *et al.* (48)	1	64 codon-anticodon interaction efficiencies; given the tRNA levels only 5 parameters should be estimated.	The parameters can be inferred by optimizing the correlation between the *tAI* and gene expression measurements. The model was applied to at least 126 genomes, ranging from archaea to eukaryotes.	Deterministic
A measure based on slow codon clusters	Zhang *et al.* (66)	1	tRNA concentration	The model was applied to *E. coli* and *B. Subtilis*. tRNA codon pairing ratios were inferred from experiments.	Deterministic
The Ribosome Flow Model (RFM), a mean field approximation of the TASEPC	Reuveni *et al.* (46)	2	61 nominal codon decoding rates, initiation rate(s), codon chunk size.	For example, the parameters can be inferred based on Ribo-Seq or the tAI values. A global initiation rate can be used. The model was applied to at least *S. cerevisiae*, *E. coli*, *S. pombe*, and *H. Sapiens*.	Deterministic
Deterministic TASEP	Zhang *et al.* (128)	2	Ribosome size, initiation rate, termination rate, codon decoding rates.	See previous comment. The model was originally applied to *E. coli*.	Deterministic
Simple TASEP with mRNA decay	Deneke *et al.* (70)	3	Simple TASEP parameters only the termination rate equals the elongation rate, two degradation rates: one proportional to the fraction of the mRNA covered by the ribosomes and the other to the unshielded.	The model was applied to *E. coli* and presents improved fitting of published decay assay data.	Stochastic
Simple TASEP with codon specific rates (TASEPC)	Mitarai *et al.* (53)	3	Simple TASEP parameters, 2 sets of codon rates: uniform, and one of three values for each codon specific translation rate, average mRNA translation number.	The model was applied to *E. coli* focusing on the *lacZ* mRNA. It simulates an in-vivo experiment.	Stochastic
TASEPC	Ciandrini *et al.* (54)	3	TASEPC parameters, the codon dependent rates are composed of: tRNA recognition rate and codon translocation rate (codon independent), the initiation rates are predicted using experimental ribosomal density.	This model was applied to *S. cerevisiae*, integrating genome-wide experimental data of ribosomal density. tRNA-capture rates can be estimated from data on tRNA abundances, which are assumed to be proportional to their gene copy numbers.	Stochastic
Simulating translation as a set of chemical equations via a version of the of the Gillespie SSA algorithm (208)	Chu *et al.* (140,208)	3	61 codon frequencies, 61 tRNA levels, 20 tRNA re-charging parameters, forward/backwards rate of cognate tRNA binding to the ribosome and of near-cognate tRNA binding to the ribosome.	The model was parameterized for *S. cerevisiae* based on experiments.	Deterministic
A multi step model of translational elongation consisting of deterministic and stochastic parts based on the concentrations of ternary complexes and competing mRNAs on a finite ribosomal pool	Zhang *et al.* (51)	3	Experimentally quantified tRNA concentrations, fraction of charged tRNA, enzymatic parameters of tRNA acylation catalyzed by the aaRS, rate constant of the complex formation between the charged tRNA with the EF, overall rate constant of EF-GTP regeneration, number of ribosomes.	The model was applied to *E. coli*. The authors experimentally quantified tRNA concentrations, and polysomal profiles.	Deterministic and Stochastic
RFM network with a pool (RFMNP)	Raveh *et al.* (129)	3	RFM parameters, number of ribosomes, mRNA levels.	Can be easily parallelized. This model was applied to *S. cerevisiae* here.	Deterministic
TASEPC with finite resources	Brackley *et al.* (149)	4	TASEPC parameters, number of available tRNAs, tRNA recharging rate constants, mRNA levels.	This model was applied to some of the mRNAs in *S. cerevisiae*.	Stochastic
TASEPC with finite resources	Chu *et al.* (150,158)	4	TASEPC parameters, mRNA levels, number of ribosomes, ribosome recycling rounds, rate of reacharging tRNA, number of tRNAs.	This model was applied to *S. cerevisiae*.	Stochastic
TASEPC with finite resources	Siwiak *et al.* (67)	4	tRNAs decoding specificities, average cell volume, average number of active ribosomes in a cell, ribosome size/coverage. Each gene is attributed with a set of translational parameters, namely the absolute number of transcripts, ribosome density, mean codon translation time, total transcript translation time, total time required for translation initiation and elongation, translation initiation rate, mean mRNA lifetime (optionally), and absolute number of proteins produced by gene transcripts.	This model was applied to *S. cerevisiae*.	Stochastic
TASEPC with finite resources	Shah *et al.* (62)	4	Many parameters including, number of ribosomes, number of mRNAs, number of tRNAs, tRNA competition coefficient, diffusion coefficient of ribosome, gene specific initiation probability.	This model was applied to *S. cerevisiae*.	Stochastic

Table 1. Continued

Short description of the model	Reference	Relative running time	Parameters fit to real data	Additional info. and data	Deterministic/ stochastic
Whole cell computational model	Karr *et al.* (156),	4	Many parameters (not only translation). The functionality of the cell was divided into 28 modules, modeled independently and integrated.	This model was applied to the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. They simulated 192 wild-type cells and 3011 single-gene deletants. Translation is only small part of the model.	Stochastic / Deterministic

*DM—deterministic model. The relative running times estimations are derived based on the analysis performed and depicted in Figure 2; the estimated running ranking is 1–4 where 1 is typically related to a very fast which is based on averaging estimations of codon rates; 2 is typically related to deterministic TASEP models; 3 is typically related to simple stochastic TASEP models; and 4 is related to models expected to be slower than a simple stochastic TASEP such as comprehensive whole cell simulations. The number of parameters reported is the minimal one used for conventional usage of the model. We considered only models that were used to analyse large scale genomic data. A simple TASEP (as described above) includes an initiation rate, codon independent elongation rate, termination rate, mRNA length (there is a single mRNA), number of codons covered by a ribosome. A simple TASEP with codon dependant rates for simplicity will be termed TASEPC.

Among others based on the tAI (48) it was suggested that genome size and tRNA gene redundancy determine the action of natural selection on codon usage in all living organisms (Figure 3H). Their findings suggest that an optimal combination of these factors exists, for which the action of translational selection is maximal. Moreover, they stress how the lack of duplicated tRNA genes might explain the absence of translationally selected codons in bacteria with small genomes.

A different study employed the tAI to suggest that cell cycle regulated genes tend to include codons that are not translationally efficient to generate cell cycle-dependent oscillations in protein levels (167). They suggest that due to oscillations of tRNA species during the cell cycle, these codons enable oscillations of the protein levels of these genes.

More recent studies based on genomic profiles (or profiles of single genes or gene groups) of measures of nominal translation elongation speed, such as the tAI, enabled discovering various translation signals that are selected for and are related to translation efficiency, ribosomal and other factors allocations, and co-translational protein folding. For example, it was suggested that the first 30–50 codons are slower to improve ribosomal allocation and protein folding (6–7,12,98,168) (Figure 3A), and that evolution tends to cluster codons that can be decoded by the same tRNA to improve tRNA recycling via improving their effective diffusion to the ribosome (6,12,42). Based on various methods for generating profiles of the normalized translation efficiency index, it was recently proposed that the folding of the protein is partially effected by the local translation elongation speed (66,68) (Figure 3D). For example, it was suggested that translation elongation is lower downstream of the protein domain boundaries probably to improve the fidelity of co-translational domain-wise protein folding (57,66). Similar ideas related to elongation and co-translational folding have been suggested based on additional techniques including the tAI and TASEP based models (98,169).

Numerous studies have employed various TASEP models to understand how evolution shapes the distribution of codons to improve ribosomal allocation and prevent ribosomal collisions. For example, it was suggested that clusters of rare codons near the 5′ of the coding region decreases the effective initiation rate, the number of ribosomes on the mRNA, and the probability of ribosomal collisions and abortions (7,128). Similarly, it was proposed that higher codon usage bias (or adaptation to the tRNA pool) helps prevent traffic jams and improves ribosome allocation (69,170), and that codon arrangement, rather than simply codon bias, has a key role in determining translational efficiency (54). Furthermore it was shown that codon bias may act as a gene expression regulator by favouring codons with high tRNA gene copy numbers in highly expressed genes, and with low tRNA gene copy numbers in lowly expressed genes (46,171).

