ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

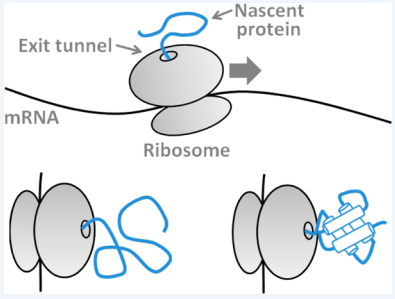
Understanding the Influence of Codon Translation Rates on Cotranslational Protein Folding

Edward P. O'Brien,*[,†,§] Prajwal Ciryam,[†,‡] Michele Vendruscolo,[†] and Christopher M. Dobson*[,†]

[†]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

[‡]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, United States

CONSPECTUS: Protein domains can fold into stable tertiary structures while they are synthesized by the ribosome in a process known as cotranslational folding. If a protein does not fold cotranslationally, however, it has the opportunity to do so post-translationally, that is, after the nascent chain has been fully synthesized and released from the ribosome. The rate at which a ribosome adds an amino acid encoded by a particular codon to the elongating nascent chain can vary significantly and is called the codon translation rate. Recent experiments have illustrated the profound impact that codon translation rates can have on the cotranslational folding process and the acquisition of function by nascent proteins. Synonymous codon mutations in an mRNA molecule change the chemical identity of a codon and its translation rate without changing the sequence of the synthesized protein. This change in codon translation rate can, however, cause a nascent protein to malfunction as a result of cotranslational misfolding. In some situations, such dysfunction can have profound implications; for example, it can alter the substrate specificity of an ABC transporter protein, resulting in patients who are nonresponsive to chemotherapy treatment. Thus, codon translation rates are crucial in coordinating protein folding in a cellular environment and can affect downstream cellular processes that depend on the proper functioning of newly synthesized proteins. As the importance of codon translation rates makes clear, a necessary aspect of fully understanding cotranslational folding lies in considering the kinetics of the process in addition to its thermodynamics.



In this Account, we examine the contributions that have been made to elucidating the mechanisms of cotranslational folding by using the theoretical and computational tools of chemical kinetics, molecular simulations, and systems biology. These efforts have extended our ability to understand, model, and predict the influence of codon translation rates on cotranslational protein folding and misfolding. The application of such approaches to this important problem is creating a framework for making quantitative predictions of the impact of synonymous codon substitutions on cotranslational folding that has led to a novel hypothesis regarding the role of fast-translating codons in coordinating cotranslational folding. In addition, it is providing new insights into proteome-wide cotranslational folding behavior and making it possible to identify potential molecular mechanisms by which molecular chaperones can influence such behavior during protein synthesis. As we discuss in this Account, bringing together these theoretical developments with experimental approaches is increasingly helping answer fundamental questions about the nature of nascent protein folding on the ribosome.

■ INTRODUCTION

The question of how best to understand, model, and predict the influence of codon translation rates on cotranslational folding[1−10] (Figure 1a) has recently come to prominence as a result of new experimental strategies[3,4,11] and developments in theory and modeling.[12−18] We discuss here how physical chemistry, computer simulations, and mathematical modeling can provide useful concepts to describe this phenomenon and how knowledge from these fields can be utilized to model and predict the effect of codon translation rates, which can vary as a result of synonymous codon usage (Figure 1b), on cotranslational folding. For these reasons, it is crucial for the advancement of the "nascent proteome" field (which concerns the proteome-wide behavior of newly synthesized proteins) to bring together the advances being made experimentally with the tools and insights being developed through theoretical and computational investigations.

■ THE IMPORTANCE OF TIME TO COTRANSLATIONAL PROTEIN FOLDING

A living cell is not in equilibrium with its environment. The foremost implication of this fact is that the behavior of a cell, or a component of it, is determined by the rates of the processes that occur within it rather than by its thermodynamic properties. As it has been put recently, "kinetics can trump thermodynamics" in a system that is out of equilibrium.[19] Since the elementary reaction rate of a single-step reaction is the inverse of the average time it takes for this reaction to occur, an equivalent statement is that the properties of a system are determined by the time scales of the processes occurring within that system. Time is therefore a central factor in understanding out-of-equilibrium processes in living cells. For this reason, to understand the impact of changing codon translation rates, it is

Received: January 21, 2014
Published: May 1, 2014

   dx.doi.org/10.1021/ar5000117 | Acc. Chem. Res. 2014, 47, 1536−1544

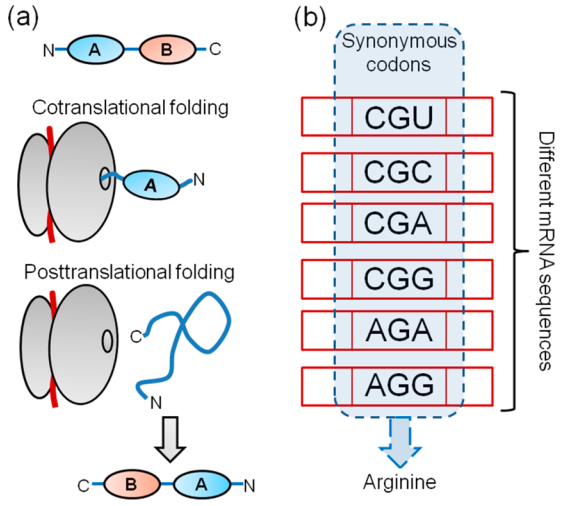


Figure 1. (a) A protein with two domains ("A" and "B") can exhibit either folding of each domain as synthesis by the ribosome progresses (termed cotranslational folding) or folding of both domains after synthesis is completed (termed post-translational folding). (b) Of the 20 naturally occurring amino acids, 18 are encoded by more than one type of codon; for example, six different synonymous codons, which are translated at different rates, encode for arginine. This codon code leads to a combinatorial explosion in which an astronomically large number of different mRNA sequences can encode for the same protein sequence.

crucial to identify first the essential features of the processes associated with translation and cotranslational folding.

