Summer 8-5-2012

Development of a Novel Method for Biochemical Systems Simulation: Incorporation of Stochasticity in a Deterministic Framework

Amit Sabnis
Georgia State University

DEVELOPMENT OF A NOVEL METHOD FOR BIOCHEMICAL SYSTEMS SIMULATION: INCORPORATION OF STOCHASTICITY IN A DETERMINISTIC FRAMEWORK

by

AMIT SABNIS

Under the Direction of Robert W. Harrison

ABSTRACT

Heart disease, cancer, diabetes and other complex diseases account for more than half of human mortality in the United States. Other diseases such as AIDS, asthma, Parkinson's disease, Alzheimer's disease and cerebrovascular ailments such as stroke not only augment this mortality but also severely deteriorate the quality of human life experience. In spite of enormous financial support and global scientific effort over an extended period of time to combat the challenges posed by these ailments, we find ourselves short of sighting a cure or vaccine. It is widely believed that a major reason for this failure is the traditional reductionist approach adopted by the scientific community in the past. In recent times, however, the systems biology based research paradigm has gained significant favor in the research community especially in the field of com-

plex diseases. One of the critical components of such a paradigm is computational systems biology which is largely driven by mathematical modeling and simulation of biochemical systems. The most common methods for simulating a biochemical system are either: a) continuous deterministic methods or b) discrete event stochastic methods. Although highly popular, none of them are suitable for simulating multi-scale models of biological systems that are ubiquitous in systems biology based research. In this work a novel method for simulating biochemical systems based on a deterministic solution is presented with a modification that also permits the incorporation of stochastic effects. This new method, through extensive validation, has been proven to possess the efficiency of a deterministic framework combined with the accuracy of a stochastic method. The new crossover method can not only handle the concentration and spatial gradients of multi-scale modeling but it does so in a computationally efficient manner. The development of such a method will undoubtedly aid the systems biology researchers by providing them with a tool to simulate multi-scale models of complex diseases.

INDEX WORDS: Systems biology, Biosimulation, Numerical methods, System dynamics

DEVELOPMENT OF A NOVEL METHOD FOR BIOCHEMICAL SYSTEMS SIMULATION: INCORPORATION OF STOCHASTICITY IN A DETERMINISTIC FRAMEWORK

by

AMIT SABNIS

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2012

DEVELOPMENT OF A NOVEL METHOD FOR BIOCHEMICAL SYSTEMS SIMULATION: INCORPORATION OF STOCHASTICITY IN A DETERMINISTIC FRAMEWORK

by

AMIT SABNIS

Committee Chair: Robert W. Harrison

Committee: Irene T. Weber

Chung-Dar Lu

Xiaolin Hu

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2012

DEDICATION

This work is dedicated to my parents, Abhilasha Vijay Sabnis and Vijay Kamalakant Sabnis.

ACKNOWLEDGEMENTS

I always liked to think of myself as a self-made man until experience proved me wrong. A self-made man does not exist but what does exist is the acute illusion of being self-sufficient. It is so easy and convenient for us to forget the countless number of people who have directly or indirectly helped us towards our goal that any semblance of independent glory rationalizes as a self-made victory in our minds. In this section, I hope to acknowledge the contributions of some of those people towards what is officially recognized as my work. First of all, I would like to mention my parents; without their emotional support I would not have pursued academia for as long as I did. I would like to thank my friends at work, especially Xianfeng (Jeff) Chen, Nael M. Abu-halaweh and Hao Wang who did an excellent job at tolerating me in the lab and of course helping me whenever I needed them to. My friends (Rebecca Ebenezer, Anita Mall and Vishal Michael) did a wonderful job giving me a life to live outside of my lab. Last but not least, I have to thank my advisor Dr. Robert W. Harrrison for giving me the opportunity to work with him, for showing me that big things can come out of small ideas and above all for being an excellent mentor. I would also like to thank my committee members Dr. Weber, Dr. Hu and Dr. Lu for their support not only for my dissertation but throughout my program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Why Study the new paradigm of systems biology?

Complex diseases such as heart disease, cancer and diabetes account for more than half of human mortality in the United States. Other diseases such as AIDS, asthma, Parkinson's disease, Alzheimer's disease and cerebrovascular ailments such as stroke not only augment the mortality but also severely deteriorate the quality of human life experience. Although the global mortality count associated with HIV-1 infections has abated, failure to produce a vaccine still permits millions of new infections worldwide. In spite of being widely studied and well characterized, infectious diseases such as influenza and tuberculosis continue to provide a threat to human life across the globe. There is no lack of financial support afforded for research in these areas and has been, in fact, very generous. Each year, for the past four years, the National Institutes of Health (NIH) in the United States has provided over 10 billion dollars for research projects involving complex diseases like heart disease, cancer and diabetes (http://report.nih.gov). In spite of the exorbitant financial investment and ever-increasing global scientific research efforts to combat the challenges posed by these multifaceted ailments, the quest to find a cure or vaccine is not yet complete (Phair R. D., 2012). It is well known that these complex diseases are not a result of a single element but a combination of genetic, environmental and lifestyle factors. Other diseases that are not officially categorized as complex, for example Tuberculosis, are also more often than not a combination of a variety of physiological factors (Kitano 2007).

1.1.1 Limitations of reductionism

For the past several decades, however, the traditional research paradigm that has been employed to study these diseases has been the one of 'reductionism'. Reductionism refers to the

method of breaking down a biological phenomenon into its constituent components in an attempt to isolate and characterize the component responsible for the physiologically observed phenotype. The conceptual rationale for reductionism is that any physical phenomena can be explained by a deterministic evaluation of its constituents. An example of this approach is trying to explain the consciousness of human mind by 'reducing' the phenomenon to a set of chemical reactions occurring in the brain and studying them in isolation (Bickle et al., 2003). This "reductionist" approach of explaining high-level observations from activity of lower level components has served the research establishment very well over several decades and has led to a very accurate functional and structural annotation of the components under consideration. A major drawback of this approach, however, is that it provides limited insights into the functional properties of the system itself which can be important for a variety of reasons.

The study of systems biology relates to studying entire biological systems consisting of many interacting networks that ultimately define the holistic character of any organism including humans. Conversely, any biological entity can be viewed as nothing but an enormous complex network of interacting sub-networks (Kitano 2000, 2002). The basic differences between reductionism and holism are explained in figure 1. It is now widely accepted that an important facet of any biological system is the property of emergence (Kitano 2002, Van Regenmortel 2004). An 'emergent' property is the one that 'emerges' as a result of the system components coming together under specific circumstances and cannot be arrived at by simply adding the components. For example, the three dimensional conformation of a protein molecule cannot be predicted by simply lining up the amino acid sequence. The unique folding pattern adopted by the protein can thus said to be an emergent property of the system. Similarly, biological systems characterizing diseases are complex systems that have their own unique emergent properties that would be im-

possible to be studied by the reductionist approach. Hence, a complex disease like cancer cannot be studied in isolation of the environmental factors without sacrificing emergent properties. Another systems property of robustness is very important in the context of drug discovery and vaccine development (Kitano 2002). A biological system is composed of biochemical pathways, gene regulatory networks, signaling pathways, protein-protein interaction networks, protein-nucleic acid interaction networks and a number of environmental factors engaged in a highly structured yet complex interaction. Robustness is the ability of this system to adapt to any perceived perturbations in its components. The effect of any one of these components on the entire network is difficult to analyze with the current molecular biology methods. With most of the reductionist techniques geared towards understanding the function of individual components, it is almost impossible to predict what effect a local disturbance introduced into the system by a disease or a drug might have on a distant pathway which may not be part of the current system but is nevertheless important in terms of function. System robustness is generally achieved via redundancy of components where several parts of the system essentially serve the same purpose and affecting one of them generally provides no appreciable difference in the overall phenotype. In reductionism such a component would be simply ignored given the lack of its effect on the phenotype and in process lose a potential drug target. Redundancy allows the system to be modular so that a failure in one part of a system does not mitigate to other areas and makes the system 'robust' to external insults. Hence from a drug discovery point of view, because diseases are a result of variation in the biological homeostasis, knowledge of the inherent robustness of the system becomes very critical. Global effects of local perturbations in a 'diseased' biochemical network are difficult to predict by reductionist techniques that are designed to ignore the property of robustness.

It is for these reasons that some researchers argue that reductionism might have hit its ceiling with regards to biological research (Mazzocchi F., 2008, Van Regenmortel 2004). Another criticism of reductionism is that it also requires studying the component of interest external to its natural environment and then attempts to extrapolate the results to the host environment. Such extrapolation rarely works as evidenced by the failure of knockout mice experiments to be able to translate to human systems. It is clear from the preceding discussion that employing the systems level paradigm for investigating biological processes holds significant potential in solving a variety of complex problems in the areas of rational drug design, vaccine development, cancer therapy, metabolic engineering and personalized medicine.

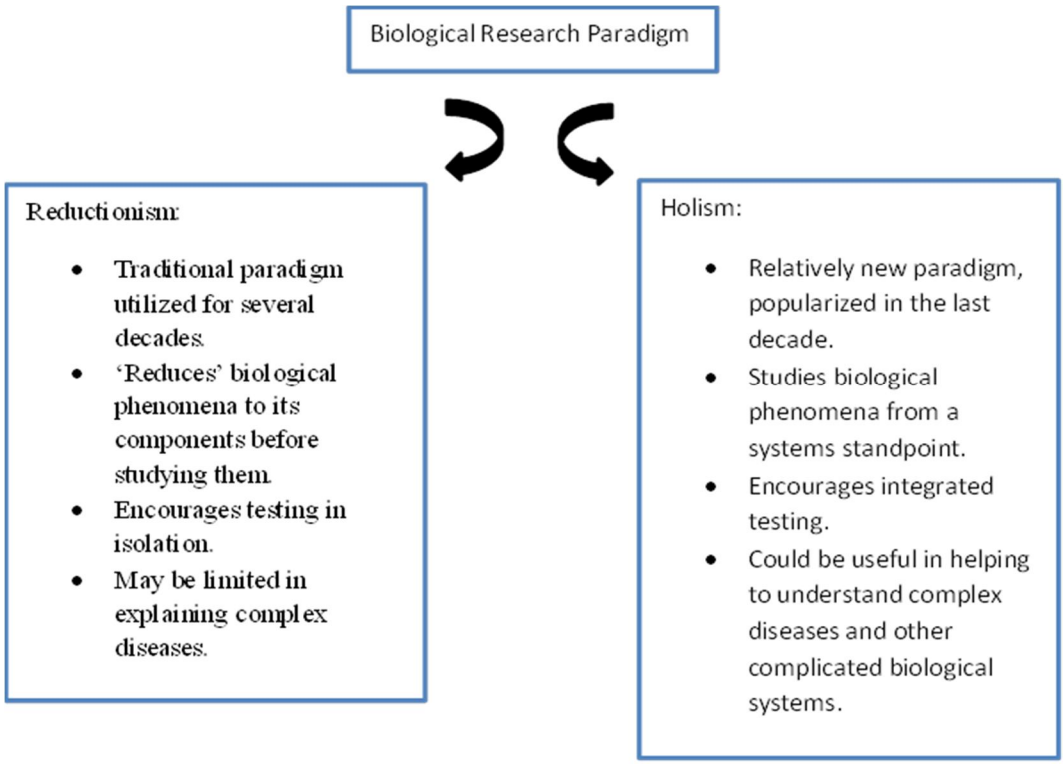

Figure 1 Conceptual differences between the two methodologies employed in biomedical research

1.1.2 Advent of high-throughput technologies

One key development that facilitated, and is arguably indispensable for, the shift from reductionism to holism is the advent of high-throughput technologies for biological experimentation (Resendis-Antonio O., 2011, Nurse P., 2011). A number of technologies such as mass spectrometry, DNA / RNA expression microarrays and next-generation sequencing platforms are now available to the molecular biologists. The literature offers a very detailed review of all the available technologies (Simpson J. C.., 2006, Segata N., 2008, Chen B. S., 2008). Mass spectrometry is widely applied in the field of proteomics where it is utilized primarily for identifying, quantifying and analyzing novel network components. In this technique, proteins are digested with proteases and then subjected to liquid phase chromatography and gas phase fractionation. The spectra thus obtained are used to determine the sequence of the protein. Mass spectrometry can identify thousands of proteins in a given sample making it an ideal fit for systems biology (Sabido E., 2011). DNA / RNA microarray technology is primarily used for detecting the up or down-regulation of RNA transcripts expressed in a biological system. As this technique can also detect thousands of transcripts at once, it fits nicely into the realm of systems biology. The RNA fragments to be detected are attached to probes and then hybridized with fluorescence. Laser is then used to detect the fluorescent intensity and determine the regulation status of RNA transcripts. The data generated by the experiment is then analyzed statistically to draw an inference and generate hypotheses that can be further tested. Microarray technology was one of the earliest technologies that helped expedite the molecular biologist's migration from reductionism to systems level interrogation. Other high-throughput technologies available target slightly different areas of biology. For instance, next generation sequencing technologies have revolutionized the area of genomics. It is now become possible to sequence the complete genome of any organism

in a matter of a few hours to days as opposed to several months required in the early part of the last decade. Although the recent high throughput experiments have been successful in providing a deluge of data to enable a systems view of a biological phenomenon, lack of adequate technology for generating meaningful hypotheses without novel experimentation still remains one of the major hindrances against successful treatment of fatal human diseases.

1.2 Application of systems biology: A literature review

One of the more obvious applications of holistic systems-level biology can be observed in the process of drug discovery. The exorbitant amount of capital spent on introducing new drugs in the market and the relatively high number of failed targets in clinical development warrants a review of current drug discovery process. Identification of drug targets and developing drugs against most diseases requires a methodical approach to help decipher the interactions among the participating biological factors, and ascertain how these factors contribute towards the etiology of the disease. The advancement of high-throughput technologies in molecular biology has allowed researchers to genotype and profile thousands of DNA markers and other molecular phenotypes simultaneously in large number of individuals, thereby permitting the reconstruction of those biological networks that are known to be associated with a particular disease. These reconstructed networks are more integrative and predictive thus providing a more insightful context for single genes that have been identified by traditional molecular biology techniques (Zhu et. al., 2008, Schadt et. al. 2009). Another advantage of systems biology based drug discovery is that in diseases such as AIDS where drug resistance is a major hindrance towards developing new drugs, systems biology may be able to provide alternate targets based on the knowledge of underlying interacting metabolic pathways (Andersen-Nissen E. et. al., 2012). The use of systems biology is also widespread in therapeutics development where mathematical techniques to

analyze and integrate large datasets are studied in order to discover novel vaccines for HIV infections (Buonaquro L., et. al., 2011, Haddad E. K., et. al., 2012). In a different approach to studying HIV pathogenesis, researchers have started focusing on the so called elite controllers of HIV. The elite controllers are individuals who have defied the presence of the virus in their systems and show no signs of transforming into full blown AIDS without the help of anti-retroviral therapy. On a physiological level, holistic approaches are currently being used to study the differential regulation of signaling pathways involved in T-cell depletion in these elite controllers of HIV (Fonseca S.G. et. al., 2011). Systems biology has also been applied to shed more light on how the sub-networks of an infected host act in concert to limit the damage to its immune system especially in elite controllers of HIV and natural carriers of SIV (Hoof I. et. al., 2011). A similar approach has also been reported for understanding the response by exposed uninfected women (Burgener A. et. al., 2010). Vaccine development using systems biology principles, however, is not restricted to one single disease and at least one study argues that using systems biology, instead of the narrowly focused 'isolate, inactivate, inject' strategy, is a more efficient option for vaccine development (Oberg A. L. et. al., 2011). Another study reports the application of systems biology principles in studying the integrated actions of innate and adapted immune response as an essential part of vaccine development (Buonaquro L. and Pulendarn B., 2011).

In the area of neuroscience, systems biology has contributed in gaining insights into mechanisms of synaptic plasticity (Kotaleski J. H. et. al., 2010) as well as addiction (Tretter F et. al., 2008). Complex neurological diseases such as neurofibromatosis type 1 (NF1) are also being explored from an integrative systems biology standpoint (Lee M. J. et. al., 2011). An integration of genome wide association studies with the gene expression data for Parkinson's disease recently provided more insight into the pathology of the ailment (Edwards Y. J. et. al., 2011). A similar

approach is adopted in the identifying new biomarkers and drug targets for the treatment of neocortical epilepsy (Loeb J. A., 2010). Geshwind offers an excellent review on how the overall concept of systems biology can be applied to various areas of neuroscience (Konopka G., 2011).

Understanding the etiology associated with complex medical conditions such as congenital heart diseases is greatly aided by systems biology (Sperling S. R., 2011) as is an elaborate understanding of the pathophysiology of heart disease (Dewey et. al., 2011). Recently systems biology has been instrumental in gaining systemic insight into cardiomyogenesis (Young D. A. et. al., 2011). Other complex diseases such as cancer also lend themselves as ideal targets for systems biology based research. Biological cellular networks are routinely deregulated in tumor metastasis. However, the resulting dynamics are not always comprehensible from an experimental output. Systems based mathematical models are hence important to help make sense of the complex behavior resulting from such a deregulation (Cloutier M. et. al., 2011). An example of this type of investigation can be cited from a recent study regarding the hyperactivation of PI3k/AKT pathway, where systems biological approaches were used to predict useful drug targets (Mosca E. et. al., 2011). Another study focusing on the dynamics of JAK-STAT pathway as related to cancer focuses on the application of systems biology in modeling cancer-relevant signal transduction networks (Vera J. et. al., 2011). Similarly, work on the role of epidermal growth factor receptors on cell migration in non-small cell lung cancer is also an extensive demonstration of the use of systems biology principles (Bianconi F et. al., 2011). From a clinical standpoint, holistic approaches are routinely utilized for identification of novel genes that could contribute to reduced efficacy of cancer treatments (Allen W. L. et. al., 2011).

