EUKASIMBIOSYS: A STOCHASTIC DISCRETE EVENT-BASED SIMULATION SOFTWARE FOR IN-SILICO STUDY OF INSULIN SIGNALING AND METABOLISM IN CARDIAC MYOCYTES

by

AMIN REZA MAZLOOM

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2008

To my wife Mehran, my mother Mina, my father Samad,

the spirit of my brother Ali Reza,

and

the land of Cyrus the great,

IRAN.

ACKNOWLEDGEMENTS

This dissertation and research has been possible with constant supports and motivations by Professor Kalyan Basu, from whom I learned to be confident and humble toward my goals, which to me is an eternal legacy. I would like to address special thanks to my supervising professor Dr. Sajal K. Das for his gentle but firm directions, leading advices and encouragements during the course of my doctoral studies.

I wish to thank Dr. Cèdric Feschotte for frequently reviewing my works, helpful discussions, and invaluable comments that helped me to improve the quality of the research. I would like to acknowledge my academic advisors Dr. Subhrangsu S. Mandal, Dr. Nikola Stojanovic, and Dr. Hriday K. Das for their interest in my research and for taking time to serve in my dissertation committee.

I would also like to extend my appreciation to Department of Computer Science and Engineering at University of Texas at Arlington for providing me the financial support, academic resources, and facilities, all of which made pursuing the doctoral degree possible for me. Also, I especially thank Dr. Bob Weems for giving me the opportunity to teach in the computer science and engineering department which was a unique experience for me.

I am grateful to all the teachers whom thought me knowledge and wisdom during the years I spent in Alamemeh Tábatábaee primary school, Tohid middle school, and Razi high school in Shiraz, Iran.

I am so blessed for having my beloved wife, Mehran. I would like to thank her by heart for her love and support during these years away from home and family.

Finally, I would like to express my deep gratitude to my mom for granting me the life, my dad for leading me in life and both of them for their sacrifice, patience and supports. I am very grateful to my uncle Rasoul for his inspirational motivations and supports during my studies.

June 27, 2008

ABSTRACT

EUKASIMBIOSYS: A STOCHASTIC DISCRETE EVENT-BASED SIMULATION SOFTWARE FOR IN-SILICO STUDY OF INSULIN SIGNALING AND METABOLISM IN CARDIAC MYOCYTES

AMIN REZA MAZLOOM, Ph.D.

The University of Texas at Arlington, 2008

Supervising Professor: Sajal K. Das

The advent of human genome annotation in the early years of third millennium has enabled the scientist to interlink the processes of life at the molecular level. The coincidence of this breakthrough along with advances in computational technology and high-throughput experimental techniques has promoted the emergence of numerous *-omics* data resources. Although for years before such discovery, scientist believed that cellular processes are the product of interaction between genes and gene products; However, any effort to exploit a comprehensive picture of cellular processes had been obscured due to the knowledge gap that avoided to correlate the cellular processes at the lowest level. Having the organisms blueprint in hand has encouraged many researchers to study a biological process as a part of a whole rather than in isolation. From an engineering point of view, the biologists interests now revolve around comprehending the system level behavior of a biological process in a complex biological network.

Studying biological systems demands for modeling and simulation tools that can capture the dynamics of these systems in time and space. Many variant of these tools have been proposed elsewhere which all try to approximate the Chemical Master Equation (CME). These modeling and simulation tools are broadly classified into deterministic and stochastic based on their temporal evolutions. In the former class, the tools that project a biological system into a set of Ordinary Differential Equations (ODE) are the most prevalent. However, it has been shown elsewhere that these models can not capture the nonlinearity and the deviant effects that exist in the biological processes, due to the inherent random environment of the cell. In latter class, majority of tools comprise strains of Gillespie algorithm, where the system is mapped into sets of chemical kinetic equations which evolves in Monte Carlo steps. The main problem that deteriorates the utilization of these simulation tools is their temporal complexity. A common drawback for both Gillespie and ODE based approaches is their oversimplification in abstracting the physiology of a process that is represented by an equation along with a single kinetic rate constant.

In this dissertation we first elucidate how the Stochastic Discrete Event Simulation (SDES) could be applied in capturing the behavior of biological processes as sets of biological events (bioevents) with random *holding times*. Then we introduce the architecture of *'eukaSimBioSys'* which is designed for system-wide simulation of a eukaryotic cell. The model repository is one of the essential components of our proposed architecture, which comprises reusable modules of parametric models. Each of these parametric models once coupled with a proper parameter set is then applied to capture the *holding time* of a specific bioevent.

These models are physicochemical models that attempt to abstract bimolecular interactions (i.e. modifications, associations, translocations, localizations, etc.) into a parametric probability distribution function of time. Typical interactions include: reac-

tion, receptor-ligand binding, protein-protein binding, chromatin remodeling, transcription, translation, splicing, etc. The previous researchers have already started building this model library and in this work we add four new models (i) ligand-receptor binding, (ii) DNA fluctuations, (ii) chromatin remodeling, and (iii) splicing. For the first one we have developed both the eukaryotic and prokaryotic variants of the model, where as the rest are specific to eukaryotes. These models have been validated with the published experimental data where empirical results were available.

Cell activity is the product of an intricate interaction among three main cellular networks: Signal Transduction Network (STN), Transcription Regulatory Network (TRN), and Metabolic Network (MTN). Each cellular function composed of one or more edges within or across these networks. Hence, system-wide study of a cell requires clear and explicit definition of these networks. We have incorporated the semantic of these networks in *'eukaSimBioSys'* by designing an object-oriented database to hold the layout of these three networks along with their inter-relationships. We have populated these databases for *'human BCell'* and *'human cardiac myocyte'* from data available in literature and other databases.

Despite the advances in health science, and discovery of new drugs, still heart disease is the most life threatening disease in both industrial and developing countries. Cardiac myocytes are the main players of the perpetual heart contraction function and are among the most energy consuming tissues in the body. Any changes in their normal metabolism can lead to severe consequences for an individual. Glucose and fatty acids comprise the major sources of energy for the myocardial cells, the interplay between these two sources is predominantly controlled by insulin. As the ultimate goal of this dissertation we have incorporated all the models developed in this dissertation and elsewhere into *'eukaSimBioSys'* and utilized that to conduct unique *in-silico* experiments for studying the effects of insulin on metabolism of heart muscles.

We exploit the features and capacities of our software by conducting six in-silico experiments where we proved its outstanding potentials in regenerating the experiential data and performing hypothesis testings by applying the experimental conditions in-silico. The biological facts that we validated in-silico briefly include: plasticity of cardiac myocytes, contributions of exogenous glucose and fatty acid in myocardial energetics, transcription regulation of insulin, and the effect of genetic null-mutations on metabolic pathways.

One of the unique features of *'eukaSimBioSys'* that was demonstrated throughout an in-silico experiment was the ability of the software to perform the system-wide simulation of myocardial cellular networks for a prolonged time (48 hours). To construct the SRN, TRN, and MTB for the experiment we incorporated the information from three major databases (i.e. KEGG, BiGG, HumanCyc) along with data from exhaustive literature searches.

'eukaSimBioSys' features variety of promising applications in the biology and health science. It could be applied to suggest the more promising experimental condition for the experimentalist or help investigating new pathways and regulatory mechanisms. Another very important application of this software is in rebuilding the disease scenarios such as hyperglycemia, diabetes, hypertension, ischemia, etc. In-silico investigation on the effects and side-effects of a new drug is another potential application of this emerging software. Note that utilizing *'eukaSimBioSys'* for the above purposes might subject to certain case based enhancements to the current version of the software.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

When the blue print of human came out of the laboratories less than half a decade ago, the start of another renaissance in bioscience was clicked. However the mission was yet to be accomplished, since thousands of questions could be answered only when the multi-billion base pairs is precisely deciphered, linked, and processed at multiple levels. According to the Genome News Network [3] the genome of more than 180 organisms have been sequenced and many yet are in the pipe line. The coincidence of genome annotation, advances in high-throughput experiment techniques and availability of powerful computational tools has given an exponential rise to the *-omics* data resources for various biological objects in the post genomic era. These data resources characterize a biological object in one or more of their aspects which the natural selection has attributed to them in the course of evolution, e.g. structure, function, interaction, localization, pathogenesis and etc. Although these data are invaluable to the science however their real contribution to the community would not be illuminated unless the invisible world in these data is revealed. Discovering this hidden intelligence is beyond intuition and requires a new realm of quantitative tools and computational techniques along with new sets of bench experiments and protocols. Computational systems biology which was coined for the first time in 2002 by *Hiroaki Kitano* [4] is a subsidiary branch of systems biology that has mainly focused in establishing new computational tools and techniques in response to the demand just mentioned. Lately this interdisciplinary field has been welcomed by many biologists, biochemists, mathematicians and computer scientists from different schools of thoughts as the base of common work-bench. This scientific alliance has awarded the

systems biology with an incredibly faster moving pace which subsequently resulted an everyday growing number of more efficient and precise tools for experimentalists and exciting research and work opportunities for the computational scientists.

Rigorous annotation of the DNA and precise definition of molecular interactions underlaying the biological functions are the precursors of a through *in-silico* experiment of the cell. *In-silico* experiment is a buzzword that the bio-community has been exposed to in passed few years. To give a concrete definition for *in-silico* experiment one can literally call that: having the experimental environment simulated in a computer and generate the same results as an empirical observation would result, should the wet lab experimental conditions is provided. However, this is a complicated task that demands for effective utilization of all aspects of -*omics* data, interlinking the processing of different biological systems for creating the complex process along with proper mathematical formalism and processing power that all together describe and mimic the experimental condition in biologically relevant and timely manner. Both the bioresearch and health science community will benefit from this practice in ways raging from recommending the most promising parameters setup for a wet lab experiment, looking at null mutations effects on homeostasis and fate of a cell, identifying and testing a drug target, helping to find an efficient therapeutic drug dosage, to do the measurements or experiments that would otherwise be impossible due to variety of reasons including of lack technology, impractical experimental condition and legal or ethical restrictions.

The envisioned road-map demands for characterizing the functionality of cellular components, which composed of all structural and molecular assets of a cell (e.g. proteins, DNA, mitochondria, lysosome, peroxisome, nucleus etc.). Also it is well understood that any function in the cell is accomplished through delicate interplay among many of the cell components. Hence, studying the dynamics of a cell is only possible by characterizing the functionality of these components as a whole not individually and in the context of

the cell's condition in time, not in isolation. This indicates that we must look in to the cell as a system in order to provide a precise picture of its functionality *in-silico*.

1.1 System-level view of a cell

In a broad sense a system is a subset of a world whose interactions with the rest of the world or another subset is properly defined. An organism could be viewed as a system or collection of systems at different hierarchical level based on the boundary of the system components which could range from organ to tissue, cell, molecule, and to atoms. The degree of complexity between levels grows exponentially from top to bottom. In the current study a cell draws the system boundary and molecules are the interacting components of this system. In order to mimic the interaction of a cell with the rest of the world the interaction of the underlying components must be properly characterized in time and space.

In an effort to create a system-wide abstract for a cell, a comparison is done by Brent and Bruck [5] between a living cell and von Neumann's *stored-program computer* architecture. To seek for homologies among the components, DNA is viewed as the executable program code that needs to be processed and run by the cell processing unit which is a composite network of Signal Transduction Network (STN), Transcription Regulatory Network (TRN), and Metabolic Network (MTN) . Each of these networks operates and cross-talks through a set-of cellular processes. The cellular processes are handled by the interaction of proteins which themselves are the output of DNA execution. The inputs to this bio-architecture are the endogenous and exogenous signals, where the output is a change in metabolism, development, proliferation and/or phenotype of the cell. An external signal is imported into the cell and then ripples in one or more route down to the nucleus and finally to the genome through cascades of complex functions. The respond of genome to the signal, spreads through the cell and/or at the higher level

outside the cell through another set of cellular functions, that might possibly overlap with the input set. It is noteworthy to mention that not all signals would necessarily result a genome level effect.

1.1.1 Signal transduction network

For a cell, an exogenous signal is composed of bacteria, ligand molecule, ions or atomic particles (virus, sugar, Fatty Acid, hormone, Ca, etc.) in the extracellular environment that binds to a cell membrane surface or receptor or enters through a membrane gate. Although, current definition of the signal does not cover cell-cell signaling which incorporates other mechanisms (e.g. notch signaling) but will serve our purposes. A cascade of physiological phenomena that coordinates the signal perception and propagation is refereed to as signaling network. The underlying biological phenomena include: protein complex formation, post transcriptional modification, diffusion, co-activation, co-repression, budding, binding of transcription factor to the DNA, etc. In biochemistry most of these phenomena is abstracted in the form of a biochemical reaction. A set of reactions that pertain to the signal transduction in a cell is referred to as a signaling pathway and set of all signaling pathways in a cell is referred to as signal transduction network of that cell (STN). It should be emphasized that the downstream effect of a signal is nothing more that an actitvated/deactivated transcription factor or protein complex or change the membrane conditions or generate an electric signal. An altered transcription factor might positively or negatively affects the expression profile of certain genes. Where, the affected genes could be determined through the transcriptional regulatory network of the cell. A protein or protein complex might trigger another signaling pathway or possibly alters the metabolic network of the cell. The major databases on Homosapiens signaling pathways include: KEGG [6], Reactome [7] and TransPath [8].

1.1.2 Transcription regulatory network

A typical eukaryotic gene is regulated at multiple levels: sequence , chromatin and nuclear [9]. The first level of regulation include the activation or inhibition of the promoter of a certain gene. The basal and upstream promoters are located ∼40 and ∼ 200 base pairs upstream of the open reading frame (ORF) that could be activated through the binding of transcription factors, i.e. proteins or protein complex. In eukaryotes the regulatory mechanism is very complex which is not completely understood; however, it has been shown that more distant regions upstream/downstram of the main regulatory region can have effect on expression profile [10] of the target gene. Multiple genes might be regulated though the same regulatory region, these set of genes are referred to as operon. A transcription factor might be activated by another protein or small molecule binding to its activation domain this process is referred to as trans-activation and the second party is called trans-activator. The opposite form of the event is also valid for co-repression where an active transcription factor is deactivated by a co-repressor.

In eukaryotes the DNA is wrapped around beads of protein referred to as nucleosome core particle. Nucleosomes contribute to the second level of regulation by confining the cell transcription machinery access to the regions of DNA trapped in nucleosome. The eukaryotic cells are equipped with certain mechanisms to overcome the repressive effect of nucleosomes. The third level of the regulation is the nuclear level, which involves the 3D folding of chromatin fiber in the nucleus and is controlled by the epigenetic regulatory mechanism.

The group of transcription factors that directly interact with DNA belong to the first level of regulation hierarchy. These proteins along with their target genes form a *bipartite graph* which we refer to as tier one transcription regulatory network (TRN). By including the feed forward of other protein on the tier one transcription factors, tier two TRN could be reconstructed and this recursive process could be applied for higher

tiers. These multi-tire transcription regulatory network along with feed back of the gene products to different layers of TRN can further reveal the genome wide effect of the TRN. The transcription factors at the highest level of an organisms TRN, if exist at all, are referred to as global transcription factors. Building a comprehensive TRN for a target cell is subjected to complete genome annotation of the organism and precise information for all the protein-protein interactions within a the cell. To the date, availability of transcription regulatory network for eukaryotic cells is very limited. There are very few databases for transcription factors (TFs) where one can query for the genes that are activated by a certain TF: TransFac [11] and bZip [12]. Rebuilding the transcription regulatory network for any of Homo-sapiens cell as well as other eukaryotes is impractical from the available databases. The sole integrated piece of data that we found which could be used to rebuild a human cell TRN is reported by Tucany *et al.* [13] on human BCell. Hence, rebuilding any cell specific TRN demands an exhaustive search in literature for discrete published records on a gene regulation and its TFs for a specific cell type.

1.1.3 Metabolic network

Metabolic network of a cell is composed of sets of pathways that produce the energy substrates and precursors for all activities of the cell. Based on the chemical family, metabolites (carbohydrates, vitamins, lipids, amino acid, nucleotides, etc.) are separated into different classes. Each class is further divided into subclasses based on precursor metabolite molecule and ultimates product(carbohydrates: glycolysis/gluconeogenesis, citrate cycle, pentose phosphate, Fructose and mannose metabolism etc.). These classification is not standardized but their variations are subtle among popular databases. Each metabolic pathway is composed of one or more reactions that convert one or more reactant molecules to one or more products. Metabolic reactions are normally catalyzed by one or more metabolic enzymes. These enzymes are composed of one or more proteins

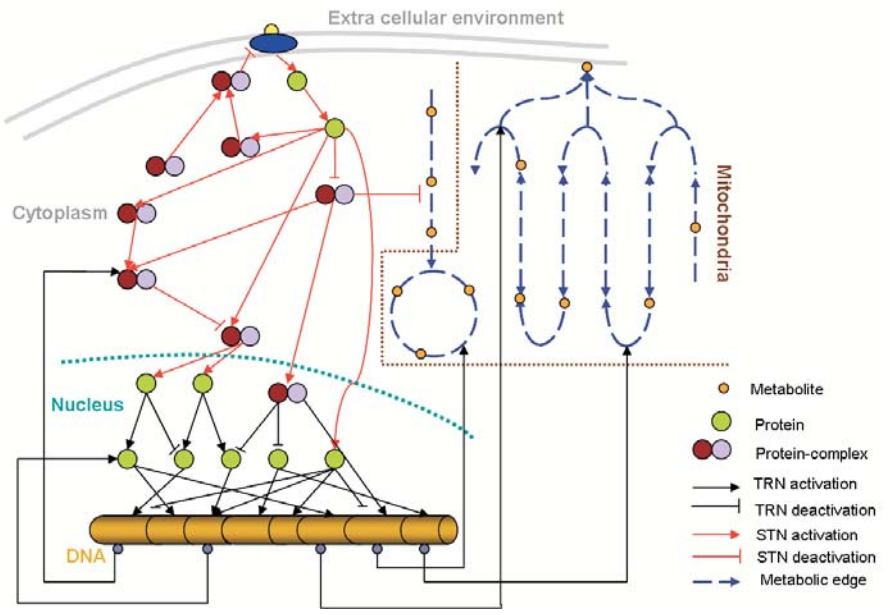


Figure 1.1. An abstract representation of STN, TRN and MTB in a eukaryotic cell.

which are gene products. Major databases for Homo-sapiens metabolic pathways include KEGG database [6], Reactome database [7], BiGG database, and HumanCyc [14] data base.

1.1.4 Interaction of the three networks

The interaction of the above networks is depicted in the graph of Fig. 1.1 where each network node corresponds to molecular body and each arc represents a physiological activity within the cell. A sequence of such activities represents a cell process. We have designed a database that stores the above networks as well as their interrelationship. This database could be queried on every node and arc of the graph and their properties. To properly characterize the spatiotemporal dynamics of the above networks having such database on hand is necessary but not sufficient. An edge of this graph is the manifestation of the physiology of a cellular activity that governs the rules of change in dynamics of a cell activity. Therefore, to gain a system-wide understanding of a cell each

of these individual physicochemical activities should be rigorously studied and relevantly abstracted by a quantitative formalism.

1.2 Systematic modeling of biological process

In 1930's the Hungarian mathematician John von Neumann made his famous quote *'Truth is too much complicated to allow anything but approximation'*. His word is the essence of any modeling effort irrespective of the knowledge domain that the model could belong to. In biology a model is often referred to a cartoon representation or verbal description of a mechanism underlying a biological activity or process, these models are known as qualitative models [15]. Mathematicians and computer scientists' version of model is a mathematical formalism or an algorithm that could reproduce the real world data, given the initial parameters and input to the algorithm. A systematic modeling of a bio-process requires both discipline along with relevant empirical data. The former is used to understand the mechanism underlaying the process, and the latter is applied to map the mechanism into a systematic and logical language. The experimental data is used to validate and optimize the model in an iterative process. Possible knowledge gaps in the mechanism of a bio-process process is filled by incorporating fair assumptions. These assumptions might further be revised during validation and optimization phases of the model design. In a modeling task should some details be unimportant to the objective of the project, or the modeling task becomes impractical due to excessive complexity imposed by the details, valid simplification assumptions will replace the unnecessary details.

1.2.1 Chemical kinetics based modeling of biological processes

Mathematical models have been particularly successful for modeling complex biochemical reaction networks, using deterministic chemical kinetics [16, 17]. In chemical

kinetic based modeling every biological process is considered as a set of reactions. The temporal progress of such reactions is governed by an empirical rate constant. These biochemical reactions are categorized in to fast and slow reactions. A fast reaction (e.g. metabolic, kinase, dimerization) happens in the order of *nano* to *mili* seconds and a slow reaction often happens in the order of seconds and minutes. With this approach, every interaction between molecules is considered as a reaction and a pathway is represented as a sequence of reversible or nonreversible reactions. Each reaction is associated with four data sets: reactants, products, stoichiometry values and the forward and backward kinetic rate constants of the reaction. This information is used to form set of Ordinary Differential Equation (ODE) corresponding to the biochemical reactions. Therefore, each pathway is modeled by a set of ODEs, which will be solved analytically or numerically to determine the temporal dynamics of that pathway. Most often analytical solution to the set of ODEs is impractical and numerical methods will be applied (e.g. Euler, Mid-point, Runge-Kutta,etc.) to approximate the solution. Many variants of ODE solvers is available, where amongst these tools Matlab ODE solvers developed by Matworks® Inc. are quite well known. List of Matlab ODE solvers is given in Table 3.1. The precursor assumption common to all ODE based models is the *steady state* of reactions. This assumption implies that all the reactions in a pathway are in their average dynamics and obviously undermines any deviation from the average. Intuitively such assumption should stay valid for the cases where the concentrations are high with respect to reaction buffer. However, Arkin and Samiolov (2006) have shown that non-classical behavior of biological networks cause their dynamics to substantially diverge from their averages [18]. In such networks Classical Chemical Kinetics (CCK) based dynamics correspond to the mode of the reactions rather than their averages. Also they have identified the potential failure patterns of CCK models even for high reactants concentrations. An example of such patterns is the condition when divergence of a reactant concentration from

Table 1.1. Matalab ODE Solvers

Solver Name	Specifications
Ode45	Applied for non-stiff problems, one step solver, best to apply as a first try for most problems ,based on explicit Runge-Kutta$4^{th}/5^{th}$-order method
Ode23	Applied for non-stiff problems, one step solver, Often quicker but less accurate than ode45, based on explicit Runge-Kutta$2^{nd}/3^{rd}$-order method
Ode113	Multi-step solver for non-stiff problems
Ode15s	Multi-step solver for stiff problems, Uses a variable order method, advised to be applied in case of Ode45 failure
Ode23s	One step solver for stiff problems that could not be solved by Ode15s
Ode23t	For moderately stiff problems
Ode23tb	For stiff problems and is more efficient that Ode15

its average hits zero, obviously lack of reactant resource will stop the progression of the reaction, where such event is not captured through CCK models. Another failure point of such models is their inability to capture the non-linearity of irreversible reactions such as transcription or translation. The fact that immature transcription leads to nascent transcript (i.e. not all nucleus synthesized mRNAs would be successfully transported in to cytosol, or not every cytosolic mRNA results production of protein) are the burstyness in of protein synthesis [19] are the manifestations of such non-linearity and thus, nonclassical effects.

Such deviant effects characterize a stochastic behavior for biological networks and suggest a random environment for the cell. Therefore, deterministic modeling does not have the capacity to properly characterize the behavior of a cellular system that has a random nature.

Also deterministic approaches assume infinite volume to map the discrete spatial distribution of molecules to a homogenous continuous variable, i.e. concentration. However, many intracellular reactions occur in small volumes, hence this assumption could

be significant to the accuracy of the model. Furthermore, the homogeneity of cell environment, assumed by the deterministic modeling approach is usually violated by the cellular processes.

1.2.2 Stochastic modeling of biological processes

The inherent randomness associated with cell environment is due to the *stochastic resonance* [20]. Such stochasticity causes random (stochastic) interactions and collisions between the discrete molecular entities (predominantly proteins and protein complexes). Chemical Master Equation (CME) [21] is used to stochastically determine the species molecular counts, in homogeneous environment within a cell. Species molecular counts describe the internal cell states and biochemical reactions governed by deterministic *'rate law'* account for transitions among the internal states of the cell. Each species corresponds to one dimension in the state space, hence number of states follows an exponential growth as the number of chemical species increases in the model. Numerous stochastic approximation algorithms to CME have been proposed, some of which have been more practical include: Gillespie [22], Arkin and Rao [23], Burrage *et al.*[24], Gibbson and Bruck [25], Le Novre and Shimizu [26], Salis and Kaznessis [27], Haseltine and Rawlings[28], Alfonai *et al.* [29], and Gadgil *et al.* [30]. The application of these algorithm is well suited for environments with low copy number molecules and small system environment.

These modeling techniques have provided an *in-silico* framework to study the dynamics of cellular processes. However, much of their success are limited to specific biological systems and suitable under certain conditions. Applying such frameworks for system-wide study of cellular behavior faces major pitfalls including combinatorial explosion of large number of molecular species involved in a cell, time-scale of difference between different types of reactions in a cell, importance of spatial localization, etc. Par-

ticularly, the complexity of signaling, regulatory and metabolic pathways, cross-talks and noise between the pathways and knowledge gap in different parts of the pathways (e.g. undiscovered reaction, unidentified enzyme, unmeasured kinetic rate, etc) together with the differences in time-scale between gene regulatory and metabolic pathways make the system-level study of their interaction dynamics an even more challenging computational problem [31].

Stochastic models for approximating CME are largely base on Gillespie algorithm [32, 33]. The original Gillespie algorithm is considerably slow, because numerous amount of random numbers need to be generated to specify the reactions that take place in each Monte Carlo step of the simulation. *Stochsim* is a stochastic simulation tool developed based on the Gillespie algorithm [26]. Some variants of Gillespie based tools that have applied approximation techniques like *tau-leaping* to overcome the computational complexity and enhance their efficiency could be found in [34, 31, 22].

Although reaction based stochastic modeling is trying to capture the random behavior of cellular processes, however every action in the cell is still mapped into a biochemical reaction by this class of models. This naive mapping strategy could be hazardous to the accuracy and efficiency of system-wide *in-silico* analysis of cellular processes in two folds. Firstly, by over simplification that results ignoring the important factors in a process that has a complex physiology (i.e. transcription/translation): the entire dynamics of any reaction-mapped biological process is captured through a rate constant in this class of stochastic modeling. Secondly, by imposing computational complexity for simpler processes such as metabolic reactions.

In this dissertation we introduce a new approach in molding biological processes which is stochastic in nature and has *mesoscopic scale*. In our approach every cellular process or interaction is viewed as an event whose temporal dynamics follows a random distribution with a known mean and variance. In each modeling effort, first we study the

qualitative model that characterize the physiology of the objective cellular process. Subsequently, try to use all possible mathematical, physical and mechanical tools, theorem and laws along with experimental data to map the qualitative model into the stochastic information domain. In the light of current modeling approach, any cellular process (e.g. signaling pathway, metabolite uptake, protein synthesis, etc) could be presented as sequence of one or more events whose temporal behaviors are govern by the appropriate models associated with each event.

1.2.2.1 Example 1: Stochastic modeling of human MAPK signaling pathway

As an example we model a process (pathway) through which an erogenous signal is propagated down to the nucleus. One of the well known signaling pathways in human signal transduction network is the MAPK signaling pathway which is predominantly engaged in the proliferation and differentiation cellular functions. The simplified version of the temporal evolution of signal propagation through this pathway modeled by a sequence of events is depicted in Fig. 1.2. The rectangles represent proteins and arrows are the events, next to each event is the event-type associated to the event. Each of these events characterize a stochastic temporal behavior which form one of the t_i realizations. The inputs to each event are the molecular counts participate in that event along with the parameters specific to the model. The outputs for each event are the changes in the molecular counts of those molecules that were effected by that event. Hence, the probability distribution function (PDF) for the time to get an exogenous signal propa-

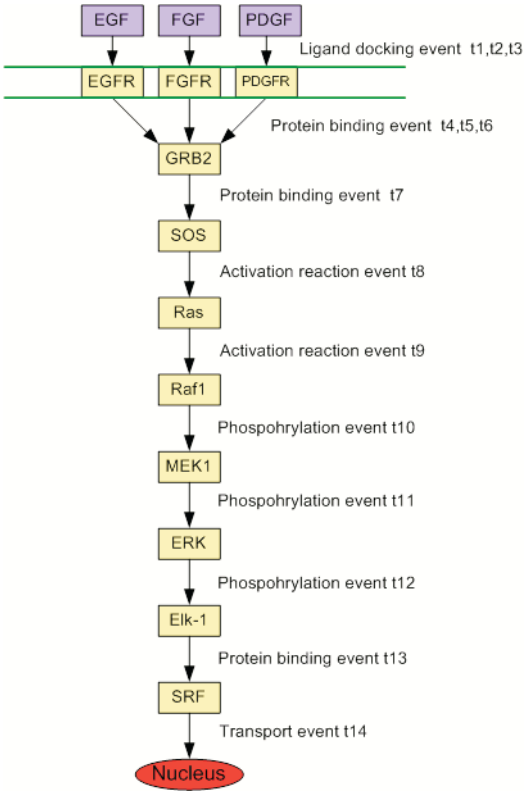


Figure 1.2. Simplified representation of the temporal evolution of MAPK signal propagation to nucleus. All t_is are the stochastic times generated by the event model associated with it.

gated down to the nucleus through the human MAPK pathway is the convolution of the random t_is PDFs, whose realizations is given by:

$$T = \sum_{i=7}^{14} t_i + \left\{ \begin{array}{c} t_1 + t_4 \\ t_2 + t_5 \\ t_3 + t_6 \end{array} \right\}$$

Every t_i is an independent stochastic time generated by the associated stochastic model. The trifurcation in the total time taken for the signal transduction is the due to of variety of exogenous signals transduced through this pathway. It should re-emphasized that the given pathway is simplified for the sake of understandability.

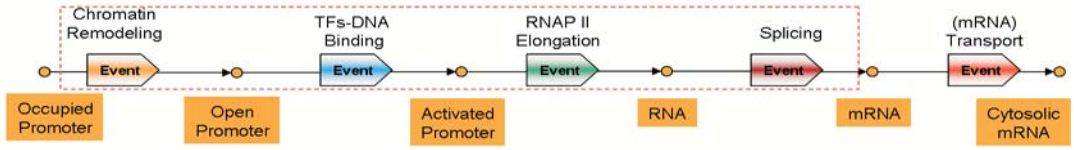


Figure 1.3. The event diagram of a eukaryotic gene expression.

1.2.2.2 Example 2: Stochastic model for inter-arrival time of mRNA in eukaryotes cytosol

In eukaryotes, translation takes place in cytosol which requires the messenger RNA (mRNA) to be available in cytosol to be translated to protein by the ribosomes. For eukaryotes, transcription which precedes the mRNA synthesis happens in nucleus and is a complex process with many of its details, yet remained to be understood. Eukaryotic genome is located in nucleus and is wrapped around histone protein cores ubiquitous across the genome. The DNA-histone complex is referred to as nucleosome that modulates the access to DNA. If a part of a gene or its regulatory regions are blocked by one or more nucleosomes, prior to its transcription the blocking histones should be displaced. Another level of complexity is the distant *cis* regulatory regions upstream and down stream the ORF which is not considered in this work. After clearing the critical regions from histone cores, the promoter of the target gene needs to be activated by the proper transcription factors. Then the RNA-Polymerase II (RNA PII) holoenzyme will bind to the transcription start site and initiates the transcription. The output of a gene transcription is the pre-mRNA of the target gene which is a nucleotide sequence of protein coding regions(exons) interleaved with non-coding regions (introns). A splicing process is needed to convert the pre-mRNA to a messenger (mRNA) by removing the introns. The intron free mRNA should be transported the cytosol to initiate the protein synthesis. We abstracted the above scenario as a sequence of stochastic events as depicted in Fig. 1.3, Table1.2 shows the stochastic model proposed for each events.

Table 1.2. Models Corresponding to the Events Preceding mRNA Translation

Event Name	Model
Chromatin Remodeling	Chromatin remodeling model (chapter 5)
TF-DNA binding	Protein-DNA binding model [35]
RNAPII elongation	Transcription model [36]
Splicing	Constitutive Splicing model (chapter 6)
mRNA transport	Diffusion [37]

Obviously the process in the second example has much more complicated physiology than the one in first example but such complexity is vanished with the state of arts of stochastic event base modeling approach. However, the models underlying each event in the second example are more complex and developing those models demands significant effort.

One of the strongest features of this modeling approach is its *mesoscopic scale* property which allows to every granular molecular detail of the actual physiology of the process to be included into the model. At the same time, it allows to scale up and eliminates any unnecessary detail from the model with the cost of accuracy, at the designer's will. For instance in the second example the entire set of events encompassed by the red rectangle could be abstracted into a single macro-transcription event. Oppositely, one can bring further details in by breaking down the elongation of RNAP II on the DNA into a sequence of more granular events.

To exploit the stochastic event base approach in designing a framework to conduct system-wide *in-silico* experiments for eukaryotes cells, we endeavored to design several models that stochastically quantify the temporal dynamics of such events. We have developed a stochastic event based facilitated diffusion model which could be used for any membranae receptor binding event for both eukaryotes and prokaryote with different

paramors sets. The validation results is primarily shown for glucose uptake in *E.Coli* in chapter 2. The same model is used for insulin receptor binding in the *in-silico* experiment of mammalian cardiac myocyte. A variant of the same model is also applied to model fatty acid uptake in cardiac myocytes in chapter 6 and 7. To estimate the temporal dynamics of constitutive pre-mRNA splicing we developed a stochastic model that will be discussed in chapter 6.

Also we conducted a comprehensive study on chromatin dynamics and proposed stochastic discrete event based models for passive DNA access, DNA thermal fluctuations and chromatin remodeling in later chapters of this dissertation.

System-wide modeling of complex biological networks such as a cardiac myocyte cellular network requires execution of assortment sets of events pertaining to different molecular processes. Each event should be executed in a consistent order and timely manner meanwhile proper input resources should be supplied to the event being executed. Such capability is attributed to our work by designing and utilizing a Stochastic Discrete-Event Simulation engine.

