AFRL-RH-WP-TR-2008-0065

Stochastic Simulation of Biomolecular Reaction Networks using the Biomolecular Network Simulator Software

John Frazier
Applied Biotechnology Branch
Biosciences and Protection Division

Yaroslav Chusak
Biotechnology HPC Sofware Applications Institute
US Army Medical Research and Materiel Command
Wright-Patterson AFB OH 45433-5707

Brent Foy
Department of Physics
Wright State University
Dayton OH 45435

February 2008
Final Report for October-2002 – February 2008

Air Force Research Laboratory
Human Effectiveness Directorate
Biosciences and Protection Division
Applied Biotechnology Branch
Wright-Patterson AFB OH 45433-5707

NOTICE

AFRL-RH-WP-TR-2008-0065

THIS REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

_//SIGNED//_____
LEAMON VIVEROS, Work Unit Manager
Applied Biotechnology Branch

//SIGNED//_____
MARK M. HOFFMAN, Deputy Chief
Biosciences and Protection Division
Human Effectiveness Directorate
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
February 2008	Final	1 Oct 02 – 28 Feb 08

Stochastic Simulation of Biomolecular Reaction Networks Using the Biomolecular Network Simulator Software

In House

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
* John Frazier, ** Yaroslav Chushak, *** Brent Foy

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
D4

5f. WORK UNIT NUMBER
07

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

**Biotechnology HPC Software Applications Institute
U.S. Army Medical Research and Material Command, WPAFB OH 45433-5707
***Department of Physics
Wright State University, Dayton OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command*
Air Force Research Laboratory
Human Effectiveness Directorate
Biosciences and Protection Division
Applied Biotechnology Branch
Wright Patterson Air Force Base OH 45433-5707

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RHPB

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RH-WP-TR-2008-0065

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; Distribution unlimited.

13. SUPPLEMENTARY NOTES
88th ABW/PA cleared 21 May 08, WPAFB-08-3324.

14. ABSTRACT

We developed a software package, the Biomolecular Network Simulator (BNS), to simulate the stochastic behavior of complex biomolecular reaction networks on single and multi-processor computing systems. The software uses either exact or approximate stochastic simulation algorithms for generating Monte Carlo trajectories that describe the time evolution of the behavior of biomolecular reaction networks. This software uses a combination of MATLAB and C-coded functions and can be run on either single processor desk top computers or on multi-processor high performance computing hardware. In the later case, the code is parallelized with the MPI library to allow for multiple simultaneous simulations. The software can be run either in an interactive or in a batch job mode. The graphical user interface of BNS allows users to easily set model and simulation parameters for single or multiple simulation sessions. Furthermore, BNS contains a comprehensive set of data processing tools for post-simulation analysis of the results. The behavior of a single gene *in vitro* transcription-translation reaction network is investigated as an application example.

.

15. SUBJECT TERMS
Biomolecular Network Simulator stochastic behavior multi-processor data processing tools

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Leamon Viveros
U	U	U	SAR	69	19b. TELEPONE NUMBER (*Include area code*)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS

FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK.

INTRODUCTION

All biological processes at the cellular level are the consequence of a series of chemical-physical reactions at the molecular level that occur within the micro-volume of the cell. The collection of molecular species and the reactions among them is referred to here as a biomolecular reaction network. The complete biomolecular reaction network for a cell includes thousands of molecular components and reactions involved in transcription, translation, molecular self-assembly, metabolic reactions, transport and physical movements. Since these reactions occur in an extremely small reaction volume, the number of molecules of any one molecular species that can participate in a given reaction can range from single copies of genes to several hundred molecules of chemicals at the μM concentration to several hundred thousand molecules of chemicals at the mM concentration. As a consequence of the fact that a subset of all the reactions in the system involve low copy numbers of substrate molecules, the behavior of individual instances of the system cannot be modeled accurately using continuous deterministic (C-D) approaches.. Thus, these natural micro-systems should be modeled and simulated using basic theory of discrete stochastic (D-S) chemical kinetics.

With the evolution of systems biology in recent years, interest in modeling and simulating the behavior of engineered genetic circuits in bacterial cells has increased. In addition to living cells, nano-biotechnology researchers are exploring the possibility of developing and using artificial cellular constructs employing natural and engineered biological processes (Ishikawa, et al., 2004; Noireaux and Libchaber, 2004; Noireaux, et al., 2005; Oberholzer, et al., 1995; Pohorille and Deamer, 2002; Yu, et al., 2001). In order to predict the behavior of these constructs, modeling and simulation of their biomolecular reaction networks are needed to enable the design and fabrication of both the constructs themselves and physical devices based on these constructs.

In the past ten years, several software packages have been developed and released to the general public that are focused on simulation and analysis tools for modeling and simulating biological systems (e.g., Adalsteinsson, et al., 2004; Dhar, et al., 2004; Ramsey, et al., 2005; Takahashi, et al., 2004). Each of these software products has its advantages and disadvantages for different modeling needs. We developed a software package – the Biomolecular Network Simulator (BNS) – that is specifically designed to operate on either single or multiple processor hardware. The software allow one to build a model of a synthetic biomolecular reaction network and to investigate its behavior using several different stochastic algorithms. In this paper, we focus on the application of the Biomolecular Network Simulator software to an example model to illustrate the advantages of using multiprocessor computational resources. It should be recognized that many of the features of BNS can be found in other simulation software, but, to our knowledge, the unique combination of features in BNS cannot be found in any other software currently available.

METHODS

Stochastic Simulation Algorithm

The mathematical description of the behavior of stochastic biomolecular reaction

networks is based on Markov process theory (Gillespie, 1992). The system behaves as a multi-variant, discrete state, Markov jump process and is governed by the chemical master equation (CME). The solution of the CME is in fact the mathematically exact description of the behavior of the system. For our purposes, we will consider a biomolecular reaction network consisting of N_S identifiable molecular species, denoted S_i (i = 1, 2, ..., N_S). These molecular species can undergo N_R fundamental chemical reactions r_k (k = 1, 2, ..., N_R) and are confined to a fixed reaction volume, V_R. It is assumed that the system is well-mixed (homogenous) and at constant volume and temperature. Let $\mathbf{s}(t)$ be an N_S dimensional state vector whose elements $s_i(t)$ (i = 1, 2, ..., N_S) are the number of molecules in the system of each molecular species S_i at time t.

The stochastic process that describes the behavior of the biomolecular reaction network is characterized by the state density function $P(s,t)$. This function gives the probability that the system is in state $\mathbf{s}$ at time t, where $\mathbf{s}$ can take on any value in the allowable state space. $P(s;t)$ is the solution of the CME:

$$\frac{dP(s,t)}{dt} = \sum_{k=1}^{N_R} \left[a_k(s-\nu^k) * P(s-\nu^k,t) - a_k(s)P(s,t) \right] \qquad (1)$$

where $a_k(\mathbf{s},t)$ is the propensity of the k^{th} fundamental reaction and $\boldsymbol{\nu}^k$ is the state change vector, a N_S dimensional vector that specifies the changes in the number of molecules of each state variable when the k^{th} reaction occurs. Note, the sum is over all of the N_R possible reactions that can occur. The specification of the initial condition for the biomolecular reaction network of interest, $P_0(s) = P(s, t = 0)$, depends on the precision and accuracy of the measurement techniques used to experimentally characterize the system. In theory, the system is in a single well defined state $\mathbf{s_0}$ at time t_0, where the number of molecules of each molecular species is equal to the exact number of molecules of that species contained in the reaction volume V_R at time t_0. In this case, $P_0(s)$ is defined by the Kronecker delta function as

$$P_0(s) = P(s.t = 0) = \delta(s, s_0) \qquad (2)$$

For our purposes, it will be assumed that the initial condition as defined by Equation (2) will hold and the state density function that is the solution of the CME can be written as the conditional probability density function $P(s,t|s_0,t_0)$.

Usually, an analytical solution of the CME is not possible and direct numerical computation of the solution is computationally overwhelming due to the large state space. However, the direct simulation of exact (theoretically possible) trajectories in state space is feasible (see Appendix A for additional details). The time evolution of the state vector $\mathbf{s}(t)$ for a theoretically possible instance of the system can be calculated using various algorithms proposed for Monte Carlo simulations of stochastic trajectories. The Gillespie direct stochastic algorithm, (Gillespie, 1977) is used in this report to illustrate the stochastic behavior of a simple gene expression system. The Gillespie direct stochastic algorithm theoretically generates exact simulations of system trajectories in state space if and only if all reactions in the biomolecular reaction network are fundamental reactions (Gillespie, 1977). In the limit of an infinite number

of simulations, the statistical properties of the ensemble of exact simulations approaches those of the exact solution of the CME, i.e., for the first moment (mean) of $\boldsymbol{s}$ we have

$$\langle s(t)\rangle = \sum_{s} s * P(s,t \mid s_0,t_0) = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{s_i(t)}{n} = \lim_{n\to\infty} \langle s(t)\rangle_n \tag{3}$$

where $\langle s(t)\rangle_n$ is the estimate of the mean based on an ensemble of n simulations, the left hand sum is over all possible states in state space and the right-hand sum is over all values of the state vector observed in the n simulation runs. In addition, the variance of $\boldsymbol{s}$ is

$$\text{var } s(t) = \sum_{s} (s(t) - \langle s(t)\rangle)^2 * P(s,t \mid s_0,t_0) = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{(s_i(t) - \langle s_i(t)\rangle)^2}{n} = \lim_{n\to\infty} {\sigma_n}^2(t) \tag{4}$$

where $\sigma_n(t)$ is the estimate of the standard deviation based on the ensemble of n simulations.

Although the basic biochemical reactions in a biomolecular reaction network are discrete, jump Markov processes and thus stochastic in nature, if the number of molecules in the system is large then the process can be approximated by a continuous Markov process (Gillespie, 1992). Furthermore, if the number of molecules and the volume increase in proportion such that the concentration of each species is constant (the so-called thermodynamic limit), then the solution describing the behavior of the state variables can be written as the sum of a sure variable that is the solution of the classical rate equations and a variable factor that decreases in magnitude as $1/\sqrt{V_R}$. Thus, for sufficiently large reaction volume, keeping concentrations constant (consequently large number of molecules), the first moment of the probability density function of the state variables approaches the classical continuous deterministic solution of the reaction rate ODEs. However, if there are only a few molecules of any given species, as is often the case in gene expression, this approximation will not accurately describe the instantaneous state of the system. Furthermore, the C-D approach will provide no information concerning the temporal fluctuations of state variables of a given system nor the variability between multiple instantiations of the system with identical initial conditions.

Biomolecular Network Simulator Software

The Biomolecular Network Simulator software was developed to allow for stochastic simulations on either desktop or multi-processor hardware (see Appendix B for additional details on the software or http://www.bioanalysis.org for complete documentation). The front-end graphical users interface (GUI) and the backend data analysis tools are written in MATLAB. This allows the user to exploit the interactive features and visualization tools of MATLAB for setting up simulations and analyzing and interpreting the resulting data. The simulation engine itself is written in the C language to maximize speed for the computationally intensive part of a simulation run.

The BNS software accepts two types of model definitions: (1) Systems Biology Markup

Language (SBML) format (Huska, et al., 2003) and (2) BNS format where models are defined by a set of MATLAB m-files. There are two types of output files: snapshot data and event log data. Snapshot data files contain the state of the system (number of molecules of each molecular species) and the number of reaction occurrences in each reaction channel since the last snapshot at user specified time intervals. The second type of output files – the event log files – contain the record of every discrete event that occurs during the simulation.

Parallelization of the BNS code for simulations runs on high performance computing hardware is accomplished using the Message Passing Interface (MPI). MPI consists of a set of MATLAB scripts that implements a subset of the Message Passing Interface standard and allows MATLAB programs to run on multiprocessor architectures. In our parallelization scheme, the 'master' processor divides the total number of simulation runs into a set of jobs depending on the number of available processors and sends a job to each of the 'worker' processors. The snapshot data from the workers are sent back to the master processor for the interactive graphics but the event log files are saved to the hard drive by the workers. In this approach to parallelization, the power of multiple processors is utilized to run a large number of simulations simultaneously and thus speedup the overall clock time for the batch job.

BNS allows the user to select the appropriate 'Model' and 'Parameters' directories and set the 'Run' mode for each simulation session. If simulations are run in the interactive mode, the current results of the simulation appear on the monitor at specified plotting intervals during a simulation run. Usually, HPC centers allocate limited resources (in terms of the number of processors and running time) for interactive simulations, therefore BNS can be run in 'Batch' mode. In this mode all output data are stored directly on the hard drive for *post hoc* analysis.

The BNS software has a comprehensive set of tools for post-simulation analyses. The most frequently used type of analysis is to plot the number of molecules of a particular molecular species versus time. The number of molecules versus time plots can be created with both types of output files: snapshot data or event log data with the event log data giving an exact description of the behavior of the selected state variable. A time-weighted average analysis provides for the calculations of the average number of molecules of a particular molecular species during a user selected time-interval. The average is weighted according to the amount of time the compound exists in each state during the selected time-interval. The averaging analysis can be performed for a single simulation run or for an ensemble of runs. In the latter case, the between run average (the average of the individual time-weighted average over the ensemble of simulation runs) and standard deviation are plotted.

Complex biomolecular reaction networks that involve gene expression are usually stiff systems, i.e., contain reactions that occur on different time scales; some reactions have a low propensity and occur rarely while other reactions have a high propensity and occur frequently. A unique feature of the BNS software is that the data stored allows the user to perform various event rate analyses on the simulation data to learn more about the basic nature of the system. Event rates (number of reaction events per unit time) in each reaction channel can be calculated for user-selected time-averaging intervals and plotted versus time. These analyses provide important information about the behavior of the system, e.g., relative event rates for important reactions. Furthermore, the event rate data can be used to calculate the rate of energy utilization

in selected reaction channels.

Exemplar model

In order to investigate the simulation of a biomolecular reaction network with BNS, a simple model of a generic self-assembling catalytic ligation reaction in a cell-free bacterial transcription-translation (CFTT) system is explored. The biomolecular reaction network consists of the transcription and translation of a single gene (geneA) to form an active catalytic enzyme (Pro_A) using a commercial gene expression system in an artificial vesicle. The system is assumed to be contained in a spherical liposome the size of a bacterial cell (reaction volume = 5 x 10^{-16} L). The catalytic enzyme is transcribed from a plasmid vector and the expressed protein catalyzes the ligation of substrates Sub_A and Sub_B to form the product Prod_A. The CFTT system contains all of the necessary bacterial components for transcription of a target gene from a plasmid containing the T7 bacteriophage RNA polymerase promoter. In addition, the system contains all the necessary ingredients for successful translation of the mRNA generated by the T7-polymerase into the expressed protein.

To formulate the simplest, yet biochemically reasonable, model of the kinetics of the self-assembly of the examplar biomolecular reaction network, the conceptual system model illustrated in Figure 1 was proposed. This system consists of 45 state variables and 12 reactions (see Supplementary Material for a more complete description of the model). Transcription consists of three reactions (*r1 - r3*) that include association and dissociation of the T7-polymerase (T7_RNAp) and the T7-promoter site for geneA (T7_P) to form the promoter-polymerase complex (T7_RNAp_T7_P) and the subsequent formation of the mRNA (geneA_mRNA). The mRNA can either be degraded by a generic RNase (*r4*) or used as a template for protein synthesis. Translation also consists of three reactions (*r5 - r7*) that include association and dissociation of the small ribosomal unit (Rib_s) and the ribosomal binding site on the geneA_mRNA to form the ribosomal-mRNA complex (Rib_s_geneA_mRNA) and the subsequent formation of the protein product (Pro_A). The protein product (Pro_A) is capable of catalyzing the ligation of Sub_A and Sub_B to form the metabolic product Prod_A via reaction *r8*. All proteins can be competitively degraded by a generic protease (Prot), reactions *r9 - r12*.

Since gene expression reactions involve a single plasmid contained in the micro-volume of the vesicle, the transcription and translation reactions are stochastic in nature. As discussed above, the most accurate way to model the biomolecular reaction system is to use a stochastic approach to solve the CME, with the number of molecules of each molecular species present in the micro-volume as variables. However, the CME for this system cannot be solved explicitly. Here we use the Gillespie direct stochastic simulation algorithm to demonstrate the advantages of using the BNS software to obtain sufficient numbers of probabilistically correct trajectories consistent with the CME through the use of Monte Carlo simulations.

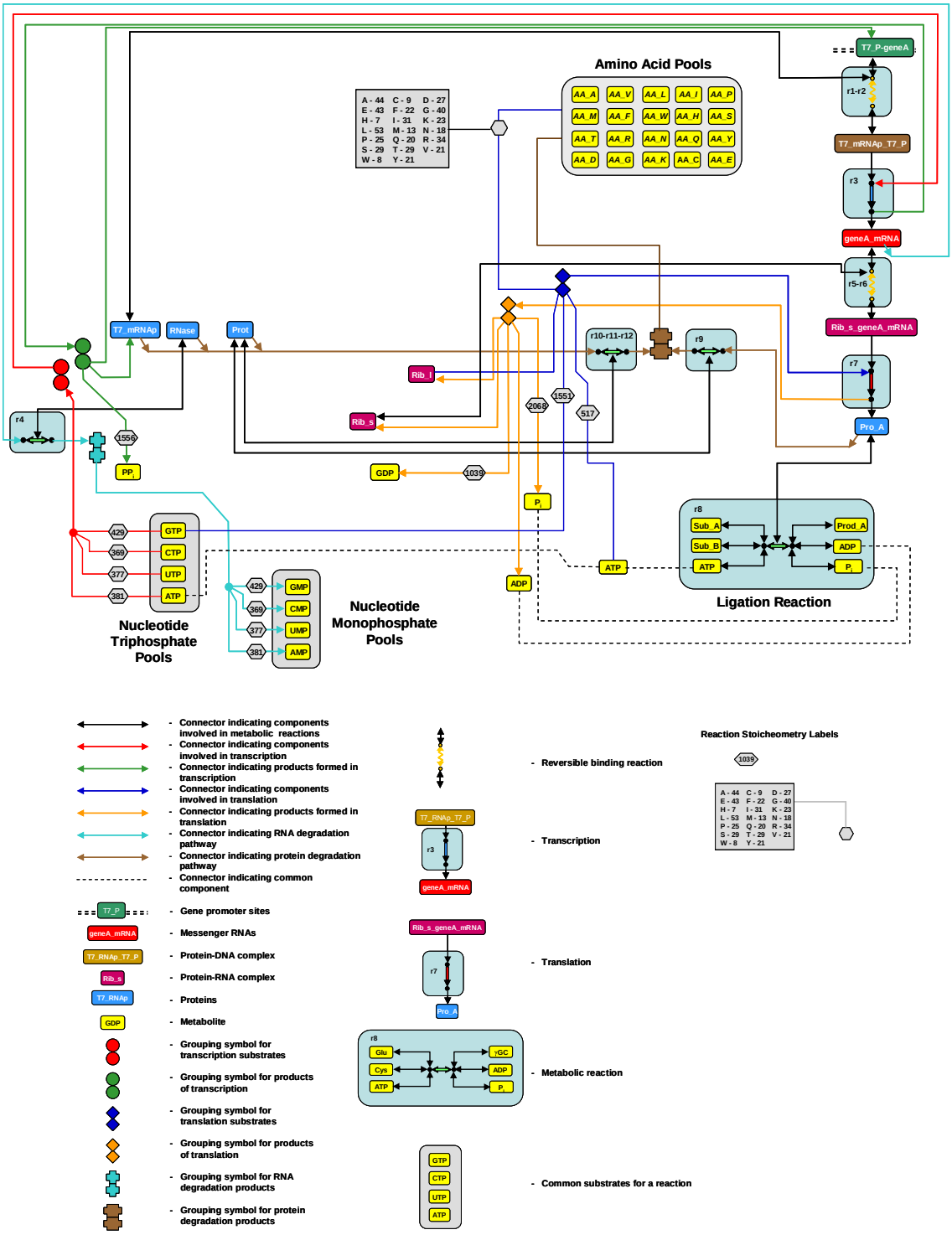


Figure 1: Schematic diagram of a single gene biomolecular reaction network.