It is clear from the above discussion that, systems biology has indeed been very useful in advancing biomedical related research as regards to disease. On a cellular level though, this par-

adigm has been equally well employed. Two of the most impacted areas of molecular biology at this level are the genetic regulatory system and the signal transduction system. On the genetic regulation level, systems biology has been used to better understand the role of microRNA in catalyzing the gene regulation process (Watanabe Y., 2011). As a specific example, in patients with pancreatic cancer certain miRNAs involved in metastasis were found to be deregulated and the epigenetic connection for the regulation of these miRNAs was investigated by using holistic systems biology (Azmi A. S. et. al., 2011). In yeast, the GAL regulon encodes for genes that allow the processing of galactose as an energy source. Recently, a systems level interrogation of this genetic network uncovered certain network properties of substrate regulation and autosensing that are important for the adaption of yeast to its environment (Pannala V. R. et. al., 2010). Experimental and computational systems biology has also been applied in the area of gene therapy (Mac Gabhann F. et. al., 2010). High throughput approaches are routinely used in identifying new members of the signal transduction family of G-protein-coupled receptors (GPCR) that are intimately involved in signal transduction in the regulation of normal mammalian physiological function (Wu J et. al., 2012). In plants, the dynamics of abscisic acid (ABA) signaling pathway are better interrogated by using transcriptome analysis and 'phosphoproteomics' approach (Umezawa T., 2011). Systems biology has also helped better understand the apoptotic signaling network in eukaryotes (Lavrik I.N., 2010).

1.3 Why is computational systems biology important?

Experimental systems biology facilitated via high-throughput experiments are just one part of the holistic process. The other part is to use a computational modeling methodology to come up with experimentally testable hypothesis. As pointed out by Kitano in his review on the subject (Kitano H., 2002), computational and experimental systems biology complement each

other and engage in an iterative process where the quantitative predictions are constantly tested in wet laboratories and the results are fed back into the computational model to generate refined hypothesis. Although it is difficult to pinpoint the exact time when simulation of biological systems originated, it has its roots in the work done on quantitative modeling of kinetics in the period from 1900-1970. In 1952, Nobel Prize winners Alan Lloyd Hodgkin and Andrew Fielding Huxley successfully constructed a mathematical model describing the action along the axon of a neuronal cell (Hodgkin et. al., 1952), which was probably the first notable application of theoretical biology. However, lack of good quality data hindered this area of study from achieving its full potential, validating the claim that a theoretical model is only as good as the data it works with. This all changed when high-throughput experiments developed in the 1990s brought a deluge of genomic and proteomic data that could be used for quantitative modeling. When this development was coupled with a revolution in the computation technology available to scientists, numerical simulation once again topped the list in scientific discussions. An important consequence of this inclusion of computation technology in life science research was the idea of a systems level integration of biological components to quantitatively understand the exact behavior of a biological system. Opposing the traditional view, systems biology tends to analyze any biological process as a network of interacting systems. There is now increasing consensus among the scientific community that this systems level perspective is poised to answer a wide range of biological questions that have immediate consequences in areas of rational drug design, cancer therapy and personalized medicine. Several methodologies including kinetic modeling, biosimulations, predictive metabolism, data mining, and disease modeling are a critical part of systems biology. Computational systems biology can hence be defined as a part of systems biology that employs model-based approaches, to integrate data extracted from existing sources while

using mathematical techniques to provide the ability to make predictions about future experimental hypotheses (Rodriguez et al. 2010). A typical project starts with a reconstructed network map of all the components of the network which is then translated into a mathematical model (Sible et. al. 2007). Although the term 'mathematical model' can have various implications, in computational systems biology a mathematical model usually refers to a system of ordinary differential equations (ODEs) that has been formulated using the kinetic data available for every non-constant component of that system. This system of ODE is then solved using appropriate initial conditions to realize a time course evolution of each component. Once the simulation results are consistent with previously known experimental observations, this model can be used to generate predictions for future experiments. The various kinetic parameters that are a part of the ODE system can be tweaked and twiddled to represent a local perturbation in the network and the global effects due to the artificial disturbance can be observed from the simulation results. Thus the ability to simulate a mathematical model that can provide experimentally testable hypotheses is central to the systems level study of a biological process. For example, a detailed mathematical model for TNFα – NFκB signaling was recently developed in conjunction with a protein-protein interaction map to quantitatively describe the signaling mechanism (Visvanathan M., 2010). Such a model can now be tweaked around and 'played with' to mimic specific scenarios as regards to disease or drug intervention and the response in turn could lead to development of new hypotheses. Similarly, in the drug discovery process, computational systems biology based network analysis enables the action of drug targets to be considered in the context of the whole genome thus making them an important tool in comprehending the complex relationship between a drug and the target genes (Berger et.al. 2009, Materi W. et. al., 2007). In general, various aspects of biology are now being tested in accordance with system biology principles and

mathematical / computational modeling is an integral part of such an effort (Ewing G. W., et. al., 2011, Meyer-Hermann M., et. al., 2009, Groh A., et. al., 2008, Palme K., 2006).

2 MATHEMATICAL MODELING AND SIMULATION

As observed in the previous chapter, systems level investigation of biological processes holds significant potential in solving a variety of complex problems in the areas of rational drug design, cancer therapy, metabolic engineering and personalized medicine. The computational aspect of such an investigation routinely utilizes mathematical modeling techniques to test the dynamics of a biochemical system and generate experimentally testable hypotheses from them.

The intention to mathematically model a physical system is to create a quantitative representation of the role of participating species (reactants) and their interdependent behavior (reactions). The first stage of modeling is to create a stoichiometric model by extracting pertinent information regarding the proportions of reactants in the system. This information is easily obtained from network diagrams and technical literature. Next, a mathematically rigorous description of the system based on classical theory of mass action kinetics and reaction stoichiometry is obtained. This mathematical description can vary depending on the framework used for subsequent simulation of the model. In the field of biochemical reactions, two simulation frameworks, deterministic and stochastic, are more popular than others. The work by Crampin E.J. (2004) provides an excellent overview of the modeling process.

2.1 Deterministic simulation framework

In the widely popular deterministic framework, the goal is to try to represent the behavior of a homogeneous physical system by a system of ordinary differential equations (ODE) derived from the law of mass action. The biological species in such a model are usually represented in

terms of concentration (usually moles/liter) which allows the overall dynamics of the system to evolve continuously in time. Developed several decades ago, the law of mass action states that the velocity of an elementary chemical reaction (i.e. a reaction without any intermediates) is directly proportional to the product of the concentration of the reactants participating in the reaction. The constant of proportionality, called rate constant, is most often a function of the reaction environment. It logically follows that, the net change in concentration of a reactant of a biochemical system will be the sum of the changes in concentration for every elementary reaction the reactant is a part of. Mathematically, for 'm' reversible elementary reactions and 'n' reactants, the statement can be represented as,

$$\sum_{i=1}^{n} u_{i,j} S_i \leftrightarrow \sum_{i=1}^{n} v_{i,j} S_i \quad j = 1,2 \ldots m \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

where, 'u' and 'v' in equation (1) are stoichiometric coefficients of reactants and products (S) respectively. In terms of a differential equation, the net rate of change in reactant 'S' can then be written as:

$$\frac{dS_i}{dt} = \sum_{j=1}^{m} R_{i,j} \quad i = 1 \ldots n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

where, 'R' is the rate of reaction 'j'. Applying the law of mass action, we get

$$R_{i,j} = (v_{i,j} - u_{i,j}) \left[c_j \prod_{i=1}^{n} S_i^{u_{i,j}} - c_{-j} \prod_{i=1}^{n} S_i^{v_{i,j}} \right] \quad i = 1 \ldots n, j = 1 \ldots m \quad \ldots\ldots\ldots \text{(3)}$$

c_j and c_{-j} are the reaction constants for the forward and reverse reactions respectively. For example, in the following hypothetical elementary reaction:

$$A + A \xrightarrow{c} B + C$$

The rate of change of reactant 'A' in ODE form would be written as,

$$\frac{dA}{dt} = -c[A][A] = -c\,[A]^2 \quad \ldots \quad (4)$$

In equation (4), [A] is the concentration of reactant A while ‘c’ is the reaction rate constant. The negative sign indicates that ‘A’ is being consumed as the reaction proceeds. Similar equations can be written for ‘B’ and ‘C’ as well. Assuming it to be the only reaction in the system, the three differential equations (each for A, B and C) taken together would form a system of ODE for the entire biochemical network and when integrated as a function of time would yield time course trajectories for all three species reflecting the dynamics of the biochemical network. The literature offers a more detailed description of kinetic modeling in biological systems (Grima R., 2011, and Crampin E. J. et. al., 2004). Unfortunately, biochemical reactions cannot always be characterized by simple elementary kinetics and complex kinetics such as the Michelis-Menten scheme is often used to describe the reactions more accurately.

The most common situation where the reactions have to be represented by complex kinetics is when a biochemical reaction is catalyzed by an enzyme. The resulting kinetics from such an encounter are inherently non-linear and hence difficult to analyze quantitatively. The most common enzyme catalyzed reaction scheme is as described in Segel (1975):

$$S + E \overset{c_{1r}c_{1f}}{\Longleftrightarrow} ES \overset{c_2}{\Rightarrow} E + P$$

‘S’ in the above scheme is the substrate, ‘E’ is the enzyme catalyzing the reaction, ‘ES’ is the enzyme-substrate complex and ‘P’ is the product. c_{1f}, c_{1r} are the reaction rate constants for the forward and reverse reactions of the first half of the reaction while c_2 is the rate constant for the final half of the interaction. The rate of reaction can only be derived using a set of assumptions collectively known as the quasi-steady state assumption (QSSA). In the above scheme, it can be reasonably assumed that the overall reaction is limited by the rate of product formation step. In other words,

$$\frac{dP}{dt} = c_2 \cdot [ES] = rate\ of\ reaction \quad \ldots\ldots (5)$$

The rate of change of [ES] can be written as,

$$\frac{d[ES]}{dt} = c_{1f} \cdot [S][E] - c_2 \cdot [ES] - c_{1r} \cdot [ES] \quad \ldots\ldots (6)$$

The first of the QSSA states that, the concentration of the enzyme-substrate complex does not change over time and hence the first derivative, d[ES]/dt, will be equal to zero. Applying this assumption to equation (6), we get,

$$\frac{d[ES]}{dt} = c_{1f} \cdot [S][E] - c_2 \cdot [ES] - c_{1r} \cdot [ES] = 0 \quad \ldots\ldots (7)$$

Solving for [ES] gives,

$$[ES] = \frac{c_{1f} \cdot [E][S]}{c_2 + c_{1r}} \quad \ldots\ldots (8)$$

For simplicity, equation (7) can be rewritten as,

$$[ES] = \frac{[E][S]}{K_M} \quad \ldots\ldots (9)$$

where, KM is known as the Michelis-Menten constant and is equal to the ratio $((c_2 + c_{1r})/c_{1f})$

Substituting equation (8) into equation (5) gives,

$$rate\ of\ reaction = c_2 \cdot \frac{[E][S]}{K_M} \quad \ldots\ldots (10)$$

Equation (10) is an acceptable and theoretically correct form of reaction rate expression except for the fact that the transient concentration of enzyme, [E], cannot be readily measured in a laboratory. Hence for practical reasons, it is more desirable to use the [E] as a function of total enzyme concentration $[E_{total}]$. This is achieved by using the second QSSA which states that, at

any given time during a reaction, the total enzyme concentration is equal to the sum of the transient concentration and the enzyme associated with the enzyme-substrate complex. Therefore,

$$E_{total} = [E] + [ES] \quad \ldots\ldots \quad (11)$$

Substituting [E] in terms of [ES], as obtained from (11), in equation (9) gives,

$$[ES] = \frac{\{[E_{total}] - [ES]\}}{K_M} \cdot [S] \quad \ldots\ldots \quad (12)$$

Solving equation (12) for [ES] gives,

$$[ES] = \frac{[E_{total}] \cdot [S]}{[S] + K_M} \quad \ldots\ldots \quad (13)$$

Hence, the rate of reaction becomes,

$$rate\ of\ reaction = \frac{c_2[E_{total}] \cdot [S]}{[S] + K_M} \quad \ldots\ldots \quad (14)$$

The third and final QSSA states that, the concentration of enzyme, [E] is far less than the concentration of substrate, or [S] >>> [E], thereby making the maximum reaction rate limited by the total concentration of enzyme ([E_{total}]). So the maximum reaction rate (V_{max}) equals the product of c_2 and [E_{total}]. With these adjustments, the rate equation finally takes form as,

$$rate\ of\ reaction = \frac{[V_{max}] \cdot [S]}{[S] + K_M} \quad \ldots\ldots \quad (15)$$

Equation (15) is known as the Michelis-Menten equation and quantitatively relates the initial rate of reaction with the substrate concentration for enzyme catalyzed biochemical reactions. This is the simplest reaction scheme for enzyme catalyzed reactions. Other enzyme mediated reactions include competitive, uncompetitive and non-competitive enzyme inhibition as well as multi-substrate reactions. Needless to say, the rate expressions for those schemes are much more complicated than equation (15) but are nevertheless based on Michelis-Menten kinetics. Finally, there are some enzymes which do not follow Michelis-Menten type of kinetics and a separate

class of kinetics called sigmoidal kinetics is used to describe them. An expression called the Hill equation is used to represent the dynamics of such enzymes.

$$\theta = \frac{S^n}{k_d + S^n} = \frac{S^n}{K_H^n + S^n} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots .. (16)$$

In (16), θ is the fraction of binding sites occupied on the enzyme. ‘S’ is the substrate and kd is the dissociation constant for the enzyme.

Enzymes obeying the Hill equation are generally allosteric in nature and promote co-operative substrate / ligand binding. A parameter known as the Hill co-efficient (n) determines the level of co-operation in binding. It can be clearly seen from the situations above that a system of ODE derived from such kinetics would rarely exhibit a linear relationship and as such its analytical solution may be extremely difficult to attain if not impossible. In such situations, numerical methods have to be employed to approximate a solution of the ODE system. Two main classes of numerical methods are described below:

2.1.1 Explicit numerical methods

The most intuitive and straightforward methods to solve an ODE are the explicit methods belonging to a family of methods known as the Runge-Kutta methods. Euler’s method (a.k.a 1^{st} order Runge-Kutta method) is the simplest explicit numerical solution available to solve an ODE. The mathematical problem can be stated as an initial value problem: Given an ODE of the form,

$$\frac{dy}{dt} = f(y), \quad y(t_0) = y_0 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots .. (17)$$

it is required to find an approximate value of the function ‘y’ at time t_1 so as to ‘simulate’ the exact solution of equation (17). If we consider the limit of the differential as $\Delta t \rightarrow 0$,

$$\frac{\Delta y}{\Delta t} = f(y) = \frac{y_1 - y_0}{t_1 - t_0} \quad \ldots\ldots (18)$$

$$y_1 = y_0 + h \cdot f(y_0) \quad \ldots\ldots (19)$$

In general, for i = 1, 2 … n

$$y_{i+1} = y_i + h \cdot f(y_i) \quad \ldots\ldots (20)$$

Equation (20) is the working equation for Euler's method, where, y_{i+1} is the value of function 'y' at time t_{i+1}; y_i is the value of 'y' at time t_i (which is known from previous step); 'h' is the arbitrarily chosen constant time step equal to $t_{i+1} - t_i$ and f (y_i) is the kinetic function derived from the law of mass action. Clearly, equation (7) has to be repeated for 'n' steps covering the entire time period of simulation, T = n x h.

Although Euler's method is extremely straightforward to implement as an algorithm, it is also the most impractical of all numerical methods. As can be clearly seen in the above derivation, the method is most accurate when Δt (or 'h') → 0, indicating the necessity of a very small time step for it to deliver an acceptable approximation to the exact solution. In doing so, because the period of simulation T is constant, the number of iterations 'n' can become very large (as h → 0, n → ∞). Although the local truncation error per step is proportional to h^2, it can be an accuracy nightmare when accrued over large number of iterations. Also, because of the limitation on the step size, this method is computationally inefficient.

The mid-point method (a.k.a second order Runge-Kutta) is more accurate than the Euler method. Assuming the problem statement to be the same as before, the equation for the mid-point method is,

$$y_{i+1} = y_i + h \cdot f(y_{h/2}) \quad \ldots\ldots (21)$$

To be able to use equation (9), an additional step to evaluate $y_{h/2}$ has to be performed.

$y_{h/2} = y_i + \frac{h}{2} \cdot f(y_i)$ ……………………………………………………………………… (22)

Equation (22) is used to evaluate function f ($y_{h/2}$) which in turn enables the use of equation (21).

The advantage that mid-point method has over the Euler method, in terms of increased accuracy, comes at a steep price in computational cost. It can be clearly seen that the use of an additional step per iteration essentially doubles the computational effort while not providing any significant improvement in accuracy (local error is of order h^3 compared to h^2 for Euler) or step size. Unless accuracy is of extreme importance, mid-point method is generally not a good choice to solve an ODE. Also, even if accuracy is more important, 4th order Runge-Kutta method is generally twice as accurate as mid-point method.

Runge-Kutta 4th order (RK4) is the most popular explicit method available to simulate a system of ODE. The working equation for this method can be written as,

$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$ …………………………………………………… (23)

$k_1 = h \cdot f(y_i)$ …………………………………………………………………………… (24)

$k_2 = h \cdot f(y_i + \frac{k_1}{2})$ ……………………………………………………………………… (25)

$k_3 = h \cdot f(y_i + \frac{k_2}{2})$ ……………………………………………………………………… (26)

$k_4 = h \cdot f(y_i + k_3)$ ……………………………………………………………………… (27)

RK4 has a fourth order global error (O (h^4)) and a local truncation error of O (h^5). It is one of the most accurate explicit methods because of the additional 'k' factors which essentially work as correctors for the slope obtained for Euler's method. RK4 also has a problem with being computationally intensive but its major criticism stems from its inherent inability to handle 'stiff' systems of equations.