1.2.3 Stochastic discrete-event based simulation

Stochastic Discrete Event-Based Simulation (SDES) has been successful in capturing the dynamic behavior of complex systems like internet traffic engineering and manufacturing systems. From there we were inspired to borrow the same concept to abstract the behavior of the complex system of a eukaryotic cell. In SDES the temporal evolution of the system between the state variables is govern by set of events with stochastic *holding time*. The *holding time* for an event is the duration of stay at current state, or more formally is the inter-arrival time between two events.

System engineering endeavors to abstract the complexity of a system in a set of discrete spacial and temporal variables [38] that can capture the behavior of system in

time. In a cell the entire system is a collection of interacting genes, gene products and small molecules distributed across compartments. The temporal behavior of these interactions that are inherently random in time and space is capture trough the events. We have designed a stochastic discrete event-based simulator architecture that can capture the dynamics of a eukaryotic cell functions at the system level by utilizing a class of stochastic models, such as those discussed earlier.

1.2.4 Modeling the human cardiac myocyte *in-silico*

Cardiac myocytes are essential in the perpetual blood pumping function of the heart and any damage to these cells could have sever consequences for the person. Hundreds of millions of people are suffering from chromic hear disease across the world, and Heart Failure (HF) mortality rate is still the highest in our time among all disease, despite the lack of ultimate cure for major chronic disease such as cancer and AIDS. Therefore having a heart cell *in-silico* which could used to conduct rigorous pathological and therapeutical *in-silico* experiments would indeed be advantageous for health science community as well as pharmaceutical industry. Hence, towards the horizon of an *in-silico* heart, in this dissertation by designing a novel software based on our simulation approach, we strive to bring the myocardial insulin signaling pathway and its metabolic effects *in-silico*.

Cardiac myocyte is among the most energy consuming cell types across the body. They are capable to exploit different molecular sources to meet energy requirements for their cellular functions. The two major sources of energy for these cells are Fatty Acid (FA) and Glucose. Amongst these two, glucose comes from normal diet of carbohydrate containing meals and the fatty acid is predominantly released to the blood from adipocytes and partially from the intracellular storage of triglyceride. Cardiac myocyte demonstrates a delicate interplay between these two sources of energy during exercise, normal and fasting condition. The intake of glucose to the cell is mainly controlled by

insulin, which is a hormone that secretes from the pancreas in the endocrine system. Insulin and its signaling pathway is a well studied topics in biochemistry, biology and medicine. Hence, to demonstrate the promising capacities of our simulation software in modeling the complex biological networks, following the design of the software we will conduct several *in-silico* on cardiac myocytes. Some of these *in-silico* experiments include: contributions of energy substrates to the myocardial energetics, plasticity of cardiac myocytes, the transcription effect of insulin signaling, and analysis of metabolic gene nullifications on production of energy precursors.

1.3 Motivation and Research Challenges

1.3.1 Motivation

The motivation for this research can be summarized as follows:

- Applying the network of connected components from network theory, along with abstraction concepts in system engineering to complex biological networks in eukaryotes to provide a system-wide picture of interacting processes in the biological networks.
- To Establish a platform that can realistically define the interaction between genes and genes products in the evolution of biological processes and easy to understand by biologists and medical professionals
- Often the physiology of bio-processes is too complex to be captured through a single rate constant. To properly mimic a the physiology of a cellular process a model should incorporate the inherent stochastically of biological process. Therefore, We need parametric stochastic model which is defined in the context of the cell not just as an isolated physical process. Developing biologically relevant models that

are modular (reusable) and would be applicable in system-level analysis of the cell in utmost desirable.

- Two major draw backs in current stochastic and deterministic simulation tools for biological networks include: (i) lack of their ability to overcome the stiffness problem that arises from several orders of magnitude deference among the temporal dynamics of cellular processes, and (ii) failure to capture the non-linear property of the cellular processes. Solving these two issues is essential for simulating the cellular functions at system-wide.
- Most important is to integrate the SDES models and its simulation algorithm into a software platform that could be unutilized for conducting *in-silico* experiments whose results could be validated with the equivalent empirical data.

1.3.2 Research challenges

Following is the list of major challenges we confronted in this research:

- Many knowledge gaps that exist in understanding of the physiology of a biological process or the molecular part pertaining to a process.
- Designing parametric models that can capture the stochastic behavior of cellular processes usually come across solving complex mathematical problems that demands significant efforts.
- The transcription regulatory network is different from cell in one tissue to the other tissue. Identifying the cell specific genes expression profiles and their transcription factors is formidable task and demand for extensive literature search.
- In eukaryotes the cellular processes are distributed among compartments, incorporating multiple compartments into system-level abstraction is a complex task.

- Lack of empirical data at the molecular level due to the technology and protocol limitations add another fold of complexity both for model validation as well as *in-silico* experiment verifications.
- The level of complexity that exits in and between the STN, TRN and MTN of the myocardial cell makes the mapping of the process to the SDES a particularly challenging task.

1.4 Contributions of this dissertation

The contributions of this dissertation could briefly short listed as follows:

- We devised a stochastic model for binding of exogenous molecule to the cellular membrane receptor that could be applied for both eukaryotes and prokaryotes.
- We have conducted a detailed study on nucleosome dynamics and provided probabilistic model for find the target DNA site genome wide. This model also incorporates the thermal fluctuation of DNA and its affect in gene regulation into the picture.
- We developed a Stochastic event based model for chromatin remodeling and combined that with the binding reaction model that is presented elsewhere. This combined model is utilized in the gene expression process of eukaryotic cell.
- A stochastic model is proposed for estimating the temporal behavior of the constitutive splicing.
- We designed an event based model for function of protein synthesis function in eukaryotes.
- An object-oriented database is proposed that stores the objects from the layouts of Signal Transduction Network, Transcription Regulatory Network and Metabolic Network for a typical eukaryotic cell.

- We have designed the *eukaSimBioSys*: a software that integrates our models, the database and the SDES simulation engine that could by utilized to study the dynamics of the biological networks in the cardiac myocytes.
- An objective based transcription regulatory network for the cardiac myocytes has been constructed by studying the insulin signaling pathway in human and rat cardiac myocytes from literature. The metabolic network of human obtained from an external database and was hand-curated to exclude unnecessary nodes and edges. The insulin and fatty acid signaling pathways in the cardiac myocytes are mapped into the event diagram and the event table for these pathways has been constructed.
- We have designed six in-silico experiments and utilized *eukaSimBioSys* to conduct these experiments. Five of these in-silico results have been validated with data published from wet lab experiments.

1.5 Organization of the dissertation

The rest of this dissertation is organized as follows: In chapter 2 we describe the architecture of our software and the stochastic discrete event simulation algorithm for a eukaryotic cell along with brief overview on the available simulation tools. Chapter 3 describes a stochastic model for the signal sensing on the membrane and facilitated molecule uptake (ligand-receptor docking). In chapter 4, first we give a detail description of nucleosome dynamics and outline its mechanism in regulation of gene expression from an event-based perspective. Then, a probabilistic model for target site access across a eukaryotic genome is derived. Finally a stochastic model for the effect of thermal fluctuations of DNA on functional DNA sites access is proposed. Throughout chapter 5 we propose a stochastic event based model for *in-cis* chromatin remodeling. We further utilize the proposed model to show that probability distribution function of the chromatin remodeling time follow a gaussian distribution. Chapter 6 contains a divers assortment

of materials required for preparing the in-silico experiment setup. These materials briefly include: the eventology of insulin signaling pathway, glucose and fatty acid uptake processes, a stochastic model for constitutive splicing, the eventology of protein synthesis, designing the event-table and event-diagram for *eukaSimBioSys*, and persistent signaling reaction model. Chapter 7 discusses six in-silico experiments that we conducted to study the effect of insulin on metabolism in cardiac myocytes using *eukaSimBioSys* software. We close the dissertation in chapter 8 by concluding our work and highlighting some of the future directions of this research.

CHAPTER 2

THE ARCHITECTURE OF THE STOCHASTIC DISCRETE-EVENT SIMULATION SOFTWARE FOR A EUKARYOTIC CELL *(eukaSimBioSys)*

2.1 Introduction

To comprehend the behavior of a complex biological system one must study them as a collection of interacting entities with interdependent regulatory control with feedback rather than isolated physicochemical objects. Such a system-wide view to a biological system could only become possible through simulation tools and techniques that indeed are capable to preserve such entity's collectivity and integrality throughout the spatiotemporal evolution of the system. With the speed and sophistication of computational devices, *in-silico* modeling and simulation techniques have become a powerful tools for biologists that are challenged with understanding system complexity of biological processes. The everyday knowledge being extracted from the genome and biological pathways by applying network and/or graph theory concepts could be coupled with mechanistic model to devise new simulation platforms. Discussions on hierarchy and taxonomy of different simulation and modeling techniques of complex biological systems could be found in [**?**, 39, 40, 41] and have been reviewed in [42].

2.1.1 Biological systems *in-Silico*

In recent years, the field of systems biology has been flourished with a wide spectrum of *in-silico* modeling and simulation methodologies to pursue the system-wide study of biological processes. Specifically physicochemical models based on equations from physical and chemical properties that describe the biochemical transformations such as

covalent modification, intermolecular association, intracellular transformation, etc. [43]. In [44] a nice categorization of these models and simulations have been proposed based on properties of four key parameters :

- *Time*: The temporal domain of the system can either be continuous (C) or discrete (D).
- *Space*: The state space of the system can be continuous or discrete.
- *System Evolution*: The evolution of the system can be considered in terms of being deterministic (D) or stochastic (S).
- *Physical Scale*: The model can abstract the system at a microscopic scale as in molecular dynamic simulations, macroscopic scale such as classical chemical kinetics base systems, or mesoscopic scale where individual molecules bodies are represented as in the current approach.

Most prevalent models of molecular networks including metabolic pathways, gene regulatory and protein-protein interaction networks, subscribed to the above paradigm. Such models consider the system as a set of coupled ordinary differential equations (ODE network) and use numerical approximation methods to capture the system dynamics deterministically in continuous time and space. A large number of computational tools, which provide a software platform for building, storing, parameterizing and solving sets of biochemical reactions using numerical techniques are available, like CyberCell [45](2006) ,Gepasi [46](2005) ,Promot/DIVA [47](2001), STODE [48](2001), Jarnac [25](2000). These kinetic rate based simulation tools have been widely applied to study gene expression and other molecular reaction systems [49]. Other non-kinetic rate based mathematical frameworks used for building computational models and simulation tools include stochastic Petri nets [50], stochastic process algebra PEPA[51], stochastic *pi*-calculus [52][53]. Agent based and object oriented simulation [54] is also used for simulation of biological systems. Agent Cell in [55] is a tool that complies agent based

simulation technique. A detail list of prominent modeling and simulation tools developed for studying the dynamics of complex biological systems along with their specifications is provided in Table 2.1.

Table 2.1. Comparative List of Biological Modeling and Simulation Softwares

Tool	***Modeling Technique***	***Spatial Representation***	***Temporal Evolution***	***Underlaying Model***
Agent Cell	Agent based simulation of biological systems	Not explicitly defined	Time-step driven	Agents model molecular behavior
Cellerator	Mathematical package for automatic equation generation and simulation for network of cells	Not explicitly defined	Continuous time	CCK
CyberCell	Atomistic modeling of biological processes with dynamic membrane geometry	Off lattice	Interparticle collisions	MD based
Dizzy	Stochastic simulation package	Compartmental	Supports CCK as well as SSA	CCK, CME
E-Cell	Object-oriented software suite for modeling, simulation, and analysis of large scale complex systems	Compartmental	Supports CCK as well as SSA	CCK, CME
FURM	Functional unit representation of biological processes	Not explicitly defined	Continuous time	Functional modeling
iSimBioSys	Stochastic modeling of discrete events	Compartmental	Event driven discrete time steps	Based on CME, explicit models of reaction time
JARNAC	ODE based	Not explicitly defined	Continuous time	CCK
M-Cell	Mont Carlo simulator of cellular micro physiology	Off lattice	Time-step driven	At surfaces, CME
MesoRD	Stochastic domain	Compartments, sub volumes	Event- driven	CME
Promot/ DIVA	Object-oriented paradigm	Not explicitly defined	Continuous time	CCK
SimBiology	Primarily ODE based simulation package	Not explicitly defined	Supports CCK as well as SSA	CCK, CME

Table 2.1 - continued

Smoldyn	Atomistic modeling of biological processes with dynamic membrane geometry	Off lattice	Interparticle collisions	MD based
Statecharts & DEVS	Discrete event system specification	Distinction between system and environment	Continuous/discrete time step	Atomic models (state transitions) and coupled models (component interactions)
Stochastic Π calculus	Abstract model of system based on DEVs	Cellular compartments	Continuous/discrete time steps	Processes model molecules and domain, communications model reactions
Stochastic Petrinets	Stochastic model of molecular interactive networks	Compartments	Continuous/discrete time steps	Graphical model
V-Cell	Continuous domain	Compartments, sub volumes	Continuous time	CCK, Mass action

2.2 Stochastic Discrete Event-based Simulation (SDES)

One of the powerful tools used extensively in *abstracting* system behavior for complex systems like traffic engineering and router design, complex social agent interaction behavior simulation, etc. is Stochastic Discrete Event based simulations (DES) [56]. In this approach the key notion is to abstract the complexity of the system as a set of discrete time and space random variables, which capture the behavior of the system in time. The entire system is a collection of functional blocks or modules, which are driven by a set of *events*.

Event: An *event* defines state transitions between a set of state variables accomplished within the event execution time. The system behavior evolves in time through the dynamic interaction of these discrete events, which change the system state (i.e. the random variables associated with them) based on the probability distribution of their execution times, more detail on SDES is could be found in [56]. The event paradigm provides the flexibility to the modeler to abstract the system at different levels of granularity based

on available data and the focus of investigation. A stochastic model associated with each event determines the *holding time* of that event based on the parametric probability distribution function encompassed by the model.

Holding time: For an event the *holding time* is a random delay whose distribution is govern by the associated model of that event (e.g. diffusion, , splicing, remodeling, reaction, protein conformation change, etc.), which represents the duration that system stays at the current state. On the other hand, the *holding time* of an event is the execution time of that event.

2.2.1 Bioevent identification and definition

Observations confirm that at the molecular level the cellular behavior arises from the stochastic interaction between molecular parts. Such observation in the intracellular environment is the key motivation in applying stochastic discrete event simulation in capturing the dynamics of a cellular function. Hence, identifying molecular functions and mapping those into sets of discrete events is a fundamental step to apply this approach to study the complex biological systems. In our simulation platform we refer to these events as *bioevents* and each *bioevent* has three attributes:

- The parametric stochastic model of the underlying physiological process associated with the *bioevent*. This model captures the physicochemical properties of the event through mathematical parameters and characterizes the holding time, defined earlier.
- The molecular resources associated with the *bioevent*, i.e the input to the model and the output from the model.
- The comparment(s) within which the event in being executed. i.e. cytosol, nucleus, peroxiosome, mitochondria, etc., one event might be executed across multiple compartments, e.g. diffusion.

Sensing signal S at the plasma membrane is a biological function through which the external signal is being transferred into the cell. Such a function could be used as an illustrative example of a *bioevent*. To identify this event we first identify the parametric model that can capture the holding time of this *bioevent*, which is the membrane receptor ligand ducking. The parameters for this model are the cell size, protein size, ratio of active receptors, etc. Input resources to this bioevent are counts of ligand (S) and receptor molecules (R). The output of the model will be the holding time and the change in the number of active receptors. This holding time will be randomly changing as the system states (counts of molecular entities involved) change and will accurately reflect the stochasticity of the cellular function. At the end of the holding time, the activated receptor molecule can trigger the subsequent event to cascade the signal further down into the cell. As the simulation proceeds at an event level, the resources states are determined in terms of the molecular counts of the various resources within individual compartment that has been affected by the executed event. For example, after the successful completion of the signal sensing, the count of extra cellular ligand and the free receptor in the system is decreased by one while that of activated membrane receptor is increased by one, one S molecule is utilized and one activated membrane receptor molecule (R_a) is created. This scenario shows how basic biological molecules and their events are identified, modeled and interlinked in a discrete event simulation framework.

2.2.2 Modularity and module reuse

The wide variability and complexity of cellular functions, resources, and possible sets of bioevents in the biological systems exacerbates the complication of the problem. However, there exists a core set of basic cellular functions that are redundant across wide variety of biological processes. Identification and modeling of these functions can greatly facilitate the study of complex processes of life and forms an integral part of the

computational effort. Some of the basic biomolecular functions, which are associated with key biological functions in eukaryotes, include:

- Reaction
- Diffusion
- DNA protein binding
- Transcription
- Splicing
- Chromatin Remodeling
- Receptor ligand binding
- Translation
- Transport
- Protein decay
- Protein folding

These set of functions could be used as functional modules to express the biological processes as sequence of reusable modules. Identifying the proper *'modularity'* of biological processes and characterizing their associated biomolecular events is a key in the SDES *in-silico* study of a cellular process and can speed up the *in-silico* experiment set up many folds.

2.2.3 Temporal evolution of SDES and extraction of intelligence

In discrete event simulation, *simulation time* is the representation of the *physical time* of the system being modeled. Each event is associated a *time-stamp* indicating when that event occurs in the physical system being simulated. The event *time-stamp* is computed from the knowledge of the previous event that has triggered the current event, together with the event holding time which is a realization of the random number characterize the event dynamics. The dynamics of resource utilizations with progression in

time unveil the complete internal picture of a complex biological process at the molecular level. The check-list for characterizing the system parameters is performed as follows:

- Identify the list of discrete events that can be included in the model based on the available knowledge of the system.
- Identify the resources of interest for the execution of the event function which are being used by the biological process for each discrete event.
- Compute the time taken to complete this biological discrete event, i.e. the holding time of the discrete event. For this purpose, it is important to mathematically relate all the event parameters which affect the interaction of the resources in a particular biological function. The event execution time is a random number drawn from a probability distribution identified by its two significant moments (mean and variance) only.
- Identify the next biological event or set of biological discrete events initiated on the completion of an event. If multiple discrete events are possible after completion of an event, the next event can be chosen probabilistically, based on the biological pathway of the process that is being modeled. This probability calculations depends on the event set and the properties of the events in the set.

Once the above check list is satisfied the discrete event simulator scheduler which is a time prioritized event queue pops individual events from the queue and system proceeds. As each event is executed, it updates the molecular resources of the system, updates the system time and populates new events into the event queue from its *next- events* list. In this manner, as the execution of events evolves across time and the temporal changes in cell characteristics associated with the biological process is extracted. The temporal evolution of the events is shown in Fig. 2.1 where it depicts the order in which the event arrive to head of the queue being executed. The algorithm of the simulation is depicted in Fig. 2.2.

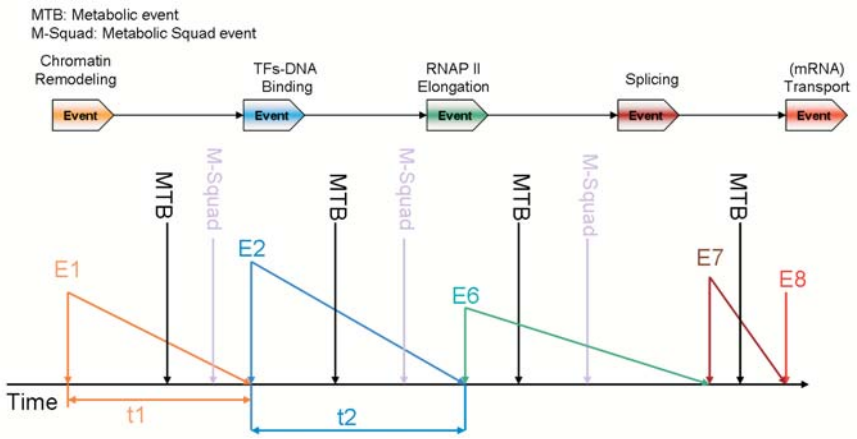


Figure 2.1. Temporal evolution of a typical system in SDES.

This algorithm briefly functions as follows: Takes the event from the head of the queue, identifies its resources, associated parametric models for their holding times and the list of next triggered events. If all the resources associated with current event were available then it executes the event which consists of updating the molecular resources of the system, updates the system time and populates new events into the event queue from its *'next events list'*. Then the algorithm checks current simulation for the total simulation time and on the failure repeats the above cycle. This algorithm forms the heart of the *eukaSimBioSys* and is integrated in the simulator engine whose architecture is depicted in Fig. 2.3.

2.2.4 Incorporating compartmentalization in SDES

In eukaryotes distribution of processes within or across compartments adds another level of complexity to the SDES of cellular processes. In a compartmentalized environment there will be two superclass of events apart from their original types, the *local event* (e_L) and *cross compartment event* ($e_{(c_i,c_j)}$), where (c_i, c_j) are the source and destination compartment across which the event is being executed. Hence the execution

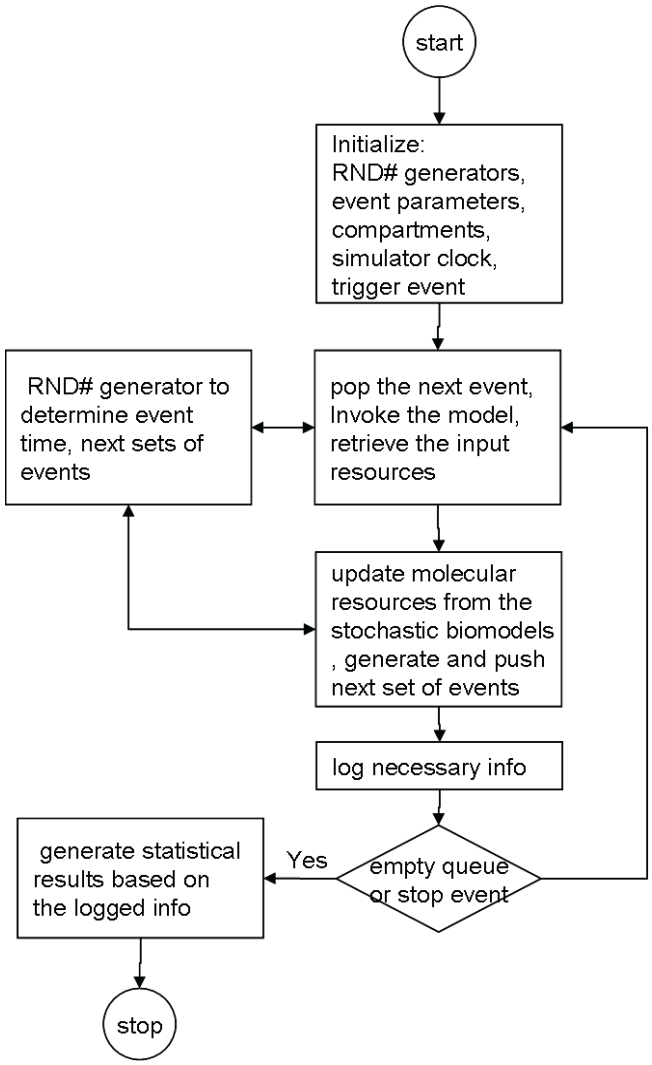


Figure 2.2. The SDES algorithm utilized by *eukaSimBioSim*.

of a local event will effect only a single table within the resource data base, where a cross compartment event will affect multiple resource tables.

For metabolic reaction event we follow the *lumped metabolic event* which is based on flux balance analysis and will be discussed on chapter 6. Applying such strategy demands to keep all the metabolic events local to the participating compartments. Keeping all metabolic events local will cause metabolite explosion in some compartment (e.g. mitochondria) and metabolite starvation in the others (e.g. cytosol). To overcome this com-

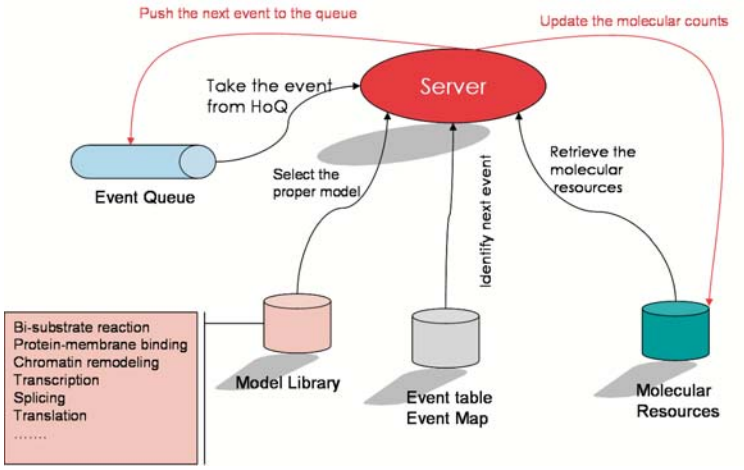


Figure 2.3. Architecture of *eukaSimBioSim* engine.

plexity we define a new cross compartment event called *metabolite squad event (MetabSquad)* which avoids metabolite explosion or starvation. The *MetabSquad* event will be executed regularly every τ *squad* unit of time and redistribute the metabolite across pairs of *neighbor compartments* with a defined ratio $\Omega(c_i, c_j) \leq 1, \quad \forall i, j : i \neq j$ where c_i and c_j are neighbor compartments. We should clarify that any pair of cellular compartments that can have direct molecular transport between themselves are called *neighbor compartments*. In the current version of the software the above distribution is manipulated manually in an iterative try and error fashion.

2.3 Architecture of the *eukaSimBioSys*

The architecture of the current simulation software for biological systems in eukaryotic cell, *eukaSimBioSys*, This architecture is has a common base with *iSimBioSys* [57] which was originally developed for prokaryotic cells. Some of the new features that are added to the architecture of *eukaSimBioSys* include:

- Extended simulation engine that can handle significantly more complex event networks.

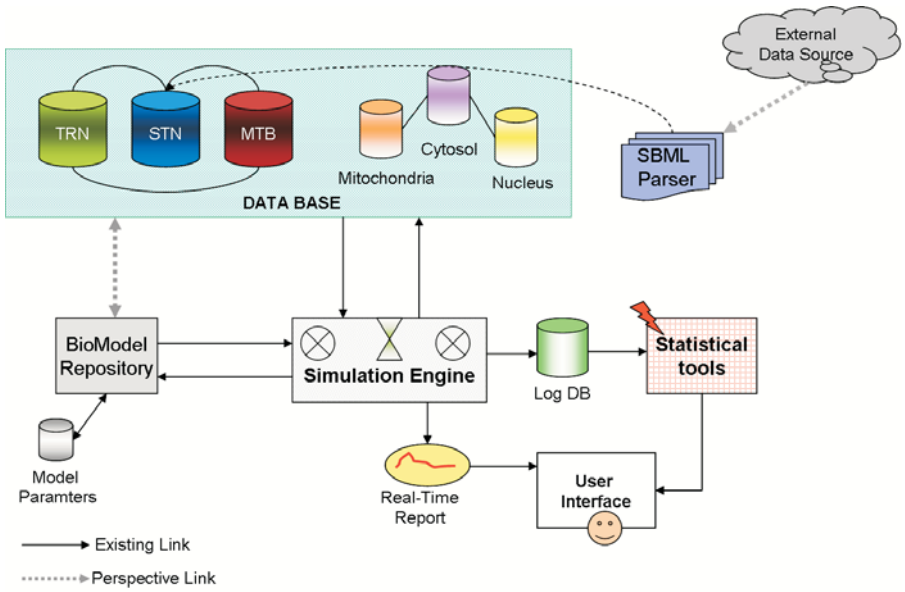


Figure 2.4. The architecture of *eukasimBioSys*.

- Ability for the engine to use dynamic model parameters (change the parameters on fly) based on the given event table.
- Enhanced model repository to handle the eukaryotic cell processes by adding new evetologies[1] for complex processes and approximation mechanism.
- Approximation techniques included to avoid stiffness and enhance the simulation efficiency.
- Implementation of the compartmentalization in the simulation engine.
- SBML enabled connectivity with external databases.

The architecture of the *eukaSimBioSys* is depicted in Fig. 2.4.

In this architecture the information which includes the layout of the signal transduction, transcription regulatory and metabolic networks along with genes, substrates, protein, compartments, etc. and their attributes would store in an object oriented database. There is a data processing module that enables the system to directly access the Excel

[1]We define *'evetology'* as the process of identifying and executing the sequence of explicit and dynamic events for a biological function which will be described in chapter 6.

sheets and parse the information from the SBML files or export the current database into SBML files. The model repository stores all the stochastic physicochemical models along with their parameter sets and grants the engine to access them to manipulate the events' holding times. There is a separate database that stores all system logs including event execution log, the resource consumption log, etc. The user interface provides a unfriendly environment for the user along with realtime reports and graph on the simulation and system status. Static tools are collection of Matlab® scripts that could be utilized to provide statistical analysis on output data.

2.4 Summary

we started the chapter by giving an overview on *in-silico* simulation of biological system. Then a comparative list of prevalent simulation tools and techniques was presented. Later we introduced the concept of stochastic discrete event simulation (SDES) for biological systems along with key terms and components. The mapping of a biological process into an event based formalism applicable in SDES is further discussed and the algorithm of the SDES was also presented. The modularity advantage of SDES approach in addition to the complexity that compartmentalization imposed to the system was discussed in details. The variant of stochastic discrete event simulation algorithm that could be applied in studying the complex biological system was introduced. Finally we proposed the architecture of *eukaSimBioSys* which is a software tool designed based on discrete event simulation paradigm for *in-silico* study of the dynamics of a eukaryotic cellular functions.

CHAPTER 3

STOCHASTIC MODEL FOR LIGAND DOCKING TO THE MEMBRANE-RECEPTOR

3.1 Introduction

Mathematical modeling provides a systematic formalism for capturing molecular details in a physiological context. These models could further be stored in dynamic repositories and applied in computational studies for uncovering biological insights. In a challenge to construct a model which is biologically relevant and computationally efficient, in this chapter we endeavor to devise a stochastic model to characterize the temporal behavior of membrane receptors and bring about new aspects of the bio-process. Traditionally these aspects has been shadowed by the kinetic rate law based approach by using a single rate constant parameter. In this work we first we present a novel random-walk approach to detect the temporal behavior of docking an exogenous ligand to a membrane receptor. Then we further elaborate this model to determine the flux of facilitated molecule uptake (e.g. glucose, mannose, Ca, Mg, etc.). Such a model could be applied for capturing the behavior of variety of cellular functions including: sensing a hormone signal, agonist ,antagonist, metabolite uptake, etc. The model that we propose here, is receptor limited stochastic discrete event based model and aimed to capture the temporal behavior for process of the ligand docking to the receptor in the membrane of the cell. Two variants of this model will be proposed for eukaryotic and prokaryotic cells.

3.2 Physiology of the process

The fate of a cell in any organism dependents highly on its communication with extracellular environment. For a multicellular organism this communication comprises the exchange of signals between organs, cells, and even between organelles inside the cell is of vital importance. An exogenous signal to cell might include: olfactory, gustatory signals, food substances, antigens, toxins, drugs, change of temperature, ions, pressure, etc. [58].

Majority of membrane receptors are transmembrane domain receptors, which indicates that such receptors can have a molecule docked to one domain (cytoplasm or extra cellular) and triggers a response on the other domain. Such a response could result a conformation change, autokinesis, phosphorylation, etc. Also the ligand and the membrane protein may form an ion pore and once the ligand binds to the membrane receptor the ion channel becomes accessible to the ions. The polypeptide chain of the transmembrane receptors might cross the membrane lipid-bilayer one or multiple times like G-protein coupled receptors (GPCR). One consequence of such structure is the increase in surface of membrane-domain for the receptor, thus higher sensitivity to the signal would be expected.

Should the transmembrane receptor be used to transport external species across the plasma membrane, such transport could be classified into (i) passive which does not require further energy for the transport, and (ii) active transport where the transport process require extra energy resource, therefore it is coupled with hydrolysis of ATP molecules. In the current model we assumed that all the transports are of the passive form.

3.3 Model overview

In a typical receptor ligand binding, the particle in extra cellular environment will need to come to the vicinity of the cell and passes through the interphases and their intermediate environments to reach the organism. Each of these environment(s) and interphase(s) that segregate them characterize a distinct interaction with the particle. Therefore we need to define a cell abstract that addresses these environments and their interphases. For this purpose we employ variants of a cell abstraction as proposed in [59] for eukaryotes and prokaryotes, as follows:

eukaryotic cell abstract:consists of following interphase and environment

1. diffusion-layer (environment)
2. plasma membrane (interphase)
3. cytoplasm (environment)
4. nuclear envelope (interphase)
5. nucleus (environment)

prokaryotic cell abstract:consists of following interphase and environment

1. diffusion-layer (environment)
2. bilayer membrane (interphase)
3. periplasmic region (environment)
4. plasma membrane (interphase)
5. cytoplasm (environment)

We establish our model based on the above cell abstractions throughout the chapter. Diffusion layer (radius) is a sub-region of the medium around the cell in which the particles will effectively participate in the process of diffusion. A quantitative definition for this region could be the average distance that a substrate travels to reach the plasma membrane or periplasmic region; we will refer to this stage of diffusion as the *medium phase*. On the plasma membrane the ligand will need to bind to the receptor, possibly

activated in advance, to conduct the signal or be transported to the cytoplasm. This step of the process is articulated in the *plasma phase* modeling step. Notice that for the prokaryotic cell prior to reach the plasma membrane in narrow periplasmic region the substrate will float down the concentration gradient to reach the plasma membrane. A sequence of cytoplasmic reactions (e.g. kinesis) change the state of a receptor from inactive to active and in our model is abstracted by the *cytoplasmic phase*.

In the course of modeling, the first two phases form two microevents for which we come up with individual equations that characterize their temporal behavior. The effect of the third stage will be captured through a parametric constrain function, which will be integrated to its preceding micro-event as a limiting factor. Ultimately, by integrating these parts we form the total time equation for the entire process. Also to better embark the idea and avoid ambiguity, meanwhile molding the process we do not follow the above phase order in establishing the model.

Without loss of generality we make following initial assumptions for the modeling purpose: 1-Spherical shape for particle and the cell 2-Flat circular receptor cite for membrane bounded protein. 3-Uniform distribution of receptors across the membrane surface. Also any other assumption will be addressed in the respective models place.