The process of translation involves the unidirectional translocation of a ribosome along an mRNA molecule one codon at a time[20] (Figure 2). The ribosome decodes the

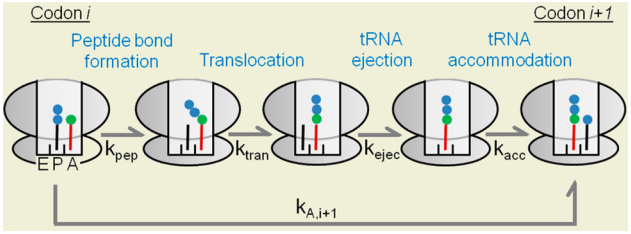


Figure 2. Translation of codon i requires a series of sequential processes, including peptide bond formation, translocation of the mRNA molecule by one codon, ejection of the deacylated tRNA from the E site, and accommodation of a charged tRNA to the A site. The overall rate of this process can be characterized by $k_{A,i+1}$, with an average rate that typically varies between 3 and 22 AA/s in prokaryotic and eukaryotic cells.[52]

genetic information contained in each codon and covalently attaches the amino acid corresponding to that codon to the growing nascent chain, which moves through a channel that is known as the exit tunnel that is located within the large ribosomal subunit.[21] The ribosome exit tunnel is about 10 nm long, and examination of its structure indicates that the first 8 nm are too narrow to allow the formation of tertiary structure by the nascent chain. As the ribosome only adds residues to the nascent chain but does not remove them, the process of translation within a cell is irreversible and therefore out of equilibrium.[22] The nascent chain, prior to its release from the ribosome, is physically attached to a ribosome-bound tRNA molecule (Figure 2); cotranslational folding is therefore also an out-of-equilibrium process in a cell. Thus, the underlying reaction rates of codon translation and of domain folding and unfolding have the potential to strongly influence the cotranslational folding behavior of a nascent chain.

We can more clearly define the specific rates that are most relevant to cotranslational folding by considering the details of the process. An elongation cycle of the ribosome, which starts with the P site of the ribosome centered on codon i and ends with it centered on codon $i + 1$, involves many intermediate steps, such as tRNA accommodation, peptide bond formation, mRNA translocation, and ejection of the deacylated tRNA from the E site[20] (Figure 2). We can subsume these intermediate steps, each with its own elementary reaction rate, into a composite rate $k_{A,i+1}$, which we call the rate of amino acid addition of the $i + 1$ residue to the growing nascent chain. This process of amino acid addition extends the nascent chain from i residues to $i + 1$ residues in length (Figure 2). The rate $k_{A,i+1}$ is an important quantity because it defines the average speed at which codon i is translated.

The other relevant rates are those associated with the process of domain folding at codon i. For example, a domain that folds in a two-state manner and is not influenced by neighboring domains has two distinct rates, $k_{UF,i}$ and $k_{FU,i}$, that characterize these intradomain structural transitions at each nascent chain length.[14] $k_{UF,i}$ and $k_{FU,i}$ are, respectively, the folding and unfolding rates of a domain embedded in a nascent chain that is i residues long at a given point during its synthesis. These rates depend on the nascent chain length because the chemical environment experienced by the domain changes as the nascent chain is elongated during synthesis.[11,14,23] More complex domain folding mechanisms involving intermediates involve an even larger number of rates that can vary at each nascent chain length.[17]

This reasoning, based purely on physicochemical principles, suggests therefore that the rates $k_{A,i+1}$, $k_{FU,i}$, and $k_{UF,i}$ have the potential to determine the cotranslational folding behavior of a protein in a cell because of the out-of-equilibrium nature of the translation process.

EXPERIMENTAL EVIDENCE FOR THE INFLUENCE OF CODON TRANSLATION RATES ON COTRANSLATIONAL FOLDING

There is a substantial amount of evidence that cotranslational folding occurs both in vitro and in vivo.[1,24,25] There have also been a smaller but still significant number of experiments probing the consequences for cotranslational folding of changing codon translation rates. Those experiments demonstrate that codon translation rates (i.e., the $k_{A,i+1}$ values) can determine the probability of cotranslational folding[3] and misfolding[4,5,26] as well as whether or not a protein ends up functioning properly.[7,8,27,28]

The biological importance of codon translation rates is also manifested across the genomes of organisms in terms of their codon usage in mRNA molecules. Natural selection has resulted in the presence of particular synonymous codons in particular positions in genes, and in some cases it appears very likely that the result is to coordinate cotranslational folding.[29] Experimental support for this concept includes the observation that one of the most successful strategies for designing synonymous mRNA sequences for heterologous protein expression reproduces the codon translation rate profile from the source organism in the recipient organism.[4] It is hypothesized that this "codon harmonization" procedure

dx.doi.org/10.1021/ar5000117 | Acc. Chem. Res. 2014, 47, 1536−1544

Table 1. Thermodynamic Classification of Cotranslational Folding Behavior

classification	definition	relevance to cotranslational folding
equilibrium	The system under study contains only reversible processes and exhibits properties that do not change over long time periods.	Equilibrium cotranslational folding behavior is exhibited by arrested ribosome molecules that are arrested for times longer than those for all of the other processes.
quasi-equilibrium	The system consists of one or more irreversible processes that prevent the establishment of equilibrium, yet the system still exhibits equilibrium-like properties.	Quasi-equilibrium cotranslational folding behavior is exhibited when the folding probability at each nascent chain length during continuous translation is comparable to the folding probability on an arrested ribosome.
nonequilibrium	The system consists of one or more irreversible processes and does not exhibit equilibrium properties. If the properties do not change with time, the system is said to be in a steady (or stationary) state.	Nonequilibrium cotranslational folding behavior is exhibited when the folding probabilities at some nascent chain lengths during continuous translation differ from the folding probabilities on an arrested ribosome.
out of equilibrium	The system consists of one or more irreversible processes, but whether it exhibits equilibrium-like behavior is not specified.	Knowledge that a protein is being continuously translated in a cell is sufficient information to know that cotranslational folding is out of equilibrium. However, this information is insufficient to determine whether the protein will exhibit quasi-equilibrium or nonequilibrium cotranslational folding behavior.

maintains the evolutionarily optimized timing of the synthesis of different segments of a protein, thereby maximizing cotranslational folding.[30]

DISTINCTIONS AMONG EQUILIBRIUM, QUASI-EQUILIBRIUM, AND NONEQUILIBRIUM REGIMES OF COTRANSLATIONAL FOLDING

The three rates $k_{A,i+1}$, $k_{FU,i}$, and $k_{UF,i}$ give rise to three broad regimes of cotranslational folding: an equilibrium regime, a quasi-equilibrium regime and a nonequilibrium regime.[15] As discussed below, an out-of-equilibrium process can be either quasi-equilibrium or nonequilibrium in nature (Table 1).

