JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Perspective

pubs.acs.org/JACS

Timing Is Everything: Unifying Codon Translation Rates and Nascent Proteome Behavior

Daniel A. Nissley and Edward P. O'Brien*

Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States

ABSTRACT: Experiments have demonstrated that changing the rate at which the ribosome translates a codon position in an mRNA molecule's open reading frame can alter the behavior of the newly synthesized protein. That is, codon translation rates can govern nascent proteome behavior. We emphasize that this phenomenon is a manifestation of the nonequilibrium nature of cotranslational processes, and as such, there exist theoretical tools that offer a potential means to quantitatively predict the influence of codon translation rates on the broad spectrum of nascent protein behaviors including cotranslational folding, aggregation, and translocation. We provide a review of the experimental evidence for the impact that codon translation rates can have, followed by a discussion of theoretical methods that can describe this phenomenon. The development and application of these tools are likely to provide fundamental insights into protein maturation and homeostasis, codon usage bias in organisms, the origins of translation related diseases, and new rational design methods for biotechnology and biopharmaceutical applications.

INTRODUCTION

During the process of translation the ribosome synthesizes a protein molecule by unidirectionally translocating along an mRNA molecule one codon at a time (Figure 1A). A number of processes involving the nascent chain occur before it has been fully synthesized. These processes, referred to as cotranslational processes, include cotranslational folding,[1,2] molecular chaperone binding,[3] translocation between cellular compartments,[4,5] and the ubiquitination[6,7] and glycosylation[8] of the nascent protein. Recently, experiments from a number of different laboratories have demonstrated that changing the rate at which codon positions in an open reading frame (ORF) are translated by the ribosome can dramatically affect such cotranslational processes and consequently alter the fate of the nascent protein *in vivo* (Figure 1B), even though the primary structure of the nascent protein is unaltered. Such results demonstrate that the thermodynamic stability of a protein's native state can be less relevant to its nascent behavior in a cell than the rates of the processes acting on the protein. This property is a hallmark of nonequilibrium processes, wherein changes in kinetics can change the system's behavior despite there being no change in the system's composition.

In this perspective, we emphasize that codon translation rates can govern nascent protein behavior and that the disparate observations supporting this claim are manifestations of the out-of-equilibrium nature of cotranslational processes. Hence, any general theory that attempts to quantitatively predict nascent proteome behaviors must necessarily account for the interplay between codon translation rates and the rates at which these cotranslational processes proceed. In the following sections, we detail recent experimental examples of these phenomena, describe promising theoretical approaches to treating such situations, and discuss how these treatments might be extended to account for the broad spectrum of nascent protein behaviors affected by codon translation rates. Ultimately, the creation of such a theoretical framework will help us better understand the causes and consequences of translation-regulated nascent protein behavior in cells and the origins of various human diseases[9] as well as provide a framework to manipulate nascent proteome behavior *in vivo*.

SYNONYMOUS CODON SUBSTITUTIONS CAN ALTER THE RATE OF TRANSLATION

The genetic code is degenerate, with the 20 naturally occurring amino acids encoded by 61 unique codon types that compose the ORFs of transcripts (Figure 1A). All amino acids, with the exception of methionine and tryptophan, are encoded by at least two and as many as six different codons.[10] Codons that encode for the same amino acid are said to be synonymous with one another. Synonymous codons are not translated by the ribosome at identical rates; experiments indicate that the average translation rates between the synonymous codons GAA and GAG, which both encode glutamic acid, differ 3-fold in *E. coli*.[11] For those codons where translation rates have not yet been measured, theoretical modeling suggests even greater variability may exist.[12] Therefore, the translation rate at a particular codon position in an ORF can be altered by substituting a synonymous codon at that position, leaving the amino acid sequence of the synthesized protein unaltered. Indeed, as discussed below, synonymous codons have been extensively utilized to examine the influence of codon translation rates on nascent protein behavior. Furthermore, there is evidence that evolutionary pressures have shaped codon usage in the transcriptomes of organisms (Figure 1C) in some cases to modulate nascent proteome behavior,[13−16] indicating just how important the influence of codon translation rates can be to an organism's phenotype.

COTRANSLATIONAL FOLDING CAN BE INFLUENCED BY CODON TRANSLATION RATES

Introducing synonymous codon mutations within a transcript, which alters the distribution of translation rates along the ORF,

