My research interests focus on applying stochastic processes to model complex systems. My primary current interest is in modeling small-scale biophysical systems, although I am interested in larger scale systems as well. In my dissertation research, I am investigating -1 programmed ribosomal frameshifting (PRF), a phenomenon that occurs during the translation of viral RNA into proteins in infected cells. I have developed a stochastic model of RNA translation and frameshifting that predicts frameshifting efficiencies and translation rates based on the known sequence of a given strand of RNA. Previous models of translation have not accounted for the possibility of frameshifts.

Programmed ribosomal frameshifting is largely specific to viruses; the mechanism is not required to synthesize any known human, plant, animal, or fungal proteins. This makes -1 PRF a potential target for new antiviral drugs. Understanding PRF and protein synthesis in more detail is also of interest to molecular and cellular biologists and to biophysicists who study single-molecule dynamics. Mathematically, stochastic processes are powerful tools for modeling systems in which the actions of individuals and the occurrence of rare events have significant consequences for the long-term behavior of the system.

Background

In the translation process, a ribosome attaches to a messenger RNA (mRNA) and progresses along it through a series of conformational changes, reading the sequence of bases one codon at a time. Transfer RNAs (tRNA) bring the amino acids that correspond to each codon. Since codons are triplets of bases, there are three possible reading frames in any mRNA. A start codon establishes the open reading frame for the ribosome and determines the set of codons to be read and thus the sequence of amino acids to be constructed. If the ribosome slips forward or back by one base during translation, it will read a different set of codons from that point on, and the ribosome will produce a different chain of amino acids. This is called a frameshift. Error-correcting mechanisms in the translation process normally either prevent unwanted frameshifts from occurring or cause the ribosome to disengage from the mRNA after a frameshift, thus aborting the protein synthesis.

Certain viral genes overlap in such a way that the second gene cannot be translated unless the ribosome does shift backward by one base before finishing the translation of the first gene [1]. These frameshifts are stimulated by structures in the mRNA. Empirical studies have determined that three components within the mRNA are necessary for -1 PRF to occur: (1) a slippery site, which is a sequence of seven bases in a specific pattern that allows slippage to be energetically possible; (2) a short section of 5 to 8 bases called the spacer, immediately following the slippery site; and (3) a stimulatory structure following the spacer, in which the mRNA has folded and bonded to itself [2]. In effect, the mRNA programs the ribosome to frameshift at a specific point. The efficiency of this frameshifting varies from 1-50%, depending on the species of virus [3]. This efficiency determines the ratio of the two proteins that are produced, a feature that is critical to the reproduction of the virus.

None of the mathematical models of translation reported in the literature to date allow for the possibility of frameshifting (e.g. [4]). One challenge in modeling translation with frameshifting is that when one considers only the ribosome's position on the mRNA as the state space, frameshifts destroy the Markov property of the process: both the ribosome's position and its prior path affect the probabilities of transition to the next codon, as does the nonlocal effect of the configuration of the stimulatory structure. Furthermore, the details of RNA folding and unfolding kinetics are not well known, but have a significant effect on frameshift efficiencies. Only one model in the literature

addresses ribosomal frameshifting efficiency [5]; however, that model looks only at averages over large numbers of events and provides no information about the timing or rates at which the events happen. I have developed a model that incorporates both translation rates and probabilities of frameshifting.

Current Work

The model

In my current work, the cycle of translation steps is modeled as a continuous time, discrete state Markov process. Each state represents a configuration of the ribosome-mRNA-tRNA complex within a four-dimensional state space, which includes the position of the ribosome, the location of the downstream structure in the mRNA, and the pair of tRNA molecules that are interacting with the ribosome and mRNA. The time required for each step is chosen from an exponential distribution whose average value depends on biophysical parameters relevant to that step. A diagram of the cycle of steps is shown in Figure 1.

