31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Multiscale Models for Synthetic Biology

Yiannis N. Kaznessis,

Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Ave SE, Minneapolis, MN 55112

Abstract—**Reacting systems away from the thermodynamic limit cannot be accurately modeled with ordinary differential equations. These continuous-deterministic modeling formalisms, traditionally developed and used by chemical engineers can be distinctly false if the number of molecules of reacting chemical species is very small, or if reaction events are very rare. Then stochastic-discrete representations are appropriate. Importantly, in cases where in a network of reactions there are some parts that must be modeled discretely and stochastically, yet others can be modeled continuously and deterministically, the need for development of multiscale models emerges naturally. In computational synthetic biology, such cases arise often. In this work we present the development of multiscale models for synthetic biology applications, demonstrating accuracy, computational efficiency and utility.**

Index Terms—**synthetic biology, stochastic simulations, kinetic Monte Carlo, Chemical Langevin Equations.**

I. Introduction

The nascent field of synthetic biology offers the promise of engineered gene networks with novel biological phenotypes. Numerous synthetic gene circuits have been created in the past decade, including bistable switches, oscillators, and logic gates [1-7]. Designing synthetic gene regulatory networks can take advantage of an ever-expanding toolbox of molecular components becoming known thanks to genome projects and the developed technologies for inexpensively manipulating DNA sequences. Biomedical and biotechnological applications abound: from protein production optimization, bioenergy generation and biosensing, to stem cell differentiation and gene therapies.

Despite a booming field and although recently developed designs of regulatable gene networks are ingenious, there are limitations in routinely engineering synthetic biological systems, i.e. designing a specific DNA sequence that will give rise in a targeted dynamic phenotype. Indeed, there is a need for rationalizing the design of novel regulatable gene networks that can be used in useful applications.

Manuscript received Jun 6, 2008. This work was supported in part by the National Science Foundation (CAREER award to YNK)

Y.N.K is with the Department of Chemical Engineering and Materials Science at the University of Minnesota, Minneapolis, MN 55455 USA (phone: 612-624-4197; fax: 612-626-7246; e-mail: yiannis@cems.umn.edu).

We are developing multiscale mathematical tools to rationalize synthetic biology [8-18]. Why are new multiscale models necessary to assist synthetic biology, and not simply apply the mathematics developed by physicists, chemists and chemical engineers to model kinetic and thermodynamic processes in living organisms? Because, although the principles of thermodynamics, kinetics and transport phenomena apply to biological systems, these systems differ from industrial-scale chemical systems in an important, fundamental way: they are occasionally far from the thermodynamic limit. This theoretical limit is attained when the number of molecules of molecular species in the system increases toward infinity. However, the fact that biomolecular systems can be very far from the thermodynamic limit, with reactants/products numbering only very small numbers of molecules in the system, hinders the use of continuous-deterministic models. Indeed, using ordinary differential equations for simulating the reaction kinetics of these systems can be distinctly false. The need arises then for stochastic models that account for inherent, thermal noise, which is manifest as phenotypic distributions at the population growth/interaction levels.

This assessment is not new. The importance of modeling formalisms appropriate for systems away from the thermodynamic limit was recognized more than 50 years ago by McQuarrie, Moyal and Oppenheim [19-23], among others. These physical chemists developed the chemical master equation that follows the time changes of the probability distribution the state is at any point in the available state space. In the next section, we will discuss the CME and the difficulties to solve it for complex systems.

In 1976, Daniel Gillespie developed a computer algorithm that could sample the master probability distribution with numerical simulations of networks of reactions [24, 25]. Although Gillespie's methods were not widely recognized for almost 20 years, his algorithms found fertile ground for development in efforts to model biological systems. Nowadays, a community of scientists and engineers is working on improving the computational efficiency and accuracy of algorithms that simulate chemical reacting systems [26-35].

II. Theory of Multiscale Models for Biomolecular Systems

The large number of molecular components and interactions involved in dynamic biological phenotypes requires sophisticated computational modeling. Computer simulations enable exhaustive searches of different network connectivities and molecular thermodynamic/kinetic parameters, greatly advancing the development of design principles.