Received: September 30, 2014
Published: December 8, 2014

dx.doi.org/10.1021/ja510082j | *J. Am. Chem. Soc.* 2014, 136, 17892−17898

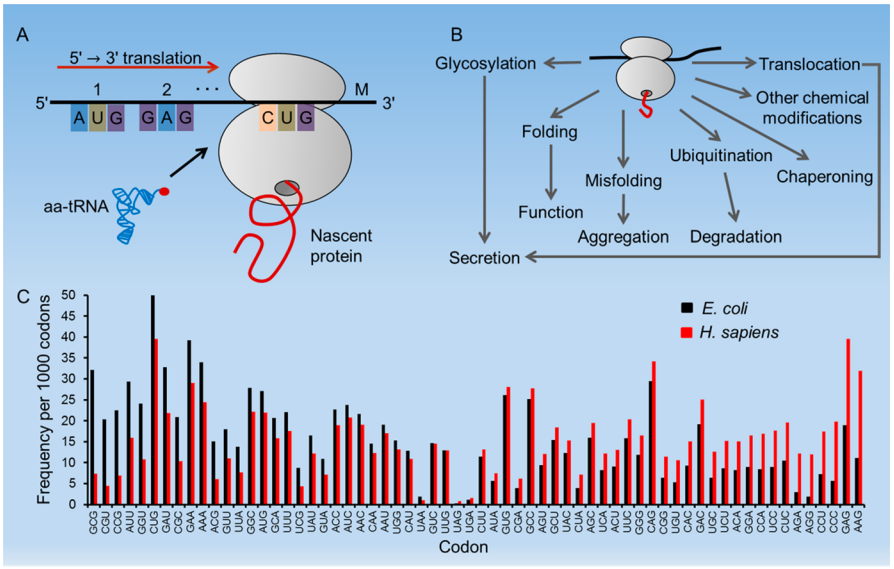

Figure 1. Codon translation rates influence a broad range of nascent protein behaviors. (A) The codons comprising the mRNA are the template directing the ribosome as to which protein sequence to synthesize. Aminoacyl-tRNA (aa-tRNA) delivers the correct amino acid by selectively binding to the codon that is complementary to its anticodon. The nascent protein emerges from the exit tunnel, N-terminus first, as the ribosome translocates along the mRNA in the 5′ to 3′ direction. (B) The range of cotranslational and post-translational processes which may occur for a nascent protein. (C) The codon usage, measured as the frequency of a particular codon per 1000 codons, is shown for the genomes of *E. coli* and *H. sapiens* (reproduced from the NCBI GenBank database).

can change the ability of a nascent protein to cotranslationally fold. The process of cotranslational protein folding consists of the concomitant folding of one or more domains in a protein into a stable tertiary structure during the time it takes to synthesize the full-length protein.[1] Co-translational folding can be a biologically beneficial process because it allows the individual segments that compose multidomain proteins to fold in the absence of nascent chain segments from other domains, thereby minimizing the chances of interdomain misfolding.[1,17,18] Consequently, proteins that fold cotranslationally often display decreased levels of misfolding and aggregation and, in some cases, populate on-pathway structures that enhance the chances that the protein will attain the correct folded structure.[19,20]

Slower codon translation speeds can favor the correct folding of eukaryotic proteins by affording domains more time to fold while bound to the ribosome.[21] For example, the exchange of two rare codons for their most common synonymous codons, which presumably translate more quickly, in a normally slow-translating region within the ORF of the three-domain SufI protein was found to decrease its ability to cotranslationally fold in an *in vitro* synthesis system.[22] In several cases, an increase in cotranslational folding has been facilitated by the positioning of rare codon clusters on the mRNA downstream of the cotranslationally folding domain.[22,23] An extensive bioinformatics study, however, indicates this arrangement of rare codons relative to domain boundaries is not typically found in the transcriptomes of organisms, though there is evidence that protein domain boundaries are enriched in fast-translating codons.[24] A recent theoretical study using kinetic models also suggests that there are scenarios in which fast-translating codons could increase the probability of correct cotranslational folding when they are positioned in the ORF downstream of nascent protein segments prone to misfolding.[25] These results illustrate how altering codon translation rates at specific positions in an ORF can modulate the cotranslational folding of domains in a protein.

CODON TRANSLATION RATES MODULATE PROTEIN FUNCTION, MISFOLDING, AND AGGREGATION

Altering codon translation rates can also alter the ability of a newly synthesized protein to carry out its biological function.[18,22,26−28] So-called "silent" single nucleotide polymorphisms, which are naturally occurring synonymous codon substitutions in the genome of an organism, have been found to affect protein expression levels,[29] the final folded structure a nascent protein attains *in vitro*,[29] and downstream processes such as aggregation[30] and the substrate specificity of newly synthesized proteins.[31] Each of these changes in the behavior of the newly synthesized protein can be explained by the changes that these synonymous codon substitutions have on the rates at which codons are translated.

The expression levels of soluble, functional protein molecules can be controlled by tuning the rates at which codons are translated. One highly successful codon optimization strategy for increasing the yields of eukaryotic proteins expressed recombinantly in prokaryotic cells is to maintain the eukaryotic codon usage profile along the ORF in the context of the prokaryotic organism.[32] The success of this approach is consistent with the notion that by preserving the evolutionarily optimized relative timing at which different segments of a nascent chain are produced, the cotranslational acquisition of correct structure and function can be achieved while minimizing the tendency of the nascent protein to aggregate. Optimizing codon usage, however, is not always beneficial to protein expression. For example, the expression of functional *Neurospora* clock protein FREQUENCY (FRQ) is dependent on the nonoptimal codon usage present within the naturally occurring transcript, and though optimizing the codon usage in the FRQ transcript results in increased yields of protein, it also causes the loss of FRQ's proper periodic expression with time.[29] More generally, a multigene codon optimization study in *E. coli*[33] found that while the majority of the genes studied showed 2- to 3-fold increases in expression relative

dx.doi.org/10.1021/ja510082j | J. Am. Chem. Soc. 2014, 136, 17892−17898

to the wild-type transcripts, 20% had decreased expression levels. These results indicate that the optimization of a codon sequence does not guarantee an increase in successful expression of a given transcript. A more robust understanding of the molecular origin of both variable codon translation rates themselves and their myriad effects on cotranslational processes may enhance the ability of optimization procedures to control nascent protein behavior and increase expression levels.