2.1.2 Implicit numerical methods

It is common in biochemical systems to have some reactions operate at a vastly different time scale than other simultaneous reactions. This disparity in time scales translates mathematically into what is described as 'stiffness' in the ODE system. Although an exact definition of stiffness does not exist, it is generally observed when the solution of an ODE does not change significantly over time but attempting to increase the step-size in order to speed up the simulation causes the solution to become unstable. Explicit methods, regardless of their accuracy, are prone to instability when applied to stiff equations. Implicit methods, on the other hand, are unconditionally stable and generally faster.

The backward Euler formula is the most basic implicit method available. Again considering the problem statement to be the same as in section 2.1.1, the solution equation can be written as,

$$y_{i+1} = y_i + h \cdot f(y_{i+1}) \quad \text{(28)}$$

A quick comparison of equation (28) with equation (20) reveals that the earlier expression is 'explicit' in the right hand side where all the terms are known beforehand. In (28) however, the function 'f' has to be evaluated at time t_{i+1} before proceeding. There are several ways to handle this challenge but the most common is to replace it with a linear approximation obtained from Taylor series.

The problem with backward Euler is the same as with the forward Euler method described in 2.1.1. Although, backward Euler is stable for stiff problems, its accuracy is limited. Also, it is much more sensitive to its step-size than other implicit methods. Computationally, backward Euler is difficult to encode and the marginal improvement in running time is usually not worth the effort.

One of the more popular implicit methods is the Adam-Moulton 2nd order method (aka trapezoidal rule) whose equation is shown below.

$$y_{n+1} = y_n + \frac{h}{2}\left[f(y_{n+1}) + f(y_n)\right] \quad \text{(29)}$$

This method enjoys several advantages including stability, speed of computation (because of large step sizes) as well as accuracy (being a second order method). Its accuracy can be drastically improved by combining an explicit method to 'predict' the solution and then 'correct' it using the trapezoidal method. Such methods are called predictor-corrector methods and are extremely useful for differential equations with complex functions.

2.2 Stochastic simulation framework

Deterministic methods have been extremely useful in modeling biochemical systems and have been used by physical scientists to model cellular behavior for several decades. These methods, however, are not without limitations (Wilkinson 2006). First, there exist several biochemical systems where the number of interacting reactants is extremely low. A gene regulatory system is an excellent example of such a system. The concentration of gene molecules in these systems is so low compared to the overall volume of the cell that using concentration units to predict the dynamical behavior of the system usually leads to erroneous results. Secondly, some biological systems are bi-stable and deterministic simulation in general fails to describe the dynamics of such systems (Zhang 2010). More importantly, random fluctuations could be physiologically important to the dynamics of the overall cell. For example, failure to capture random fluctuations in genetic regulation might interfere with the prediction of protein translation dynamics. Thus random behavior in certain situations can actually be a desired property rather than anomalous behavior.

2.2.1 Exact stochastic simulation

It is for the need to accurately simulate the biochemical network that stochastic simulation framework was introduced in 1977 by Daniel T. Gillespie. The underlying logic of a stochastic framework is that biochemical reactions occur due to the random collisions between two or more reacting species. The biochemical species in a stochastic model are represented in terms of discrete number of molecules instead of concentration and hence the entire system dynamics can be updated discretely rather than continuously as in a deterministic framework. The reaction rate parameters in this framework are linearly related to the deterministic kinetic parameters and serve as hazard functions that identify the probability of a reaction event occurring. If and when a reaction event occurs, only the reactants corresponding to that particular reaction are updated while others are left unaltered. Multiple runs of this random procedure yield a mean value for the state of the system which is then reported. As can be clearly seen, because the system is allowed to evolve discretely in time, the dynamics of species present in lower numbers of molecules (and hence lower concentration) can be effectively captured. From a mathematical standpoint, the stochastic framework attempts to adopt a Monte Carlo approach to solve the chemical master equation (Gillespie D. T., 1992).

$$\frac{\partial P(n, t\,|n_0, t_0)}{\partial t} = \sum_{\mu=1}^{M}\left[c_\mu h_\mu(n - v_\mu)\, P(n - v_\mu, t\,|n_0, t_0)\right] - \sum_{\mu=1}^{M} c_\mu h_\mu(n) P(n, t\,|n_0, t_0)] \quad . (30)$$

Equation (29) describe the rate of change of probability of system ‘n’ having ‘M’ reactions each having parameters ‘c’ and ‘h’. The solution of the equation (30) as obtained by the stochastic simulation algorithm (SSA) is a probability density function that is a product of two mutually exclusive and statistically independent random events. The first event is the random selection of time step at which the chemical systems evolves and the second event is the random

selection of the reaction event that brings about this evolution. The probability functions for the first event (p_1) is an exponential decay function with a decay factor of 'a' which is a function of the parameters 'c' and 'h'. The function for the second event (p_2) is a linear function of 'c' and 'h' and is similar to the one used by the roulette wheel selection process of a genetic algorithm.

$$p_1 = a_\mu e^{-a_0 \tau} \ and \ p_2 = \frac{c_\mu h_\mu}{a}$$

As the two events are mutually exclusive and independent, the total probability will be given by $P = p_1 \cdot p_2$.

In this way, the next reaction chosen is always the one having either the most number of molecules or the one that has the highest propensity based on reaction rate constant or both. On the other hand, the time step is selected from a continuous function rather than from a discrete one and that makes this method an 'exact' solution to the chemical master equation (Kierzek A. M., 2002). The algorithm is implemented by the following steps:

- Initialize the system with number of reactions, number of initial molecules for every reaction, the type of reaction and the parameters 'c' and 'h' for each reaction.
- Use the parameters to calculate the hazard functions for each reaction.
- Pick two random numbers from a uniform distribution.
- Use the first random number to generate a time step τ according to the distribution p_1
- Use the second random number to generate the index for the next reaction according to p_2
- Update the system according to the reaction stoichiometry.
- Update the hazard functions for each reaction.
- Repeat the above steps until end of time T

The algorithm has to be run several times to produce trajectories that generate statistically significant results.

2.2.2 Approximate stochastic simulation

Computational time can be a drawback for exact stochastic simulation. To tackle this problem, methods that attempt to combine the best of both deterministic and stochastic frameworks called hybrid methods are developed (Salis, 2006). Hybrid methods operate on the premise that any reaction system can be categorized into a subset of 'fast' (high concentration) and 'slow' (lower concentration) reactions. Both subsets are then simulated simultaneously using the appropriate simulation method i.e. deterministic for 'fast' and stochastic for 'slow' reactions. For a more information on hybrid methods, the work by Bentele M, (2004) provides a much detailed description. Hybrid methods tend to sacrifice accuracy of simulation results for gain in computational speed and are hence known as approximate methods. Hybrid methods are not the only approximate stochastic solutions available. The Tau-leaping algorithm and slow-scale stochastic simulation algorithm are just a couple of examples of attempts to provide approximate solutions to stochasticity. For further details the reader is referred to a review by Daniel T. Gillespie (Gillespie D. T., 2007).

3 DEVELOPMENT OF THE CROSSOVER METHOD

3.1 Why develop a new computational method?

3.1.1 The need for multi-scale modeling

Studying biology at a systems level to investigate emergent network properties encompasses a wide spectrum of time scales and physiological detail and as such mandates the use of multi-scale modeling techniques to construct mathematical models (Meier-Schellersheim M. et. al., 2009). For example, a seemingly trivial problem of testing the role of a particular enzyme in an organism would have to include description of genetic regulation and protein-protein interaction at an intracellular level as well as cellular dynamics at a cell population level. The corresponding time scale of reactions can possibly range over a few orders of magnitude and addressing this layer of complexity is imperative for effective systems biology.

Multi-scale modeling is typically implemented via either a bottom-up or top-down method. A bottom-up method starts with the molecular interactions at an intracellular level and makes its way to the physiological function while top-down method goes the other way. Both methods have their own benefits. While bottom-up methods seem to be an intuitive approach to build a model, the biochemistry of a significant number of biological processes is not very well characterized at the cellular level. This forces the model building efforts to stall at the very elementary stage even if the ultimate physiological phenotype is well understood. Conversely, if the biochemical reaction parameters are known, a bottom-up approach can be used to construct a comprehensive model to generate meaningful experimental hypotheses about the physiological function. A top-down approach, on the other hand, is designed to take an observed higher level phenotype and make testable hypotheses about its underlying molecular mechanism. In this approach, a model is first built to replicate the higher level observations and then compounded with

additional details at every relevant layer of biological complexity until the lowest level of detail is reached. Hypotheses are generated and tested systematically at every level to determine the optimum path to the following layer. This approach, because it starts from the actual observation, is reliable in regenerating the phenotype. However, at sub-cellular level, there are several channels to achieve a desired experimental observation and therefore it is difficult to pin-point the exact pathway or mechanism responsible for the observed phenomenon. A typical systems biology project actually involves a combination of both approaches and the experimental output to form a hybrid solution. Regardless of the approach used, it is clear that a systems biology application has to account for multi-scale complexity as it is the only way to correctly investigate emergent network properties.

The issue of multi-level complexity becomes even more critical for studying complex diseases where the mechanism of pathogenesis itself is complex across biological strata (Vicini P., 2010, Dewey F. E., 2011, Hatzikirou H., 2011). One of the most insightful descriptions of this issue is observed while studying genesis and metastasis of cancerous cells (Hatzikirou H., et. al., 2011). As tumor cells originate from a single cell but grow to form a mass of tissue made up of an ensemble of various cell types, it is imperative to focus on the subcellular scale during modeling tumor genesis. At this scale, in addition to carcinogenic biochemical mechanisms, one must also consider other factors that influence cancerous behavior such as epigenetic regulation. Next, at the cellular level the interactions between tumor cells and their microenvironment such as intercellular signaling and transport become dominant and have to be accounted for in the model. Although the spatial scales at these levels do not present any complexity, the time scales do and the modeling framework has to be able to handle it. Finally, at the tissue scale, the spatial scales become important as do the time scales and thus provide the ultimate complexity from a model-

ing perspective. Using agent based modeling in conjunction with sensitivity analysis to predict therapeutic targets for cancer is another recent area where multi-scale modeling is applied (Wang Z., et. al., 2011). The use of multi-scale modeling is also relevant in the area of diabetes research. For example, the interplay between insulin secretion on a physiological scale and the inter-cellular events on a higher scale can be understood by simulating models incorporating an integrated scale of detail (Pederson M.G., et. al., 2011). Most of the common heart diseases like myocardial ischemia and arrhythmia can be studied most effectively by looking beyond the simple genetic association. Building integrative models of mechanism of ventricular arrhythmia in a healthy as well as diseased heart, models for initiation of arrhythmia and image-based models are all examples of extensive modeling efforts spanning multiple scales (Trayanova N. A. et. al., 2009). Such efforts not only help understand the complex mechanisms of heart disease but also allow testing the efficacy of antiarrhythmic drugs at multiple biological levels (Dux-Santoy L. et. al., 2011).

The success of systems biology as a paradigm is inherently coupled with the necessity to develop integrative multi-scale models. Complex diseases, especially, can be studied effectively only with such a multi-scale approach. While it may seem intuitive to use multi-scale models as an antithesis to reductionism, it is still a fairly complicated process which is not very well understood. A bigger problem, however, is the lack of an appropriate computational framework to simulate such a multi-scale model. To encourage researchers into applying systems biology, it is absolutely essential for the computational biology community to provide them with the tools that make the effort of building multi-scale models worthwhile. As discussed earlier, building a quantitative model and its subsequent simulation are tied together and one is useless without the other. While one can argue in favor of developing a simulation framework to handle mathematical

complexities in isolation from building a model, the reverse is not true. It is quite likely that the frequent influx of novel experimental data from high-throughput biological experiments will have a qualitative impact on the nature of the model being simulated. This would require having to rebuild the model from scratch as none of the current infrastructure supports partial modification of any model. For real time systems biology applications that have the capability to support experimental research, it is highly imperative that an appropriate simulation mechanism be developed that can inherently adapt to the complexities presented by any biological model multi-scale or not.

3.1.2 Inadequacy of existing simulation frameworks

As discussed in the previous chapter, a plethora of integration methods employing deterministic, stochastic and hybrid techniques exists to simulate mathematical models. However, their suitability for systems biology, particularly in context to multi-scale modeling, is woefully inadequate. Here we look at the most popular methods and discuss how their shortcomings can limit the potential of systems biology.

Limitations of deterministic methods:

The continuous deterministic methods based on ODEs are the most popular methods to simulate a biological network. All methods in this family are easy to implement and almost all physical scientists are familiar with them. These methods are generally robust and can handle a wide range of complex kinetic expressions. In fact, these methods would be a perfect fit for systems biology if not for one glaring disadvantage: deterministic methods cannot replicate the random fluctuations that are ubiquitous in biomolecule copy numbers. Such randomness also referred to as stochasticity or ‘noise’ is known to play very important role from a biological point of view. In the last decade a considerable body of work has been dedicated towards understand-

ing how noise at genetic and molecular level leads to variability in gene expression and hence protein translation (Stewart-Ornstein J. et. al., 2012, Munsky B. et. al. 2012). The effects at molecular level are often transduced to the metabolic level and therefore it is critical for any simulation method to acknowledge the existence of such noise. Logically it could be deduced that in multi-scale modeling, where the models often start with a description of lower level phenomenon, ignoring stochasticity can lead to erroneous results further up the levels. Deterministic methods are fueled by kinetic rate expressions which do not change with time or with concentration and hence output a smooth solution to the system of ODE. The parameters of the kinetic expressions are often a representation of the average behavior of the system and hence the noise within the system is impossible to capture by pure continuous functions. The amplitude of these random fluctuations is generally not very high and at higher concentration of biomolecules the fluctuations may not make a significant impact thereby making the deterministic solution closely match the experimentally observed values. At lower concentration, however, the fluctuations can appear dominant and a deterministic solution may not seem satisfactory. For example, if the numbers of molecule X are fluctuating between 1000 and 1001, a deterministic value of say 1000.30 seems like a close match. If, however, a molecule Y is varying between 1 and 2, the deterministic solution of say 0.95 seems insufficient and maybe even incorrect.

To address this issue, there have been efforts directed towards using stochastic differential equations (SDE) to model the biochemical reactions. While a detailed discussion about SDEs is beyond the scope of this work, it is critical to note that SDEs act by simply adding a ‘noise’ term to the regular differential equation and integrating it as an *Ito* differential term. The results are encouraging at higher concentration where the output does become noisy but at very low

concentration of less than 3 molecules, the results are generally inaccurate. ODE or SDE based methods are therefore unsuited for systems biology based applications.

Limitations of stochastic methods:

One of the key developments in the 1970s, with regards to biochemical simulation, was the development of a Monte Carlo method to accurately solve the chemical master equation. Popularly called the Gillespie algorithm after its developer Daniel T. Gillespie, this method predicts which biochemical reaction, from a given biochemical system, is "most likely" to occur and the time at which it will occur. The fundamental idea behind this method is that if we reduce any biochemical system to set of elementary molecular reactions it is possible to simulate, within accepted statistical boundaries, the Brownian motion of molecules. In this way, it ensures that any random fluctuations that might arise because of the random motion of molecules are effectively captured.

There are two critical problems, however, regarding its applicability to systems biology. First, the Gillespie algorithm derives its statistical parameters from the available kinetic data for the simulated reactions. There is an innate assumption here that these kinetic data are readily available for every single reaction and that they are accurate. This may not be always true. Experimental determination of the kinetic parameters is sometimes not possible for certain reactions while the ones that are possible might have experimental errors in them. The quality of stochastic simulation trajectory is then dependent on the quality of kinetic data available just as in the deterministic method. During the process of modeling if the modeler encounters missing kinetic data, it is a common practice to lump the missing reactions together and generalize its behavior with an approximate kinetic expression. The Michelis-Menten kinetic expression is an excellent example of such an approach. The Gillespie algorithm cannot be used in the case of Michelis-

Menten type of kinetics without significantly sacrificing the accuracy of the solution. It is reasonable to assume that in a complex multi-scale model, Michelis-Menten type of kinetic rate expressions would be more of a norm than exception and the stochastic method will fall far short of need.

Second, and more important, issue is with the computational efficiency of the Gillespie algorithm. This method updates the biochemical system one reaction at a time based on the propensity of the reaction to occur. This is a major handicap for system with large number of reactions especially when a system scales to a larger size. Another issue is the manner in which the time step is chosen for the occurring reactions. As shown in figure 2, the algorithm randomly picks a number from an inverse exponential distribution. The scaling factor (λ) of the distribution depends upon the kinetics of the reactions and hence the time step cannot be adjusted to make the simulation run faster. Any attempt to select a bigger time step in essence jeopardizes the accuracy of the method.

Moreover, stochastic simulation methods are not built to handle 'stiff' differential equations. Stiffness in an ODE system occurs when a majority of the reactions evolve very slowly (low reaction rate) compared to a few fast ones (high reaction rate). As a result, the time evolution of the system slows down considerably. The reason for this slow down can be explained with the help of figure 2.