Equilibrium cotranslational folding can be studied in a test tube by stalling indefinitely a ribosome at codon i (i.e., setting $k_{A,i+1} = 0$) and measuring the properties of the nascent chain at that length.[11] Because in this case the nascent chain is not growing, there is no irreversibility in the system, and equilibrium can be achieved. By studying nascent chains arrested at different chain lengths, the equilibrium cotranslational folding curve can be measured. Such equilibrium behavior, where thermodynamics governs the properties of folding at each length, can be used as a reference process against which to compare in vivo cotranslational folding.

Carrying out such comparisons has revealed that two different out-of-equilibrium classifications are possible (Figure 3a).[15] The first is a quasi-equilibrium regime, in which cotranslational folding exhibits equilibrium-like behavior despite being an out-of-equilibrium process because of the irreversible nature of in vivo translation. The other is a nonequilibrium regime, in which cotranslational folding exhibits properties markedly different from those found at equilibrium. A protein can switch between these quasi-equilibrium and nonequilibrium regimes by alteration of the ratio of the folding and codon translation rates, as discussed below. For example, the use of coarse-grained molecular dynamics simulations, in which residues are attached to the growing nascent chain every 60 ms, results in the quasi-equilibrium cotranslational folding of protein G, whereas higher elongation rates result in nonequilibrium cotranslational folding behavior (Figure 3a).

The existence of these three regimes is significant because it can tell us in specific cases whether or not the translation rate is more important to cotranslational folding than the thermodynamic properties of the ribosome−nascent-chain complex (RNC). Furthermore, the nonequilibrium regime implies that the cotranslational folding process itself can depend in a sensitive manner on the initial conditions and the history of the RNC, leading to altered folding pathways, structures, and populations from one ribosome molecule to the next in living

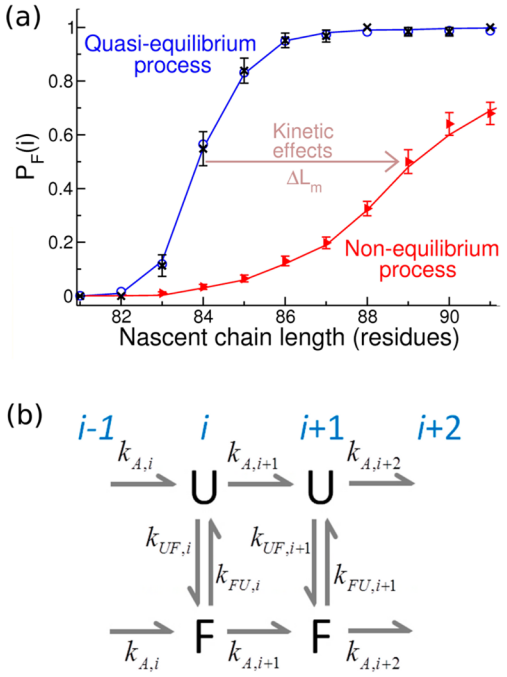


Figure 3. (a) Probability of folding of the protein G domain as a function of nascent chain length immediately before addition of the next amino acid at different codon translation times.[14] The cases shown are for 60 ms/AA (blue circles), 1.3 ms/AA (red triangles), and indefinitely arrested translation (black ×'s). The data are from coarse-grained molecular dynamics simulations. Despite being out of equilibrium, the process with the global translation time of 60 ms/AA reproduces the equilibrium behavior for this protein. Not all proteins exhibit quasi-equilibrium behavior at 60 ms/AA.[15] (b) Chemical reaction scheme representing the cotranslational folding of a domain that folds in an apparent two-state manner. The rates indicated are those that are defined in the text. Image reproduced from ref 15. Copyright 2012 National Academy of Sciences of the United States of America.

cells. Such "memory effects" do not occur in the equilibrium and quasi-equilibrium regimes because there is sufficient time for the nascent chain to equilibrate at each nascent chain length and hence "forget" the initial conditions under which it was prepared. A practical implication of this phenomenon is that in a single-molecule experiment involving quasi-equilibrium cotranslational folding, each nascent chain exhibits the same time-averaged behavior at each length i from one ribosome molecule to the next, whereas in the nonequilibrium regime each nascent chain can exhibit different time-averaged behavior.

dx.doi.org/10.1021/ar5000117 | Acc. Chem. Res. 2014, 47, 1536−1544

THEORETICAL TOOLS FOR UNDERSTANDING AND PREDICTING THE INFLUENCE OF CODON TRANSLATION RATES AT THE MOLECULAR AND CELLULAR LEVELS

Significant advances have recently been made in theory and in simulation techniques to understand, model, and predict cotranslational folding behavior at levels ranging from individual proteins[14,17] to entire proteomes.[15]

On the theoretical front, analytical solutions have been determined for chemical reaction schemes representing cotranslational folding that are difficult to solve because of their complexity (Figure 3b).[14,17] These advances are significant for several reasons. The equations involved enable predictions to be made without having to resort to numerical simulations of the reaction schemes, which can introduce statistical inaccuracies in the calculated quantities. These results also provide a direct means of utilizing the unfolding and folding rates ($k_{FU,i}$ and $k_{UF,i}$, respectively) measured on arrested ribosomes[11] to predict the behavior of nascent proteins during continuous translation.[14] They also allow predictions to be made concerning cotranslational folding scenarios that might occur when codon translation rates are changed.[17] In combination with models for estimating protein folding and unfolding rates,[31] it is possible to use these analytical solutions to predict the cotranslational folding behavior of most of the proteins in the proteome of an organism.[15] Finally, these equations provide a potential means of designing mRNA sequences using synonymous codons to maximize cotranslational folding.[15]

To exemplify these advances, we consider eq 1, which describes the influence of codon translation rates on the ensemble-averaged cotranslational folding probability of a domain that folds via a two-state mechanism:[14]

$$\langle P_F(i)\rangle = \sum_{j=1}^{i} \frac{k_{UF,j}}{k_{A,j+1}} \prod_{l=j}^{i} \frac{k_{A,l+1}}{k_{A,l+1} + k_{UF,l} + k_{FU,l}} \quad (1)$$

where i is the nascent chain length at a given point during synthesis. To use eq 1, three rates must be known at various nascent chain lengths to make predictions for a specific protein. Although for a 200-residue protein this requirement could correspond to up to 600 rates, far fewer are needed in practice, as only those rates for chain lengths in the transition region are required. The transition region is defined as the range of nascent chain lengths over which a domain in the protein of interest shifts from a primarily unfolded to a primarily folded conformation at equilibrium (Figure 4). The transition region can only be reached once the domain has emerged from the exit tunnel, and this can be taken into account in eq 1 by setting $k_{UF,i} = 0$ for chain lengths at which segments of a domain reside inside the tunnel. For a domain that folds in a two-state manner, the size of this transition region can be small, spanning just six residues (Figure 4a).[14] At nascent chain lengths shorter than the onset of this transition region, $k_{FU,i} \gg k_{UF,i} \approx 0$ (i.e., $k_{UF,i}$ is negligible relative to $k_{FU,i}$), and at lengths greater than this transition region, $k_{UF,i} \gg k_{FU,i} \approx 0$ (Figure 4b). Moreover, if correlations among the rates exist in this transition range, even fewer rates need to be measured. In this case, it may be possible to measure, for example, just the rates corresponding to every other chain length in the transition region, estimate the unmeasured rates, and still make accurate predictions using eq 1.