RESULTS

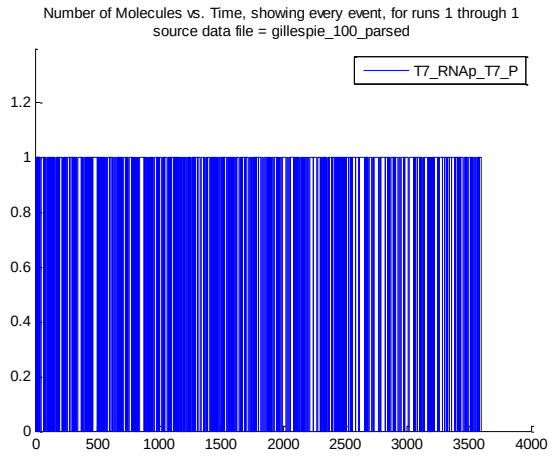
Simulation of exemplar model using the Gillespie Direct Algorithm

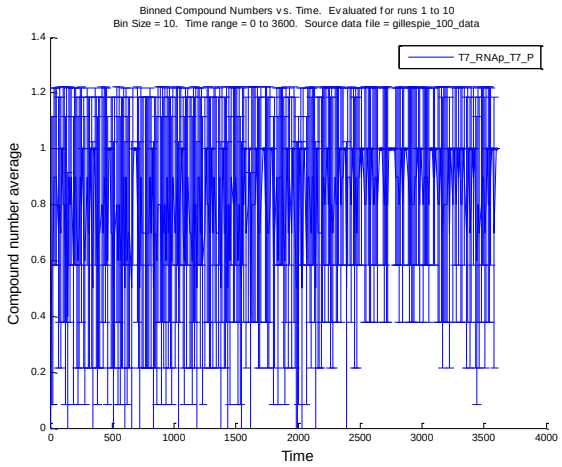
In order to investigate the general behavior of the exemplar model, a series of simulations were run using the following conditions: (1) the Gillespie direct stochastic simulation algorithm, (2) an SBML model definition, (3) the stochastic reaction parameters and initial conditions in Tables C.2 and C.3, respectively, in Appendix C, and (4) the following simulation parameters: duration of simulation = 3600 sec, snapshot interval = 10 sec (giving a total of 360 snapshots), and number of simulations = 10. Due to the scale of the model (45 state variables), it is not possible to show the total set of data for all state variables, but a few selected and important state variables are shown in Figure 2 (remember, these are simulation data for a generic model and do not necessarily represent the behavior of actual state variables and/or reaction rates). The data presented show the trajectory for a single simulation and the estimated mean (first moment) and standard deviation of the state density function $P(\mathbf{s},t \mid \mathbf{s}_0,t_0)$ for each selected state variable. Since the biomolecular reaction system under investigation is a closed system, when critical substrates are depleted, the affected reactions stop. In this particular system, three substrates prove to be critical: (1) AA_Q (glutamine) is depleted at about 1400 sec, (2) GTP is depleted at about 2500 sec, and (3) Sub_A at about 3000 sec. Thus, even though there is adequate geneA_mRNA present, protein synthesis terminates at about 1400 sec when the limiting amino acid, AA_Q, is depleted. Messenger RNA synthesis terminates at approximately 2500 sec when one of the nucleotides, GTP, is depleted. Note, GTP is utilized by both mRNA synthesis and protein synthesis, thus if protein synthesis had not terminated at 1400 sec due to depletion of one of the amino acids, it would have terminated at 2500 sec due to the depletion of GTP. Finally, formation of the metabolic product Prod_A terminates when one of its substrates, Sub_A, is depleted at 3000 sec.

Each simulation run provides a probabilistically accurate trajectory of the system in state space. However, the likelihood that any actual system would follow the simulated trajectory is small. Thus, comparison of an single simulation run with time-series experimental data from a single vesicle is not particularly useful, except in the general sense of trends. The value of individual simulation runs is to provide some intuitive insight into the possible behavior of the system under investigation. For example, Figure 3 shows the state space trajectories for protein Pro_A as generated by 10 individual simulations. In each case, the ultimate level of protein Pro_A is 108 molecules in the vesicle (this is determined by the limiting amino acid AA_Q). However, the time when protein synthesis is completed varies over a significant range, approximately 300 sec, from 1100 to 1400 sec. As a consequence of this stochastic variability, when real-time experimental data from individual vesicles are obtained, the only meaningful comparison is between the experimental data and the simulation ensemble mean ± the standard deviation (right-hand panels in Figure 2). Two thirds of the time, the experimental data should fall within the envelop of the mean ± the standard deviation. However, significant excursions from the envelop can occur even when the model is a correct representation of the experimental system. A better comparison between single vesicle experimental data and model simulations is between the experimental mean ± the standard deviation obtained from multiple (many) single vesicle observations versus the mean ± the standard deviation of an ensemble of a large number of simulations runs (see discussion below on the effect of the number of simulation runs on

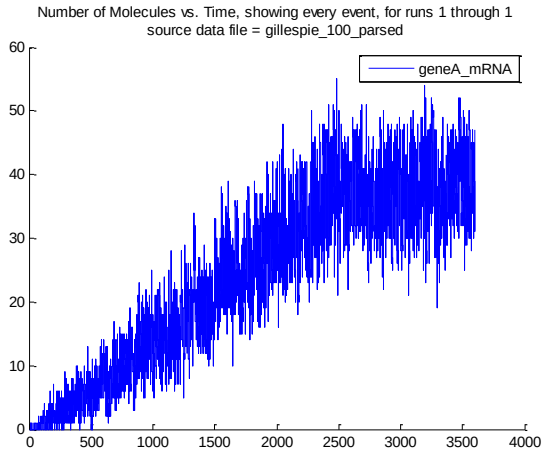
estimates of the mean and standard deviations of the probability density function for the system). If experimental data is only obtained as the mean of a large sample of vesicles, i.e., a grab sample consisting of many vesicles, then the only meaningful comparison is between the macro-sample mean and the mean of a large number of simulations at corresponding time points. In this case, no data concerning the variability between individual vesicles can be obtained. Note, the standard deviation obtained from multiple macro-mean experiments still would not correspond to the fluctuations exhibited in model simulations, but rather would be the result of experimental uncertainties (e.g., experimental measurement errors and non-identical systems), which are not simulated. In fact, if there were no experimental error, then the macro-means of multiple experiments on identical systems would be identical.

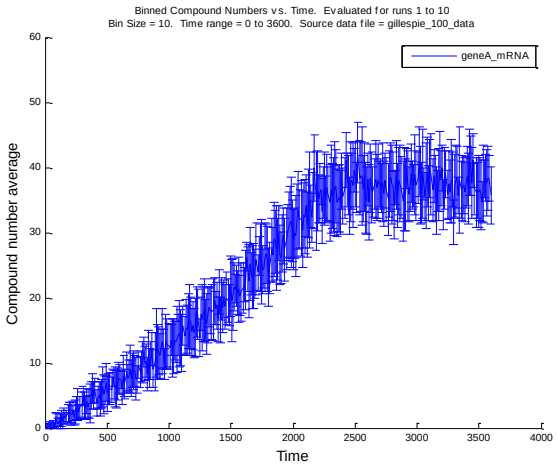
Figure 2: Selected results for simulations of the exemplar model. The left-hand panel is a plot of the number of molecules of the selected state variable versus time for a single simulation run. These plots were obtained from the event log data and include every event that influenced the particular state variable. The right-hand panel is an approximation to the state density function obtained by averaging the number of molecules over 10 simulation runs at selected time intervals





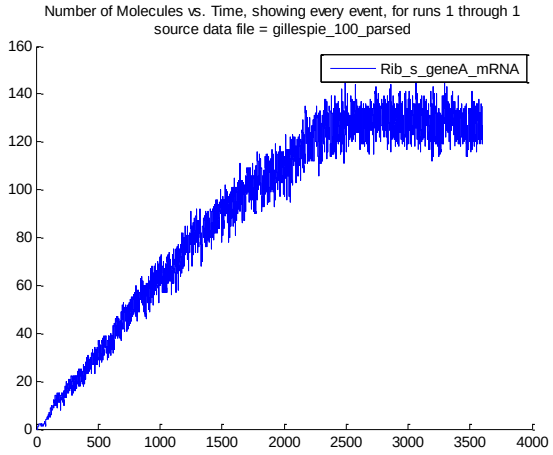
(every 10 sec) using the snapshot data.

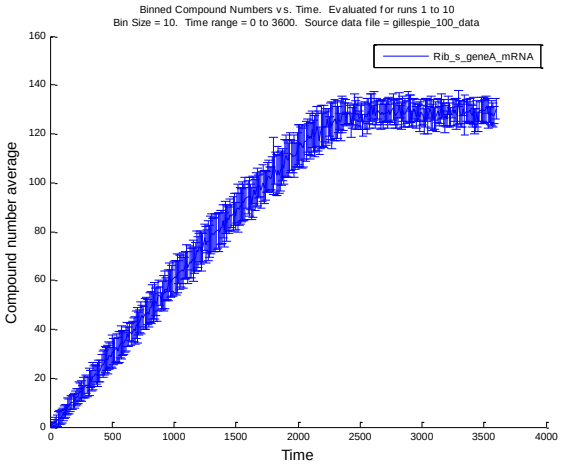




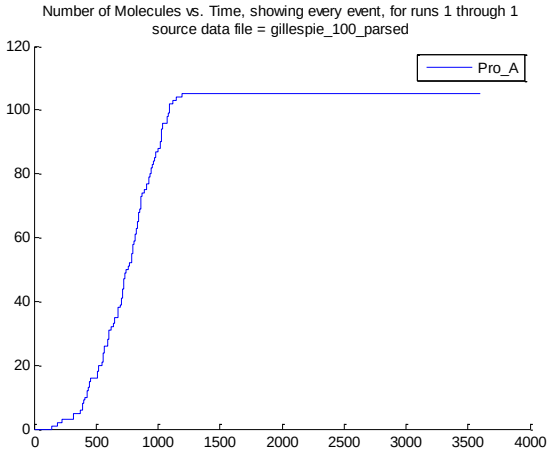
(A) T7_RNAp-T7_P Complex

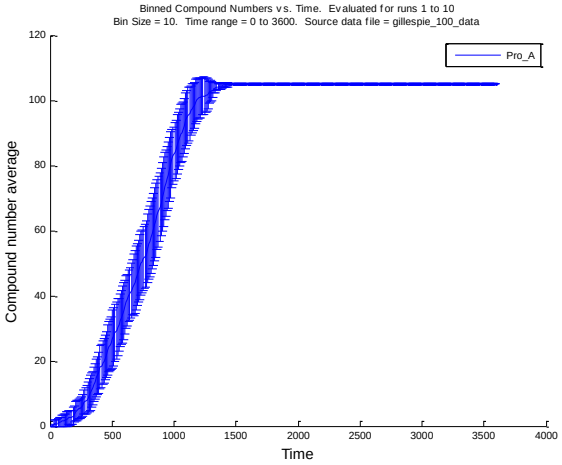
(B) geneA_mRNA



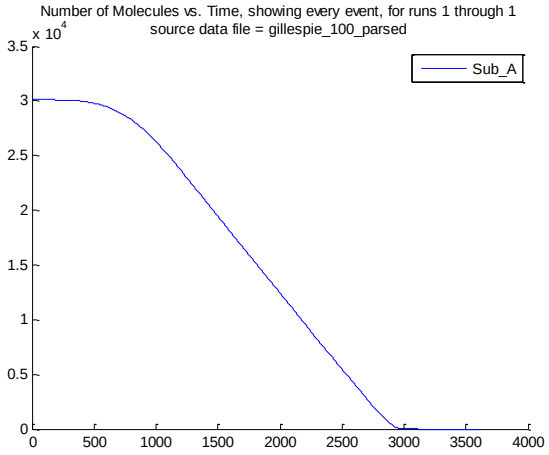


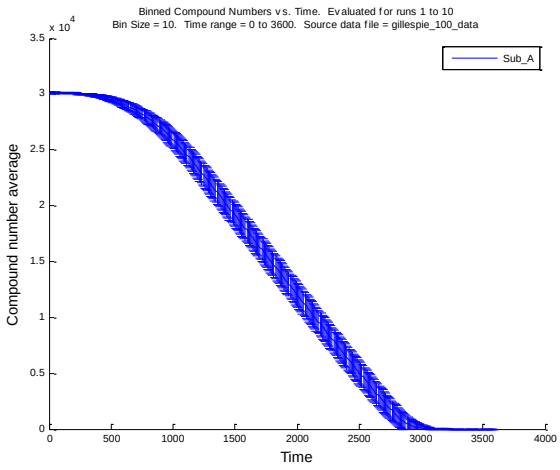
(C) Ribo_s-geneA_mRNA



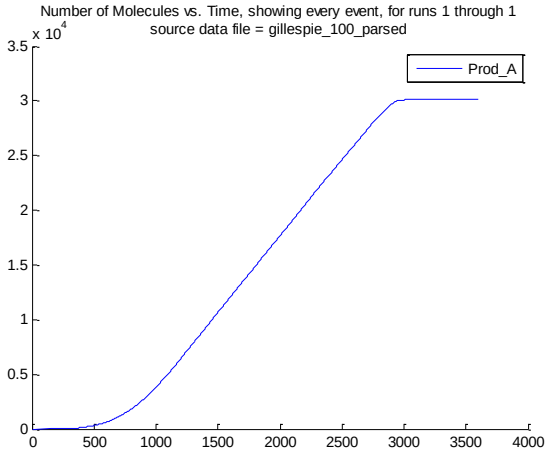


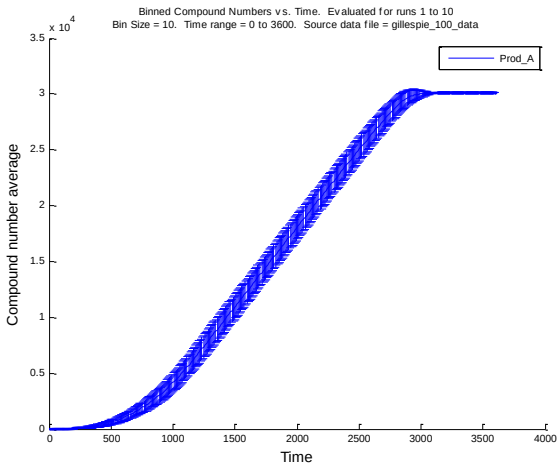
(D) Pro_A (Ligase - geneA expression product)





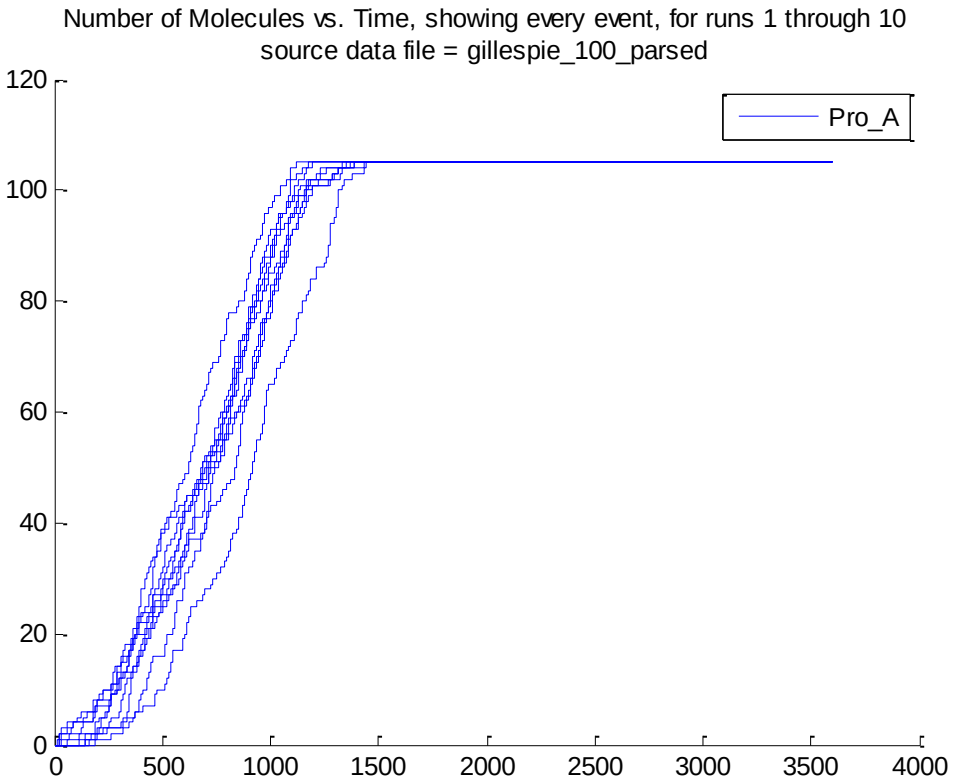
(E) Sub_A





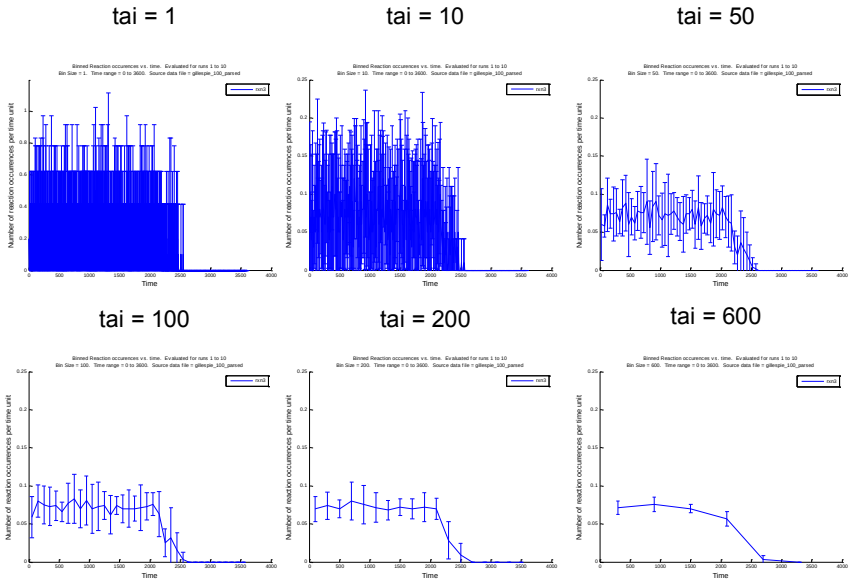
(F) Prod_A

Figure 3: Simulation data for possible trajectories in state space for the number of molecules of protein Pro_A. Ten individual simulations were run and the number of molecules of Pro_A versus time are plotted for each simulation. Event log data were used for these plots, therefore every translation event that produced a molecule of Pro_A is shown for each trajectory.