3.3.1 Random-walk of the particle

A particle in an aqueous environment will constantly step in different directions enormous number of times per second. The probability of a particle to move x unit(s) in one direction will follow the Gaussian distribution [60]. If $p = q = 1/2$ be the probability of fluctuating back and forth in the same direction $\hat{i}$, we will have:

$$P(x,t)dx = \frac{1}{(4\pi Dt)^{1/2}} e^{-x^2/4Dt} dx \tag{3.1}$$

from (3.1) we can find the *expected movement* of particle in time t in any euclidian direction $\hat{i}$, giving:

$$\bar{x_t} = \int_0^\infty \frac{x}{(4\pi Dt)^{1/2}} e^{-x^2/4Dt} dx = \sqrt{\frac{Dt}{\pi}} \tag{3.2}$$

from (3.2) and simple rules of geometry for a motion in 3D space we can write:

$$\bar{r}^2 = \bar{x}^2 + \bar{y}^2 + \bar{z}^2 = \frac{3Dt}{\pi} \Rightarrow t = \frac{\pi \bar{r}^2}{3D} \simeq \frac{\bar{r}^2}{D} \tag{3.3}$$

3.3.2 Medium phase

3.3.2.1 Eukaryotic cell

The ligand substrate would float in vicinity of the cell prior to reaching the membrane surface and binding to the receptor. We envision a motion pattern for the particle as the onein Fig. 3.1. The presented floating-dome is a hypothetical cone volume to represent the motion space of the transport mechanism for a molecule. In our analysis, we are considering a single molecule traversing through the space in random-walk. The floating particle will move in this environment and maintain a constrained random-walk with monotonically decreasing distance from the surface of plasma membrane. With this assumption, the landing area of the particle will be the outer surface of a spherical cap [61]. This area is bounded between the two tangents from the initial starting point of the moving molecule. We also assume that the molecule has uniform probability to hit the collision surface of the cell spherical cap. If x be the vertical distance of the particle

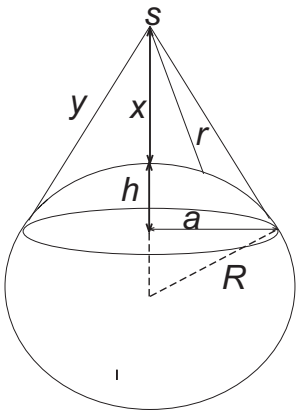


Figure 3.1. Floating-dome: the hypothetical cone which includes the ultimate floating pattern that a substrate might follow in the vicinity of cell before reaching the plasma membrane.

from the plasma membrane, hence considering Fig. 3.1 and spherical cap properties we can infer that:

$$\begin{aligned} y^2 &= x^2 + 2xh_{max} + 2Rh_{max} \\ y^2 &= (R+x)^2 - R^2 \\ \Rightarrow\ h_{max} &= \frac{Rx}{R+x} \end{aligned} \tag{3.4}$$

To find the average time for touching down on the membrane we consider and infinitesimally small segment of the spherical cap (from a ring formed by $r_x \leq y$) having equidistant from point x, then we get:

$$r_x = \int_r r_x p(r) dr \tag{3.5}$$

Also the probability of each path r will depend on their share of the relative segment area in the entire sphere cap, so:

$$p(r)dr = \frac{2\pi R dh}{2\pi R h_{max}} = \frac{dh}{h_{max}} \tag{3.6}$$

By substituting the above expression in Eqn.3.5, integrating the expression and applying the boundary condition: $h\epsilon[0, \frac{Rx}{R+x}]$, we will have:

$$\bar{r}_x = \frac{(x^2 + 2Rx)^{3/2} - x^3)}{3Rx} \tag{3.7}$$

Eqn.3.7 gives the mean distance for only one realization of the x, if we integrate Eqn.3.7 over x, divide by R_p, and apply the boundary condition $x\epsilon[0, r_p - r_l]$, then we get the expression for general case r. Now if r_N be the width of the diffusion layer around the cell and r_L be the ligand radius; since $r_L << r_N$ we can neglect that in calculation, therefore:

$$\bar{r} = \int_x \bar{r}_x p(\bar{r}_x) d\bar{r}_x = \frac{R}{18\delta} \times \Bigg(3ln(1 + \delta + \sqrt{\delta(2+\delta)} - 2\delta^3) + \sqrt{\delta(2+\delta)}(3 + 7\delta + 2\delta^2) \Bigg) \tag{3.8}$$

In Eqn.3.8 δ is the ratio of diffusion layer to the cytoplasm radius R. Using U to abstract the coefficient of R in Eqn.3.8 and substituting that in Eqn.3.3 we get:

$$\tau_p = \frac{\pi U^2 R^2}{3D} \tag{3.9}$$

In the above expression τ_p is the mean travel time of the molecule to the plasma membrane, e.g. mean time a substrate molecule requires to reach the plasma membrane surface.

3.3.2.2 Prokaryotic cell

For prokaryotic cell the plasma membrane is a periplasmic interphase between the cell wall and the plasma membrane therefore a particle needs to diffuse to the cell wall through pores and float down the periplasmic region to reach the plasma membrane, therefore for the periplasmic region we can still use Eqn.3.8 by replacing δ with ratio of periplasmic region to cytoplasm radius.

3.3.3 Plasma phase

This section could be applied for both eukaryotes and prokaryote with appropriate parameter for each case. On the plasma membrane, random motion of the metabolite forms a two dimensional random-walk. This random-walk is projected on a hypothetical flat grid that covers the entire membrane, the flatness is due to significant size difference between the particle and the cell. The area of each grid element is equal to one swept during the Brownian motion in Δt. This schema is shown in Fig. 3.2. Since the Brownian motion has a thermal character, each grid area $S \propto f(T)\Delta t$, where $f(T)$ is the Maxuell-Boltzmann molecular velocity at temperature T and $\Delta t(\simeq 10^{-13})$ is a single Brownian motion time step. Our objective is to find τ_r, the average time for a metabolite to find an active receptor. If p be the probability of gird element (site) be an active receptor, therefore, the site access time τ_r will be:

$$\begin{aligned} \tau_r &= \sum_{i}^{\infty}\sum_{j}^{\infty}(i+j)p(1-p)^{i+j-1}\Delta t \\ \Rightarrow &= \frac{2\Delta t}{p^2} \end{aligned} \tag{3.10}$$

Also τ_r^2 the second moment of site access time can be obtained from:

$$\begin{aligned} \tau_r^2 &= \sum_i^\infty \sum_j^\infty (i+j)^2 p(1-p)^{i+j-1} \Delta t \\ \Rightarrow &= \frac{(3(2-p)-p^2)\Delta^2 t}{p^2} \end{aligned} \tag{3.11}$$

In Eqns.(9,10) i and j are the number of steps in x and y coordinates respectively. The

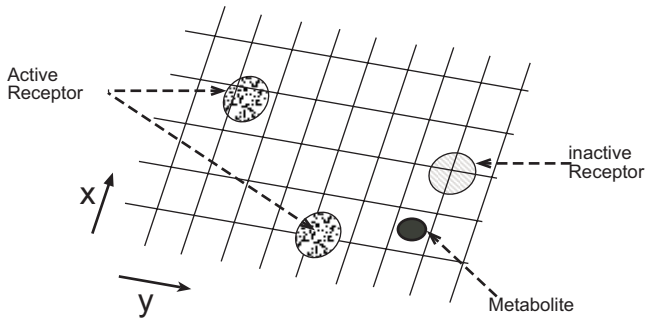


Figure 3.2. Two dimensional random-walk of the particle on the plasma membrane grid, a receptor might span over more that one grid element, where either an active or silent receptor.

probability of a site being an active receptor is a composite stochastic $p = \rho \cdot \varphi$, where ρ is the ratio of area that is swept by the cross section of substrate on membrane to the over all surface area of the membrane occupied by $[C_e]$, receptor concentration. Next parameter is φ, which is the probability of a site being active depends on cytoplasmic parameters that will be discussed in the next section.

$$f(U,T)du = 4\pi(\frac{m}{2\pi k_b T})^{3/2} U^2 e^{\frac{-mU^2}{2k_b T}} du \tag{3.12}$$

Following the Maxuell-Boltzmann distribution of molecular velocity U in Eqn.3.12, for a particle of mass m at temperature T, the average velocity will be as follow:

$$\begin{aligned} f(T) &= \int_0^\infty 4\pi(\frac{m}{2\pi k_b T})^{3/2} U^2 e^{\frac{-mU^2}{2k_b T}} dU \\ \Rightarrow &= \sqrt{\frac{8k_b T}{\pi m}} \end{aligned} \tag{3.13}$$

The mapping of substrate motion on the membrane during arbitrary time t will resemble a pattern shown in a Fig. 3.3. This pattern in an infinitesimally snap shut of Δt could roughly approximated by the area of the rectangle whose sides are $f(T) \cdot \Delta t$ and $2(R_e + R_L)$, in which Re and R_L are the radiuses of the receptor and substrate, respectively. Now if R_{cyt} represents the cytoplasmic radius we can write the expression for ρ as following:

$$\rho = \frac{2000}{3}\sqrt{\frac{8k_b T}{\pi m_L}} N_A C_e (R_L + R_e) R_L \Delta t \tag{3.14}$$

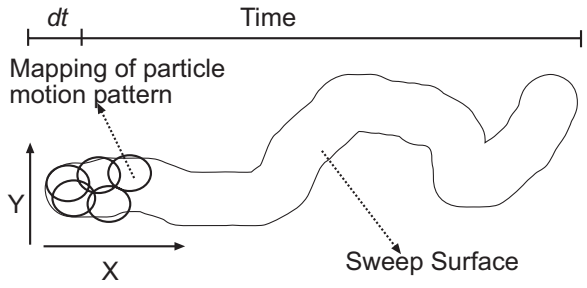
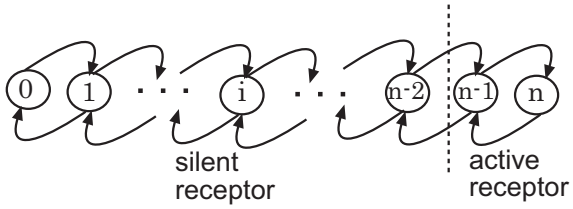


Figure 3.3. Projection of random motion pattern of the metabolite on the plasma membrane, in a dt could be approximated by a rectangle area.

3.3.4 Cytoplasmic Phase

The cytoplasmic domain of the membrane receptor might be subjected to one or more reactions (e.g. kinase, phosphatase, etc.) prior to extracellular domain of a membrane receptor becomes sensitive to a ligand or metabolic substrate. To better advise the modeling approach we give an abstract definition to this manifold as follow: A *n-ligands* activated receptor site, is an individual receptor that must have n ligand molecules docked to their sites on protein structure to switch the designated receptor '*ON*' for performing a objective task. From the biochemical perspective, n reaction of the type in expression 3.15 must take place in the cytoplasm prior to the activation of a target site. We assume that the activating reactions can happen in arbitrary order. Therefore, we can write:



$$\mathrm{R_i + E{-}L_j \rightleftharpoons R_i{-}L_j{-}E \rightleftharpoons R_{i+1} + E} \tag{3.15}$$

In the above reaction R_i is the receptor with i ligands already docked to its structure, E is catalyzing enzyme and L_j is the ligand of type j. Our objective is to find φ the ratio of sites activated by ligands. For this purpose we use a linear state diagram in which node i is the concentration of the receptors with i ligands already docked. Also each intermediate product will be considered as a separate node, therefore for a *n-ligands* activated site we will have *2n* states. Since the states *n-1* and n are the only states

that have all the ligands docked to the receptor; Hence, if $C_e(i)$ be the concentration of receptor e in state i then:

$$\varphi = \frac{C_e(n)C_e(n-1)}{C_e} \tag{3.16}$$

Values for $C_e(i)$ will be taken form the corresponding experimental results. Also we can find the average number of active receptors from:

$$n_e = 10^3 C_e V_{cyt} N_A \varphi \tag{3.17}$$

Now from Eqns.(3.8,3.16) we can get the expression for p and complete the τ_r equation. Notice that for the class of receptors that does not require cytoplasmic activation we always need to set $p = 1$.

3.4 Bulk diffusion and flux

So far all the temporal quantities that we have calculated consider only a single metabolite substrate uptake; however, in a real scenario the uptake happens in bulk and this will significantly affect the time equations that we derived earlier. Since our model follows an enzyme (receptor) limited approach, at any time multiple substrate will target one receptor. This ratio could be calculated from:

$$\begin{aligned} \kappa &= \frac{C_{mtb} N_A 4/3\pi((R_p + R_{cyt})^3 - R_{cyt}^3)}{\varphi C_e N_A 4/3\pi R_{cyt}^3} \\ \Rightarrow &= \frac{C_{mtb}((\delta+1)^3 - 1)}{\varphi C_e} \end{aligned} \tag{3.18}$$

Having κ multiple substrate molecules present simultaneously at each phase per receptor can reduce the time corresponding to that phase by the same factor. Earlier we calculated

τ_r for a single ligand but usually $\kappa \geqslant 1$. Here we need to revise the τ_r such that it can incorporate both the one-to-one and one-to-many substrate-receptor matchings. We break down the problem into $\kappa = 1$ and $\kappa > 1$ and offer a model for the former and then generalize that to solve the latter. In the case where $\kappa = 1$ we have $[Ce]$ simultaneous receptor finding events in Δt. To circumvent this parallelism, we project the n parallel matching (receptor finding) events into n serialized atomic matching events (microevents), with this projection we would have different matching time for each atomic event. Since number of receptors for each later microevent would be lessened by one form immediate earlier one therefore we can write:

$$\tau_{r,1} = \frac{n_e \tau_r}{n_e} < \tau_{r,2} = \frac{n_e \tau_r}{n_e - 1} < \dots$$
$$\dots < \tau_{r,i} = \frac{n_e \tau_r}{n_e - i + 1} < \dots < \tau_{r,n} = n_e \tau_r$$

since we assumed that n events are atomic event, therefore no new receptor will be added in the meantime. Generalizing this fact we can infer the average time required for κ substrate to one receptor as:

$$\begin{aligned} \tau_{r,\kappa} &= \frac{1}{\kappa n_e}\left(\frac{n_e \tau_r}{n_e} + \frac{n_e \tau_r}{n_e - 1} + \dots + \frac{n_e \tau_r}{2} + n_e \tau_r\right) \\ \Rightarrow &= \frac{H(n_e)\tau_r}{\kappa} \end{aligned} \tag{3.19}$$

From Eqn.3.19 we can find t_γ, the cumulative diffusion bulk time for the floating and plasma phase as follows:

$$t_\gamma = \tau_{p,\kappa} + \frac{\tau_p}{\kappa} \tag{3.20}$$

Now if we assemble all the manipulated times for different stages and incorporate the affecting bulk parameter we can write the expression for flux of the signal transport or metabolite uptake to the cytoplasm for a cell as following:

$$J = \frac{n_e \kappa}{\kappa \tau_{r,\kappa} + \tau_p + l^2 D^{-1}} \tag{3.21}$$

3.4.1 Estimating the width of diffusion layer

A substrate in the extra-cellular environment travels a random distance to approach the cell. Due to the homogenous distribution of substrate in the medium, for higher concentrations of substrate lesser time is required since there are sufficient number of substrate available in the vicinity of the cell. By increasing the receptor concentration cell becomes more aggressive in absorbing the substrate; however, as the ratio of the extra cellular substrate to the receptor decreases, the effect of the diffusion layer as a constraining factor becomes more significant. In order to capture this behavior we utilized the well-known capillary diffusion model [62] for particle diffusing form an aqueous environment to the vicinity of the plasma membrane (eukaryotic cell), or to periplasmic region (prokaryotic cell). Considering t_m the time for particle at distance l to diffuse to the cell we can have:

$$t_m = \frac{l^2}{D} \tag{3.22}$$

In the above, l is the capillary length (diffusion radius) and D is the diffusion constant. To find the diffusion radius we define $W(\Re, D, t_\gamma, \varepsilon)$ as a training function to train the value for diffusion radius from the real experimental data set $\Re$, diffusion constant D and t_γ. Here t_γ is the cumulative bulk model time for the *plasma phase*. Also, ε denotes the

acceptable root mean square of error (***RMSE***) for the trained value. Now if we have s sets of $(\Re(i), t_\gamma(i))$ tuples by using the following function one can train the value of l:

$$W(\Re, D, t_\gamma, \varepsilon) = \sqrt{\frac{\sum_{i=1}^{s}(\frac{l^2}{D\kappa} + t_\gamma(i) - \Re(i))^2}{s}} < \varepsilon \tag{3.23}$$

3.5 Model validation and results

The validation of the model for the prokaryotic case will be for *E. Coli* which is a Gram-negative bacteria. The eukaryotic validation will be addressed along with the results of in-silico experiments in chapter 7.

In Gram-negative bacteria for any salute or particle to reach the cytoplasm must pass trough a lipid bilayer (outer membrane), the cell wall (periplasmic space) and the plasma membrane. The former maintains the shape and protects the organism from invasions and osmoticlysis [58]. Periplasmic space in the Gram-negative bacteria maintains the free enzymes that work as components of transport system; these receptor proteins will move specific molecules down a concentration gradient trough the membrane [58]. The receptor proteins are spanned over the surface of plasma membrane where these receptor proteins receive the substrate from the medium, couple them to a chemical conversion and release them in the cytoplasm, referred to as passive metabolite transport [63]. Most often this mechanism in an organism is used for number of different substrate transport with the different receptor proteins, for this reason they are also referred to as group transporters. *E.Coli's* phosphoenolpyruvate transferase system (PTS) is among the most well-known group transporters. The *PTS* pathway is shown in Fig. 3.4 where the details for the pathway is available in [64, 65]. We will use the pathway in Fig. 3.4 along with glucose as the substrate to validate our model.

$$\mathrm{pep} + \mathrm{Hpr} \overset{\mathrm{E\ I}}{\rightleftharpoons} \mathrm{P{-}Hpr} + \mathrm{Pr}$$
$$\mathrm{pep} + \mathrm{Cys} \overset{\mathrm{E\ I}}{\rightleftharpoons} \mathrm{P{-}Cys} + \mathrm{Pr}$$
$$\mathrm{P{-}Hpr} + \mathrm{IIA} \rightleftharpoons \mathrm{Hpr} + \mathrm{P{-}IIA}$$
$$\mathrm{P{-}Cys} + \mathrm{IIB} \rightleftharpoons \mathrm{Cys} + \mathrm{P{-}IIA}$$
$$\mathrm{P{-}IIA} + \mathrm{IIC} \rightleftharpoons \mathrm{P{-}IIC} + \mathrm{IIA}$$
$$\mathrm{P{-}IIC} + \mathrm{S_{out}} \rightleftharpoons 6\,\mathrm{P{-}S_{in}} + \mathrm{IIC}$$

Figure 3.4. (a) and (b) denote two different phosphoenolpyruvate PTS pathways in *E.coli*; Enzyme II is substrate specific which follows the capability of PTS in group transferase.

Table 3.1. Parameters Value

quantity	value	unit
intra cellular volume	2.5	μliter/mg
enzyme m [1]	63.652	kDalton
glucose m	0.1822	kDalton
glucose r [2]	5	$\mathring{A}$
enzyme r	50	$\mathring{A}$
cytoplasm r	0.543 μm	
periplasmic width	15.0	nm
D	6.7^{-6} cm^2/s	
T	300	kelvin

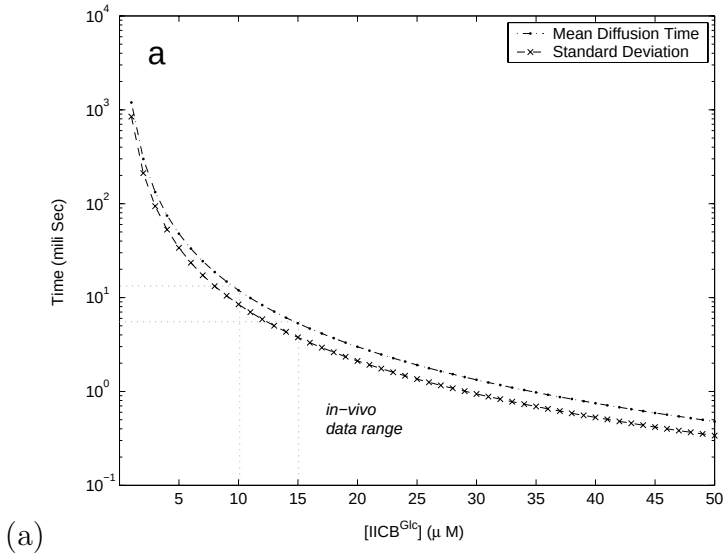
3.5.1 Results

In this section we show numerical calculation for validation of our model with comparison to some of the available kinetic models and *in-vivo* assays. Note that Kinetic models are not closed form parametric expressions and thus not suitable for fast computation. The numerical calculations is done for the *E.Coli*, the cell physical and biochemical specifications are based on data reported in Ecocyc [65] and CCDB [66]. Other parameters used in the calculations are listed in Table 3.1 and are taken from literature.

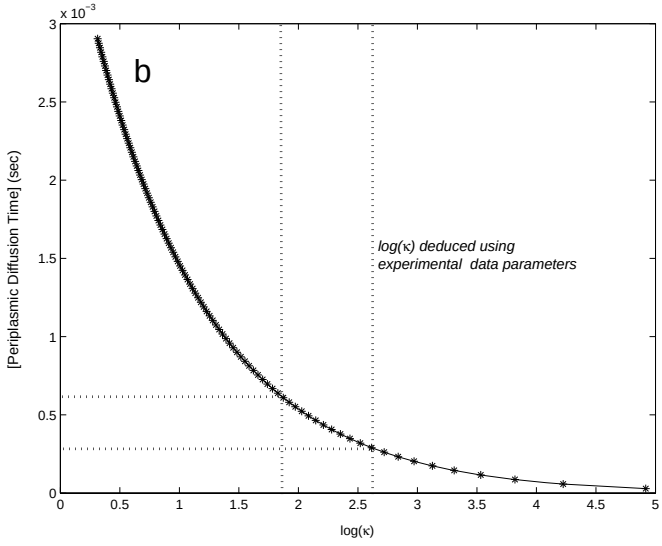
Fig. 3.5.(a) shows that for a single molecule τ_r follows the shape of an exponential decay function. Also, since the standard deviation of time is very close its mean τ_r, thus, we can approximate the probability distribution function of τ_r, average time to

find an active receptor single molecule model, with a negative exponential distribution of function as follows:

$$p(\tau) = e^{10^{-9}[C_e]\tau} \tag{3.24}$$



(a)



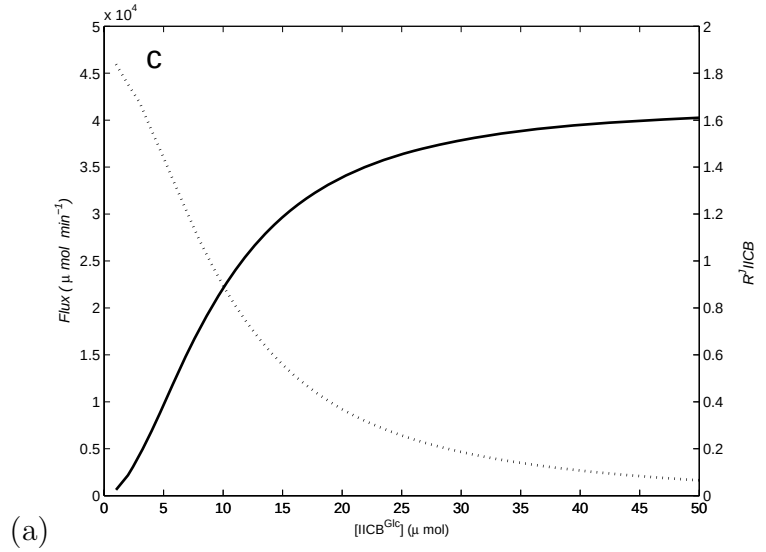
(b)

Figure 3.5. (a) average time along with its standard deviation to find an active receptor in the single metabolite molecule mode for various enzyme concentrations; (b) periplasmic float time versus $log(\kappa)$ which is the metabolite per enzyme ratio.

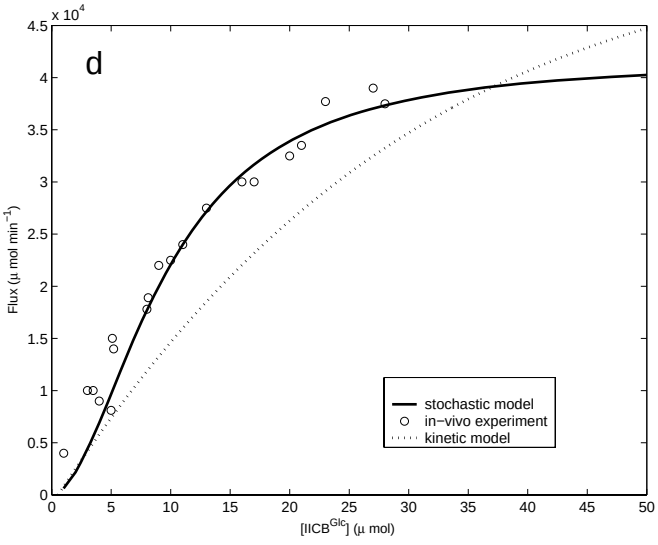
Fig. 3.5.(b) depicts the predicted time for the metabolite in the periplasmic region to follow the conical hypothetical path in the bulk model with respect to the *log* ratio of metabolite to the receptor counts. This result indicates that the sensitivity of the periplasmic time to the *log* ratio is significantly decreased after $\kappa = 12$, where that lays in the region reported from experimental observations in [67]. The reported experimental data in [67] is obtained at equilibrium which confirms that the model has captured the equilibrium condition by demonstrating a low sensitivity in inducing the metabolite concentration.

$$R^J_{q_i} = \left(\frac{\partial |J|}{\partial q_i}\right)_{q_n} \times \frac{q_i}{J} = \left(\frac{\partial ln|J|}{\partial ln q_i}\right)_{q_n} \tag{3.25}$$

Fig. 3.6.(a) demonstrates the sensitivity of the flux and flux response coefficient (FRC) with respect to the enzyme variations. Flux response coefficient is calculated form the Eqn.3.25, where q_i is the modulated parameter and p_n represents all other parameters that remain constant. We can observe that these two quantities are inversely proportional. This observation follows the same behavior as reported in [68]. This fact is intuitively true as well, since due to the effect of diffusion radius and constant metabolite concentration in a closed system addition of a small amount of receptor enzyme as the system reaches its maximum flux capacity will lose its perturbational effect to the system.



(a)



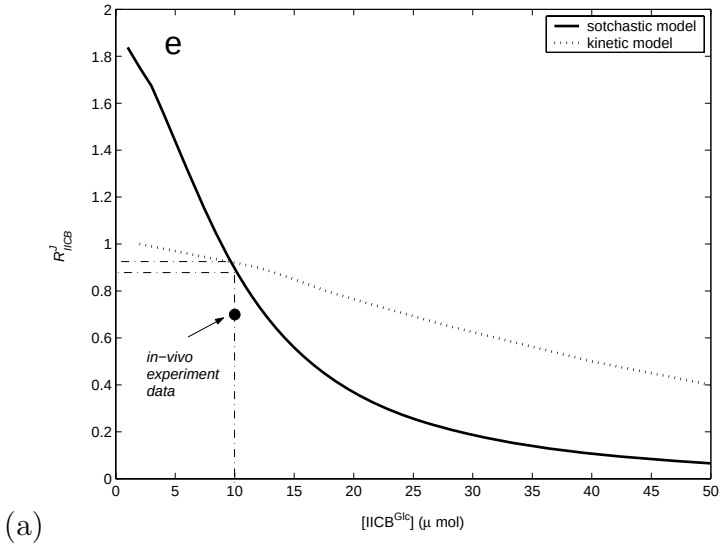
(b)

Figure 3.6. (a) the solid line on the left-axis shows the flux and the dotted line on the right-axis is the FRC; (b) comparison between the glucose flux for current model, available kinetic rate-base approach and experimental data.

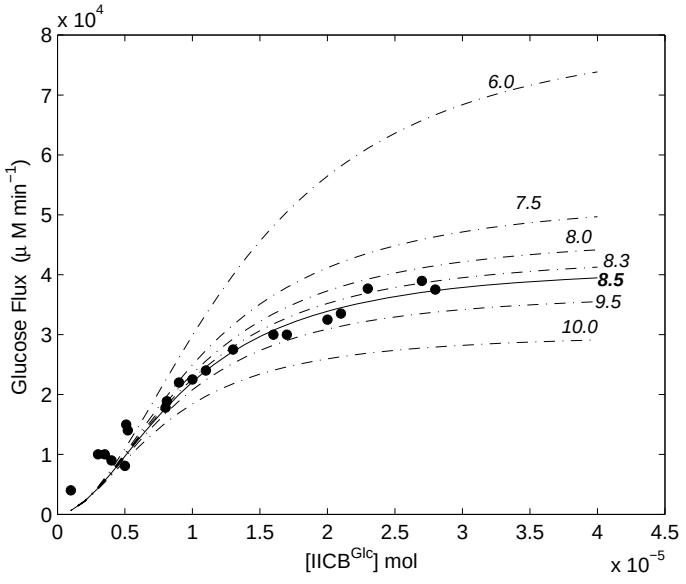
Fig. 3.6.(b) is a cross comparison between the stochastic model, kinetic rate-based model and the experimental data, as illustrated our model has well performed and captured the nature of the flux fairly better than the kinetic model. Also as its observed for higher receptor concentrations our model performs fairly better estimation of the experimental results than the kinetic rate-based approach.

Fig. 3.7.(a) compares R^J_{IICB} flux response coefficient (FRC) between the current and kinetic model. Although the differences between these models in very low concentration is significant; however the trends in both models follow the same behavior also at the experimental threshold both models are very close although the stochastic model performs slightly better. Two results for the models tend to become closer as the enzyme concentration further increases.

Training of the diffusion radius is shown in Fig. 3.7.(b) where the resultant diffusion radius for the *E.coli* is $8.5\mu m$; this value is very close to the bacterial diffusion layer $10\mu m \leq l \leq 100\mu m$ that is reported in [59].



(a)



(b)

Figure 3.7. (a) comparison of the R^J_{IICB} from kinetic and the stochastic model; (b) shows the training steps for finding the diffusion radius, the number above each line indicates the applied diffusion radius.

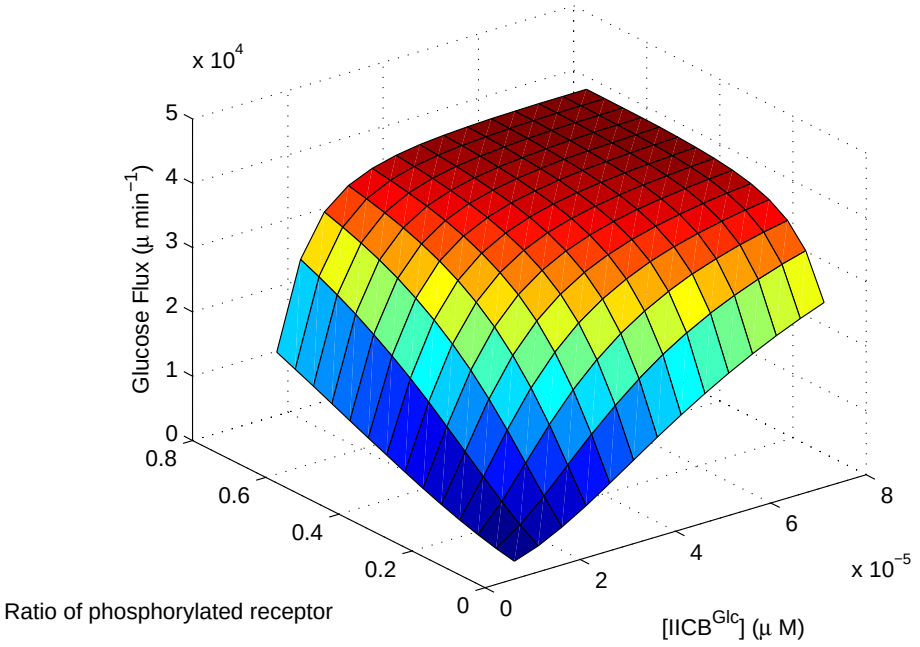


Figure 3.8. The effect of the model parameters on the flux of the metabolite to the cell.

In the last chart (Fig. 3.8) we have depicted the sensitivity of the of flux to cytoplasmic phase in various enzyme concentration, and as one can observe the sensitivity to the phosphorylation is more prominent at low concentration. Also after some threshold for concentration the phospho-factor looses its influence on the flux. For instance for $[C_e] = 10\mu mol$ this threshold is around 0.6 which is in agreement with the equivalent empirical reported values in [68, 67].

3.6 Summary

In this chapter, first we briefly introduced the process of ligand-receptor binding and outlined some of its contribution to cellular process in both eukaryotes and prokaryotes. Two abstraction variants for eukaryotic and prokaryotic cells were introduced which

founded the base of our modeling effort. Then the process of ligand docking to receptor was abstracted as two consecutive microevents along with the boundary condition function. For each micro-event we proposed a stochastic model to capture the temporal behavior of that microevent. The differences for application of each micro-event in eukaryotes and prokaryotes was pointed out where it was applicable. Also we used the proposed model to estimate temporal effect of the signal transduction and metabolite flux into the cytoplasm. Finally we used the *E.coli* phosphoenolpyruvate transferase pathway as a case study to validate our results for the prokaryotic case. The validation for the eukaryotic variant of the model was postponed to chapter 7.

CHAPTER 4

STOCHASTIC MODELS FOR PASSIVE DNA ACCESS AND THERMAL FLUCTUATIONS OF DNA IN EUKARYOTES

4.1 Introduction

DNA access is the key to cell protein synthesis machinery both in prokaryotes and eukaryotes. The long DNA chain of the eukaryotes use a systematic hierarchical compression. In the lowest compaction level the genetic material comprises arrays of coiled DNA around globular octamer of cationic nucleus proteins (histone) [58]. Each of these array elements is referred to as nucleosome and the chain of the ~ 1.65 left-handed superhelical turn is known as chromatin [69] [70].

A nucleosome component consists of ~ 200 base-pairs (~ 147 nucleosomal-DNA base-pairs), ~ 153 linker-DNA base-pairs) in euchromatin which is the most permissible chromatin conformation in gene expression [58]. Higher order DNA structures are also available in the form of 30nm chromatin fiber [71] and heterochromatin [72], all of which encapsulate the multi-billion base pairs of higher organism in the nano-metric volume of nucleus. Extensive research in the past thirty years on the chromatin structure and dynamics has revolutionized our knowledge on chromatin and its dynamics, yet much of the actual mechanistic is still not completely understood. Although many qualitative models have been proposed for the chromatin remodeling where the *in-vivo* and *in-vitro* assays have proven their accuracy with great deal of confidence, yet they are speculative in details of the process.