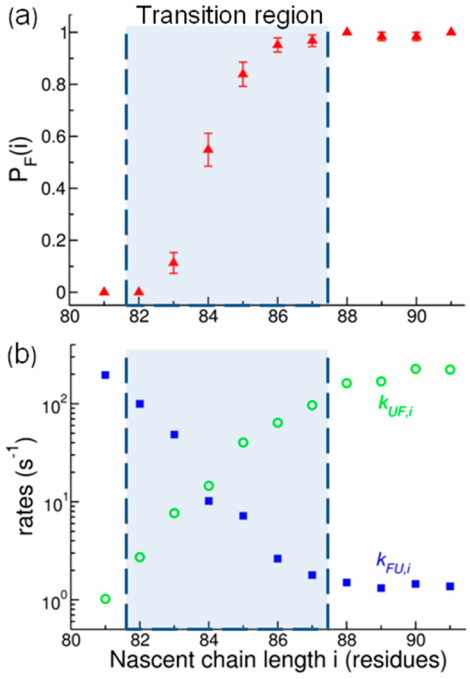


Figure 4. (a) Illustration of the transition region in the cotranslational folding curve (blue region) over which a domain is converted from a predominately unfolded to a predominantly folded state at equilibrium. (b) Outside the transition region, one rate (either $k_{UF,i}$ or $k_{FU,i}$) predominates and the other is very low, making it unnecessary to measure the low rate in order to make predictions using eq 1. The width of the transition region can vary among proteins, with noncooperative domains that populate intermediates more likely to exhibit larger transition regions.[14] Image reproduced from ref 14. Copyright 2012 Nature Publishing Group.

We note that eq 1 deals only with the elongation phase of translation, not the initiation phase. As the initiation rate determines the quantity of protein synthesized per unit time, eq 1 cannot be used to calculate the translation efficiency of individual transcripts. Furthermore, eq 1 does not account for the rates of mRNA and protein degradation. This fact, however, should not affect the predictions significantly because transcripts being actively translated are less likely to be degraded than those not undergoing translation and cotranslational protein degradation is estimated to be minor, affecting up to 16% of nascent chains in yeast.[32]

On the simulation front, to probe the molecular influence of codon translation rates on the cotranslational folding process, new coarse-grained models of the ribosome have been developed, as conventional all-atom molecular dynamics simulations cannot currently be used to simulate the time scales (tens of seconds) that are relevant to cotranslational folding.[33,34] As a result of the smoother energy landscape created by coarse graining[35] and the acceleration of dynamics by the low-friction Langevin dynamics approach,[36] coarse graining of the ribosome can effectively model behavior taking place on these time scales.

We discuss here two coarse-grained models that have recently been introduced to simulate the behavior of the ribosome,[12,13] although several other models on lattices have also been introduced.[37,38] In both models, ribosomal proteins are represented as either one or two interaction sites per residue, a description based on models previously developed for protein folding in vitro.[39,40] The two models differ, however, in

dx.doi.org/10.1021/ar5000117 | Acc. Chem. Res. 2014, 47, 1536−1544

their representation of the rRNA. In one model, an all-atom representation of the RNA is used,[12] while in the other model the RNA is represented as up to four interaction sites per nucleotide[13] (Figure 5). Both models treat solvent effects

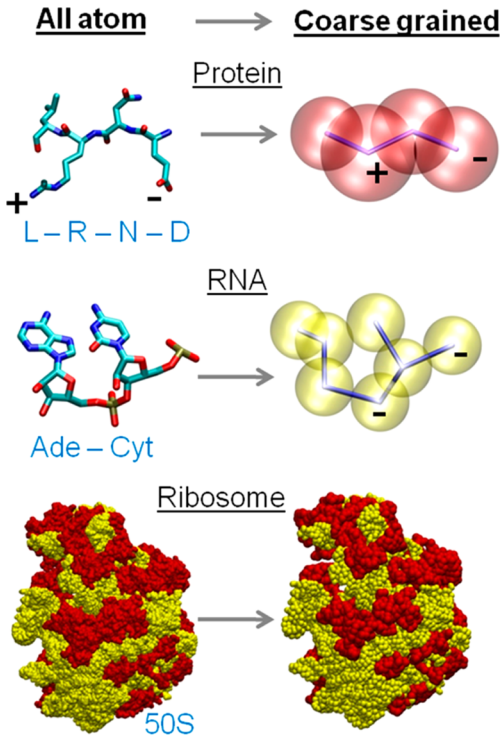


Figure 5. Coarse-grained model of the *E. coli* ribosome. Ribosomal proteins are represented as one or two interaction sites per residue, and rRNAs are represented as three or four interaction sites. As a consequence, the 50S subunit is reduced from nearly 150 000 atoms to 15 000 interaction sites.[18]

implicitly using Debye–Hückel theory to account for counter-ion charge screening and Langevin dynamics to account for solvent buffeting of the protein and RNA molecules. This type of coarse graining for domains that are relatively small in size (≤100 residues) allows hundreds of cotranslational folding events to be simulated.[14,18]

Such coarse graining has been extended to other cellular components associated with cotranslational folding, including the chaperone Trigger factor that interacts with the nascent chain cotranslationally[18] (Figure 6). In this case, the relevant pairwise binding energies between the components of the ribosome, Trigger factor, and nascent chain were set by determining the nonbonded force-field parameters that reproduced the experimentally measured dissociation constants, K_D, between the different components.[41,42] We discuss below the insights that these kinetic and simulation models have provided about cotranslational folding.