Another line of research connects biophysical models of translation and coding sequence evolution (44,65,170). Among others, based on such models it was suggested that a combination of nonsense errors and codon usage bias can have a large effect on the probability that a ribosome will complete transcript translation. In addition, they showed position dependent selection on codons, and suggested that nonsense errors can play an important role in shaping codon usage bias and can be used for predicting protein levels.

Moreover, Ciandrini *et al.* (54) used TASEP to identify the initiation rate, ribosome traffic dynamics and response to ribosome availability, of different classes of mRNAs, demonstrating that this classification based on translational dynamic maps onto known classes of gene functionality (Figure 3B and G).

An important question studied with TASEP models is the relative rate limiting step of translation, these studies have demonstrated that depending on the condition or organism both initiation and elongation may be rate limiting (7,54,60,62). Specifically, nowadays there is much emphasis on initiation being the rate limiting step and the principle determinant of translation 'efficiency'. However, the behavior of TASEP is a function of its three main parameters/stages (initiation, elongation, termination); two TASEP parameters/stages (initiation, elongation) are relevant when the termination is not rate limiting or is negligible (as usually assumed). A related topic is the phase diagram of the TASEP as a function of the initiation and typical/mean elongation rate (see, for example, (172)). It is common to divide the dynamics of the TASEP to three phases (see, Figure 3F): (i) Low density phase—which

is related to the case where the initiation is rate limiting (i.e. much lower than the minimal elongation rate); thus, the number of ribosomes on the mRNA and traffic jams is very low/negligible. (ii) High density phase—which is related to the case where the initiation is high (higher than the typical/mean elongation); thus, elongation is rate limiting and the number of ribosomes on the mRNA and traffic jams is high. (3) The third phase is maximal current—in this case, given the elongation rate, the initiation rate has similar levels which maximizes the translation rate. It is important to remember that ribosomes are extended objects that cover around 10 codons each; this contributes to the shrinking (relatively to the case where the TASEP particles occupy only one codon/cell) of the parameter space of the 'low density' phase, so that they render the cases where the initiation is 'rate limiting' less likely (111).

Furthermore, when considering only the mean measured initiation rate and elongation rates and a model of a single mRNA the initiation is clearly more rate limiting (123) (Figure 3G). However, more complicated models demonstrate that this is not the case (52,54,60,71,109). For example, recent experimental results show that eukaryotic ribosomes may translate mRNAs in multiple cycles before entering the free ribosome pool (71,173). This would greatly affect the translation control, and compared with single-cycle models, it could transfer significant levels of control to the elongation stage. Another example is related to the fact that mRNAs actually 'compete' over a finite ribosome pool; thus, initiation and elongation are coupled: higher elongation rate is related to faster release of ribosomes and higher effective initiation rate (129,149–150,153–157).

At what ribosome depletion level translation control is transferred to elongation depends in complex ways on the codon composition of the genome. The average speed of translation is not only a function of the number of slow codons in a message, but also of their distribution (46,50,54,61,66,69,71,131,174–177).

Another topic (which overlaps with the previously discussed) studied by a variety of models ranging from the simpler to the more complex, is the demand vs. supply of tRNAs, amino acids, ribosomes, and other factors, or the competition of mRNAs on various factors. Various studies assumed and suggested that this issue is a major determinant shaping coding regions and organismal evolution (19,46,51,54,62,66,68–69,131,134,140–141,149).

Additional facets explored included understanding the translation regimes in various genes, and the explanations to observed variables such as protein levels, ribosome densities, mRNA levels, and mRNA half lives. These include the study of the distributions of variables such as initiation rates and ribosomal densities and collisions, that cannot be easily measured, but can be estimated based on biophysical models and gene expression measurements (53,62,67,70).

Another interesting line of research that can only be studied via biophysical models of translation is the relation between local mRNA folding dynamics and translation dynamics, although to date very few papers tackled this issue or introduced this aspect into biophysical translation modeling. It was suggested, for example, that strong local folding slows down the elongation rate—the ribosome has helicase activity and more time is needed to unfold regions

with stronger local mRNA folding (61). Another study has recently also focused on the effect of the ribosome on the mRNA structure via its unfolding of the mRNA during translation (139) (Figure 3C). In (61) a TASEP model was devised where the nominal elongation rate of a codon is proportional to $e^{k*\text{FE}}$ where FE is the (negative) local folding energy of the 40nt downstream of the translated codon, and k is a parameter that was inferred based on fitting the model (ribosome density profile) to ribosome profiling data. In (139) they assume that the dwell time of a codon is proportional to the predicted mean base pairing probability of the codon's three nucleotides. In addition, various studies connected the mRNA folding at the beginning of the ORF to translation initiation efficiency—specifically weak folding surrounding the start codon enables efficient recognition of the start codon and the initiation signals surrounding it by the pre-initiation complex, while strong folding may contribute to the pre-initiation complex missing the start codon (Figure 3E) (40,98–99,178). A thermodynamic model which calculates the initiation rate in prokaryotes based on the difference in Gibbs free energy before and after the 30S complex assembles onto an mRNA transcript was suggested by (179). In this model, various free energy terms are calculated and integrated, including: (i) The energy released when the last nine nucleotides of the 16S rRNA co-folds and hybridizes with the mRNA subsequence at the 16S rRNA-binding site. (ii) The energy released when the tRNA$^{\text{fMet}}$'s anticodon hybridizes to the start codon. (iii) The folding energy of the mRNA subsequence surrounding the start codon prior to binding with the 30S complex. Finally, the effect of folding on translation, mentioned above, may suggest that the local mRNA folding landscape has shaped transcript evolution in various ways (98,180–182). Thus, local mRNA folding is an important 'intersection' that connects the sequence composition of the transcript with the dynamics of translation initiation/elongation and the molecular spatial/temporal evolutionary patterns of the transcript.

Moreover, many studies based on the biophysical translation models aimed at predicting various gene expression variables such as mRNA levels, protein levels, protein degradation rates, and ribosome densities based only on the genomic sequences (46–47,51,54,60,66,171).

DISCUSSION AND CONCLUSIONS

We reviewed dozens of biophysical computational/mathematical models related to ribosomal movement. We offer six major points for comparison among these models/papers:

1. Number of parameters needed to be estimated: on the one hand, models based on larger numbers of variables describe fine-tuned aspects of translation and capture larger fractions of the biological knowledge related to translation; on the other hand often there is no knowledge regarding the values of many of the parameters of the model, and/or estimation of these parameters is impossible or extremely noisy. Thus, subsequently these models cannot be used for providing predictions based on or related to real biological data. Furthermore, the major disadvantage when developing a model with many

Table 2. A summary of fundamental translation phenomena/discoveries studied with predictive biophysical models of translation

Discoveries / studied phenomena	References
Codon distribution and bias	(42,44,46,54,66,128,167)
Casual relation between tRNA levels and translation elongation rates	(19,42,46,48,51,54,66,68,140,149)
Supply and demand of tRNA	(68,149)
Rate limiting steps in translation	(7,46,54,60–62,66,71,209,168,173)
The effect of a ramp of slower codons on translation and fitness	(7,12,98,211)
Ribosome drop off	(51,165)
Ribosome interactions, collisions, allocations, and traffic	(7,54,128)
The effect of mRNA folding on translation initiation and elongation	(50,61,139)
Interaction of the nascent peptide with the ribosome	(61,169)
Co-translational folding	(57,66,68,169)
Initiation rates estimation and analysis	(54)
Protein abundance prediction	(46,67,149)
Ribosome density prediction	(46,54,60)
Translation rate prediction	(46–47,51,53–54,66)
mRNA half-life prediction	(70)
Codon translation rates prediction	(161,163)

parameters is overfitting (183), or accumulation of error due to individual errors in the different parameters.