Nascent protein misfolding, which can be measured experimentally for enzymes as a decreased specific activity in the soluble protein fraction,[13,21,26,27] is also a potential consequence of synonymous codon substitutions. Recombinant expression of eukaryotic proteins (including firefly luciferase and green fluorescent protein) by *E. coli* containing streptomycin−pseudodependent ribosomes, whose global translation elongation rate increases with increasing concentration of streptomycin, was found to produce a greater fraction of correctly folded protein when codon translation rates were slowed.[21] A synonymous codon mutation in the Multidrug Resistance 1 gene was suggested to alter the conformation and drug transport function of the synthesized protein despite resulting in an identical amino acid sequence and comparable levels of expression.[31] In the case of FRQ, it was shown that optimization of its codon sequence also led to a conformational change in mature FRQ and the total abolishment of circadian periodicity within the organism.[29] These, and other experiments,[34] make clear the tangible link between cotranslational protein misfolding, codon translation rates, and subsequent cellular processes that depend on the proper functioning of nascent proteins.

When proteins are unable to reach their correct folded conformation, or any other soluble structure, they can aggregate and precipitate from solution.[35] Such aggregation can be detrimental to a cell and is the factor which unifies the amyloidosis family of diseases.[35] The aggregation propensity of nascent proteins can be altered through modification of codon translation rates, indicating a direct link between cotranslational events and the post-translational process of aggregation. For example, the replacement of three codons in a segment of the *Echinococcus granulosus* fatty acid binding protein1 transcript with synonymous codons resulted in an increase in aggregation *in vivo*,[26] suggesting that aggregation of nascent proteins is more prevalent when certain codons are translated at different rates.[21] It was also shown that optimizing the codon sequence of firefly luciferase's transcript, with all codons being replaced with their fastest-translating synonymous codon, led to an increase in the amount of aggregation relative to the wild-type codon sequence at comparable levels of expression in *E. coli*.[13] Therefore, changes in expression levels of this protein do not drive aggregation, but instead, it is likely that changes in a cotranslational process result in nascent protein misfolding. Codon translation rates are thus seen to have a significant impact on the likelihood of nascent protein aggregation.

TRANSLOCATION OF PROTEINS ACROSS CELL MEMBRANES IS MODULATED BY CODON TRANSLATION RATES

The successful cotranslational translocation of secretory proteins through the Sec-translocon has also been linked to codon translation rates. The signal recognition particle (SRP) is an abundant and universally conserved ribonucleoprotein which recognizes signal sequences located at the N-terminus of secretory nascent polypeptides and targets them to the endoplasmic reticulum (ER) for translocation.[5] The likelihood that a nascent protein is successfully translocated depends on the affinity of SRP for the signal sequence, which can be altered by modifying the amino acid sequence. In a recent study, it was shown that globally decreasing the codon translation rate in HDB52 cells via the addition of cycloheximide led to an increase in the percentage of nascent protein successfully translocated into the ER for proteins containing signal sequences with low binding affinity for SRP.[36] It was hypothesized that slower translation rates provide more time for SRP to recognize and bind to the signal sequence, which is the first step on the pathway to translocation. Additionally, an earlier study[37] found that the topology of membrane proteins can also be altered by globally slowing translation rates. Though these studies used an external means (cycloheximide) to manipulate the translation rate, it is likely that altering the codon usage in the transcript could provide a similar decrease in translation rate and aid in the cotranslational translocation of secretory proteins.

Co-translational protein translocation can also be influenced by changes in other cotranslational processes that are sensitive to codon translation rates. A zinc-finger domain, the folding of which can be induced by the presence of Zn^{2+}, was used to show that the ability of protein-conducting channels to cotranslationally translocate proteins across the ER membrane is inhibited by the cotranslational folding of the passenger sequence.[38] In constructs that were designed to allow the zinc finger to cotranslationally fold, the passenger protein that was covalently attached to zinc finger was observed to be diverted to the cytosol instead of entering the ER lumen. On the other hand, the inhibition of zinc-finger folding allowed translocation into the ER to occur. This suggests that altering codon translation rates could alter the likelihood of cotranslational folding and, thereby, affect the translocation efficiency of secretory proteins. More generally, this observation illustrates how codon translation rates can modulate not only individual cotranslational processes but also multiple cotranslational processes that occur in concert. Such multifactorial effects of codon translation rates illustrate the pressing need for a theoretical framework which will facilitate a quantitative understanding of the consequences of codon translation rates for nascent proteome behavior.

OTHER COTRANSLATIONAL PROCESSES CAN ALSO DEPEND ON CODON TRANSLATION RATES

The processes of nascent protein glycosylation,[8] chaperoning interactions,[3] enzymatic modification,[39] and ubiquitination[6,7,40] can also occur cotranslationally. While the influence of codon translation rates on these cellular processes has not yet been examined, any process that occurs on the time scale of translation could also have its outcomes influenced by the kinetics of translation elongation.

The covalent attachment of carbohydrates to proteins is known as glycosylation, and most eukaryotic proteins that are secreted to the exterior of the cell are glycosylated.[8] This modification to the nascent chain often occurs cotranslationally, with the ribosome inserting the nascent protein from the cytoplasmic side of the ER into the SEC translocon and glycosylation occurring as nascent chain segments emerge into the ER lumen.[4,8] Cotranslational glycosylation of the nascent protein may be much more efficient as compared to post-translational glycosylation because nascent protein segments are more likely to be unstructured during the period of their synthesis on the ribosome, thereby offering greater exposure of potential glycosylation sites.[8] This suggests that slowing down translation could increase the probability of nascent chain

segments folding as they emerge from the translocon, thereby decreasing the chances for glycosylation.