THE INFLUENCE OF CODON TRANSLATION RATES ON COTRANSLATIONAL FOLDING CAN BE PREDICTED QUANTITATIVELY

The ability of an analytical solution such as eq 1 to predict accurately the effect of codon translation rates on cotranslational folding on the basis of arrested ribosome data has recently been illustrated.[14] Data on $k_{FU,i}$ and $k_{UF,i}$ were generated from explicit coarse-grained molecular simulations

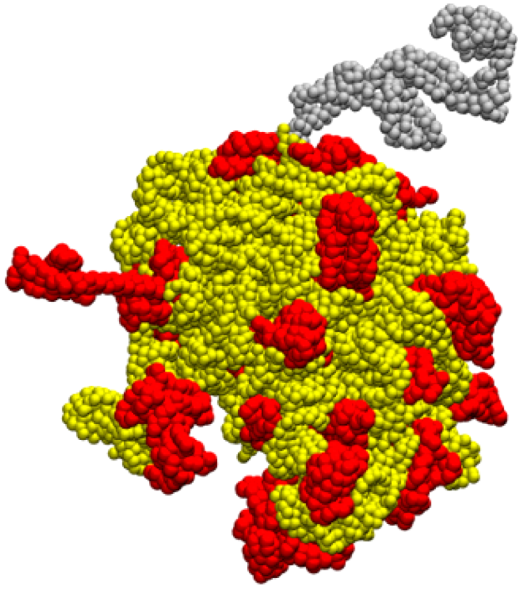


Figure 6. Representation of the 432-residue chaperone Trigger factor (gray) using a coarse-grained model, which was used to probe its mode of action on cotranslational folding.[18] Image reproduced from ref 18. Copyright 2012 American Chemical Society.

of a nascent chain, that of protein G,[43] arrested on an *Escherichia coli* ribosome at different nascent chain lengths (Figure 4b). The values of these rates were then used as the arguments in eq 1 to predict how the folding of this protein would behave during continuous translation; these predictions were then tested against explicit molecular dynamics simulations of continuous translation. The results revealed quantitative agreement between the predictions and the results from the explicit simulations (Figure 7a). Even the effect of changing the translation rate of a single codon on the cotranslational folding curve of this protein was accurately predicted by eq 1 (Figure 7b). The results illustrate in a dramatic manner how measurements[11] or estimates[15] of $k_{FU,i}$, $k_{UF,i}$, and $k_{A,i+1}$ can be used to predict accurately the influence of codon translation rates on cotranslational folding behavior.

FAST-TRANSLATING CODONS CAN COORDINATE COTRANSLATIONAL FOLDING BY AVOIDING MISFOLDING

It is commonly assumed that decreasing codon translation rates monotonically increase the probability that a domain folds before the protein is released from the ribosome.[1] Indeed, as mentioned previously, there are several experimental reports that support this conclusion. An important question, however, is whether or not this inverse relationship between translation elongation speed and cotranslational folding is a universal phenomenon.

In order to test whether other scenarios are possible, the dynamic regimes that can occur in the analytical models describing the effects of codon translation rates on domains that fold via two-state or three-state cotranslational folding mechanisms involving both on- and off-pathway intermediates were examined.[17] This objective was achieved by applying the first-derivative test for monotonicity to the analytical equations.[17] The results indicate that, counter to conventional wisdom, slow-translating codons can sometimes decrease

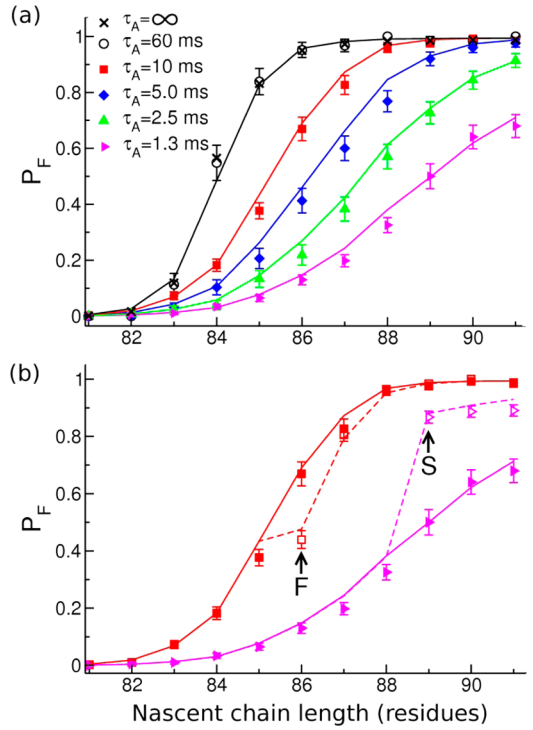


Figure 7. Prediction of the influence of codon translation rates on cotranslational folding. Predictions were made using eq 1 and tested against explicit continuous translation simulations using the coarse-grained translation model; the case of protein G is shown here.[14] (a) Probabilities of folding of the protein G domain at different codon translation times, as indicated, predicted using eq 1 (solid lines) and molecular simulations (symbols). (b) Probability of folding upon insertion of a fast translating codon ("F") and a slow translating codon ("S"). The effect of these single-codon translation rate changes are accurately predicted by eq 1. Image reproduced from ref 14. Copyright 2012 Nature Publishing Group.

