# A signal processing-based model for analyzing programmed frameshifts

[1]L. Ponnala, [2]T. Barnes, [1]D. Bitzer, [1]M. Vouk and [1]A. Stomp
[1]North Carolina State University, Raleigh, NC, USA
[2]University of North Carolina at Charlotte, NC, USA

***Abstract*— The process of translation in prokaryotes involves the interaction of the 3' tail end of the 16S ribosome with the messenger RNA (mRNA) sequence to produce a polypeptide chain. The set of free energy estimates for the binding of the ribosomal tail end along the mRNA sequence constitutes a discrete periodic signal. There are locations along the mRNA in certain genes where the ribosome changes reading frame with high efficiency. Such an event is referred to as a "programmed" translational frameshift. We present a model that uses the parameters of the free energy signal to analyze the frameshifting phenomenon. We discuss preliminary results obtained using the *prfB* gene in E.coli K-12.**

## I. Introduction

One of the most important processes in the cell is the conversion of genetic information into proteins. This process is initiated by transcription of information contained in the DNA to mRNA, which later is translated into proteins. Translation occurs via the synchronized action of ribosomes, mRNA, aminoacylated transfer RNAs, and a huge number of different protein factors [1]. It is a complex energy-requiring process that also depends on transfer RNA (tRNA) molecules that adapt specific codons in mRNA to their correct amino acids. The tRNA molecules are first charged under the direction of aminoacyl tRNA synthetases, and then chemically linked to their respective amino acids. Since prokaryotes do not have nuclei, it is probable that translation takes place as soon as mRNA is created.

In standard translation, ribosomes read mRNA in steps of three bases. However, specific mRNAs carry "programmed" frameshifting signals. If the ribosome skips a single base in the downstream (3') direction, it is referred to as a $+1$ frameshift. A $-1$ frameshift occurs when the ribosomal reading frame overlaps one base in the upstream (5') direction. Two general features of known programmed frameshifting sites are a slippery sequence in the mRNA that favors tRNA movement or misalignment, and a stimulator that enhances the process, probably by induction of a ribosomal pause [2] [3]. Notable examples of such sites occur in the *prfB* gene ($+1$ frameshift) and the dnaX gene ($-1$ frameshift) in E.coli [4].

## II. Mathematical Model

The ribosome's movement along the mRNA requires energy. Our hypothesis is that a small but significant part of this energy is supplied by the release of free-energy that arises from hybridization between the mRNA and the 3'-end of the 16S rRNA.

A particular alignment of the 3' exposed tail end against the messenger RNA is referred to as a conformation. The binding energy released in this conformation, also referred to as free energy, is estimated using the method of base-doublets [5]. This calculation penalizes mismatches and rewards consecutive base pairing in the conformation. A shift along the mRNA by one base position results in a new conformation, and the calculation of the free energy estimate is repeated. The binding energies for matching doublets are determined by experiment and are listed in [6]. The set of free energy estimates for all possible conformations along the mRNA sequence constitutes a discrete signal that can be analyzed using methods of discrete-time signal processing [7]. This signal is also referred to as the "synchronization signal" since its 3-base periodicity appears to reflect the movement of the ribosome in steps of 3 bases [5].

We fit a sine-wave to the free-energy signal cumulatively in increments of 3 bases, i.e. one codon, in length. The minimum point on this sine wave is considered the "minimum-energy" position. We estimate the magnitude and phase of the signal at each increment [8][9]. The cumulative signal vector is thus computed at each codon position. The differential vector (magnitude $D_k$, phase $\psi_k$) at codon $k$ is calculated by computing the vector difference between the cumulative vector at codon $k+1$ and that at codon $k$.

We now relate changes in the phase of the free energy signal to the displacement of its minimum-energy point. If the position of the minimum-energy point changes between one codon and the next, we interpret this as a "push" or "pull" in the movement of the ribosome. When this displacement is too large, we believe that the probability of frameshifting becomes much greater. Our hypothesis is that the component of the differential vector in a direction orthogonal to the cumulative vector contributes to the change in position.

Let $\Delta x_k$ represent the displacement caused at the $k^{th}$ increment. If we accumulate these $\Delta x_k$ values at every codon, we obtain a cumulative signal $x_k$ as a function of codon position. This signal tells us how the position of the minimum-energy point varies by codon. The following equations are used to compute the function $x_k$:

$$\Delta x_k = AD_k sin(\phi_k - \psi_k) \tag{1}$$

$$x_k = x_{k-1} + \Delta x_k \tag{2}$$

$$\phi_{k+1} = (\pi/3)x_k - |\Phi_{sp}| \tag{3}$$

where $\phi_k$ is the phase of the cumulative signal vector. We use constants $A = -0.1$, $\Phi_{sp} = -\pi/6$ for E.coli determined by examining the signals of a number of verified genes.