In this domain the experimental data have played an important role in the success of biologists and biochemists in one or more of following aspects: (a) explaining the nature

of the nucleosome related bio-processes, (b) proposing rational speculations where no explicit indication of actual mechanism could be observed, or (c) providing some useful empirical data in different domains (time, concentration , gene expression level, etc.) that would help further discoveries and validations. Having these crucial data on hand, the collaborative effort of bio-scientist along with physicist and mathematician have led to the development of growing number models that could address some of processes that govern the nucleosome dynamics.

In this chapter we use the same stochastic event based modeling approach as outlined in the earlier chapters to develop parametric understanding of nucleosome dynamics specifically the passive mechanism of accessing DNA sites.

The dynamics of nucleosome establishes certain pathways for accessing the hindered portions of nucleosomal DNA (nDNA). From the energetic point of view, a mechanism that requires energy for its progression is called an *'active mechanism'* and if it is spontaneous, is referred to as a *'passive mechanism'*. For the latter which is the focus of this chapter we incorporate spontaneous *unwrapping and rewrapping* and also *nucleosome sliding* mechanism. We model two variants of the nucleosome sliding mechanism: twist--defect [73] and planar-bulge inchworm [74]. In this work, we first find the probabilities of finding a motif of length n across the genome. Second, we propose a stochastic model for *unwrapping and rewrapping* and come up with a closed form solution. Thirdly, for twist-defect and planar-bulge inchworm mechanisms we propose two stochastic models, where the first model is sequence specific.

4.2 Nucleosome dynamics

Nucleosome is not a monolithic static assembly. In literature the dynamics of nucleosome is segregated into three categories; however there is a belief that conformational fluctuation is the forth category. Therefore, a nucleosome exhibits at least four dynamics:

1. compositional alternation
2. covalent modification
3. translational repositioning
4. conformational fluctuation

Compositional alternation is done by some remodeling enzymes to promote gene activation [75]. For example, although canonical nucleosomes are deposited during cell replication, H2A variant H2A.Z are highly enriched in promoter area which is deposited by SWR1 chromatin-remodeling complex [76]. Post translational modifications including acetylation, methylation, phosphorylation and ubiquitination [77] are among the covalent modifications that can destabilize the histone cores and exploit DNA access to the biological processes. ATP-dependant remodelers use energy derived form ATP hydrolysis to loosen the contacts between the coiled DNA and the histone core. All these remodelers have at least one ATPase domain that supplies the energy necessary to alter the nucleosome conformation. For instance, it takes less than one second for SWI/SNF to remodel a nucleosome[1]. A nucleosome remodeling mechanism is the hallmark of a chromatin remodeling complex. The remodeling protein complexes classified into different families and each family share many similarities in the mechanisms they implement in remodeling the core component.

The variants of chromatin remodeling mechanism (excluding covalent modifications) that is implemented by the remodeling complexes include: (1) creation of translational repositioning of histone octamer, (2)ejection and spacing of nucleosomes, (3) histone octamer transfer, (4) creation of remodeled di-nucleosome species (altosome formation), and (5) altered restriction enzyme access [78]. In translational repositioning the base-pair position of the core particle in the genome is altered to enhance the target site access. This process could happen both intrinsically or by the aid of remodelers.

[1]The mechanisms implemented by this remodeler is presented and modeled in chapter 5

Conformational fluctuation is a minor change to the conformation of a canonical nucleosome that happens periodically at the room temperature. This alternation could make an otherwise occluded DNA site temporarily available to the transcription factors. The mechanism underlying this thermal fluctuation will be discussed in proceeding sections.

Not all the times remodeling is necessary to access nucleosomal DNA. *In-vitro* assays in physiological salt concentration has strongly suggested nucleosomes are in an equilibrium fluctuation between the fully wrapped and semi-unwrapped states *in-vivo*, unless otherwise enforced [70]. The interval between two wrapped states gives a chance to the regulatory factor in the close vicinity of DNA to bind to the motif that would otherwise be inaccessible. Another spontaneous process of this category is the nondissociative nucleosome mobilization where histone core fluctuate between neighboring sites [74]. All these fluctuations are driven by thermal forces in physiologically relevant conditions [79, 80].

4.3 Event based pathway for target site access

Coiled structure of DNA in nucleosome confines the accessibility of the portion of DNA where the minor groove of helical turn faces the histone core. In euchromatin, passive and active varmints of mechanism for chromatin remodeling that makes a buried DNA site accessible, could be classified into three nucleosome conformation alternator categories: (1) Thermal spontaneous conformational fluctuations [70, 74], (2) ATP-dependent chromatin remodeling [81], and (3) covalent nucleosome modifications [78]. Each of these categories themselves might have different classes and implement distinct mechanisms. Furthermore they could work individually or in a concerted fashion. In the latter, they expedite each others' functions, hence allowing multiple options to be considered during the evolution of a specific mechanism for an individual motif access. The

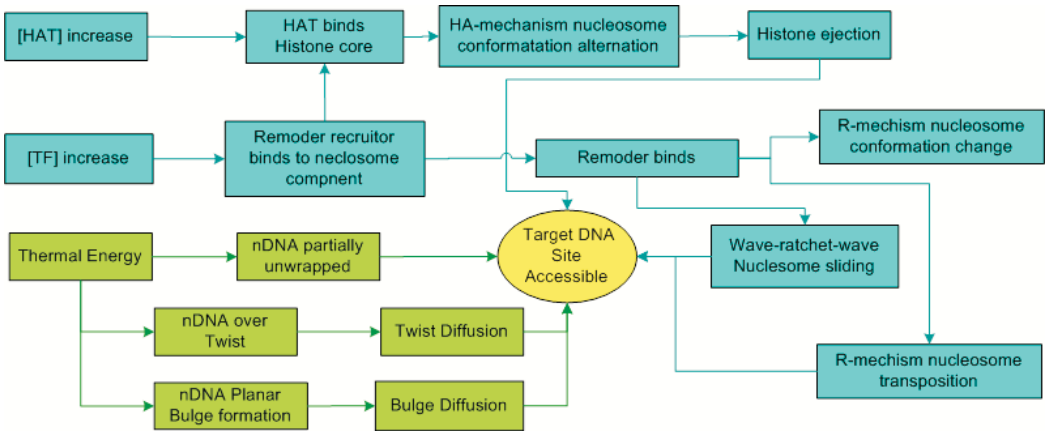


Figure 4.1. nDNA access event pathway; each pathway starts with one trigger event; *HA-mechanism nucleosome alternation* block represents the mechanism implemented by histone acetylation; *R-mechanism nucleosome conformation alternation* and *R-mechanism nucleosome transposition* depicts two classes of remodeling mechanism that derive the remodeling process.

eventual goal is to have a properly structured template with all compulsive transcription initiation complexes, available in a timely manner.

4.3.1 Event pathway model

In eukaryotes assembly of transcription apparatus and associated basal transcription factors comprises the access of promoter and several other functional target sites. Every access pathway is distinguished based on the mechanism that the pathway employs. Each of these pathways could further be broken into several protein complex specific pathways. Fig.4.1 depicts a set of plausible mechanism level motif access event pathways, only part this set is the focus of our investigation. In this chapter we only concentrate on passive pathways that is colored in light green in Fig.4.1.

Each block in a pathway is an event with a stochastic holding time. The collaborative effect of these stochastic events in a concerted fashion will led the target sites to be come accessible to the transcription factors. Therefore, we abstract each pathway as

a series of uncorrelated random events in the time and space and estimate a probability distribution for them.

4.4 Assumptions and stochastic components

In order to manipulate the temporal behavior of the pathways in Fig.4.1, we need to have execution time of individual components (event) and the respective probability distribution of each event. For that purpose we need to make certain rational assumptions without losing the generality. We also assume the properties of the set of abstract bodies that can represent physical topology of the individual participant of the remodeling process. The abstract bodies that participate in the modeling process are as follows:

4.4.1 DNA

The DNA model that we are using is the rod-like model from [82]. Such a DNA model is formed from beads that are inter-connected by harmonic spring in a spatially harmonic fashion. This model defines three dynamics for the DNA: longitudinal, rotational and bending. These dynamics are usually used in the mechanical models.

4.4.2 Nucleosome

Nucleosome consists of two components, the histone core and nDNA. Throughout this chapter, histone core is assumed to be a rigid cylinder with no structural alternation. The nDNA is attached to the histone core at 7 absorption points per super-helical turn, starting from the dyad [83] axis down to the end of core cylinder. These contact points are perpendicular to the super helix axis of nucleosome. In [84] the nucleosomal DNA is mapped into Frenkel-Kontorova chain and the total energy of the canonical nucleosome in equilibrium is manipulated based on three energy components: elasticity energy (E_l),

docking point energy (E_d), and sequence energy (E_s). We follow the same energy concept for manipulating the energy barriers in our model, where applicable.

4.4.3 Proteins

In biological functions, proteins appear in two forms: (a) single protein which is one polypeptide of amino acids molecule within its proper tertiary structure ,or (b) protein complex that is made of number of protein molecules in their quaternary structure prior to binding to the target site. RNAP II holoenzyme and SWI/SNF remodeling complex are the typical examples of a protein complex. In either case we abstract the molecules with a spherical body. For the former case we approximate the diameter of the sphere by the average width of the protein, where for latter this is done by averaging over width of the complex.

4.4.4 Chromatin wide collision

For any binding to happen first the has to be a collision between the parties, for our work we follow the collision theory approach by considering an infinitesimally small time Δt between successive collisions. We assume that the continuous nucleosome chain inside the nucleus is not repositioning. Also since the nucleosome array are available in the form of continuous chain, therefore their population density is non-uniformly distributed and is directly proportional to the density of the chain . For this purpose, we divide the nucleus into s equal volume partitions $\nu_s = \nu/s$ where each partition will have $\rho_i = N_i/\nu_i$ nucleosome density, N_i being the number of nucleosomes in partition i and ν_i as the respective partition volume. We further assume that the second colliding party, is uniformly distributed inside the nucleus. Hence, one can find probability of collision

between a protein (e.g transcription factor) and any nucleosome component in partition i during Δt , by applying the same approach as discussed in [85] as follows:

$$p_{col}^{i} = \pi(r_p + 1/2(r_{nuc} + r_\ell))^2 \sqrt{\pi 8 k_B T} \Delta t [pr] m_p^{-1/2} \nu_s \tag{4.1}$$

In Eqn.(4.1), r_{nuc} and r_ℓ are the nucleosome and linker DNA (lDNA) radiuses, r_p and m_p are protein radius and mass, k_B is the Botzmann constant, T is kelvin temperature, and [pr] is the protein concentration.

In the partition borders, the structure of chromatin chain enforces a close correlation among the nucleosome population distribution in the neighboring partitions. Hence, choosing the distribution form either one of the neighboring partitions would not significantly effect the final results. Considering Eqn.(4.1), in order to get the collision probability anywhere in the nucleus we can use, $p_{col}^{\nu} = 1/s \sum_i \rho_i p_{col}^{i}$. Also, the probability to collide with a specific nucleosome is:

$$p_{col}^{N} = \frac{p_{col}^{\nu} v_{nuc}}{\nu} = \frac{p_{col}^{\nu} r_{nuc}^2 L_{nuc} \lceil G/200 \rceil}{1/2(r_{nuc} + r_\ell)^2 (l_{nuc} + l_\ell)} \tag{4.2}$$

In the above expression, l_{nuc} is the nucleosome length where l_ℓ is the lDNA length and G is the base-pair length of genome.

Each complete helical turn (360°) consists of ∼10.2 base-pairs [74] in the nucleosome ground state, where for our calculation we round this number to 10 base-pairs per helical turn. The outer face of each symmetric half of the helical turn forms 6 DNA histone core association points, which along with an additional contacts between the N terminal of $\alpha N - helix$ and the N termini of histone H3 tail is referred as Super-Helical- Locations (SHL) [74],(see Fig.4.2). Therefore, in each nucleosome we have 14 association points. This fashion in DNA packaging inhibits the access to those base-pairs that face the

histone octamer where the minor groove of DNA contacts the core particle, at SHLs, this structural conformation is further elucidated in [83]. Having these in mind, we derive few useful probabilities. Probability that a motif i ($i \leqslant 10$) base-pairs lays within one helical turn ,p_h, is:

$$p_h = \frac{14(10 - i + 1) \times L}{(L - 147 + 1)(147 - \ell)} \tag{4.3}$$

In the above expression, L is the length of DNA in base-pairs and ℓ is the average linker DNA length. In each helical turn supercoiled around the canonical nucleosome only half of the base pairs are accessible; therefore, the probability that a motif of length i, ($i \leqslant 5$) to be directly accessible could be written as:

$$p_m^i = p_h \frac{5 - i + 1}{10 - i + 1} \tag{4.4}$$

Seemingly the probability of having the motif of length i, ($i \leq 147$) in one nucleosome, p_n^i will be: $p_n^i = \frac{L(147-i+1)}{(147-\ell)(L-i+1)}$ and if $L \gg i$ then,

$$p_n^i \approx \frac{147 - i + 1}{147 - \ell} \tag{4.5}$$

To continue with modeling, we make another rational assumption for the structure of the suppercoiled DNA around the histone octamer: DNA helical turn on entry and exit points of nucleosome has 180° phase difference. On the other hand, if we assume that on entry point DNA enters with major groove and leaves the nucleosome with minor groove, therefor the leading flanking tail of the nDNA is always accessible where the lagging one always remains trapped. Also we assume all of the motifs that we are working with in this chapter would not span over more than one super-helical turn. Hence, this avoids trapping the motif nucleosome in a DNA knot [86]. This assumption enforces a maximum motif length $i_{max} = 147/2 \approx 73$. Another useful probability isthe probability of accessing

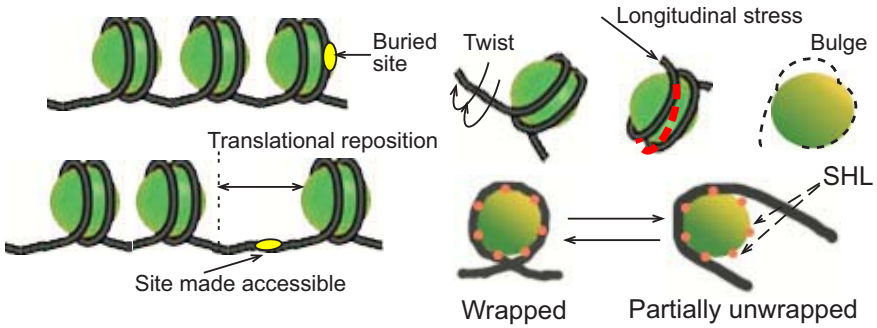


Figure 4.2. Left side: The upper image shows the site buried in the nucleosome,and the lower one shows site access made possible through translational repositioning; Right side: The upper shows the dynamics that might lead to translational repositioning of nucleosome, and the lower image plots the partially uncoiled nDNA that made the DNA site accessible.

a motif of length i on the linker DNA. For the average length of $\ell \approx 53$ base-pairs per linker DNA and the accessible length of $g = 3.5$ base-pairs for lagging flanking nDNA, the maximum length of motif on the linker DNA that could be accessed without displacing the nucleosome is:

$$i_\ell = \ell + 5 + g \approx 62 \tag{4.6}$$

Furthermore, the probability of finding a motif of length $i \leq 63$ in the linker DNA could be written as:

$$p_{\ell m} = \frac{(\ell_j + 9 - i + 1) \times (N - 1)}{L - i + 1} = \frac{\ell_j + 10 - i}{147 - \ell} \tag{4.7}$$

in Eqn. 4.7, ℓ_j is the lDNA between nucleosome j and $j + 1$.

4.5 Spontaneous remodeling mechanisms for DNA access

The spontaneous mechanisms that alternate the nucleosome are slow phenomena driven by the thermal molecular energy that happen in the room temperature. In this chapter, we try to use the discrete event simulation approach to model three widely

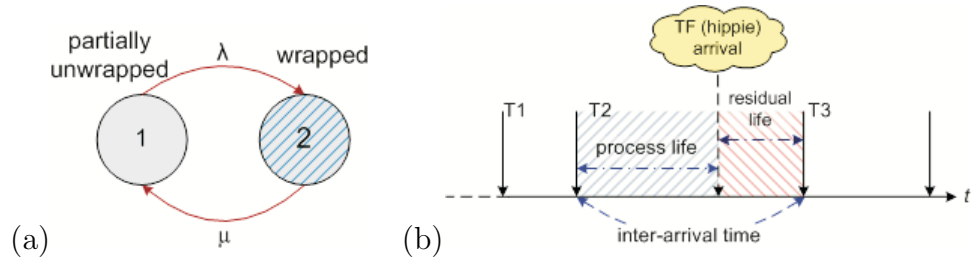


Figure 4.3. (a) Projection of the partial unwrapping to a *renewal process*; (b) *Process life/residual life concept*: process life lays between the two states arrival time, where the residual life is the interval between present time and the arrival moment of next state.

accepted mechanisms of this kind including: (1) partial unwrapping/rewrapping of nucleosomal DNA, (2) twist-defect nucleosome shifting, and (3) planar- bulge inchworm sliding of nucleosome.

4.5.1 Partial Unwrapping/Rewrapping of nucleosomal DNA

The propensity of nucleosomal DNA (nDNA) to partially unwrap on either end, can give temporary access to a DNA sequence whose access was originally blocked in the canonical conformation. This scaffold is mediated by the thermal force in a periodic manner [70, 81].

As reported in [70], nDNA remains fully wrapped for $\sim$250 ms then spontaneously becomes partially unwrapped, it remains in that state for $\sim 10-50$ ms and then rewraps again. In the latter state, if any of the protein machinery transcription factors find their target sites on the DNA , which was originaly hindered, binding becomes highly probable.

For instance if RNAP II reaches the nucleosome before it becomes fully wrapped the sheer size of polymerase would avoid further rewrapping and the process of RNAP II elongation would continue more rapidly. This rate could even get close to the linker DNA rate of elongation which is ~ 23 base-pairs s^{-1}, which otherwise would be $\sim$ 6-7 base-pairs, [70]. The outlined scaffold manifests a behavior of renewal processes [87],

where the active service time of the process is the period in which process reside in state 2 of Fig.4.3. We assume this service time follows a negative exponential distribution, $p(\tau_1) = e^{-\mu\tau_1}$. Therefore, based on the data in [70] one can say: $\mu \in [20, 100]$ and $\lambda \approx 4$. Considering a general distribution for the arrival rate of the binding species, in order to receive service (binding) it must arrive meanwhile the process is in state 2 and $\tau_2 > 0$, where τ_2 is the residual life of state 2. Using the *M/G/1* queue service model and *hippie arrival* concept as in [87], we find

$$p(\tau_2) = \frac{\mu + e^{-\mu\tau_2} - 1}{\mu^2} \tag{4.8}$$

however, we need to have $\tau_2 > 0$, thus:

$$p(\tau_2 > 0) = 1 - p(\tau_2 = 0) = \mu^{-2}(\mu^2 - \mu + 1) \tag{4.9}$$

From Eqn.(4.9) and $p_2 = \mu/(\lambda+\mu)$, the probability of state 2, we can find the probability for species catching the unwrapped nDNA, p_{uw}, in time τ from:

$$p_{uw} = \frac{\mu^2 - \mu + 1}{\mu(\lambda + \mu)} \tag{4.10}$$

As we see the p_{uw} has an stationary probability because the two integrations, one in calculating the laplace transform of residual life, and other the implicit integration in Eqn.(4.9) have averaged over the time.

4.5.2 Nucleosome sliding through Twist-Defect

Imposing an anomaly into the super-helix conformation of nucleosomal DNA is an alternative mechanism to translocate the nucleosome. The concerted translational and rotational motion of DNA that has high torsional flexibility, leads to injection of this

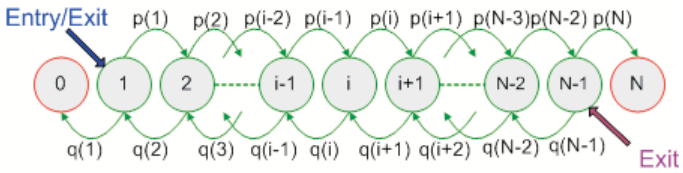


Figure 4.4. Markovian random-walk chain with absorbing states: The absorbing states guaranties two conditions: (a) there will be no re-enter for each walker, and (b) access to the two exit points are mutually exclusive to every walker.

anomaly to the nDNA. Since this phenomena is impelled by thermal energy at the room temperature, it is only limited to the very efficient torsions. Considering the DNA with 10 base-pairs per helical turn would result $\sim 36^\circ/pb$ for nDNA in ground state. Therefore, twisting or unwinding the DNA double helix by 36° would add or remove one base-pair to or from the 360° helical turn which is docked to the histone core between two SHLs [74]. The torsion introduced at the entry or exit point of a nucleosome has to successfully travel all 147 base-pairs around the nucleosome to result a once pase-pair shift for the nucleosome. Also, since the manipulations for over-twist and unwound are fairly similar we limit our discussion to the former case.

The fluctuation of the twist between each pair of docking points forms a mono-dimensional random-walk (RW). To model this random-walk, we define 13 states that resemble the location of the twist between each pair of 14 docking points at arbitrary time t. Our approach is different from the one proposed in [84] in following aspects: (i) We have used an-isotropic RW where they have isotropic RW (ii) our Markovian step process has two absorbing state at both ends but they did not include any absorbtion (iii) our model is fully stochastic where as they started with a stochastic process and then switched to deterministic domain (iv) we end up with a stochastic process rather than diffusion constant.

In order to move in either direction, twist will require to overcome docking energy (E_d), elastic (E_l) energy, and sequence energy (E_s) barriers. The first two energy barriers are already included in the defect potential energy cost expression that is provided in [84] which is based on the Peierls-Nabarro (PN) potential energy concept [88]. By assembling all parts of twist PN equation and integrating the expression given in [88] the average energy of a twist could be written as:

$$\Delta U_{twist} = \frac{1}{6}C + \frac{9}{10}U_0 \tag{4.11}$$

Where in this expression $C \approx (84 - 120)k_bT$ is the effective *Frenkel-Kontorova* (FK) combined twist and stretch constant, and $U_0 \approx 6k_bT$ is the SHL docking energy. In ΔU_{twist} the an-isotropic bendability of DNA is not included. This propensity of DNA is sequence dependant and the respective energy cost could be calculated from the following equation which is obtained from [84] after proper substitutions:

$$E_s(k) = \sum_{i=1}^{10} -\alpha^s_{i+10(k-1)} cos(i\pi/5) + \beta^s_{i+10(k-1)} \tag{4.12}$$

Eqn.(4.12) defines a 10 base-pairs periodic energy field around the histone and linearly assigns two bending energy charges to each dinucleotide, isotropic (β^s_i) and an-isotropic (α^s_i). Here $1 \leqslant k \leqslant 13$ represents the state of defect in Fig.4.4, (states 0 and 14 are the absorbing states). The two approximate bending energy charges per dinucleotide variants is reflected in Table 4.1 where of energy is in k_bT unit.

The defect follows a one dimensional random-walk on the nucleosome to exit from either end. If the defect that is inserted at one end (entry point) leaves the nucleosome from the other end (exit point) , this would result a one base-pair sliding of the nucle-

Table 4.1. Dinucleotide Bending Charges

Dinucleotide	$\alpha\ k_bT$	$\beta\ k_bT$
A/T	-43	-1.5
G/C	48	10
other	0	3.5

osome. We use *Mean First Passage time (MFPT)* to get the defect leaving time from entry point (t_{en}) or exit point (t_{ex}).

Using the splitting probability concept given in [21] we can find probability of reaching one site before being absorbed by the other. Therefore, we will have the probability to exit from entry point (p_{en}), and probability to exit from the exit point (p_{ex}) as follows:

$$p_{ex} = \left(1 + \sum_{i=1}^{13} u(i)\right)^{-1} \quad \text{where} \quad u_j(i) = \left\{ \begin{array}{c} \mu_i u(i-1), \quad \text{for } i > j \\ \mu_i, \quad \text{i=j} \end{array} \right\}$$

$$p_{en} = \frac{1 + \sum_{i=2}^{13} v(i)}{1 + \sum_{i=1}^{13} v(i)} \quad \text{where} \quad v_j(i) = \left\{ \begin{array}{c} \mu_i^{-1} v(i+1), \quad \text{for } i < j \\ \mu_i, \quad \text{i=j} \end{array} \right\}$$

In Eqns.(4.13,4.13) $\mu_i = p_i/q_i$, where p_i and q_i are probabilities of the defect to move one step to the right or to the left of state i in the unit of time, respectively. Considering the energy barrier, E_s, we can define p_i and q_i as:

$$p_i = \frac{e^{-E_s(i+1)\omega_i}}{e^{-E_s(i-1)\omega_i} + e^{-E_s(i+1)\omega_i}} \qquad q_i = \frac{e^{-E_s(i-1)\omega_i}}{e^{-E_s(i-1)\omega_i} + e^{-E_s(i+1)\omega_i}} \tag{4.13}$$

$\omega_i = (|E_s(i+1)| + |E_s(i-1)|)^{-1}$ is the projection coefficient that project the energy barrier to $[-1, 1]$, thus, minimizes the effect of $exp(\cdot)$ function on the fate of the probability.

Now we apply the concept of Mean First Passage Time (MFPT) for one dimensional random-walk on a random lattice form [89]. By applying this approach, the pair of exit

times, t_{en} and t_{ex} which denote the MFPT from exit and entry points, respectively, could be expressed as follows:

$$t_{ex} = \Sigma_{k=1}^{N-1} p_k^{-1} + \Sigma_{k=1}^{N-2} \Sigma_{i=k+1}^{N-1} \Pi_{j=1}^{i} q_i p_j^{-1} \tau_s \tag{4.14}$$

$$t_{en} = \Sigma_{k=1}^{N-1} p_k^{-1} \Pi_{i=1}^{k} \mu_i \tau_s \tag{4.15}$$

Here τ_s is one RW step. Also since

$$p_i^s = \prod_{k=0}^{i-1} \frac{q_k}{p_{k+1}} p_1^s$$

is the stationary probability of non-absorbing state i, therefore:

$$p_0^s = q_1 \left(\sum_{i=1}^{N-1} \left(\prod_{k=1}^{i-1} \frac{q_k}{p_{k+1}} \right) + q_1 + \prod_{k=1}^{N-2} \frac{q_k}{p_{k+1}} p_{N-1} \right)^{-1} \tag{4.16}$$

Arrival rate of the defect, λ_d is confined to: (a) our original assumption that at most one defect could exist in the nucleosome at any moment which is satisfied by p_0 and (b) the energy dependant probability of a twist formation which is given by Boltzmann factor, $p(\Delta U) = e^{-\Delta U / k_b T}$. Another implicit indication of (a) is $\lambda_d < \mu_d$ where,

$$\mu_d = (p_{ex} t_{ex} + p_{en} t_{en})^{-1} \tag{4.17}$$

μ_d is the export rate of defect from either end of the nucleosome. Having μ_d, Eqns.(4.13,4.13) and considering (a) and (b), we can give the following upper bound for nucleosome sliding rate to the right ($5'$ to $3'$):

$$\lambda_d \lesssim (p_0 + p_N) \cdot p_{ex} \cdot \mu_d e^{-\Delta U_{twist} / k_b T} \tag{4.18}$$

So far we elucidated the forward sliding ($5^{'}$ to $3^{'}$ with respect to DNA) of the nucleosome, likewise approach could also be applied to the reverse sliding ($3^{'}$ to $5^{'}$ with respect to DNA).

Since we have considered an-isotropic random-walk and directly included the effect of sequence-depended energy in the random-walk, our model is more accurate than the one proposed in [84] where the authors ignored both of these facts. Also the diffusion constant was originally reported from their model was not consistent with the experimentally observed values. Hence, they decided to bring the sequence energy factor,E_s, through the modified Bessel function to exponentially reduce the diffusion constant. However, that still would not compensate for the effect of the absorbtion states that they did not include in their model. Switching to deterministic domain in the mean term of modeling process has induced the approximation overhead and consequently less accuracy to their model.

4.5.3 Planner-bulge Inchworm Nucleosome Sliding

Formation of bulge is a prominent feature of the nDNA that is cognate with the high DNA curvature. If during the final interphase of state 1 (partial unwrap) in section 4.5.1 a more distal sequence of DNA would be pulled in to the nucleosome and adsorbed to the entry SHL then a bugle would be formed. There are two types of bugle: planar and topological bulge. In the planar bulge the entire bulge falls in a same plane, while the topological bulge has a more complex structure where the DNA crosses over itself and creates a twist [83]. In the current work we just consider the small planar bugle since in the passive paradigm it is highly unlikely to have topological bulge. Small planar bulges are 10 or 20 base-pairs in length, where the rationale for having a multiply of 10 base-pairs in their length is to encompass an even number of 360° helical turns. Because a

different number base-pairs in their length would impose a phase shift to the nucleosomal DNA which energetically is very costly [74].

In this mechanism sliding of histone octamer is the result of sequential formation and annihilation of bulge along the nDNA. This resembles the motion of an inchworm creeper around the nucleosome. To model the fluctuation of the bulge we use a random-walk but with some distinctions from the one we applied earlier for twist defect. A major difference between is the isotropy of the steps in the current random-walk; therefore, in the step process diagram of Fig. 4.4 for all steps the probabilities $p_i = q_i = 1/2$ and E_s would not have any effect in fluctuation of the bulge. Adopting the energy barrier model from [90] and applying proper substitutions, the energy cost ΔU for a bulge with the base-pair length of l will be:

$$\Delta U_{bulge} = \left(20\pi^4 R_N^4 E_a^5 \sigma k_b T\right)^{1/6} \left(\frac{3.4 l \mathring{A}}{R_N}\right)^{1/3} \tag{4.19}$$

In Eqn.4.19 R_N is the nucleosome radius, E_a is the docking energy per unit of length, k_b is the Boltzmann constant, T is the absolute temperature, and σ is persistence length of DNA. Persistence length of DNA is the limit beyond which DNA will lose its physical properties as a pure elastic rod. DNA bend persistence length is $\sim$150 base-pairs (51 nm).

By applying the same random-walk algorithms, while $p_i = q_i = 1/2, \quad \forall i \epsilon [1, 13]$, we can find the t_{ex} and t_{en} from the same approach as for twist defect. Again the export rate of the bulge is $\mu_b = (p_{ex} t_{ex} + p_{en} t_{en})^{-1}$. For calculation of arrival rate of bulge λ_b is very similar to the one for twist: $\lambda_b \lesssim (p_0 + p_N) \cdot p_{ex} \cdot \mu_b e^{-\Delta U_{bulge}/k_b T} \times l$. The factor l on the left side of expression is the base-pair length of the bulge. The argument that we made earlier for direction of nucleosome sliding as a result of twist-defect random-walk remains valid for the bulge as well.

4.6 Results

In this section we briefly present the numerical results that was derived by applying the proposed models to study and predict the nucleosome translocations. For nucleosome sliding as the result twist-defect, we used two sequences: *PSEN1* base-pair position $\epsilon\ [-700, -200]$ (human gene responsible for producing precenilin-1 protein), and *AHI1* base-pair position $\epsilon\ [+181, +681]$ (gene responsible for Jouberin protein in human). Fig.4.5.(a) shows the variation of nucleosome sliding rate for different base-pair in the sequences. As we observe the PSEN1 sequence has larger variation domain. This biologically implies that the sequence has more an-isotropic segments (more populated with A/T and G/C di-nucleotides), hence predicting a lower sliding rate for the nucleosome. In both sequences the high picks could imply a rotational trap in the DNA where ATP-dependent remodeling might be required and intrinsic sliding possibly halts. Due to the granularity of the measurement, finding the experimentally confirmed values sliding rates is a challenging task. Although we could not find any explicit sliding rate for twist defect, we could infer some thresholds form assays in [79]; These thresholds are depicted by dashed lines. The sole theoretical model that we could find to the date of this reported, is a diffusion based deterministic model that reported two diffusion constant: $D = 580\ base - pairs/s^2 \rightarrow \lambda_d \approx 0.79\ base - pairs/s$ and $D = 10^{-6}\ base - pairs^2/s \rightarrow \lambda_d \approx 1.2 \times 10^{-8}\ base - pairs/s$ for isotropic and an-isotropic sequences, respectively. These diffusion constants are several fold skewed from experimentally reported range.

Fig.4.5.(b) depicts the time required for nucleosome to slide on same section of *PSEN1* in forward ($5^{'}$ to $3^{'}$) or reverse ($3^{'}$ to $5^{'}$) direction. The net relative displacement of nucleosome with respect to the reference base-pair at (-700) position is also depicted in the same figure (plotted in green color).

For the bulge inchworm model we got $\lambda_b = 2.14 \times 10^{-7}$ for $l = 10$ base-pairs and $\lambda_b = 4.56 \times 10^{-9}$ for a bulge length of 20 base-pairs. This sliding rate results indicate that within one hour epoch almost no sliding happened. We concluded that nucleosome sliding through bulge mechanism is a very rare event. We were not able to find any experimental data for the bulge model. We used time step , $\tau_s = 10^{-6}$ throughout our calculations.

4.7 Summary

In the current chapter we created a systematic view of the pathways which comprises the synergy of events that effect nucleosome dynamics and procures nDNA access to transcription machinery proteins. We further elucidated three of such pathways all of which were passive and tried to model by applying a stochastic discrete event based approach. We also used the collision theory to estimate probability of finding the target nucleosome component genome wide. We used the renewal process along with hippie arrival concept to model the spontaneous unwrapping pathway for DNA thermal fluctuations. For nucleosome sliding by twist-defect we used an-istorpic random-walk that takes the DNA sequence curvature energy barrier into the account to find the MFPT in state diagram of the process and ultimately the sliding rate of the nucleosome. Bulge-inchworm is another mechanism for intrinsic nucleosome sliding that we investigated in this chapter and proposed a stochastic model to capture its temporal dynamics. Also, we discussed that our proposed models are more realistic and closer to actual prices compared to the available diffusion constant based models. Finally we validated our models with available experimental data and made a predication on spontaneous translocation of nucleosome for one of the human genes.