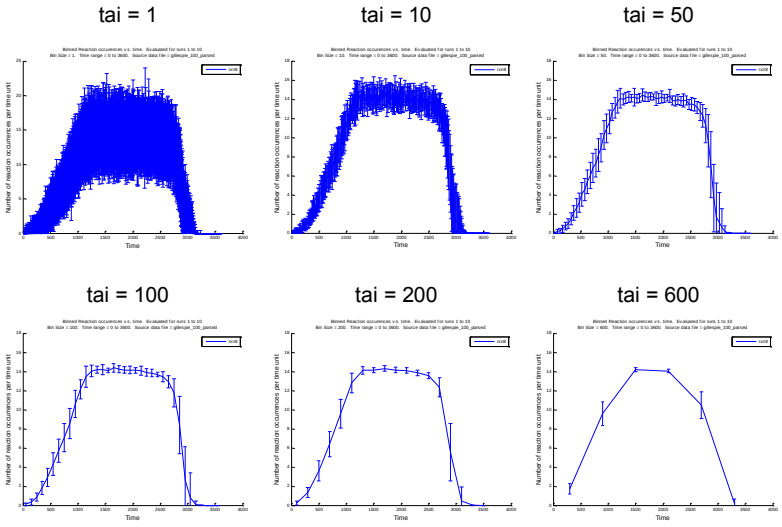


To further investigate the behavior of the system, the event rates of selected reactions were investigated. As a consequence of the system behaving as a discrete jump Markov process, each event occurs instantaneously and the value of associated state variables change discontinuously at the time of the event. As a consequence, there is no derivative of the state variables that would correspond to the C-D concept of rate of change. Hence, for these processes, the 'reaction rate' is defined as the number of events counted during a time-averaging interval (TAI) divided by that time interval, giving an estimate of the event rate (number of events per unit time). These estimates will depend on the TAI as illustrated in Figure 4. A small time-averaging interval results in counting individual events and dividing by a small time interval giving large fluctuations within a individual simulation run and between multiple simulation runs depending on whether a particular time interval contains an event or not. This is obvious in the TAI = 1 sec panel where the between run variability is large. On the other hand, a large time-averaging interval will reduce the variability thus smoothing the data, but will affect the time resolution of dynamical changes in rates due to the averaging over longer intervals. For the results below, a time-averaging interval of 10 sec was selected to maximize time resolution of system dynamics without significant artifacts due to too small a time-averaging interval.

Figure 4: Effect of time-averaging interval (TAI) on estimated reaction event rates. Estimated event rate data was calculated using various TAIs from 1 to 600 sec and averaged over 10 simulations for selected reactions. The mean is the average of the estimated event rate for all 10 simulation runs at the given time interval and the standard deviation reflects the variability between runs. Note the difference in scale between the TAI = 1 sec panel and the other panels.



(A) r3 - transcription

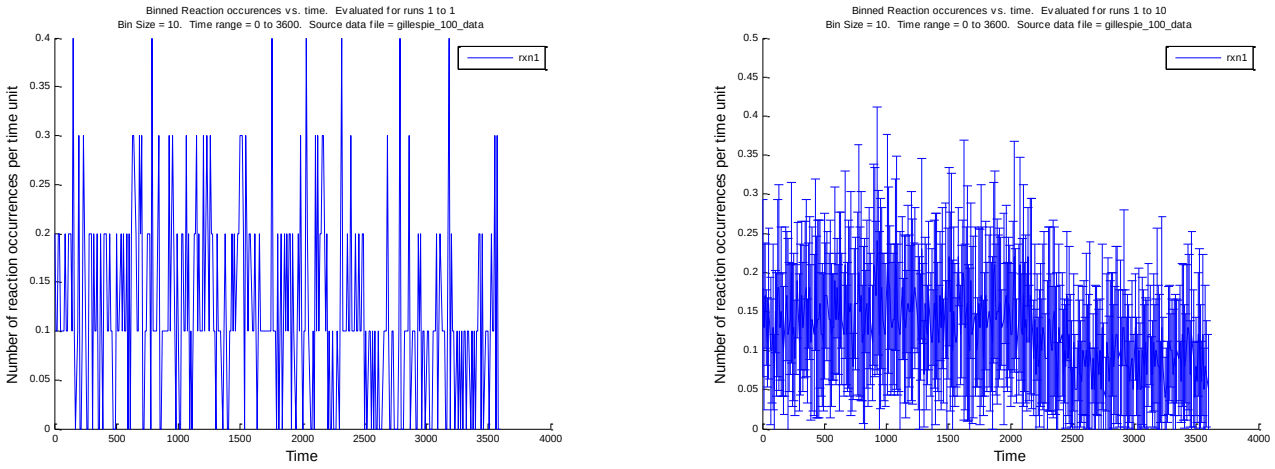


(B) r8 - catalytic ligation

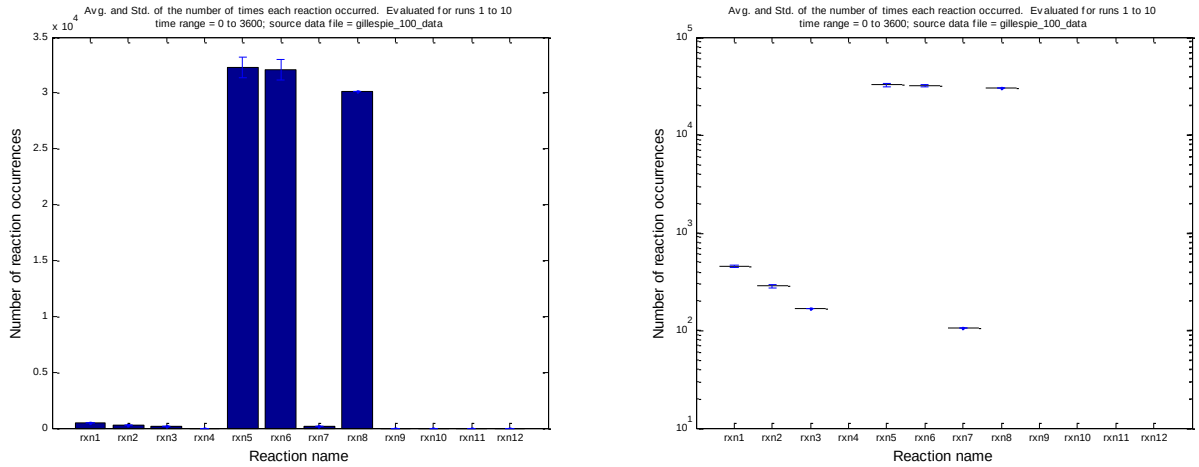
The reaction event rate was computed for selected reactions using a user defined time averaging interval of 10 sec as discussed above and the results are shown in Figure 5. In Figure 5(A) the total number of reaction events in each reaction channel is shown, averaged over the 10 simulation runs. In this examplar model, reactions *r5*, *r6* and *r8* dominate the behavior of the system. Reactions *r5* and *r6* are the association and dissociation of the small ribosomal unit Ribo_s and the ribosomal binding site on gene_A messenger RNA, geneA_mRNA, and reaction *r8* is the catalytic ligation reaction. In figures 5(B) through 5(F), both the time-averaged event rate for a single simulation run (left-hand panel) and the mean ± one standard deviation for the ensemble of 10 simulations (right-hand panel) are shown. The reaction event rates vary during the simulation depending on the availability of substrates (and enzymes where required) and range from 0 - 0.3 reactions per sec for reaction r3 (transcription) to 0 - 18 reactions per sec for reaction r8 (catalytic ligation). Thus, the fastest reaction is about 100 times faster than the slowest reaction. A unique feature of stochastic systems is that the timing of specific events varies from one instance to the next. An example of this effect is seen in Figure 6, where the reaction event rate for reaction *r3* (transcription) is shown for each of the 10 simulation runs. These plots were obtained from the snapshot data with a time-averaging interval of 10 sec. Above each plot the time of the last transcription event is displayed. The transcription reaction terminates when the available GTP is depleted and ranges from 2180 to 2551 sec with a mean and standard deviation of 2337 ± 136 sec. Thus, the timing of any specific event in a stochastic process will always appear as a distribution rather than a fixed time as would be the case for a C-D process. This effect will be addressed further in the discussion of the C-D approximation below.

Figure 5: Time-averaged event rates of selected reactions. (A) Total number of reactions in each reaction channel during simulation (left hand panel is plotted with a linear scale, the right hand panel uses a log scale). (B) - (F) Time-averaged event rates for selected reactions - number of events per sec averaged over 10 sec intervals. Left hand panel shows the averaged rate for a single simulation run. The right hand panel is the mean ± SD for the ensemble of 10 simulation runs.

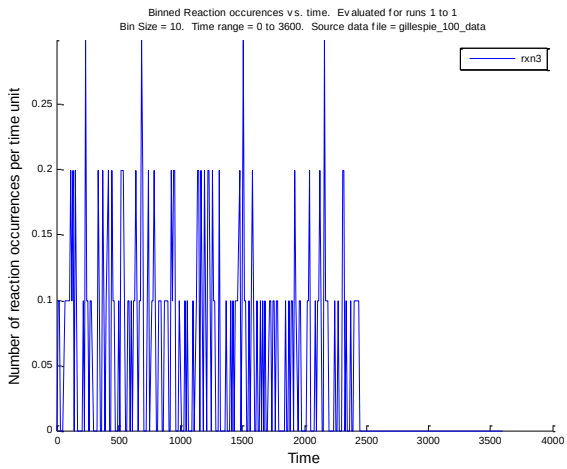
(A)

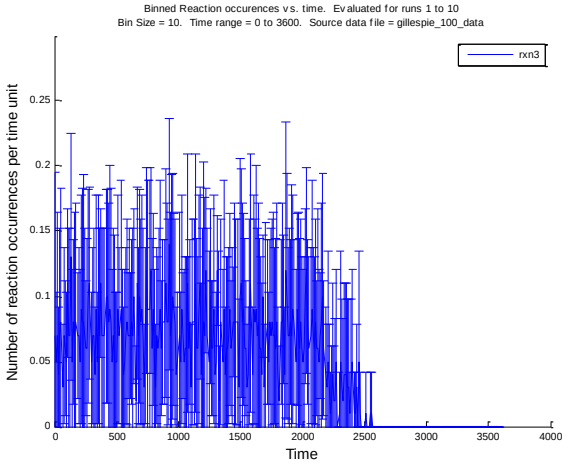


(B) Reaction r1 - Association of T7-RNAp and T7-P on geneA to form the T7_RNAp-T7_P complex

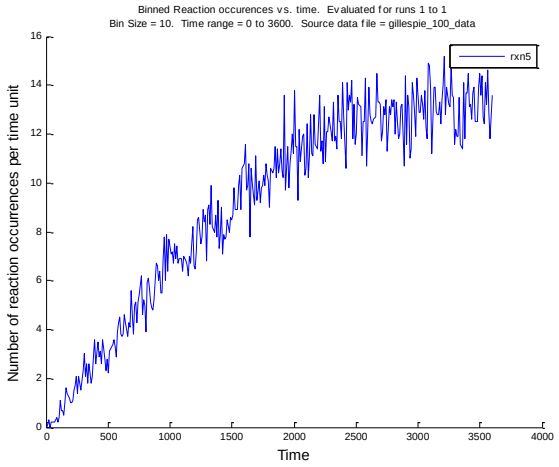


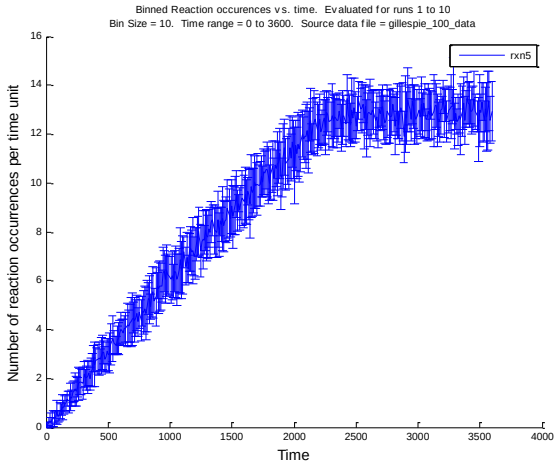
(C) Reaction r3 - Transcription



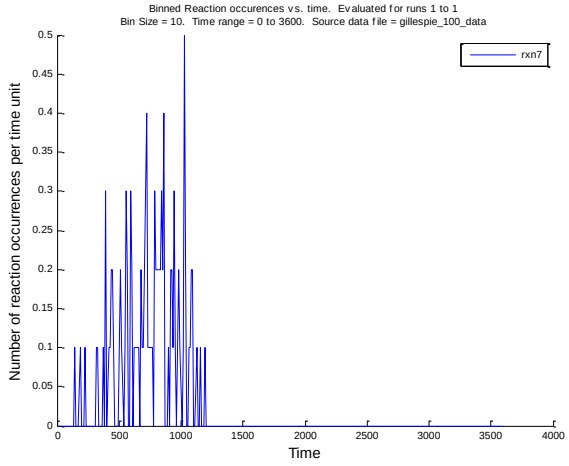


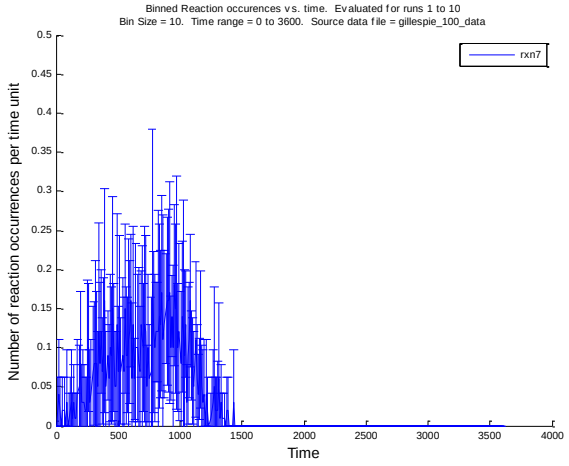
(D) Reaction r5 - Association of Rib_s with geneA_mRNA to form the Rib_s_geneA_MRNA





(E) Reaction r7 - Translation





(F) Reaction r8 - Catalytic Ligation

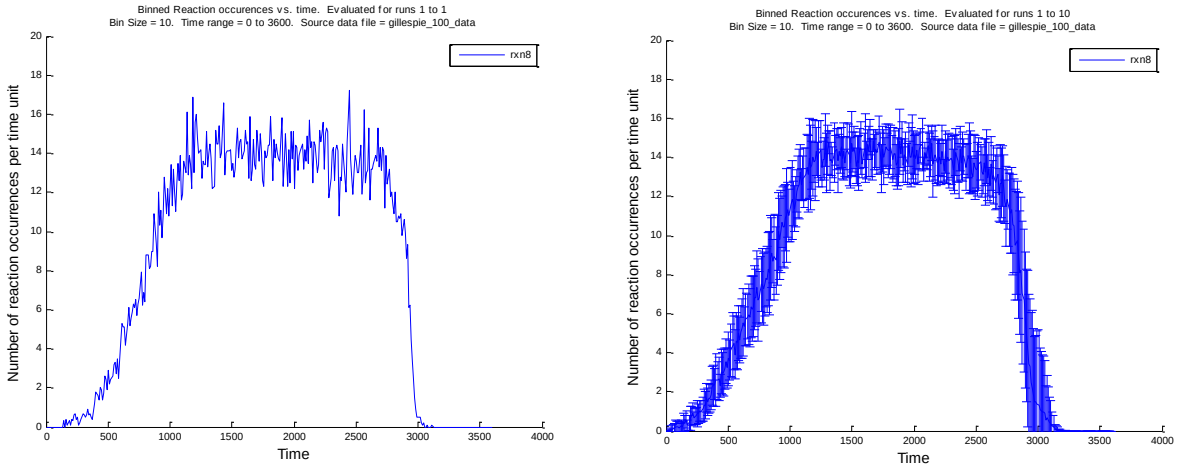
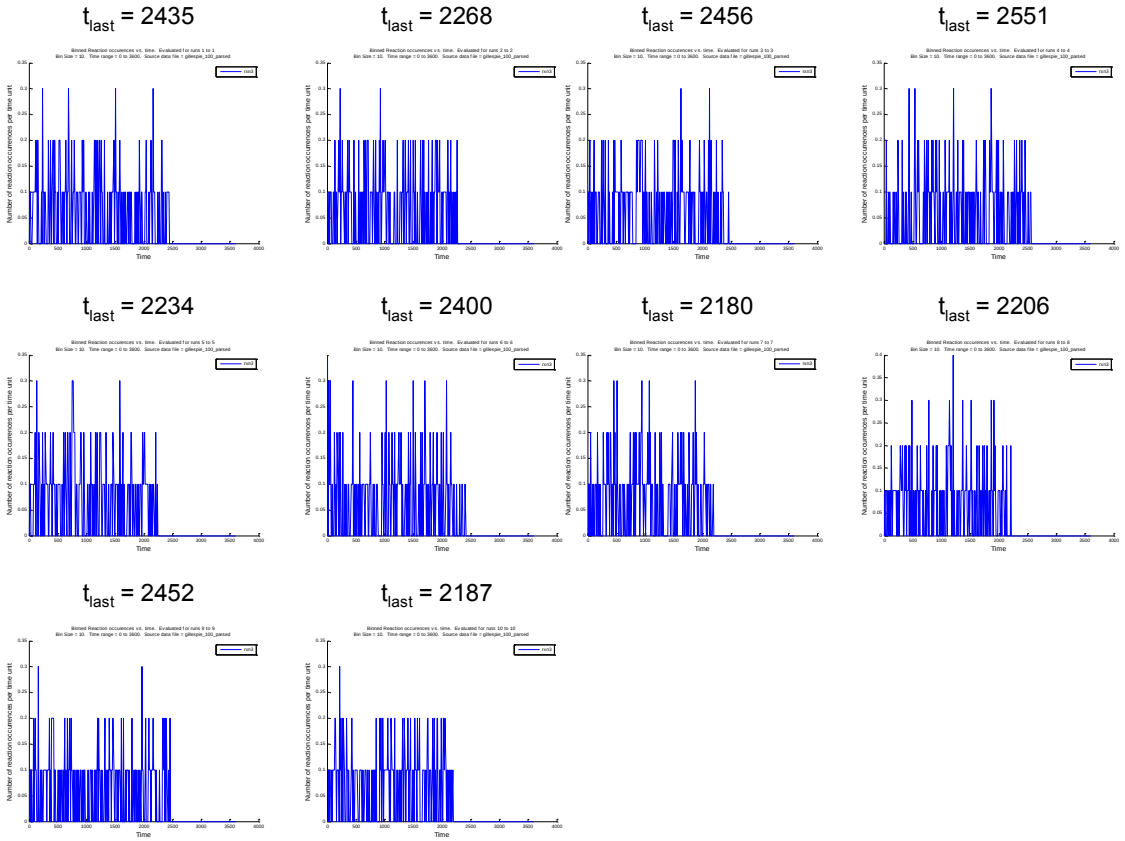


Figure 6: Individual reaction event rate plots for reaction r3 (transcription) for 10 simulation runs. Reaction event rates were calculated with a TAI of 10 sec. The time of the last reaction event is displayed above each plot.