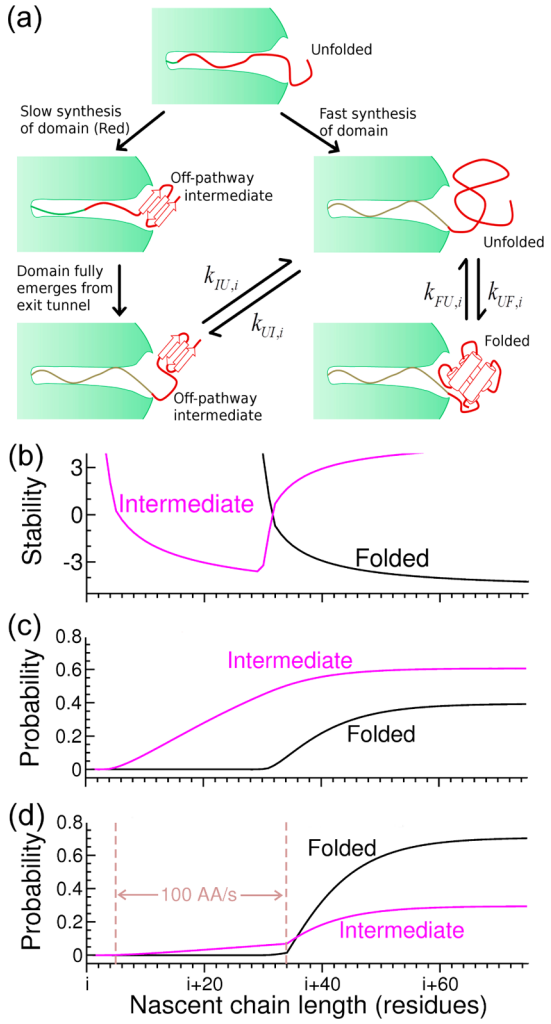


Figure 8. Fast-translating codons can coordinate cotranslational folding by avoiding misfolded intermediates. (a) Cross-section of the ribosome (green) with a nascent chain shown emerging from the exit tunnel. For a domain that can populate an off-pathway intermediate before the full domain segment has emerged (red), increasing the translation elongation rates can decrease the population of the intermediate and increase the population of the folded state. (b) Free energies of stability of the intermediate state and the folded state of a domain as functions of the nascent chain length during synthesis. It should be noted that in this scenario the intermediate can become stable at shorter nascent chain lengths than the folded state. (c) Cotranslational probabilities of populating the folded and intermediate states for the system shown in (b) when all of the codon translation rates are 10 AA/s, a typical value found in *E. coli*. (d) Same as (c) except that the translation elongation rates were increased to 100 AA/s in the region where the intermediate can become stable. Comparison of (c) and (d) reveals that this is a situation where fast-translating codons can increase the probability of cotranslational folding.[17] Image reproduced from ref 17. Copyright 2014 Nature Publishing Group.

cotranslational domain folding and fast-translating codons can increase it.

Within these models, such behavior can occur when nonmonotonic changes in domain stability occur as a function of the nascent chain length or when off-pathway intermediates can be populated during protein synthesis.[17] In the case of domains that fold in an effective two-state manner, speeding up translation through regions of a polypeptide chain where the stability of the folded state of a domain switches from increasing stability to decreasing stability can significantly increase the probability that the domain is folded upon release from the ribosome. For domains that can form an off-pathway intermediate, it was found that speeding up translation through segments of the nascent chain where the intermediate state is stable can increase the final domain folding probability (Figure 8). Thus, for proteins that fold by effective two-state and three-state mechanisms, there are scenarios where fast-translating codons can increase the probability of folding.

The likelihood that intermediates are populated during the folding process in bulk solution increases with increasing protein size, with domains that are less than 80 residues in length typically not populating any intermediate state.[44] In both prokaryotes and eukaryotes, however, protein sizes are on average between 200 and 300 residues,[45] suggesting that the potential to form intermediates could be common. Therefore, nature may have found widespread use for fast-translating codons in coordinating the cotranslational folding of proteins. This hypothesis leads to the prediction that there could be patterns of fast-translating codon usage across the transcriptome of organisms that are consistent with this newly identified function of fast codons. This prediction can be tested by using genomic engineering[4] to systematically search for fast-translating codons for which synonymous substitutions to slower-translating codons significantly decrease the probability of cotranslational folding.

dx.doi.org/10.1021/ar5000117 | *Acc. Chem. Res.* 2014, 47, 1536−1544

PROTEOME-WIDE COTRANSLATIONAL FOLDING PROPERTIES IN *E. coli* ARE GOVERNED BY TRANSLATION RATES

Kinetic modeling of cotranslational folding combined with large biological data sets is opening up new ways to predict cotranslational folding behavior of entire proteomes under varying cellular conditions. This situation is possible at least in part because of the large number of known protein structures[46] and genomic sequences. This information enables reasonable estimates of $k_{A,i+1}$, $k_{FU,i}$, and $k_{UF,i}$ to be made for each protein domain[31] and mRNA molecule[47] in an organism. Using these rates in eq 1 enables the prediction of in vivo cotranslational folding curves for all of the protein domains in a given proteome.

This approach was recently applied to the cytosolic proteome of *E. coli* to estimate the fraction of the proteome that cotranslationally folds and the degree to which kinetic effects govern cotranslational folding.[15] The results predict that close to 40% of proteins (not weighted by expression levels) in the cytosolic proteome exhibit cotranslational folding at the translation rates found in vivo, with at least one domain in each protein acquiring its stable native structure before the full-length nascent chain is released from the ribosome. In that study, the predicted behavior of individual protein domains was reported, providing a data set against which high-throughput techniques for monitoring cotranslational folding[48] can be compared.

That study also indicated that for the majority of cytosolic proteins, there is likely a deviation between cotranslational folding curves generated on arrested ribosomes ($k_{A,i+1} = 0$), where thermodynamic factors govern the nascent protein behavior, and the curves that occur at in vivo translational rates ($k_{A,i+1} > 0$). For these proteins, cotranslational folding is governed by the translation kinetics at one or more nascent chain lengths. This analysis suggests that around 20% of cytosolic proteins contain domains that fold cotranslationally on an arrested ribosome but fold post-translationally at natural translation rates. That is, synthesis in vivo can be so fast that some domains may not fold cotranslationally despite a thermodynamic driving force to do so. Thus, the translation kinetics can have profound effects on the likelihood of cotranslational folding in vivo.

The basis for the predominance of kinetics in *E. coli* cotranslational folding is that the time scale of the folding of the majority of domains is longer than the time scale of amino acid addition ($\tau_F > \tau_A$). As a consequence, these domains cannot equilibrate at every nascent chain length and thus are unable to sample effectively the extent of conformational space that is in principle available to them. According to the study in ref 15, larger domains rich in β-sheet structure tend to exhibit the largest deviations from equilibrium, while small domains containing primarily α-helical structure are more likely to exhibit quasi-equilibrium behavior. The reason for this difference is that large β-sheet domains tend to have much larger τ_F values than small α-helical domains, and this time scale, relative to τ_A, determines the cotranslational folding properties (Figure 9).

The importance of translation kinetics for cotranslational folding has other possible consequences. Some proteins may get kinetically trapped in metastable states during synthesis, which in extreme cases could result in substantially different final structures and assembly pathways from those observed in vitro. Thus, the nonequilibrium nature of cotranslational folding in vivo is likely to lead to richer self-assembly behavior than found in vitro. This scenario warrants further experimental investigation, especially with respect to the probability of nascent chain misfolding and the biological and cellular consequences for the function of newly synthesized proteins. Indeed, it has recently been shown that changing the translation rates of a designed protein in vivo can alter the conformation the protein adopts,[49] lending support to the theoretical result that at in vivo translation rates many proteins can fold under a nonequilibrium regime.

