

# Detailed Stochastic Simulation of Transcription

Xiang Jerry He, Williams College  
Brent Foy, Wright State University

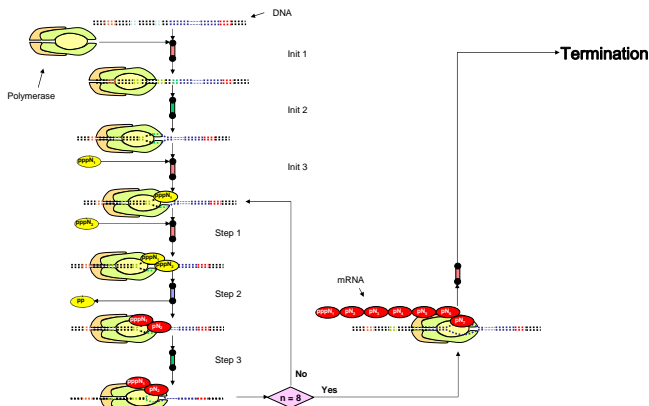


## Introduction:

Biologists have succeeded in giving existing cells new properties, such as inserting a gene into E. Coli to make it produce insulin. But existing cells are complex and changes made to it often have unintended consequences. An alternative approach undertaken by the Cell-like Entity project (CLE) is to create cells from scratch. The hope is to mix the right amount of the right components: DNAs, RNAs, ribosomes, etc into a membrane, and have the material exhibit cell-like behavior. To design a CLE, we need mathematical models to simulate the biochemical processes involved. My part of the whole project is to write a detailed transcription module to the existing Matlab program that models a cell-like entity.

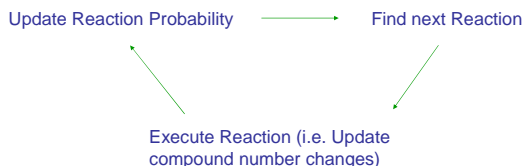
## Transcription and Our Interpretation of it:

Transcription is the process through which a DNA sequence is enzymatically copied by an RNA polymerase to produce a complementary RNA sequence (known as mRNA). It is part of the larger biological process where the genetic code is used to produce proteins. However, transcription is a very complicated process, and since we do not have enough computational power to simulate it in its full physical context, we use the following simplified model.



## Gillespie's method:

Detailed Stochastic Simulation of Transcription (DSST) is meant to be a module to the larger CLE simulation program, which uses Gillespie's algorithm (Gillespie, 1977), a technique for modeling chemical reactions. In this algorithm, each reaction is assigned a probability based on the multiplicity of the reactants and the reaction rate constant. At each iteration, Gillespie's algorithm picks the next reaction according to the probability distribution of all reactions and then "executes" it by changing the compound numbers of the reactants/products. Finally, since some compound numbers have changed, the probability of those reactions use these compounds as reactants will have to be updated before we go to the next iteration. The following diagram illustrates this scheme.



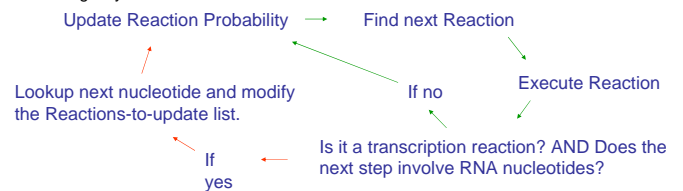
## Method:

The Detailed Stochastic Simulation of Transcription (DSST) is a simulation that does the following

- 1) Keep track of transcription on multiple genes.
- 2) Keep track of transcription on multiple copies of the same gene.
- 3) Keep track of multiple polymerases on the same copy of a gene.

Along with the original CLE simulation, DSST simulates the process of transcription as a series of chemical reactions. However, in transcription, not all reactions can be made available. For example, for the reaction that describes Step 2,  $\text{PolymeraseSite2Empty} + \text{Nucleotide} \rightarrow \text{PolymeraseSite2Full}$ , we need 4 different reactions to accommodate all possibilities where Nucleotide = ATP, CTP, GTP, or UTP. Yet only one of those reactions can happen for a particular gene, so we designed a way to block other three reactions from entering into the Update Reaction Probability process.

The DSST divides transcription into three sections: initialization, stepping, and termination. The initialization consists of the three init reactions, which happens only once for every gene. So the DSST looks ahead into the gene sequence and selects the correct next reaction. As for the stepping process, the DSST intercepts the original system in the following way:



## Results:

I ran the simulation 3 times with a 3-gene system for 600 seconds, and although no two results are identical, all results are somewhat qualitatively similar. One of the benefits of a stochastic algorithm is that it can mimic inherent fluctuations often found in natural systems. We can even measure this fluctuation by running the simulation hundreds of times.

Table 1. Positions of Polymerases on the 3-gene system

Genes	Trial 1	Trial 2	Trial 3
glutamate cysteine ligase	397, 469	513	476
glutamate receptor (1st copy)	264, 521	36	183
glutamate receptor (2nd copy)	445	20, 393	144, 402

## Conclusion and Analysis:

The DSST successfully simulates transcription using a stochastic algorithm, and this simulation will be used at AFIT to facilitate a 2-gene system used for test-tube production of glutathione. The accuracy of our Stochastic model can be improved, however, if the model is made to run in a parallel processing environment in the following way:

- all the compounds are evenly divided amongst CPUs.
- all CPU's run for one time step, and then compound numbers are compared across computers and molecules from "high-density" areas can be allowed to diffuse into "low-density areas".
- after the redistribution through diffusion, the CPU's run for another time step.

This new formulation, which allows reactions to occur simultaneously, is more realistic, and may generate more accurate results.

References: Gillespie, D., T., The Journal of Physical Chemistry, Vol. 81, No.25, 1977

Acknowledgments for support and funding:



National Science Foundation



Physics REU Program



Department of Defense