2. Computational running times of model simulations: the running time of some of the measures/models is very short (48), while in others (111) the running time can be several orders of magnitude longer (as compared in (46)). The running time becomes an important issue when the objective is to analyze the entire genomes of complex eukaryotes with various/multifarious sets of variables.

3. The biophysical aspects modeled: which aspects of translation are modeled and which are not. The simplest models include mainly/only the effect of tRNA levels on ribosomal codon decoding time, while the newer more complex models include many additional aspects. For example, notable extensions that were introduced to the TASEP include multiple competing mRNA species (62,67,71,129,156), up to complete, genome-wide transcriptomes (51–52,71); mRNA transcription and decay (70,184); ribosome-induced peptidyl-tRNA hydrolysis in response to translational errors ('ribosome editing', (185)); aminoacyl-tRNA synthesis (19,46,51,54,62,66,69,131,134,140–141), ribosomal slow-down at mRNA secondary structures (57,66,138); competition for a finite pool of ribosomes (151–152,186); and the use of rate constants and species concentrations rather than rates (71,109), to mention a few.

4. Exact mathematical solutions versus simulations: The possibility of mathematically analyzing a model is a great advantage. Currently, there is no exact solution for a simple TASEP with non-homogenous hopping rates; thus, there is no solution for more comprehensive models based on the TASEP. However, in some cases a mathematical analysis is possible via various approximations, and answering questions in an indirect manner such as mean field approximations (see, for example, (46,72,176,187–190)).

5. Discrete models *vs.* an approximation of the models as continuous: The original TASEP based models are discrete; the number of ribosomes on the mRNA is an integer. However, there are various approximations for describing translation which are continuous (e.g. the RFM (46,191)). In these cases, the ribosomal densities are defined as real numbers (for example, it can be a probability

to see a ribosome in a certain region). Continuous models are often easier to analyze mathematically (187), and are thus more advantageous in that sense, but are a cruder approximation of the reality.

6. What kind of predictions do the models provide in comparison to real data: connecting the model to real measurements is important. Specifically, the possibility to show that the model prediction correlates/matches with biological measurements should help evaluating the model, and predictive models have various advantages in biomedical research. Some of the models/papers mentioned here were not based on a comprehensive analysis of large scale biological data and validation of the prediction they provide, while others are more strongly connected to real data.

The points above demonstrate that it is not trivial to rank the different models mentioned in this review. Almost all the models have relative advantages in some cases. Thus, the user of such models should carefully assess his specific research aims before choosing a model.

We would like to emphasize that this review was focused on translation (and specifically translation elongation), since this is by far the most studied gene expression aspect in this context. We believe that in the near future models similar to the ones mentioned here can be used to study and focus on additional gene expression steps such as transcription (192), intracellular transport (193), and translation initiation (63,194–195).

There are many challenges and open questions in the field:

First, in many cases some of the predictions (e.g. correlation with protein levels) provided by biophysically more comprehensive models of mRNA translation are similar (or even inferior) to the ones obtained for much simpler models (1,118). However, there are many aspects that can be provided only by comprehensive biophysical translation models; for example, these include ribosomal traffic jams, interactions between ribosomes, ribosomal drop-off rate, stochasticity in translation rate, and more. One major challenge in this context is related to the fact that there are currently no measurements related to these variables, thus, they cannot be accurately evaluated.
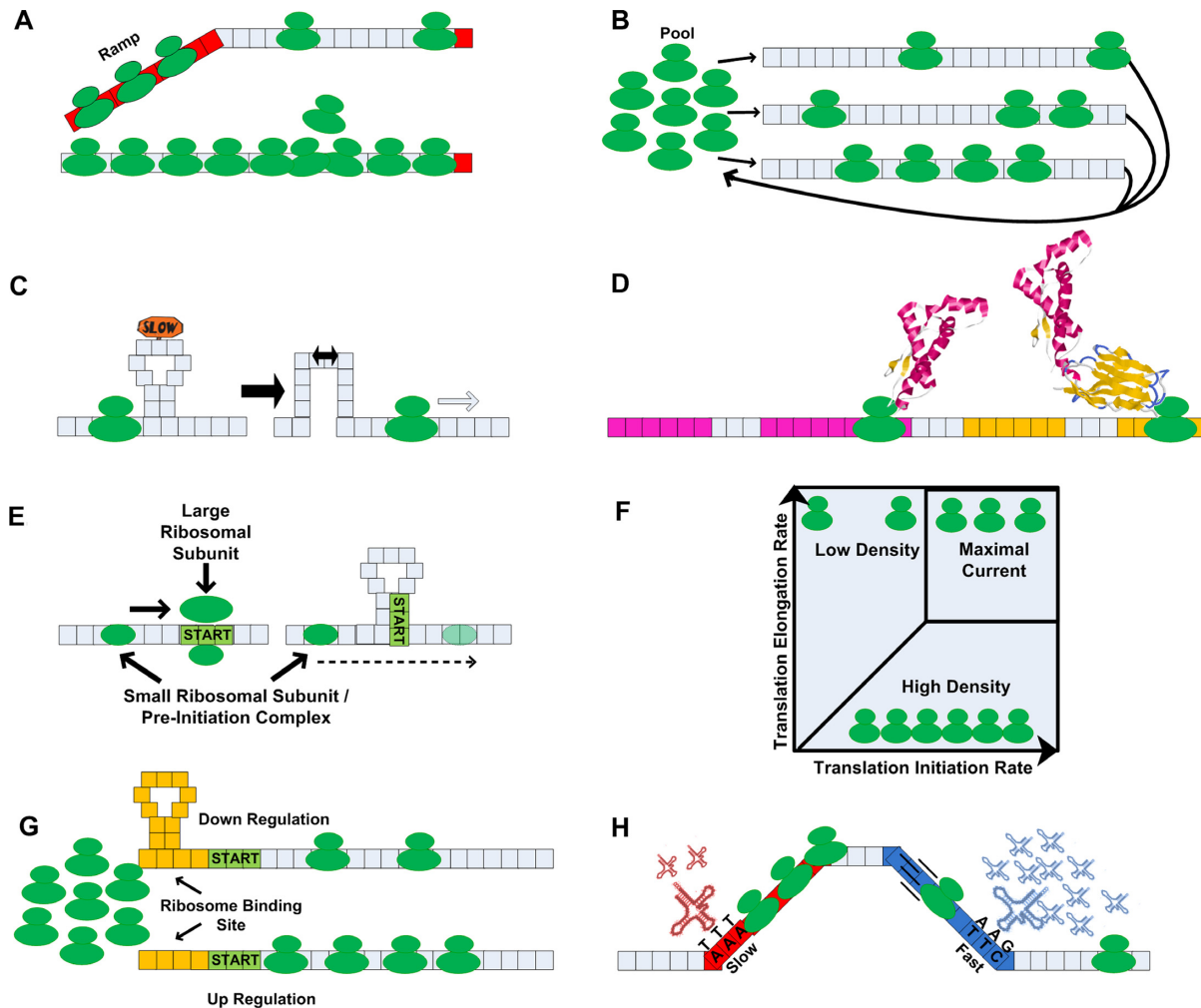
Figure 3. Illustration of some central and fundamental biological questions that are studied with predictive biophysical translation modeling. (**A**) Ramp and ribosome allocation: the ramp is a region characterized by slower elongation speed and codons less adapted to the tRNA pool, and may provide several physiological benefits, such as increasing the distances between ribosomes, promoting improved ribosomal allocation, and reducing ribosomal collisions and jamming, thus reducing the cost of wasted ribosomes and of spontaneous or collision-induced abortions (7,211). Red indicates a slow codon decoding rate (as illustrated in H). (**B**) Competition between ribosomes: since the amount of ribosomes in a cell is finite, the ribosome consumption of one mRNA will affect all others. For example, mRNAs consuming many ribosomes due to their properties (such as strong folding, or relatively slow codons), prevent them from re-entering the pool, thus reducing the initiation rates of all the other mRNAs (129,212). (**C**) mRNA folding and translation elongation: mRNA folding tends to slow down ribosome elongation, however ribosomes also tend to unfold the mRNA (61,139). (**D**) Co-translational folding: local discontinuous translation rates temporally separate the translation of segments of the peptide chain and actively coordinate their co-translational folding, to promote accurate folding of the peptide. Pink represents the codon rates enabling correct folding of the α-helix domain, while orange the β-sheet domain (15,68). (**E**) The mRNA folding near the start codon affects translation initiation efficiency. Specifically, weak folding surrounding the start codon enables efficient recognition of the start codon and the initiation signals surrounding it by the pre-initiation complex (left), while strong folding may contribute to the pre-initiation complex missing the start codon (right). (**F**) The three different phases of the TASEP as a function of the initiation and elongation rate. (**G**) Rate limiting steps: for instance, efficient/non-efficient ribosome binding sites and weak/strong mRNA structure at the end of the 5′ UTR can promote either up/down regulation of translation (7,46,54,60–62,66,71,209,168,173). (**H**) The effect of the tRNA levels on codon decoding rates: codons that are recognized by low-abundance tRNA isoacceptors are decoded more slowly than those recognized by high-abundance tRNAs (19,42,46,48,51,54,66,68,140,149). Codon decoding rates are illustrated by the following colouring scheme - dark blue represents fast, red represents slow, light blue represents neutral.

