Influence of translation rate on co-translational folding intermediates

Thomas Clarke IV[1] **and Patricia L. Clark**[2]

Keywords: codon usage, co-translational folding, tailspike

1 Introduction.

Translation in bacteria occurs at an average rate of 20 aa/sec. Translation rate, however, is not uniform; transient pauses can arise due to rare codons (*4*), mRNA secondary structure (*2*), or depletion of a particular aminoacyl-tRNA (*1*). These pauses may represent points at which co-translational events, such as structure formation, domain organization, or recognition factor binding can occur. To investigate the role of translation pausing in co-translational folding, we have developed a novel algorithm for evaluating codon usage (and, by extension, translation pause sites arising from the use of rare codons) in mRNA sequences.

To date, we have focused on protein nascent chains from organisms utilizing similar (*Salmonella typhimurium* phage P22) or disparate (*Aequorea victoria*) codon usages (*3*) and translational processes from the model organism *Escherichia coli*. Significant rare codon clusters have been identified, especially within the tailspike gene from *S. typhimurium* phage P22 (Figure 1); the difficulty with exogenous expression of wild-type GFP in *E. coli* is also implied by the differences produced by using the codon frequencies from *E. coli* and *A. victoria* (Figure 2).

The verification of co-translational pausing has been performed utilizing *ex vivo* purification of nascent chains from *E. coli* (Figure 3). The relationship between rare codon clusters, as measured by the algorithm, and the appearance of co-translational pausing has been established experimentally using silent mutagenesis of rare codon clusters. Application of this codon usage algorithm to mRNA sequences encoding a variety of structural topologies has been used to develop a hypothesis regarding the effects of translation pausing on co-translational structure formation.

2 Software and files

Potential co-translational pauses induced by clusters rare codons was performed utilizing an algorithm based on the simple statistical distance between the codon frequency for the actual mRNA sequence and the frequency extremes (rare or common) over a sliding window. Codon usage frequencies were obtained from publicly available databases (*3*).

3 Figures and tables.

[1] Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556; Email: tclarke1@nd.edu

[2] Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556; Email: pclark1@nd.edu

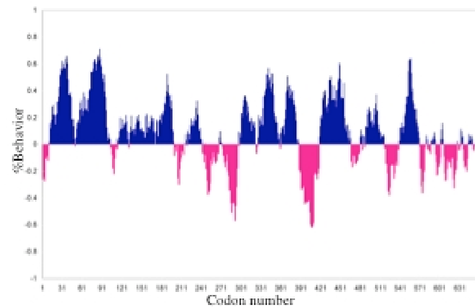

Figure 1. Output from the %MinMax algorithm. The output from the %MinMax algorithm for *S. typhimurium* phage P22 tailspike (A) shows several significant negative (magenta) peaks, denoting clusers of rare codons. Areas above the average axis (blue) utilize more common codons and would not be predicted to generate co-translational pauses based on their codon usage.

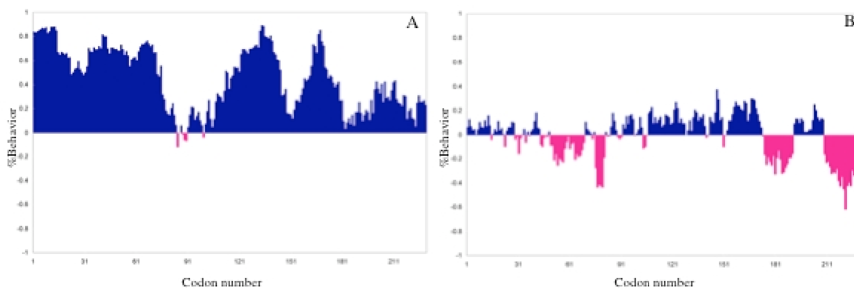

Figure 2. Output from %MinMax for green fluorescent protein. Values for green fluorescent protein (GFP) in the native host (A, *A. victoria)* are extremely positive, indicating optimization for high levels of expression; exogenous expression in *E. coli*, however, produces a graph with significant negative peaks (B), hampering over-expression.

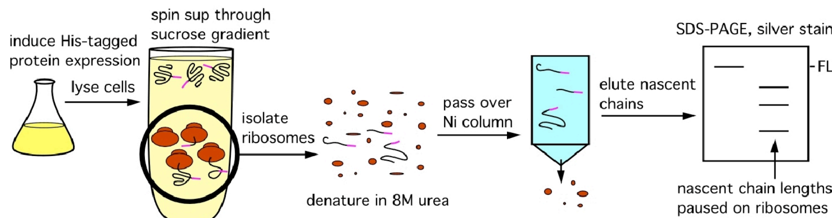

Figure 3. Protocol for nascent chain isolation. An N-terminal His-tag allows for all target nascent protein chains to be specifically isolated. Slower translating regions will be more populated, forming visible bands after silver staining.

4 References

1. Bula, C., and Wilcox, K. W. (1996) Negative effect of sequential serine codons on expression of foreign genes in Escherichia coli. *Protein Expr Purif 7*, 92-103.
2. Guisez, Y., Robbens, J., Remaut, E., and Fiers, W. (1993) Folding of the MS2 coat protein in Escherichia coli is modulated by translational pauses resulting from mRNA secondary structure and codon usage: a hypothesis. *J Theor Biol 162*, 243-52.
3. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res 28*, 292.
4. Pedersen, S. (1984) Escherichia coli ribosomes translate in vivo with variable rate. *Embo J 3*, 2895-8.