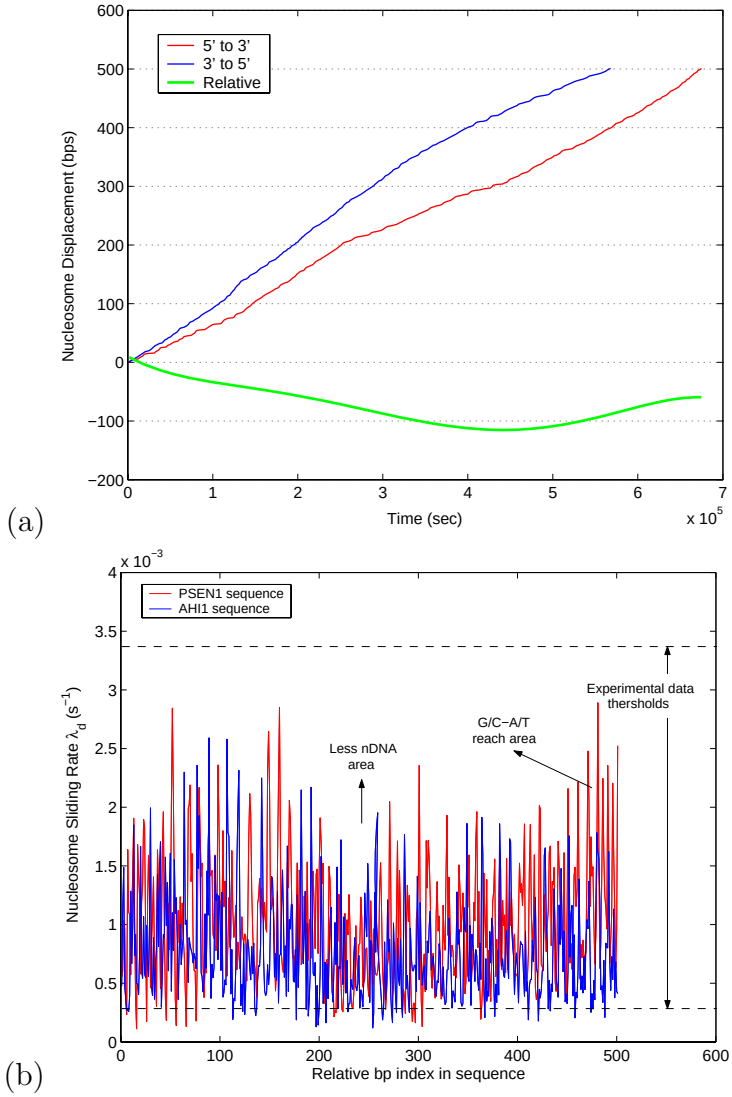


Figure 4.5. The twist-defect mechanism used for manipulations in both graph; (a) The red line depicts the forward displacement relative to position -700 in *PSEN1* gene versus time; The blue line shows the amount of reverse displacement relative to same reference position across time; Green line shows the net displacement of nucleosome, where a negative value indicates the intrinsic sliding of nucleosome tends to slide in reverse direction in this sequence. (b) shows the *forward* sliding rates of the nucleosome for *PSEN1_HUMAN* and *AHl1_HUMAN* sequences, the reported experimental thresholds are shown in dashed-lines.

CHAPTER 5

AN EVENT BASED STOCHASTIC MODELS FOR CHROMATIN REMODELING

5.1 Introduction

Coiled structure of DNA in nucleosome confines the access to the portions of nucleosome DNA, where the minor groove of the helical turn faces the octamer. Certain complexes impede the DNA access which would be temporally exhaustive and biologically irrelevant, should the process be driven by thermal energy, as we showed in previous chapter. These complexes and their underlaying mechanisms could work individually or in a concerted fashion, to establish a modified nucleosome structure that concedes the access to nucleosomal DNA, [78]. Generation of local chromatin topology conducive to gene transcription pertains to one following classes of chromatin modifications based on their mode of action: ATP-dependent chromatin remodeling and covalent histone modification. The former class, has been classified into five families (SWI/SNF, ISWI, INO80, NURD /Mi-2/CHD, SWR1) based on their protein complex subunits. Among these re-

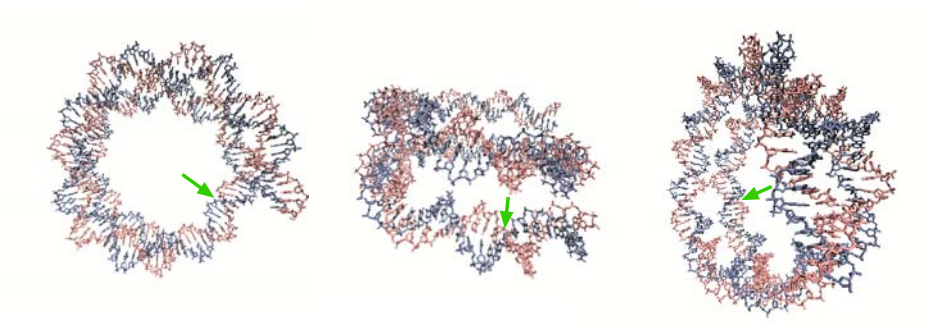


Figure 5.1. The green arrows show the typical base-pairs whose access is confined in canonical nucleosome structure.

modeling families SWI/SNF has been more intensively studied for two reasons: Firstly, because of the biochemical characterization of SWI/SNF with 8 to 11 subunits (varies among different species), [91], enables this remodeler to employ diverse mechanisms to alter the chromatin conformation, [92]. Secondly, SWI/SNF shares the same highly conserved ATPase domain (SWI2/SNF2) with remodelers of over thirty other organism, [1]. Also all eukaryotes contain all five families of chromatin remodelers, [93]. Furthermore no sequence specificity has been reported for SWI/SNF complex, [94]. These and many more evidences reported elsewhere strongly suggest that similar mechanisms might possibly be employed by many other species. The collection of above reasonings was enough convincing for us to choose this family of chromatin remodeler as the next target of our modeling effort.

Reported *in-vitro* data suggests that SIW/SNF perturbs the nucleosome positions in a phased array of nucleosome, [93]. The nucleosome position rearrangement is archived through two primary pathways: histone sliding *in-cis* and histone displacement *in-trans*. The former which is the main focus of this paper does not change the local nucleosome occupancy (LNO) of the DNA , [95]. The latter will reduce the LNO, hence more naked DNA would be available to the cell protein synthesis machinery complexes.

cis pathway is the most prevalent chromatin remodeling pathway and was first reported by [80] from the observation of the simian virus 40 minichromosome in the DNA fragment. Two variants of *cis* pathway are available to the remodelers: a memoryless pathway where the continuous hydrolysis of the ATP required for continuation of remodeling, and with-memory variant in which a dimer of neighboring nucleosome is formed and exposes portions of the nucleosomal DNA of intermediate structure which referred to as altosomes, [96]. The altosome would still keep its conformation for some time even in absence of ATP hydrolysis. In the current study we will only consider the memoryless pathway.

Having a comprehensive picture of the chromatin, its metabolism and modification is essential for comprehending of many other cell level process in eukaryotes. Also, since the chromatin is not a monolithic ensemble *per se* individual models are required to capture its dynamics. These models should be parametric and integrative such that ultimately a systematic view of the whole process that can represent different states of the ensemble could be provided. In the previous chapter we devised individual models for passive variants of chromatin remodeling mechanism. To the date of this dissertation no single parametric model is propped for ATP-dependant chromatin that is in the context of the cell and has a systematic view of the process.

Kulic and Schissel [86] did a detail study on the nucleosome loop formation. They proposed an estimation for nucleosome loop energy; however they did not a provide a concise and biologically relevant formalism for the remodeling process in their work. In [2] a deterministic model is proposed for nucleosome repositioning that ends with a diffusion constant. Although the model is physically complex; however they did not incorporated any of physicochemical properties of the remodeler and it's subunits. The bulge step-size was ignored in nucleosome translocation, and most important they did not bring the stochastic nature of the process in the picture. Their reported data is several time skewed according to the experimental data in [1, 97], and the error in results aggravates as the loop size grows larger. Wang *et. al* [98] provided a Brownian dynamics simulation of the DNA loop traversal around the nucleosome. To describe the system they used Langevian equation with Morse potential and random force. They applied Runge-Kuta algorithm to numerically solve the equation and determine the system states in the course of simulation. There two major drawbacks to their model: Firstly their simulation is purely a non-parametric physical model simulation, therefore is not possible to couple their simulation with a systematic modeling approach of a biological process. Secondly, their model is computationally intensive. In summary, the common problem of the above

approaches is that they are not biologically very significant and could not be embedded in systematic modeling project.

In this chapter we will devise a *de-novo* parametric model that can capture the dynamics of the nucleosome remodeling through the memoryless version of SWI/SNF *in-cis* pathway. The proposed model takes into account both the sequence curvature of the DNA and ATPase domain of the remodeling complex in the nucleosome remodeling process. With the loop formation any transcription factor would have two opportunities of binding to the otherwise occluded DNA site:(i) catch the loop, i.e. the excess length of DNA between two histone docking points or sits on tail, or (ii) wait until the target site leaves the nucleosome by as the consequence of histone core translocation.

5.2 Approach

Straining a super helical torsion on the nucleosome DNA is the hallmark of swi2/snf2, a member of SF2 superfamily of helicase like proteins , [99]. Experimental assays use number of diverse methods to detect the accessible DNA loops and histone translational repositioning resulted from catalyzed remodeling. Most of these methods benefit from one or both following properties: nucleosome DNA cleavage by micrococcal endonuclease of the remodeled nucleosome that would otherwise be buried in the nucleosome, and electrophoretic mobility shift-assays that prevail the discrepancies between the canonical and remodeled mobility patterns, [97, 100]. In a more recent *in-vitro* assay, Bustamanate *et al.*[1] used optical tweezers to measure the real time response of the nucleosome and loop formation under stress. Based on *in-vivo* and *in-vitro* observations and biological principles, authors in [93] proposed a qualitative model that describes the mechanism of DNA translocation around the nucleosome by the aid of remodeler. This model which is referred to as *'wave-rachet-wave'*, forms the physiological base of our modeling effort and is depicted in Fig5.2.

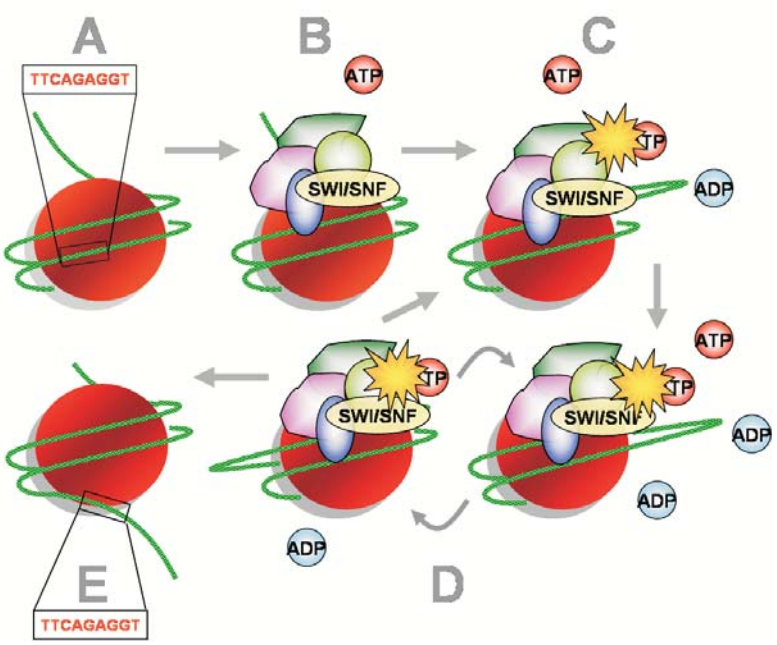


Figure 5.2. The physiology of the nucleosome translocation *in-cis* is abstracted in four states:(A) a canonical nucleosome which has limited the access to a motif, (B) SWI/SNF remodeling complex bound to the nucleosome component close to the dyad axis, (C) remodeling complex uses the ATP hydrolysis to pull extra base-pairs into the nucleosome from linker DNA (bulge formation), (D) the tracking sub-domain of remodeling complex tracks the bulge around the nucleosome until it exists from the distal entry point. At the end of tracking the nucleosome is repositioned equal to the bulge size. The repositioned nucleosome can undergo another remodeling to state (C) or the remodeling complex gets detached from the core component as shown in (E).

Traces of the random behavior in the dynamics of chromatin remodeling could be observed in the uncertainty within the reported data on quality of the remodeling process by SWI/SNF, ISWI and RSC family of remodelers in [100, 1, 97, 101, 102].

As a corollary to this randomness we characterize each step in 5.2 by one or more microevents that are stochastic in time and space and offer a probability distribution for them. The convolution of these stochastic events can provide an appropriate picture of the remodeling process dynamics.

The key to translocation of nucleosome in the outlined model is conversion of the ATP hydrolysis energy to the mechanical force that can alter the nucleosome conformation. Torsion and tracking sub-domains are the two sub-domains of remodelers translocase. The tracking sub-domain pulls the proximal linker DNA into the nucleosome, traps the extra base-pairs in the fissure between the two sub-domains. The ATP hydrolysis then would enforce a conformation change to the tracking sub-domain and drive the loop on the unidirectional path around core particle to the distal linker by taking the bases from the leading end and adding to lagging side. The locality of engagement of the SWI/SNF to dyad axis of the nucleosome would be a an important factor in the quality of the remodeling. We initially loosen the constrains suggested by [101] and [102] and allow the remodeling complex to be attached at any super helical location (SHL) in the vicinity to the dyad axis on the proximal linker DNA side (SLH_i, $2 < i < 8$). Also the fissure size between the sub-domains can vary in our model from entry point (SLH_1) of the nucleosome to $SLH_{(i-1)}$. These additional degrees of freedom is granted for two reasons: firstly because the definitive proof of these constrains is still missing in experimental observations elsewhere, secondly the energy-translocation plane for each scenario that will be provided by the model could provide further evidence for the efficiency of the structural preference. Since the evolutionary lineage of a bio-process favors efficiency, therefore the results could argue in-favor-of or against the available hypothesis.

5.3 Modeling the histone core translocation *in-cis*

As outlined earlier, the qualitative model for SWI/SNF remodeling suggests three stages for remodeling including: remodeler-Nucleosome binding, bulge formation (torsion strain), and intra-nucleosome bulge tracking. To abstract these stages we map them from the qualitative domain to probabilistic information domain by defining one or more microevents corresponding for each stage. The sequence of microevents that drive the

remodeling process is serialized, i.e. no two microevents would have temporal overlap. We define three class of microevents: (a) Binding, (b) Bulge Formation, (c) Tracking. The execution order of these micro-event classes is deterministic as depicted in Fig5.3.(b). However, the *holding time* of each microevent is a stochastic process govern by sub-model that is associated with each class of microevents. We define *Nucleosome Coordinate (NC)*, x, as the position of base-pair on the nucleosome where: $1 \leqslant x \leqslant 147$. Now we introduce state of the remodeling process with following parameters: The availability of remodeler on the nucleosome (b), leading nucleosome coordinate of the bulge (x_s), the lagging nucleosome coordinate of the bulge (x_e), number base-pairs pulled into the nucleosome (n). Therefore, at any time the state of the process could be presented by the (x_s, x_e, n, b), where at any time, except the last element which has boolean value the other three have non-negative integer values. The state diagram is a random graph as depicted in Fig. 5.3.(a). Transition between each pair of adjacent states is driven by a microevent that has a stochastic holding time. For each microevent the holding time is the execution time of that microevent. A proper sequence of such stochastic microevents forms one realization of *in-cis* translocation of the core particle. The canonical nucleosome structure that we incorporate along this study is assumed to have 147 base-pairs in the ground state. The transition form the state (0,0,0,0) to (0,0,0,1) is barely and protein-DNA binding event, we assume that once a remodeling complex associated with nucleosome, it commits to conduct at least one translocation step. Therefore, the binding event would merely characterize a signaling event to the process, though it would not affect its progress. Hence, we would not consider the time for binding in our probability distribution function manipulation. Our next assumption for the model is the tracking size across one iteration of the remodeling remains constant.

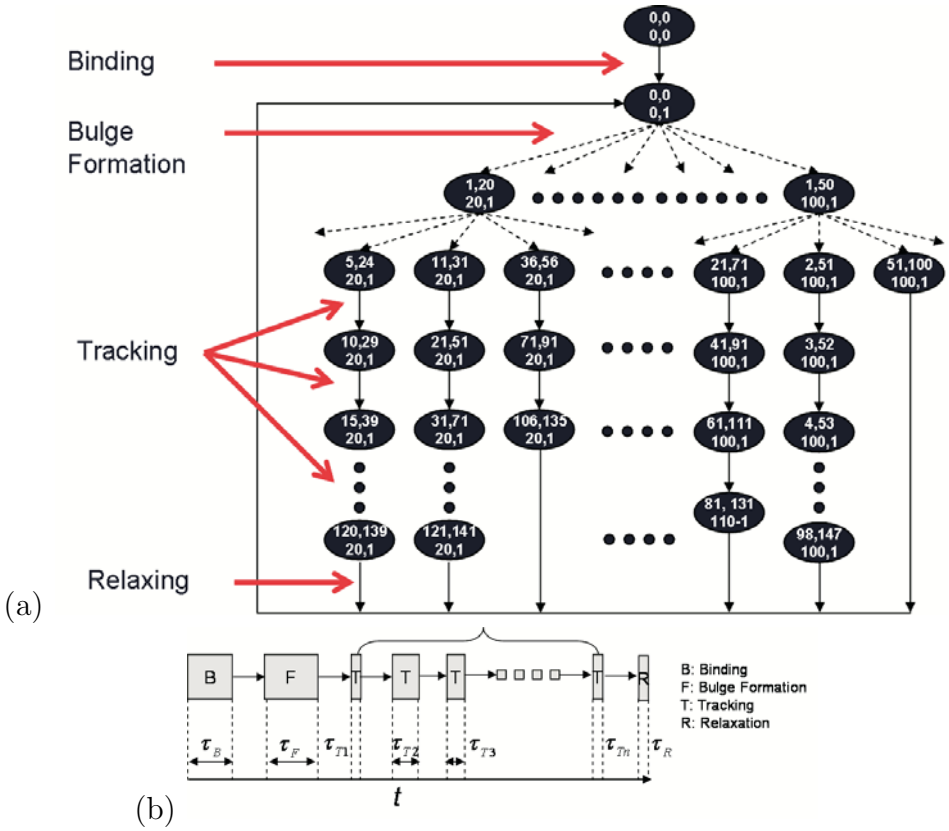


Figure 5.3. (a) State diagram of the remodeling process forms a random graph, edges are the microevents that moves the process from one state to the next, the state quadrable should be read from top-left in clockwise order; (b) Event diagram of the remodeling process, each microevent has a random holding time with mean and variance.

5.3.1 Bulge formation analysis

In order to provide a rigorous analysis of the bulge formation we need to look in to zoology of the bulge and their energy profile. Loops that are formed from the excess length of a rod like object are classified into three structural categories:

- Simple loop: no self contact point within the loop.
- Crossed loop: there is a single self contact point between within the loop.
- Entropic loop: where are is more than one self contact point within the loop.

These geometry of these loops is depicted in Fig. 5.4. In [86] researchers proposed approximations for the energy profiles of these loops, however biological feasibility of the

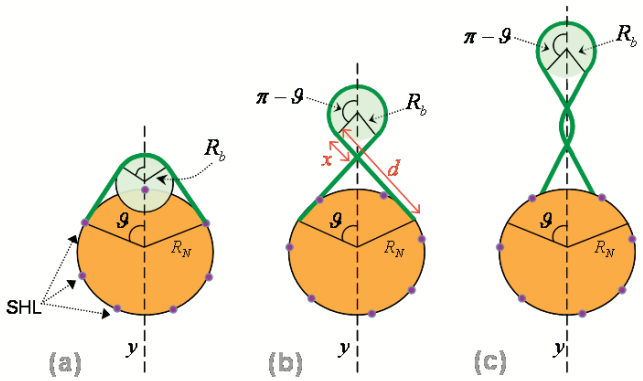


Figure 5.4. Three structural loop conformations on the nucleosome:(a) simple loop, (b) crossed loop, (c) entropic loop. Entropic loop conformation is not limited to the one shown in (c).

formation for each class of loops in the nucleosome context was not studied. One can argue that, since tracking mechanism of the bulge is archived by removing the base-pairs from lagging end and adding them to the leading end (the side with hight SHL), a remodelers tracking sub-domain task, should the bulge has an entropic structure this mechanism fails due to writhe induced on the DNA. Therefore, we make a fair assumption that a nucleosome bulge has either simple or crossed structure. To study the formation feasibility of such loops we use the nucleosome angle ϑ which is the min angle between the y axis and the tangent point of the DNA on the histone core. Using ϑ and expanding the geometry that is provided in [86] along with comparison to the experimental data we propose the boundary condition on the structure and length of the DNA loop.

Intrinsic curvature and bendability of stacks of nucleotides adds another complexity to the dynamics of energy profile of the loop formation. To reflect this complexity in the bulge formation energy we used the Eqn. 5.1 from [103] to manipulate the expected Young's modulus for the bulge of n base-pairs length.

$$E_b = \frac{\sum_1^{n-2} e_i^{-1}}{n-2} \tag{5.1}$$

In the above expression e_i is trinucleotide sequence stack Young's modulus based on the Table 1 in the [103]. Now we can find the bending rigidity of an arbitrary DNA sequence of length n, B_n from:

$$B_n = \frac{\pi}{16} E_b R_{DNA}^4 \times 10^{30} \tag{5.2}$$

Noteworthy, that in Eqn. 5.2 B_n is in $k_B T \cdot nm$, R_{DNA} in meter, and E_b in N/m^2. Then, we replace our terms in the simple bulge energy, U_S, and crossed bulge energy, U_X, approximation expressions adopted from [86] an rewrite them as follows:

$$U_S(\vartheta, n) = 2\vartheta \left(B_n \frac{\omega}{2R_N\omega - n' \cdot bp} + R_N \varepsilon_{ads} \right) \tag{5.3}$$

$$U_X(\vartheta, n) = 2\vartheta B_n \left(\frac{\pi + \omega}{n' - 2R_N\omega)} + R_N \varepsilon_{ads} \right) + B_n \frac{2R_{DNA} tan^{-1}(\frac{4R_{DNA}R_N tan\vartheta}{\kappa})}{\kappa} \tag{5.4}$$

In the above expressions $\varepsilon_{ads} = 0.7 k_B T/nm$ is DNA-histone interaction density , $\omega = tan\vartheta - \vartheta$, bp is singe DNA base pair length and

$$n' = n - \frac{88.90\vartheta}{\pi}$$

is the number of extra base-pairs inserted into the nucleosome. Also in Eqn. 5.4, $\kappa = R_N^2 \tan^2\vartheta - 4R_{DNA}^2$.

5.3.1.1 Constrain analysis of ϑ and n boundary conditions

A lower bound for the bulge size is inferred from Eqn. 5.5 where we need to have $n' > 0$, analytical graph is provided in supplementary section. For simple bulge the binding location of the remodeler will impose and upper bound, since the SWI/SNF binds 2 SHLs from nucleosome dyad [101] and [102]. Hence, $\vartheta < \frac{3.1\pi \times 4}{13 \times 2} \approx \frac{\pi}{2}$. Also to

have a valid U_s from 5.3 procures $2R_N(tan\vartheta - \vartheta) - n'bp > 0$. Obviously the new boundary condition dominates the former one.

In Fig. 5.4, d can be manipulated as follows:

$$tan\vartheta = x/R_b, \quad tan\vartheta = (d-x)/R_N$$
$$\Rightarrow x = R_b d/(R_n + R_b) \Rightarrow d = tan\vartheta(R + R_b) \tag{5.5}$$

also the bulge radius R_b is:

$$R_b = \frac{nbp - 2R_N tan\vartheta}{2(\pi - \vartheta + \tan\vartheta)} \tag{5.6}$$

Having the above expressions we can apply the $R_b > 0$ boundary condition for the crossed which from 5.6 this would lead to $\vartheta < \pi/2$, also we need to satisfy following condition:

$$n' \cdot bp > 2R_N \tan\vartheta \quad \sim \Rightarrow n > 29.417(\tan\vartheta + \vartheta) \tag{5.7}$$

The recent condition on n (Fig. 5.5) should be considered in conjunction with the global boundary condition for n' (see 5.3.1.2). Also to avoid extreme the deformation of the bulge condition [84] proved that $R_N tan\vartheta > R_D NA$ which results $\vartheta_{min} \approx 14°$.

5.3.1.2 Global boundary condition for n'

Apart from the structure of bulge, (n, ϑ) characterizes the nucleosome bulge. The size of a bulge should always be larger than the number of base-pairs across the same nucleosome angle in the canonical nucleosome, therefore for the following preposition should always stay valid:

$$\{(n, \vartheta) | \forall n, \vartheta : ln(n) - ln(\vartheta) > 3.365\} \tag{5.8}$$

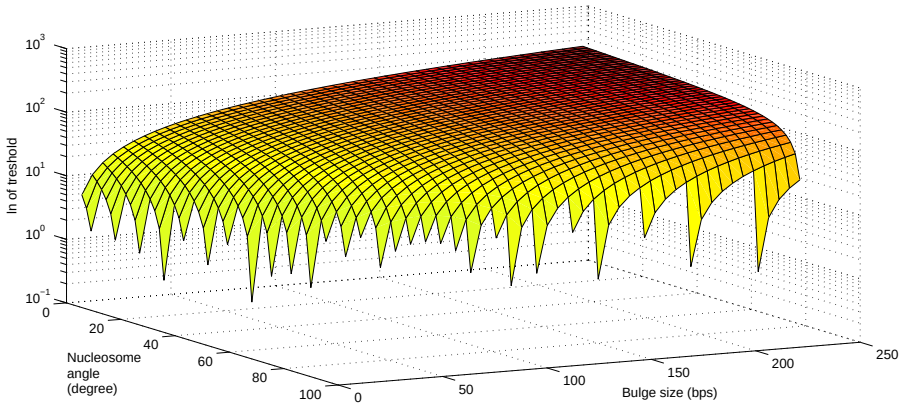


Figure 5.5. In this graph each point has coordinate of $x = n, y = \vartheta, z = ln(n - 29.417(\tan\vartheta + \vartheta))$. Natural logarithm of the threshold is used to ease the elimination of invalid (n, ϑ) tuples process, those with undefined $ln(\cdot)$.

Fig. 5.6 depicts this condition by intersecting the two plane. All points above the critical plane fits into the Eqn5.5 and could be a potential bulge parameter tuple.

5.3.1.3 Effect of DNA sequence and length on the its energy profile

In order analyze the effect of DNA sequence range for the relevant nucleosome angle we generated 10^5 individual sequence and manipulated the average bulge energy

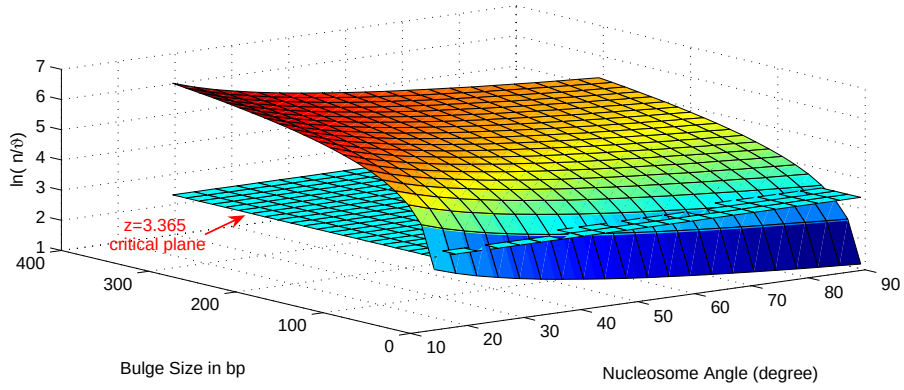


Figure 5.6. The critical plane shows the threshold for the valid (angle, bulge size) tuples..

along with standard deviation deviation for range of 10 to 200 base-pairs bulges, this manipulation is done using three curvature models of the DNA sequences in [103]:

- *Isotropic model*: DNA is equality bendable in all four planar geometric direction.
- *Anisotropic symmetric model*: DNA is equally bendable in major and minor groove direction and 10 time more rigid in the other two directions.
- *Anisotropic asymmetric model*: DNA is more bendable in the direction of major groove and 10 times more rigid in all other directions.

In Fig. 5.7 the bending rigidity for three different models is depicted. As one can observe the energy difference for sequences with < 50 base-pairs in length, could vary up to 10 times among them (as in isotropic model). However, the rigidity tends to converge to the mean as the sequence length increases. Therefore, the energy profile of *in-cis* translocation process is more sensitive to small bulges than larger ones. Furthermore, although three bending model characterize the same behavior; However there is still a noticeable difference among three, e.g. for the 269 base-pairs sequence the bending energy density isotropic rigidity could be as high as 73.66 $k_bT \cdot nm$ where the asymmetrical anisotropic rigidity could be as low as 19.37. These observations indicate that DNA sequence effect should be incorporated only in sliding whose average bulge is < 50. As a conclusion we can say that as the sequence length increases the deviation in the bulge energy significantly decrease and this observation is intuitive, firstly because for very small DNA sequence bending will average only few base-pairs and this will vary significantly due to the E_b for the small sequence. For a large sequence especially where no short repeat pattern exists the bending energy will averages over larger number of base-pairs with more diverse trinucleotide sequence stack. Also the difference between the energy level among the three bending models is noticeable for small bulges however for the medium and large bulges these differences are negligible. Therefore choosing any of

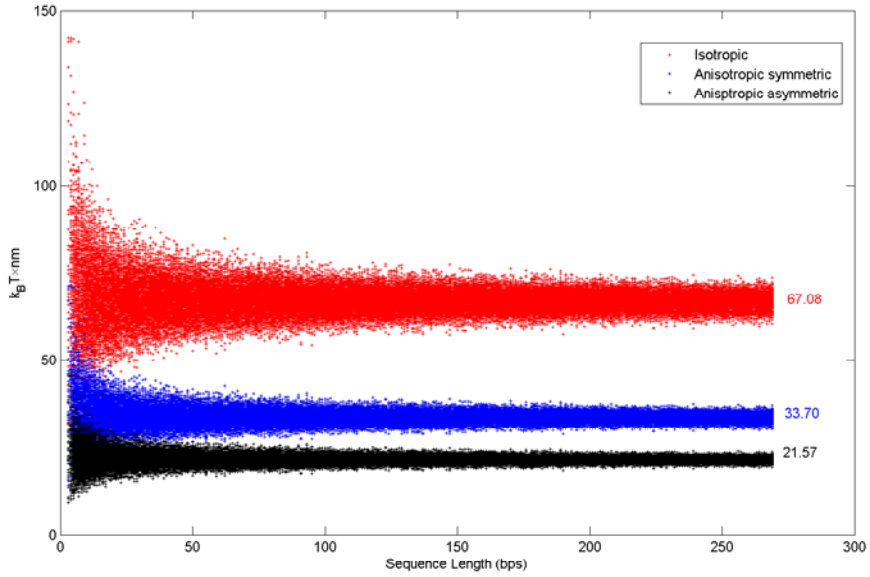


Figure 5.7. Sensitivity of the DNA curvature energy with respect to the length and sequence for three differed bending models. 10^5 random sequences is plotted for each model where each sequence length has an upper bound of 357 instances. Red, blue and black plot depicts the data for isotropic model, anisotropic symmetric model, and anisotropic asymmetric model, respectively.

the three models would not skew the final result. We will use the anisotropic asymmetric model for the rest of the manipulations in this study.

5.3.1.4 Drag force

The sedimentation of the complex molecules in a viscus environment such as nucleus and cytoplasm impinges a viscus drag force which is approximately proportional to the motion velocity of the macromolecule. Since the sedimentation of the linker DNA is in the order of nm/s, we can use the *Gabriel Stokes* approximation for drag force, by using

the equivalent spherical object moving in the nucleus viscus environment. Therefore, the required drag force in bulge formation with n' extra base-pairs will be:

$$f(n') = 3\sqrt{2}\pi\eta\left(R_{DNA}(R_{DNA} + n' \cdot bp)\right) v \tag{5.9}$$

in the above expression η is the fluid viscosity and $1 \leq v \leq 6 \;\; nm/s$ is the velocity at which the molecule displaces in the environment, [2]. The drag force energy factor for a known v and n' could be maximized when the required displacement is maximized, such drag energy factor is obtain from:

$$U_f(n') = f(n') \cdot n' \cdot bp \tag{5.10}$$

Fig. 5.8 shows the drag energy factor for three nucleosome angles, $v = 6 \; nm/s$ and nucleus viscosity is considered 6.6 *cp*, [104]. As we can see this energy factor is in the order of of mk_BT, which is negligible compare to the other energy factors. Therefore, we will exclude the drag force energy from our calculations.

5.3.2 Model for bulge transition

Following the bulge formation by torsion sub-domain, the tracking sub-domain uses ATP hydrolysis energy to track the bulge around the nucleosome toward the distal entry point on a unidirectional path. To the date of this report no *in-vitro* or *in-vivo* observation has been found to address the step-size and the quality of the bulge tracking, therefore we propose a model to abstract the bulge tracking phase. For this purpose we make an assumption that at least one ATP molecule is required for each tracking step. Hence, based on this hypothesis should a hydrolysis reaction releases more energy than demand, it would be either consumed for protein conformations or dissipated from

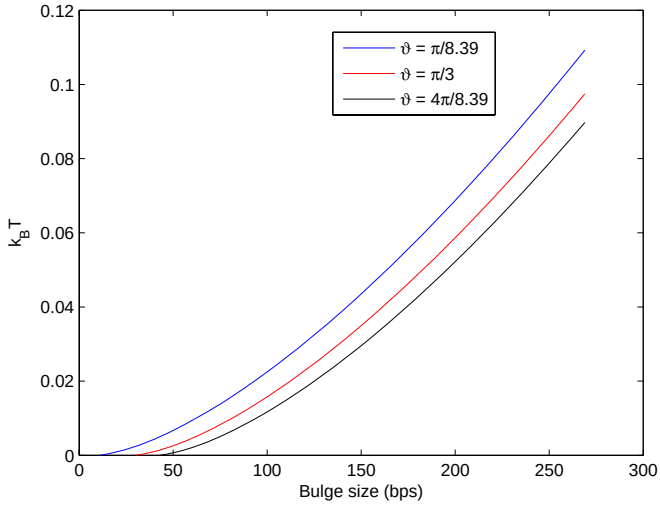


Figure 5.8. The force impinged on the DNA by the torsion sub-domain of remodeler in the order of pN [1], this force should ultimately keep the nucleosome displacement pace with motion of RNAP on the DNA, hence we have chosen $v = 6nm/s$ which according to the same article and [2] is a legitimate upper bound. In this figure red line shows drag energy factor for the small $\vartheta = \pi/8.39$, blue is for $\vartheta = \pi/3$ and black one has a large nucleosome angle of $\vartheta = 4\pi/8.39$. The coefficient is proportional to the velocity and square root of excessive DNA length.

the system. Noteworthy, the last tracking step for both models is a relaxing step where the nucleosome returns to the ground state. On the other hand, the relaxing step of a tracking microevent requires the energy just to disrupt the remaining DNA-histone interactions.