As mentioned, one of the future challenges in the field is developing a whole cell simulation related to all mRNA translation aspects. Today there are models that consider the competition among many mRNAs (51–52,129,148–150,156–157) but omit additional central phenomena (e.g. interaction with the mRNA folding, the nascent peptide, ribosome drop off, initiation mechanisms, ribosome recycling, ribosome frame shift); some of these aspects were analyzed in simpler, single mRNA based, mod-

els (e.g. see (46,53–55,57,59,61,66,69,131,135,137–138,142–144,147)). A later challenge includes developing comprehensive detailed models that connect all the gene expression steps and not only translation.

Another challenge is related to developing organism and tissue specific models of translation. As mentioned, a useful biophysical model of translation should include parameters that reflect the intracellular regime (e.g. number of ribosomes, concentration of translation factors, de-

coding rates, etc.), since these parameters vary among different organisms and cellular conditions an important step in the field will be to infer them for various organisms/tissues/conditions. Specifically, since most of the studies in the field are based on popular model organisms such as *E. coli* and *S. cerevisiae*, it will be important to develop such models for organisms from the three domains of life.

The organism specific models should provide organism/tissue specific answers to the biological questions and debates mentioned above (see Figure 3).

Moreover, there are central aspects related to translation that are today not considered in any of the models. The limited resources mentioned in the context of translation models in this review include mainly the ribosome pool and other translation factors. However, the cellular energy, i.e. ATP, is also a very important rate limiting factor for translation initiation and elongation which is not studied explicitly in our context.

Translation is known to be the most energy consuming intracellular process, consuming most (up to 75%) of the cellular energy (196).

Thus, it makes sense to develop models that explicitly consider the metabolic cost of translation together with the dynamics of the process. These models can, for example integrate metabolic modelling (197) with translation dynamics models to answer and consider aspects such as: the metabolic cost of traffic jams (e.g. due to ribosome drop off, and the energetic cost of jammed ribosomes); the energetic cost of synthesizing ribosomes (i.e. translating the ribosomal proteins); the energetic cost of generating other translation factors (e.g. tRNAs, ribosomal RNA); the relation/connection between competition of mRNAs for translation factors and their competition for cellular energy.

Initial attempts in this direction have been recently suggested (198); however, these studies haven't considered the fundamental questions mentioned above.

The energetic aspects of translation, specifically its very high energetic cost, are clearly the central relation between the intracellular dynamics of translation and its effect on the way evolution shapes transcripts. Thus, considering explicitly energy consumption as a rate limiting resource is particularly important for molecular evolution studies that study translation via its biophysical modelling.

Another central challenge in the field is related to studying/modelling translation dynamics *not* under the assumption of steady state. While there are some theoretical papers that aimed at studying the dynamics of the translation models (199–201), currently there are no translation modelling studies that are based on biological data which focus on questions related to the dynamics of translation. The studies overviewed in this review focused on reporting steady state estimations related to the translation process. The assumptions in all these cases are that the number of translation factors, mRNAs, ribosomes, etc, are close to constant (small fluctuations around steady state), the mRNA life times are much longer than their translation times, and translation is studied not during the cell cycle or major changes in the system (e.g. G0—cell cycle arrest), etc. Thus, typical values of variables such as translation rates,

initiation rates, and the number of ribosomes can be estimated and studied.

Little attention has been dedicated in the literature to study the processes beyond steady-state. For example, during transition from one state to the other of cell-cycle/cell division/cell growth we expect to observe significant absolute/relative changes in the levels of translation factors such as tRNAs, elongation factors, initiation factors, number of ribosomes, number of mRNAs and more (5,167,202–204). Thus, if we study the translation dynamics during the cell cycle it does not make sense to assume steady state and compute single typical values for each of the state variables during the entire cell cycle. Currently the main bottleneck related to the study of translation dynamics via computational biophysical models not during steady state is the data—there are no high resolution measurements of variables (e.g. tRNA levels, number of ribosomes, Ribo-Seq, mRNA levels, etc.) during different states of the cell, such as at different cell cycle steps, that enable inferring the relevant model parameters. In addition, it is important to remember that the data available today related to translation measurements (e.g. Ribo-Seq) is an average over many cells/mRNAs; accurate measurements of translation in real time (205–207) or single cell/single transcript measurements of translation should contribute towards the accurate modelling of translation dynamics without the steady state assumption.

ACKNOWLEDGEMENT

FUNDING

REFERENCES

1. Vogel,C., de Sousa Abreu,R., Ko,D., Le,S.Y., Shapiro,B.A., Burns,S.C., Sandhu,D., Boutz,D.R., Marcotte,E.M. and Penalva,L.O. (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.*, **6**, 400.
2. Tuller,T., Kupiec,M. and Ruppin,E. (2007) Determinants of protein abundance and translation efficiency in S. cerevisiae. *PLoS Comput. Biol.*, **3**, e248.
3. Ghaemmaghami,S., Huh,W.K., Bower,K., Howson,R.W., Belle,A., Dephoure,N., O'Shea,E.K. and Weissman,J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.
4. Cheng,Z., Teo,G., Krueger,S., Rock,T.M., Koh,H.W., Choi,H. and Vogel,C. (2016) Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.*, **12**, 855.
5. Alberts,B., Johnson,A., Lewis,J., Morgan,D., Raff,M., Roberts,K. and Walter,P. (2014) *Molecular Biology of the Cell*. Garland Science, NY.
6. Plotkin,J.B. and Kudla,G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
7. Tuller,T., Carmi,A., Vestsigian,K., Navon,S., Dorfan,Y., Zaborske,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.

8. Stergachis,A.B., Haugen,E., Shafer,A., Fu,W., Vernot,B., Reynolds,A., Raubitschek,A., Ziegler,S., LeProust,E.M. and Akey,J.M. (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Science*, **342**, 1367–1372.

9. Kimchi-Sarfaty,C., Oh,J.M., Kim,I.W., Sauna,Z.E., Calcagno,A.M., Ambudkar,S.V. and Gottesman,M.M. (2007) A 'silent' polymorphism in the MDR1 gene changes substrate specificity. *Science*, **315**, 525.

10. Zhang,F., Saha,S., Shabalina,S.A. and Kashina,A. (2010) Differential arginylation of actin isoforms is regulated by coding sequence–dependent degradation. *Science*, **329**, 1534.

11. Iost,I. and Dreyfus,M. (1995) The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.* **14**, 3252.

12. Fredrick,K. and Ibba,M. (2010) How the sequence of a gene can tune its translation. *Cell*, **141**, 227–229.

13. Zhou,T., Weems,M. and Wilke,C.O. (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol. Biol. Evol.*, **26**, 1571–1580.