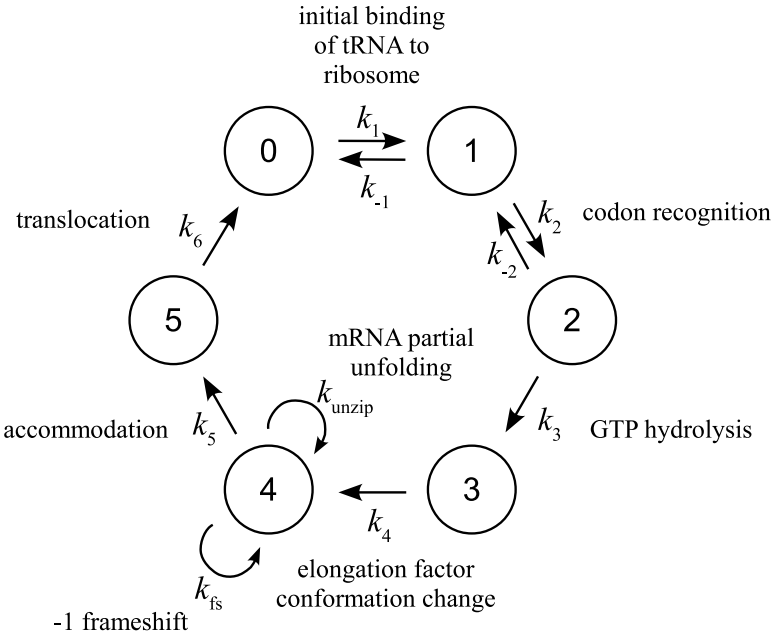

Figure 1: A diagram of the model of the translation cycle.

Values of the fixed parameters come from a combination of physical and biochemical theory of molecular interactions [6] and empirical values taken from the literature (see for example [7]).

The fifth step in the cycle, which forms the central focus of this research, determines the frameshifting efficiency and thus the ratio of viral proteins produced. At this step, either the ribosome will frameshift (with rate constant k_{fs}), the mRNA structure will partially unfold (with rate constant k_{unzip}), or the system will go on to the next step (with rate constant k_5). The probability

of a frameshift occurring is then given by

$$P[FS] = \frac{k_{\text{fs}}}{k_{\text{fs}} + k_{\text{unzip}} + k_5}, \tag{1}$$

with similarly defined probabilities for the other two possible events. The outcome is then determined stochastically via the Gillespie algorithm [8].

Each rate constant is calculated based on an associated energy E_i:

$$k_i = Z_i \exp\left(E_i/RT\right) \tag{2}$$

where k_i denotes the rate constant for the ith transition and Z_i is a prefactor that may depend on temperature, species concentrations, diffusion rates, and steric factors. The energies E_i at the fifth step depend on the free energy of the bonds between the current codons and anticodons, the change in energy of the stimulatory structure due to partial unfolding, and the integrated tension in the spacer. These energies are recalculated at each step in the Markov process. The calculation of k_{fs} is a modification of the model put forth by Cao and Chen [5], based on the mechanical model described by Plant et al. [9].

Key results

The model can reproduce frameshift efficiencies and translation times similar to those observed in the laboratory under reasonable assumptions for parameter values. In simulations of up to 1000 runs of the model with a test RNA sequence that includes the HIV-1 slippery site and structure, the frameshift efficiency is 30%, compared to 11% measured *in vitro* at a lower temperature by Jacks et al. [2]. The mean translation rate is 4 codons/sec, which is within the experimentally determined range of 2-20 codons/sec [10]. The longest pauses occur when the ribosome is at the position of the downstream structure, as expected. Furthermore, frameshifts occur at the expected sites and not elsewhere on the mRNA, based only on the calculated probabilities described above.

The model gives similar results for test sequences involving slippery sites and stimulatory structures from the infectious bronchitis (IBV) and sugarcane yellow leaf (ScYLV) viruses. Sequences with no structure do not frameshift, and sequences with a structure but no slippery site frameshift at efficiencies on the order of 1%, allowing us to tease apart the relative contributions of each component of the frameshift signal.

Predictions

The frameshift efficiency varies with both the type and position of the downstream structure and the slippery sequence. This feature allows us to predict the effects of mutations in these regions on frameshift efficiency, which can be tested in the lab. For example, increasing the spacer before the HIV-1 hairpin by 1 base decreases the predicted frameshift efficiency by a factor of 10 when the hairpin is paired with the IBV or ScYLV slippery sites.

The unknown parameters were varied in order to evaluate the sensitivity of the model to those parameters and identify possible biologically relevant parameter regimes. I have identified narrow ranges of values for the spacer length and extension of the mRNA in which a frameshift may occur. The extension depends on the size of the ribosome, which differs in eukaryotic (animal and plant) and bacterial systems. Although many viruses of interest infect eukaryotic cells (including humans),

bacterial ribosomes are commonly used to investigate gene expression in laboratory experiments. The model also predicts a sensitive dependence of frameshift efficiency on the type of tRNA available. This prediction is to be tested at the lab of Dominique Fourmy at CNRS next summer.

The model predicts that the ribosome pauses after a frameshift but before translocation to the next codon, rather than before the frameshift, as has been assumed [11]. The translation time is most strongly correlated with the rate constants for initial binding of the tRNA and for unfolding the downstream structure. Variations in all other model parameters show no significant effect on the model predictions.