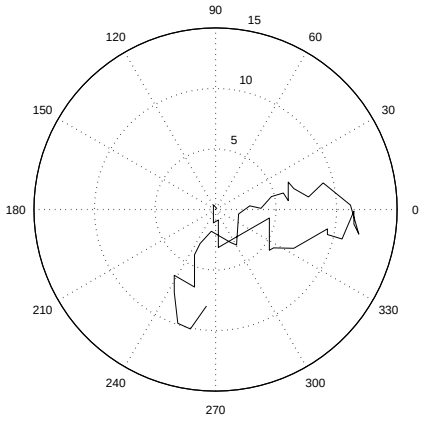


Fig. 1. Polar plot of *prfB* gene.

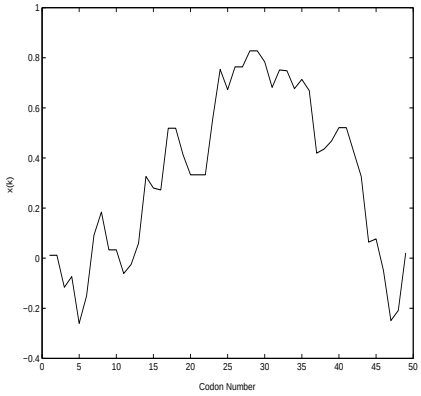


Fig. 2. Cummulative position $x$ as a function of codon number.

## III. Results and Discussion

The *prfB* gene in E.coli has a stop-codon at the $26^{th}$ codon-position. The ribosome skips one base in translation, and avoids encountering the stop codon. The polar plot Figure 1 shows the cumulative signal $\phi_k$ of the *prfB* gene starting out in the E.coli-specific phase angle of $-\pi/6$. There is a precipitous change in the phase of the signal close to the frameshift location. We also find that the position of the minimum-energy point Figure 2 gradually changes until the ribosome goes out-of-frame at codon 26. After the frameshift position, the differential vectors indicate a restoration of the minimum-energy point of the signal to its original position. This appears to indicate that the ribosome has now "adapted" itself to the new reading frame.

A validation of this model for other cases of programmed translational events, such as $-1$ frameshifts and ribosomal hops is pending. This method may be particularly promising in understanding programmed frameshifts in prokaryotes, and also in understanding substitute amino acids, "stress" proteins, and overall protein production efficiency.

## References


[1] W. S. Klug and M. R. Cummings, *Concepts of Genetics*, Prentice Hall, Upper Saddle River, NJ, 5th edition, 1997.

[2] Farabaugh, P., "Programmed translational frameshifting", *Annual Review of Genetics*, 30:507-528, 1996.

[3] H. Kontos, S. Napthine, and I. Brierley, "Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency", *Mol Cell Biol*, 21(24): 8657-8670, 2001.

[4] Engelberg-Kulka H. and Schoulaker-Schwarz R., "Regulatory implications of translational frameshifting in cellular gene expression", *Mol Microbiol.*, 11(1):3-8, 1994.

[5] D.I. Rosnick, "Free energy periodicity and memory model for genetic coding", PhD thesis, North Carolina State University, Raleigh, NC, 2001.

[6] S. M. Freier et al., "Improved free-energy parameters for predictions of RNA duplex stability", *Proc. Natl. Acad. Sci. USA*, 83: 9373-9377, 1986.

[7] A. V. Oppenheim and R. W. Schafer, *Discrete Time Signal Processing*, Prentice Hall, March 1989.

[8] L. Ponnala, T. Barnes, D. Bitzer and M. Vouk, "The search for the optimal ribosome 3' tail end in E.coli", *Proc. IEEE-EMBS*, September 2004.

[9] M. Mishra, S.Vu, D. Bitzer and M. Vouk, "Free energy periodicity in prokaryotic coding and its Role in Identification of +1 ribosomal frameshifting in the escherichia coli K-12 gene prfB", *Proc. IEEE-EMBS*, September 2004.

[10] Trifonov E., "Recognition of correct reading frame by the ribosome, Biochimie", *Biochimie*, 74(4):357-62, 1992.

[11] Atkins J., Weiss R. and Gesteland R., "Ribosome gymnastics - degree of difficulty 9.5, style 10.0", *Cell*, 62, 413-423, 1990.

[12] Crick,F., Barnett,L., Brenner,S. and Watts-Tobin,R., "General nature of the genetic code for proteins", *Nature*, 192, 1227-1232, 1961.

[13] Farabaugh, P. and Bjork, G., "How translational accuracy influences reading frame maintenance", *EMBO J.*, 18, 1427-1434, 1999.