14. Fulle,S. and Gohlke,H. (2009) Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. *J. Mol. Biol.*, **387**, 502–517.

15. O'Brien,E.P., Vendruscolo,M. and Dobson,C.M. (2014) Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat. Commun.*, **5**, 2988.

16. Nissley,D.A. and O'Brien,E.P. (2014) Timing is everything: Unifying codon translation rates and nascent proteome behavior. *J. Am. Chem. Soc.*, **136**, 17892–17898.

17. Bahir,I., Fromer,M., Prat,Y. and Linial,M. (2009) Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.*, **5**, 311.

18. van Weringh,A., Ragonnet-Cronin,M., Pranckeviciene,E., Pavon-Eternod,M., Kleiman,L. and Xia,X. (2011) HIV-1 modulates the tRNA pool to improve translation efficiency. *Mol. Biol. Evol.*, **28**, 1827.

19. Shah,P. and Gilchrist,M.A. (2010) Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet.*, **6**, e1001128.

20. Bulmer,M. (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics*, **129**, 897.

21. Sharp,P. and Li,W.H. (1987) The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol. Biol. Evol.*, **4**, 222–230.

22. Vogel,C. (2013) Protein expression under pressure. *Science*, **342**, 1052–1053.

23. Silvera,D., Formenti,S.C. and Schneider,R.J. (2010) Translational control in cancer. *Nat. Rev. Cancer*, **10**, 254–266.

24. Arava,Y., Wang,Y., Storey,J.D., Liu,C.L., Brown,P.O. and Herschlag,D. (2003) Genome-wide analysis of mRNA translation profiles in Saccharomycescerevisiae. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 3889.

25. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218.

26. Dittmar,K.A., Sørensen,M.A., Elf,J., Ehrenberg,M. and Pan,T. (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.*, **6**, 151–157.

27. Uemura,S., Aitken,C.E.Ã., Korlach,J., Flusberg,B.A., Turner,S.W. and Puglisi,J.D. (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature*, **464**, 1012–1017.

28. Taniguchi,Y., Choi,P.J., Li,G.W., Chen,H., Babu,M., Hearn,J., Emili,A. and Xie,X.S. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533.

29. Newman,J.R.S., Ghaemmaghami,S., Ihmels,J., Breslow,D.K., Noble,M., DeRisi,J.L. and Weissman,J.S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, **441**, 840–846.

30. Lu,P., Vogel,C., Wang,R., Yao,X. and Marcotte,E.M. (2006) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.*, **25**, 117–124.

31. Baerenfaller,K., Grossmann,J., Grobei,M.A., Hull,R., Hirsch-Hoffmann,M., Yalovsky,S., Zimmermann,P., Grossniklaus,U., Gruissem,W. and Baginsky,S. (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science*, **320**, 938.

32. Wen,J.D., Lancaster,L., Hodges,C., Zeri,A.C., Yoshimura,S.H., Noller,H.F., Bustamante,C. and Tinoco,I. (2008) Following translation by single ribosomes one codon at a time. *Nature*, **452**, 598–603.

33. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.

34. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.

35. Lackner,D.H., Beilharz,T.H., Marguerat,S., Mata,J., Watt,S., Schubert,F., Preiss,T. and Bahler,J. (2007) A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell*, **26**, 145–155.

36. Shalem,O., Dahan,O., Levo,M., Martinez,M.R., Furman,I., Segal,E. and Pilpel,Y. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.*, **4**, 223.

37. Narsai,R., Howell,K.A., Millar,A.H., O'Toole,N., Small,I. and Whelan,J. (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell.*, **19**, 3418–3436.

38. Futcher,B., Latter,G., Monardo,P., McLaughlin,C. and Garrels,J. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.*, **19**, 7357.

39. Dittmar,K.A., Goodenbour,J.M. and Pan,T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, e221.

40. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255.

41. Aviner,R., Geiger,T. and Elroy-Stein,O. (2013) Novel proteomic approach (PUNCH-P) reveals cell cycle-specific fluctuations in mRNA translation. *Genes Dev.*, **27**, 1834–1844.

42. 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.

43. Karim,M.M., Hughes,J.M.X., Warwicker,J., Scheper,G.C., Proud,C.G. and McCarthy,J.E.G. (2001) A quantitative molecular model for modulation of mammalian translation by the eIF4E-binding protein 1. *J. Biol. Chem.*, **276**, 20750–20757.

44. Gilchrist,M.A. (2007) Combining models of protein translation and population genetics to predict protein production rates from codon usage patterns. *Mol. Biol. Evol.*, **24**, 2362–2372.

45. Huang,T., Wan,S., Xu,Z., Zheng,Y., Feng,K.Y., Li,H.P., Kong,X. and Cai,Y.D. (2011) Analysis and prediction of translation rate based on sequence and functional features of the mRNA. *PLoS One*, **6**, e16036.

46. Reuveni,S., Meilijson,I., Kupiec,M., Ruppin,E. and Tuller,T. (2011) Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput. Biol.*, **7**, e1002127.

47. Brockmann,R., Beyer,A., Heinisch,J.J. and Wilhelm,T. (2007) Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput. Biol.*, **3**, e57.

48. dos Reis,M., Savva,R. and Wernisch,L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.

49. Jansen,R., Bussemaker,H.J. and Gerstein,M. (2003) Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res.*, **31**, 2242–2251.

50. Tuller,T., Waldman,Y., Kupiec,M. and Ruppin,E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3645.

51. Zhang,G., Fedyunin,I., Miekley,O., Valleriani,A., Moura,A. and Ignatova,Z. (2010) Global and local depletion of ternary complex limits translational elongation. *Nucleic Acids Res.*, **38**, 4778–4787.

52. Mehra,A. and Hatzimanikatis,V. (2006) An algorithmic framework for genome-wide modeling and analysis of translation networks. *Biophys. J.*, **90**, 1136–1146.

53. 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.

54. Ciandrini,L., Stansfield,I. and Romano,M.C. (2013) Ribosome traffic on mRNAs maps to gene ontology: genome-wide quantification of translation initiation rates and polysome size regulation. *PLoS Comput Biol*, **9**, e1002866.

55. Dimelow,R.J. and Wilkinson,S.J. (2009) Control of translation initiation: a model-based analysis from limited experimental data. *J. Roy. Soc. Interface*, **6**, 51–61.

56. Fluitt,A., Pienaar,E. and Viljoen,H. (2007) Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.*, **31**, 335–346.

57. Zhang,G., Hubalewska,M. and Ignatova,Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.*, **16**, 274–280.

58. Chou,T. (2003) Ribosome recycling, diffusion, and mRNA loop formation in translational regulation. *Biophys. J.*, **85**, 755–773.

59. Sharma,A.K. and Chowdhury,D. (2011) Distribution of dwell times of a ribosome: effects of infidelity, kinetic proofreading and ribosome crowding. *Phys. Biol.*, **8**, 026005.

60. Racle,J., Picard,F., Girbal,L., Cocaign-Bousquet,M. and Hatzimanikatis,V. (2013) A genome-scale integration and analysis of Lactococcus lactis translation data. *PLoS Comput. Biol.*, **9**, e1003240.

61. Tuller,T., Veksler-Lublinsky,I., Gazit,N., Kupiec,M., Ruppin,E. and Ziv-Ukelson,M. (2011) Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.*, **12**, R110.

62. Shah,P., Ding,Y., Niemczyk,M., Kudla,G. and Plotkin,J.B. (2013) Rate-limiting steps in yeast protein translation. *Cell*, **153**, 1589–1601.

63. Berthelot,K., Muldoon,M., Rajkowitsch,L., Hughes,J. and McCarthy,J.E. (2004) Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.*, **51**, 987–1001.

64. Kotlar,D. and Lavner,Y. (2006) The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genomics*, **7**, 67.

65. Gilchrist,M.A. and Wagner,A. (2006) A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *J. Theor. Biol.*, **239**, 417–434.