A number of molecular chaperones and processing enzymes have been observed to interact with nascent proteins during translation.[39,41] In some cases, molecular chaperones, such as trigger factor in *E. coli*, assist nascent protein folding by sequestering aggregation-prone polypeptide sequences that can become exposed as a result of protein misfolding. Although some chaperones act post-translationally, those which associate with proteins during translation must do so on the time scale of protein synthesis and thus may be kinetically dependent on codon translation rates. Processing enzymes in *E. coli*, such as PDF and MAP,[42] act on nascent chains during translation, suggesting that their binding and enzyme kinetics could also be affected by codon elongation rates. Translation elongation rates have also been linked to observed differences in the correlated processes of arginylation and ubiquitination of γ- and β-actin nascent proteins *in vivo*.[43]

Ubiquitination is carried out by ubiquitin ligases, enzymes that can covalently attach an ubiquitin protein molecule to another protein, which in some cases targets that protein for degradation.[6] Estimates of the relative amounts of nascent proteins that are cotranslationally ubiquitinated range from 1% to 30%.[6,7,39] Just as glycosylation, chaperoning, and enzymatic processing may be dependent on codon translation rates, it is also possible that cotranslational ubiquitination may be affected by the kinetics of translation elongation. Since misfolded proteins are more likely to be ubiquitinated, cotranslational ubiquitination and folding are another set of processes where changes in translation rates could affect more than one cotranslational process.

SEVERAL HUMAN DISEASES HAVE BEEN LINKED TO CODON TRANSLATION RATES

A number of human diseases that affect the lungs and blood, as well as various cancers, have been linked to the variability of codon translation rates.[9] The recent sequencing of the lymphocyte mRNAs of five Swedish families with hemophilia B led to the finding that a synonymous single nucleotide polymorphism (SNP) in the F9 gene was the only difference between their genotype and healthy individuals, with mRNA splice variation ruled out as a cause.[44] This leaves open the possibility that this SNP alters nascent protein behavior through a change in the translation elongation rate. Similarly, a synonymous SNP in the cystic fibrosis transmembrane conductance regulator gene, one of the most frequent causes of cystic fibrosis, was found to alter the expression level of the mutant protein via an observed change in protein synthesis rates;[34] however, this could also potentially occur due to a change in the translation initiation rate as compared to a change in elongation rate. Various cancers, including lung carcinoma and cervical and vulvar cancer, have also been linked to synonymous SNPs, suggesting a possible role of codon translation rates.[9,45] A molecular perspective connecting changes in cotranslational processes due to alteration in codon translation rates and the progression of these diseases is lacking. Such a perspective could help us understand the origin of these diseases.

RECENT APPROACHES TO MODELING CODON TRANSLATION RATE EFFECTS ON COTRANSLATIONAL PROTEIN FOLDING AND TRANSLOCATION

The numerous examples discussed previously illustrate the importance that codon translation rates can have in governing nascent proteome behavior and indicate an area of biology where the application of theoretical and simulation techniques could significantly advance knowledge and understanding. Chemical kinetics and coarse-grained molecular dynamics simulations are two techniques that have recently been applied to study codon translation rate effects on cotranslational protein folding and translocation through the SEC translocon.

Using a Markov chain probability approach, the chemical reaction schemes representing cotranslational domain folding mechanisms involving either two or three thermodynamic states were analytically solved.[25,46] The resulting equations provide the capability to predict how individual codon translation rates in an ORF will influence the probability of a domain populating its unfolded, intermediate, or folded states at different points during its synthesis. Extending this type of approach to other nascent protein behaviors (Figures 1B and 2A) is particularly promising, as it would allow for the rapid prediction of the time evolution of the different states that a nascent chain populates as it is being synthesized, the probability of various cotranslational processes occurring, and ultimately could provide predictions about the fate of the nascent protein *in vivo*.

An alternative to solving such reaction schemes analytically is to numerically integrate the differential equations that describe the relationships between the cotranslationally populated states and the underlying rates of interconversion between those states (Figure 2B). This approach was recently used to study the post-translational behavior of newly synthesized proteins in *E. coli*, including their degradation, aggregation, and interactions with mainly post-translationally acting chaperones.[47] This numerical approach can also, in principle, be applied to cotranslational behavior. There are benefits and costs to such a numerical approach as compared to the analytical approach previously described. Numerical approaches can be readily adapted to any complex reaction scheme, especially with their implementation in widely used and robust software packages such as Mathematica and Matlab. On the other hand, analytical solutions sometimes must be solved on a case-by-case basis for different cotranslational situations.[25] Numerical integration, in principle, can take longer to calculate, as a larger number of integration steps must be executed to obtain converged results, whereas the analytic approach often involves fewer steps. A significant advantage that analytical approaches have is that they can rapidly identify the dynamic regimes within such models through the use of derivative tests,[25] while an exhaustive numerical search of the kinetic regimes of cotranslational behavior can be prohibitive due to the large kinetic-parameter space of the representative reaction networks. Therefore, the benefits of analytical approaches are quite extensive, although this can be mitigated by the one-time cost of obtaining the analytic solution.