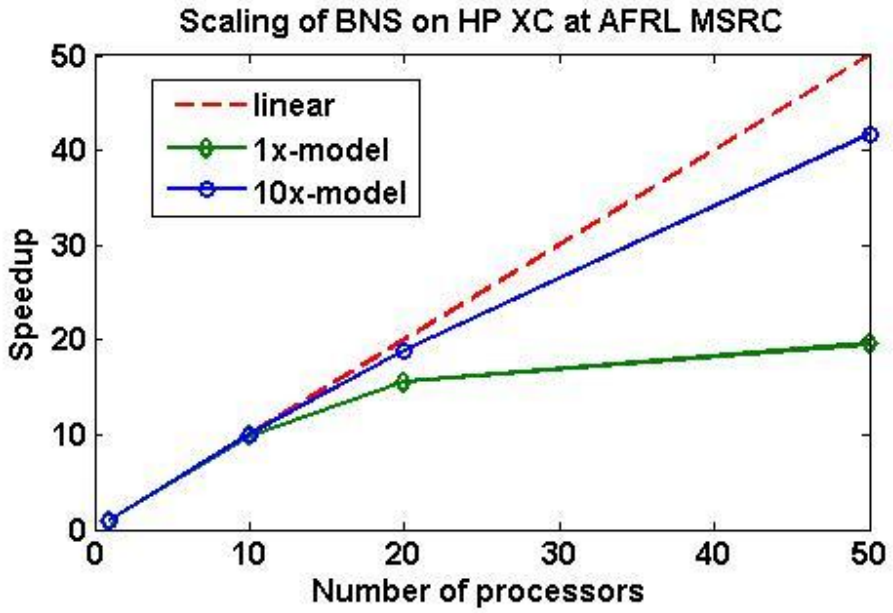


Comparison between single and multi-processor simulation runs

Running a simulation session as a batch job on multi-processor HPC hardware entails a certain amount of overhead, e.g., the time it takes to breakup the job into smaller tasks and assign the problem to each processor on the front end and the collection and data storage on the back end. As a result, the speed-up gained by using multi-processor hardware is to a degree dependent on how computationally intensive the problem is. For a relatively simple problem that is not particularly computationally intensive, the majority of the clock time for the simulation session is taken up with overhead. Whereas, for a problem that is computationally intensive, the computations involved in the actual simulation are the time consuming component of the simulation process. To test this effect, we ran a batch job with the exemplar model using multi-processor HPC hardware to evaluate the speed-up in clock time with increasing numbers of processors. Specifically, we executed 10000 simulation runs of the exemplar model as a batch job on an HP XC machine with distributed memory architecture using the Gillespie direct stochastic simulation algorithm and various numbers of processors (Figure 7). Speed-up was calculated as the clock time it took to run the batch job on a single processor divided by the clock time for the same batch job using multiple processors. As a consequence of the manner in which parallelization using multiple processors was implemented (parallel simulations on multiple processors), full utilization of the BNS software should result in a speed-up proportional to the number of processors used. Up to 10 processors, the speed-up was approximately linear with the number of processors for this computationally simple model. However, the speedup observed by running the model using 20 and 50 processors in the batch mode was only 15.6- and 19.6-fold, respectively. This drop-off in performance is due to the significant role that set-up overhead plays in the total batch run time. For this simple model, the actual computation of the state variable trajectories for each simulation run is very small compared to the time involved in compiling and distributing the model to each processor. Thus, the performance using more than 10 processors results in diminishing returns when the computational demand of the simulation session is small.

To further explore this effect, we repeated the test with a '10x' exemplar model, where initial values of all state variables were increased by a factor of 10. This is equivalent having 10 plasmids containing geneA present in the same reaction volume with ten times the number of substrate molecules available. The speed-up results using the 10x model are also given in Figure 7. For this computationally more complex problem, the value of additional processors is clearly apparent even when 50 processors are accessed. Thus, the value of multi-processor hardware is clearly dependent on the computational dimensions of the problem.

Figure 7: Scaling of simulation run time with the number of processors. Each model was run 10000 times as a batch job using the BNS software on an HP XC machine with distributed memory architecture and the Gillespie direct stochastic simulation algorithm and various numbers of processors. Speed-up was calculated as the run time for the batch job on one processor divided by the run time with the given number of processors.



Improvement in estimating the mean and standard deviation of state variables and reaction rates with the number of simulation runs

The mean and standard deviation of the number of molecules averaged over the ensemble of simulation runs at time t is an estimate of the first moment and variance of the random variable $\boldsymbol{s}$ as defined by the solution of the CME, $P(s,t \mid s_0,t_0)$. As the number of simulations increases, these estimates improve. This can be seen by inspecting the estimated mean ± SD between batch jobs with increasing numbers of simulation runs (Figure 8). The estimated ensemble mean ± SD for the number of molecules of the polymerase-promoter complex (T7_RNAp_T7_P) is shown in Figure 8(A). For this state variable, the possible states in state space are either 0 or 1, thus, the number of molecules of the complex fluctuate over time from 0 $\rightarrow$1 or 1 $\rightarrow$ 0 in any given simulation (Figure 8(A), top left plot). At any given time, the mean over simulation runs fluctuates significantly from one sample time to the next when averaged over a small number of simulation runs - i.e., the mean appears to be noisy when the number of simulations are small (lower panel of Figure 8(A)). However, this is merely a consequence of

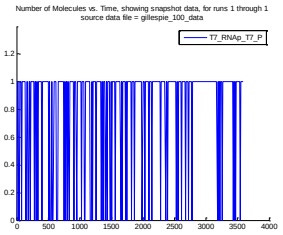
the approximate statistical estimate of the first moment of the solution of the CME using a small number of simulations and the standard error of the mean will decrease with increasing *n* as $SD/\sqrt{n}$ (where SD is the standard deviation of the ensemble distribution). In fact, the exact mean, $\langle s(t) \rangle$, is a smooth function of time as the series of approximations with increasing *n* in the lower panel of Figure 8(A) suggests. Only for estimations of the mean with $n \geq 100$ runs does the shift in the mean at approximately 2300 sec become well defined. This shift is due to the cessation of mRNA synthesis. Another point to note from the top panel of Figure 8(A) is that the estimates of the SD of the ensemble, $\langle \sigma(t) \rangle_n$, also fluctuate significantly from one time point to the next when *n* is small, but tends to smooth out with increasing *n* as the estimates of the SD improve.

Figure 8(B) shows the behavior of geneA_mRNA as *n* increases. Here, the estimates of the ensemble mean and SD again shows significant fluctuations from one time point to the next when *n* is small due to the inaccuracies in each estimate of $\langle s(t) \rangle_n$ and $\langle \sigma(t) \rangle_n$. As *n* increases, each individual estimate of the mean of ***s(t)*** improves and the plot approaches the exact smooth curve for $\langle s(t) \rangle$. Also, the estimates of the SD also improve with increasing *n* and the SDs from one time to the next smooth out. The dependency of the accuracy of the estimates of the mean and SD on the number of simulations is an issue that must be taken into consideration when dealing with stochastic simulations; model predictions of experimental observations will only be exact in the limit of $n \to \infty$ simulations. Thus, it is necessary to use a large number of simulations to investigate the behavior of the of the system if one wished to fit model predictions to experimental data. The larger the number of simulations the better the estimate of the model prediction, thus reducing an additional source of error that is not present when fitting solutions of C-D ODEs to experimental data.

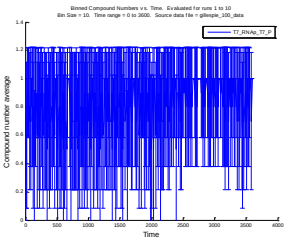
Figure 8: Comparison of estimates of the mean and standard deviation of selected state variables with increasing numbers of simulation runs. For each state variable, the top panel is the mean ± SD for various numbers of simulations plotted at 10 sec intervals and the bottom panel is only the mean. In each lower panel, the solution of the C-D ODE solution is also given. (A) the T7_RNAp-T7_P complex and (B) geneA_mRNA.

(A) T7_RNAp-T7_P

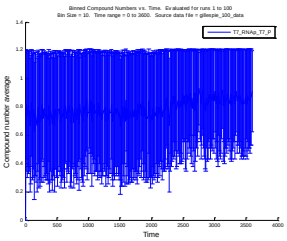
n = 1 run



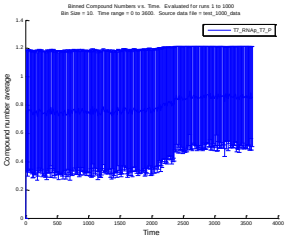
n = 10 runs



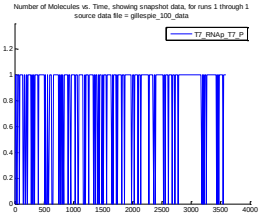
n = 100 runs



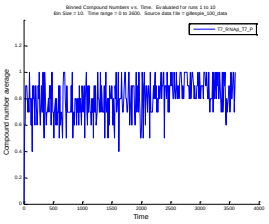
n = 1000 runs



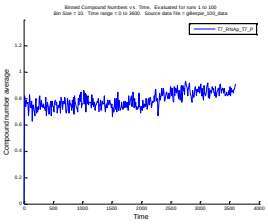
n = 1



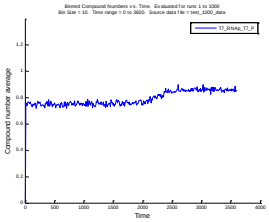
n = 10



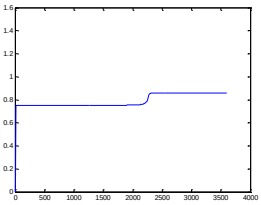
n = 100



n = 1000

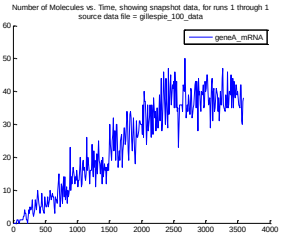


C-D

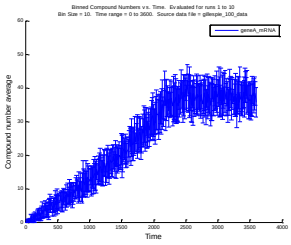


(B) geneA_mRNA

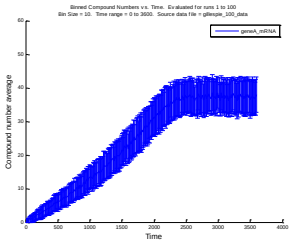
n = 1 runs



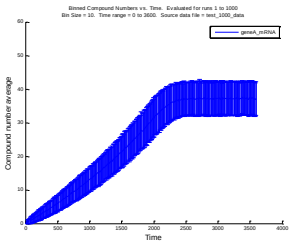
n = 10 runs



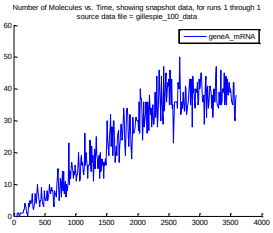
n = 100 runs



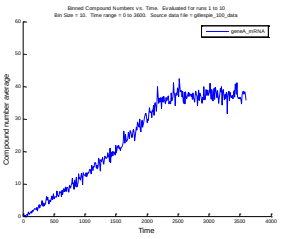
n = 1000 runs



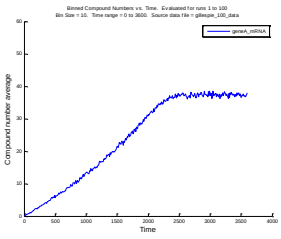
n = 1



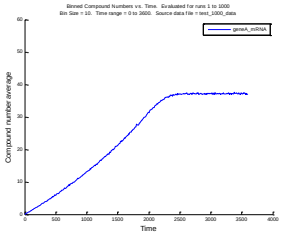
n = 10



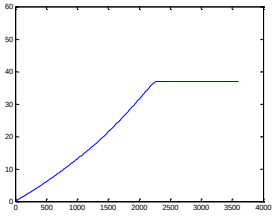
n = 100



n = 1000



C-D



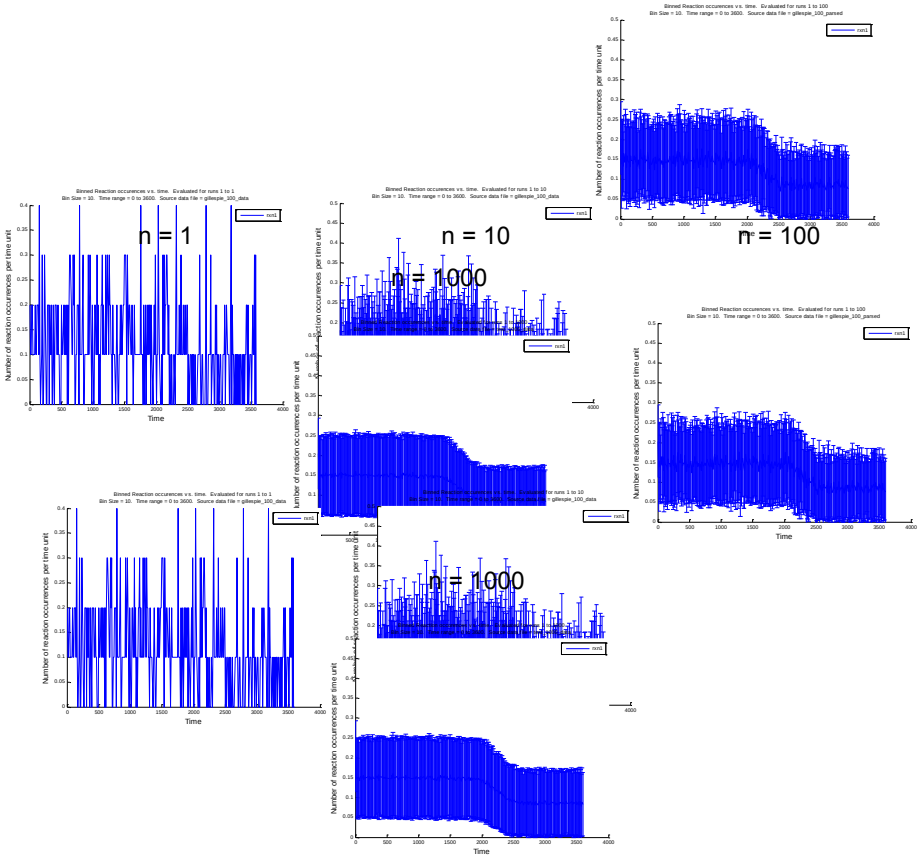
Similar issues arise when investigating reaction event rates. Estimates of the event rate using a time-averaging interval of 10 sec for reactions *r1* and *r8* are given in Figure 9. For reaction *r1* (the association of the polymerase, T7_RNAp, with the promoter for geneA, T7_P, to form the T7_RNAp-T7_P complex), a single simulation, $n = 1$, indicates that the reaction occurred anywhere from 0 to 4 times in any 10 sec counting intervals (corresponding to event rates of 0 - 0.4 events/sec) with large fluctuations from one time point to the next. If multiple simulations are run, the estimated event rate can be averaged over the ensemble of simulations. As can be seen from Figure 9(A), averaging over multiple runs gives a more consistent estimate of the mean and SD of the event rate as a function to time. Even for a reaction that occurs at a significantly greater rate than reaction *r1*, e.g., reaction *r8* (Figure 9(B)), the effect of averaging over multiple simulations is still apparent.

Comparison between exact simulations and the C-D approximation

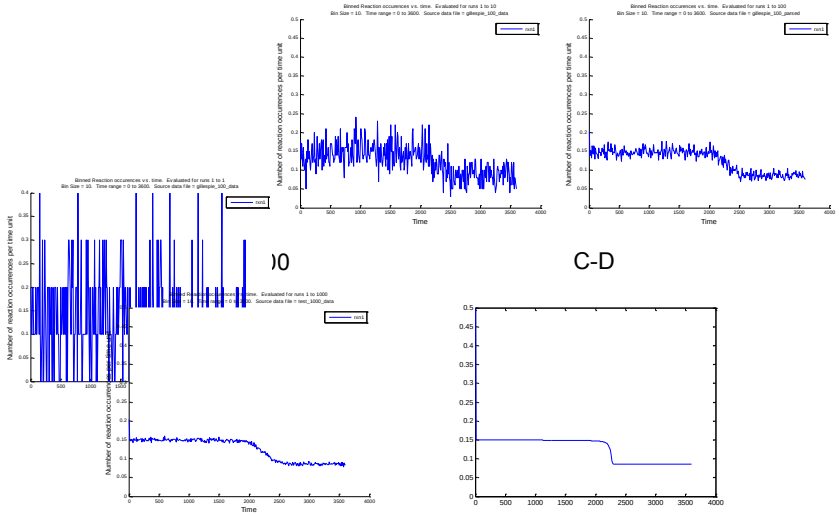
Although the basic biochemical reactions in a biomolecular reaction network are stochastic in nature, the fact that some of the molecular species in the system are present in relatively large numbers should allow for the approximation of the first moment of the state variables by the continuous deterministic approach. To investigate this possibility, the exemplar model was simulated using the C-D ODE approach (see Supplementary Material for reaction parameters). The results are shown in Figure 8 for the selected state variables and in Figure 9 for the selected reaction event rates. As is evident, with one exception, the C-D approximation gives a reasonable representation of the ensemble average for the state variables for this particular model. The one noticeable difference between the two approaches can be seen in the regions where there is a transition in the dynamics due to the termination of certain reactions. In these regions, the ensemble means of the S-D simulations tend to have smooth transitions whereas the C-D simulation has a sharper discontinuity. This effect is due to the variability in the timing of the transition in the S-D approach as discussed above. Each individual S-D simulation has a rather sharp transition when these reaction terminate, but because the time of the transition varies from simulation to simulation, the ensemble mean has a smooth transition. An additional limitation of the C-D approximation is that no information on the variability in the number of molecules of state variables in individual instances can be obtained from this approach.

Figure 9: Comparison of estimates of the time-averaged reaction event rates with increasing numbers of simulation runs. The time-averaged event rates (time-averaging interval = 10 sec) averaged over *n* simulation runs are plotted for: (A) reaction *r1* - association of T7_RNAp and T7_P to form the T7_RNAp-T7_P complex, and (B) reaction *r8* - the catalytic ligation reaction.

(A) Reaction *r1*

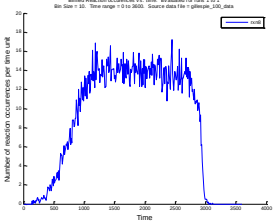


n = 1 *n* = 10 *n* = 100

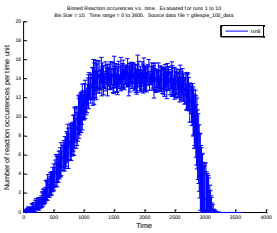


(B) Reaction r8

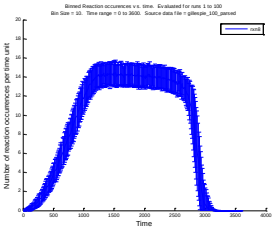
n = 1



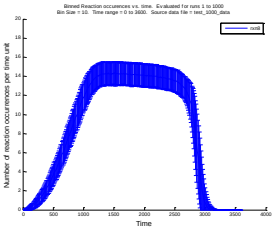
n = 10



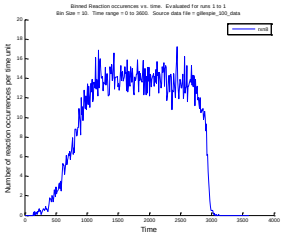
n = 100



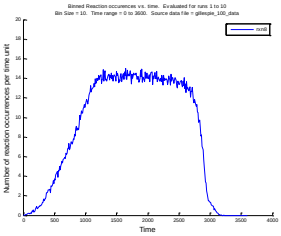
n = 1000



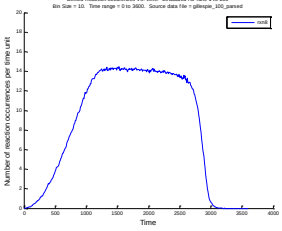
n = 1



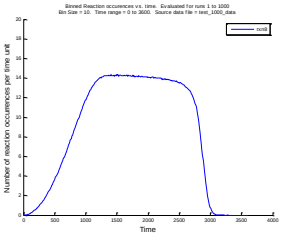
n = 10



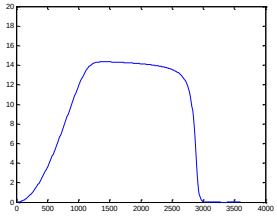
n = 100



n = 1000



C-D



Discussion

The Biomolecular Network Simulator (BNS) software developed in this laboratory allows users to simulate the stochastic behavior of complex biological reaction networks with the flexibility of utilizing either single or multi-processor hardware. Some of the more important features of the BNS software are:

- the usage of MATLAB and C-coded functions allowing the user to combine intensive visualization of data with high speed computations;
- the implementation of parallelized code for multiple simultaneous simulations allowing the user to run BNS on multi-processor machines;
- the availability of the option to run the code in the interactive or batch mode;
- a user friendly graphical user interface allowing the user to easily select and modify models, model parameters, simulation parameters and analysis tools; and
- a comprehensive set of analysis tools providing for post-simulation analysis and plotting of results.

These features allow the user to perform a wide range of simulation tasks in a relatively transparent environment.