66. Zhang,G. and Ignatova,Z. (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS One*, **4**, e5036.

67. Siwiak,M. and Zielenkiewicz,P. (2010) A comprehensive, quantitative, and genome-wide model of translation. *PLoS Comput. Biol.*, **6**, e1000865.

68. Pechmann,S. and Frydman,J. (2012) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.*, **20**, 237–243.

69. Mitarai,N. and Pedersen,S. (2013) Control of ribosome traffic by position-dependent choice of synonymous codons. *Phys. Biol.*, **10**, 056011.

70. Deneke,C., Lipowsky,R. and Valleriani,A. (2013) Effect of ribosome shielding on mRNA stability. *Phys. Biol.*, **10**, 046008.

71. von der Haar,T. (2012) Mathematical and computational modelling of ribosomal movement and protein synthesis: an overview. *Comput. Struct. Biotechnol. J.*, **1**, e201204002.

72. Zia,R.K., Dong,J. and Schmittmann,B. (2011) Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *J. Stat. Phys.*, **144**, 405–428.

73. Churchill,G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.*, **32**, 490–495.

74. Huh,W.-K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.

75. Dittmar,K.A., Mobley,E.M., Radek,A.J. and Pan,T. (2004) Exploring the regulation of tRNA distribution on the genomic scale. *J. Mol. Biol.*, **337**, 31–47.

76. Liao,J.Y., Ma,L.M., Guo,Y.H., Zhang,Y.C., Zhou,H., Shao,P., Chen,Y.Q. and Qu,L.H. (2010) Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ trailers. *PLoS One*, **5**, e10563.

77. Kawaguchi,R., Girke,T., Bray,E.A. and Bailey-Serres,J. (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *Plant J.*, **38**, 823–839.

78. Mustroph,A., Zanetti,M.E., Jang,C.J.H., Holtan,H.E., Repetti,P.P., Galbraith,D.W., Girke,T. and Bailey-Serres,J. (2009) Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 18843–18848.

79. Aspden,J.L., Eyre-Walker,Y.C., Phillips,R.J., Amin,U., Mumtaz,M.A.S., Brocard,M. and Couso,J.-P. (2014) Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *Elife*, **3**, e03528.

80. Michel,A.M. and Baranov,P.V. (2013) Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip. Rev. RNA*, **4**, 473–490.

81. Bazzini,A.A., Johnstone,T.G., Christiano,R., Mackowiak,S.D., Obermayer,B., Fleming,E.S., Vejnar,C.E., Lee,M.T., Rajewsky,N. and Walther,T.C. (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.*, e201488411.

82. Smith,J.E., Alvarez-Dominguez,J.R., Kline,N., Huynh,N.J., Geisler,S., Hu,W., Coller,J. and Baker,K.E. (2014) Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. *Cell Rep.*, **7**, 1858–1866.

83. Zupanic,A., Meplan,C., Grellscheid,S.N., Mathers,J.C., Kirkwood,T.B., Hesketh,J.E. and Shanley,D.P. (2014) Detecting translational regulation by change point analysis of ribosome profiling data sets. *RNA*, **20**, 1507–1518.

84. Gerashchenko,M.V., Lobanov,A.V. and Gladyshev,V.N. (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 17394–17399.

85. Menschaert,G., Van Criekinge,W., Notelaers,T., Koch,A., Crappé,J., Gevaert,K. and Van Damme,P. (2013) Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol. Cell. Proteomics*, **12**, 1780–1790.

86. Juntawong,P., Girke,T., Bazin,J. and Bailey-Serres,J. (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E203–E212.

87. Han,Y., Gao,X., Liu,B., Wan,J., Zhang,X. and Qian,S.-B. (2014) Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Res.*, **24**, 842–851.

88. Guttman,M., Russell,P., Ingolia,N.T., Weissman,J.S. and Lander,E.S. (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, **154**, 240–251.

89. Brar,G.A., Yassour,M., Friedman,N., Regev,A., Ingolia,N.T. and Weissman,J.S. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*, **335**, 552–557.

90. Ingolia,N.T., Brar,G.A., Stern-Ginossar,N., Harris,M.S., Talhouarne,G.J., Jackson,S.E., Wills,M.R. and Weissman,J.S. (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.*, **8**, 1365–1379.

91. Michel,A.M., Choudhury,K.R., Firth,A.E., Ingolia,N.T., Atkins,J.F. and Baranov,P.V. (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.*, **22**, 2219–2229.

92. Lee,S., Liu,B., Lee,S., Huang,S.-X., Shen,B. and Qian,S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2424–E2432.

93. Ingolia,N.T., Lareau,L.F. and Weissman,J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

94. Yewdell,J.W., Lacsina,J.R., Rechsteiner,M.C. and Nicchitta,C.V. (2011) Out with the old, in with the new? Comparing methods for measuring protein degradation. *Cell Biol. Int.*, **35**, 457–462.

95. Belle,A., Tanay,A., Bitincka,L., Shamir,R. and O'Shea,E.K. (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 13004–13009.

96. Johnson,A.E. (2005) The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer. *FEBS Lett.*, **579**, 916–920.

97. Clark,P.L. and Ugrinov,K.G. (2009) Measuring cotranslational folding of nascent polypeptide chains on ribosomes. *Methods Enzymol.*, **466**, 567–590.

98. Tuller,T. and Zur,H. (2015) Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Res.*, **43**, 13–28.

99. Ben-Yehezkel,T., Atar,S., Zur,H., Diament,A., Goz,E., Marx,T., Cohen,R., Dana,A., Feldman,A. and Shapiro,E. (2015) Rationally designed, heterologous S. cerevisiae transcripts expose novel expression determinants. *RNA Biol.*, **12**, 972–984.

100. Tuller,T. (2014) Challenges and obstacles related to solving the codon bias riddles. *Biochem. Soc. Trans.*, **42**, 155–159.

101. Burgess-Brown,N.A., Sharma,S., Sobott,F., Loenarz,C., Oppermann,U. and Gileadi,O. (2008) Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr. Purif.*, **59**, 94–102.

102. Welch,M., Govindarajan,S., Ness,J.E., Villalobos,A., Gurney,A., Minshull,J. and Gustafsson,C. (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One*, **4**, e7002.

103. Goodman,D.B., Church,G.M. and Kosuri,S. (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science*, **342**, 475–479.

104. Ceroni,F., Algar,R., Stan,G.-B. and Ellis,T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods*, **12**, 415–418.

105. Diament,A. and Tuller,T. (2016) Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biol. Direct*, **11**, 1.

106. Greenbaum,D., Jansen,R. and Gerstein,M. (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics*, **18**, 585–596.

107. Lithwick,G. and Margalit,H. (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res.*, **13**, 2665–2673.

108. MacDonald,C.T., Gibbs,J.H. and Pipkin,A.C. (1968) Kinetics of biopolymerization on nucleic acid templates. *Biopolymers*, **6**, 1–5.

109. Heinrich,R. and Rapoport,T.A. (1980) Mathematical modelling of translation of mRNA in eucaryotes; steady states, time-dependent processes and application to reticulocytest. *J. Theor. Biol.*, **86**, 279–313.

110. Margaliot,M. and Tuller,T. (2012) Stability analysis of the ribosome flow model. *IEEE/ACM Trans. Comput. Biol. Bioinformatics (TCBB)*, **9**, 1545–1552.

111. Shaw,L.B., Zia,R. and Lee,K.H. (2003) Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Phys. Rev. E.*, **68**, 021910.

112. MacDonald,C.T., Gibbs,J.H. and Pipkin,A.C. (1968) Kinetics of biopolymerization on nucleic acid templates. *Biopolymers*, **6**, 1–25.

113. Sabi,R. and Tuller,T. (2014) Modelling the efficiency of codon–tRNA interactions based on codon usage bias. *DNA Res.*, **21**, 511–526.

114. Sharp,P.M. and Li,W.-H. (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.

115. Ikemura,T. (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.*, **151**, 389–409.

116. Bennetzen,J.L. and Hall,B. (1982) Codon selection in yeast. *J. Biol. Chem.*, **257**, 3026–3031.

117. 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|Genetics*, **5**, 73–80.