Monod's and Jacob's assertion can be adopted that biological complexity emerges as a result of biomolecular interactions. We can then represent all gene expression molecular level events with reactions. For any two molecular species A and B (proteins, DNA, RNA, signaling molecules, etc.) interacting in solution to form a complex A*B (e.g. a repressor protein and the corresponding DNA operator site) we write

$$[A]aq + [B]aq \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [A*B]aq \quad (1)$$

with k_1 and k_{-1} the association and dissociation kinetic constants, respectively.

One could then generate networks of chemical reactions that incorporate all individual molecular species and interactions known to be involved in gene expression. All the steps in transcription, translation, regulation, induction, degradation can be expressed mechanistically with chemical reactions.

Sets of ordinary differential equations (ODEs) could be written for the chemical reaction kinetics, sets of initial conditions assigned and numerical simulations of biomolecular systems conducted.

Nonetheless, the underlying assumption of ODEs, which are continuous-deterministic models, is that the number of molecules approaches the thermodynamic limit (i.e. that the volume of the system and the number of reacting molecules are infinite).

This assumption can be invalid for biological systems, since for some components (DNA sites for example) there are only a few copies available. An alternative way to model reactions involving very dilute reactants is to treat the system kinetics as a Markov chain. In particular, it will be a Markov chain with a discrete set of possible states, or "state space", occurring in continuous time. Here, states refer to numbers of molecules present in the system. In general this will be a vector:

Xi(t) = number of molecules of the ith unique chemical where i = 1…N chemical species (2)

Transitions between states of the Markov chain occur when a chemical reaction occurs. Reactions in biological systems may include covalent reactions, bindings, conformational changes, transcriptional elongation events, etc. If the N chemical species engage in M distinct reactions, then the reactions are described by a stoichiometric matrix:

ν_{ij} = the M x N stoichiometic matrix (3)

Macroscopic rate constants are replaced by a vector of reaction propensities:

$a_j(X)dt$ = the probability that the jth reaction will occur in the system in a dt (4)

Thus, starting from a particular initial state at t_0, at some later time there exists a conditional probability distribution $P(\underline{X},t|\underline{X}_o,t_o)$ of possible states at a later time t [19-23]. The time-evolution of this probability density is described by the "Master Equation,"

$$\frac{dP(\underline{X},t\,|\,X_o,t_o)}{dt} = \sum_{i=j}^{M}\left[a_j(\underline{X}-\underline{\nu}_j;t)P(\underline{X}-\underline{\nu}_j,t\,|\,X_o,t_o) - a_j(\underline{X};t)P(\underline{X},t\,|\,X_o,t_o)\right] \quad (5)$$

If the master equation could be solved, one would know the probability of each state as a function of time, and the probabilistic behavior of the system would be completely characterized. Unfortunately, this is rarely possible with systems of even modest complexity. Rather than attempting to calculate the time-dependant probability density function, it is often useful to calculate individual stochastic reaction trajectories in state space. A numerical stochastic simulation algorithm (SSA) to calculate these trajectories was described by Gillespie [24-25]. Although accurate in capturing the dynamic of biomolecular interaction systems, SSA becomes computationally intractable, if the time scales of involved interaction events are disparate, because it simulates every single biomolecular interaction event, spending inordinate amounts on fast reactions for very few simulated occurrences of slow reactions. There have been numerous attempts to improve the efficiency of the SSA 26-35]. Usually though algorithms are not computationally efficient, especially for complex biological systems.

III. HYBRID STOCHASTIC-DISCRETE AND STOCHASTIC-CONTINUOUS ALGORITHMS

We devised numerous multiscale hybrid methods to simulate complex biomolecular interaction networks [8-18,36]. For the sake of brevity we will only describe one to set the stage for the presented work [8].

Given the system definition of the previous section, the system is dynamically portioned into two subsets, the fast/continuous and slow/discrete reactions. Namely M is now the sum of M^{fast} and M^{slow} respectively. Propensities are also designated as fast ($\underline{\alpha}^f$) and slow ($\underline{\alpha}^s$). Two conditions must be met for the jth reaction to be classified as "fast": i) The reaction occurs many times in a small time interval. ii)The effect of each reaction on the numbers of reactants and products species is small, when compared to the total numbers of reactant and product species. In equation form, respectively: $\alpha_j\left(\underline{X}(t)\right) \geq \lambda >> 1$ and $X_i(t) > \varepsilon \cdot \left|\nu_{ji}\right|$, where the i-th species is either a product or a reactant in Both conditions ensure that the fast subsystem can be approximated as a continuous Markov process, instead of a jump or discrete Markov process. The two parameters λ and ε define respectively the numbers of reactions occurring within time Δt and what is the upper limit for the effect of a reaction to be negligible in the number of molecules of the reactants and products. This approximation becomes valid when both λ and ε become infinite i.e. in the thermodynamic limit. In practice, we have found that typical values for λ and ε can be $O(10^2)$

and $O(10^4)$ respectively [8].