The BNS software was used to investigate the generic behavior of gene expression in a cell-free transcription-translation system in a closed system consisting of an artificial vesicle. The ability to simulate this system allows one to identify the critical factors that limit the behavior of the system. Although the model for the system is relatively crude, it is clear that the availability of limiting amino acids controls the ultimate expression of proteins, the availability of GTP limits transcription of the plasmid gene to form mRNA and the availability of substrates for the catalytic ligation reaction limits the generation of the product. This quantitative knowledge can be used to attempt to optimize the system to maximize production of products, either proteins or metabolites. As the models for the CFTT-vesicle system become more sophisticated, a more detailed understanding of its behavior will evolve.

ACKNOWLEDGMENT

The work of Yaroslav Chushak was made possible by a grant from the Department of Defense High Performance Computing Modernization Program Office (HPCMPO). The work of Brent Foy was made possible by a grant from the Air Force Office of Scientific Research (AFOSR) and by the Air Force Summer Faculty Fellowship Program.

References

Adalsteinsson, D., McMillen, D. and Elston, T.C., 2004, Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMCBioinform.* **5**: 24.

Dhar, P., Meng, T.C., Somani, S., Ye, L., Sairam, S., et al., 2004, Cellware - a multi-algorithmic software for computational systems biology. *Bioinform.* **20**: 1319-1321.

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S., 2002, Stochastic gene expression in a single cell." *Science* **297**: 1183 – 1186.

Gillespie, D., 1977, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem*. **81**: 2340-2361.

Gillespie, D., 1992, *Markov Processes: An Introduction for Physical Scientist*. Academic Press. pp. 565.

Gillespie, D., 2001, Approximate accelerated stochastic simulations of chemically reacting systems. *J. Chem. Phys*. **115**: 1716.

Hucka, M., Finney, A., Sauro H. M., Bolouri, H., et al., 2003, The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinfor.* **19**: 524-531.

Ishikawa, K., Sato, K., Shima, Y., Urabe, I., and Yomo, T., 2004, Expression of a cascading genetic network within liposomes. *FEBS Letters* **576**: 387–390.

Kaern, M., Elston, T. C., Blake, W. J. and Collins, J. J., 2005, Stochasticity in gene expressions: From theories to phenotypes. *Nat. Rev. Genetics* **6**: 451-464.

Noireaux, V. and Libchaber, A., 2004, A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. National Acad. Sci.USA* **101**: 17669–17674.

Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H. and Libchaber, A., 2005, Toward an artificial cell based on gene expression in vesicles. *Phys. Bio.* **2**: 1-8.

Oberholzer, T., Albrizio, M. and Luisi, P. L., 1995, Polymerase chain reaction in liposomes. *Chem. Bio.* **2**: 677-682.

Pohorille, A. and Deamer, D., 2002. Artificial cells: prospects for biotechnology. *TRENDS in Biotech.* **20**: 123-128.

Ramsey, S., Orrell, D. and Bolouri, H., 2005, Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol*. **3**: 415-436.

Takahashi, K., Kaizu, K., Hu, B. and Tomita, M., 2004, A multi-algorithm, multi-scale method for cell simulation. *Bioinform.* **20**: 538-546.

Voit, E. O., 2000. *Computational Analysis of Biochemical Systems*. Cambridge University Press, Cambridge, UK.

Yu, W., Sato, K., Wakabayashi, M., Nakaishi,T., Ko-Mitamura, E.P., Shima, Y., Urabe, I. and Yomo, T., 2001, Synthesis of functional protein in liposome. *J. Biosci. Bioeng.* **92**: 590-593.

Appendix A - Stochastic Simulation Algorithm

The mathematical description of the behavior of stochastic biomolecular reaction systems is based on Markov process theory. The system behaves as a multi-variant, discrete state, Markov jump process and is governed by the chemical master equation (CME). The solution of the CME is in fact the mathematically exact description of the behavior of the system. For our purposes, we will consider a biomolecular reaction network consisting of N_S identifiable molecular species, denoted S_i (i = 1, 2, ..., N_S). These molecular species can undergo N_R fundamental chemical reactions R_k (k = 1, 2, ..., N_R) and are confined to a fixed reaction volume, V_R. It is assumed that the system is well-mixed (homogenous) and at constant volume and temperature. Let $\mathbf{s}(t)$ be an N_S dimensional state vector whose elements $s_i(t)$ (i = 1, 2, ..., N_S) are the number of molecules of each molecular species S_i at time t.

The probability that the k^{th} fundamental reaction R_k will occur in the next time interval dt, $prob\ R_k(t)$, is physically defined by molecular collision theory in solution and the quantum mechanics of molecular interactions. It has been shown (Gillespie, 1977) that a reasonable mathematical formulation of this probability is defined by:

$$prob\ R_k(t) = a_k(s,t) * dt + o(dt) \tag{A1}$$

where $a_k(s,t)$ is the propensity of the k^{th} fundamental reaction (see Table A1). Furthermore, for biomolecular reaction systems at constant temperature and volume, the system is temporally homogenous which means that $a_k(s,t)$ is not explicitly dependent on time.

$$a_k(s,t) = a_k(s) \tag{A2}$$

However, $a_k(\mathbf{s})$ is implicitly dependent on time through its dependency on $\mathbf{s}(t)$. he relationship in Equation (A1) is exactly correct to the order of dt.[1]

The N_R by N_S dimensional stoichiometry matrix $\boldsymbol{\nu}$ provides the necessary information as to the change in the number of each molecular species when a particular reaction occurs. The k^{th} row of the stoichiometry matrix, $\boldsymbol{\nu}^k$, is a N_S dimensional vector, referred to as the state change vector, that specifies the changes in the state variables when the k^{th} reaction occurs.

The stochastic process that describes the behavior of the biomolecular reaction network is characterized by the state density function $P(s,t)$. This function gives the probability that the system is in state $\mathbf{s}$ at time t, where $\mathbf{s}$ can take on any value in the allowable state space. $P(s;t)$ is the solution of the CME:

[1] Order of dt, indicated by $o(dt)$, means $\lim_{dt\to 0} \frac{o(dt)}{dt} = 0$.

$$\frac{dP(s,t)}{dt} = \sum_{k=1}^{N_R} \left[a_k(s-\nu^k) * P(s-\nu^k,t) - a_k(s)P(s,t) \right] \quad (3)$$

Note, the sum is over all of the N_R possible reactions that can occur. The first term when summed over all possible reactions give the increase in probability due to transitions into state ***s*** from states one reaction away, and the second term give the decrease in probability due to reactions that originate in state ***s*** and result in transitions to new states one reaction away.

The initial condition for the biomolecular reaction network of interest, $P_0(s)$, depends on the precision and accuracy of the measurement techniques used to experimentally characterize the system. In theory, the system is in a single well defined state $\boldsymbol{s_0}$ at time t_0, where the number of molecules of each molecular species is equal to the exact number of molecules of that species contained in the reaction volume V_R at time t_0. Thus, $P_0(s)$ can be defined by the Kronecker delta function as

$$P_0(s) = \delta(s, s_0) \quad (A4)$$

For our purposes, it will be assumed that the initial condition as defined by Equation (A4) will hold and the state density function can be written as the conditional probability $P(s,t|s_0,t_0)$.

An analytical solution of the CME is not possible usually and direct numerical computation of the solution is computationally overwhelming due to the large state space (the dimension of the state space is $d = \prod_{i=1}^{N_S} (s_i^{Max} + 1)$ where s_i^{Max} is the maximum number of molecules of s_i that can theoretically exist in the system). However, the direct simulation of exact (theoretically possible) trajectories in state space is feasible. For a multi-variant, discrete state, Markov jump process, it is possible to define the *next-jump density function* $p(\tau,k|s,t)$ such that: $p(\tau,k|s,t) * d\tau$ is the probability that the next jump will occur between $t + \tau$ and $t + \tau + d\tau$ and will involve reaction R_k, i.e., the event will change the state of the system to $\boldsymbol{s} + \boldsymbol{\nu}^k$ given that the system was in state ***s*** at time *t*. It can be shown for a temporally homogenous, multi-variant, discrete state, Markov jump process that

$$p(\tau,k|s,t) = p_1(\tau|s,t) * p_2(k|s,t) \quad (A5a)$$

where

$$p_1(\tau|s,t) = \sum_{k=1}^{N_R} p(\tau,k|s,t) = a_0(s) * e^{-a_0(s)*\tau} \quad (A5b)$$

$$p_2(k|s,t) = \int_0^\infty p(\tau, k|s,t)\, d\tau = \frac{a_k(s)}{a_0(s)} \tag{A5c}$$

and

$$a_0 = \sum_{i=1}^{N_R} a_i \tag{A5d}$$

$p_1(\tau|s,t)$ is the conditional density function for the waiting time to the next reaction with the mean waiting time of $1/a_0(\mathbf{s})$. $p_2(k|s,t)$ is the conditional density function that the next reaction will be the k^{th} reaction and is equal to the ratio of the propensity for the k^{th} reaction to the total propensity for all reactions in the system.

The evolution of the state vector $\mathbf{s}(t)$ can be calculated using various algorithms proposed for Monte Carlo generation of stochastic trajectories. The Gillespie direct simulation algorithm, (Gillespie, 1077; Table A1), which is one of the algorithms implemented in the Biomolecular Network Simulator software, answers two questions: (1) which reaction will occur next, and (2) what is the waiting time for the next reaction to occur. To answer these questions, two random numbers uniformly distributed over the interval (0,1) – r_1 and r_2 – are generated. The first random number is used to determine the next reaction R_k according to the probability density in Equation (A5c). Thus, k is selected to satisfies the condition

$$\sum_{i=1}^{k-1} a_i < r_1 * a_0 < \sum_{i=1}^{k} a_i \tag{A6}$$

where a_0 is given by Equation (A5d).

The distribution of the waiting time is given by the probability density function Equation (A5b). The waiting time for the next reaction is calculated as (Gillespie, 1977)

$$\tau = \frac{1}{a_0} \log \frac{1}{r_2} \tag{A7}$$

Once the next reaction and its waiting time are determined, the reaction is executed and the state of the system is changed according to the state-change vector $\boldsymbol{\nu}^k$. The simulation time is increased by τ and any reaction propensities affected by the reaction are recalculated. If the simulation time is less than the total duration of the simulation the algorithm is repeated to generate the next simulation step, otherwise, the simulation is terminated. The Gillespie stochastic algorithm is exact for the fundamental reactions indicated in Table A2. independent of

the number of reacting molecules in the system.

A second algorithm based on the modifications of Gibson and Bruck (2000) is also included in the BNS software. This algorithm, which is also mathematically exact, allows for the reuse of random numbers to minimize the computational time required to generate new random numbers for each reaction step.

The third algorithm available in the BNS software is the τ-leaping algorithm of Gillespie (Gillespie, 2001). The tau-leaping algorithm calculates a time interval τ which encompasses more than one reaction event and satisfies the Leap Condition, i.e., the expected state change induced by the leap must be sufficiently small that no propensity function changes its value by a significant amount. Several methods have been proposed recently to choose the size of the time interval for tau-leaping (Chatterjee, et al., 2005; Gillespie, D., 2001, Gillespie and Petzold, 2003; and Tian and Burrage, 2004). We implemented the tau-leaping method proposed by Cao et al. (2006). In that method a tau selection formula is given by

$$\tau' = \min_{i \in I_{rs}} \left\{ \frac{\max\{\varepsilon x_i / g_i, 1\}}{|\mu_i(x)|}, \frac{\max\{\varepsilon x_i / g_i, 1\}^2}{\sigma_i^2(x)} \right\} \tag{A8}$$

where g_i is the highest order of reaction in which species S_i appears as reactant, ε is an error control parameter, I_{rs} is the set of indices of all reactant species, and

$$\mu_i(x) = \sum_j \nu_{ij} a_j(x),$$

(9)

and

$$\sigma_i^2(x) = \sum_j \nu_{ij}^2 a_j(x).$$

(10)

After the time interval τ′ has been selected, the number of firings of each reaction channel during this time interval is approximated as a Poisson random variable. The Poisson random variable can have arbitrary large sample value and it is possible that the population of some of the molecular species can run negative. Therefore, a critical number of molecules n_c (typically in the range of 5-20) was introduced. If the number of molecular species gets less then n_c, all reaction in which that species appears as reactant are defined as critical. These reactions are simulated by the stochastic simulation algorithm. We calculate the sum of propensity functions of all the critical reactions a_0^c and generate a second candidate time τ′′ according to

$$\tau'' = \frac{1}{a_o^c} \log \frac{1}{r}.$$

(11)

$$p_2(k|s,t) = \int_0^\infty p(\tau, k|s,t)\, d\tau = \frac{a_k(s)}{a_0(s)} \tag{A5c}$$

and

$$a_0 = \sum_{i=1}^{N_R} a_i \tag{A5d}$$

$p_1(\tau|s,t)$ is the conditional density function for the waiting time to the next reaction with the mean waiting time of $1/a_0(\mathbf{s})$. $p_2(k|s,t)$ is the conditional density function that the next reaction will be the k^{th} reaction and is equal to the ratio of the propensity for the k^{th} reaction to the total propensity for all reactions in the system.

The evolution of the state vector $\mathbf{s}(t)$ can be calculated using various algorithms proposed for Monte Carlo generation of stochastic trajectories. The Gillespie direct simulation algorithm, (Gillespie, 1077; Table A2), which is one of the algorithms implemented in the Biomolecular Network Simulator software, answers two questions: (1) which reaction will occur next, and (2) what is the waiting time for the next reaction to occur. To answer these questions, two random numbers uniformly distributed over the interval (0,1) – r_1 and r_2 – are generated. The first random number is used to determine the next reaction R_k according to the probability density in Equation (A5c). Thus, k is selected to satisfies the condition

$$\sum_{i=1}^{k-1} a_i < r_1 * a_0 < \sum_{i=1}^{k} a_i \tag{A6}$$

where a_0 is given by Equation (A5d).

The distribution of the waiting time is given by the probability density function Equation (A5b). The waiting time for the next reaction is calculated as (Gillespie, 1977)

$$\tau = \frac{1}{a_0} \log \frac{1}{r_2} \tag{A7}$$

Once the next reaction and its waiting time are determined, the reaction is executed and the state of the system is changed according to the state-change vector $\boldsymbol{\nu}^k$. The simulation time is increased by τ and any reaction propensities affected by the reaction are recalculated. If the simulation time is less than the total duration of the simulation the algorithm is repeated to generate the next simulation step, otherwise, the simulation is terminated. The Gillespie stochastic algorithm is exact for the fundamental reactions indicated in Table A1. independent of

Table A2: Summary of the main steps in the Gillespie stochastic simulation algorithm.

Step 1: Initialization. Set the state vector $\boldsymbol{s} = \boldsymbol{s}_0$ (the initial number of molecules of each molecular species) and set the time $t = 0$.
Step 2: Calculate the propensity $a_i(\boldsymbol{s})$ for each reaction R_i and their sum $a_0(\boldsymbol{s})$
Step 3: Generate two random numbers r_1 and r_2 uniformly distributed over the interval (0,1).
Step 4: Find the next reaction R_j according to Eq. (1).
Step 5: Calculate the waiting time τ_j for the next reaction according to Eq. (4).
Step 6: Update the number of molecules of all molecular species affected by reaction Rj according to the state-change vector $\boldsymbol{v}_j$.
Step 7: Update the simulation time $t = t + \tau_j$.
Step 8: Calculate a new propensity $a_i(\boldsymbol{s})$ for each reaction that was affected by Step 6 and their sum $a_k(\boldsymbol{s})$
Step 7: Return to Step 3 or termination.
Termination. Simulations are terminated when the simulation time exceeds the maximal time of the simulations.

References

Cao, Y., Gillespie, D. and Petzold, L., 2006. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys*. **124**: 44109.

Chatterjee, A., Vlachos, D. and Katsoulakis, M., 2005. Binomial distribution based τ-leap accelerated stochastic simulation. *J. Chem. Phys*. **122**: 24112.

Gibson, M.A. and Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**: 1876-1889.

Gillespie, D., 1977, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem*. **81**: 2340-2361.

Gillespie, D. 1992. *Markov Processes: An Introduction for Physical Scientist*. Academic Press. pp. 565.

Gillespie, D., 2001. Approximate accelerated stochastic simulations of chemically reacting systems. *J. Chem. Phys*. **115**: 1716-1733.

Gillespie, D. and Petzold, L., 2003. Improved leap-size selection for accelerated stochastic simulations. *J. Chem. Phys*. **119**: 8229.

Tian, T. and Burrage, K. 2004. Binomial leap methods for simulation chemical kinetics." *J. Chem. Phys*. 121: 10356.

Appendix B - Biomolecular Network Simulator Software

The Biomolecular Network Simulator software is designed to provide an environment to perform stochastic simulations and to analyze the resulting data. BNS can be run on any computer platform where MATLAB 6.5 or newer is installed. Complete details of the BNS software can be found at http://www.bioanalysis.org.

.

Input Data

A model is a set of mathematical relationships that describe the behavior of biochemical reactions that control cellular biological processes. Each of the 'Model' directories contains one or more subdirectories with model description files and/or different set of parameters for the same model and an 'Output' directory where the results of simulations are stored. There are two types of model directories: one for models defined in the Systems Biology Markup Language (SBML) format (Hucka, et al., 2003) and one for models defined in BNS format as a set of MATLAB m-files. There are two additional m-files in each 'Parameters' directory – general_constants.m and storing_and_plotting.m, which contain information about simulation conditions and storing of the output data, respectively. In addition, BNS allows one to perform simulations with multiple parameter sets, with each parameter set being run multiple times. Simulations with multiple parameter sets can be used for optimization and sensitivity analysis of the model.

Output Data

There are two types of output files: snapshot data and event log data. Both of these files are in MATLAB format. Snapshot data files contain the state of the system (number of molecules of each molecular species) at user specified time intervals. The information stored in the snapshot files are used to create runtime interactive graphics and for *post hoc* analysis of the data. The second type of output files – the event log files – contain the record of every discrete event that occurs during the simulation. The user should be aware that event log files may require considerable memory or hard disk space and, therefore, may create memory management problems for simulations involving a large number of long runs or for large reaction networks.

Running Simulations

The BNS can be run either in command line mode or via a GUI. The GUI allows the user to modify model parameters at runtime and to execute simulations in the interactive or batch mode on HPC resources. The main dialog window of the BNS GUI is shown in Figure B1. It allows the user to select the appropriate 'Model' and 'Parameters' directories and set the 'Run' mode. A click on the 'Details' button next to the 'Parameters' directory opens a new window, shown in Figure B2. This dialog window allows the user to modify model parameters and to set parameters for the simulation. If simulations are run in the interactive mode, the results of

simulations will instantly appear on the screen. Usually, HPC centers allocate limited resources (in number of processors and running time) for the interactive simulations, therefore BNS can be run in 'Batch' mode. In this mode all output data are stored on the hard drive for further analysis.

Analysis

The BNS has a comprehensive set of tools for the post-simulation analysis. A GUI for the analysis tools allows the user to easily select the data and to set conditions for the analysis. Multiple types of post-simulation analyses are available.

Plots of number of molecules vs. time

The most frequently used type of analysis is a plot of the number of molecules vs. time. Such plots are available in the interactive mode or as post-simulation analysis. There are two ways to create plots: each compound is plotted on a separate figure or multiple compounds are plotted on the same figure window (grouping mode). The plot in Figure B3 is in grouping mode and shows the behavior of two molecular species, S1 and S2, over the time interval of 1500 seconds for a biomolecular reaction network containing transcription, translation and metabolic reactions. The number of molecules vs. time plots can be created with both types of output files: snapshot data or event log data.

Time-weighted average analysis

A time-weighted average analysis refers to the calculations of the average number of molecules of a particular molecular species during a user selected time-bin. The average is weighted according to the amount of time the compound exists in each state during the selected time-bin. The time weighted average is then plotted versus time. The averaging analysis can be performed for a single run or for a selected set of runs.