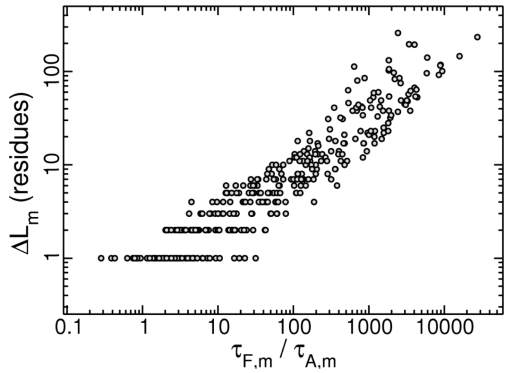


Figure 9. Deviation of cotranslational folding from equilibrium as a function of the separation of time scales. ΔL_m, a measure of the deviation from equilibrium (Figure 3A), is equal to the number of residues separating the midpoint of the cotranslational folding curve for a domain on an arrested ribosome from that for continuous translation at the in vivo translation rate. ΔL_m is plotted as a function of the ratio of the domain folding time to the amino acid addition time at the midpoint nascent chain length.[15] Image reproduced from ref 15. Copyright 2012 National Academy of Sciences of the United States of America.

CONCLUDING REMARKS

A variety of experiments have established that the rate at which individual codons are translated can strongly influence whether some proteins fold to functional states or misfold and malfunction.[1,3,5,7,9,10,14,15,50] As misfolded species often arise from partially folded intermediates, some proteins are particularly sensitive to conformational disruption due to inefficient cotranslational folding. Therefore, determining the points during translation where proteins are likely to remain unfolded or misfold could shed light on the molecular mechanisms of some forms of cellular stress. In this Account, we have summarized recent efforts to develop a molecular perspective and a theoretical framework to understand, model, and predict the influences of codon translation rates on these cellular processes. The approaches being developed and utilized range from techniques in the areas of molecular simulations to kinetic analyses and systems biology.

Each approach makes its own assumptions, which define the range of questions that it can effectively address and form the basis for new developments. Thus, for example, conventional chemical kinetic models are limited in their ability to predict single-molecule behavior, at least in part because of the effects of stochasticity, but hybrid approaches that blend chemical kinetics and a discrete time formalism suggest one way forward.[14] Estimating codon translation rates under varying cellular conditions is also a significant challenge, but ribosome profiling data may be very useful, as the number of "reads" (i.e.,

the signal) from such experiments is a function of the codon elongation rates.[51] Other methods and models of translation rates may also prove to be very useful.[4] It is important to test by experiment the predictions from these approaches, and we hope that this review will help to promote quantitative experimental measurement of the coupling of cotranslational folding to elongation rates. For example, to our knowledge only two in vivo cotranslational folding curves have been experimentally measured.[24] Advances in single-molecule techniques and in vivo imaging methods hold out the promise that more examples will be reported in the future.

Thus, theory and computation are making vital contributions to our understanding of cotranslational folding and proteome metastability in living cells. With the continuing advance of quantitative, real-time measurements of processes in living cells, the contributions and necessity of these theoretical approaches will undoubtedly grow in the future.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: epo2@psu.edu.
*E-mail: cmd44@cam.ac.uk.

Present Address

[§]E.P.O.: Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA.

Notes

The authors declare no competing financial interest.

Biographies

Edward O'Brien received his B.S. in Biochemistry from University of the Sciences and his Ph.D. in Chemical Physics from the University of Maryland College Park, where he was a NIH GPP Biophysics Fellow and also an NSF EAPSI Fellow in the summer of 2007. He was an NSF Postdoctoral Fellow and EPSRC grant recipient while at the University of Cambridge. He is now an Assistant Professor at Pennsylvania State University, University Park.

Prajwal Ciryam received his B.A. in Biological Sciences from Northwestern University and is completing a Ph.D. in Chemistry from the University of Cambridge, where he was a Fulbright Scholar at St. John's College, and a Ph.D. in Neuroscience from Northwestern University, where he is also enrolled in the M.D. program as part of the Medical Scientist Training Program.

Michele Vendruscolo received his Ph.D. in Condensed Matter Physics from Scuola Internazionale Superiore di Studi Avanzati di Trieste. He has been a Royal Society University Research Fellow at the Department of Chemistry, University of Cambridge where he is now a Professor of Biophysics.

Christopher M. Dobson received his B.A. and D.Phil. in Chemistry from the University of Oxford. He is now the John Humphrey Plummer Professor of Chemical and Structural Biology and the Master of St. John's College at the University of Cambridge. He is a Fellow of the Royal Society and of the Academy of Medical Sciences and a Foreign Associate of the National Academy of Sciences.

ACKNOWLEDGMENTS

We thank John Christodoulou and Lisa Cabrita for very useful discussions concerning cotranslational folding and Attila Szabo for stimulating suggestions for modeling approaches. We thank the Engineering and Physical Sciences Research Council (UK) for financial support (E.P.O., M.V., C.M.D.) and the National Science Foundation (USA) for the award of a postdoctoral fellowship to E.P.O.