The subset of "fast" reactions can be approximated as a continuous time Markov process. Under the assumptions of the previous section, we can also partition the CME into fast and slow subsets. One can make a volume expansion to the CME governing the fast/continuous subset of reactions ending up with a Chemical Langevin Equation (CLE). The CLE is an Itô stochastic differential equation with multiplicative noise and represents one possible solution of the Fokker-Plank equation. In our case, of a multidimensional Fokker-Plank equation we end up with a system of Itô stochastic differential equations:

$$dX_i = \sum_{j=1}^{M^{fast}} \nu_{ji} \alpha_j^f \left(\underline{X}(t)\right) dt + \sum_{j=1}^{M^{fast}} \nu_{ji} \sqrt{\alpha_j^f \left(\underline{X}(t)\right)} dW_j \quad (6)$$

where α_j^f, ν_{ji} are the propensities and the stoichiometric coefficients of only the fast reactions, M^{fast} is the number of fast reactions and W is a Wiener process [23], which is a a continuous-time stochastic process producing a Gaussian-distributed noise perturbation. The system of CLEs is the one used to propagate the subsystem of fast/continuous reactions over time. Though not trivial to solve numerically, it is much easier to deal with it rather than looking into the CME.

A system of differential jump equations is used to calculate the next jump of a slow reaction. The jump equations are defined as:

$$\begin{aligned} dR_j(t) &= \alpha_j^s \left(\underline{X}(t)\right) dt, \\ R_j(t_0) &= \log\left(URN_j\right), \qquad j = 1,\ldots, M^{slow} \end{aligned} \quad (7)$$

where denote the residual of the jth slow reaction, M^{slow} are the propensities of only the slow reactions, is the number of slow reactions and URN is a uniform random number in the interval (0,1). Equations (7) depict the rate at which the reaction residuals change. Note that the initial conditions of all are negative. Equations (7) are also Itô differential equations even though they do not contain a Wiener process, because the propensities of the slow reactions depend on the state of the system, which in turn depends on the system of CLEs. Due to the coupling between the system of CLEs (6) and differential jump equations (7) they can be integrated simultaneously using SDE numerical schemes. This results in very significant computational gains, while retaining accuracy, as discussed in [8].

IV. COMPUTER-AIDED SYNTHETIC BIOLOGY

Using the multi-scale models, we have modeled and engineered in silico multiple synthetic gene networks: bistable switches, oscillators, tetracycline-inducible networks, and AND logical gates [36-40].

Recently we built the Synthetic Biology Software Suite, a software suite that automates the steps for building models of and conducting numerical simulations synthetic biological systems. There are three components in SynBioSS: *Designer, Wiki* and *Simulator.*

With SynBioSS Designer, gene network models are created automatically after the user enters molecular components and their relationships. Expression then progresses for any gene in a systematic way, following the molecular biology dogma: RNAp binds the promoter site, forms an open complex, proceeds with transcriptional elongation, and synthesizes mRNA; then ribosome binds the RBS of mRNA and proceeds with translational elongation and polypeptide synthesis; the protein is formed, it folds and functions. In the process, mRNA and protein molecules are degraded, enriching the pools of RNA bases and amino acids. This system of reactions can be built for any particular sequence of DNA with defined genetic components.

Every reaction in the model has a corresponding kinetic rate that describes the rate of association of its reactant molecules and the formation or destruction of any covalent bonds or stable non-covalent interactions. SynBioSS Wiki has been specifically created to store and recall just this sort of kinetic data. SynBioSS Wiki has two components: i) a web interface based on the MediaWiki package and ii) a database for storing molecular components, their interactions, and pertinent biological information. The SynBioSS Wiki goes beyond the MediaWiki software in storing kinetic information in a formatted (and therefore machine-searchable) format. The database of kinetic constants is easily searchable for participating species, reaction type, etc. Users can search or browse the web site and select reactions to interactively build a model that can be exported in a SBML format. Each kinetic constant entered in the database is correlated with a reference field in the database as well as type-specific reference information (pdb ID for proteins, CAS ID for small molecules, PubMed ID for everything, etc).