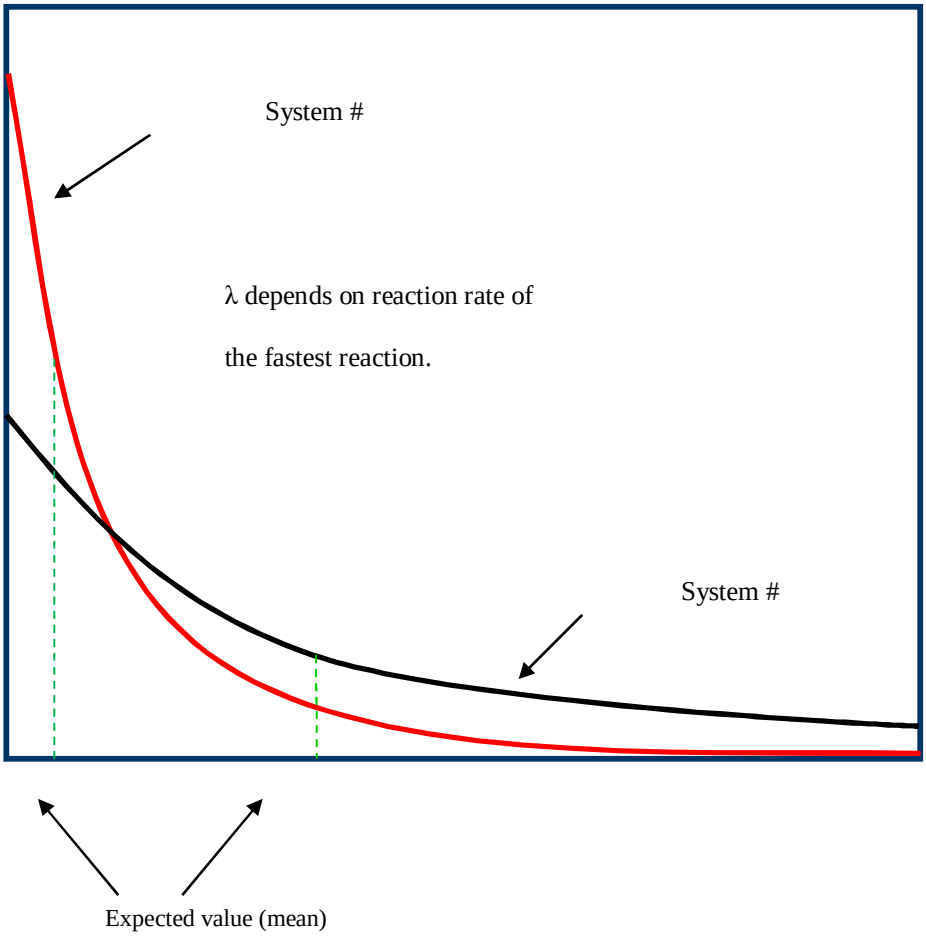

Figure 2 Comparison of the probability distributions for two hypothetical biochemical systems used for predicting the time-step in the Gillespie algorithm. System # 2 evolves slower than system # 1.

The two curves in figure 2 represent the probability distribution of the time step for systems 1 & 2. Assume that the two biochemical systems are evolving independently of each other. Also, assume that system # 2 has a reaction that evolves much faster than others making the scaling factor λ higher than system # 1 which consists of reactions with rate constants in the same order of magnitude. The Gillespie algorithm randomly chooses a value (μ) from this distribution as its time step for the respective system. The expected value or the value most likely to be chosen for the distribution can be clearly seen to be larger for system # 1 than system # 2. In other words, system # 1 is more likely to choose a larger time step and hence evolve faster than system

2. In the above example, system # 2 can be described as being “stiff” and is a common occurrence in multi-scale models.

Limitations of hybrid methods:

The accuracy afforded by pure stochastic methods such as the Gillespie algorithm is important for a variety of simulation based applications. To overcome the drawbacks outlined in the previous section, there has been a rapid development of hybrid methods. These methods can usually be divided into two categories: The algorithms in the first category classify the biochemical system into groups of fast and slow reactions and treat them with either deterministic (fast) or stochastic (slow) methods. The problem with such a classification is that it is not entirely clear as to what criteria is suitable to justify such a division. Also, in some instances it might be important to study fluctuations for fast reactions and it would be impossible to do so because they are solved with a continuous method. Additionally, for stiff systems only a few reactions are fast which means the total computational efficiency is still controlled by the slow stochastic reactions thereby nullifying the apparent speed up in performance. The other category consists of methods that modify the time step calculation in the Gillespie algorithm. These methods generally referred to as ‘Tau-leaping’ methods approximate a time step for the Gillespie method without losing the statistical significance of the outcome. The limitation of such an approximation is twofold: 1) the use of an “approximation” step for a stochastic algorithm is essentially a sacrifice of accuracy for gain in performance. This loss of accuracy defeats the purpose of using stochastic simulation in the first place. 2) Even if we accept that the loss of accuracy is minimal, which is true for some cases, question marks still remain on whether the performance enhancement is scalable with the size of the system. Most of the Tau-leaping methods are tested on smaller hy-

pothetical biochemical pathways and so their adaptability for large systems biology applications is open for debate.

To summarize, none of the methods described above really have an all-round suitability when it comes to simulating multi-scale systems biology models. It is extremely important for the integration method to be unbounded by the issues of spatial and temporal complexity which is a hallmark of multi-scale models. At the same time, in spite of state-of-the-art computational infrastructure available, the computational performance cannot be ignored. It worth noting, however, that none of these methods were initially developed with systems biology as a focal point and the scientific community as a whole is merely trying to adapt them to suit the new paradigm of systems biology. In that context, it would be more prudent to develop a simulation technique with the sole focus on simulating integrative models.

3.2 The crossover method

3.2.1 Rationale

The lack of a single unified integration method for solving the ODE system of multi-scale models is a hindrance in the path of exploiting the full potential of systems level biology. A quick glance over the various available methods reveals that, the deterministic method is the only method that comes close to the ubiquity desired in multi-scale simulation. In addition, the computational performance of deterministic method is beyond any debate as implicit methods allow a system to be simulated with large time steps. Given this background it is reasonable to propose that if the deterministic method is somehow manipulated to include stochastic fluctuations, at least qualitatively, then it can be an ideal choice for simulating a multi-scale model. Hence, the most important questions to be asked in this investigation are: can we develop a method that operates within the premise of the deterministic framework yet outputs a trajectory that includes

random fluctuations? If we can, will it be computationally faster than the stochastic simulation algorithm? The formulation of such a method is not trivial as there are several ramifications to introducing a deliberate modification. In the following section, a detailed analysis of how such a task could be achieved is presented in terms of the development of “the crossover method”.

3.2.2 Methodology

Concept of a Bernoulli trial:

The conceptual basis of the crossover method lies in an event termed as a Bernoulli trial. A Bernoulli trial or event is a statistical experiment the outcome of which has only two possible values. Figure 3 shows the graphical depiction of Bernoulli trial examples. Every such trial or experiment is an independent event that can be repeated any number of times without affecting the outcome of previous or future events. The outcomes always have a fixed probability of occurring and that value is independent of the number of trials conducted. A coin flip is the best example of a Bernoulli trial. When a coin is flipped only two outcomes are possible: a heads or a tails. Each result has a constant probability of 50% or 0.5 of coming up each time. Another example would be that of rolling a dice. A roll of a dice has only two possible outcomes: either the desired number comes up or it doesn’t. The probability of the desired number coming up each time the dice is rolled (positive outcome) here is 1/6 and the probability of something else coming up (negative) is 5/6. These probability values however are independent of each trial which means regardless of how many times the dice is rolled, the outcome probability remains constant.

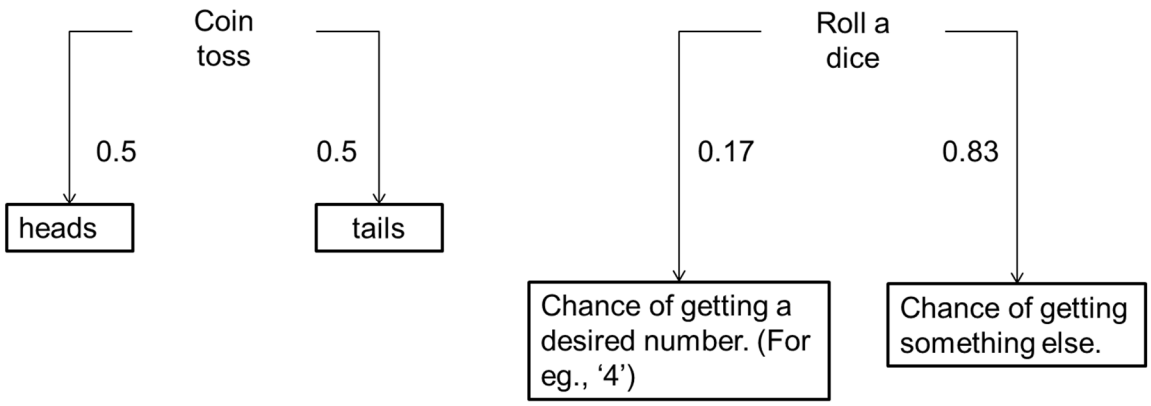

Figure 3 Illustration of a Bernoulli trial. A coin toss and a roll of dice are both examples of a Bernoulli event.

Modification of the deterministic framework with Bernoulli trial:

The rationale behind the crossover method is actually quite straightforward. It is common knowledge that a deterministic method generates trajectories which are continuous curves made up by real numbers. One way to force a fluctuation in an otherwise continuous function is to round up or down a fractional real number encountered at every time step. This process of rounding is not a trivial step because of its mathematical implications. The crossover method achieves this by performing a Bernoulli trial at every time step of the deterministic solution where the outcome of the event is either rounding up (success) or rounding down (failure) the real number. It can be shown that if a continuous state variable is to be replaced by its discrete equivalent the only way to do it is to have a series of Bernoulli trials with probability being the fractional part of the real number. This can be verified by elementary statistical theory. If large enough Bernoulli trials are conducted with the fractional part of the real number, the binomial distribution converges to a normal distribution and the expected value of such a distribution comes out to be the average of all outcomes which will be the original real number itself.

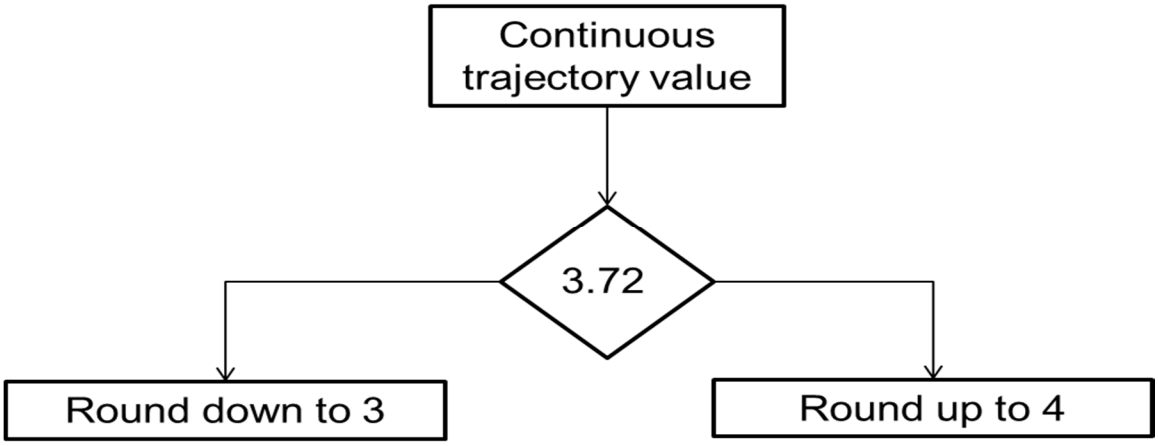

Figure 4 Illustration of the use of Bernoulli trial to modify a deterministic framework.

Consider the example in figure 4. Say a continuous trajectory value of 3.72 needs to be discretized. A Bernoulli trial is conducted with the fractional part i.e. 0.72 as the probability and an outcome of either 3 or 4 is obtained. If we conduct enough trials and look at the average expected outcome, it would be 3.72. Thus, the crossover method essentially replaces a continuous value by a set of discrete integers averaging to the continuous value.

Mathematical basis:

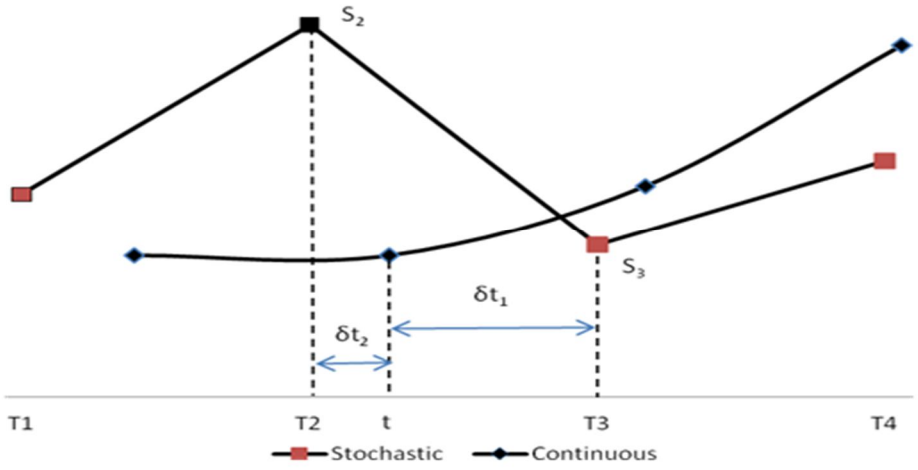

Figure 5 Schematic comparison of continuous and exact stochastic solutions to a system of differential equations. δt_1 and δt_2 are the errors introduced in the time step 't' of the continuous solution by the crossover method.

The simplistic description of the crossover method begs for a more formal and mathematically rigorous definition. There are several ways to do so: First, observing the two trajectories of figure 5, it can be noted that, assuming both trajectories are solutions of the differential equations, introducing error by rounding up or down the deterministic number is equivalent to introducing an error ($t+\delta t_1$ or $t-\delta t_2$) in the time step of the numerical method. For the solution of the crossover method to be one of the true solutions of the differential equation, the 'rounded' step size should correspond to a solution of the chemical master equation which is obtained by the stochastic simulation algorithm. The standard error in such a rounding is

$$\frac{\partial t}{\left|\frac{\partial f}{\partial t}\right|} \leq \frac{1}{\left|\frac{\partial f}{\partial t}\right|}$$

Since the probability function for a Bernoulli trial is a binomially distributed, the variance is given by,

$$\sigma^2 \propto \frac{1}{N}$$

So the standard error will be,

$$\sigma \propto \frac{1}{\sqrt{N}}$$

For a large number of Bernoulli events, i.e N $\rightarrow \infty$, this error will be,

$$\lim_{N\to\infty} \frac{1}{\left|\frac{\partial f}{\partial t}\right|} \frac{1}{\sqrt{N}}$$

$$= 0$$

Thus, as number of trials increase, the error associated with this modification goes down and eventually approaches the exact stochastic result. In other words, the solution of a crossover

method is mathematically viable and bounded by the deterministic solution on one end and the stochastic on the other.

More support can be obtained is found in the theory of stochastic differential equations. A stochastic differential equation is of the form,

$$\frac{dS_i}{dt} = f(s,t) + (noise) \quad \ldots\ldots \quad (31)$$

Where S_i is any variable and f (s,t) is the function to be integrated. The RHS of equation (31) is made of two terms: a regular function that is to be integrated (i.e. f(s,t)) and a noise term to add stochasticity. The equation can then be integrated as

$$dS_i = f(s,t) \cdot dt + W(t) \cdot dt \quad \ldots\ldots \quad (32)$$

$$\int_{S_n}^{S_{n+1}} dS_i = \int_{t_n}^{t_{n+1}} f(s,t) \cdot dt + \int_{t_n}^{t_{n+1}} W(t) \cdot dt \quad \ldots\ldots \quad (33)$$

Where, W (t) in equation (33) is the white noise. The first term on RHS is a regular integral while the noise term integral is called an “Ito integral”. A detailed discussion on Ito calculus is beyond the scope of this work, however, the integral can be solved to give a product of square root of the time step and a normally distributed random number. So the solution to the integral looks like,

$$\Delta S_i = \int_{t_n}^{t_{n+1}} f(s,t) \cdot dt + \sqrt{\tau} \cdot N(0,1) \quad \ldots\ldots \quad (34)$$

Equation (34) is known as the Chemical Langevin equation and the second term on RHS is the mathematical representation of the Brownian motion. In the case of crossover method an equivalent expression looks like,

$$\Delta S_i = \int_{t_n}^{t_{n+1}} f(s,t) \cdot dt + Binomial \quad distribution \quad \ldots\ldots \quad (35)$$

While a binomial distribution can never be confused with Brownian motion, an argument can be made that with large number of trials, a binomial distribution can be approximated by a normal distribution. Thus, although not an exact solution but crossover method does utilize a mathematically valid avenue to introduce randomness in the solution.

Finally, there is the concept of shadowing lemma which can be borrowed from dynamical systems theory. In a chaotic system, a solution generated by a numerical method always has rounding off errors in it. Therefore, to ascertain the validity of such a solution it has been proved that for a fixed bounded error ε at every time step of a numerical method, there exists a true solution that 'shadows' the numerically generated one (Grebogi C et. al., 1990). In case of the crossover method, the deterministic solution can be thought of as the true solution that shadows the one generated by crossover method.

Algorithm development and analysis:

The fundamental premise of the crossover method rests on the fact that deterministic methods are capable of accurately simulating biochemical networks across a varying spectrum of species concentration and time scales provided that its accuracy at lower concentration is not compromised. As will be seen, this requirement can be satisfied by incorporating a controlled degree of randomness in an otherwise fully deterministic simulation. The overall method can be partitioned into two stages: In the deterministic stage, the rate of change of a reacting species, A, can be given as,

$$dA/dt = \sum f(k_i, [R]_i) - \sum f(k_j, [R]_j) \quad \text{..... (36)}$$

Where 'i' is any reaction producing species A and 'j' is any reaction consuming 'A'. 'k' is the reaction rate constant of the respective reaction and [R] is the concentration of the reactants for that reaction. 'f' is the deterministic function that is dependent on the type of the reaction. For

simplicity, if we consider a system where i = j =1 and the reaction is first order, the equation can be rewritten as,

$dA/dt = f_1(k_1, R_1) – f_2(k_2, R_2)$ ..... (37)

The integral of the above ODE for a known time interval will yield the solution of 'A' at the end of that interval. As most of the functions of interest on right hand side of the above ODE are analytically intractable, numerical methods are employed to approximate the solution of 'A'. In the present work, we have used forward Euler's method to maintain simplicity and ease of implementation of the solution. In Euler's method, for a given infinitesimal time step dt, approximated by $\Delta t = t1 – t0$, the change in concentration of A can be given as $\Delta A = A1 – A0$. The concentration of reactants can also be expressed as number of molecules (N), if the reaction volume is known. So the ODE can now be changed to,

$\Delta A/\Delta t = f_1(k_1, N_1) – f_2(k_2, N_2)$ ..... (38)

$\Delta A = f_1(k_1, N_1)* \Delta t – f_2(k_2, N_2)* \Delta t$ ..... (40)

$\Delta A = D_1 – D_2$ ……..…….…. (41)

where D equals the product $f(k, N)* \Delta t$ and signifies the partial change in species A due to a single reaction. Clearly, summation of all partial changes in species A will eventually yield the net change in 'A' for the given time step. So,

$A1 = A0 + \sum D$ ……..… (42)

It can also be noted that the function f(k, N) is a continuous function and hence the result of D will be a continuous value (real number) as well. This is undesirable from a physical standpoint because it is unreasonable for the number of molecules of any species to be anything other than discrete integers. This requirement necessitates the inclusion of a stochastic effect to the solution.