5.3.2.1 Energy profile of the tracking phase

following the earlier assumptions and definitions, for all the tracking steps bulge coordinates should follow:

$$\forall x_s, x_s : x_s < x_e, x_s \epsilon [1..136], x_e \epsilon [2..137]$$

The number of steps can still be calculated from:

$$k_{f,i}^{\vartheta} = \lfloor \frac{3.1\pi - 2\vartheta}{\beta(i)} \rfloor \tag{5.11}$$

where $\beta(i) = \frac{2\pi i}{88.9} \cdot bp \ | i \epsilon \{1, 2, ..., n\}$. Remaking that $\beta(i_{max}) = min(\frac{n\pi}{88.9}, 3.1\pi - 2\vartheta)$. Therefore, total number hydrolysis reaction required for the bulge of size n for the all tracking steps in one instance of remodeling is:

$$N_{n,i}^{\vartheta} = \lceil \frac{U(\vartheta, n) + U_{Rlx}}{\triangle_{ATP}} \rceil + k_i^{\vartheta} \lceil \frac{\beta(i) U(\vartheta, n)}{2\vartheta \triangle_{ATP}} \rceil \tag{5.12}$$

In Eqn. 5.12 to choose the conformation with lower energy barrier (to increase the entropy of the ensemble) we set $U(\vartheta, n) = min\{U_S(\vartheta, n), U_X(\vartheta, n)\}$, and $\triangle_{ATP}$ is the potential energy of the phosphate bond of a single adenosine triphosphate (ATP) molecule.

It should be emphasized that hydrolysis reaction of SWI/SNF is very fast reaction since it is an enzymic reaction, this claim could be observed from kinetics rate constants (K_m , k_{cat} , V_{max}) that are experimentally measured in [97]. Therefore, the main factor that would effects the time between two successive hydrolysis reaction is the arrival rate of the fuel molecules, ATP. It is very common to assume inter-arrival times, which is the inter-arrival of two successive hydrolysis reaction, follows an exponential distribution [105]. Considering such distribution with the mean time $\bar{t} = \frac{1}{\lambda}$, where the λ is the rate, each remolding process is a sequence of N_n^{ϑ} exponential inter-arrival time of hydrolysis reaction. The distribution of these cumulative inter-arrival times will follow an erlangian distribution:

$$q(N_n^{\vartheta}, t|\lambda) = \frac{\lambda(\lambda t)^{N_n^{\vartheta}-1}}{(N_n^{\vartheta} - 1)!} e^{-\lambda t} \tag{5.13}$$

5.3.2.2 Bugle size distribution

The other component that would affect the remodeling process is the bulge size. In [1] researchers used optical tweezers on ySWI/SNF and yRSC to conduct their experiments and collected the sample data on the SWI/SNF and RSC bulge size. We were able to roughly fit their data into Generalized Extreme Value (GEV) distribution , $g(x|\mu,\sigma,\xi)$. The estimated parameter and their respective standard error for this distribution are as follows: location parameter μ: $(59.6917, 2.47457)$, shape parameter σ: $(32.3734, 2.33)$, and scale parameter ξ:$(0.485177, 0.0719)$. In order to find the probability mass function of the bulge size n, we discrete the GEV probability distribution G_X, on the vicinity of n as follows:

$$\begin{aligned} j(n) &\approx \frac{G_X(n+\frac{1}{2}|\mu,\sigma,\xi) - G_X(n-\frac{1}{2}|\mu,\sigma,\xi)}{\sum_{i=1}^{\infty}\left[G_X(i+\frac{1}{2}|\mu,\sigma,\xi) - G_X(i-\frac{1}{2}|\mu,\sigma,\xi)\right]} \\ &\Rightarrow = \frac{G_X(n+\frac{1}{2}) - G_X(n-\frac{1}{2})}{G_X(1+\frac{1}{2}) - G_X(1-\frac{1}{2}) + G_X(2+\frac{1}{2}) - G_X(2-\frac{1}{2}) + \ldots} \\ &\Rightarrow = \frac{G_X(n+\frac{1}{2}) - G_X(n-\frac{1}{2})}{1 - G_X(\frac{1}{2})} \end{aligned} \tag{5.14}$$

For the estimated parameters $G_x(\frac{1}{2},\mu,\sigma,\xi) << 1$, therefore, the denominator will be roughly equal to unity and will be eliminated from the expression. Hence the estimated probability distribution for bulge size of length n would be as follows:

$$j(n) \approx \exp\left[-(1+\xi z)^{\frac{-1}{\xi}}\right] - \exp\left[-(1+\xi \acute{z})^{\frac{-1}{\xi}}\right] \tag{5.15}$$

In the Eqn. 5.14, G_X is the cumulative distribution function of GEV, $z = \frac{x+0.5-mu}{\sigma}$, and $\acute{z} = \frac{x-0.5-mu}{\sigma}$. Now from Eqns. (5.14, 5.13) we can write the expression for probability distribution function (PDF) of n base-pairs transposition in time t as following:

$$P(n, t|\vartheta, \lambda) = j(n) \cdot q(N_n^{\vartheta}, t) \tag{5.16}$$

From 5.16 one can manipulate $T(\vartheta)$, the average time required for a remodeling with mean bulge length whose probability distribution functions follows the Eqn. 5.14 from:

$$T(\vartheta) = \frac{N_M^{\vartheta}}{\lambda} \quad where \quad M = \lceil \mu + \frac{\sigma}{\xi}(\Gamma(1-\xi) - 1) \rceil \tag{5.17}$$

The only missing part of the Eqn. 5.17 is ϑ. To the day of this report there no published data is available on angle and its distribution. Hence, in Eqn. 5.17 for each constant nucleosome angel ϑ and mean ATP arrival rate λ we will have an instance of probability distribution function. The boundary condition derived earlier in addition to the experimental data on micrococcal DNA cleavage of nucleosome DNA will be used in the result section to find the proper nucleosome angle.

5.3.3 SWI/SNF binding time

To estimate the time for binding of the SWI/SNF complex to the target nucleosome we adopted the collision theory based stochastic reaction model proposed by [85] and manipulated the reaction rate as follows:

$$\rho = C_{SWI/SNF}(r_{SWI/SNF} + r_N)^2 N_A \sqrt{\frac{8k_B T(m_{SWI/SNF} + m_C)}{m_{SWI/SNF} \cdot m_C}} e^{-\frac{E_{act}}{k_B T}} \tag{5.18}$$

In above expression parameters are as follows: $C_{SWI/SNF}$ concentration of the remodeling complex, $r_{SWI/SNF}$ radius of remodeling complex, r_N nucleosome radius, N_A Avogadro

number, $m_{SWI/SNF}$ remodeling complex mass, m_C is the cumulative mass of nucleosome and DNA sequence, and E_{act} is the activation energy of the reaction.

5.3.4 Estimating the tension

The tension which is applied on the nucleosome by the remodeler to induce the DNA loops of size n in each remodeling cycle, estimated as follows:

$$\begin{aligned} w(n) &= f_n \cdot d \\ \Rightarrow E_{rem} 4.11 \times 10^{-21} &= f \times n 0.34 \times 10^{-9} \\ \Rightarrow f_n &= 12.88 \times E_{rem} n \end{aligned} \tag{5.19}$$

In Eqn. 5.19 the force f_n is in pico-Newton and remodeling energy E_{rem} in $k_B T$. The remodeling energy is proportional to the number of hydrolysis reactions ($E_{rem} \propto N_n^{\vartheta}$) from Eqn. 5.12. Hence, we can write the following relation for the tension:

$$f_n = 12.88 \times C_{eff} n \Delta_{ATP} \tag{5.20}$$

In this equation $C_{eff} < 1$ is the energy efficiency coefficient of the system and Δ_{ATP} is manipulated based on the 29 $kcal/mol$ of ATP.

5.4 Results

This section first we provide a quantitative analysis on the model and then validate our result with experimental data.

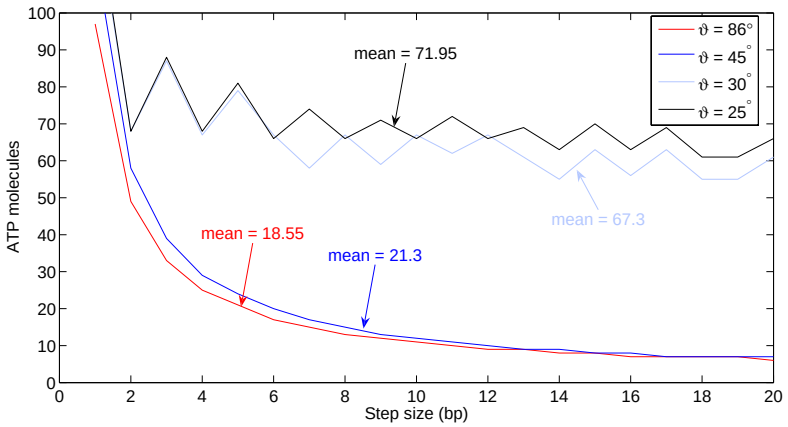
5.4.1 Analysis of the remodeling ATP profile

To analyze the ATP profile of one remodeling cycle we manipulated the energy profile for selected paths from the random graph 5.3. While selecting these paths we tried to choose those that could represent the diversity of the energy profile that exists within graph paths.

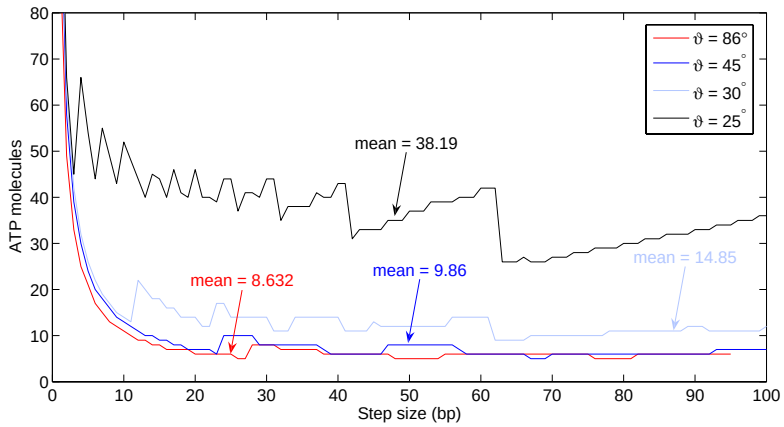
In Fig. 5.9 we initially chose 12 bulge formation state: *(1,12,L,1), (1,15,L,1),(1, 22,L,1),(1,42,L,1)* with *L=20, 50, 150*. Then we manipulated the ATP profiles from all possible tracking paths descended from the 12 initial paths. Hence, Fig. 5.9 comprises ATP analysis of 960 paths of the entire remodeling state space. The fluctuation observed on individual ATP profiles demonstrated in Fig5.9 is due to the elimination of the last tracking step and replacement of that with a relaxation step which energetically is less costly. Such circumstance is possible when addition of one base-pair to the step size can convert the last state to a relaxation step(is favorable for the process), or impose fewer but accumulatively more expensive tracking phase to the remodeling process. Small bulges with acute angle are energetically very expensive therefore addition of one step could be significant. For bending rigidity of DNA sequence we used the third bending model outlined earlier and $1/2(e_{max} + e_{min})$ from a pool of 350 randomly generated sequences with the target length given by the distribution in Eqn. 5.6

5.4.2 Sensitivity analysis of the remodeling PDF

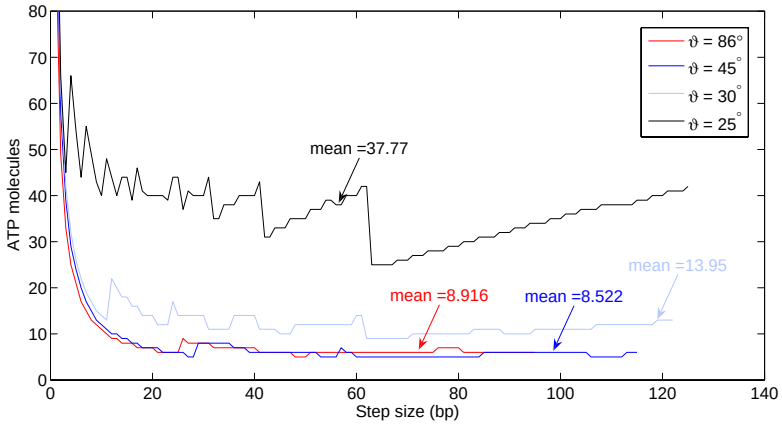
Bulge size , n, and time, t, are the two stochastic components of the model whose joint PDF was derived in 5.16 their effect of the remodeling PDF is sketch in Fig. 5.10.(a) for a constant nucleosome angle. In the probability distribution function of the remodeling cycle, nucleosome angle ϑ works concurrently as shift and scaling parameters. Fig. 5.10.(b) shows this concurrent effect for a fixed $n = 100$, and as one can observe such effect is not liner. This none linearity could be predicated as well as justified from



(a)



(b)



(c)

Figure 5.9. The nonlinear dependency of the remodeling ATP profile of a remodeling cycle on the bulge size, corresponding nucleosome angle, and the tracking step size;(a) small size bulge size of 20 pbs, (b) medium size bulge of 100 bps, (c) large bulge of 150 bps. The topology of the bulge is chosen based on minimum energy profile of the bulge.

the impression of nucleosome angle on the energy profile of process which were analyzed earlier.

5.4.3 Analysis of sliding rate versus λ

We have simulated a period of 700 seconds of back to back remodeling cycles. During first 200 seconds of simulation the stochastic nature of bulge size led counter intuitive remodeling of higher rate for smaller λ, in Fig. 5.11.(a) where the end of initial phase is depicted by a vertical dashed line. We simulated the sliding process for different values of λ out of which three instances are shown in Fig5.11.(a). To further analyze the fluctuation of sliding rate we magnified the results for a 100 seconds time frame of Fig. 5.11.(a) (specified by the dotted rectangle) in Fig5.11.(b), where the contribution of the large and small size bulges to the sliding rate is clearly observable. As Fig. 5.11.(b) shows the formation of bulges with extreme sizes is a rare phenomena, this fact is endorsed by observations in [1, 102]. Fig. 5.12 shows the rate of octamer *in-cis* translocation versus ATP arrival rate λ. As we can observe sliding rate and ATP arrival rate have a linear dependency.

5.4.4 Sensitivity of reaction time to the E_{act}

Fig. 5.13 shows the valid range of the activation energy for biological reactions, reported in [85], versus reaction time for three different molecular collision times, δt.

5.4.5 Validation of the model

To investigate the nucleosome translocation, experimentalist use endonuclease restriction enzyme to cleave the binding sites that are normally restricted by the nucleosome. One of the most prevalent practical approach is to use the synthesized mononucleomsome such as 343 base-pair DNA fragment with 601 nucleosome positioning at the

center of the DNA. In Fig. 5.14 this mononucleosome is demonstrated and each arrow points to the site sensitive to a particular restriction enzyme indicated by a label on base of the arrow. The general method for such assays briefly include mixing a known concentration of the mononucleosome, remodeling complex, and restriction enzyme in a reaction buffer and monitor the percent of DNA cut over time. The position of the site along with the percentage of DNA cut over time could be used to estimate the translocation rate of the histone core. We also used the event diagram of Fig. 5.3 to mimic the same process in-silico. For every in-silico experiment we conducted 10 different realization of the experiment with individual seeds. However, such strategy will add *DNA nuclease* event to the process whose temporal behavior should to be quantified. To quantify the nuclease event we reviewed different experimental assay of this kind in [97, 101, 102, 100, 92] and noticed that in all assays the concentration of the restriction enzymes were significantly higher than the mononucleosome concentrations. The concentration of restriction enzyme in these assays are in the order of tens of restriction enzyme units (U). Each U is the amount of restriction enzyme that is needed to completely digest one microgram of DNA in hour in 50 μl of reaction volume. This indicates a DNA weight of 0.02g per liter, having this amount of DNA completely digested in 1 hour will result an approximate net rate of $1.825 \times 10^{18} \;\; base - pairs/hour/U$ or $< 1fsec/base - pairs/s/U$. As it is observed this time is infinitesimal and could be ignored in our calculations.

5.4.5.1 Validation scenario 1

For this in-silico experiment we used the wet lab experiment in [97] as the reference for validation. Based on that we used *Hhal* restriction enzyme, therefore the required displacement is +75 or -79 base-pairs. The sign of the displacement distinguishes the direction of repositioning with respect to the position of the target DNA site on the nucleosome. We used the same concentration of mononucleosome and SWI/SNF as

Table 5.1. Experiment 1 Parameters

Parameter	Value
Nucleosome	5 nM
SWI/SNF	0.5 nM
ATP	1mM
Temperature	300 K
mass SWI/SNF	2M Dalton
DNA length	343 bps
E_{act}	11 k_BT
ϑ	86°

reported on page 5886 of the same paper, all the simulation parameters are reflected in and reflected in Table5.1.

To calibrate the simulation we had to plug in different values for λ, our results best matched the experimental data where $\lambda = 10.53$. More interesting outcome of this in-silico experiment is that we learned that our objective ATP arrival rate agrees with $k_{cat}/60$ which is reported in the same paper. Fig. 5.15 shows the comparison between the in-silico result and data from wet lab experiments. In Fig. 5.19 we have shown the same measurements as in Fig. 5.15 but from the reverse angle, which is the time versus percent of DNAsed nucleosomes.

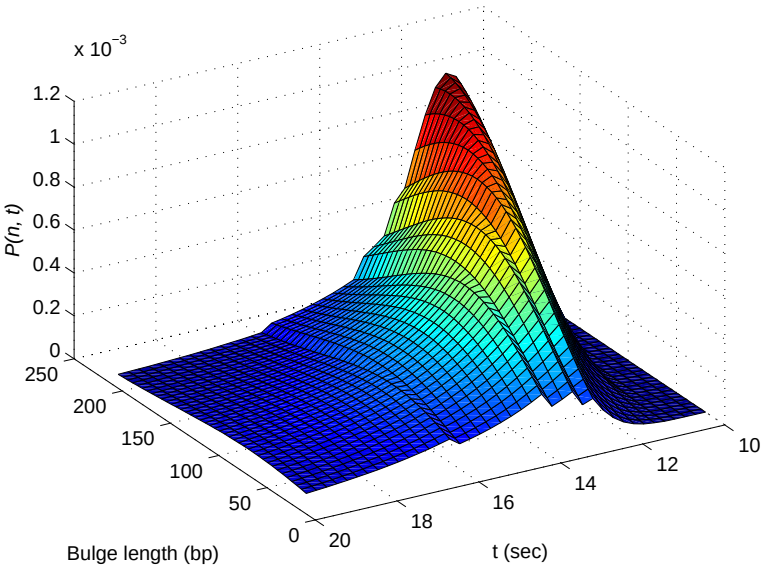
5.4.5.2 Validation scenario 2

For this in-silico experiment we used the assay in [1] as reference for validation, where the researchers used the optical tweezers to mimic the SWI/SNF tension strained on the nucleosome during the remodeling and measure the force threshold for creating a nucleosomal bulge. They also estimated the sliding rate of the nucleosome under certatin ATP and SWI/SNF concentrations. We used the same concentrations for the SWI/SNF and ATP and applied our model and discrete event simulation to estimate both the

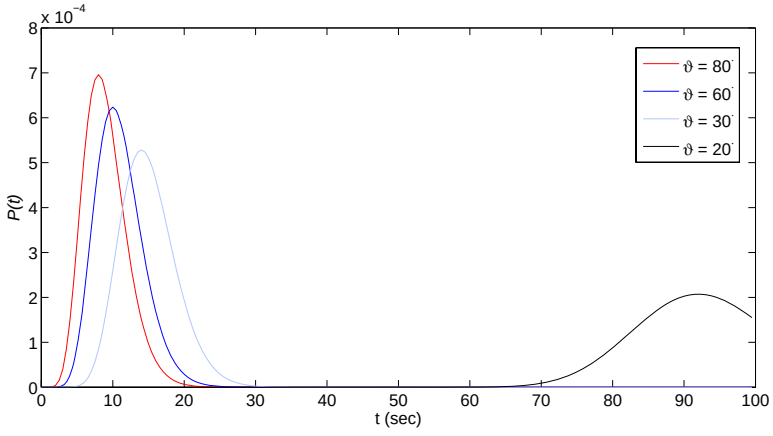
Table 5.2. Experiment 2 Parameters

Parameter	Value
Nucleosome	5 nm
SWI/SNF	0.5 nM
ATP	1mM
Temperature	300 K
mass SWI/SNF	2M Dalton
DNA length	3M bps
E_{act}	11 k_BT
dt	10^{-8} s

induced force by SWI/SNF from Eqn. 5.20 and sliding rate of the histone core. The molecular concentrations and model parameters for this experiment are listed in Table 5.2.

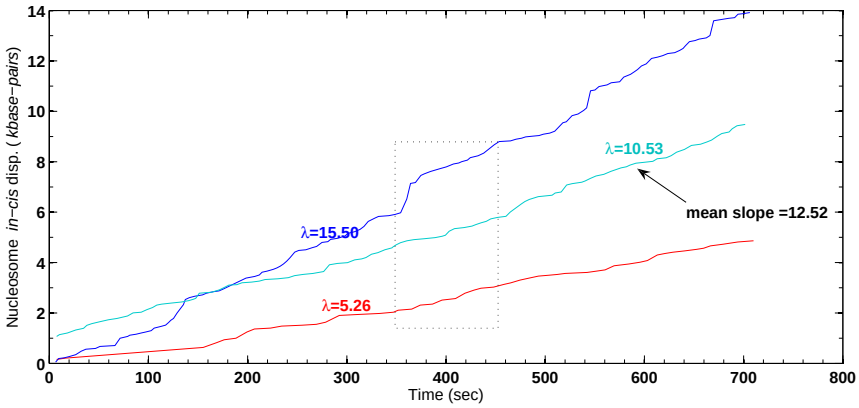


(a)

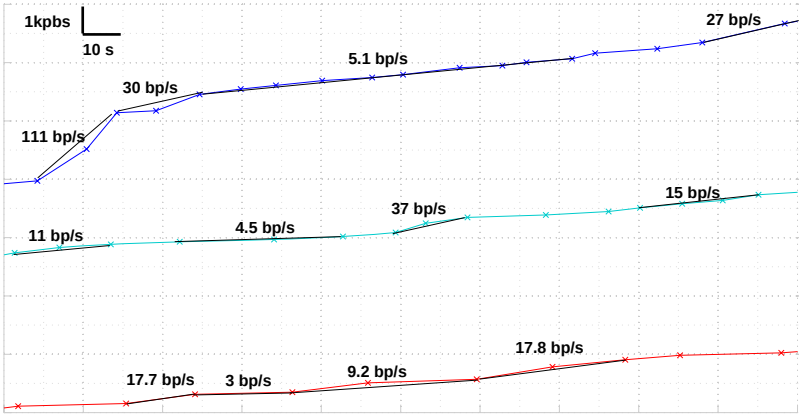


(b)

Figure 5.10. (a) PDF of a remodeling cycle $P(n,t)$ for constant nucleosome angle $\vartheta = 30°$ for a reasonable range of bulge size;(b) depicts the effect of the nucleosome angle on the PDF of remodeling. In fact graph (b) is the front view of the (a) for $n = 100$ and if fixed nucleosome angle.



(a)



(b)

Figure 5.11. (a) 700 seconds simulation of a remodeling cycles burst for three different λ values with nucleosome angle $\vartheta = 86°$;(b) shows the magnification for (a) where $350 \leq t \leq 450$ sec.

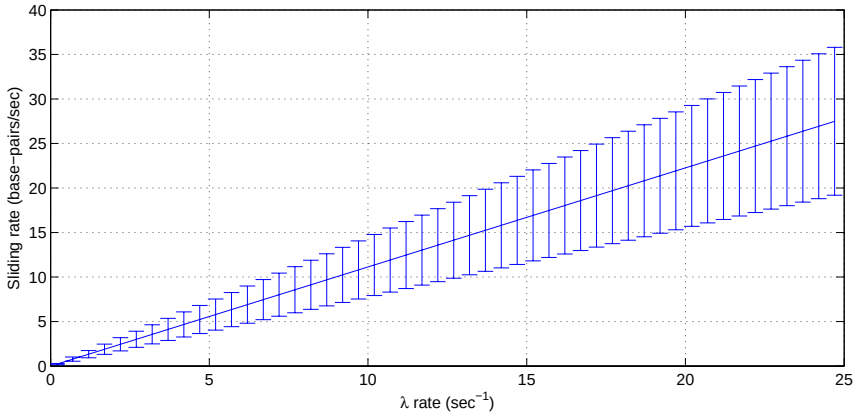


Figure 5.12. The effect of ATP arrival rate on histone core sliding rate.

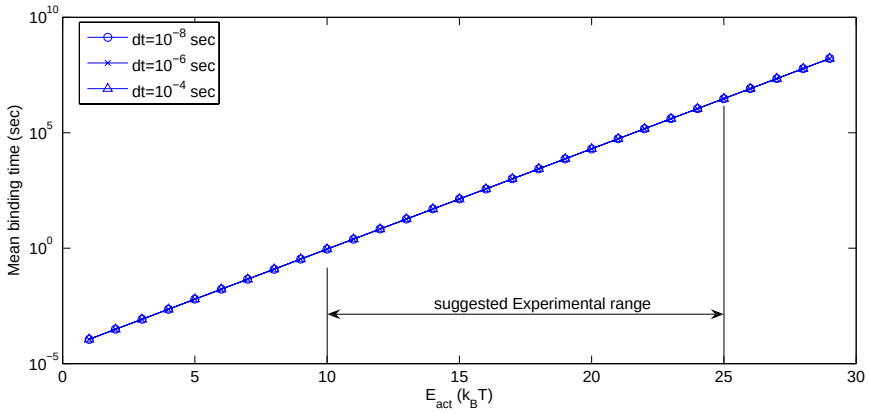


Figure 5.13. Sensitivity of the model to activation energy for the remodeler-nucleosome binding.

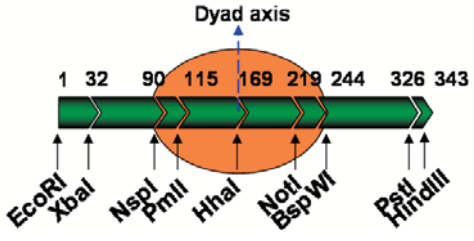


Figure 5.14. Synthesized mononucleomsome with 601 sequence nucleosome positioning.

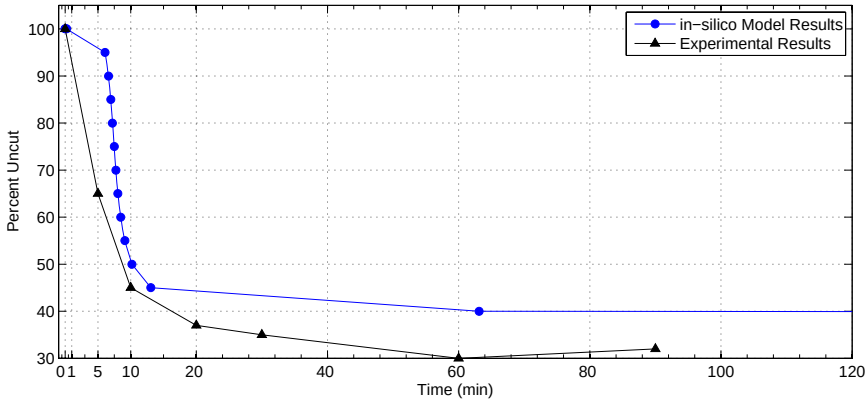


Figure 5.15. Comparing the in-silico and wet lab results for percent of mononucleosome cut versus time, for restriction enzyme *Hhal*.

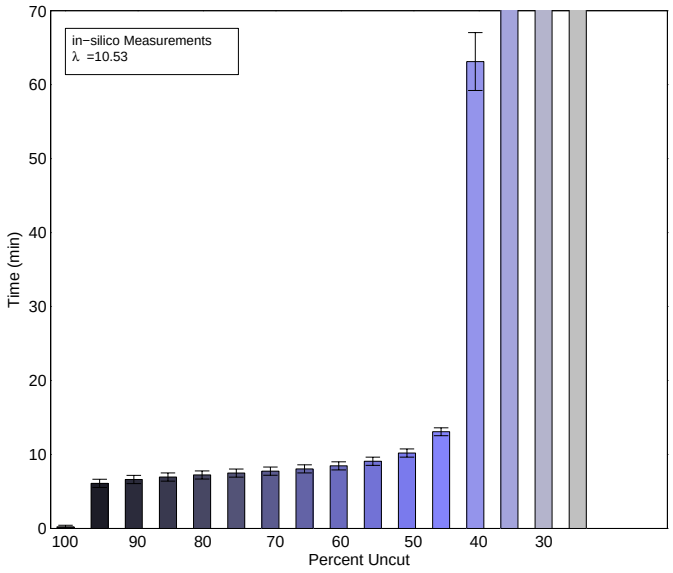
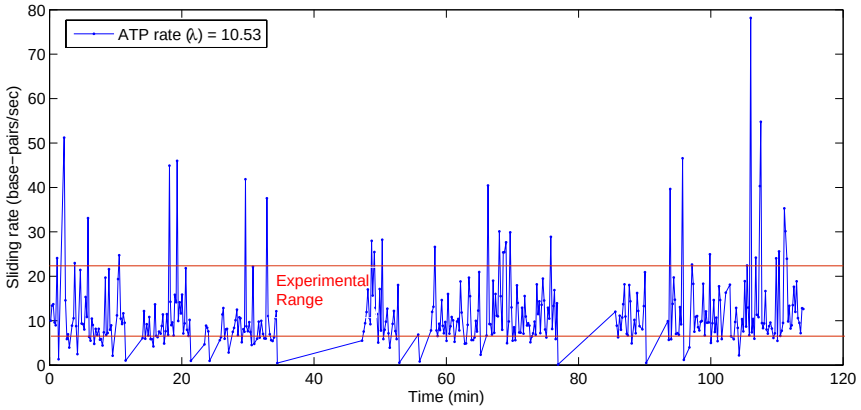


Figure 5.16. Time versus percent of sequences cut.

(a)



(b)

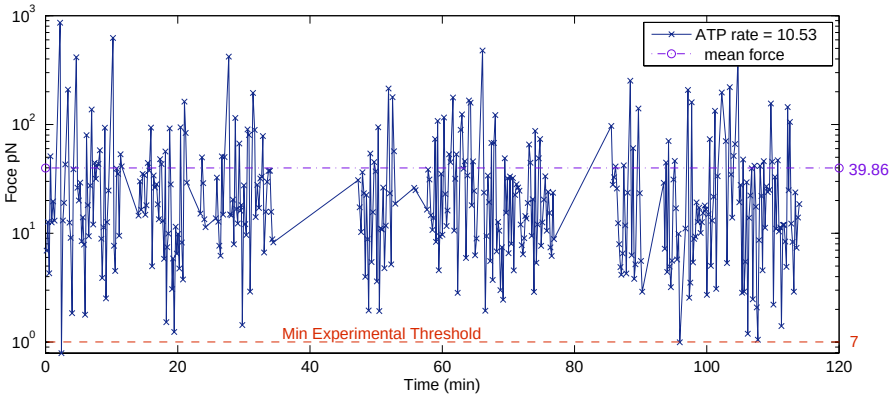
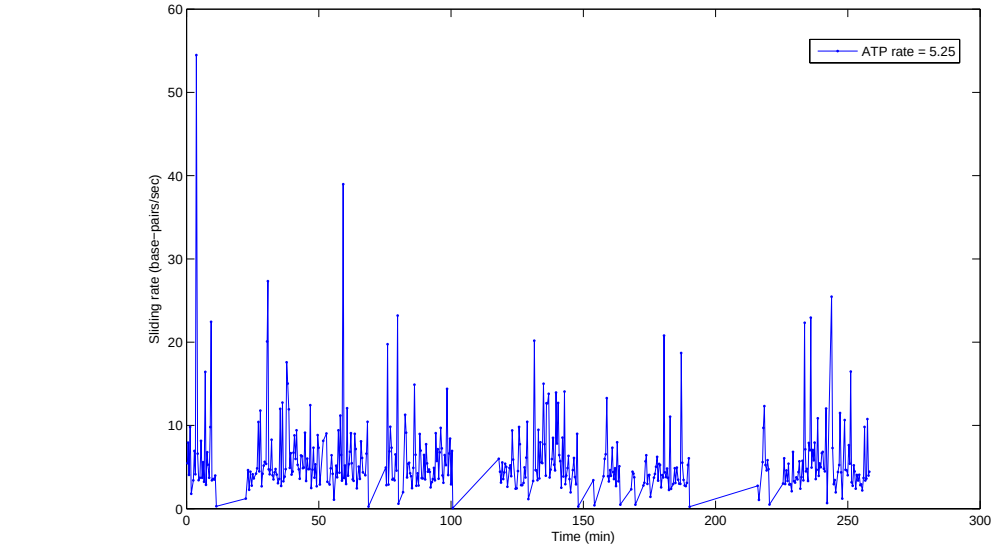
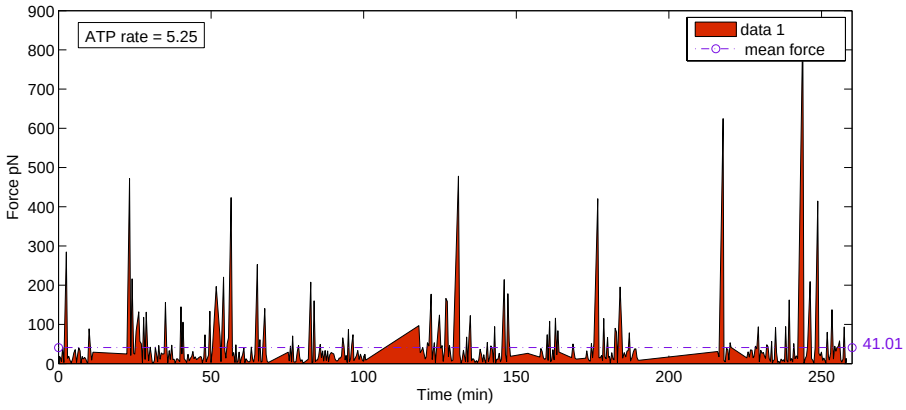


Figure 5.17. (a) is the realtime in-silico slide rate of the nucleosome and (b) shows the quantified tension applied tension on the nucleosome for $\lambda = 10.53$.