When the analysis is applied to multiple runs, the plot shows the between run average (the average of each run's time-weighted average) and standard deviation. As in the previous case, the user can plot each compound on a separate figure or multiple compounds on the same figure. Figure B4 shows the between run average of the time weighted average number of molecules for the same 50 runs as shown on Figure B3 using a time averaging interval of 10 sec.

Reaction frequency analysis

Complex biomolecular reaction networks usually contain reactions that occur on different time scales: some reactions have a low propensity and occur rarely; other reactions are very fast and occur frequently. The data stored in the event log files allow the user to perform various reaction frequency analyses of the simulation data to learn more about the basic nature of the system. One type of analysis creates plots of the total number of times each reaction occurred during the simulation. Figure B5 shows an example of a histogram of the average number and standard deviation of the number of reaction occurrences in each reaction channel averaged over the 50 runs in Figure B3. The number of reactions is shown in the logarithmic scale.

Reaction rates (number of reactions per unit time) in each reaction channel can be calculated for user-selected time-bins and plotted versus time. Such types of analyses provide important information about the behavior of the system.

References

Hucka, M., Finney, A., Sauro H. M., Bolouri, H., et al., 2003, The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinfor.* **19**: 524-531.

Figure B1: A screen shot of the main BNS GUI dialog window.

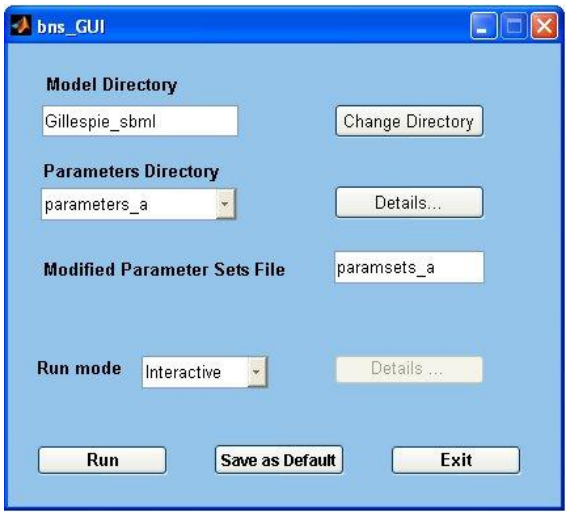


Figure B2: The parameters dialog window of the BNS GUI allows the user to modify the model parameters and to set simulation parameters.

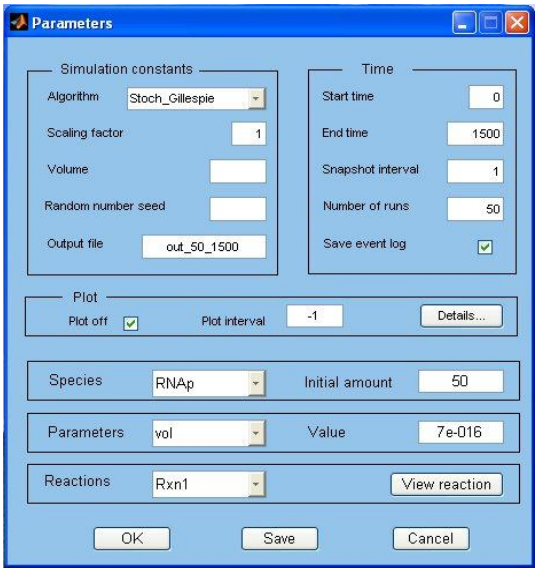


Figure B3: The evolution of the number molecules of molecular species S1 and S2 with time. The snapshot data for 50 runs are shown.

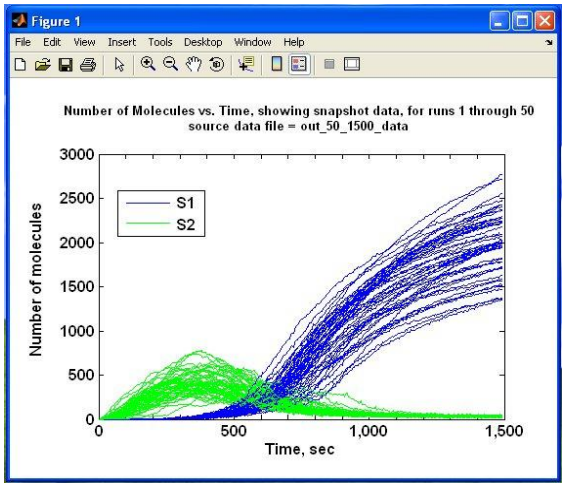


Figure B4: The averaged number of compounds S1 and S2 in the time interval (0, 1500) for the same simulation runs as in Figure 4. For each simulation, the time weighted average was calculated using a 10-sec time-bin and the time weighted averages were averaged over the 50 simulation runs. Data for the mean ± SD are shown.

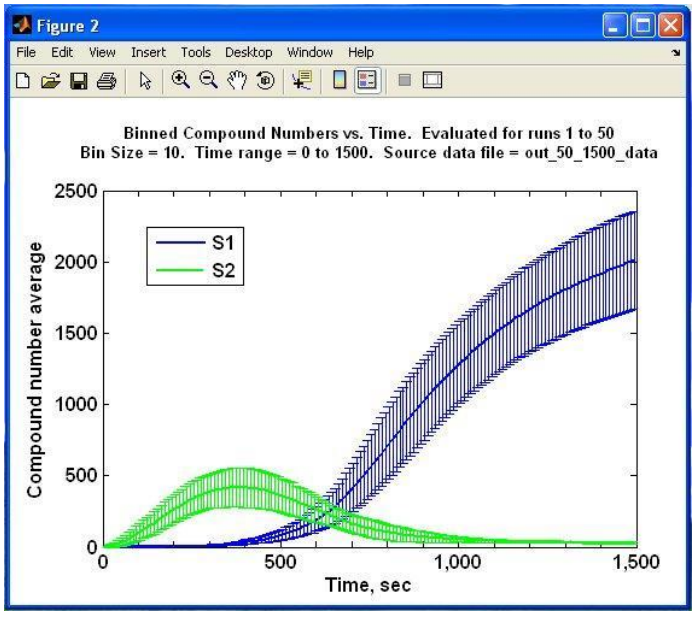
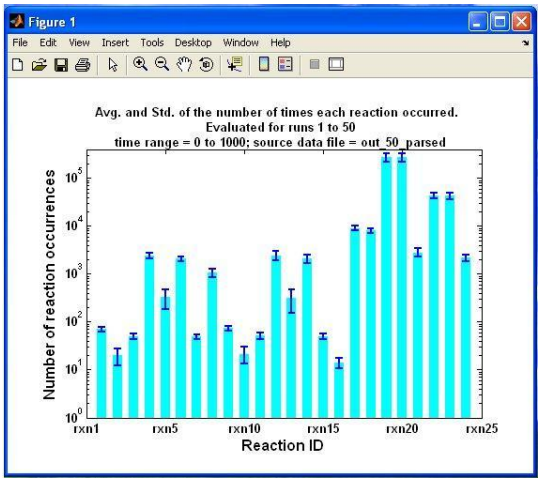


Figure B5: The average total number of reaction occurrences in each reaction channel. The data are for the mean ± SD for the 50 simulation runs.



Appendix C - geneA_CFTT_0p0 Model Documentation

Conceptual Model

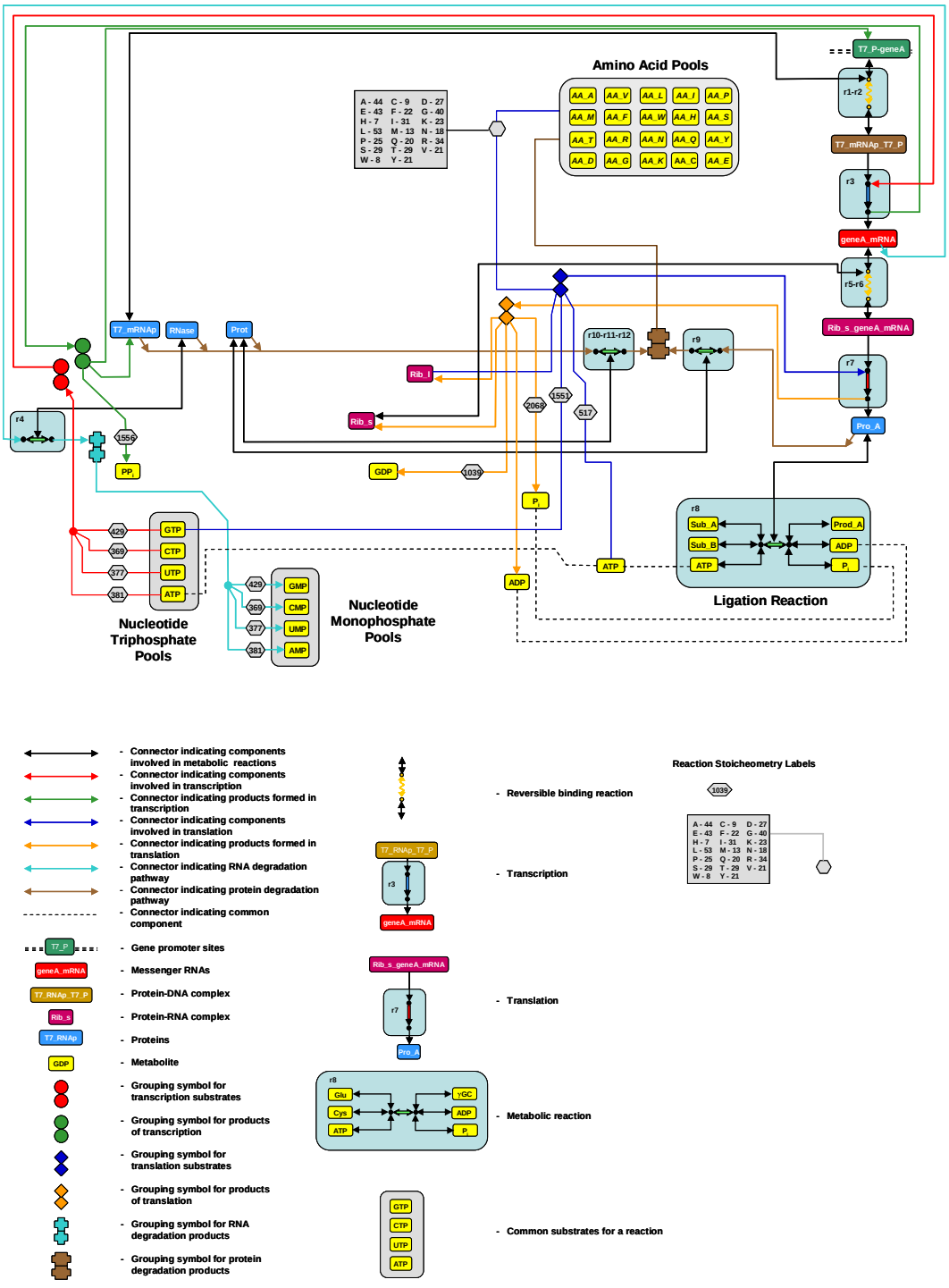


Figure C.1: The schematic diagram of the geneA self-assembling catalytic reaction model.

The diagram in Figure C.1 is a schematic of the various reactions and material flow connections between state variables of the geneA_CFTT_0p0 model. This is the simplest model for the biomolecular reaction system that retains the basic features of the system. Because many of the reactions are described at the conceptual level as lumped, macro-reactions, this model represents an approximation to the exact fundamental representation of the biomolecular reaction system that would be rigorously compliant with the basic tenants of the Markov process theory of multi-variant, discrete state, temporally homogenous, Markov jump processes.

To transform the conceptual model into a schematic more representative of the mathematical description of the system, the model schematic in Figure C.1 can be redrawn as in Figure C.2.

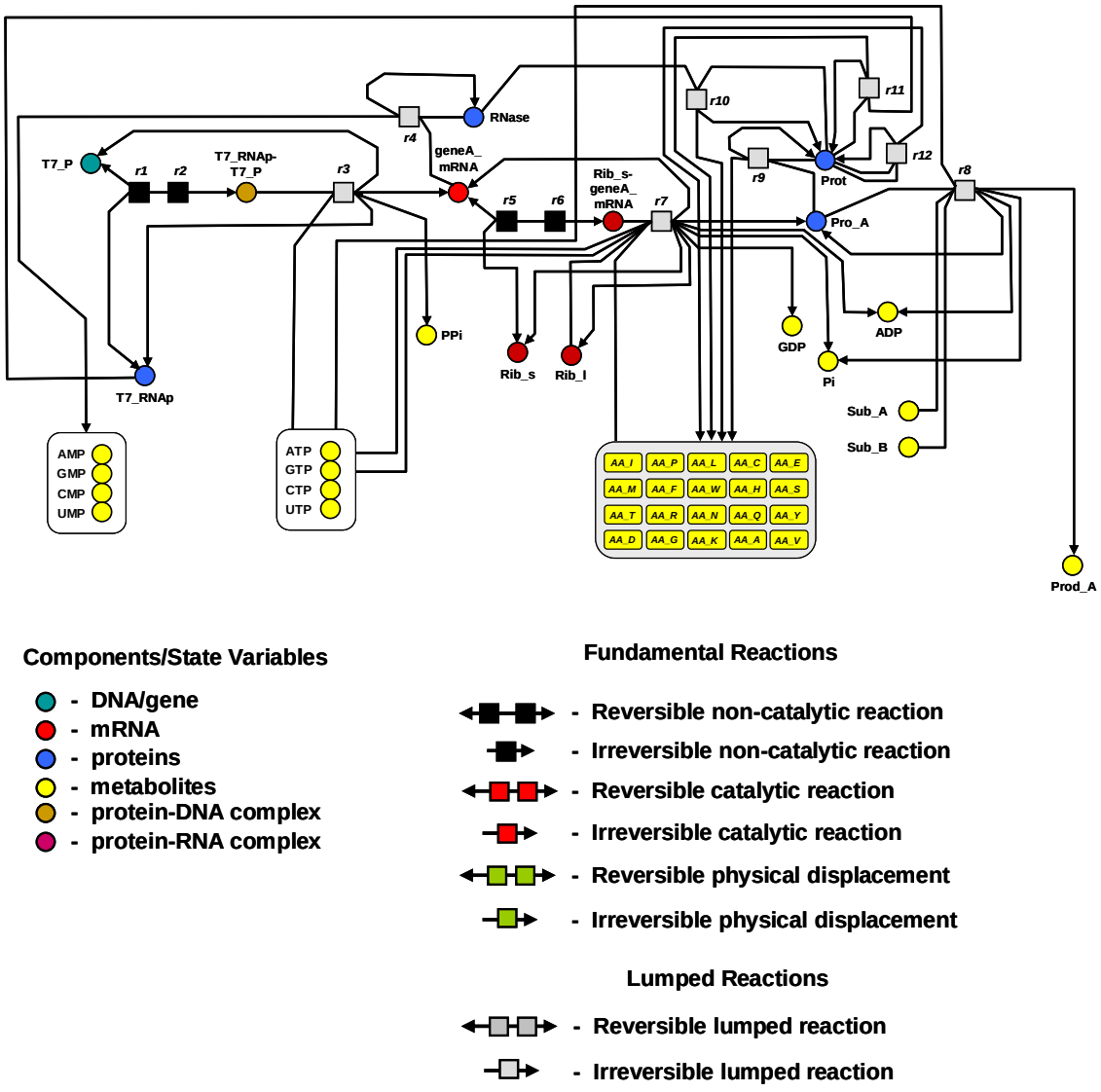


Figure C.2: Schematic diagram of the mathematical model of the geneA_CFTT_0p0 model.

In this representation, each reaction is defined and all of the substrates for each reaction are easily identified. In this case the mathematical model of the geneA-CFTT system is a mixture of fundamental and lumped reactions.

Reactions for a Minimal Model of geneA Self-assembly in a Cell-Free Transcription-Translation System

Transcription (Reactions r1 - r3))

In the simplest model (denoted by the postfix _0p0), the transcription process in the cell-free transcription-translation (CFTT) expression system is treated as a two step reaction: (1) reversible association and dissociation of the T7 RNA polymerase (T7_RNAp) to the T7 promoter site (T7_P) on the plasmid to form the initiation complex (T7_RNAp-T7_P), and (2) creation of the geneA mRNA (geneA_mRNA) in an instantaneous event at a time determined by the stochastic probability. The formation of the mRNA utilizes the appropriate number of nucleotide triphosphates (NTPs: ATP, GTP, CTP and UTP) and creates one pyrophosphate molecule (PPi) as byproduct per NTP incorporated. The number of each NTP used is determined by the sequence of the geneA gene and the number of PPi formed will be equal to the total number of nucleotides in the translated portion of the geneA gene.

The specific reactions involved in transcription are:

Reaction *r1*: The initial step in transcription of the geneA gene is the association of T7 RNA polymerase with its promoter site on the plasmid containing the geneA gene.

T7_P + T7_RNAp -> T7_RNAp-T7_P

Reaction *r2*: Dissociation of the polymerase-promoter complex.

T7_RNAp-T7_P -> T7_P + T7_RNAp

Reactions *r1* and *r2* are fundamental stochastic reactions that are described by standard Markov propensities for bi- and uni- substrate reactions, respectively:

$$a_1 = c_1 * T7\_RNAp * T7\_P / V_R$$

and

$$a_2 = c_2 * T7\_RNAp - T7\_P$$

As there is only one plasmid in the reaction volume, the state variable T7_P that represents the T7 promoter on the plasmid can fluctuate only between values of 1 and 0, where T7_P = 1 implies the geneA T7 promoter site is unoccupied and T7_P = 0 when T7_RNAp is associated with the promoter site. Thus,

$$a_1 = \begin{cases} c_1 * T7\_RNAp / V_R & when \quad T7\_P = 1 \\ 0 & when \quad T7\_P = 0 \end{cases}$$

and

$$a_2 = \begin{cases} c_2 & when \quad T7\_RNAp - T7\_P = 1 \\ 0 & when \quad T7\_RNAp - T7\_P = 0 \end{cases}$$

Reaction *r3*: Elongation and ultimate formation of the geneA_mRNA polymer by the incorporation of the nucleotide tri-phosphates (UTP, CTP, ATP and GTP) into the RNA polymer. Inorganic pyrophosphate (PPi) is formed as a byproduct.

T7_RNAp-T7_P + 381 UTP + 429 CTP + 377 ATP + 369 GTP ->
geneA_mRNA + T7_P + T7_RNAp + 1556 PPi

where the nucleotide composition of a generic geneA has been used. Reaction *r3* is a lumped, macro-reaction and must be treated as an approximation to the exact series of fundamental reactions that constitute the transcription reaction. Again, as the consequence of there being only one plasmid in the reaction volume, when T7_RNAp is associated with the promoter site, T7_RNAp-T7_P = 1, and the transcription reaction (*r3*) can occur. The propensity for reaction *r3* is

$$a_3 = c_3 * T7\_RNAp - T7\_P * f_3(ATP, GTP, CTP, UTP) / V_R$$

which can take on only two values

$$a_3 = \begin{cases} 0 & when \quad T7\_RNAp - T7\_P = 0 \\ c_3 * f_3(ATP, GTP, CTP, UTP) / V_R & when \quad T7\_RNAp - T7\_P = 1 \end{cases}$$

where $f_3(ATP, GTP, CTP, UTP)$ describes the dependency of the polymerization reaction on the NTP substrates. Obviously, a detailed model of the transcription reaction would be more exact and the natural dependency of the reaction on the availability of substrates would appear as a consequence of the fundamental reactions involved in transcription. However, for the purposes of this simple model, we will treat *r3* as a lumped macro-reaction and use an approximate phenomenological formulation for the propensity based on the Michaelis-Menten approximation for isomerization reactions. Thus,

$$f_3(ATP, GTP, CTP, UTP) = \left(\frac{ATP}{ATP + km3\_ATP} \right) * \left(\frac{GTP}{GTP + km3\_GTP} \right) * \left(\frac{CTP}{CTP + km3\_CTP} \right) * \left(\frac{UTP}{UTP + km3\_UTP} \right)$$

Note, $0 \le f_3(s4,s5,s6,s7) \le 1$, thus this formulation guarantees that the propensity will be zero if any one of the substrates is depleted, and will be limited to a maximum value of c_3 even when there is excess of the NTP substrates.