The second stochastic stage simply seeks to determine the integer value that D might attain based on the fractional value predicted by the deterministic step. This is achieved by conducting a Bernoulli trial with the fractional part of the real number, obtained in the previous step, as the probability of success. A true outcome of the trial rounds up the real number of D to the nearest integer and a false outcome rounds it down. For instance, if the real number obtained for D1 is 1.26 then, a Bernoulli trial with probability 0.26 will be conducted. A true outcome will assign D1 with 2 and a false outcome will assign it with 1 thereby ensuring D1 to be always an integer. It can be clearly seen that the physical significance of such a Bernoulli trial is to determine whether or not a reaction has occurred as 'D' represents partial change in the state of species 'A' due to a single reaction. The method can be generalized in the following simple algorithm.

- Formulate the deterministic functions (based on kinetics of the reaction) for all reactions in the network.
- At time t = 0; Initialize the reactant species with their initial number of molecules.
- Assume a time step small enough for using Euler's method.
- Compute D values for all reactions based on their deterministic functions and the time step.
- Conduct Bernoulli trials on all 'D' values and update the D values based on outcome of the trial.
- Update reactant species by summing up their respective D values and report the results.
- Repeat steps 4 thru 6 until $t < T_{max}$ (total simulation time).

The method detailed above is subject to two constraints that ensure the law of mass conservation is not violated.

- The stoichiometry of the reactions should be conserved.
- The number of molecules of any species at any time should not be less than zero. In other words, if A1 computes to be negative in the previous example, it should be constrained to zero.

There are two points that need to be addressed in the analysis of the crossover algorithm. First, it is necessary to show that the use of a simple Bernoulli trial introduces sufficient randomness, and second, it is necessary to show that the conversion between microscopic scale, i.e. numbers of atoms, and macroscopic scale, i.e. concentrations, does not introduce errors.

Fluctuations in a system governed by Bernoulli trials follow the binomial distribution. The variance of the binomial distribution follows $np \cdot (np-1)$, where n is the number of trials, and p is the probability associated with the trial. For any non-zero p, the limit in large n of the variance is $(np)^2$ which rises to infinity. Therefore, even this simple stochastic step is sufficient for the model system to visit any accessible state.

Conversion between macro- and microscopic systems can be analyzed as well. Under the Grand Canonical Ensemble, the expected number of molecules of a chemical species is estimated by taking an expectation over an exponential distribution defined in terms of the chemical potential (μ), a constant expected kinetic energy or temperature, and other physical terms as needed for the specific system. The chemical potential describes the difference in free energy associated with the creation or destruction of a chemical species.

The expected value for the number of molecules of type R is (by definition as μ is adjusted to make the expected value correct):

$$\langle R \rangle = V[R] = \frac{\sum_n n e^{-\mu(n-\langle R \rangle)/kT}}{Z} \quad \text{..... (43)}$$

where k the Boltzmann constant, Z is the partition function that normalizes the distribution and the sum is over all possible values of Discrete algorithms for simulation use an ODE estimate of d<R>/dt . By definition the expected value of dR/dt over the molecular distribution is given by:

$$\langle dR/dt \rangle = \frac{\sum_n \frac{d\langle R \rangle}{dt} e^{-\mu(n-\langle R \rangle)/kT}}{Z} \quad \text{..... (44)}$$

Taking the derivative of equation (43) results in

$$d\langle R \rangle / dt = \frac{\sum_n \mu / kt \frac{d\langle R \rangle}{dt} e^{-\mu(n-\langle R \rangle)/kT}}{Z} \quad \text{..... (45)}$$

$$d\langle R \rangle / dt = \mu / kt \langle dR / dt \rangle \quad \text{..... (46)}$$

Equation (46) shows that the expected value of the derivative of <dR/dt> and the derivative of the expected value of d<R>/dt are related by a constant. Since they are linearly related, changes in macroscopic and microscopic pictures are directly comparable.

4 TESTING AND VALIDATION

The crossover method outlined in the previous chapter was used to study the dynamics of a variety of biochemical systems including both hypothetical systems and experimentally verified ones. The algorithm was implemented as a C + + code adhering to the object-oriented principles. The choice of C + + over other languages such as Python and Java was based on the ability of C + + to present a strongly typed language platform to reduce code ambiguity and its integration into the Linux system. A well-stocked standard reference library and prior familiarity also played a role in the eventual selection. Having said that, any of the programming languages including either Python or Java, could have been used to develop this program. The program accepts the model parameters and other relevant data from flat file or directly from the user and writes the results into a '.csv' file to facilitate further analysis by analytical software such as MS Excel and Gnumeric. The entire program was developed on a Red Hat Linux distribution using 'vi' as the editor and 'gcc' as the compiler. Wherever appropriate, the stochastic simulation algorithm was also written in C + + and implemented on the same system. Other times, the stochastic simulator from the software package 'Dizzy' was used to perform exact stochastic simulation. Software called XPPAUT (http://www.math.pitt.edu/~bard/xpp/xpp.html) was used to perform deterministic simulation using either the Euler method or Runge-Kutta method.

4.1 Specific Aim 1

Can the crossover method qualitatively recreate the trajectories generated by the stochastic simulation algorithm without compromising the biological relevance of the fluctuations?

Rationale: Stochastic simulation algorithm (SSA) can recreate the random walk of molecules and thus capture the random fluctuations in a biochemical reaction that may or may not be

relevant from a physiological standpoint. Any new method has to be able to match this capability at least on a qualitative level. The purpose of this experiment was to test if the crossover method can generate the required trajectories and capture the relevant fluctuations. This can be achieved by using both the SSA and crossover method to simulate the same model and compare the output trajectories.

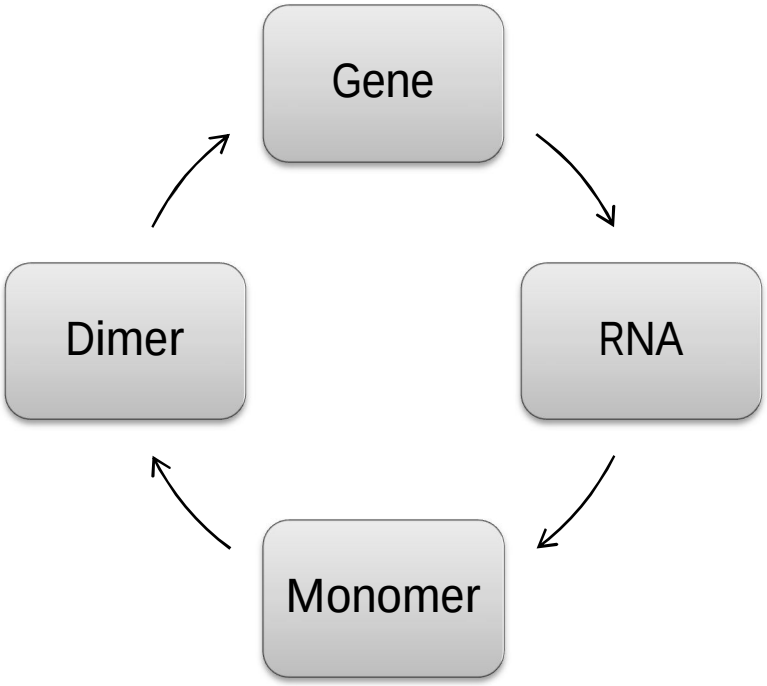

Figure 6 The scheme for an auto-regulatory gene network. The dimer negatively regulates the gene.

Procedure: The auto-regulatory network is a simple example of a mostly feed-forward loop coupled with a feedback regulator. It is a classic low concentration hypothetical biochemical system first used by D.J. Wilkinson (Wilkinson D. J. 2006) to demonstrate the importance of stochasticity in biochemical pathways. In this pathway, Gene, RNA, monomer and dimer are in a feed forward sequence where the concentration of any species depends upon the one before it. The dimer however negatively regulates 'gene' and creates a feedback loop. This example net-

work was used to test if the crossover method provides biologically relevant random fluctuations. The network consists of eight reactions and five species. The reaction rate constants (based on number of molecules) and the deterministic functions for these reactions are as reported in Table 1.

Table 1 Reaction details and kinetic parameters for the Auto-regulatory gene network

Description	**Reaction**	**Rate con-stant (k)**	**f (k,N) for crossover method**
Transcription	Gene → Gene + Rna	1	0.01*Gene
Translation	Rna → Rna + Monomer	10	10* Rna
Dimerization	2 Monomer → Dimer	0.01	1*Monomer*Monomer
Dissociation	Dimer → 2 Monomer	10	1*Dimer
Complex formation	Gene + Dimer → Gene.Dimer	1	1*Gene*Dimer
Complex dissociation	Gene.Dimer → Gene + Dimer	1	10*Gene.Dimer
RNA degradation	Rna → 0	0.1	0.1*Rna
Monomer degradation	Monomer → 0	0.01	0.01*Monomer

Results: The network was simulated with initial conditions of 0 molecules for all species except for the species ‘Gene’ which was equal to 10 molecules. The reaction volume was as-

sumed to be 1E-15 liters. Figure 7 (left panel) shows the time course evolution of all species involved in this network obtained from a single run of the crossover method ($\Delta t = 0.001$) and compares it with the solution obtained from the SSA (single run of direct method) and deterministic method. The right panel of Figure 7 shows the mean behavior of the species from multiple runs. It is apparent that the solution from the crossover method is in reasonable agreement with the other two methods.

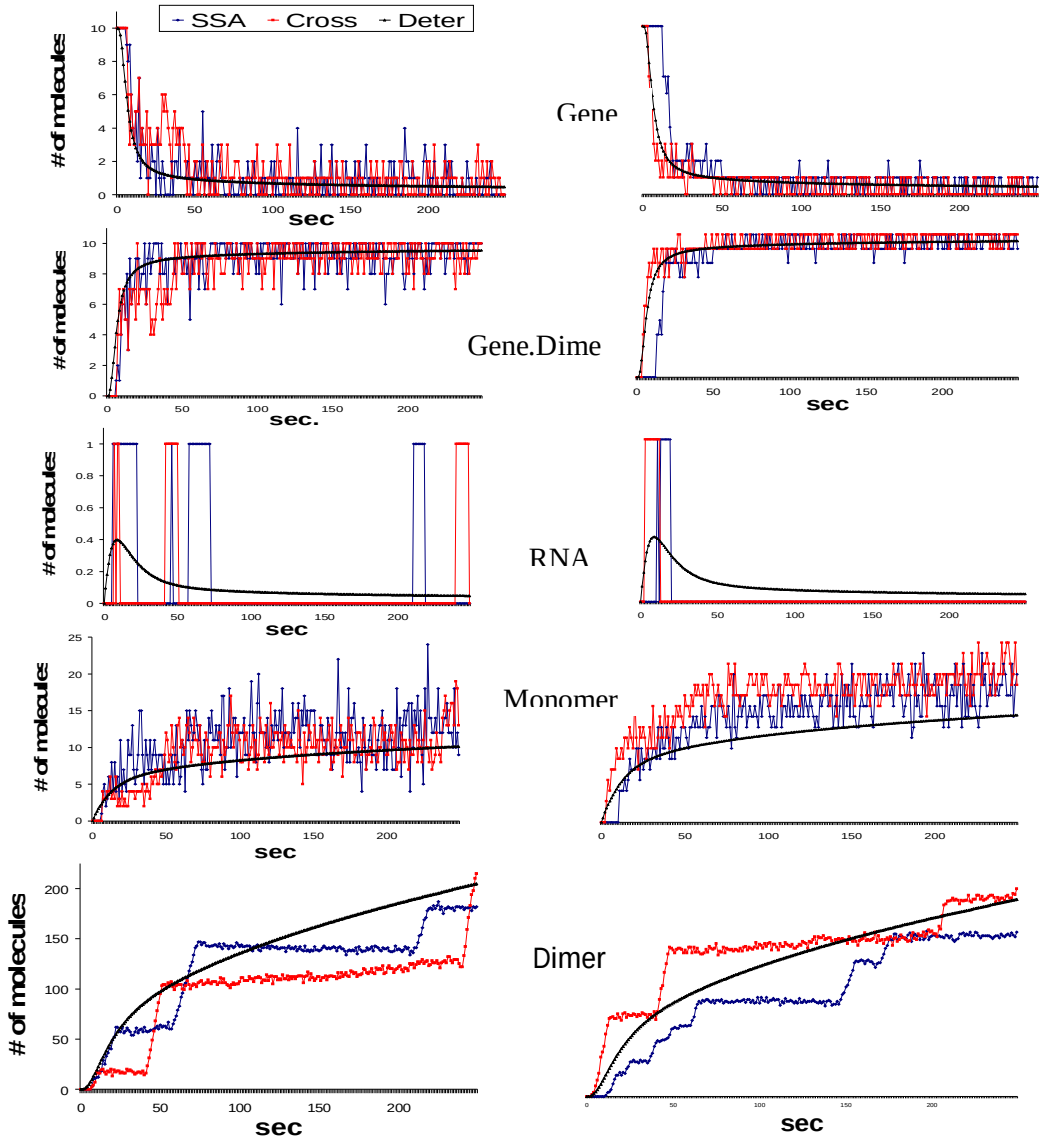

Figure 7 Results of a single run (left panel) and median of 9 runs (right panel) of crossover method (red) compared with the results from SSA (blue) and deterministic (black) method.

To test for the stoichiometric consistency of the crossover method, the solution of species ‘Gene’ and ‘Gene.Dimer’ is plotted in figure 8. As expected, the two solutions are mirror images of each other (figure 8 left) and their sum is always 10 molecules (figure 8 right). This is a direct consequence of the fact that because species ‘Gene’ operates in a closed system (i.e. there is no external production or degradation of ‘Gene’), its net concentration inside the system represented by ‘Gene + Gene.Dimer’ should remain constant throughout the course of the simulation.

Finally, figure 9 demonstrates that in spite of being based on deterministic functions, the controlled randomness introduced by the Bernoulli trials allows the crossover method to reflect those fluctuations (figure 9 right) which are routinely observed in stochastic simulation (figure 9 left) but completely missed by the deterministic simulation.

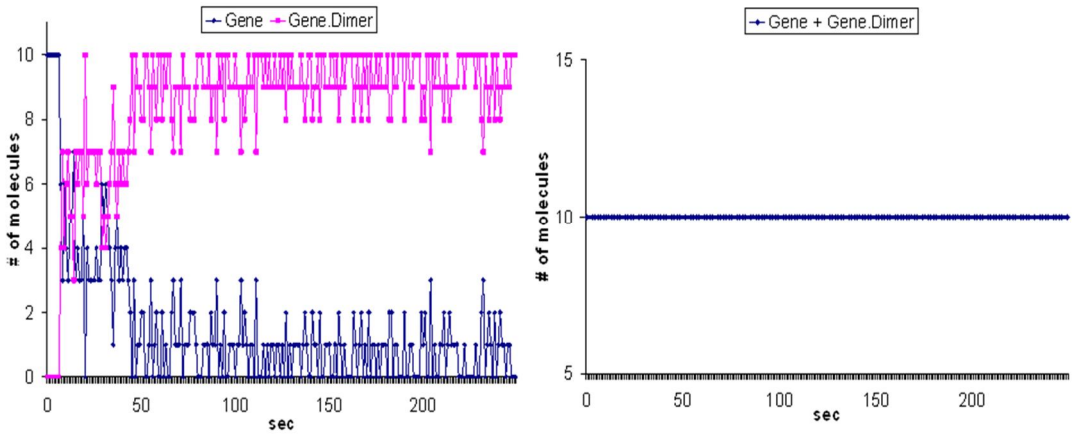

Figure 8 The sum of ‘Gene’ and ‘Gene.Dimer’ stays constant throughout the simulation.

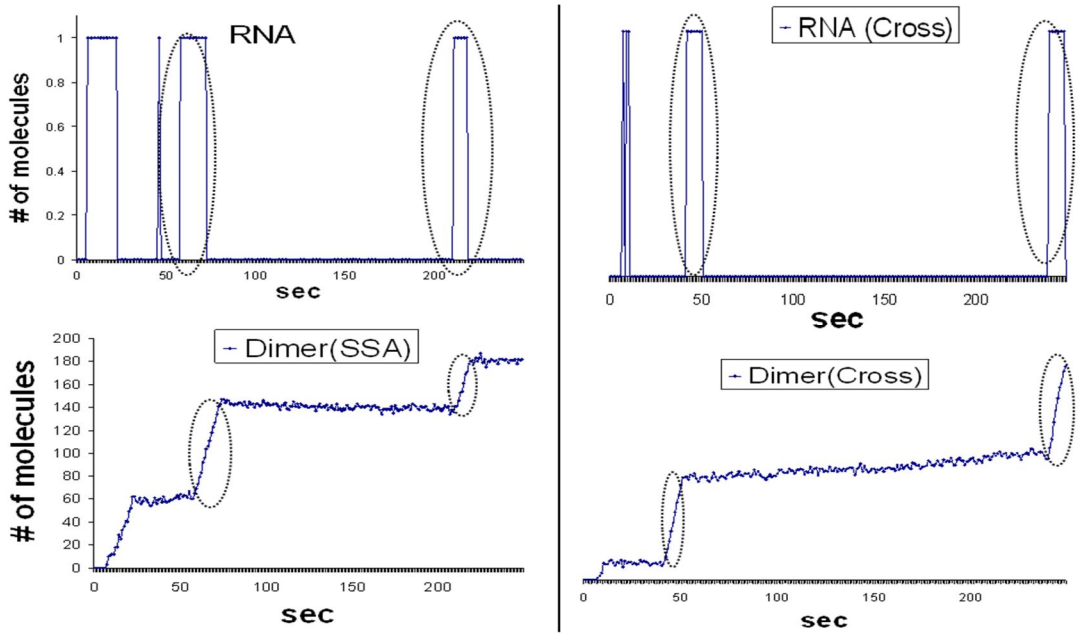

Figure 9 Dimer (bottom) responds to the random fluctuations in RNA (top). These stochastic effects (dotted ovals) that are normally observed in SSA (left panel) are also observed in the crossover method (right panel).