Coarse-grained simulations (Figure 2C) of the translation process offer a more detailed molecular view of the consequences codon translation rates have on cotranslational folding than mathematical modeling.[48−50] Like mathematical models, coarse-grained models can be used to calculate the probabilities of nascent chains being in different states at different translation elongation rates,[46] but they also offer additional information,

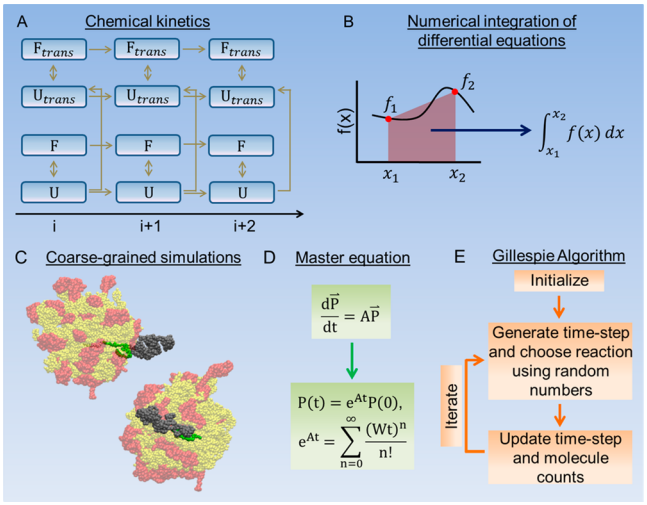

Figure 2. A range of methods for describing the influence of codon translation rates on nascent protein behavior. (A) A chemical kinetic reaction scheme that describes cotranslational translocation. At each nascent chain length, the states that a nascent chain segment may populate include unfolded (U), folded (F), unfolded and translocated (U_{trans}), or folded and translocated (F_{trans}) states. Note that translocation cannot occur if the nascent protein segment is folded and that translocation is irreversible. (B) Numerical integration relies on mathematical estimation to solve systems of differential equations that represent the time evolution of chemical kinetic reaction schemes. (C) Coarse-grained simulations of ribosome nascent chain complexes allow for a detailed molecular perspective on cotranslational folding. Snapshots from simulations in which ribosomes (red and yellow) are engaged in the translation of nascent proteins (green). Trigger factor (black) is shown associated with the ribosome-nascent chain complex.[50] (D) The master equation approach can solve a set of differential equations to reveal the time-dependent evolution of the system on the reaction scheme. (E) The Gillespie Algorithm is used to solve a stochastic version of chemical kinetics for single molecules and consists of three key steps. The initialization step requires defining the states of the system and the rates of interconversion between these states. Random numbers are then used to generate the time step and starting reaction, and the reaction is then modeled. Next, the time step and number of molecules in each state are updated. The final two steps are then repeated until a predetermined stop condition is met.

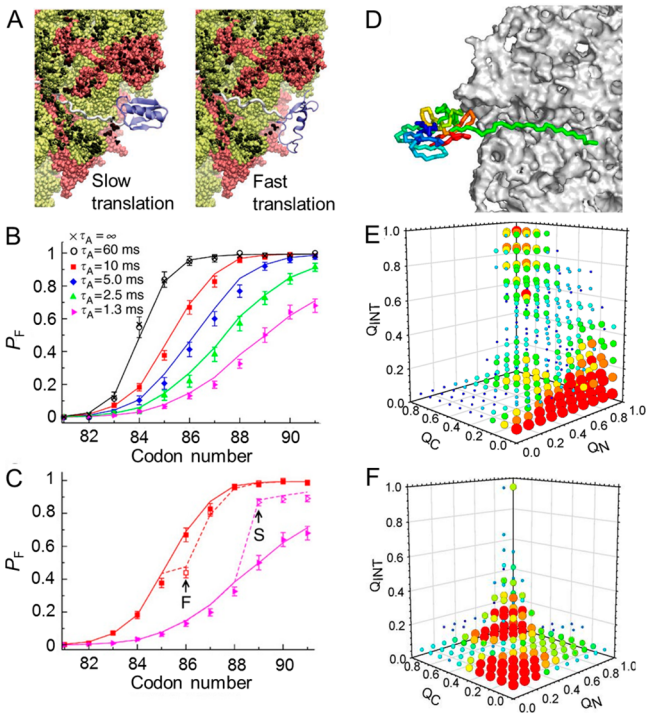

Figure 3. Codon translation rate effects on nascent protein folding behavior predicted by chemical kinetic or coarse-grained simulation methods. (A) Cotranslational folding on the ribosome is favored by slow translation conditions (left) which allow the nascent chain additional time to adopt the native conformation.[46] Fast translation conditions (right) can result in unfolded or misfolded confirmations. (B) The effect of globally altering codon translation rates on the probability that a protein will cotranslationally fold as determined by both coarse-grained simulations (symbols) and chemical kinetics (solid lines) for a wide range of codon translation rates. The figure legend lists the codon translation times, τ_A, which are equivalent to the inverses of the rates. (C) The probability that a domain cotranslationally folds as a function of nascent chain length determined by both chemical kinetic (solid lines) and coarse-grained simulations (symbols) as affected by the introduction of a single fast- (orange) or slow- (pink) translating codon, indicated by "F" or "S", respectively. (D) A snapshot of the time step at which a coarse-grained model of the cotranslationally folding Semliki Forest Virus Protein (SFVP) adopts its native fold.[51] (E) The pathways of cotranslational folding of SFVP on the ribosome. Red points symbolize highly probable states along SFVP's co-translational folding pathways, where Q_N, Q_C, and Q_{int} are the fraction of native contacts in the protein's N-terminal, C-terminal, and interfacial regions, respectively. (F) Same as E but for refolding trajectories started with the SFVP molecule free in solution. Panels A, B, and C were produced from ref 46 with permission from Nature Publishing Group. Panels D, E, and F were produced from ref 51 with permission from PLOS.