The probability that transcription will occur is non-zero as long as the polymerase binds to the promoter and there are sufficient NTP substrates available to form complete mRNA polymers. As transcription progresses, the geneA_mRNA formed serves as either the substrate for the translation process or for the mRNA degradative process.

Translation (Reactions *r5-r7*)

As with the transcription process, in this simplest conceptual model (geneA_CFTT_0p0), the translation process is treated as a two step process: (1) reversible association-dissociation of the ribosomal small unit (Rib_s) to the ribosomal binding site (RBS) on geneA_mRNA, and (2) translation of the mRNA into the Pro_A protein by an instantaneous event at a time determined by the stochastic probability.

The three translation reactions are:

Reaction *r5*: Association of the ribosomal small-unit (Rib_s - 30S subunit) with the ribosomal binding site (RBS) on the geneA_mRNA.

geneA_mRNA + Rib_s -> Rib_s-geneA_mRNA

Reaction *r6*: Dissociation of the ribosomal small subunit from the RBS on the geneA-mRNA.

Rib_s-geneA_mRNA -> geneA_mRNA + Rib_s

As with transcription, the translation reactions r5 and r6 are fundamental stochastic reactions that are described by standard Markov propensities for bi- and uni- substrate reactions, respectively:

$$a_5 = c_5 * Rib\_s * geneA\_mRNA / V_R$$

and

$$a_6 = c_6 * Rib\_s - geneA\_mRNA$$

Reaction *r7*: Elongation and ultimate formation of the Pro_A protein by the incorporation of the appropriate number of amino acids into the protein polymer. Upon completion of the reaction, the Pro_A protein is released and the geneA_mRNA, ribosomal large subunit (Rib_l) and the ribosomal small subunit (Rib_s) are returned to the available pools for reuse. Guanidine diphosphate, adenine diphosphate and inorganic phosphate (Pi) are formed as by-products.

Rib_s-geneA_mRNA + Rib_l + 44 A + 9 C + 27 D + 43 E + 22 F + 40 G + 7 H + 31 I + 23 K + 53 L + 13 M + 18 N + 25 P + 20 Q + 34 R + 29 S + 29 T + 21 V + 8 W + 21 Y + 1552 GTP + 517 ATP -> geneA_mRNA + Rib_s + Rib_l + Pro_A + 1552 GDP + 517 ADP + 2068 Pi

Note, the amino acid composition of Pro_A is based on the nucleotide sequence of geneA. Once Rib_s is bound to the geneA_mRNA forming the Rib_s-geneA_mRNA complex, the protein product, Pro_A , can be assembled via the translation reaction (*r7*) with a phenomenological propensity of

$$a_7 = c_7 * Rib\_l * f_7(Rib\_s - geneA\_mRNA, ATP, GTP, AA\_I, \ldots, AA\_V) / V_R$$

where the substrate dependency is given by

$$f_7(Rib\_s - geneA\_mRNA, ATP, GTP, AA\_I, \ldots, AA\_V) = \left(\frac{Rib\_s - geneA\_mRNA}{Rib\_s - geneA\_mRNA + km7\_Rib\_s - geneA\_mRNA}\right) * \left(\frac{ATP}{ATP + km7\_ATP}\right) * \ldots * \left(\frac{AA\_V}{AA\_V + km7\_AA\_V}\right)$$

Here, the propensity of the translation reaction is modulated by the availability of the substrates, including the Rib_s-geneA_mRNA complex, energy substrates GTP and ATP, and all of the amino acids (AA_I - AA_V). The modulation factor f_5, $0 \le f_5 \le 1$, is of the form of the product of hyperbolic factors and accounts for certain realistic features of the reaction, namely, a zero propensity when any of the substrates is not available to complete the polymerization reaction and a maximum propensity, $c_7 * Rib\_l / V_R$, when all substrates are available in saturating concentrations.

Metabolic Reaction

The enzyme mediated metabolic reaction catalyzed by Pro_A results in the ligation of Sub_A and Sub_B to form Prod_A using ATP as a source of free energy.

Reaction r8: Catalytic ligation of Sub_A and Sub_B to form Prod_A using ATP as a cofactor:

Pro_A + Sub_A + Sub_B + ATP -> Pro_A + Prod_A + ADP + Pi.

The synthetic reaction forming Prod_A is mediated by the product of the geneA gene. The enzyme Pro_A utilizes ATP to carry out the ligation of Sub_A and Sub_B. The sequence of micro-reactions that constitutes the ligation reaction - association and dissociation of substrates ATP, Sub_A and Sub_B and the enzyme, splitting ATP to ADP and Pi to provide free energy for the ligation reaction, and dissociation of Prod_A and by-products from the enzyme - is treated as

a single lumped macro-reaction in this version of the model. The reaction propensity is expressed as

$$a_8 = c_8 * \mathrm{Pr}o\_A * f_8(ATP, Sub\_A, Sub\_B) / V_R$$

where the f-factor

$$f_8(ATP, AA\_C, AA\_E) = \left(\frac{ATP}{ATP + km8\_ATP}\right) * \left(\frac{Sub\_A}{Sub\_A + km8\_Sub\_A}\right) * \left(\frac{Sub\_B}{Sub\_B + km8\_Sub\_B}\right)$$

again modulates the reaction propensity to assure that it does not exceed the maximum propensity, c_7*Pro_A/V_R, when all substrates are saturating and that the propensity goes to zero when any substrate is depleted. The product of the reaction is Prod_A along with the by-products ADP and inorganic phosphate.

Degradation Reactions

There are five degradation reactions in this model: one for degradation of mRNA and four for proteolytic degradation of the proteins present in the biomolecular reaction network. Since there is only one mRNA present, that reaction is treated as an independent reaction. However, since there are four proteins present in the system - Pro_A, T7_mRNAp, the generic RNase, the generic protease Prot, they are all subject to degradation by the single generic protease Prot. Since this model assumes that the same protease degrades all four proteins, then it is necessary to use a slightly modified formulation for the propensity that takes into consideration competition between the substrates for degradation.

Reaction *r4*: Degradation of geneA_mRNA by a generic RNase resulting in the formation of the constituent nucleotide mono-phosphates.

geneA_mRNA + RNase -> RNase + 381 UMP + 429 CMP + 377 AMP + 369 GMP

The mRNA degradation reaction (r4) introduces the possibility that geneA_mRNA will be degraded by the generic RNase. The products of the degradation process are the nucleotide monophophates (NMPs: AMP, GMP, CMP, UMP). Here, the mRNA degradation reaction is treated as a lumped macro-reaction again using an approximate phenomenological formulation for the propensity.

$$a_4 = c_4 * RNase * f_4(geneA\_mRNA)$$

where

$$f_4 = \frac{geneA\_mRNA}{geneA\_mRNA + k4\_geneA\_mRNA}$$

the hyperbolic functional form. The stochastic reaction parameter c_4 for this reaction is set low in the current model so that there is little degradation of mRNA.

Reaction *r9*: Degradation of the Pro_A protein by a generic protease (Prot) resulting in the formation of the constituent amino acids.

Pro_A + Prot -> Prot + 44 A + 9 C + 27 D + 43 E + 22 F + 40 G + 7 H + 31 I + 23 K + 53 L + 13 M + 18 N + 25 P + 20 Q + 34 R + 29 S + 29 T + 21 V + 8 W + 21 Y

Since this reaction is treated as a lumped, macro-reaction and the fact that the protease is assumed to degrade all proteins present in the biomolecular reaction network, the phenomenological propensity is given by

$$a_9 = c_9 * \Pr ot * f_9(\Pr o\_A,\ RNase,\ T7\_RNAp,\ \Pr ot)$$

where

$$f_9 = \frac{\Pr o\_A}{\Pr o\_A + km9\_\Pr o\_A * \left(1 + \frac{RNase}{km9\_RNase} + \frac{T7\_RNAp}{km9\_T7\_RNAp} + \frac{\Pr ot}{km9\_\Pr ot}\right)}$$

This expression is a phenomenological factor that modulates the propensity for reaction r9 to compensate for utilization of the enzyme by the other substrates.

Reaction *r10*: Degradation of RNase protein by the generic protease resulting in the formation of the constituent amino acids. (The amino acid composition of the RNase in the following reaction is an arbitrary placeholder)

RNase + Prot -> Prot + 50 A + 50 C + 50 D + 50 E + 50 F + 50 G + 50 H + 50 I + 50 K + 50 L + 50 M + 50 N + 50 P + 50 Q + 50 R + 50 S + 50 T + 50 V + 50 W + 50 Y

Since this reaction is treated as a lumped, macro-reaction and the fact that the protease is assumed to degrade all proteins present in the biomolecular reaction network, the phenomenological propensity is given by

$$a_{10} = c_{10} * \Pr ot * f_{10}(RNase,\ \Pr o\_A,\ T7\_RNAp,\ \Pr ot)$$

where

$$f_{10} = \frac{RNase}{RNase + km10\_RNase * \left(1 + \frac{\mathrm{Pro\_A}}{km10\_\mathrm{Pro\_A}} + \frac{T7\_RNAp}{km10\_T7\_RNAp} + \frac{\mathrm{Prot}}{km10\_\mathrm{Prot}}\right)}$$

Reaction *r11*: Degradation of T7 polymerase protein by the generic protease resulting in the formation of the constituent amino acids. (The amino acid composition of the T7 polymerase in the following reaction is an arbitrary placeholder)

T7_RNAp + Prot -> Prot + 50 A + 50 C + 50 D + 50 E + 50 F + 50 G + 50 H + 50 I + 50 K + 50 L + 50 M + 50 N + 50 P + 50 Q + 50 R + 50 S + 50 T + 50 V + 50 W + 50 Y

Since this reaction is treated as a lumped, macro-reaction and the fact that the protease is assumed to degrade all proteins present in the biomolecular reaction network, the phenomenological propensity is given by

$$a_{11} = c_{11} * \mathrm{Prot} * f_{11}(T7\_RNAp, \mathrm{Pro\_A}, RNase, \mathrm{Prot})$$

where

$$f_{11} = \frac{T7\_RNAp}{T7\_RNAp + km11\_T7\_RNAp * \left(1 + \frac{\mathrm{Pro\_A}}{km11\_\mathrm{Pro\_A}} + \frac{RNase}{km11\_RNase} + \frac{\mathrm{Prot}}{km11\_\mathrm{Prot}}\right)}$$

Reaction *r12*: Self-degradation of the generic protease protein resulting in the formation of the constituent amino acids.(The amino acid composition of the generic protease Pro in the following reaction is a placeholder)

Prot + Prot -> Prot + 50 A + 50 C + 50 D + 50 E + 50 F + 50 G + 50 H + 50 I + 50 K + 50 L + 50 M + 50 N + 50 P + 50 Q + 50 R + 50 S + 50 T + 50 V + 50 W + 50 Y

Since this reaction is also treated as a lumped, macro-reaction and the fact that the protease is assumed to degrade all proteins present in the biomolecular reaction network, the phenomenological propensity is given by

$$a_{12} = c_{12} * \mathrm{Prot} * f_{12}(\mathrm{Prot}, \mathrm{Pro\_A}, T7\_RNAp, RNase)$$

where

$$f_{12} = \frac{\mathrm{Prot}}{\mathrm{Prot} + km12\_\mathrm{Prot} * \left(1 + \frac{\mathrm{Pro\_A}}{km12\_\mathrm{Pro\_A}} + \frac{T7\_RNAp}{km12\_T7\_RNAp} + \frac{RNase}{km12\_RNase}\right)}$$

Code for the SBML model description of the geneA_CFTT_0p0 model

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level1" level="1" version="1">
  <model name="geneA">
    <listOfCompartments>
      <compartment name="" volume="1"/>
    </listOfCompartments>
    <listOfSpecies>
      <specie name="T7_RNAp" compartment="C1" initialAmount="301"/>
      <specie name="T7_P" compartment="C1" initialAmount="1"/>
      <specie name="T7_RNAp_T7_P" compartment="C1" initialAmount="0"/>
      <specie name="geneA_mRNA" compartment="C1" initialAmount="0"/>
      <specie name="Rib_s" compartment="C1" initialAmount="301"/>
      <specie name="Rib_l" compartment="C1" initialAmount="301"/>
      <specie name="Rib_s_geneA_mRNA" compartment="C1" initialAmount="0"/>
      <specie name="Pro_A" compartment="C1" initialAmount="0"/>
      <specie name="Prot" compartment="C1" initialAmount="301"/>
      <specie name="RNase" compartment="C1" initialAmount="301"/>
      <specie name="GC" compartment="C1" initialAmount="0"/>
      <specie name="ATP" compartment="C1" initialAmount="189329"/>
      <specie name="ADP" compartment="C1" initialAmount="29197"/>
      <specie name="AMP" compartment="C1" initialAmount="0"/>
      <specie name="GTP" compartment="C1" initialAmount="234178"/>
      <specie name="GDP" compartment="C1" initialAmount="0"/>
      <specie name="GMP" compartment="C1" initialAmount="0"/>
      <specie name="CTP" compartment="C1" initialAmount="150500"/>
      <specie name="CMP" compartment="C1" initialAmount="6622"/>
      <specie name="UTP" compartment="C1" initialAmount="166754"/>
      <specie name="UMP" compartment="C1" initialAmount="0"/>
      <specie name="Pi" compartment="C1" initialAmount="0"/>
      <specie name="PPi" compartment="C1" initialAmount="0"/>
      <specie name="AA_A" compartment="C1" initialAmount="43645"/>
      <specie name="AA_C" compartment="C1" initialAmount="94213"/>
      <specie name="AA_D" compartment="C1" initialAmount="61705"/>
      <specie name="AA_E" compartment="C1" initialAmount="79163"/>
      <specie name="AA_F" compartment="C1" initialAmount="45752"/>
      <specie name="AA_G" compartment="C1" initialAmount="76454"/>
      <specie name="AA_H" compartment="C1" initialAmount="33411"/>
      <specie name="AA_I" compartment="C1" initialAmount="52976"/>
      <specie name="AA_K" compartment="C1" initialAmount="21973"/>
      <specie name="AA_L" compartment="C1" initialAmount="44247"/>
      <specie name="AA_M" compartment="C1" initialAmount="37324"/>
      <specie name="AA_N" compartment="C1" initialAmount="12943"/>
      <specie name="AA_P" compartment="C1" initialAmount="205884"/>
      <specie name="AA_Q" compartment="C1" initialAmount="2107"/>
      <specie name="AA_R" compartment="C1" initialAmount="49665"/>
      <specie name="AA_S" compartment="C1" initialAmount="35217"/>
      <specie name="AA_T" compartment="C1" initialAmount="8428"/>
      <specie name="AA_V" compartment="C1" initialAmount="37926"/>
      <specie name="AA_W" compartment="C1" initialAmount="29498"/>
      <specie name="AA_Y" compartment="C1" initialAmount="56588"/>
    </listOfSpecies>
    <listOfParameters>
```

```
    <parameter name="vol" value="5e-016"/>
   </listOfParameters>
   <listOfReactions>
    <reaction name="r1" reversible="false">
     <listOfReactants>
      <specieReference specie="T7_RNAp"/>
      <specieReference specie="T7_P"/>
     </listOfReactants>
     <listOfProducts>
      <specieReference specie="T7_RNAp_T7_P"/>
     </listOfProducts>
     <kineticLaw formula="c1*T7_RNAp*T7_P/vol">
      <listOfParameters>
       <parameter name="c1" value="1e-018"/>
       <parameter name="vol" value="5e-016"/>
      </listOfParameters>
     </kineticLaw>
    </reaction>
    <reaction name="r2" reversible="false">
     <listOfReactants>
      <specieReference specie="T7_RNAp_T7_P"/>
     </listOfReactants>
     <listOfProducts>
      <specieReference specie="T7_RNAp"/>
      <specieReference specie="T7_P"/>
     </listOfProducts>
     <kineticLaw formula="c2*T7_RNAp_T7_P">
      <listOfParameters>
       <parameter name="c2" value="0.1"/>
      </listOfParameters>
     </kineticLaw>
    </reaction>
    <reaction name="r3" reversible="false">
     <listOfReactants>
      <specieReference specie="T7_RNAp_T7_P"/>
      <specieReference specie="ATP" stoichiometry="381"/>
      <specieReference specie="GTP" stoichiometry="429"/>
      <specieReference specie="UTP" stoichiometry="369"/>
      <specieReference specie="CTP" stoichiometry="377"/>
     </listOfReactants>
     <listOfProducts>
      <specieReference specie="T7_RNAp"/>
      <specieReference specie="T7_P"/>
      <specieReference specie="geneA_mRNA"/>
      <specieReference specie="PPi" stoichiometry="1556"/>
     </listOfProducts>
     <kineticLaw
formula="c3*(T7_RNAp_T7_P/vol)*(ATP/(ATP+km3_ATP))*(GTP/(GTP+km3_GTP))*(CTP/(CTP+km3_C
TP))*(UTP/(UTP+km3_UTP))">
      <listOfParameters>
       <parameter name="c3" value="5e-017"/>
       <parameter name="vol" value="5e-016"/>
       <parameter name="km3_ATP" value="301"/>
       <parameter name="km3_GTP" value="301"/>
       <parameter name="km3_CTP" value="301"/>
       <parameter name="km3_UTP" value="301"/>
```

```
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r4" reversible="false">
      <listOfReactants>
        <specieReference specie="geneA_mRNA"/>
        <specieReference specie="RNase"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="RNase"/>
        <specieReference specie="AMP" stoichiometry="381"/>
        <specieReference specie="GMP" stoichiometry="429"/>
        <specieReference specie="UMP" stoichiometry="369"/>
        <specieReference specie="CMP" stoichiometry="377"/>
      </listOfProducts>
      <kineticLaw
formula="c4*(Prot/vol)*RNase/(RNase+km4_RNase*(1+(T7_RNAp/km4_T7_RNAp)+(Pro_A/km4_Pro_A)
+(Prot/km4_Prot)))">
        <listOfParameters>
          <parameter name="c4" value="5e-033"/>
          <parameter name="km4_RNase" value="301"/>
          <parameter name="km4_T7_RNAp" value="301"/>
          <parameter name="km4_Prot" value="301"/>
          <parameter name="km4_Pro_A" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r5" reversible="false">
      <listOfReactants>
        <specieReference specie="geneA_mRNA"/>
        <specieReference specie="Rib_s"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Rib_s_geneA_mRNA"/>
      </listOfProducts>
      <kineticLaw formula="c5*Rib_s*geneA_mRNA/vol">
        <listOfParameters>
          <parameter name="c5" value="1e-018"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r6" reversible="false">
      <listOfReactants>
        <specieReference specie="Rib_s_geneA_mRNA"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="geneA_mRNA"/>
        <specieReference specie="Rib_s"/>
      </listOfProducts>
      <kineticLaw formula="c6*Rib_s_geneA_mRNA">
        <listOfParameters>
          <parameter name="c6" value="0.1"/>
        </listOfParameters>
      </kineticLaw>
```