4.2 Specific Aim 2

Will the crossover method be able to handle the concentration transitions within the model during simulation?

Rationale: It is a requirement of systems biology that any method applied to multi-scale modeling has to be able to transition between systems of varying concentrations, especially since none of the existing methods do so. The purpose of this experiment was to test if the crossover method faithfully simulates the trajectories in accordance with the concentration regime of the model. This can be done by applying the deterministic, SSA and the crossover method to the same biological model and compare the solution trajectories, first with the deterministic method in a high concentration zone and then with the SSA in a lower concentration zone.

Procedure: We applied the crossover method to a high concentration system which describes a dimerization pathway. This model system was used by Gillespie D. T. to demonstrate his Tau-leap algorithm. The rationale behind using a high concentration system was twofold: First, we wanted to test the crossover method on a model in which the macroscopic behavior of the system was not dependent upon the microscopic stochastic changes. Such dependency was prevalent in the auto-regulatory gene network of the previous section, where the evolution of the dimer (higher concentration) was dictated by the fluctuations of RNA (present at lower concentration). The current model allowed us to test the crossover method without the influence of background random fluctuations. Second, we wanted to test if the solution from the crossover method would qualitatively resemble a stochastic trajectory when the high concentration system eventually transitions to a lower concentration. The system details are described in table 2.

Table 2 Details of a dimerization pathway

Description	**Reaction**	**Rate constant (k)**	**f (k,N) for crossover method**
Degradation	S1 → 0	1	1*S1
Dimerization	2 S1 → S2	0.002	10*S1 *S1
Dimer Dissociation	S2 → 2 S1	0.5	0.5*S2
Transformation	S2 → S3	0.04	0.04* S2

Results: The initial conditions were 100,000 molecules for S1 and 0 for S2 and S3. The simulations results can be observed in figures 10 and 11. The top half of both figures compares

the higher concentration behavior of deterministic and crossover method for species ‘S1’ and ‘S2’ respectively. Results from stochastic simulation (SSA direct method) are also included for comparison and can be observed to be quantitatively and qualitatively different from the other two trajectories. On the other hand, as expected the crossover method agrees very well with the continuous solution. The bottom half of the figures depict the trajectories at low species concentration. In both cases it can be seen that at a lower concentration when stochastic fluctuations can be dominant, the qualitative behavior of the crossover method is almost the same as stochastic except for the fact that stochastic simulation is about 5 to 8 seconds slower than crossover. This difference is reasonable considering that stochastic and crossover methods are based on different core equations and are bound to yield slightly offsetting results.

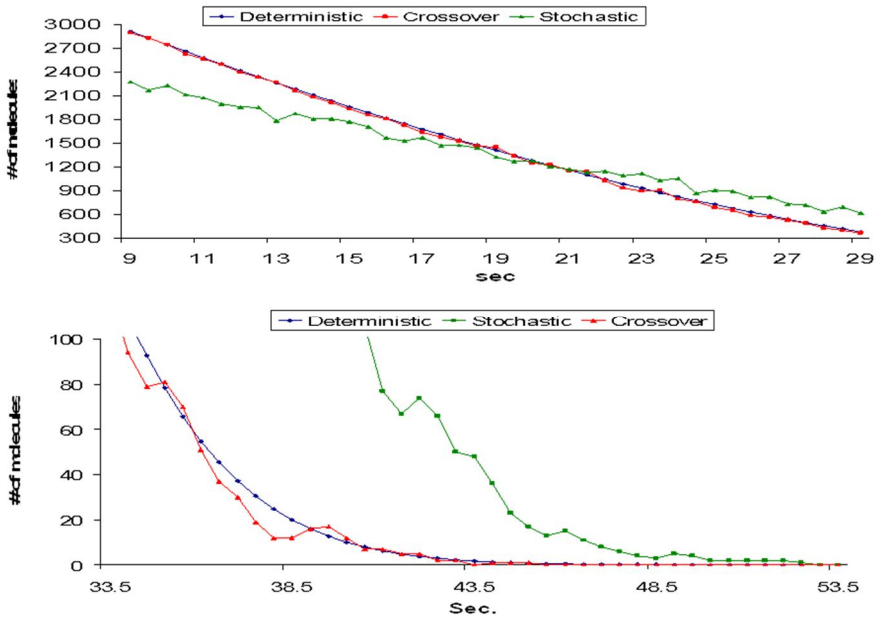

Figure 10 Crossover method displays stochastic effects as the S1 changes into a low concentration (bottom) from a high concentration (top).

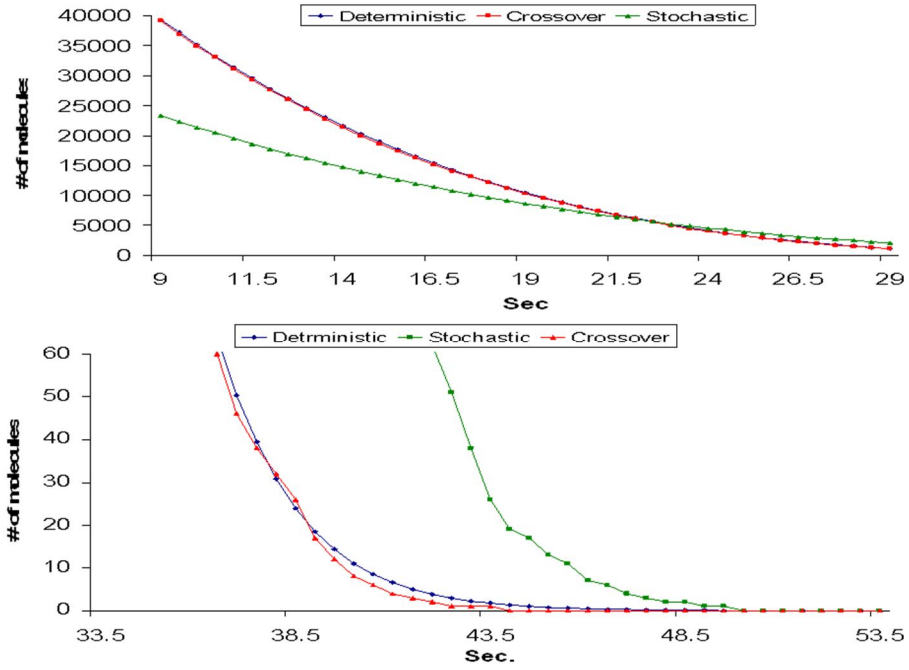

Figure 11 Stochastic effects are observed when S2 transitions into a lower concentration (bottom).

4.3 Specific Aim 3

Can the crossover method replicate experimentally verified stochastic events occurring in biochemical systems?

Rationale: The crossover method was so far tested only on hypothetical models whose predictions were not physically tested. To purpose of this experiment was to further augment the applicability of the crossover method by testing it on a model whose results have been experimentally verified. This can be achieved by simulating an experimentally verified mathematical model with the SSA and crossover method and compare the solution trajectories for similarities. A deterministic trajectory can be used as a background to contrast the stochastic results.

Procedure: This experiment was divided into stages: 1) A B-cell differentiation model was used to demonstrate that the crossover method can generate as much molecular noise as is

experimentally observed. 2) A model for predicting the dynamics of blood testosterone levels was used to show that stochastic effects generated *in vivo* can be simulated by the crossover method.

B-cell differentiation model: A recently reported mathematical model for predicting the differentiation of activated B cells into plasma cells presents itself as an ideal system for such a test for several reasons. First, it is an extremely relevant system from a molecular biologist's point of view as the mechanism underlying heterogeneous differentiation of B cells into plasma cells is not well understood. Secondly, because the model predictions clearly demonstrate that the ramification of stochastic events generated at a molecular level can be clearly manifested as a physiological phenotype, it is important that the crossover method be able to replicate those results.

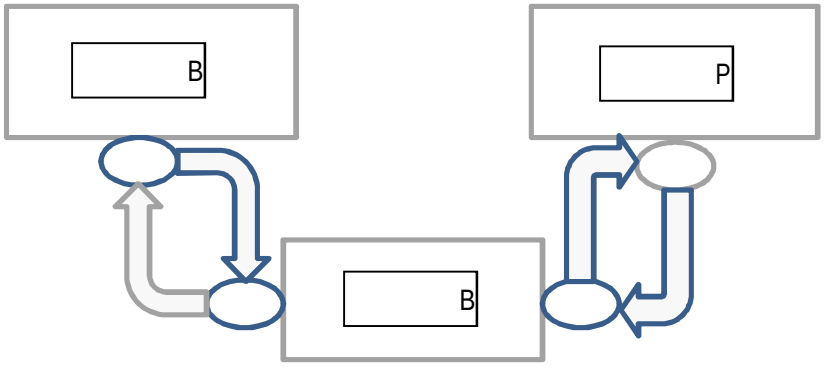

Figure 12 The interaction scheme for Blimp-1, Bcl-6 and Pax-5. Arrows indicate negative regulation.

Terminal differentiation of B cells into antibody producing plasma cells is controlled by three key transcription factors: Bcl-6, Blimp-1 and Pax5.(Zhang et. al., 2010). Random time evolution of the genes of these proteins promotes stochastic expression patterns for the proteins that

eventually leading to terminal irreversible classification of B cells. The initial conditions and other parameters are also the exact same as used in Zhang et. al. 2010.

Results of B-cell differentiation model: The results are presented as a trajectory of the three proteins evolving through 100 hours of simulation (Fig. 6). The plots are a median of 11 simulation runs and clearly show a noisy expression pattern which can never be obtained from a continuous state solution. The level of noise present in the expression patterns is represented as a ratio of standard deviation (σ) to the average (μ) termed as coefficient of variation. As table III suggests the coefficient of variation is in good agreement with the results obtained from for all three proteins.

Table 3 Noise levels in critical proteins.

Protein	**Co-efficient of variation (crossover)**	**Coefficient of variation (from SSA)**
Bcl-6	0.32	0.28
Blimp-1	1.37	1.93
Pax5	0.31	0.28

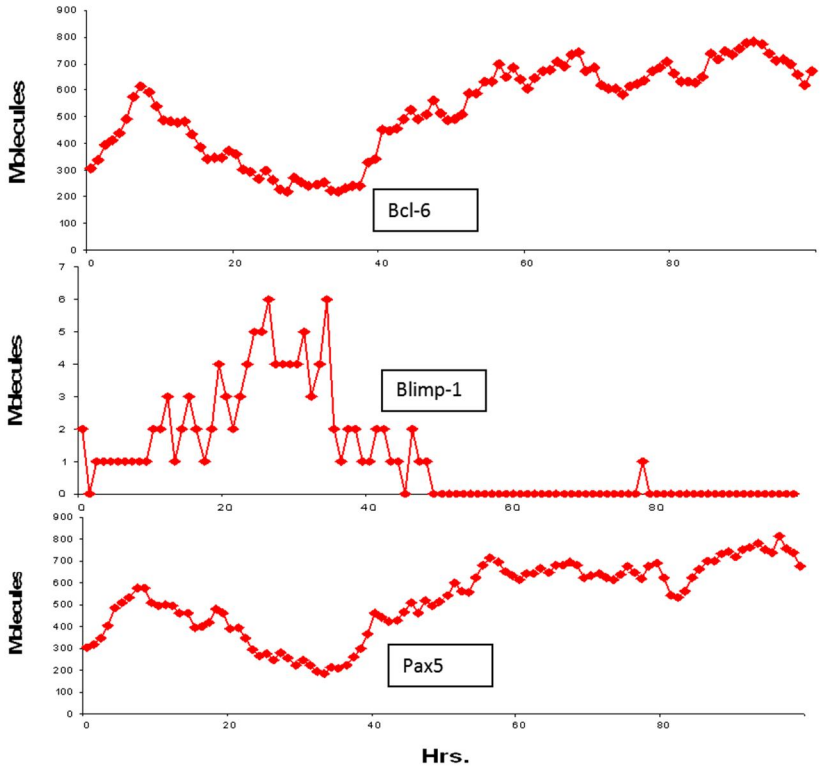

Figure 13 Evolution of three proteins critical for terminal B cell differentiation into plasma cells.

Oscillatory dynamics of testosterone: Mammalian circadian clock controls a variety of biological processes including oscillations of the gonadal hormones. From a biological perspective, the study of hormonal oscillations is non trivial as they are known to affect the morpho-physiology of reproductive organs. Moreover, variation in the testosterone levels in mammals is known to influence Ca2+ oscillations, which are implicated in neuronal apoptosis. Thus, from a computational systems biology point of view, the ability to accurately and efficiently simulate hormonal oscillations is imperative to the success of any modeling and simulation tool. The fluctuations observed in the secretion of the gonadal hormones are a direct result of the physiology of the hypothalamic-pituitary-gonadal (HPG) axis in vertebrates (figure. 13). The hypothalamic neurons secrete gonadotropin releasing hormone (GnRH) via a GnRH pulse generator circuit,

which in turn promotes the secretion of luteinizing hormone (LH) from the pituitary gland using a calcium dependent mechanism. LH then stimulates a cyclic AMP dependent transport of cholesterol into the testicular leydig cells where it is converted to testosterone and oestradiol. Testosterone and oestradiol together provide a feedback signal that negatively regulates the production of GnRH and LH giving rise to the oscillations observed *in vivo*.

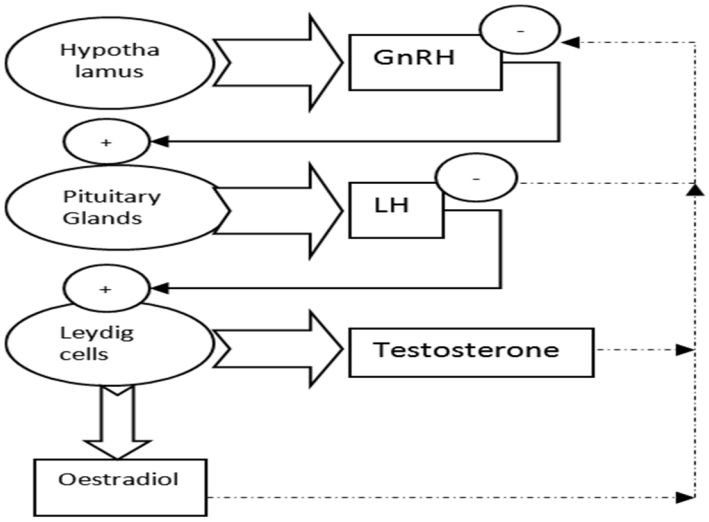

Figure 14 Schematic representation of the HPG axis in vertebrates and its regulation. Dashed lines indicate negative feedback.

Table 4 Details of the HPG axis

Reaction	**Deterministic function (same as reaction rate)**
T → GnRH + T	A/(k+T)
GnRH → 0	b1*GnRH
GnRH → LH + GnRH	g1*GnRH
LH → 0	b2 * LH

LH → T + LH	g2*LH
T→ 0	b3*T

Table 5 Kinetic parameters used in the H-P-G system.

Parameter	Set 1	Set 2	Set 3
A	0.0001	0.0001	0.1
k	0.00010001	0.00010001	0.0001
b1	0.23	0.23	0.23
g1	0.2618	0.2618	0.2618
b2	0.032	0.07	0.032
g2	0.9015	0.9015	0.9015
b3	0.046	0.1	0.046

The HPG biochemical system is typically classified as a high concentration system since the reacting species are often present at a high copy number. A simple mathematical model does

exist (Heuett W. J., and Qian H. 2006) for simulating the HPG axis *in silico;* however a continuous-time continuous-state deterministic solution to this model which is traditionally used for high concentration systems, does not yield the desired oscillatory dynamics. A continuous-time discrete-state stochastic solution, on the other hand, is known to satisfactorily reproduce the experimentally reported sustained oscillations. This discrete event algorithm for biochemical systems was first developed by Daniel T. Gillespie in 1977 and is popularly known as the stochastic simulation algorithm (SSA) or the Gillespie algorithm. Owing to the differences in the physical basis of both methods, the SSA is more accurate than the continuous deterministic method at lower system concentration; however, the accuracy obtained for a high concentration system does not vary significantly. Moreover, regardless of the system concentration, the SSA is computationally inefficient as it requires generating a large amount of random numbers and multiple simulation runs for every reaction. Another significant disadvantage of this technique is that it only works for reactions obeying elementary kinetics. Reactions that follow, for example, the complex Michaelis-Menten kinetics cannot be simulated by SSA without having to break those reactions into a series of elementary reactions which only add to its already inefficient computational performance. As a result, a significant amount of work is now focused on improving the computational efficiency of stochastic simulation. In spite of these improvements, there is growing evidence to show that deterministic methods are yet the most efficient solutions available for simulating biological models. The only significant drawback of a deterministic method, as exposed in earlier works, is its inability to reproduce the stochastic effects that are especially dominant as a system transitions into lower concentration. In an attempt to address this problem, we have previously reported the development of a deterministic-stochastic crossover method that allows the incorporation of stochastic effects in an otherwise deterministic simulation. The solutions ob-

tained from the crossover method were shown to be qualitatively identical to those obtained from SSA while retaining a deterministic implementation.