including structural information, the energetics associated with various cotranslational processes,[49] and information on dynamics that might be neglected in the chemical kinetic modeling. The trade-off for this more detailed picture is a dramatically increased cost of computation. An ensemble of coarse-grained simulations of cotranslational folding took ~800 CPU days (= 4 days per trajectory × 200 trajectories), while the kinetic model predicted this result in a computation requiring just a few seconds.[25] Thus, coarse-grained simulations and mathematical modeling can be used to address a wide range of questions, but where there is overlap in the question being addressed it is often more efficient to use kinetic modeling.

Chemical kinetic and molecular dynamic computations have recently been used to address a number of questions concerning the fates of nascent proteins, especially with regard to their cotranslational folding and translocation. A chemical kinetic model describing a two-state cotranslational folding mechanism (Figure 3A) allowed for the probability of cotranslational folding to be calculated as a function of nascent chain length.[46] For Protein G, it was found that both globally changing the codon translation rate (Figure 3B) and the insertion of single slow- or fast-translating codons (Figure 3C) dramatically alter predicted folding behavior. Furthermore, coarse-grain models (Figure 3D) were used to probe the intrinsic differences between *in vivo* and *in vitro* protein folding, with results suggesting that, for multidomain proteins, ribosome-mediated folding *in vivo* (Figure 3E) can follow significantly different pathways from that of *in vitro* (Figure 3F) refolding.[51]

Furthermore, the molecular regulation of cotranslational protein translocation through the Sec translocon was probed using a coarse-grained model, revealing that multiple kinetic pathways exist for the integration of proteins into lipid membranes (Figure 4A).[52] This study also demonstrated that in such models increasing the global codon translation rate by a

dx.doi.org/10.1021/ja510082j | J. Am. Chem. Soc. 2014, 136, 17892−17898

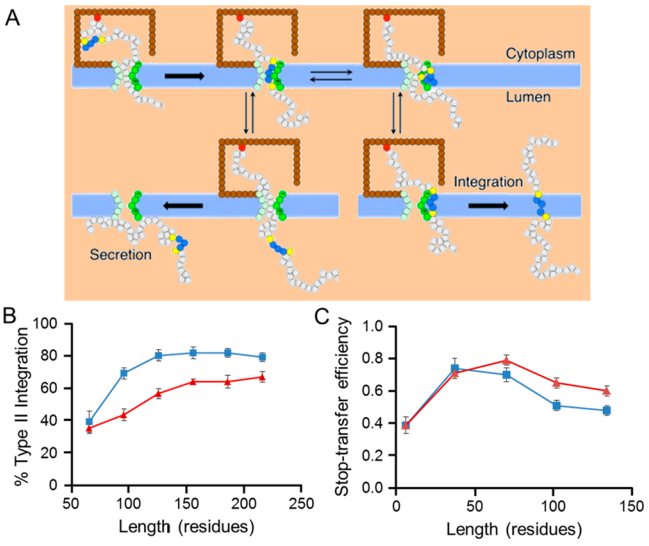

Figure 4. Influence of codon translation rates on the translocation or insertion of proteins into membranes as described by coarse-grained models. (A) Schematic of a kinetically controlled translocation/membrane-insertion coarse-grained model used to study the influence of codon translation rates.[52] (B) The percent of nascent protein which is inserted by a Type II mechanism in coarse-grained simulations at global translation rates of 6 AA/s (blue) and 24 AA/s (red) as a function of nascent chain length. (C) Same as B except stop-transfer efficiency is plotted as a function of nascent chain length. The figures and data displayed in (A), (B), and (C) are reproduced from ref 52 with permission from Cell Press.

factor of 4 resulted in a large reduction in the percent of nascent protein successfully inserted into the membrane by a Type II mechanism (Figure 4B) and an increase in the amount of protein directed into the membrane (referred to as the stop-transfer efficiency; Figure 4C). Related all-atom models of cotranslational translocation also suggest that the kinetics of membrane insertion versus codon translation play a critical role in the cell's regulation of nascent protein translocation.[53,54] Such chemical kinetic and coarse-grain modeling techniques will continue to be important tools for the expansion of our understanding of the impact of codon translation rates on nascent protein behavior.

UNIFYING CODON TRANSLATION RATES AND NASCENT PROTEOME BEHAVIOR

As we have shown, the kinetics of translation and various cotranslational processes can be the key factors governing nascent protein behavior. As such, a requirement of any theoretical framework that is capable of predicting the range of nascent protein behaviors must be that it utilizes knowledge of these various rates to make useful predictions. The foundations for such a framework could lie in other fields, where similar challenges have previously been solved. In the field of *in vitro* protein folding, for example, significant advances have recently been made in predicting the time course of protein folding by utilizing master equations in combination with short time-scale molecular dynamics simulations.[55−59] In this approach, Markov states are first identified from simulation trajectories,[56] and the rates of interconversion between those states are calculated and then inserted into master equations to predict the state probabilities as a function of time (Figure 2D). Master equations consist of differential equations that can be solved in a prescribed manner to describe the time evolution of the probability of populating different states within a system. As the idea of Markov states is generalizable to the different states populated during cotranslational processes, the master equation approach could potentially be applied to the full-range of nascent protein behaviors and is an attractive option because it would provide a formalism for calculating the effect of codon translation rates on cotranslational processes and nascent protein behavior.