(a)



(b)

Figure 5.18. (a) is realtime in-silico slide rate of the nucleosome and (b) the measured applied tension on the nucleosome for $\lambda = 5.25$.

Fig. 5.17.(a) is the realtime slide rate of the nucleosome for the given experimental parameter and ATP arrival rate of 10.53. The deviation observed on the image is due to the randomized bulge size. Fig. 5.17.(b) is the projection of (a) on the force field. As we can observe all the reported forces are well above the minimum of $7pN$ which is the required force threshold for remodeling reported in [1]. The average magnitudes of the tension is 39.86 is twice greater than the reported experimental results. Fig. 5.18.(a,b) are the same as Fig. 5.17.(a,b) but for ATP rate of 5.25. There is no experimental reported data for this arrival rate; however the interesting observation is the linear dependency of the sliding rate of nucleosome to the λ which is more obvious in chart of Fig. 5.18.(a). Also in Fig. 5.18.(b) as we can see the mean of the force is roughly remained the same only the number of higher picks have remarkably dropped. This is physically relevant, because the magnitude of force that is required to unbind the DNA from nucleosome should not change; however, its should span over a longer time epoch that leads to lower number of high pick and hence slower sliding rate.

5.4.5.3 Validation scenario 3

For the third in-silico experiment we refer to a Nature Structural and Molecular Biology published article by Wang *et. al* [98], where the authors endeavored to probe the catalytic efficiency (K_m/K_{cat}) of the SWI/SNF. For this purpose they investigated and reported the temporal dynamics of nucleosome *in-cis* translocation. As advised by the authors we used a 774 base-pairs DNA segment with 601 nucleosome positioned at 400 base-pairs from the beginning. In this experiment we monitor number of the nucleosomes that had any displacement (remodeled) over 15 minutes of reaction time. The concentration of species and model parameters are listed in Table 5.3.

Fig. 5.19 shows the fraction of the nucleosome remodeled for 15 minutes of the experimental reaction time as well as the simulation time. For in-silico results we stayed

Table 5.3. Experiment 3 Parameters

Parameter	Value
Nucleosome	5 nm
SWI/SNF	0.5 nM
ATP	1mM
Temperature	300 K
Mass SWI/SNF	2M Dalton
DNA length	774 base-pairs
E_{act}	11 kBT
dt	10^{-8} s

consistent with the perviously validated $\lambda = 10.54$. The characteristic displacement (CD) is a parameter that is introduced by the reference article which is the average displacement of the remodeled nucleosome over 1 minute of reaction time per reaction event. The reported experimental $CD = 66.5 \pm 5.3$ base-pairs, and the in-silico $CD = 132.36 \pm 90.41$. The in-silico CD is about twice larger that the experimental counterpart because in this assay the average bulge size for SWI/SNF was not consistent with $\approx$ 100 base-pairs reported in many other experimental data elsewhere. To modify the average bulge size a new set of data samples for fitting the distribution curve is required necessary.

5.4.5.4 Model Verification

After validating the model with experimental data we have used the model to predict the time which is required for the SWI/SNF to make an occluded region of a genome accessible to the other proteins and complexes. This region may include the an upstream promoter region or a Open Reading Frame (ORF) of a target gene. To continue with our estimation we need to know *'nucleosome occupancy'* [106] of the target region or the average *'global nucleosome occupancy'* across the genome. Nucleosome occupancy

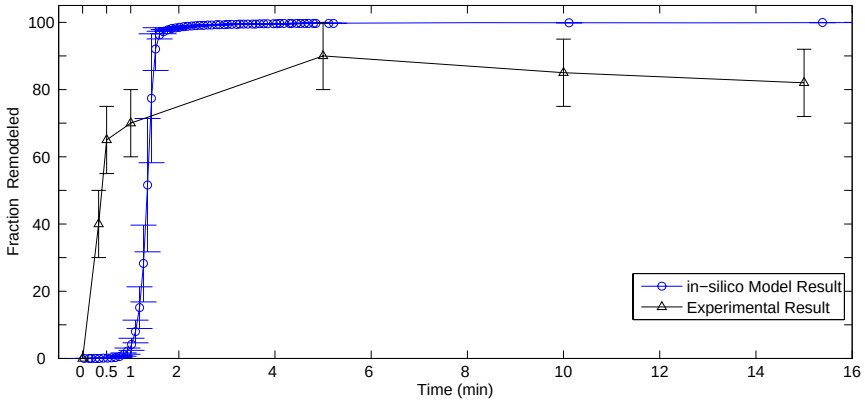


Figure 5.19. Fraction of nucleosomes remodel versus time; the blue line with circle markers are the in-silico results and the black line with triangle markers are the experimental results.

defines the probability that a base pair in a target section of a genome is covered by a nucleosome.

For this experiment we used the following parameters: a reaction volume of human cell nucleus $V = 4 \times 10^{-15} m^3$, [SWI/SNF] = 250 molecules based on the data in [94], the other parameters are same as Table 5.11. For this predication we make the following assumptions:

- minimum length of each linker DNA is 53 base-pairs.
- nucleosomes are uniformly distributed across the genome.
- a bound SWI/SNF does not disassociate from a nucleosome until the sliding is completed.
- SWI/SNF has an equal chance to binds to the 3' and 5' DNA strand.

If we denote the nucleosome occupancy with O, the number of nucleosome over a L base-pair genome section will be $\lfloor \frac{L \times O}{200} \rfloor$, and the adjacent nucleosome base-pair distance $s = \lfloor \frac{(1-O) \times L}{k} \rfloor$. Therefore, the cardinality of the tranlocase events can be obtained from :

$$|tranlocase \ \ events| = \lceil \frac{1}{2} L \cdot (k+1) \cdot (1-O) \rceil \tag{5.21}$$

Also we set the simulator such that the size of all any bulge would not exceed s. Fig. 5.20.(a) shows average realization of time required to open a genome region ranging from 1 base-pairs up to 3 kilo base-pairs length for three different nucleosome occupancy. The result are counter intuitive since we expect that lower occupancy takes lower time. However, in most data points all three cases shows the same temporal behavior, because of the parallelism that existed in the remodeling events. Since we had 250 copies of the remodeling complex so there were an upper bound 250 parallel remodeling that could happen simultaneously. Also as we can observe for the sequence with 0.3 occupancy there are few case where the process takes remarkably longer time than the heavier occupied equivalent length sequence. By rigorously looking into the data and the monitoring the process we found out that this counter intuitive phenomena happens because the remodelers in the lesser occupied sequence have a larger range of bulge size to pick from. Also since the arrival process of ATP follows and poisson process the once a very large size bulge is selected the cumulative time grows remarkably faster. In an attempt to avoid this problem we forced the simulator to avoid bulges larger that 250 base-pairs. The smoothed results is shown in Fig. 5.20.(b). Another observation from this experiment is the time for opening the occupied sequence grows linearly with sequence length irrespective of the occupancy ratio.

5.5 Discussion

In this work we aimed to capture the dynamics of nucleosome translocation *in-cis* in the presence of SWI/SNF which is an ATP dependent $\acute{3} \rightarrow \acute{5}$ remodeler. In the proposed model $\vartheta,\ n,\ \lambda$, are the three critical parameters. Should these parameters set properly one can leverage model to other classes of remodelers that implement the same mechanism. ISWI family has a remodeling mechanism that is very similar to SWI/SNF,[101, 93], their main difference is the smaller size of bulges for ISWI. RSC,

another close relative of SWI/SNF that implements the same mechanism, [101], but possibly with different bulge size distribution. Hence, given the quantitative data regarding the bulge size the model could be adopted accordingly. ATP arrival rate, λ, is directly proportional to the concentration of the ATP where is in the range *mili-M* for cytoplasm and hundreds of *micro-M* for the nucleus. Due to the sensitivity of the model to λ precise adoption of this value for the mimicking the experimental condition is crucial.

Also in our model we focused on the most well-known structure of the nucleosome that accommodate $\sim$147 base-pairs. However, it is a trivial task to use this model for a nucleosome with different base-pair density. One just needs to replace the density of base-pairs and use the corresponding adsorption energy density, ε.

Structural conformation of SWI/SNF avoids the reverse loops, [107], consequently no intermediate jumps to the death state is not considered in the state space. Furthermore, the relative direction of nucleosome depends on the tracking strand. The tracking strand of DNA is determined by the binding side of nucleosome to the protein complex. Hence, this fact would not bring any limitation to our model. It should emphasized that all known remodelers that follow the *wave-rachet-wave* mechanism slide the nucleosome from $\acute{3} \rightarrow \acute{5}$ across the tracking strand. There is no sequence preference in our model, which is supported by the evidence that SWI/SNF has no sequence specificity, [94].

One of the challenges in our modeling process was the to find the experimental data at the molecular level. With a countable exception, [93, 102], all other available wet lab studies reported the data in batch. Specifically in chromatin remodeling the analysis of remodeling process is based on the quality of cutting of restriction enzyme activity, e.g. Endo VII, HhaI, PmlI, etc., on often synthesized DNA sequence with mononucleosome. Although such results will show the evolution of the process over time though reaching an explicit value is not possible all the times.

One of the interesting features of our model which stays through for all models with mesoscopic scale is that one can alternate abstraction elevation based on the desired degree of accuracy. In our model the entire bulge formation and tracking could be fitted with a negative exponential distribution with an average rate equal to the mean of the simulation results. The other way around is to increase the accuracy by not using the average step size, but to devise an other distribution and incorporate that for the step size.

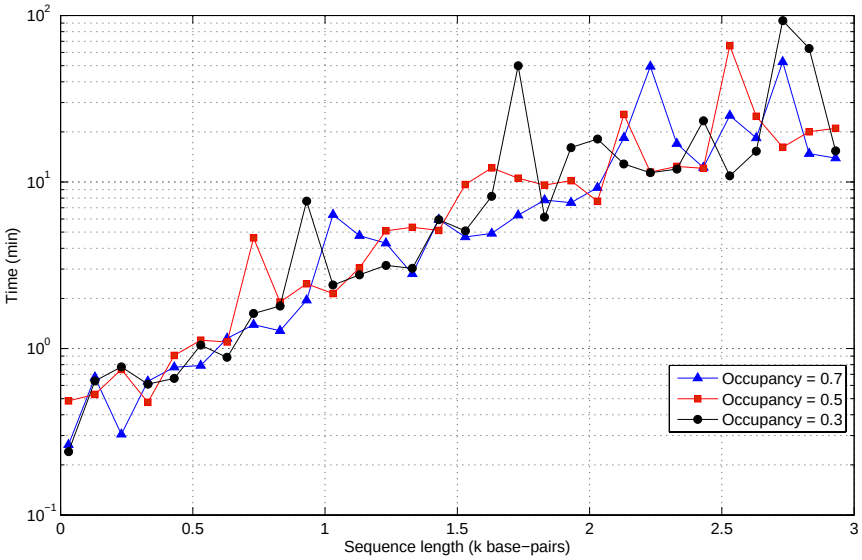
5.5.1 Claims

I. in-cis Translocation rate of nucleosome follows a Gamma distribution.

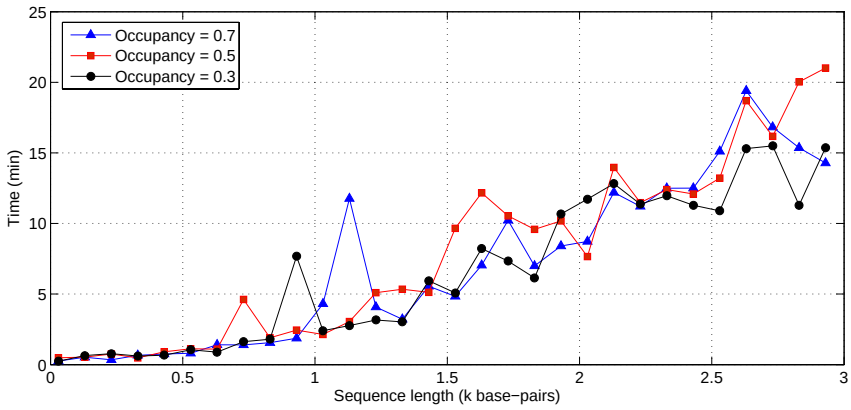
We collected the SWI/SNF induced sliding rates of nucleosome, analyzed their distribution and were able to fit a Gamma distribution into the sliding rate of these in-silico experiments with mean $\mu = 11.70$ which is $\approx \lambda$, shape parameter $\alpha = 2.6 \pm 0.17$ and scale parameter $\beta = 4.67 \pm 0.35$. The distribution and curve fit is shown in Fig. 5.21.

II. SWI/SNF mediated chromatin remodeling manifests a bizarre energy efficiency.

We measured the C_{eff} from Eqn. 5.20 and observed that remolding process has a an average energy efficiency of $\%35 \pm 20$ which is quiet low. Fig. 5.22 shows the C_{eff} for experiments where $\lambda = 10.53$. We hypothesized three possibilities for such a weak energy utilization as follows:(i) the rest or part of ATPase energy coupled with remodeler conformation changes, (ii) energy dissipated from the system, (iii) unknown mechanism involved in the remodeling which our model does not account for that.



(a)



(b)

Figure 5.20. Time required to open a region of genome trapped in nucleosome: (a) unrestricted bulge size, (b) bulge size restricted to maximum of 250 base-pairs.

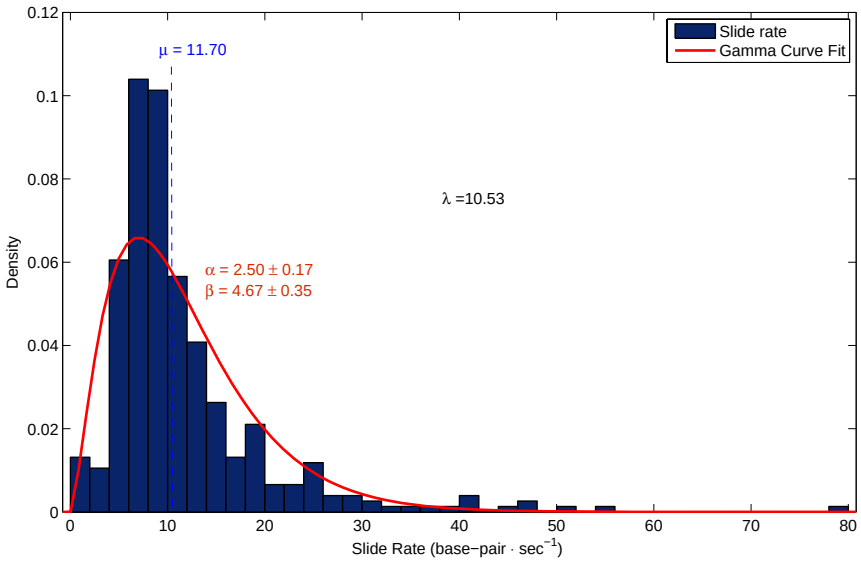


Figure 5.21. The density of different sliding rates when $\lambda = 10.53$ and the Gamma curve fitted to the distribution.

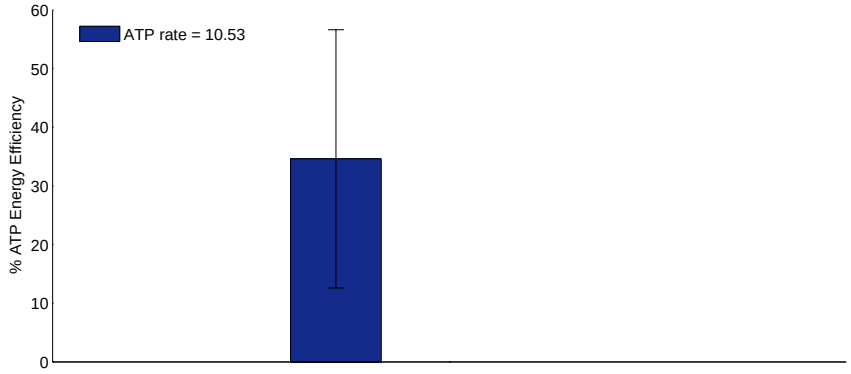


Figure 5.22. Energy efficiency of SWI/SNF mediated remodeling process.

5.5.2 Model limitations

There are certain limitations to our model we could be listed as follows:

- *Hinderance effect:* The hinderance effect of the neighboring nucleosome does not taken in to consideration.
- *Out-of-phase nucleosome effect:* An explicit assumption in our model is that the target nucleosome is not out-of-phase with itself (i.e. the minor grooves of the histone core stay at the right place). Impeding some bulge sizes or at certain nucleosomes positions (i.e. x_e) might push the nucleosome out-of-phase . This phenomena would effect the energy profile of the nucleosome for the next remodeling event. This fact is not captured by our model.
- *Elliptical bulge:* We have not considered bulges with elliptical geometry in our model, [74] because: Firstly, no experimental observation of any kind is reported on such bulge geometry. Secondly, the available approximation to the energy profile of does not have enough accuracy,[74].

5.6 Summary

In this chapter first we gave a brief introduction on SWI/SNF and ISWI remodeling complexes and their subunits. Then the *wave-rachet-wave* qualitative model was reviewed. Following the description of the qualitative model, we defined the state of nucleosome based on the nucleosome loop size and loop coordinates. We abstracted a state space of the remodeling process as random graph where each state (graph node) was represented by a quadruple and transition between two states (graph edge) was handled by a microevent. Nucleosome translocation *in-cis* is the consequence of execution of a sequence of valid microevents in a proper order. We applied three variants of available biophysical models to analyze the curvature of the DNA. An energy profile for the

DNA loops was adopted and coupled with the state diagram to propose the probability distribution function for temporal evolution of the remodeling process. The remodeling process is initiated by binding of remodeling complex to nucleosome and evolves by a loop formation event and a sequence of loop tracking event. The stochasticity of the model is on the loop size, tracking step and inter-arrival time of ATP molecules. We also showed that the sedimentation drag force has a negligible effect on the remodeling process. We conducted three in-silico experiments to mimic the wet lab reported results and successfully validated our simulation results. Finally we showed that translocation rate of nucleosome follows a gamma distribution. We demonstrated that based on our model, SWI/SNF induced chromatin remodeling has a low efficiency in utilizing ATP energy. At the end we browsed through the pros and cons of the proposed model.

CHAPTER 6

SETTING UP *In-silico* EXPERIMENT WITH *eukaSimBioSys*

6.1 Introduction

In previous chapters we covered our proposed methodology for in-silico simulation of a eukaryotic cell along with some of the physicochemical models apt to the same approach. We also described the architecture of *eukaSimBioSys* software tool which is developed based on our methodology for simulating biological systems. However, in order to bring the in-silico simulation of a eukaryotic cell from the workshop to the show case a detailed and clear picture of the underlying biological system should be provided. Also the are many fine grain pieces of assumptions, abstractions and models that needed to be put together with the body of main system to fill in the gaps. Some of these pieces of information could potentially be regarded as primitive model for a complex biological process which with future enhancement could let us have an even more comprehensive and accurate simulation as an integral part of the *eukaSimBioSys*.

In this chapter we intend to discuss the remaining concepts which are necessary for getting the *eukaSimBioSys* to run and mimic the experimental conditions. These concepts fall in divers domains including physiology of the insulin signal transduction network, estimating metabolic reaction flux, pre-mRNA splicing model, *eventology* of biological function, *eukaSimBioSys* design issues, etc. Although we promise to keep the flow of materials in the best consistent manner; however, due to the interdisciplinary nature of the work anticipating heterogeneous discussions is inventible.

Table 6.1. Substrates Contributions to the Heart Energy

Energy substrate	Energy share(%)	Preference situation	Major pathway controllers/points of regulation
Fatty acids	60–90	High energy needs	FAT/CD36; CPT1; Malonyl-CoA
Glucose	10–40	HF; ischemia; DM; cardio hypertrophy, Exercise	GLUT-1,4; PFK1; PDH
Ketone bodies	<1-40	Starvation	β-OH-Butyrate dehydrogenase
Lactate	<1–40	Exercise	LDH
Amino acids	<1-5	Protein breakdown	Aminotransferases
Ethanol	<1-5	Never	Acetyl-CoA synthase

6.2 Substrate metabolism in cardiac myocytes

Cardiac myocytes have diverse sources of energy substrates, however there preference to this metabolites is very skewed as pointed out in Table 6.1, data is taken from [108]. As its illustrated in the table fatty acid and glucose are the main energy substrates of cardiac tissues. Hence, in this work we only consider these two substrates for the following reasons: 1) their dominant share in energy provision for the cell, 2) to avoid excessive complexity in the underlying system. Following this section we briefly describe the physiology of fatty acid and glucose uptake, biochemistry of their metabolism and elaborate their abstraction in our in-silico experiments.

6.2.1 Glucose source

6.2.1.1 Glucose uptake

Glucose is taken into the cell mainly by glucose transporter 1 (Glut1) and glucose transporter 4 (Glut4) [109], the former is dominant during the fetal and its activity replaced gradually replaced by latter after birth. In an adult heart Glut1 normally participates in keeping the glucose at its basal level. Insulin signal stimulates the glucose uptake through a complex mechanism where the details of whole mechanism is not completely

understood. However it has been widely accepted that insulin promotes the Glut4 membrane transport through two parallel merging pathways. The tasks of two pathway which both originates from the *IRS* protein activation are complementary in mediating the glucose uptake. Phosphorylated $aPKC\lambda/\xi$ is a down stream product of a phosphorylation signaling cascade from the first pathway which enables the Glut4 vesicle transporters (GSV) to move to the vicinity of the membrane [110]. The trafficking of Glut4 between the cytoplasmic compartments (tubulovesicular and vesicular) and the plasma membrane forms a recycling type of process for this glucose transporters, more detail in the mechanism of this transport is available in [111, 112] . The second signal also originates from the same insulin receptor *INSR* activated protein, *INSR* phospho-activates the *APS* protein. Activation of *APS* initiates a sequence of activations and interactions that involve *Cbl,CAP, Crk, C3G, CIP4, Exo70, TC10* proteins [111]. Through a sequence of complex interactions the Glut4 which at the time is present in vicinity of the plasma membrane is first docked then tethered and ultimately fussed into the membrane [112].

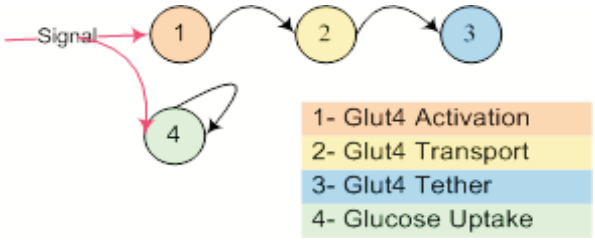


Figure 6.1. The event diagram for glucose uptake process.

6.2.1.2 In-silico abstraction of the process

In our in-silico experiment we abstract the outlined process into three events. These three event includes GSV activation event, GSV transport event, and Glut4 tether events. Fig. 6.1 shows the discrete even representation of the glucose uptake and Table 6.2 shows

Table 6.2. Glucose Uptake Event List

Event	Input	output	model
GSV activation	aPKC, Glut4.C	Glut4-a.C	Reaction
GLUT4 Transport	Glut4-a.C	Glut4-a-t.C	Diffusion
GLUT4 Tether	Glut4-a-t.C	Glut4	Reaction
Glucose Uptake	Glut4, Glucose	D-Glucose, Glut4.C	Glucose Uptake

these events along with their associated models. To account for the glucose uptake holding time, with the n number of Glut4 fussed into the membrane we make a subtle modification to model described in chapter 3 and use that to manipulate the respective holding time. The modification includes adding a capacity to each Glut4 which adopts the original model to work for group uptake. We assumed a capacity g for each Glut4, hence each Glut4 will be able to intake a group of g glucose molecules. Notice that this capacity will not effect rate of finding and binding to the membrane receptor for metabolite; however, it assumes that sugar molecules bind to the receptors in bundle of g molecules. Also since details of docking, tethering and fussing is unimportant to our investigations we consider all three as sequential microevents where their aggregate effect could be formulated as one reaction event.

6.2.2 Glycolysis

Upon the glucose uptake it will be converted to *D-Glucose-6Phosphate* through non-reversible reaction catalyzed by Hexokinase I and II. Each product molecule can enter one of the three pathways 1)Glycolysis, 2)Pentose-phosphate pathway or 3)Glycogenesis. In this work we focus on genesis of ATP from exogenous metabolites, hence we include the glycolysis pathway and first phase of fructose-mannose metabolism pathway but will not cover the Glycogenesis and Pentose Phosphate pathways. The reason for partial incorporation of fructose-mannose pathway is mainly because of an interesting

experimental data which is available that we will use in the next chapter to validate our in-silico experiment, and also to show the metabolite flux division across multiple metabolic pathways.

The product of hexokinase reaction increase the flux across glycolysis pathway which consists of 10 reactions as well as flux across *phosphofructo-2-kinase/fructose-2,6-biphosphatase* reaction from fructose-mannose pathway. We consider the aerobic respiration where along the glycolysis pathway, 6 ATP molecules is produced from each glucose molecule and the final product of the pathway which is *pyruvate*. *Pyruvate* could be converted to lactate by *lactate dehydrogenase(LDH)* (reversible reaction) or transported to the mitochondrial by pyruvate dehydrogenase (PDH) which is available on the mitochondrial matrix. In this work we do not incorporate the contribution of LDH in the metabolic network of cardiac myocyte. Fig. 6.2 shows the glycolysis pathway and a portion of fructose-mannose metabolism pathway. The oxidation of each pyruvate feeds one *AcetylCoA* to the tricarboxylic acid cycle (TCA) which will be used to produce GTP, $NADH$ and $FADH_2$. Each $NADH$ and $FADH_2$ leads to production of 3 and 2 ATP molecules in electron transport chain within the mitochondrial membrane, respectively.

6.2.2.1 In-silico abstraction of the Glycolysis reactions

As discussed in chapter 2 all metabolic reactions including Glycolysis reactions are handled in batch during the execution of the metabolic event whose details will be discussed in this chapter.

6.2.3 TCA cycle

TCA (Tricarboxylic acid cycle) produces citrate in one or two steps from *pyruvate* or in one step from *AcetyleCoA*. The oxidation of each pyruvate feeds one *AcetylCoA*

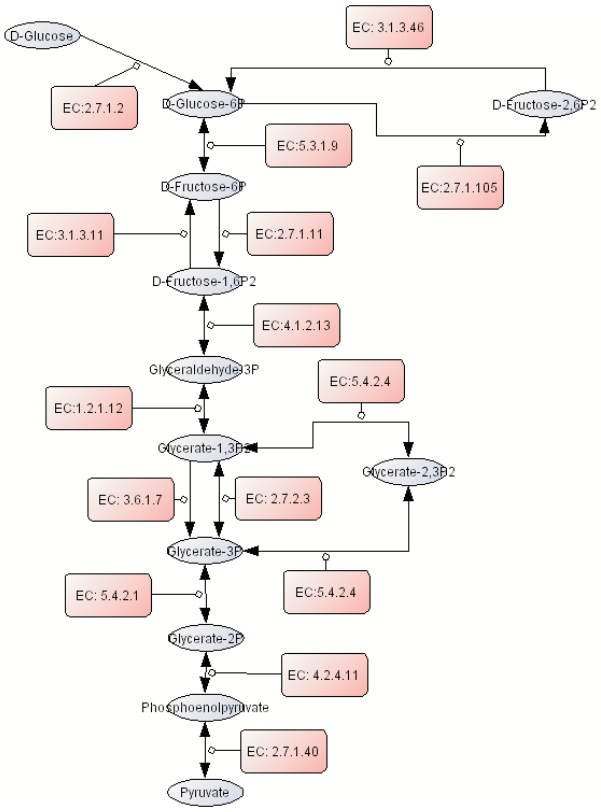


Figure 6.2. The Glycolysis I pathway along with EC:3.1.3.46 and EC:2.7.1.105 (*6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4*) reactions from fructose and mannose metabolism pathway.

to the which will be used to produce GTP, $NADH$ and $FADH_2$ where the latter two leads to production ATP molecules in electron transport chain within the mitochondrial membrane. Although *pyruvate* oxidation supply *AcetylCoA* to the TCA cycle, though the main source of *AcetylCoA* for this pathway is β-Fattyacid oxidation pathway. The net effect of ATP production for each glucose molecule could be up to a factor of 36 which come from 6 ATP molecules through aerobic oxidation, 2 molecules from anaerobic oxidation, and 15 ATPs per pyruvate molecules supplied to TCA cycle. We only consider the aerobic net share of ATP contribution which is 36 molecules. The TCA cycle pathway is depicted in Fig. 6.3.

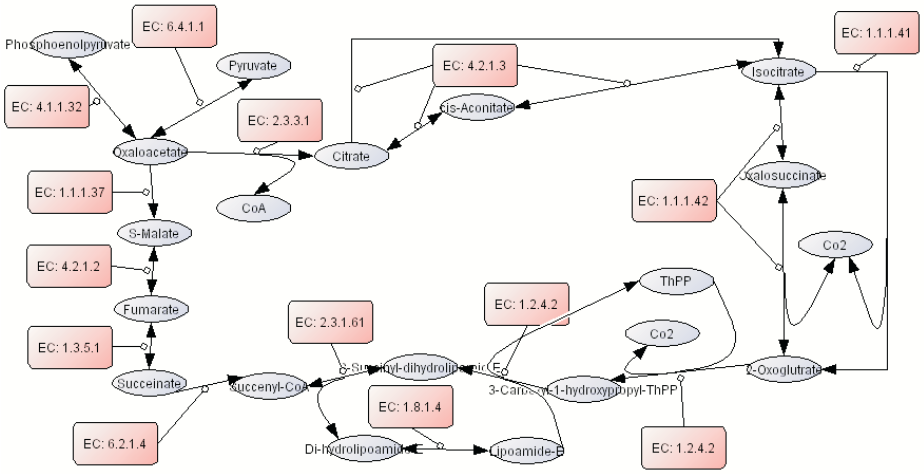


Figure 6.3. TCA (Tricarboxylic acid cycle) cycle pathway.

6.2.3.1 In-silico shift of pyruvate localization

PDH (*pyruvate dehydrogenase*) is available on the mitochondrial matrix, in our abstraction any substrate that is available on the mitochondrial matrix is considered to be a part of mitochondria. *Pyruvate* is the product of glycolysis which takes palace in cytosol, thus the shift in localization of pyruvate is a prerequisite to participate in the PDH reaction. We seek to make this adjustment with Metabolic Squad event that was described in chapter 2. Also the underlying reactions are incorporated as a part of the metabolic event.

6.2.4 Fatty acid source

There are two main sources of fatty acids for energetics of body cells, the exogenous fatty acids which are stored in from of fat in adipocyte (fat tissues) and endogenous fatty acid which comes from the intracellular storage of triglyceride. In this dissertation we only consider the latter source during our investigations.

6.2.4.1 Fatty acid uptake

The mechanism of Fatty Acid (FA) uptake in cardiac myocyte is not fully understood yet. Scientist postulate three main mechanism for fatty acid transport across cardiac myocyte membrane:

- Passive diffusion: diffusion through the membrane and subsequent deposition to the cytoplasm, this mechanism could be implemented by the Free Fatty Acids (FFA) which have very low abundance in blood ($< \%1$). Hence the contribution of this mechanism to the cytoplasmic FA is negligible [113].
- Albumin receptor mediated transport: in which the albumin-bound FA gets an anchor point to the cell membrane, subsequently the fatty acid could be released in the vicinity of on the membrane. This mechanism promotes a rich pool of free fatty acid across the membrane that could be eventually deposited to cytoplasm.
- FA mediated transport system: In this mechanism which still is not completely carried out of the theoretical workshop, three proteins interacts through a not well characterized mechanism. The three proteins include: Fatty Acid Binding Protein (FABP) which has membrane and cytoplasmic isoforms, Fatty Acid Transport Protein (FATP) and Fatty Acid translocase (FAT/36).

The plasma isoforms of FATBP and FAT/36 both advocate for passive diffusion by increasing the dissociation rate of albumin, and facilitated transport by interacting with FATP and importing the FA into cytoplasm [114, 115]. This system adjusts the rate of FA uptake with mitochondrial demand to avoid accumulation of FA in cytoplasm which could be hazardous for the cell. Once the FA entered the cytoplasm, it then binds to cytoplasmic isoform of FATBP and will be transported to the vicinity of the mitochondrial outer membrane. *Acyl-CoA* Synthase (ACS) converts the Long Chain Fatty Acid (LCFA)to *LC_acyl-CoA*. To participate in the β-Fatty Acid oxidation

LC_acyl-CoA should be transported into the mitochondria. To cross the impermeable mitochondria membrane the fatty acid transport pathway utilizes the *Carnitine palymitoyltransferase (CPT)* system. CPT composed of $L-$carnitine, *acylcarnitnie translocase* (ACT) and two transfer proteins i.e. *CPT1* and *CPT2* [116]. *Carnitine palymitoyltransferase 1* is the transmembrane protein that is localized on the outer membrane of mitochondria and delivers the *LC_acyl-CoA* to carnitine to form *LC_acylcarnitine*. ACT hands the *LC_acyl-CoA* over to CPT2 through the intermembrane space. The second transfer protein replaces the carnitine group of *LC_acylcarnitine* with *CoA* and releases the *LC_acyl-CoA* in the mitochondria to participate in the β-fatty acid oxidation pathway [116, 117]. *CPT1* is sensitive to *Malonyl_CoA* which is the product of *Acetly_CoA* carboxylation in cytoplasm this reaction is catalyzed by *Acetyl_CoA Carboxylase (ACC)*. Hence *Malonyl_CoA* is a negative regulator of β-fatty acid oxidation.