Future Work

Within cells, multiple ribosomes translate each mRNA simultaneously. Zouridis and Hatzimanikatis [12], among others, have previously modeled the action of multiple ribosomes. These models investigate the effects of changes in molecular concentrations or reaction rates of individual steps on ribosome spacing and the overall protein synthesis rate. They do not consider ribosomal frameshifting. Ribosome spacing on the mRNA may affect the ability of the downstream structure to refold between ribosomes, thus altering the observed frameshift efficiency. I plan to address this question within the next few months.

The model can also be extended to explore the effects of further details of translation and the intracellular environment. For example, Mg^{2+} ions stabilize RNA structures. It is also likely that the stimulatory structure interacts with the entrance of the ribosome via specific intermolecular forces, thus promoting the deformations and tension needed for frameshifting to occur [13].

A related future project is to investigate other types of recoding events that have been observed in protein synthesis, including forward (+1) frameshifts and hopping of the ribosome over several bases on the mRNA. An approach similar to the one I have used to model −1 PRF could be fruitful in exploring these translation mechanisms as well.

Other research interests

Apart from extending my current work, I am also interested in applying modeling techniques to processes at larger scales. One prior research project investigated the orbital dynamics of a class of small bodies in the outer solar system. The orbits of these objects are perturbed by gravitational encounters with the giant planets, leading to chaotic behavior on timescales of a few thousand years. We found that the dynamics of one sub-class of these objects can be described by a generalized diffusion model. The results of this study were published in the journal *Icarus* in September, 2009 [14]. In another project, I developed an agent-based model of swarm behavior that simulated the movement of herds in two dimensions. I am still fascinated by these topics and intend to revisit them and related questions someday.

Involving Undergraduates in Research

As an undergraduate, I spent one summer doing lab work in support of a field ecology research project and two summers in Research Experience for Undergraduates programs in astronomy. I especially enjoyed the astronomy experiences because I was given the chance to learn about all aspects of the projects: the astronomy, the image processing techniques, the computer skills, and

the goals of the research. Those experiences gave me a far better appreciation for research than simply being given isolated tasks to do in a lab. As a supervisor for undergraduates doing research, I would like to facilitate similarly rich experiences. As I mentioned in my teaching statement, I believe that a successful research experience for undergraduates should help the students understand the time it takes to carry out a project and the steps involved in beginning to explore a new question, as well as learning some specific skills and content related to the project. In my area of mathematics research, the basic ideas of stochastic processes and dynamical systems are well within the grasp of undergraduates. Both biophysics and solar system dynamics are rich sources of modeling problems. Some specific project ideas include the kinetics of ribosome translocation, vaccination patterns and the spread of measles, and orbital dynamics of Earth-crossing asteroids. I would also like to support students in pursuing questions that they themselves find interesting.

References

[1] T. Jacks and H. E. Varmus. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science*, 230(4731):1237–1242, 1985.

[2] T. Jacks, M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331:280–283, 1988.

[3] J. D. Dinman, M. J. Ruiz-Echevarria, and S. W. Peltz. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends in Biotechnology*, 16(4):190–196, 1998.

[4] A. Heyd and D. A. Drew. A mathematical model for elongation of a peptide chain. *Bulletin of Mathematical Biology*, 65(6):1095–1109, 2003.

[5] S. Cao and S.-J. Chen. Predicting ribosomal frameshifting efficiency. *Physical Biology*, 5:1–10, 2008.

[6] I. Tinoco, P. T. X. Li, and C. Bustamante. Determination of thermodynamics and kinetics of RNA reactions by force. *Quarterly Reviews of Biophysics*, 39(4):325–360, 2006.

[7] T. Pape, W. Wintermeyer, and M. V. Rodnina. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. *The EMBO Journal*, 17(24):7490–7497, 1998.

[8] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[9] E. P. Plant, K. L. M. Jacobs, J. W. Harger, A. Meskauskas, J. L. Jacobs, J. L. Baxter, A. N. Petrov, and J. D. Dinman. The 9-Å solution: How mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA*, 9(2):168–174, 2003.

[10] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. New York: Garland Science, 5th edition, 2007.

[11] J. D. Lopinski, J. D. Dinman, and J. A. Bruenn. Kinetics of ribosomal pausing during programmed −1 translational frameshifting. *Molecular and Cellular Biology*, 20(4):1095–1103, 2000.

[12] H. Zouridis and V. Hatzimanikatis. A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. *Biophysical Journal*, 92(3):717–730, 2007.

[13] O. Namy, S. J. Moran, D. I. Stuart, R. J. C. Gilbert, and I. Brierley. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090):244–247, 2006.

[14] B. L. Bailey and R. M. Malhotra. Two dynamical classes of Centaurs. *Icarus*, 203:155–163, 2009.