In this work we have used the crossover method to simulate the dynamics of the HPG axis using the model reported in and compared the results with those obtained from the SSA. The purpose of such an investigation was twofold: First, because oscillations of hormones from the HPG axis observed *in vivo* are a direct result of the random fluctuations of testosterone molecules, we wanted to test if the crossover method was able to correctly reproduce these stochastic effects. Secondly, because the crossover method is fundamentally based on a computationally efficient deterministic method, it would be informative to test its computational performance against that of the SSA.

Earlier, the work of Heuett W. J. and Qian H (2006). was successful in demonstrating the occurrence of sustained oscillations by simulating an ordinary differential equation model of the HPG axis using the SSA for three separate sets of parameters. In this work, we have used the same model and parameters as reported in Heuett W. J. and Qian H (2006). to allow for rational comparison of the crossover method with the SSA. The details of the reaction network derived from and the corresponding functions are summarized in table 4 while the parameters are noted in table 5.

The network of the HPG axis consists of six reactions and seven parameters. For every set of parameters described in table 2, the network was simulated first with the crossover method, followed by the SSA and finally a continuous deterministic (Gear method) algorithm. Furthermore, the execution time for the SSA and the crossover method was noted and compared. The initial conditions were maintained the same for all simulation runs at 100 molecules of GnRH, 10 molecules of LH and 1 molecule of testosterone. The time step was assumed to be 0.1

min for the crossover method and 1 min for the deterministic method. Each simulation run was conducted for a total period of 478 mins. so as to allow enough time for the occurrence of sustained oscillations.

Results for the oscillatory dynamics of testosterone model: The simulation results are plotted in Fig. 13, 14 and 15 for the three sets of parameters respectively. In all three figures parts A, B and C correspond respectively to the deterministic solution, the SSA and the crossover method. The left panel of each of these parts shows the trajectory of testosterone while the right panel shows the trajectories of hormones GnRH and LH of the HPG axis. The deterministic simulation implemented via Gear method can be clearly seen to lack any sustained oscillations which can be attributed to the fact that the mathematical model being simulated did not have any explicit higher order differential terms. The peak observed in the trajectories is primarily due to presence of an implicit higher order differential term in the model that responds to a perturbation but does not oscillate. The SSA was implemented internally without the assistance of any external software to allow for a fair comparison with the crossover method. The SSA solution of the model can clearly replicate the oscillatory dynamics expected from the system mainly as a direct consequence of the random fluctuations instigated by degrading testosterone levels. The qualitative and quantitative nature of these oscillations is similar to those reported. This is especially true for Fig. 3 where the period of oscillation (about 120 mins.) is close to the experimentally reported values.

Although, the simulation output from the SSA seems to be accurately reflecting *in vivo* observations, it also highlights a major drawback of the SSA. The SSA randomly selects a time step for updating the number of molecules present in a system based on an exponential distribution. Owing to the intrinsic design of the SSA, this exponential distribution always generates a

very small time step (0.001 – 0.01 mins.) to ensure that none of stochastic effects are missed. This leads to very inefficient computing time. The crossover method, on the other hand, can be noted to reproduce all of the oscillations previously produced only by the SSA albeit with a bigger and uniform time step of 0.1 mins. Indeed, as observed in part B (the output from SSA) of figures, the data points are more than twice of those required for part C (the output from crossover) of the same figures for the same period of simulation. This ability of the crossover method to generate stochastic effects while implementing an efficient deterministic solution makes it particularly attractive for these biological systems where the traditional SSA seems impractical. It can also be noted that the solution from SSA (part B) "appears" to have non-uniform period of oscillation as opposed to the crossover method (part C). However, a close examination of the SSA will reveal that this visual anomaly is a direct consequence of the unequal time steps adopted by SSA during simulation and the actual period of oscillation is in fact comparable for either method.

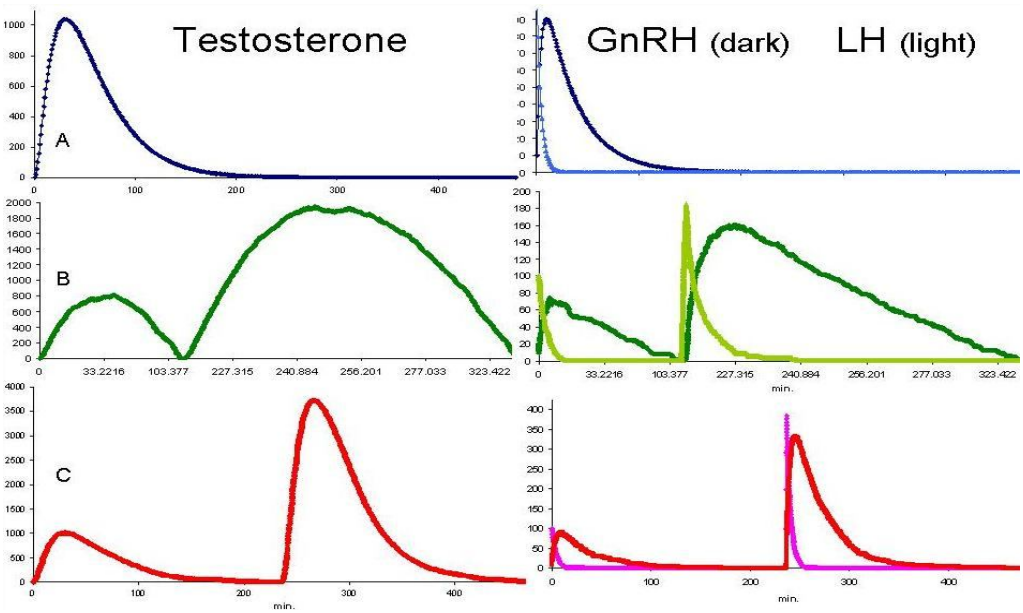

Figure 15 The simulation output of gonadal hormones of the HPG axis for parameter set 1. Testosterone (left), GnRH and LH (right). Trajectories from A) Deterministic solution (blue), B) SSA (green) and C) the crossover method (red).

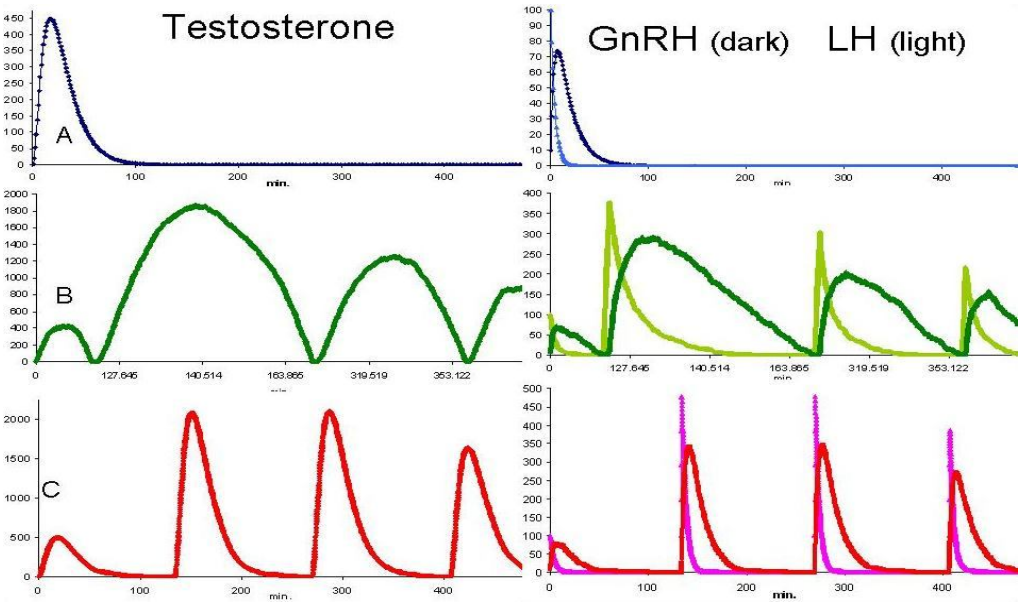

Figure 16 The simulation output of gonadal hormones of the HPG axis for parameter set 2. Testosterone (left), GnRH and LH (right). Trajectories from A) Deterministic solution (blue), B) SSA (green) and C) the crossover method (red).

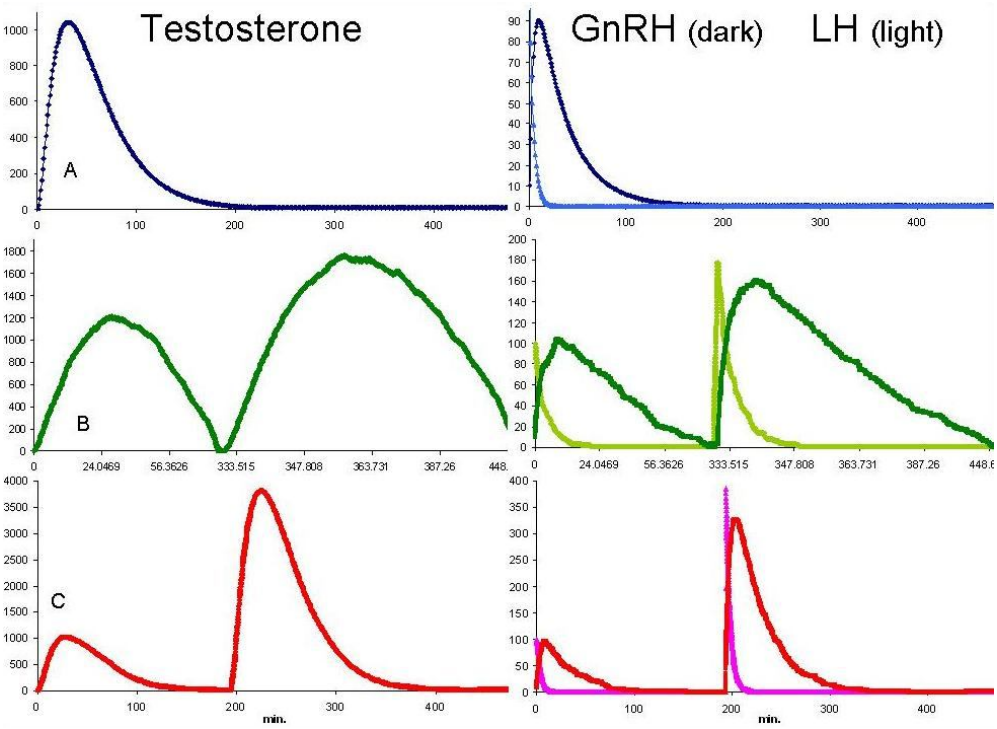

Figure 17 The simulation output of gonadal hormones of the HPG axis for parameter set 3. Testosterone (left), GnRH and LH (right). Trajectories from A) Deterministic solution (blue), B) SSA (green) and C) the crossover method (red).

4.4 Specific Aim 4

Is the crossover method computationally more efficient than the stochastic simulation algorithm?

Rationale: One of the drawbacks of the SSA is the computational burden it puts on the machinery. It is also one of the reasons, it cannot be considered for systems biology based simulation. The purpose of this experiment was to test if the novel method was faster than the SSA. This can be achieved by implementing the SSA and the crossover method on the same computational platform and compare the run times of both methods while simulating the exact same mathematical model. Only the execution times should be considered for a fair comparison.

Figure 18 Comparison of the execution times of the crossover method and the SSA for the three sets of parameters.

Procedure: The code for the SSA was rewritten in C+ + so as to allow its implementation it on the same machine as the crossover method thereby making the comparison as fair as possible. Both methods were then used to simulate the dynamics of blood testosterone of the previous experiment.

Results: Fig. 16 shows the execution times for the two methods. This time does not include the time taken for input or output as that is not a part of either algorithm. The execution time for the crossover method was 0.01 seconds for all three parameter sets mentioned in table II indicating that the run time may not be parameter dependent. The SSA, on the other hand consumed 0.05 seconds for the first and second set of parameters and 0.03 seconds for the third set. In either case, the crossover method was at least thrice as fast as the SSA and on average was more than four times faster than the SSA. It is worth reiterating that these numbers are for execution of the algorithms only and do not include the time required for data input or output that can alter the net run time and lead to erroneous results.

4.5 Discussion

We have presented in this work a simple crossover method for the simulation of biochemical networks. The results obtained after testing on model biochemical systems clearly suggest the plausibility of this novel method in correctly simulating the time course evolution of biochemical pathways. Although hybrid algorithms that use both deterministic and stochastic simulation techniques for the respective parts of a biochemical system have already been described in the literature, the crossover method has two unique advantages; First, it does not require the partitioning of a biochemical system into 'fast' and 'slow' reactions which is a basic requirement of all hybrid algorithms. Moreover it can be noted that, unlike hybrid methods, the crossover method does not 'combine' deterministic and stochastic techniques but merely introduces a certain degree of randomness in an otherwise purely deterministic solution. Secondly, none of the current methods (hybrid or otherwise) deal with the possibility of a 'fast' reaction, normally operating at a higher concentration, transitioning to a 'slow' reaction where it is forced to operate at a lower concentration due to some physiological disturbance in the system. Such scenarios are not altogether trivial or rare as it is common for experimental system biologists to purposely alter the local intracellular levels of enzymes in order to study their global effects. The crossover method is better equipped to predict these scenarios because the intrinsic randomness of the algorithm allows it to seamlessly transition between varying degrees of concentration without failing to capture the random fluctuations when necessary. Also, because the crossover method is based on deterministic functions, the system reactions need not be restricted to have simple elementary kinetics but can be represented by more complex expressions as evident in the case of B cell differentiation model.

The ability to simulate molecular circadian clock dependent gonadal oscillations is important for any computational systems biology application. Considering the fact that oscillations observed in blood testosterone levels are a direct consequence of the random activities that occur at a molecular level, only the stochastic simulation technique has so far been successful in replicating the dynamics of the HPG axis. As this technique is not deemed to be computationally efficient, in this work we have demonstrated the feasibility of a novel approach to simulate similar sustained oscillations. This approach, previously reported by us as the crossover method, is a predominantly deterministic method with the inclusion of a controlled degree of randomness. This inherent randomness bestows the crossover method with a unique ability to reflect the molecular fluctuations of a biochemical system that are only captured by the pure stochastic methods, while retaining the computational efficiency afforded only by a deterministic implementation. To support our hypothesis, we performed model simulations using the same sets of parameters as previously reported and obtained results that were qualitatively and quantitatively similar to the published reports. Furthermore, we showed that, the crossover method was at least three times and in some cases up to five times faster than the SSA. With that being so, the crossover method is yet in a nascent stage of development and its ability to handle complex networks across multiple time scales as well as computationally stiff systems is yet to be tested. Nevertheless, the holistic approach adopted by systems biology introduces a dynamic wherein the effect of random molecular fluctuations in a biochemical system often propagates to the metabolic level. This caveat has to be adequately addressed by any simulation algorithm focused on computational systems biology. Towards that end, the efficiency and accuracy demonstrated by the crossover method in this work qualifies it as a viable option for systems level investigation of biological processes.

5 SUMMARY AND CONCLUSION

Complex diseases are the biggest cause of human mortality all over the world. Traditional reductionist techniques are not suitable for generating a cure for these diseases as the isolation encouraged by reductionism fails to capture the complicated interplay between the disease physiology and environmental factors. The new paradigm of systems biology may help to solve some of these issues as it encourages a holistic approach towards solving complex diseases. As a concept, systems biology is not just limited to complex diseases but any biological phenomenon that warrants a systems level investigation. Systems biology is broadly classified as either experimental, which employs high-throughput experiments to generate data from multiple dimensions (e.g. micro array), or computational, which uses mathematical and statistical techniques to analyze that data and make testable hypotheses from it. There often exists an iterative process where the testing of hypotheses in a lab leads to improved data for computational use which in turn results in better quality of future hypotheses.

Mathematical modeling and simulation is a very important tool in computational systems biology used exclusively for predictive hypotheses. It allows representing seemingly abstract biological processes in quantitative terms so as to obtain physiologically significant clues from them. Simulation thus makes it possible to realize the full potential of systems biology. It is important to point out that the use of simulation in biology is by itself not a novel concept but has been used for more than fifty years. It is only now that the expanded computational infrastructure has unshackled the potential of simulation techniques which were often held back in the past by limited computational power. The simulation process typically starts with building a mathematical model which is often a set of differential equations. These equations are usually complex

themselves and do not have an analytical solution to generate a time series trajectory. Hence, numerical methods are employed to create an approximate solution to the set of differential equations. Numerical methods are typically implemented as computer programs which make it easier to perform the repetitive calculations that go into using a numerical method. Some of the widely studied and popular numerical implementations include explicit methods, implicit methods and the predictor-corrector family of methods. Each method has its own advantages and disadvantages when it comes to solving differential equations. Explicit methods are very simple to implement and easy to follow but suffer from instability and lack of sufficient accuracy. Implicit methods are very stable, highly accurate and computationally efficient; however they are extremely complicated in terms of implementation. Predictor-corrector methods combine the best of both worlds, the simplicity of an explicit formula along with the robustness of an implicit method. These methods collectively are known as deterministic methods because the solutions of the differential equations obtained using these methods are predicated by the equations themselves. Biological simulation is heavily dominated by the use of deterministic methods as its simplistic framework appeals to a variety of researchers, especially physical scientists.

A second class of methods known as stochastic methods is recently being used for biochemical simulation as an alternative to the deterministic methods. These methods are used by assuming that the function to be integrated is a discrete function and its evolution is subject to random events occurring in time. For example, the chemical molecules that participate in a reaction are discrete and hence any differential equation involving them can be solved stochastically. The premise of such integration is to figure out the random Brownian motion of molecules and the “chance” of collision between them. It is important to understand here that stochastic meth-

ods are developed strictly for chemical reactions and cannot be generally used for just any mathematical functions. Its application in biochemical simulation, however, has had extraordinary success. While fluctuations in molecule numbers are a common occurrence in biological reactions, its effect is not very appreciable at higher concentration of molecules. At lower concentrations, however, the fluctuations can present some unique events that are usually important for downstream processes. As stochastic simulation methods are based on predicting the random collisions between molecules, they easily capture these fluctuations and present them as a part of its solution. This unique ability of a stochastic simulation algorithm to accurately capture the natural randomness of a chemical reaction makes it very useful for biological simulations. Stochastic methods are not without its share of issues though. They are known to be extremely inefficient and problematic during scale-up. Moreover, owing to their quest for pinpoint accuracy, they are built to handle only simple elementary kinetics.