118. Zur,H. and Tuller,T. (2013) Transcript features alone enable accurate prediction and understanding of gene expression in S. cerevisiae. *BMC Bioinform.*, **14**, S1.

119. Line,S.R., Liu,X., de Souza,A.P. and Yu,F. (2013) Translational signatures and mRNA levels are highly correlated in human stably expressed genes. *BMC Genomics*, **14**, 268.

120. Warner,J.R., Rich,A. and Hall,C.E. (1962) Electron microscope studies of ribosomal clusters synthesizing hemoglobin. *Science*, **138**, 1399–1403.

121. Simha,R., Zimmerman,J. and Moacanin,J. (1963) Polymerization kinetics of biological macromolecules on templates. *J. Chem. Phys.*, **39**, 1239.

122. Zimmerman,J. and Simha,R. (1965) The kinetics of multicenter macromolecule growth along a template. *J. Theor. Biol.*, **9**, 156–185.

123. MacDonald,C.T. and Gibbs,J.H. (1969) Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers*, **7**, 707–725.

124. Hiernaux,J. (1974) On some stochastic models for protein biosynthesis. *Biophys. Chem.*, **2**, 70–75.

125. Ordon,R. (1969) Polyribosome dynamics at steady state. *J. Theor. Biol.*, **22**, 515–532.

126. Vassart,G., Dumont,J. and Cantraine,F. (1971) Translational control of protein synthesis: a simulation study. *Biochim. Biophys. Acta (BBA)-Nucleic Acids Protein Synth.*, **247**, 471–485.

127. Li,K., Kisilevsky,R., Wasan,M. and Hammond,G. (1972) A computer simulation of in vivo protein synthesis. *Biochim. Biophys. Acta (BBA)-Nucleic Acids Protein Synth.*, **272**, 451–462.

128. Zhang,S., Goldman,E. and Zubay,G. (1994) Clustering of low usage codons and ribosome movement. *J. Theor. Biol.*, **170**, 339–354.

129. Raveh,A., Margaliot,M., Sontag,E.D. and Tuller,T. (2016) A model for competition for ribosomes in the cell. *J. R. Soc. Interface*, **13**, 20151062.

130. Shaw,L.B., Sethna,J.P. and Lee,K.H. (2004) Mean-field approaches to the totally asymmetric exclusion process with quenched disorder and large particles. *Phys. Rev. E*, **70**, 021901.

131. Chou,T. and Lakatos,G. (2004) Clustered bottlenecks in mRNA translation and protein synthesis. *Phys. Rev. Lett.*, **93**, 198101.

132. Zouridis,H. and Hatzimanikatis,V. (2008) Effects of codon distributions and tRNA competition on protein translation. *Biophys. J.*, **95**, 1018–1033.

133. Mier-y-Terán-Romero,L., Silber,M. and Hatzimanikatis,V. (2010) The origins of time-delay in template biopolymerization processes. *PLoS Comput. Biol.*, **6**, e1000726.

134. Sharma,A.K. and Chowdhury,D. (2011) Stochastic theory of protein synthesis and polysome: Ribosome profile on a single mRNA transcript. *J. Theor. Biol.*, **289**, 36–46.

135. Basu,A. and Chowdhury,D. (2007) Traffic of interacting ribosomes: effects of single-machine mechanochemistry on protein synthesis. *Phys. Rev. E*, **75**, 021902.

136. Heyd,A. and Drew,D.A. (2003) A mathematical model for elongation of a peptide chain. *Bull. Math. Biol.*, **65**, 1095–1109.

137. Godefroy-Colburn,T. and Thach,R. (1981) The role of mRNA competition in regulating translation. IV. Kinetic model. *J. Biol. Chem.*, **256**, 11762–11773.

138. Von Heijne,G., Nilsson,L. and Blomberg,C. (1977) Translation and messenger RNA secondary structure. *J. Theor. Biol.*, **68**, 321–329.

139. Mao,Y., Liu,H., Liu,Y. and Tao,S. (2014) Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **42**, 4813–4822.

140. Chu,D., Barnes,D.J. and von der Haar,T. (2011) The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **39**, 6705–6714.

141. Skjøndal-Bara,N. and Morrisb,D.R. (2007) Dynamic model of the process of protein synthesis in eukaryotic cells. *Bull. Math. Biol.*, **69**, 361–393.

142. Dong,J., Schmittmann,B. and Zia,R.K. (2007) Towards a model for protein production rates. *J. Stat. Phys.*, **128**, 21–34.

143. Artieri,C.G. and Fraser,H.B. (2014) Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.*, **24**, 2011–2021.

144. Lu,J. and Deutsch,C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.*, **384**, 73–86.

145. Marshall,E., Stansfield,I. and Romano,M. (2014) Ribosome recycling induces optimal translation rate at low ribosomal availability. *J. Roy. Soc. Interface*, **11**, 20140589.

146. Harger,J.W., Meskauskas,A. and Dinman,J.D. (2002) An 'integrated model'of programmed ribosomal frameshifting. *Trends Biochem. Sci.*, **27**, 448–454.

147. Margaliot,M. and Tuller,T. (2013) Ribosome flow model with positive feedback. *J. Roy. Soc. Interface*, **10**, 20130267.

148. Dana,A. and Tuller,T. (2012) Efficient manipulations of synonymous mutations for controlling translation rate: an analytical approach. *J. Comput. Biol.*, **19**, 200–231.

149. Brackley,C.A., Romano,M.C. and Thiel,M. (2011) The dynamics of supply and demand in mRNA translation. *PLoS Comput. Biol.*, **7**, e1002203.

150. Chu,D., Zabet,N. and von der Haar,T. (2012) A novel and versatile computational tool to model translation. *Bioinformatics*, **28**, 292–293.

151. Cook,L.J. and Zia,R. (2009) Feedback and fluctuations in a totally asymmetric simple exclusion process with finite resources. *J. Stat. Mech.: Theory Exp.*, P02012.

152. Cook,L.J., Zia,R. and Schmittmann,B. (2009) Competition between multiple totally asymmetric simple exclusion processes for a finite pool of resources. *Phys. Rev. E*, **80**, 031142.

153. Greulich,P., Ciandrini,L., Allen,R.J. and Romano,M.C. (2012) Mixed population of competing totally asymmetric simple exclusion processes with a shared reservoir of particles. *Phys. Rev. E*, **85**, 011142.

154. Heldt,S. and Thiel,M. (2009), Otto-von-Guericke Universität Magdeburg.

155. Cook,L.J. and Zia,R. (2012) Competition for finite resources. *J. Stat. Mech.: Theory Exp.*, P05008.

156. Karr,J.R., Sanghvi,J.C., Macklin,D.N., Gutschow,M.V., Jacobs,J.M., Bolival,B., Assad-Garcia,N., Glass,J.I. and Covert,M.W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell*, **150**, 389–401.

157. Mather,W.H., Hasty,J., Tsimring,L.S. and Williams,R.J. (2013) Translational cross talk in gene networks. *Biophys. J.*, **104**, 2564–2572.

158. Chu,D. and von der Haar,T. (2012) The architecture of eukaryotic translation. *Nucleic Acids Res.*, gks825.

159. Roberts,E., Magis,A., Ortiz,J.O., Baumeister,W. and Luthey-Schulten,Z. (2011) Noise contributions in an inducible genetic switch: a whole-cell simulation study. *PLoS Comput. Biol.*, **7**, 1002010.

160. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.

161. Dana,A. and Tuller,T. (2014) The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.*, **42**, 9171–9181.

162. Gardin,J., Yeasmin,R., Yurovsky,A., Cai,Y., Skiena,S. and Futcher,B. (2014) Measurement of average decoding rates of the 61 sense codons in vivo. *Elife*, **3**, e03735.

163. Gritsenko,A.A., Hulsman,M., Reinders,M.J. and de Ridder,D. (2015) Unbiased quantitative models of protein translation derived from ribosome profiling data. *PLoS Comput. Biol.*, **11**, e1004336.