Simulating gene regulatory networks becomes simple with the third component of SynBioSS, the Desktop Simulator. SynBioSS Desktop Simulator can be downloaded as an installation executable for Windows (a beta MacOS version is available). The steps are as follows:

1. Go to synbioss.sourceforge.net
2. Click on "Simulator" on the upper left corner. This will take you to http://synbioss.sourceforge.net/simulator/
3. Click on "Download" in the middle of the webpage. This will take you to the sourceforge file directory.
4. Click on SynBioSSDSInstaller-1.0.1.exe. This downloads the installation executable on your computer.
5. Run the executable. This will install the current version of SynBioSS on your computer.
6. Click on the Start Menu to find and click the SynBioSS icon. This will launch SynBioSS.

We have made available with SynBioSS all the files to simulate such synthetic biological systems as bistable switches, tetracycline-inducble networks and oscillators.

We will continue making our codes available through SourceForge.net (search SynBioSS), the *de facto* forum for dissemination of open source code. In making our software available, we are using the BSD license format from OpenSource.org.

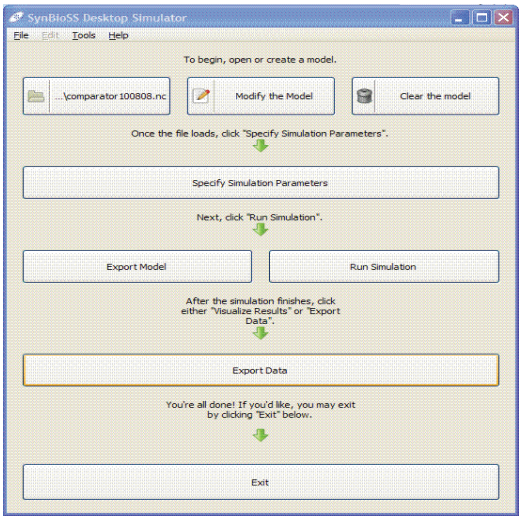

Figure 1 SynBioSS Desktop GUI. Users can quickly build or upload models of gene networks, specify simulation parameters, conduct simulations and export concentration vs. time data for analysis.

REFERENCES

[1] Andrianantoandro E, Basu S, Karig DK, Weiss R., Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol. 2006;2:2006.0028. .

[2] Keasling J (2005) The promise of synthetic biology. Bridge Natl Acad Eng 35: 18–21

[3] Alon U (2003) Biological networks: the tinkerer as an engineer. Science 301: 1866–1867

[4] Elowitz MB , Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335–338

[5] Endy D (2005) Foundations for engineering biology. Nature 438: 449–453

[6] Gardner TS , Cantor CR , Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339–342

[7] Kaern M , Blake WJ , Collins JJ (2003) The engineering of gene regulatory networks. Annu Rev Biomed Eng 5: 179–206 29.

[8] Salis H, Y. Kaznessis, "Accurate Hybrid Stochastic Simulation of a System of Coupled Chemical or Biochemical Reactions", J. Chem. Phys. Vol. 122, p. 054103 1-13, 2005.

[9] Kaznessis, Y. Models for Synthetic Biology, BMC Systems Biology 2007, 1:47

[10] H. Salis, Y. Kaznessis, "Stochastic Simulations of Gene Regulatory Modules", Computers in Chemical Engineering, Vol. 29, p. 577-588, 2005.

[11] L. Tuttle, H. Salis, J. Tomshine, Y. Kaznessis, "Model-Driven Design Principles of Gene Networks: the Oscillator", Biophysical Journal, Vol. 89(6), p. 3873-83, 2005.

[12] Tomshine J and Kaznessis YN. Optimization of a stochastically simulated gene network model via simulated annealing. Biophys J. 2006;91:3196–3205.

[13] H. Salis, Y. Kaznessis, "Computer-aided design of modular protein devices: Boolean AND gene activation." Phys Biol. 3(4):295-310, 2006.

[14] Sotiropoulos V and Kaznessis YN. Synthetic tetracycline-inducible regulatory networks: Computer-aided design of dynamic phenotypes. BMC Syst Biol. 2007;1:7.