None of these methods can really claim to suit systems biology based applications. Multi-scale modeling and simulation is the norm for systems biology and an integration method being able to handle the transition to and from a lower stochastically dominated concentration system is very critical. It is even more critical to do so in an efficient manner. The main goal of this work was to present such a technique to integrate the system of differential equations originating from a multi-scale model. Towards that end, a new method was developed within the parameters of a deterministic framework but incorporating some stochasticity as well. This was achieved by using a regular explicit method for solving the equation while employing the concept of a Bernoulli event to introduce some randomness in the solution. This new concept was termed the crossover method. The method was then tested on four different platforms. In the first case, a

proof of principle had to be established. So, a hypothetical system was considered which in the past had been used to demonstrate the presence of stochasticity in biochemical networks. It was shown that the random fluctuations in this model which were not captured by the deterministic method were promptly captured by the crossover method and confirmed by the stochastic simulation method. It was a strong indication that the crossover method might be able to generate stochastic trajectories. The second platform was used to test the transition handling of the method. As mentioned earlier, it is imperative from a systems standpoint that a simulation method be able to transition smoothly between different concentration zones. This was achieved by testing the method on another hypothetical network which evolves from a higher concentration system to a lower concentration. In this case, it was shown that the crossover method does not have any appreciable difference in the high concentration regime but as the system transitions into a lower concentration the random fluctuations become significant and are correctly captured by the crossover method. This result was pointing towards more evidence that the crossover method was able to capture stochasticity. So far the test platforms were all hypothetical models and it was logical to ask the question as to how the crossover method would stack up against the stochastic simulation in an experimental setting. To test it, we simulated a B cell differentiation model with the crossover method. The results proved that we were able to generate the same degree of noise in the system as observed in the experimental set up. To test it on a more comprehensive platform, a model for the oscillations observed in blood testosterone levels was used. Previously published study had confirmed that only pure stochastic methods were able to generate the oscillations required to accurately reflect the experimental observations. This was an ultimate challenge for the crossover method as only if the method is able to capture the minor fluctuation, will it be able to generate the oscillations. As expected, the crossover method handled

the oscillations with ease thereby proving beyond doubt the validity of this method to generate stochastic results that are also biologically relevant.

The next and final question was that of the efficiency of the crossover computation. To test this question, the stochastic simulation algorithm was implemented on the same system as the crossover method. The run times for methods were then noted to execute the testosterone model. The read / write time was ignored for both methods in order to create a fair testing platform. It was found that, on average, the crossover method was at least three times faster than the stochastic simulation method. The superior efficiency of the crossover method can be reasonably concluded form this test.

The development of a crossover method described in this work can be useful for systems biology applications. It has been shown that the new method has no issues handling the concentration gradients either in terms of accuracy of solution or robustness of the algorithm. Although the spatial transients in biological systems have not been explored with the crossover method it can be reasonable to predict that it may not be too difficult to extend the core algorithm to include it. This unique ability of crossover method makes it extremely suitable for multi-scale simulation where other methods fail. The method can be a basis for a core simulation engine specifically for systems level testing of biological hypotheses generated from wet lab experiments. As documented throughout this work, applying the systems paradigm towards work regarding complex diseases requires a powerful multi-scale simulation tool which the crossover method is well poised for. The efficiency displayed by the crossover method is another aspect of its suitability for systems biology. The scaling of biological networks, which is very common in the era of high

throughput experiments, needs a simulation engine that can scale equally well. The crossover method will not have any problems in scaling due to its primarily deterministic framework. The overall computational efficiency, stability and the ability to capture random collisions between reacting molecules shown by crossover can be an asset not just for systems biology but any application of simulation of mathematical models.

REFERENCES

Allen W. L., Stevenson L., Coyle V. M., Jithesh P. V., Proutski I., Carson G., Gordon M. A., Lenz H. J., Van Schaeybroeck S., Longley D. B., Johnston P. G. (2011). A systems biology approach identifies SART1 as a novel determinant of both 5-fluorouracil and SN38 drug resistance in colorectal cancer. Molecular CancerTherapeutics, 11, 119 – 131.

Andersen-Nissen E, Heit A, McElrath MJ. (2012). Profiling immunity to HIV vaccines with systems biology. Current opinion in HIV and AIDS, 7, 32 – 37.

Azmi A. S., Beck F. W., Bao B., Mohammad R. M., Sarkar F. H. (2011). Aberrant epigenetic grooming of miRNAs in pancreatic cancer: a systems biology perspective. Epigenomics, 3, 747 – 59.

Bentele M., Lavrik I., Ulrich M., Stösser S., Heermann D. W., Kalthoff H., Krammer P. H., Eils R. (2004). Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. The Journal of Cell Biology, 66, 839 – 851.

Berger S. I., Iyengar R. Network analyses in systems pharmacology. (2009). Bioinformatics, 25, 2466 – 2472.

Bianconi F., Baldelli E., Ludovini V., Crinò L., Flacco A., Valigi P. (2011) Biotechnology Advances, 30, 142 – 153.

Bickle J. (2003). Philosophy and Neuroscience: A Ruthlessly Reductive Account, Springer.

Buonaguro L, and Pulendran B. (2011). Immunogenomics and systems biology of vaccines. Immunological reviews, 239, 197 – 208. doi: 10.1111/j.1600-065X.2010.00971.x.

Buonaguro L., and Pulendran B. (2011). Immunogenomics and systems biology of vaccines. Immunological Reviews, 239, 197 – 208.

Burgener A., Sainsbury J., Plummer F. A., Ball T. B. (2010). Systems biology-based approaches to understand HIV-exposed uninfected women. Current HIV/AIDS Reports, 7, 53 – 59.

Chen B. S., Yang S. K., Lan C. Y., Chuang Y. J. (2008). A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. BMC Medical Genomics, 1, 46.

Cloutier M., Wang E. (2011). Dynamic modeling and analysis of cancer cellular network motifs. Integrative Biology:Quantitative Biosciences from Nano to Macro, 3, 724 – 732.

Crampin E. J., Schnell S., McSharry P. E. (2004). Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Progress in Biophysics and Molecular Biology, 86, 77 – 112.

Dewey F. E., Wheeler M. T., Ashley E. A. (2011). Systems biology of heart failure, challenges and hopes. Current opinion in Cardiology, 26, 314 – 321.

Dewey FE, Wheeler MT, Ashley EA. Systems biology of heart failure, challenges and hopes. Current opinion in Cardiology, 26, 314 – 321.

Dux-Santoy L., Sebastian R., Felix-Rodriguez J., Ferrero J. M., Saiz J. (2011). Interaction of specialized cardiac conduction system with antiarrhythmic drugs: a simulation study. IEEE Transaction for Biomedical Engineering, 58, 3475 – 3478.

Edwards Y.J., Beecham G.W., Scott W. K., Khuri S., Bademci G., Tekin D., Martin E. R., Jiang Z., Mash D. C., ffrench-Mullen J., Pericak-Vance M. A., Tsinoremas N., Vance J. M. (2011). Identifying consensus disease pathways in Parkinson's disease using an integrative systems biology approach. PLoS One, 6, e16917.

Ewing G. W., Parvez H. S. (2011). Mathematical modelling the systemic regulation of blood glucose: 'a top-down' systems biology approach. Neuro Endocrinology Letters, 32, 371 – 379.

Fonseca SG, Procopio FA, Goulet JP, Yassine-Diab B, Ancuta P, Sékaly RP. (2011). Unique features of memory T cells in HIV elite controllers: a systems biology perspective. Current opinion in HIV and AIDS, 6, 188 – 196.

Gillespie D. T. (2007). Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, 58, 35 – 55.

Grebogi, C., Hammel, S., Yorke, J. A. and Sauer, T. (1990). Shadowing of physical trajectories in chaotic dynamics: containment and refinement. Physical Reviews and Letters, 65, 1527 – 1530.

Grima R. (2011). Construction and accuracy of partial differential equation approximations to the chemical master equation. Physical Review E Statistical Nonlinear and Soft Matter Physics, 84, 056109.

Groh A., Louis A. K., Weichert F., Richards T., Wagner M. (2008). Mathematical modelling in systems biology. Simulation of the desmoplastic stromal reaction as an example. Der pathaloge, 29, Suppl 2, 135 – 140.

Haddad EK, Pantaleo G. (2012). Systems biology in the development of HIV vaccines. Current opinion in HIV and AIDS, 7, 44 – 49.

Hatzikirou H, Chauviere A, Bauer A. L., Leier A., Lewis M. T., Macklin P, Marquez-Lago T. T., Bearer E. L., Cristini V. (2012). Integrative physical oncology. Wiley Interdisciplinary Reviews Systems Biology and Medicine, 4, 1 – 14.

Heuett W. J., and Qian H. (2006). A stochastic model of oscillatory blood testosterone levels. Bulletin of Mathematical Biology, 68, 1383 – 1399.

Hodgkin A. L. and Huxley A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. Journal of Physiology, 117, 500 – 544.

Hoof I., Pérez C. L., Buggert M., Gustafsson R. K., Nielsen M., Lund O., Karlsson A. C. (2011). Interdisciplinary analysis of HIV-specific CD8+ T cell responses against variant epitopes reveals restricted TCR promiscuity. Journal of Immunology, 184, 5383 – 5391.

Kierzek A. M. (2002). STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. Bioinformatics, 18, 470 – 481.

Kitano H. (2002). Computational Systems biology. Nature, 420, 206 – 210.

Kitano H. (2002). Systems biology: a brief overview. Science, 295, 1662 – 1664.

Kitano H. (2007). A robustness-based approach to systems-oriented drug design. Nature Review, Drug Discovery, 6, 202 – 210.

Konopka G. (2011). Functional genomics of the brain: uncovering networks in the CNS using a systems approach. Wiley Interdisciplinary reviews. Systems Biology and Medicine, 3, 628 – 648.

Kotaleski J. H., Blackwell K. T. (2010). Modelling the molecular mechanisms of synaptic plasticity using systems biology approaches. Nature Reviews. Neuroscience., 11, 239 – 251.

Lavrik IN. (2010). Systems biology of apoptosis signaling networks. Current Opinion in Biotechnology, 21, 551 – 555.

Lee M. J., Cho J. H., Galas D. J., Wang K. (2012). The systems biology of neurofibromatosis type 1 - critical roles for microRNA. Experimental Neurology, 235, 464 – 468.

Loeb J. A. (2010). A human systems biology approach to discover new drug targets in epilepsy. Epilepsia, 51, Suppl 3:171 – 177.

Mac Gabhann F., Annex B. H., Popel A. S. (2010). Gene therapy from the perspective of systems biology. Current Opinion in Molecular Therapeutics, 2, 570 – 577.

Materi W., Wishart D. S. (2007). Computational systems biology in drug discovery and development: methods and applications. Drug Discovery Today, 12, 295 – 303.

Mazzaocchi F. (2008). Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. EMBO Reports, 9, 10 – 14.

Meier-Schellersheim M., Fraser I. D., Klauschen F. (2009). Multiscale modeling for biologists. Wiley Interdisciplinary Reviews Systems Biology and Medicine, 1, 4 – 14.

Meyer-Hermann M., Figge M. T., Straub R. H. (2009). Mathematical modeling of the circadian rhythm of key neuroendocrine-immune system players in rheumatoid arthritis: a systems biology approach. Arithritis and Rheumatism, 60, 2585 – 2594.

Mosca E., Barcella M., Alfieri R., Bevilacqua A., Canti G., Milanesi L. (2012). Systems biology of the metabolic network regulated by the Akt pathway. Biotechnology Avances, 30, 131 – 141.

Munsky B., Neuert G., van Oudenaarden A. (2012). Using gene expression noise to understand gene regulation. Science,336, 183 – 187.

Nurse P, Hayles J. (2011). The cell in an era of systems biology. Cell, 144, 850 – 854.

Oberg A. L., Kennedy R. B., Li P., Ovsyannikova I. G., Poland G. A. (2011). Systems biology approaches to new vaccine development. Current Opinion in Immunology, 23, 436 – 443.

Palme K. (2006). Towards plant systems biology--novel mathematical approaches to enable quantitative analysis of growth processes. The New Phytologist, 171, 443 – 444.

Pannala V. R., Bhat P. J., Bhartiya S., Venkatesh K. V. (2010). Systems biology of GAL regulon in Saccharomyces cerevisiae. Wiley Interdisciplinary Reviews. Systems Biology and Medicine, 2, 98 – 106.

Pedersen, M.G. (2011). Multiscale Modeling of Insulin Secretion. IEEE Transactions on Biomedical Engineering, 58, 3020 – 3023.

Phair, R. D. (2012). Why and how to expand the role of systems biology in pharmaceutical research and development. Advances in Experimental Medicine and Biology, 736, 533 – 542.

Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. (2011). Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. BMC Systems Biology, 5, 120.

Rodriguez B, Burrage K, Gavaghan D, Grau V, Kohl P, Noble D. The systems biology approach to drug development: application to toxicity assessment of cardiac drugs. Clinical Pharmacology and Therapeutics, 88, 130 – 134.

Sabidó E., Selevsek N., Aebersold R. (2011). Mass spectrometry-based proteomics for systems biology. Current Opinion in Biotechnology, doi: 10.1016/j.copbio.2011.11.014.

Salis H., Sotiropolous V., and Kazznesis Y. (2006). Multiscale Hy3S: Hybrid stochastic simulation for supercomputers. BMC Bioinformatics, 7, 93. doi:10.1186/1471-2105-7-93.

Schadt EE, Zhang B, Zhu J. (2009). Advances in systems biology are enhancing our understanding of disease and moving us closer to novel disease treatments. Genetica, 136, 259 – 269.

Segata N, Blanzieri E, Priami C. (2008). Towards the integration of computational systems biology and high-throughput data: supporting differential analysis of microarray gene expression data. Journal of Integrative Bioinformatics, 5,1. doi: 10.2390/biecoll-jib-2008-87.

Segel I. H. (1975). Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady state enzyme systems, Wiley. New York.

Sible J. C., Tyson J. J. (2007). Mathematical modeling as a tool for investigating cell cycle control networks. Methods, 41, 238 – 247.

Simpson J. C. and Pepperkok R. The subcellular localization of the mammalian proteome comes a fraction closer. Genome Biology, 7, 222.

Sperling S. R. (2011). Systems biology approaches to heart development and congenital heart disease. Cardiovascular Research, 91, 269 – 278.

Stewart-Ornstein J., Weissman J. S., El-Samad H. (2012). Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. Molecular Cell, 45, 483 – 493.

Trayanova N.A., Tice B.M. Integrative computational models of cardiac arrhythmias -- simulating the structurally realistic heart. Drug Discovery today. Disease models, 6, 85 – 91.

Tretter F., Albus M. (2008). Systems biology and psychiatry - modeling molecular and cellular networks of mental disorders. Pharmacopsychiatry, 41, s2 – s18.

Umezawa T. (2011). Systems biology approaches to abscisic acid signaling. Journal of Plant Research, 124, 539 – 548.

Van Regenmortel M. H. (2004). Biological complexity emerges from the ashes of genetic reductionism. Journal of Molecular Recognition, 17, 145 – 148.

Vera J., Nikolov S., Lai X., Singh A., Wolkenhauer O. (2011). Model-based investigation of the transcriptional activity of p53 and its feedback loop regulation via 14-3-3σ. IET Systems Biology, 5, 293 – 307.

Vicini P. (2010). Multiscale modeling in drug discovery and development: future opportunities and present challenges. Clinical Pharmocology and Therapeutics, 1, 126 – 129.

Visvanathan M., Baumgartner C., Tilg B., Lushington G. H. (2010). Systems Biology Approach for Mapping TNFα-NFκB Mathematical Model to a Protein Interaction Map. The Open Systems Biology Journal, 3, 1 – 8.

Wang Z., Bordas V., Deisboeck T. S. (2011). Discovering Molecular Targets in Cancer with Multiscale Modeling. Drug DevelopmentResearch, 72, 45 – 52.

Watanabe Y., Kanai A. (2011). Systems Biology Reveals MicroRNA-Mediated Gene Regulation. Frontiers in Genetic, 2, 29.

Wilkinson D. J. (2006). Stochastic modeling for systems biology, Chapman & Hall/CRC.

Young D. A., DeQuach J. A., Christman K. L. (2011). Human cardiomyogenesis and the need for systems biology analysis. Wiley Interdisciplinary reviews. Systems Biology and Medicine, 3, 666 – 680.

Zhang L., Zhao G. (2010). Superiority of single covalent modification in specificity: from deterministic to stochastic viewpoint. Journal of theoretical Biology, 264, 1111 – 1119.

Zhang Q., Bhattacharya S., Kline D. E., Crawford R. B., Conolly R. B., Thomas R. S., Kaminski N. E., Andersen M. E. (2010). Stochastic modeling of B lymphocyte terminal differentiation and its suppression by dioxin. BMC Systems Biology, 1, 4 – 40.

Zhu F., Zheng C. J., Han L. Y., Xie B., Jia J., Liu X., Tammi M. T., Yang S. Y., Wei Y. Q., Chen Y. Z., (2008). Trends in the exploration of anticancer targets and strategies in enhancing the efficacy of drug targeting. Current Molecular Pharmacology, 1,213 – 232.

APPENDIX

Amit Sabnis and Robert W. Harrison, “A Continuous-time, Discrete-state Method for Simulating the Dynamics of Biochemical Systems”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2010.

Amit Sabnis and Robert W. Harrison, “Simulation of Oscillatory Dynamics of Blood Testosterone Levels Using the Crossover Method”, Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) 2010, Montreal, Canada., pp:1 – 6.

Sabnis, A., Harrison, R.W. “A Novel Deterministic-Stochastic Crossover Method for Simulating Biochemical Networks”, Proceedings of the IEEE International Conference on Bioinformatics & Biomedicine (BIBM) 2009, Washington D.C., pp: 315 – 322.