In vivo translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome

Prajwal Ciryma,b, Richard I. Morimotoa, Michele Vendruscoloa, Christopher M. Dobsona, and Edward P. O’Brienb,c

*a*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; and *b*Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208

**AUTHOR SUMMARY**

In the process of translation, ribosome molecules synthesize proteins from the genetic information carried by mRNA molecules. These nascent proteins must self-assemble into unique, ordered structures to carry out their respective biological functions, and they can begin to do so concomitantly with their synthesis in a phenomenon referred to as cotranslational folding (1, 2) (Fig. P1). In vivo translation is time-dependent and irreversible; it is therefore out of equilibrium, which means the kinetics of translation, rather than the thermodynamics of the ribosome nascent chain (RNC) complex, can govern cotranslational folding behavior. The extent to which cotranslational folding is exhibited by the proteome of any organism is unknown. However, it is important to quantify this phenomenon because the ribosome is the first place in the cell where proteins can self-assemble, and errors in this self-assembly process can lead to the formation of potentially toxic protein aggregates. Furthermore, understanding the extent to which translation elongation kinetics affect cotranslational folding can shed light on a range of issues in synthetic and molecular biology, including codon usage bias in the transcriptome and the coupling of nascent protein synthesis to other downstream processes, such as interactions with chaperones. Using a systems biology approach, we estimate that approximately one-third of the cytosolic proteome of the bacterium *Escherichia coli* exhibits cotranslational folding and that the majority of these proteins exhibit delayed folding due to the relatively slow speed of translation in vivo.

Although the cotranslational folding properties of an increasing number of proteins are being characterized experimentally in vitro (3), little is known about cotranslational folding in vivo for entire proteomes of specific organisms. Therefore, one fundamental question that arises concerns the percentage of an organism’s proteome that folds cotranslationally. A second fundamental question is the extent to which proteins that fold cotranslationally do so under a regime governed by the kinetics of translation (“kinetic effects”) or the thermodynamics of the RNC complex (“thermodynamic effects”). These issues are important because cotranslational folding populations, pathways, structures and mechanisms can differ significantly depending on whether kinetic or thermodynamic effects predominate.

To address these questions, we focused on the cytosolic proteome of *E. coli* because its genome and proteome have been characterized extensively and the concentrations of ribosomes and tRNA molecules present at different growth rates have been measured. We restricted our study to *E. coli* proteins that are synthesized and remain in the cytosol to avoid considering additional processes, such as cotranslational protein translocation, that would complicate our modeling and analysis. Our systems approach combines a kinetic model that predicts the effects of individual codon translation rates on the cotranslational folding of protein domains (4) with biophysical data to predict cotranslational folding curves for protein domains in the *E. coli* cytosol.

We assembled a database of sequences and structures for 1,236 domains from 758 unique proteins representing ~30% of the *E. coli* K12 MG1655 cytosolic proteome. To analyze the cotranslational folding of these proteins, we used a chemical kinetic equation that takes into account the three most relevant time scales: (a) the mean time required for addition of an amino acid to the growing nascent chain, (b) the mean time of domain unfolding, and (c) the mean time of domain folding. We have thus been able to calculate cotranslational folding curves at the rates of translation found in *E. coli*, as well as the curves that would arise from an infinitely slow translation process corresponding to ribosomes in which translation has been arrested.

Our results indicate that about one-third of *E. coli* cytosolic proteins exhibit cotranslational folding. The overwhelming majority (91%) of cotranslationally folding proteins consist of multiple domains, whereas among proteins that fold posttranslationally, single and multidomain proteins are equally represented (51% vs. 49%). Further, the structural characteristics of cotranslationally folding domains differ from those of posttranslationally folding domains. Cotranslationally folding domains are smaller in size on average (106 vs. 177 residues) and are more likely to be predominantly α-helical.

Author contributions: P.C., R.I.M., M.V., C.M.D., and E.P.O. designed research; P.C. and E.P.O. performed research; P.C. and E.P.O. contributed new reagents/analytic tools; P.C., R.I.M., M.V., C.M.D., and E.P.O. analyzed data; and P.C., R.I.M., M.V., C.M.D., and E.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: eo264@cam.ac.uk.


Cite this Author Summary as: PNAS 10.1073/pnas.1213624110.
Posttranslationally folding domains, in contrast, tend to consist of mixed α/β structures.

We find that cotranslational folding in the E. coli cytosol is sensitive to the kinetics of translation, which can substantially delay folding along the nascent chain length. Of the 1,236 cytosolic domains in our database, 602 exhibit cotranslational folding at infinitely slow translation rates. The faster translation rates found in vivo in E. coli at 37 °C cause 180 of these domains to switch from cotranslational to posttranslational folding. For the 422 domains that do exhibit cotranslational folding at in vivo translation rates, the majority exhibit delays in cotranslational folding relative to their folding on arrested ribosomes. Notably, large β-strand-rich domains are, on average, particularly sensitive to these kinetic effects.

For a small percentage of proteins, we find that synonymous codon mutations can significantly decrease kinetic effects. In silico, we mutated the wild-type transcriptome to the slowest translating transcriptome possible by using synonymous codons and then recalculated cotranslational folding curves for this transcriptome. For most cytosolic proteins, these synonymous mutations have little effect on cotranslational folding. For a minority, however, synonymous mutations can significantly reduce delays in cotranslational folding caused by the in vivo translation rate, and thereby increase the folded population at shorter nascent chain lengths.

The systems approach we present here indicates that kinetic effects are widespread in the cotranslational folding of the proteome. This suggests that translation-related time scales can be manipulated for synthetic biology, biotechnology, and cellular engineering purposes. For example, our approach can be used to redesign the transcriptome using synonymous codons to increase cotranslational folding or to target particular transcripts whose proteins we predict could switch from post- to cotranslational folding in vivo. In addition to informing strategies for redesigning mRNA sequences, this analysis may shed light on the dynamics of protein homeostasis. For example, our observation that larger domains are more likely to fold posttranslationally may explain why the chaperones Trigger Factor and DnaK, which assist unfolded proteins in reaching their folded state, bind preferentially to the nascent chains of larger proteins (5–7).

This study offers a theoretical framework that can address such questions in other cells and tissues in both prokaryotes and eukaryotes, an approach that has the potential to yield insight about the mechanisms of folding in vivo and to enable the rational engineering of entire transcriptomes.