While master equations can accurately describe the behavior of a collection of a large number of molecules, the behavior of a single nascent protein molecule can be predicted by the Gillespie Algorithm (Figure 2E). The Gillespie Algorithm uses the underlying rates of interconversion between a system's states and random numbers to numerically simulate the time evolution of the states sampled by a single molecule.[60] This algorithm is based on a stochastic formulation of chemical kinetics. The master equation and Gillespie Algorithm approaches yield equivalent results in the limit of the simulation of a large number of single molecule trajectories. The master equation and Gillespie Algorithm approaches are, therefore, complementary, allowing for detailed modeling of the time-dependent evolution of nascent proteins at both the ensemble and single molecule levels.

The success of such a theoretical framework will ultimately be measured by its explanatory and predictive powers regarding open questions about codon translation rates. In our view, open questions that might be successfully addressed in the next several years include the following: (*i*) What fraction of an organism's nascent proteome is susceptible to a change in behavior due to a change in codon translation rates? (*ii*) What are the critical codon positions in an ORF that are likely to have a large effect on a protein's nascent behavior? (*iii*) Can we accurately predict, for a given protein and translation-rate profile, how changing the translation rate at one or more codon positions will alter the nascent protein's behavior? (*iv*) What are the absolute codon translation rates at each position within an ORF? (*v*) What are the quantitative, molecular origins of codon translation rates? (*vi*) How does the coupling between codon translation rates and nascent protein behavior help shape the evolution of synonymous codon usage in organisms?

CONCLUDING REMARKS

This perspective has highlighted the far reaching effects that codon translation rates have on modulating the broad spectrum of nascent protein behavior and the tools that can be used to understand and model these phenomena. The principles of chemical kinetics suggest that it can be reasonably assumed that any cotranslational process can be influenced by codon translation rates to one degree or another. Despite the recent progress toward an understanding of their importance in experimental systems, there is still a striking lack of theoretical tools available to provide the framework for modeling these codon translation rate effects.

Chemical kinetic approaches, due to their ability to predict the time evolution of nonequilibrium systems, are one of the most promising places to look for solutions, and the progress made toward a general model of cotranslational protein folding within this vein is an indicator of its potential to explain other *in vivo* nascent protein behaviors. We expect that codon translation rates will be a growing area of research interest and activity as a result of their potentially transformative implications for molecular and cellular biology, evolution, biotechnology, and biomedicine. Bringing the theoretical tools from chemistry and physics to these important phenomena is likely to provide the framework

dx.doi.org/10.1021/ja510082j | J. Am. Chem. Soc. 2014, 136, 17892−17898

for unifying codon translation rates and nascent proteome behavior.

AUTHOR INFORMATION

Corresponding Author

epo2@psu.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Professors Will Noid and Phil Bevilacqua as well as all O'Brien group members for valuable comments on the manuscript.