[15] Salis H, Y. Kaznessis "An equation-free probabilistic steady state approximation: Dynamic application to the stochastic simulation of biochemical reaction networks", Journal of Chemical Physics, Vol. 123,p. 214106, 2005.

[16] Sotiropoulos V and Kaznessis YN. An adaptive time step scheme for a system of SDEs with multiple multiplicative noise. Chemical Langevin equation, a proof of concept. J Chem Phys. 128, 014103, 2008.

[17] Salis H, V. Sotiropoulos, Y. Kaznessis "Multiscale Hy3S: Hybrid Stochastic Simulations for Supercomputers", BMC Bioinformatics, (highly accessed), Vol. 7:93, 2006

[18] Kaznessis Y, "Multi-Scale Models for Gene Network Engineering", Chemical Engineering Science, Vol. 61(3), p. 940-953, 2006.

[19] McQuarrie DA. Stochastic approach to chemical kinetics. J Appl Probab. 1967;4:413–478.

[20] Moyal JE: Stochastic Processes and Statistical Physics. Journal of the Royal Statistical Society Series B (Methodological) 1949, 11:150-210

[21] Oppenheim I and Shuler KE. Master equations and markov processes. Phys Rev. 1965;138:1007–1011.

[22] Oppenheim I, Shuler KE and Weiss GH. Stochastic theory of multistate relaxation processes. Adv Mol Relax Processes. 1967;1:13–68.

[23] Oppenheim I, Shuler KE and Weiss GH. Stochastic processes in chemical physics: The master equation. Cambridge, MA: The MIT Press, 1977.

[24] Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22:403-434.

[25] Gillespie, D. T. 1977. Exact Stochastic Simulation of Coupled Chemical Reactions. Journal of Physical Chemistry 81:2340-2361. 101)

[26] Haseltine EL and Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J Chem Phys. 2002;117:6959–6969.

[27] Gibson, M. A., and J. Bruck. 2000. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. Journal of Physical Chemistry 104:1876-1889.

[28] Cao Y, Li H and Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. J Chem Phys. 2004;121:4059–4067.

[29] Cao Y, Gillespie DT and Petzold LR. Avoiding negative populations in explicit poisson tau-leaping. J Chem Phys. 2005;123:054104

[30] Harris LA and Clancy P. A ``partitioned leaping" approach for multiscale modeling of chemical reaction dynamics. J Chem Phys. 2006;125:144107.

[31] Chatterjee A, Mayawala K, Edwards JS and Vlachos DG. Time accelerated monte carlo simulations of biological networks using the binomial {tau}-leap method. Bioinformatics. 2005;21:2136–2137.

[32] Tian T and Burrage K. Binomial leap methods for simulating stochastic chemical kinetics. J Chem Phys. 2004;121:10356–10364

[33] E W, Liu D and Vanden-Eijnden E. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. J Chem Phys. 2005;123:194107.

[34] Munsky B and Khammash M. The finite state projection algorithm for the solution of the chemical master equation. J Chem Phys. 2006;124:044104.

[35] Chatterjee A and Vlachos DG. Temporal acceleration of spatially distributed kinetic monte carlo simulations. J Comput Phys. 2006;211:596–615.

[36] Sotiropoulos V and Kaznessis YN. Synthetic tetracycline-inducible regulatory networks: Computer-aided design of dynamic phenotypes. BMC Syst Biol. 2007;1:7.

[37] Tuttle L, H. Salis, J. Tomshine, Y. Kaznessis, "Model-Driven Design Principles of Gene Networks: the Oscillator", Biophysical Journal, Vol. 89(6), p. 3873-83, 2005.

[38] Tomshine J and Kaznessis YN. Optimization of a stochastically simulated gene network model via simulated annealing. Biophys J. 2006;91:3196–3205.

[39] Salis H, Y. Kaznessis, "Stochastic Simulations of Gene Regulatory Modules", Computers in Chemical Engineering, Vol. 29, p. 577-588, 2005.

[40] Salis H., Y. Kaznessis, "Computer-aided design of modular protein devices: Boolean AND gene activation." Phys Biol. 3(4):295-310, 2006.

[41] Hill,A, J. Tomshine, E. Wedding, V. Sotiropoulos, Y. Kaznessis, "SynBioSS: the Synthetic Biology Modeling Suite", Bioinformatics 2008, 24(21):2551-2553