```
    </reaction>
    <reaction name="r7" reversible="false">
      <listOfReactants>
        <specieReference specie="Rib_s_geneA_mRNA"/>
        <specieReference specie="Rib_l"/>
        <specieReference specie="GTP" stoichiometry="1552"/>
        <specieReference specie="ATP" stoichiometry="517"/>
        <specieReference specie="AA_A" stoichiometry="44"/>
        <specieReference specie="AA_C" stoichiometry="9"/>
        <specieReference specie="AA_D" stoichiometry="27"/>
        <specieReference specie="AA_E" stoichiometry="43"/>
        <specieReference specie="AA_F" stoichiometry="22"/>
        <specieReference specie="AA_G" stoichiometry="40"/>
        <specieReference specie="AA_H" stoichiometry="7"/>
        <specieReference specie="AA_I" stoichiometry="31"/>
        <specieReference specie="AA_K" stoichiometry="23"/>
        <specieReference specie="AA_L" stoichiometry="53"/>
        <specieReference specie="AA_M" stoichiometry="13"/>
        <specieReference specie="AA_N" stoichiometry="18"/>
        <specieReference specie="AA_P" stoichiometry="25"/>
        <specieReference specie="AA_Q" stoichiometry="20"/>
        <specieReference specie="AA_R" stoichiometry="34"/>
        <specieReference specie="AA_S" stoichiometry="29"/>
        <specieReference specie="AA_T" stoichiometry="29"/>
        <specieReference specie="AA_W" stoichiometry="8"/>
        <specieReference specie="AA_V" stoichiometry="21"/>
        <specieReference specie="AA_Y" stoichiometry="21"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="geneA_mRNA"/>
        <specieReference specie="Rib_s"/>
        <specieReference specie="Rib_l"/>
        <specieReference specie="Pro_A"/>
        <specieReference specie="Pi" stoichiometry="2069"/>
        <specieReference specie="GDP" stoichiometry="1552"/>
        <specieReference specie="ADP" stoichiometry="517"/>
      </listOfProducts>
      <kineticLaw
formula="c7*(Rib_l/vol)*(ATP/(ATP+km7_ATP))*(GTP/(GTP+km7_GTP))*(Rib_s_geneA_mRNA/(Rib_s_g
eneA_mRNA+km7_Rib_s_geneA_mRNA))*(AA_I/(AA_I+km7_AA_I))*(AA_M/(AA_M+km7_AA_M))*(AA_
T/(AA_T+km7_AA_T))*(AA_D/(AA_D+km7_AA_D))*(AA_P/(AA_P+km7_AA_P))*(AA_F/(AA_F+km7_AA_
F))*(AA_R/(AA_R+km7_AA_R))*(AA_G/(AA_G+km7_AA_G))*(AA_L/(AA_L+km7_AA_L))*(AA_W/(AA_W
+km7_AA_W))*(AA_N/(AA_N+km7_AA_N))*(AA_K/(AA_K+km7_AA_K))*(AA_C/(AA_C+km7_AA_C))*(A
A_H/(AA_H+km7_AA_H))*(AA_Q/(AA_Q+km7_AA_Q))*(AA_A/(AA_A+km7_AA_A))*(AA_E/(AA_E+km7_
AA_E))*(AA_S/(AA_S+km7_AA_S))*(AA_Y/(AA_Y+km7_AA_Y))*(AA_V/(AA_V+km7_AA_V))">
        <listOfParameters>
          <parameter name="c7" value="2.5e-018"/>
          <parameter name="vol" value="5e-016"/>
          <parameter name="km7_ATP" value="301"/>
          <parameter name="km7_GTP" value="301"/>
          <parameter name="km7_Rib_s_geneA_mRNA" value="301"/>
          <parameter name="km7_AA_I" value="301"/>
          <parameter name="km7_AA_M" value="301"/>
          <parameter name="km7_AA_T" value="301"/>
          <parameter name="km7_AA_D" value="301"/>
          <parameter name="km7_AA_P" value="301"/>
```

```
          <parameter name="km7_AA_F" value="301"/>
          <parameter name="km7_AA_R" value="301"/>
          <parameter name="km7_AA_G" value="301"/>
          <parameter name="km7_AA_L" value="301"/>
          <parameter name="km7_AA_W" value="301"/>
          <parameter name="km7_AA_N" value="301"/>
          <parameter name="km7_AA_K" value="301"/>
          <parameter name="km7_AA_C" value="301"/>
          <parameter name="km7_AA_H" value="301"/>
          <parameter name="km7_AA_Q" value="301"/>
          <parameter name="km7_AA_A" value="301"/>
          <parameter name="km7_AA_E" value="301"/>
          <parameter name="km7_AA_S" value="301"/>
          <parameter name="km7_AA_Y" value="301"/>
          <parameter name="km7_AA_V" value="301"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r8" reversible="false">
      <listOfReactants>
        <specieReference specie="Pro_A"/>
        <specieReference specie="AA_E"/>
        <specieReference specie="ATP"/>
        <specieReference specie="AA_C"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Pro_A"/>
        <specieReference specie="ADP"/>
        <specieReference specie="Pi"/>
        <specieReference specie="GC"/>
      </listOfProducts>
      <kineticLaw
formula="c8*(Pro_A/vol)*(AA_E/(AA_E+km8_AA_E))*(ATP/(ATP+km8_ATP))*(AA_C/(AA_C+km8_AA_C
))">
        <listOfParameters>
          <parameter name="c8" value="2.5e-016"/>
          <parameter name="km8_ATP" value="301"/>
          <parameter name="km8_AA_E" value="301"/>
          <parameter name="km8_AA_C" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r9" reversible="false">
      <listOfReactants>
        <specieReference specie="Pro_A"/>
        <specieReference specie="Prot"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Prot"/>
        <specieReference specie="AA_A" stoichiometry="44"/>
        <specieReference specie="AA_C" stoichiometry="9"/>
        <specieReference specie="AA_D" stoichiometry="27"/>
        <specieReference specie="AA_E" stoichiometry="43"/>
        <specieReference specie="AA_F" stoichiometry="22"/>
        <specieReference specie="AA_G" stoichiometry="40"/>
```

```
        <specieReference specie="AA_H" stoichiometry="7"/>
        <specieReference specie="AA_I" stoichiometry="31"/>
        <specieReference specie="AA_K" stoichiometry="23"/>
        <specieReference specie="AA_L" stoichiometry="53"/>
        <specieReference specie="AA_M" stoichiometry="13"/>
        <specieReference specie="AA_N" stoichiometry="18"/>
        <specieReference specie="AA_P" stoichiometry="25"/>
        <specieReference specie="AA_Q" stoichiometry="20"/>
        <specieReference specie="AA_R" stoichiometry="34"/>
        <specieReference specie="AA_S" stoichiometry="29"/>
        <specieReference specie="AA_T" stoichiometry="29"/>
        <specieReference specie="AA_W" stoichiometry="8"/>
        <specieReference specie="AA_V" stoichiometry="21"/>
        <specieReference specie="AA_Y" stoichiometry="21"/>
      </listOfProducts>
      <kineticLaw
formula="c9*(Prot/vol)*Pro_A/(Pro_A+km9_Pro_A*(1+(RNase/km9_RNase)+(T7_RNAp/km9_T7_RNAp)
+(Prot/km9_Prot)))">
        <listOfParameters>
          <parameter name="c9" value="5e-030"/>
          <parameter name="km9_Pro_A" value="301"/>
          <parameter name="km9_RNase" value="301"/>
          <parameter name="km9_T7_RNAp" value="301"/>
          <parameter name="km9_Prot" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r10" reversible="false">
      <listOfReactants>
        <specieReference specie="RNase"/>
        <specieReference specie="Prot"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Prot"/>
        <specieReference specie="AA_A" stoichiometry="50"/>
        <specieReference specie="AA_C" stoichiometry="50"/>
        <specieReference specie="AA_D" stoichiometry="50"/>
        <specieReference specie="AA_E" stoichiometry="50"/>
        <specieReference specie="AA_F" stoichiometry="50"/>
        <specieReference specie="AA_G" stoichiometry="50"/>
        <specieReference specie="AA_H" stoichiometry="50"/>
        <specieReference specie="AA_I" stoichiometry="50"/>
        <specieReference specie="AA_K" stoichiometry="50"/>
        <specieReference specie="AA_L" stoichiometry="50"/>
        <specieReference specie="AA_M" stoichiometry="50"/>
        <specieReference specie="AA_N" stoichiometry="50"/>
        <specieReference specie="AA_P" stoichiometry="50"/>
        <specieReference specie="AA_Q" stoichiometry="50"/>
        <specieReference specie="AA_R" stoichiometry="50"/>
        <specieReference specie="AA_S" stoichiometry="50"/>
        <specieReference specie="AA_T" stoichiometry="50"/>
        <specieReference specie="AA_W" stoichiometry="50"/>
        <specieReference specie="AA_V" stoichiometry="50"/>
        <specieReference specie="AA_Y" stoichiometry="50"/>
      </listOfProducts>
```

```
      <kineticLaw
formula="c10*(Prot/vol)*RNase/(RNase+km10_RNase*(1+(T7_RNAp/km10_T7_RNAp)+(Pro_A/km10_Pr
o_A)+(Prot/km10_Prot)))">
        <listOfParameters>
          <parameter name="c10" value="5e-030"/>
          <parameter name="km10_RNase" value="301"/>
          <parameter name="km10_T7_RNAp" value="301"/>
          <parameter name="km10_Prot" value="301"/>
          <parameter name="km10_Pro_A" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r11" reversible="false">
      <listOfReactants>
        <specieReference specie="T7_RNAp"/>
        <specieReference specie="Prot"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Prot"/>
        <specieReference specie="AA_A" stoichiometry="50"/>
        <specieReference specie="AA_C" stoichiometry="50"/>
        <specieReference specie="AA_D" stoichiometry="50"/>
        <specieReference specie="AA_E" stoichiometry="50"/>
        <specieReference specie="AA_F" stoichiometry="50"/>
        <specieReference specie="AA_G" stoichiometry="50"/>
        <specieReference specie="AA_H" stoichiometry="50"/>
        <specieReference specie="AA_I" stoichiometry="50"/>
        <specieReference specie="AA_K" stoichiometry="50"/>
        <specieReference specie="AA_L" stoichiometry="50"/>
        <specieReference specie="AA_M" stoichiometry="50"/>
        <specieReference specie="AA_N" stoichiometry="50"/>
        <specieReference specie="AA_P" stoichiometry="50"/>
        <specieReference specie="AA_Q" stoichiometry="50"/>
        <specieReference specie="AA_R" stoichiometry="50"/>
        <specieReference specie="AA_S" stoichiometry="50"/>
        <specieReference specie="AA_T" stoichiometry="50"/>
        <specieReference specie="AA_W" stoichiometry="50"/>
        <specieReference specie="AA_V" stoichiometry="50"/>
        <specieReference specie="AA_Y" stoichiometry="50"/>
      </listOfProducts>
      <kineticLaw
formula="c11*(Prot/vol)*T7_RNAp/(T7_RNAp+km11_T7_RNAp*(1+(Pro_A/km11_Pro_A)+(RNase/km11_
RNase)+(Prot/km11_Prot)))">
        <listOfParameters>
          <parameter name="c11" value="5e-030"/>
          <parameter name="km11_RNase" value="301"/>
          <parameter name="km11_Prot" value="301"/>
          <parameter name="km11_Pro_A" value="301"/>
          <parameter name="km11_T7_RNAp" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction name="r12" reversible="false">
      <listOfReactants>
```

```
        <specieReference specie="Prot"/>
        <specieReference specie="Prot"/>
      </listOfReactants>
      <listOfProducts>
        <specieReference specie="Prot"/>
        <specieReference specie="AA_A" stoichiometry="50"/>
        <specieReference specie="AA_C" stoichiometry="50"/>
        <specieReference specie="AA_D" stoichiometry="50"/>
        <specieReference specie="AA_E" stoichiometry="50"/>
        <specieReference specie="AA_F" stoichiometry="50"/>
        <specieReference specie="AA_G" stoichiometry="50"/>
        <specieReference specie="AA_H" stoichiometry="50"/>
        <specieReference specie="AA_I" stoichiometry="50"/>
        <specieReference specie="AA_K" stoichiometry="50"/>
        <specieReference specie="AA_L" stoichiometry="50"/>
        <specieReference specie="AA_M" stoichiometry="50"/>
        <specieReference specie="AA_N" stoichiometry="50"/>
        <specieReference specie="AA_P" stoichiometry="50"/>
        <specieReference specie="AA_Q" stoichiometry="50"/>
        <specieReference specie="AA_R" stoichiometry="50"/>
        <specieReference specie="AA_S" stoichiometry="50"/>
        <specieReference specie="AA_T" stoichiometry="50"/>
        <specieReference specie="AA_W" stoichiometry="50"/>
        <specieReference specie="AA_V" stoichiometry="50"/>
        <specieReference specie="AA_Y" stoichiometry="50"/>
      </listOfProducts>
      <kineticLaw
formula="c12*(Prot/vol)*Prot/(Prot+km12_Prot*(1+(T7_RNAp/km12_T7_RNAp)+(Pro_A/km12_Pro_A)+(R
Nase/km12_RNase)))">
        <listOfParameters>
          <parameter name="c12" value="5e-030"/>
          <parameter name="km12_RNase" value="301"/>
          <parameter name="km12_T7_RNAp" value="301"/>
          <parameter name="km12_Prot" value="301"/>
          <parameter name="km12_Pro_A" value="301"/>
          <parameter name="vol" value="5e-016"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
  </listOfReactions>
 </model>
</sbml>
```

Model Parameters

Table C.1: Composition of geneA and proteins in the geneA_CFTT_0p0 model.

Stoichiometry of Molecules		
Nucleotide composition of the geneA gene	ATP	381
	GTP	429
	CTP	369
	UTP	377
Energy requirements for Pro_A formation	ATP	517
	GTP	1552
Amino acid composition of Pro_A protein	isoleucine	31
	methionine	13
	threonine	29
	aspartic acid	27
	proline	25
	phenylalanine	22
	arginine	34
	glycine	40
	leucine	53
	tryptophan	8
	asparagine	18
	lysine	23
	cysteine	9
	histidine	7
	glutamine	20
	alanine	44
	glutamic acid	43
	serine	29
	tyrosine	21
	valine	21
Amino acid composition of T7_RNAp	isoleucine	50
	methionine	50
	threonine	50
	aspartic acid	50
	proline	50
	phenylalanine	50
	arginine	50
	glycine	50
	leucine	50
	tryptophan	50
	asparagine	50
	lysine	50
	cysteine	50
	histidine	50
	glutamine	50
	alanine	50

	glutamic acid	50
	serine	50
	tyrosine	50
	valine	50
Amino acid composition of RNase	isoleucine	50
	methionine	50
	threonine	50
	aspartic acid	50
	proline	50
	phenylalanine	50
	arginine	50
	glycine	50
	leucine	50
	tryptophan	50
	asparagine	50
	lysine	50
	cysteine	50
	histidine	50
	glutamine	50
	alanine	50
	glutamic acid	50
	serine	50
	tyrosine	50
	valine	50
Amino acid composition of the generic Prot	isoleucine	50
	methionine	50
	threonine	50
	aspartic acid	50
	proline	50
	phenylalanine	50
	arginine	50
	glycine	50
	leucine	50
	tryptophan	50
	asparagine	50
	lysine	50
	cysteine	50
	histidine	50
	glutamine	50
	alanine	50
	glutamic acid	50
	serine	50
	tyrosine	50
	valine	50

Table C.2: Numerical values for model reaction parameters for the discrete-stochastic (D-S) and the continuous-deterministic (C-D) models.

Reaction Parameters			
	Parameter	D-S Model	C-D Model
r1 - association of T7 RNA polymerase (T7_RNAp) with T7 promoter (T7_P) on geneA gene	k1	1.0e-18	6.02e-01
r2 - dissociation of T7 RNA polymerase-T7 promoter complex	k2	1.0e-01	1.0e-01
r3 - transcription of the geneA gene to form geneA_mRNA (lumped reaction)	k3	5e-017	1.0e-01
	km3_ATP	301	1
	km3_GTP	301	1
	km3_CTP	301	1
	km3_UTP	301	1
r4 - degradation of geneA_mRNA by RNase (lumped reaction)	k4	5.0e-025	1.0e-09
	km4_geneA_mRNA	301	1
r5 - association of the small ribosomal unit (Rib_s) with geneA_mRNA ribosomal binding site	k5	1e-18	6.02e-01
r6 - dissociation of Rib_s-geneA_mRNA complex	k6	1e-01	1e-01
r7 - translation of geneA mRNA to form the Pro_A protein (lumped reaction)	k7	2.5e-18	5e-03
	km7_geneA_mRNA	301	1
	km7_ATP	301	1
	km7_GTP	301	1
	km7_AA_I	301	1
	km7_AA_M	301	1
	km7_AA_T	301	1
	km7_AA_D	301	1
	km7_AA_P	301	1
	km7_AA_F	301	1
	km7_AA_R	301	1
	km7_AA_G	301	1
	km7_AA_L	301	1
	km7_AA_W	301	1
	km7_AA_N	301	1
	km7_AA_K	301	1
	km7_AA_C	301	1
	km7_AA_H	301	1
	km7_AA_Q	301	1
	km7_AA_A	301	1
	km7_AA_E	301	1
	km7_AA_S	301	1
	km7_AA_Y	301	1

	km7_AA_V	301	1
r8 - production of Prod_A using Pro_A (lumped reaction)	k8	7.0e-017	1.4e-01
	km8_ATP	301	1
	km8_AA_C	301	1
	km8_AA_E	301	1
r9 - degradation of Pro_A by the generic protease (lumped reaction)	k9	5.0e-025	1.0e-09
	km9_T7_RNAp	301	1
	km9_Pro_A	301	1
	km9_RNase	301	1
	km9_Prot	301	1
r10 - degradation of T7_RNAp by the generic protease (lumped reaction)	k10	5.0e-025	1.0e-09
	km10_T7_RNAp	301	1
	km10_Pro_A	301	1
	km10_RNase	301	1
	km10_Prot	301	1
r11 - degradation of RNase by the generic protease (lumped reaction)	k11	5.0e-025	1.0e-09
	km11_T7_RNAp	301	1
	km11_Pro_A	301	1
	km11_RNase	301	1
	km11_Prot	301	1
r12 - self-degradation of the generic protease Prot (lumped reaction)	k12	5.0e-025	1.0e-09
	km12_T7_RNAp	301	1
	km12_Pro_A	301	1
	km12_RNase	301	1
	km12_Prot	301	1

Table C.3: Initial conditions for the state variables in geneA_CFTT_0p0 model.

Initial Number of Molecules in State Variables		
	D-S (# of molecules)*	C-D (μM)
T7_P	1	0.00332
T7_RNAp	301	1
T7_RNAp-T7_P	0	0
ATP	189329	629
GTP	234178	778
CTP	150500	500
UTP	166754	554
geneA_mRNA	0	0
PPi	0	0
AMP	0	0
UMP	0	0
CMP	6622	22
GMP	0	0
Rib_s	301	1
Rib_s-geneA_mRNA	0	0
Rib_l	301	1
AA_I - isoleucine	52976	176
AA_M - methionine	37324	124
AA_T - threonine	8428	28
AA_D - aspartic acid	61705	205
AA_P - proline	205884	684
AA_F - phenylalanine	45752	152
AA_R - arginine	49665	165
AA_G - glycine	76454	254
AA_L - leucine	44247	147
AA_W - tryptophan	29498	98
AA_N - asparagine	12943	43
AA_K - lysine	21973	73
AA_C - cysteine	94213	313
AA_H - histidine	33411	111
AA_Q - glutamine	2307	7

AA_A - alanine	43645	145
AA_E - glutamic acid	79163	263
AA_S - serine	35217	117
AA_Y - tyrosine	56588	188
AA_V - valine	37926	126
Pro_A	0	0
GDP	15652	52
Pi	0	0
ADP	29197	97
Sub_A	30100	100
Sub_B	45002	150
Prod_A	0	0
RNase	301	1
Prot	301	1

* Initial number of molecules of nucleotides and amino acids are based on analytical measurements in a commercial gene expression kit.