REFERENCES

(1) Komar, A. A. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **2009**, *34*, 16−24.
(2) Cabrita, L. D.; Dobson, C. M.; Christodoulou, J. Protein folding on the ribosome. *Curr. Opin. Struct. Biol.* **2010**, *20*, 33−45.
(3) Zhang, G.; Hubalewska, M.; Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **2009**, *16*, 274−280.
(4) Spencer, P. S.; Siller, E.; Anderson, J. F.; Barral, J. M. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* **2012**, *422*, 328−335.
(5) Siller, E.; DeZwaan, D. C.; Anderson, J. F.; Freeman, B. C.; Barral, J. M. Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *J. Mol. Biol.* **2010**, *396*, 1310−1318.
(6) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(7) Kimchi-Sarfaty, C.; Oh, J. M.; Kim, I. W.; Sauna, Z. E.; Calcagno, A. M.; Ambudkar, S. V.; Gottesman, M. M. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **2007**, *315*, 525−528.
(8) Edwards, N. C.; Hing, Z. A.; Perry, A.; Blaisdell, A.; Kopelman, D. B.; Fathke, R.; Plum, W.; Newell, J.; Allen, C. E.; S., G.; Shapiro, A.; Okunji, C.; Kosti, I.; Shomron, N.; Grigoryan, V.; Przytycka, T. M.; Sauna, Z. E.; Salari, R.; Mandel-Gutfreund, Y.; Komar, A. A.; Kimchi-Sarfaty, C. Characterization of Coding Synonymous and Non-Synonymous Variants in *ADAMTS13* Using Ex Vivo and In Silico Approaches. *PLoS One* **2012**, *7*, No. e38864.
(9) Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J. M.; Sachs, M. S.; Liu, Y. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **2013**, *495*, 111−115.
(10) Tsai, C. J.; Sauna, Z. E.; Kimchi-Sarfaty, C.; Ambudkar, S. V.; Gottesman, M. M.; Nussinov, R. Synonymous Mutations and Ribosome Stalling Can Lead to Altered Folding Pathways and Distinct Minima. *J. Mol. Biol.* **2008**, *383*, 281−291.
(11) Kaiser, C. M.; Goldman, D. H.; Chodera, J. D.; Tinoco, I., Jr.; Bustamante, C. The ribosome modulates nascent protein folding. *Science* **2011**, *334*, 1723−1727.
(12) Elcock, A. H. Molecular simulations of cotranslational protein folding: Fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol.* **2006**, *2*, 824−841.
(13) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. New Scenarios of Protein Folding Can Occur on the Ribosome. *J. Am. Chem. Soc.* **2011**, *133*, 513−526.
(14) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.* **2012**, *3*, No. 868.
(15) Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P. In vivo translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E132−E140.
(16) O'Brien, E. P.; Hsu, S. T. D.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Transient Tertiary Structure Formation within the Ribosome Exit Port. *J. Am. Chem. Soc.* **2010**, *132*, 16928−16937.
(17) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat. Commun.* **2014**, *5*, No. 2988.
(18) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Trigger Factor Slows Co-translational Folding through Kinetic Trapping While Sterically Protecting the Nascent Chain from Aberrant Cytosolic Interactions. *J. Am. Chem. Soc.* **2012**, *134*, 10920−10932.
(19) Chandler, D. Self-assembly where kinetics trumps thermodynamics. *Abstr. Pap.—Am. Chem. Soc.* **2007**, *233*, PHYS 132.

(20) Rodnina, M. V.; Wintermeyer, W. The ribosome as a molecular machine: The mechanism of tRNA−mRNA movement in translocation. *Biochem. Soc. Trans.* **2011**, *39*, 658−662.
(21) Fedyukina, D. V.; Cavagnero, S. Protein Folding at the Exit Tunnel. *Annu. Rev. Biophys.* **2011**, *40*, 337−359.
(22) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
(23) Knight, A. M.; Culviner, P. H.; Kurt-Yilmaz, N.; Zou, T.; Ozkan, S. B.; Cavagnero, S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **2013**, *8*, 1195−1204.
(24) Nicola, A. V.; Chen, W.; Helenius, A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell Biol.* **1999**, *1*, 341−345.
(25) Khushoo, A.; Yang, Z.; Johnson, A. E.; Skach, W. R. Ligand-Driven Vectorial Folding of Ribosome-Bound Human CFTR NBD1. *Mol. Cell* **2011**, *41*, 682−692.
(26) Chang, H. C.; Kaiser, C. M.; Hartl, F. U.; Barral, J. M. De novo folding of GFP fusion proteins: High efficiency in eukaryotes but not in bacteria. *J. Mol. Biol.* **2005**, *353*, 397−409.
(27) Sorensen, M. A.; Kurland, C. G.; Pedersen, S. Codon Usage Determines Translation Rate in *Escherichia coli*. *J. Mol. Biol.* **1989**, *207*, 365−377.
(28) Komar, A. A.; Lesnik, T.; Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **1999**, *462*, 387−391.
(29) Clarke, T. F., IV; Clark, P. L. Rare codons cluster. *PLoS One* **2008**, *3*, No. e3412.
(30) Angov, E.; Hillier, C. J.; Kincaid, R. L.; Lyon, J. A. Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. *PLoS One* **2008**, *3*, No. e2189.
(31) De Sancho, D.; Munoz, V. Integrated prediction of protein folding and unfolding rates from only size and structural class. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17030−17043.
(32) Duttler, S.; Pechmann, S.; Frydman, J. Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol. Cell* **2013**, *50*, 379−393.
(33) Whitford, P. C.; Geggier, P.; Altman, R. B.; Blanchard, S. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* **2010**, *16*, 1196−1204.
(34) Sanbonmatsu, K. Y. Computational studies of molecular machines: The ribosome. *Curr. Opin. Struct. Biol.* **2012**, *22*, 168−174.
(35) Levy, Y.; Onuchic, J. N. Mechanisms of protein assembly: Lessons from minimalist models. *Acc. Chem. Res.* **2006**, *39*, 135−142.
(36) Klimov, D. K.; Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **1997**, *79*, 317−320.
(37) Wang, P. Y.; Klimov, D. K. Lattice simulations of cotranslational folding of single domain proteins. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 925−937.
(38) Krobath, H.; Shakhnovich, E. I.; Faisca, P. F. Structural and energetic determinants of co-translational folding. *J. Chem. Phys.* **2013**, *138*, No. 215101.
(39) Karanicolas, J.; Brooks, C. L. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **2002**, *11*, 2351−2361.
(40) Best, R. B.; Chen, Y. G.; Hummer, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure* **2005**, *13*, 1755−1763.
(41) Maier, R.; Scholz, C.; Schmid, F. X. Dynamic association of trigger factor with protein substrates. *J. Mol. Biol.* **2001**, *314*, 1181−1190.
(42) Rutkowska, A.; Mayer, M. P.; Hoffmann, A.; Merz, F.; Zachmann-Brand, B.; Schaffitzel, C.; Ban, N.; Deuerling, E.; Bukau, B. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **2008**, *283*, 4124−4132.
(43) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein-G. *Science* **1991**, *253*, 657−661.
(44) Braselmann, E.; Chaney, J. L.; Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **2013**, *38*, 337−344.
(45) Brocchieri, L.; Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **2005**, *33*, 3390−3400.
(46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(47) Fluitt, A.; Pienaar, E.; Viljoen, H. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* **2007**, *31*, 335−346.
(48) Han, Y.; David, A.; Liu, B.; Magadan, J. G.; Bennink, J. R.; Yewdell, J. W.; Qian, S. B. Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12467−12472.
(49) Sander, I. M.; Chaney, J. L.; Clark, P. L. Expanding Anfinsen's Principle: Contributions of Synonymous Codon Selection to Rational Protein Design. *J. Am. Chem. Soc.* **2014**, *136*, 858−861.
(50) Netzer, W. J.; Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **1997**, *388*, 343−349.
(51) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R. S.; Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **2009**, *324*, 218−223.
(52) Liang, S. T.; Xu, Y. C.; Dennis, P.; Bremer, H. mRNA composition and control of bacterial gene expression. *J. Bacteriol.* **2000**, *182*, 3037−3044.