REFERENCES

(1) Komar, A. A. *Trends Biochem. Sci.* **2008**, *34*, 16.
(2) Nicola, A. V.; Chen, W.; Helenius, A. *Nat. Cell Biol.* **1999**, *1*, 341.
(3) Gloge, F.; Becker, A. H.; Kramer, G.; Bukau, B. *Curr. Opin. Struct. Biol.* **2014**, *24*, 24.
(4) Pechmann, S.; Willmund, F.; Frydman, J. *Mol. Cell* **2013**, *49*, 411.
(5) Walter, P. *Annu. Rev. Cell Biol.* **1994**, *10*, 87.
(6) Comyn, S. A.; Chan, G. T.; Mayor, T. *J. Proteomics* **2013**, *100*, 92.
(7) Duttler, S.; Pechmann, S.; Frydman, J. *Mol. Cell* **2013**, *50*, 379.
(8) Ruiz-Canada, C.; Kelleher, D. J.; Gilmore, R. *Cell* **2009**, *136*, 272.
(9) Sauna, Z. E.; Kimchi-Sarfaty, C. *Nat. Rev. Genet.* **2011**, *12*, 683.
(10) Lehmann, J.; Libchaber, A. *RNA* **2008**, *14*, 1264.
(11) Sørensen, M. A.; Pedersen, S. *J. Mol. Biol.* **1991**, *222*, 265.
(12) Fluitt, A.; Pienaar, E.; Viljoen, H. *Comput. Biol. Chem.* **2007**, *31*, 335.
(13) Spencer, P. S.; Siller, E.; Anderson, J. F.; Barral, J. M. *J. Mol. Biol.* **2012**, *422*, 328.
(14) Makhoul, C. H.; Trifonov, E. N. *J. Biomol. Struct. Dyn.* **2002**, *20*, 413.
(15) Clarke, T. F.; Clark, P. L. *BMC Genomics* **2010**, *11*, 118.
(16) Clarke, T. F.; Clark, P. L. *PLoS One* **2008**, *3*, e3412.
(17) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Hartl, F. U. *Nat. Struct. Biol.* **1999**, *6*, 697.
(18) Chang, H.-C.; Kaiser, C. M.; Hartl, F. U.; Barral, J. M. *J. Mol. Biol.* **2005**, *353*, 397.
(19) Evans, M. S.; Sander, I. M.; Clark, P. L. *J. Mol. Biol.* **2008**, *383*, 683.
(20) Sánchez, I. E.; Morillas, M.; Zobeley, E.; Kiefhaber, T.; Glockshuber, R. *J. Mol. Biol.* **2004**, *338*, 159.
(21) Siller, E.; DeZwaan, D. C.; Anderson, J. F.; Freeman, B. C.; Barral, J. M. *J. Mol. Biol.* **2010**, *396*, 1310.
(22) Zhang, G.; Hubalewska, M.; Ignatova, Z. *Nat. Struct. Mol. Biol.* **2009**, *16*, 274.
(23) Sander, I. M.; Chaney, J. L.; Clark, P. L. *J. Am. Chem. Soc.* **2014**, *136*, 858.
(24) Saunders, R.; Deane, C. M. *Nucleic Acids Res.* **2010**, *38*, 6719.
(25) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. *Nat. Commun.* **2014**, *5*, 2988.
(26) Cortazzo, P.; Cerveñansky, C.; Marín, M.; Reiss, C.; Ehrlich, R.; Deana, A. *Biochem. Biophys. Res. Commun.* **2002**, *293*, 537.
(27) Komar, A. A.; Lesnik, T.; Reiss, C. *FEBS Lett.* **1999**, *462*, 387.
(28) Sherman, M. Y.; Qian, S.-B. *Trends Biochem. Sci.* **2013**, *38*, 585.
(29) Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J. M.; Sachs, M. S.; Liu, Y. *Nature* **2013**, *495*, 111.
(30) Fedyunin, I.; Lehnhardt, L.; Böhmer, N.; Kaufmann, P.; Zhang, G.; Ignatova, Z. *FEBS Lett.* **2012**, *586*, 3336.
(31) Kimchi-Sarfaty, C.; Oh, J. M.; Kim, I.; Sauna, Z. E.; Maria, A.; Ambudkar, S. V.; Gottesman, M. M. *Science* **2007**, *315*, 525.
(32) Angov, E.; Hillier, C. J.; Kincaid, R. L.; Lyon, J. A. *PLoS One* **2008**, *3*, e2189.
(33) Maertens, B.; Spriestersbach, A.; von Groll, U.; Roth, U.; Kubicek, J.; Gerrits, M.; Graf, M.; Liss, M.; Daubert, D.; Wagner, R.; Schäfer, F. *Protein Sci.* **2010**, *19*, 1312.
(34) Bartoszewski, R. A.; Jablonsky, M.; Bartoszewska, S.; Stevenson, L.; Dai, Q.; Kappes, J.; Collawn, J. F.; Bebok, Z. *J. Biol. Chem.* **2010**, *285*, 28741.
(35) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333.
(36) Zhang, D.; Shan, S. *J. Biol. Chem.* **2012**, *287*, 7652.
(37) Goder, V.; Spiess, M. 2003, *22*, 3645.
(38) Conti, B. J.; Elferich, J.; Yang, Z.; Shinde, U.; Skach, W. R. *Nat. Struct. Mol. Biol.* **2013**, *21*, 228.
(39) Arfin, S. M.; Bradshaw, R. A. *Biochemistry* **1988**, *27*, 7979.
(40) Turner, G. C.; Varshavsky, A. *Science* **2000**, *289*, 2117.
(41) Kramer, G.; Boehringer, D.; Ban, N.; Bukau, B. *Nat. Struct. Mol. Biol.* **2009**, *16*, 589.
(42) Meinnel, T.; Mechulam, Y.; Blanquet, S. *Biochimie* **1993**, *75*, 1061.
(43) Zhang, F.; Saha, S.; Shabalina, S. A.; Kashina, A. *Science* **2010**, *329*, 1534.
(44) Knobe, K. E.; Sjörin, E.; Ljung, R. C. R. *Haemophilia* **2008**, *14*, 723.
(45) Supek, F.; Miñana, B.; Valcárcel, J.; Gabaldón, T.; Lehner, B. *Cell* **2014**, *156*, 1324.
(46) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. *Nat. Commun.* **2012**, *3*, 868.
(47) Powers, E. T.; Powers, D. L.; Gierasch, L. M. *Cell Rep.* **2012**, *1*, 265.
(48) O'Brien, E. P.; Hsu, S.-T. D.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 16928.
(49) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2011**, 893.
(50) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2012**, *134*, 10920.
(51) Elcock, A. H. *PLoS Comput. Biol.* **2006**, *2*, e98.
(52) Zhang, B.; Miller, T. F. *Cell Rep.* **2012**, *2*, 927.
(53) Gumbart, J. C.; Teo, I.; Roux, B.; Schulten, K. *J. Am. Chem. Soc.* **2013**, *135*, 2291.
(54) Zhang, B.; Miller, T. F. *J. Am. Chem. Soc.* **2012**, *134*, 13700.
(55) Best, R. B.; Mittal, J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11087.
(56) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. *J. Chem. Phys.* **2007**, *126*, 155101.
(57) Buchete, N.-V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057.
(58) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.
(59) Pande, V. S.; Beauchamp, K.; Bowman, G. R. *Methods* **2010**, *52*, 99.
(60) Gillespie, D. T. *J. Phys. Chem.* **1977**, *81*, 2340.