164. Zur,H. and Tuller,T. (2013) New universal rules of eukaryotic translation initiation fidelity. *PLoS. Comput. Biol.*, **9**, e1003136.

165. Sin,C., Chiarugi,D. and Valleriani,A. (2016) Quantitative assessment of ribosome drop-off in E. coli. *Nucleic Acids Res.*, **44**, 2528–2537.

166. Milo,R., Jorgensen,P., Moran,U., Weber,G. and Springer,M. (2010) BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.*, **38**, D750–D753.

167. Frenkel-Morgenstern,M., Danon,T., Christian,T., Igarashi,T., Cohen,L., Hou,Y.-M. and Jensen,L.J. (2012) Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.*, **8**, 572.

168. Tuller,T., Waldman,Y.Y., Kupiec,M. and Ruppin,E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3645–3650.

169. Nissley,D.A., Sharma,A.K., Ahmed,N., Friedrich,U.A., Kramer,G., Bukau,B. and O'Brien,E.P. (2016) Accurate prediction of cellular co-translational folding indicates proteins can switch from post-to co-translational folding. *Nature communications*, **7**, 10341.

170. Shah,P. and Gilchrist,M.A. (2011) Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10231.

171. Lavner,Y. and Kotlar,D. (2005) Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene*, **345**, 127–138.

172. Kolomeisky,A.B., Schütz,G.M., Kolomeisky,E.B. and Straley,J.P. (1998) Phase diagram of one-dimensional driven lattice gases with open boundaries. *J. Phys. A: Math. Gen.*, **31**, 6911.

173. Kopeina,G.S., Afonina,Z.A., Gromova,K.V., Shirokov,V.A., Vasiliev,V.D. and Spirin,A.S. (2008) Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. *Nucleic Acids Res.*, **36**, 2476–2488.

174. Bergmann,J.E. and Lodish,H.F. (1979) A kinetic model of protein synthesis. Application to hemoglobin synthesis and translational control. *J. Biol. Chem.*, **254**, 11927–11937.

175. Liljenström,H. and von Heijne,G. (1987) Translation rate modification by preferential codon usage: intragenic position effects. *J. Theor. Biol.*, **124**, 43–55.

176. Dong,J., Schmittmann,B. and Zia,R.K. (2007) Inhomogeneous exclusion processes with extended objects: the effect of defect locations. *Phys. Rev. E*, **76**, 051113.

177. Varenne,S. and Lazdunski,C. (1986) Effect of distribution of unfavourable codons on the maximum rate of gene expression by an heterologous organism. *J. Theor. Biol.*, **120**, 99–110.

178. Gu,W., Zhou,T. and Wilke,C.O. (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.*, **6**, e1000664.

179. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.

180. Zur,H. and Tuller,T. (2012) Strong association between mRNA folding strength and protein abundance in S. cerevisiae. *EMBO Rep.*, **13**, 272–277.

181. Mortimer,S.A., Kidwell,M.A. and Doudna,J.A. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, **15**, 469–479.

182. Park,C., Chen,X., Yang,J.-R. and Zhang,J. (2013) Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E678–E686.

183. Babyak,M.A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.*, **66**, 411–421.

184. Singh,U. (1969) Polyribosomes and unstable messenger RNA: a stochastic model of protein synthesis. *J. Theor. Biol.*, **25**, 444–460.

185. Menninger,J.R. (1983) Computer simulation of ribosome editing. *J. Mol. Biol.*, **171**, 383–399.

186. Adams,D., Schmittmann,B. and Zia,R. (2008) Far-from-equilibrium transport with constrained resources. *J. Stat. Mech.: Theory Exp.*, P06009.

187. Blythe,R. and Evans,M. (2007) Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A: Math. Theor.*, **40**, R333.

188. Poker,G., Zarai,Y., Margaliot,M. and Tuller,T. (2014) Maximizing protein translation rate in the non-homogeneous ribosome flow model: a convex optimization approach. *J. Roy. Soc. Interface*, **11**, 20140713.

189. Poker,G., Margaliot,M. and Tuller,T. (2015) Sensitivity of mRNA Translation. *Sci. Rep.*, **5**, 12795.

190. Pierobon,P., Mobilia,M., Kouyos,R. and Frey,E. (2006) Bottleneck-induced transitions in a minimal model for intracellular transport. *Phys. Rev. E*, **74**, 031906.

191. Zur,H. and Tuller,T. (2012) RFMapp: ribosome flow model application. *Bioinformatics*, **28**, 1663–1664.

192. Edri,S., Gazit,E., Cohen,E. and Tuller,T. (2014) The RNA polymerase flow model of gene transcription. *Biomed. Circuits Syst. IEEE Trans.*, **8**, 54–64.

193. Neri,I., Kern,N. and Parmeggiani,A. (2013) Exclusion processes on networks as models for cytoskeletal transport. *N. J. Phys.*, **15**, 085005.

194. Seo,S.W., Yang,J.-S., Kim,I., Yang,J., Min,B.E., Kim,S. and Jung,G.Y. (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.*, **15**, 67–74.

195. Na,D., Lee,S. and Lee,D. (2010) Mathematical modeling of translation initiation for the estimation of its efficiency to

computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst, Biol.*, **4**, 1.

196. Lane,N. and Martin,W. (2010) The energetics of genome complexity. *Nature*, **467**, 929–934.
197. Bordbar,A., Monk,J.M., King,Z.A. and Palsson,B.O. (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.*, **15**, 107–120.
198. Thiele,I., Fleming,R.M., Que,R., Bordbar,A., Diep,D. and Palsson,B.O. (2012) Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. *PLoS One*, **7**, e45635.
199. Margaliot,M., Sontag,E.D. and Tuller,T. (2014) Entrainment to periodic initiation and transition rates in a computational model for gene translation. *PLoS One*, **9**, e96039.
200. De Jong,H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, **9**, 67–103.
201. Nagar,A., Valleriani,A. and Lipowsky,R. (2011) Translation by ribosomes with mRNA degradation: Exclusion processes on aging tracks. *J. Stat. Phys.*, **145**, 1385–1404.
202. Higareda-Mendoza,A.E. and Pardo-Galván,M.A. (2010) Expression of human eukaryotic initiation factor 3f oscillates with cell cycle in A549 cells and is essential for cell viability. *Cell Division*, **5**, 1.
203. Patil,A., Dyavaiah,M., Joseph,F., Rooney,J.P., Chan,C.T., Dedon,P.C. and Begley,T.J. (2012) Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. *Cell Cycle*, **11**, 3656–3665.
204. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
205. Yan,X., Hoek,T.A., Vale,R.D. and Tanenbaum,M.E. (2016) Dynamics of translation of single mRNA molecules in vivo. *Cell*, **165**, 976–989.
206. Wu,B., Eliscovich,C., Yoon,Y. and Singer,R.H. (2016) Translation dynamics of single mRNAs in live cells and neurons. *Science*, **352**, 1430–1435.
207. Morisaki,T., Lyon,K., DeLuca,K.F., DeLuca,J.G., English,B.P., Zhang,Z., Lavis,L.D., Grimm,J.B., Viswanathan,S., Looger,L.L. *et al.* (2016) Real-time quantification of single RNA translation dynamics in living cells. *Science*, **352**, 1425–1429.
208. Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
209. Chu,D., Kazana,E., Bellanger,N., Singh,T., Tuite,M.F. and Haar,T. (2014) Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.*, **33**, 21–34.
210. Warner,J.R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.*, **24**, 437–440.
211. Lauria,F., Tebaldi,T., Lunelli,L., Struffi,P., Gatto,P., Pugliese,A., Brigotti,M., Montanaro,L., Ciribilli,Y. and Inga,A. (2015) RiboAbacus: a model trained on polyribosome images predicts ribosome density and translational efficiency from mammalian transcriptomes. *Nucleic Acids Res.*, **43**, e153.
212. Gorochowski,T.E., Avcilar-Kucukgoze,I., Bovenberg,R.A., Roubos,J.A. and Ignatova,Z. (2016) A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth. Biol.*, **5**, 710–720.