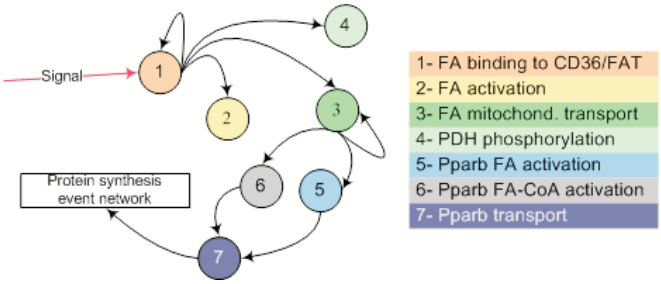


Figure 6.4. Fatty acid uptake event diagram.

6.2.4.2 In-silico abstraction of FA uptake and mitochondrial transport localization

To model the fatty acid uptake we map the physiology into the events listed in Table 6.3. These events form the event network for fatty acid uptake as demonstrated in Fig. 6.4. In the FA mitochondrial transport we have modeled the process by breaking

Table 6.3. Fatty Acid Uptake Event List

Event	Input	Output	Model
FA Uptake	LCFA CD36/FAT	LCFA-t CD36/FAT	Fatty Acid Uptake
FA Transport	LCFA-t Fatp.C	LCFA.C Fatp.C	Fast Reaction
FA Mitochondrial Transport	FA_carnitine Cpt1.C	FA.M Cpt1.C	Fast Reaction

the transport between the metabolic network and signaling network. More precisely the binding FA to CPT1 is handled by the signaling network as one bimolecular reaction, the shuttle of *LC_acylcarnitine* to the CPT2 is handled by a metabolic reaction which is catalyzed by *EC:2.3.1.21*. Release of LCFA in the mitochondria is handle by another signaling event. The reason for breaking the event between metabolic and signaling network originates from the set of metabolic reactions that we used in populating the data base. Our metabolic reaction set is consistent with **BiGG** database [118, 119]. Therefore, any metabolic reaction which is included in that database will be used by our simulation. This causes the fragmentation of the reactions that are not explicitly available in **BiGG** database but are among the precursor reactions to other metabolic reactions.

6.2.4.3 β-Fattyacid oxidation pathway

Fatty acids are carboxylic acids that are often categorized based on two properties : 1) Existence of double bonds in their carbon-backbone categorizes them to saturated and unsaturated fatty acids, 2) Length of their carbon back bone chain, most of natural fatty acids have even number of carbons in their backbone. Short Chain Fatty Acids (LCFA) have < 8 carbons, Medium Chain Fatty Acids (MCFA) have 8 to 14 carbons, Long Chain Fatty Acids (LCFA) have > 15 carbons. Table 6.4 lists some of the well-known fatty

acids along with their length and saturated property, detail discussion about the FA and their structures is beyond the scope of this dissertation and interested readers are referred to [120, 121] for further details. Our focus in this work is the metabolism of exogenous FA, hence we assume that no triglyceride is available in the cell and the entire fuel of FA oxidation pathway is provided from FA uptake. The FA residues are available in the form of *CoA thioesters* within the cell [122]. On the average the oxidation of FA comprises %60 to %90 of heart cell heart energy requirements during fasting, normal, and exercise conditions [123]. On each FA molecule oxidation onset these residues undergo up to seven serial oxidation reactions sets, based on their length during their post mitochondria import. For each set of reaction one H_2O molecule is consumed and a set of $NADH$, $FADH_2$, and $Acetyl_CoA$ is produced. As outlined earlier $NADH$ and $FADH_2$ will further be used to produce ATP in mitochondria and $Acetyl_CoA$ would fuel the TCA cycle. Each reaction set composed of following ordered reactions: FAD^+ dependent *acyl-CoA* dehydrogenase, a *2,3-enoyl-CoA* hydratase, NAD^+ dependent *3-hydroxyacyl-CoA* dehydrogenase ,and the *thiolase* cleavage reaction [123]. Fig. 6.5 depicts the β-Fatty acid oxidation pathway.

6.2.4.4 In-silico abstraction of the FA reactions

The same metabolic event that accounts for glycolytic and TCA reactions also includes the reactions for current pathway.

6.3 Metabolic Event

As pointed out in chapter 2 we issue the metabolic events in a periodic time order. The output of each Metabolic event in fact is a set of reaction fluxes that happen during the inter-arrival period of successive metabolic events.

Table 6.4. List of Well-known Fatty Acids

FA name	IUPAC name	Chemical structure	Abbr.	Saturated
Butyric	Butanoic acid	$CH_3(CH_2)_2COOH$	C4:0	Yes
Caproic	Hexanoic acid	$CH_3(CH_2)_4COOH$	C6:0	No
Caprylic	Octanoic acid	$CH_3(CH_2)_6COOH$	C8:0	Yes
Capric	Decanoic acid	$CH_3(CH_2)_8COOH$	C10:0	Yes
Lauric	Dodecanoic acid	$CH_3(CH_2)_{10}COOH$	C12:0	Yes
Myristic	Tetradecanoic acid	$CH_3(CH_2)_{12}COOH$	C14:0	Yes
Palmitic	Hexadecanoic acid	$CH_3(CH_2)_{14}COOH$	C16:0	Yes
Stearic	Octadecanoic acid	$CH_3(CH_2)_{16}COOH$	C18:0	Yes
Arachidic	Eicosanoic acid	$CH_3(CH_2)_{18}COOH$	C20:0	No
Behenic	Docosanoic acid	$CH_3(CH_2)_{20}COOH$	C22:0	No

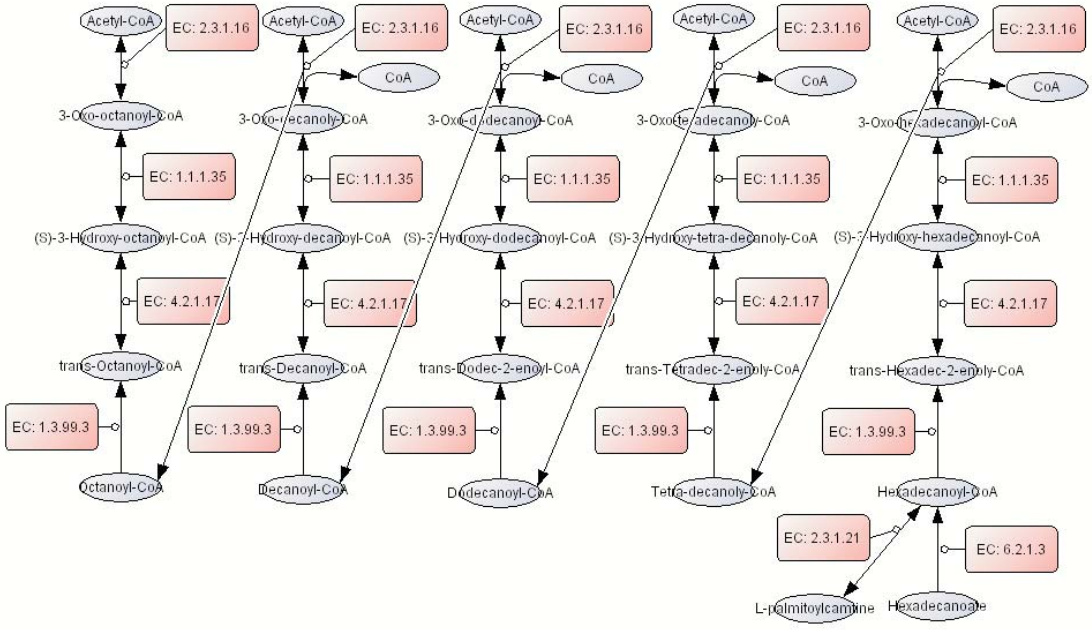


Figure 6.5. β-Fatty Acid oxidation pathway.

6.3.1 Metabolic events and metabolic reaction fluxes

We employ the a Flux Balance Analysis (FBA) [124, 125] approach to determine the flux across the metabolic reactions. From a flux we can determine the change in cardinality of specific a metabolite in the entire set of metabolic reactions over a time epoch, given the steady state condition. The essence of the FBA for a metabolic reaction founded on the assumption that the cell tends to maximize the biomass yield in the steady state condition. The emerging problem is then mapped into a linear optimization problem where the solution to this problem are optimum fluxes across sets of metabolic reactions given: the reactants, products, and enzymes concentrations. rFBA [126, 127, 128, 129] variants of FBA are enhanced version of the original FBA that takes into account the effect regulatory constrains on the metabolic fluxes. Furthermore, rFBA benefits from a dynamic search space on each instances of the search. The dynamics of the search space is driven by the gene expression profile of metabolic that possibly imposes new set constrains to the problem. One of the premium advantages of these approaches is their temporal efficiency in merging heterogenous networks i.e. metabolic and gene regulatory networks. By applying the rFBA we manipulate the metabolic flux in six steps. Before we get to these steps we introduce our we nomenclature for metabolic reaction R_i which is an ordered set of $(r_1, ..., r_m, p_1, ..., p_m, g_1..., g_z, K_{eq})$. In this ordered set $\{r_j\}$, $\{p_k\}$, and $\{g_t\}$ are sets of reactant consumed, products yielded and enzyme E_i proteins that catalyzes the reaction, respectively. The K_{eq} is the equilibrium constant of the reaction. Let

$$M = \forall R_i : \{r_j\} \bigcup \{p_k\}$$

and O_{E_i}be the turnover number of the enzyme E_i and $u(x)$ be the unit step function. Now we manipulate the flux for R_i during in time epoch t, which is denoted by $F^t_{R_i}$ from the steps below:

Step-I determine active status: $\forall\ R_i$ if $(E_i = min(|g_t|)) > 0$ then reaction is active.

Step-II reaction direction: if $\frac{\sum p_k}{\sum r_k} < K_{eq}$ then $d_i = 1$ and reaction has forward direction and $d_i = -1$ otherwise, where the reaction has a reverse direction.

Step-III build the STM $= R \times M$ which is the global stoichiometry matrix of all reactions such that if the j is substrate and $|j|$ is the stoichiometry of j in reaction i then: $STM_{i,j} = |j| \times d_i$ if j a reactant and $STM_{i,j} = -|j| \times d_i$ otherwise.

Step-IV manipulate the weight of reaction R_i for metabolite M_j as follows:

$$w_{i,j} = \frac{\lfloor O_{E_i} \times STM_{i,j} \times |M_j| \rfloor}{\sum_i (|E_i| \times O_{E_i} \times STM_{i,j} \times u(STM_{i,j}))} \times u(STM_{i,j}) \tag{6.1}$$

Notice that if the denominator in the above expression is zero then $w_{i,j} = 0$.

Step-V apply the metabolite constrains: $q_i = min(min\{|r_i|\}, min\{w_{i,j} \forall j\})$.

Step-VI determine the flux: $F^t_{R_i} = min(q_i, E_i \times O_{E_i}) \times t$.

In the last step above, t is the inter-arrival time of the metabolite events. The estimated flux is inversely proportional to t, where as efficiency of the simulation has direct proportionality to that. In this work we have 109 metabolic reactions which are imported from BiGG database. Each of these reactions is identified by a unique reaction ID borrowed from the original record in the BiGG. List of these reaction IDs along with their associated enzymes is given in Table 6.5.

6.4 A computational model for pre-mRNA splicing

To chicaneries the dynamics of pre-mRNA splicing which is an integral step of protein synthesis process in eukaryotes we have abstracted the process into a stochastic discrete event based model.

Table 6.5. List of Metabolic Reactions

	Reaction ID	Enzyme		Reaction ID	Enzyme		Reaction ID	Enzyme
1	R_PCm	6.4.1.1	38	R_FAOXC200180m	1.3.99.3	75	R_ACONTm	4.2.1.3
2	R_SUCOAS1m	6.2.1.4	39	R_FAOXC11	1.3.99.3	76	R_ACONT	4.2.1.3
3	R_LDH_L	1.1.1.27	40	R_GCC2cm	1.8.1.4	77	R_PGI	5.3.1.9
4	R_MDH	1.1.1.37	41	R_C226CPT1	2.3.1.21	78	R_PGM	5.4.2.1
5	R_ME1m	1.1.1.38	42	R_DMHPTCRNCPT1	2.3.1.21	79	R_PGMT	5.4.2.2
6	R_ME2m	1.1.1.40	43	R_DLNLCGCPT1	2.3.1.21	80	R_GlyMtas	5.4.2.4
7	R_ICDHxm	1.1.1.41	44	R_DCSPTN1CPT1	2.3.1.21	81	R_FACOAL246_1	6.2.1.3
8	R_ICDHyrm	1.1.1.42	45	R_C204CPT1	2.3.1.21	82	R_FACOAL191	6.2.1.3
9	R_ICDHy	1.1.1.42	46	R_C180CPT1	2.3.1.21	83	R_FACOAL226	6.2.1.3
10	R_GAPD	1.2.1.12	47	R_HPDCACRNCPT1	2.3.1.21	84	R_FACOAL224	6.2.1.3
11	R_PDHm	1.2.4.1	48	R_LNLNCGCPT1	2.3.1.21	85	R_FACOAL140i	6.2.1.3
12	R_AKGDm	1.2.4.2	49	R_C161CPT22	2.3.1.21	86	R_FACOAL170	6.2.1.3
13	R_SUCD1m	1.3.5.1	50	R_LNLCCPT1	2.3.1.21	87	R_FACOAL1821	6.2.1.3
14	R_FAOXC1811602m	1.3.99.3	51	R_LNELDCCPT1	2.3.1.21	88	R_FACOAL1822	6.2.1.3
15	R_FAOXC18480m	1.3.99.3	52	R_C161CPT12	2.3.1.21	89	R_FACOAL240	6.2.1.3
16	R_FAOXC140	1.3.99.3	53	R_TTDCPT1	2.3.1.21	90	R_FACOAL226i	6.2.1.3
17	R_FAOXC204	1.3.99.3	54	R_TMNDNCCPT1	2.3.1.21	91	R_FACOAL1832	6.2.1.3
18	R_FAOXC1811601m	1.3.99.3	55	R_ADRNCPT1	2.3.1.21	92	R_FACOAL184	6.2.1.3
19	R_FAOXC226205m	1.3.99.3	56	R_VACCCPT1	2.3.1.21	93	R_FACOAL1831	6.2.1.3
20	R_FAOXC16180m	1.3.99.3	57	R_STRDNCCPT1	2.3.1.21	94	R_FACOAL241	6.2.1.3
21	R_FAOXC16080m	1.3.99.3	58	R_C161CPT1	2.3.1.21	95	R_FACOAL205	6.2.1.3
22	R_FAOXC80	1.3.99.3	59	R_ARACHCPT1	2.3.1.21	96	R_FACOAL206	6.2.1.3
23	R_FAOXC2242046m	1.3.99.3	60	R_PTDCACRNCPT1	2.3.1.21	97	R_FACOAL200	6.2.1.3
24	R_FAOXC2251836m	1.3.99.3	61	R_CSm	2.3.3.1	98	R_FACOAL203	6.2.1.3
25	R_FAOXC180	1.3.99.3	62	R_ACITL	2.3.3.8	99	R_FACOAL204	6.2.1.3
26	R_FAOXC182806m	1.3.99.3	63	R_HEX1	2.7.1.1	81	R_FACOAL161	6.2.1.3
27	R_FAOXC183806m	1.3.99.3	64	R_FBPK26	2.7.1.105	101	R_FACOAL180i	6.2.1.3
28	R_FAOXC2051843m	1.3.99.3	65	R_PFK	2.7.1.11	102	R_FACOAL260	6.2.1.3
29	R_FAOXC161802m	1.3.99.3	66	R_PYK	2.7.1.40	103	R_FACOAL1812	6.2.1.3
30	R_FAOXC18280m	1.3.99.3	67	R_PGK	2.7.2.3	104	R_FACOAL160i	6.2.1.3
31	R_FAOXC2031836m	1.3.99.3	68	R_DPGase	3.1.3.13	105	R_FACOAL1813	6.2.1.3
32	R_FAOXC2252053m	1.3.99.3	69	R_FBPP26	3.1.3.46	106	R_FACOAL2251	6.2.1.3
33	R_FAOXC183803m	1.3.99.3	70	R_FBP	3.1.3.11	107	R_FACOAL2252	6.2.1.3
34	R_FAOXC160	1.3.99.3	71	R_PEPCK	4.1.1.32	108	R_SUCOASm	6.2.1.5
35	R_FAOXC170m	1.3.99.3	72	R_FBA	4.1.2.13	109	R_ACCOAC	6.4.1.2
36	R_FAOXC226	1.3.99.3	73	R_ENO	4.2.1.11			
37	R_FAOXC150m	1.3.99.3	74	R_FUMm	4.2.1.2			

6.4.1 An overview of splicing

In eukaryotic the central dogma includes an additional preprocessing step prior to have a mature transcript of the gene. This preprocessing concatenates the exons which are the protein coding units of a gene, by removing the interleaved introns that do not convey any code for protein. The splicing pathway that converts the pre-mRNA to mRNA was first revealed in mid 80s [130, 131, 132]. The protein machinery that impalements the splicing is referred to as *spliceosome* and in [133] researchers have shown that > 300 proteins intentionally involve in the splicing process. The splicing can happen cortranscriptionally [134] and based on the class of introns can implement different mechanism. There are several classes of introns including: 1)tRNA introns [135], 2)group I and II introns that could be found in protozoan rRNAs [136], mitochondria and chloroplasts of fungi and plants,and 3) spliceosomal introns that form the largest class and exclusively available across eukaryotic genomes. In this work our attention is focused on the last group and detail discussion regarding the rest is beyond the scope of this dissertation.

In eukaryotes splicing has two variants for intron excision from a premature transcript, known as *constitutive splicing* and *alternative splicing* both of which occur *in-cis* [137]. Former is the primitive mechanism where the objective is accomplished by removing the intron sequence flanking an exon and pasting the two successive exons. The latter augments the coding capacity of a single gene by reordering and/or cleaving the original exons [138, 137]. In passed decade extensive amount of research have been devoted to proteins that catalyze the splicing or signaling pathways that affect the alternative splicing pattern, etc. whose detail discussion is beyond the scope of our work an could be studied elsewhere. For our simulation purpose we seek to find a computationally efficient model that can estimate the temporal dynamics of splicing. Therefore, at this time we prefer to draw our attention to the constitutive splicing and leave the alternative variant for the future works.

6.4.2 Mechanism of constitutive splicing

The core spliceosome composed of five small nuclear ribonucleoprotein particles (snRNPs), [137], listed here: U1, U2, U4, U5 and U6 . The spliceosome assembly happens in an orchestrated fashion which leads a bistep splicing pathway [139] as follows:

Step 1: Cleavage of pre-mRNA at the $\acute{5}$ splice site which forms the two splicing intermediates: the linear first exon mRNA and intron-second exon lariat. Lariat configuration forms the $\acute{2} \rightarrow \acute{5}$ phosphodiester bond between the *Guanosine* at the $\acute{5}$ splice site and the *T-hydroxyol* of an *Adenosine* in the vicinity of the intron $\acute{5}$ end.

Step 2: Ligation of the two exons by cleavage of the $\acute{3}$ splicing site which results a uninterrupted exons and spliced intron in a lariat configuration. The evolution of these bistep pathway subjects to organized assembly/disassebly of the spliceosome components in a timely manner which forms the foundation of the splicing process.

6.4.2.1 Spliceosome assembly pathway

The objective of spliceosome assembly pathway is to make the mRNA branching possible by bringing the $\acute{5}$ branch site to the proximity of branching point. For this purpose the small nuclear ribonucleoproteins (snRNP) and pre-mRNA forms the backbone of the process. These snRNPs interact with one another and with RNA and under go conformational change to develop the process. Other non-snRNP factors (e.g. auxiliary factor U2AF) also participate in these delicate interactions and catalyze the process. We do not bring them the picture since they have a marginal role in over all evolution of process. The spliceosome assembly pathway could be segregated into four stages as follow [139]:

Stage-I U1 snRNPb binds to the 5 splice site referred to as commitment complex formation. This is the only none ATP dependent stage in splicing.

Stage-II Splicing complex A: pre-mature spliceosome recruits U2 snRNP that binds to the branching point which most often is about 18-38 nucleotides upstream the $\acute{3}$ splice site [140]. At this point the commitment of complex to splicing becomes irreversible where this is the first ATP dependent stage in splicing. It should be mentioned that binding of U2AF stabilizes the U2 binding. Also, formation of U1 pre-spliceosome is required for U2 biding since it helps to juxtapose the $\acute{3}$ and $\acute{5}$ splicing sites.

Stage-III Splicing complex B: formation of mature spliceosome by joining of U4/U5/U6 multi-snRNP to the complex A. The U4/U5/U6 is formed prior to the binding to splicing complex A.

Stage-IV Splicing complex C: U4/U6 undergo a conformational change prior to cleavage at splice site. This conformation change leads to instability and loss of U4 snRNP which forms the splicing complex C.

Following the assembly of spliceosome complex C the splicing starts using the ATPase energy to form the branch ($\acute{2} \rightarrow \acute{5}$ phosphodiester bond). Subsequently the $\acute{3}$ site is cleaved meanwhile the two exons legation occurs. The spliceosome that remains bound the intron facilitates the intron degradation and itself gets recycled for the further splicing. Fig. 6.6 depicts the spliceosome assembly stages and intron excision process.

6.4.2.2 Event based abstraction of splicing

In order to capture the temporal dynamics of splicing event, we abstracted the spliceosome four stage assembly pathway and its post assembly interactions (branching and ligations) as a sequence of microevents. Due to the uncertainty that exists in the cell environment [18] these microevents render an inherent stochastic behavior in time an space. We try to provide parametric mathematical expressions that can estimate the temporal stochasticity of each microevent. The convolution of these stochastic ex-

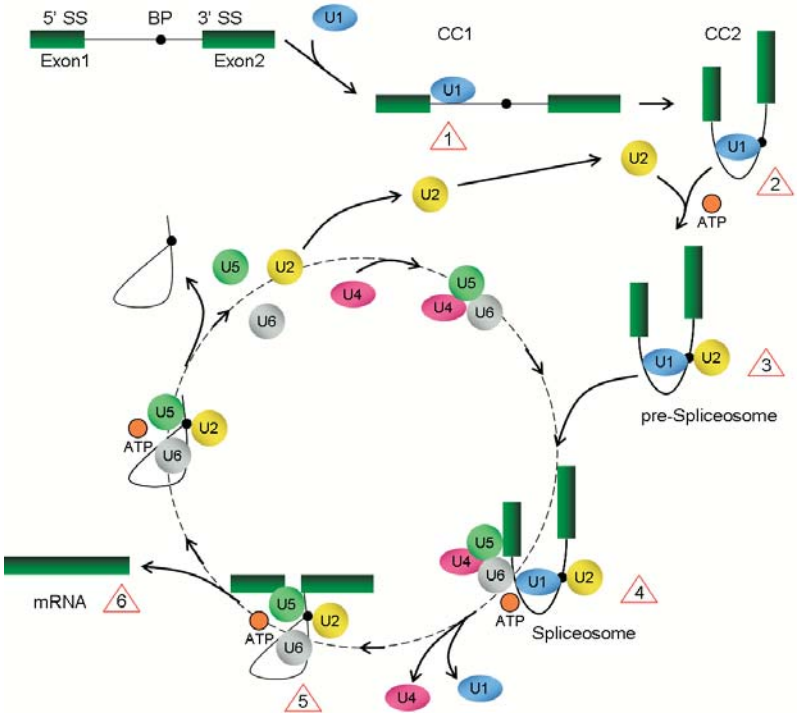


Figure 6.6. Constitutive Splicing: the branch point(BP) is in the vicinity of 3́ splicing site(SS), binding U1 forms the commitment complex 1 (CC1), joining the U2 to the commitment complex 2 (CC2) forms the pre-spliceosome, spliceosome becomes mature once the U4/U5/U6 complex binds to the pre-spliceosome. Each red triangle distinguishes an identified microevent.

pressions can represent an estimation of one stochastic splicing event. If we number the microevent based on Fig. 6.6, hence we will have 3 protein-RNA binding miscreants (1,3,4), one protein conformation change (5), one loop formation microevent (2), and the ligation microevent (6). As it is outlined earlier all the microevent except the first protein-RNA binding couple ATP hydrolysis energy for their progress. Therefore, for all but one of the microevents the ATP inter-arrival time will also be an effective factor. We make the following assumption prior to modeling the microevents: ATP molecules have poisson arrival process [105], the tri-snRNP complex is always available in the reaction environment, any protein requires one ATP molecule for its conformation change.

protein-RNA binding micro-event: To estimate the time for this microevent we follow the

approach proposed in [35] where they showed that temporal dynamics for protein-DNA binding follows an exponential distribution with respect to the number of binding sites. We use the same approach to find rate for protein-RNA binding as follows: In [35] they showed that probability of an abundant protein to binds to it's functional site s on DNA ($p_D(B, s)$). Considering the case where the set of binding sites has a unit cardinality ($|s| = 1$), then the modified $p_D(B, s)$ could be written as:

$$p_D(B,s) = \frac{4\Delta t(r_D + r_p)(n_s \cdot bp + 2B \cdot bp + 2r_p)(n_s + 2B)aa_m}{\pi V \cdot aa_p \cdot n_D}\sqrt{\frac{8k_bT}{\pi m}}e^{-\frac{E_{act}}{k_BT}} \tag{6.2}$$

in Eqn. 6.2 B is the number of nucleotides that protein slides over on RNA surface prior to binding to a functional site and $m = \frac{m_p \cdot m_R}{m_p + m_R}$ is the reduced mass. List of other parameters could be found in Table 6.6. If $|s| > 1$ then the probability to access site j is:

$$p_D(B, s_j) = p_D(B, s_j) \prod_{i=1, i\neq j} (1 - p_i) \tag{6.3}$$

$$p_D = p_D(B, s_j) + p(0, s_j) - 2p_D(0, s_j)p_D(B, s_j) \tag{6.4}$$

Eqn. 6.3 only counts for the case when 1D diffusion on RNA surface precedes binding. Eqn. 6.4 averages over both cases, binding with/without sliding. Now we can write the average binding time as follows:

$$\frac{1}{\gamma} = \frac{p_{nr}\Delta t + (1 - p_{nr})(\Delta t + \tau_{1d})}{p_D} \tag{6.5}$$

p_{nr} is the probability for no random-walk on the RNA surface by protein, and τ_{1d} is the 1-D diffusion time (due to the random-walk) of the biding protein on the RNA surface, this diffusion time has been estimated for DNA in [141], we use the same approximation for RNA since the force fields remains roughly the same. Table 6.6 lists the values for

Table 6.6. List of Parameters for Protein-RNA Binding

Parameter	Description	Value
λ	ATP arrival Rate	
V	Volume	$4 \times 10^{-18}\ m^3$
n_R	bps Length of RNA	
n_m	bps Length of sn-RNP binding site	50 base-pairs
aa_p	Length of protein amino acid chain	339 base-pairs
aa_m	Length of protein motif	154 base-pairs
r_p	Average radius of the protein	$5^{-9}\ m$
r_D	Radius of DNA	$10^{-9}\ m$
m_p	RNA weight	$n_D\ \times 330\ Da$
m_D	Protein weight	37698 Da
Δt	Inter-collision time	$10^{-8}\ sec$
E_{act}	Activation energy	$1\ k_BT$
bp	RNA nucleotide length	$0.34 \times 10^{-9}\ m$
k_bT	Boltzmann constant	$1.3806503 \times 10^{-23}\ kg\ m^{-2}\ s^{-2}\ K^{-1}$
ΔATP	ATP hydrolysis energy	$29\ kcal/mol$

parameters we used in our simulation to estimates Protein-RNA binding time.

Branching microevent: To estimate the required energy for branching we apply the same approach as in chapter 5 for DNA bulge formation. Considering the intron with n nucleotide length, form [103] the bending rigidity of the intron B_n could be written as:

$$E_b = \frac{\sum^{n-2_i} e_i^{-1}}{n-2} \Rightarrow B_n = \frac{\pi}{16} E_b R_{RNA}^4 \times 10^{30} \tag{6.6}$$

in the above expressions e_i is Young modulus of the nucleotide triplet from the Table 1 in [103]. If the resultant branch has a radius $r = \frac{n0.34\times10^{-9}}{\pi}$, then we can approximate the required energy from:

$$U = B_n \frac{\alpha}{r} \tag{6.7}$$

Subsequently required number of ATP molecule to supply the energy for this branch will be:

$$a = \lceil \frac{U}{\Delta_{ATP}} \rceil \tag{6.8}$$

In Eqn. $\Delta_{ATP} = 11.558 k_B T$. As mentioned earlier only the first stage of splicing is ATP independent, hence considering $a + 4$ will be the lower bound for the required number ATP during the spicing process. Considering our assumption for the ATP arrival process with mean rate λ and Eqns. (6.5,6.4.2.2) the probability distribution function for splicing an intron could be written as:

$$p_{spl}(t) = \frac{\lambda(\lambda t)^{a+3+b}}{a+3+b} e^{-\lambda t} \times e^{-\gamma t} \tag{6.9}$$

In Eqn. 6.9 b is the tuning parameter for that counts for ATPs above the lower bound required for one instance of splicing. Now if the pre-mRNA consists of m introns, and $l \leq m$ splicing happen simultaneously then then $\lceil m/l \rceil$ fold convolution of Eqn. 6.9 will be form the probability distribution function of the one pre-mRNA splicing event.

Since the solution to the probability distribution function of the splicing event becomes intractable, therefore we use the discrete event simulation to calibrate the parameters such the rate of pre-mRNA event comes to a close vicinity of the experimental average ($p < 0.01$). Although the reported experimental time for pre-mRNA ranges from several minutes to a couple of hours [142, 143, 144], yet in-vivo observation shows for mammalian that have an average intern length $\sim$ 3kilo-bases splicing takes place in the time order of $<$ 10 minutes [145]. For the in-silico experiments we have calibrated our model to have an average splicing time of 5 minutes.

6.5 Event based model for protein synthesis

Protein synthesis is the core process of life which involves a very complex and not completely known regulatory mechanism. Although, we might still need years to come for a comprehensive and detailed quantitative model of protein synthesis; however, with the state of art of event based abstraction and the fundamentals of *central dogma* as will be demonstrated shortly we are able to model this process to serve our purpose.

6.5.1 An overview to qualitative model of the process

Protein synthesis in eukaryotes compromise an orchestrated sequence of events including: chromatin remodeling, gene transcription, pre-mRNA splicing, mRNA nuclear transport and mRNA translation. Any one of these events involves complex evolution and regulatory mechanisms and undoubtedly, they have been topics for excessive research over passed quarter century and considerable amount of online and published resources are available on them. Comprehensive discussion over these topics will be beyond scope and capacity of this dissertation, hence we briefly browse through the major concepts that will contribute to our modeling effort. Transcription and translation in eukaryotes involve sophisticated evolution and regulation mechanisms, and much of their details yet not understood properly. General mechanism of transcription and translation is discussed in [146] and elsewhere. In [147] general concepts involved in mammalian gene transcription is detailed and a qualitative model for their assembly is proposed. Transcription and RNA II-TFIIB are structurally analyzed in [148, 149] and mechanism of RNA II elongation is discussed in [150]. Binding of TATA Box Proteins (TBP) is essential for gene expression, in [151] regulation of gene expression by TBP is elucidated. In a recent work kinetic analysis of gene transcription is provided in [152]. As outlined earlier following the gene expression the pre-mRNA will be spliced to generate messenger mRNA. Each mRNA should be transported to cytoplasm and translated by the ribosomal proteins (tRNA) to

give birth to protein that it encodes. The process of transporting the mRNA to cytoplasm is referred to as nuclear transport which it self has divers and complex mechanisms [153, 154, 155]. Also the kinetics of mRNA nuclear transport is studied in [145, 144]. Following the export of mRNA to the cytoplasm ribosomal protein (tRNA) will translate the codons in mRNA to the proper amino acids and click the protein synthesis. The mechanism of translation initiation is given in [156, 157] while the molecular mechanism of translation is described in [158]. In order to have a dependable translation, authors in [159] discussed that eukaryotic translation machinery applies a surveillance mechanism to ensure the correct reading and translating of mRNA codons. Also it has been shown that protein synthesis is non-linear and has a bursty dynamic [160], furthermore protein synthesis could also be regulated at the ribosome level [161, 162].

6.5.2 Eventology of the process

Nevertheless, the details of gene expression and its regulatory mechanism is far more complicated than described above, we have abstracted the protein synthesis process as a network of events. These events could be categorized into two classes of *explicit* and *implicit* events based on the mechanism of their initiation. Former, includes those events whose trigger is explicitly indicated in the qualitative models such as transcription event, splicing event, etc. The latter class includes those events that will be executed although they are not explicitly included in the qualitative models, examples of those include: protein decay, mRNA decay, transcription termination, etc. The protein synthesis is the product of collaboration between the transcription regulatory and signaling networks. Therefore, to inspire a clear event view (*'eventology'*) of the whole process we seek to extrapolate into a step of signaling prior to the gene expression. Suppose an external signal in its downstream activates the transcription factor α, upon activation α is transported into the nucleus. The effect of α on gene expression is interpreted by the

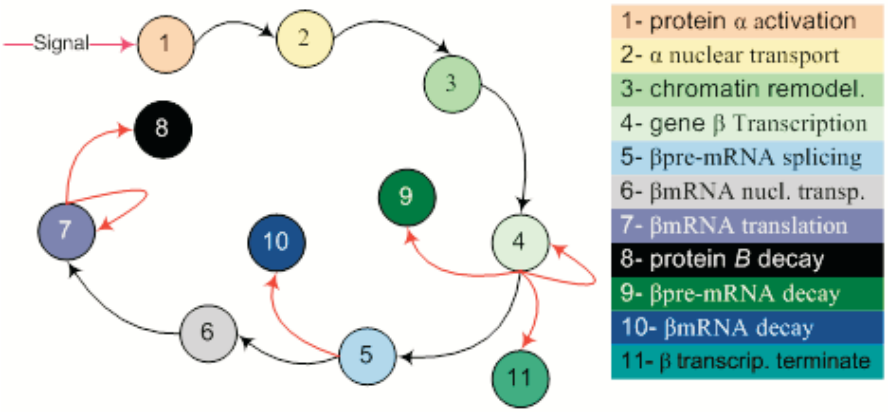


Figure 6.7. Event diagram of protein synthesis in eukaryotes.

gene regulatory network, our abstracted mechanism of gene regulation will be discussed later in this chapter. Back to our model, further assume that as result of transcription regulation gene β will be expressed. Ultimately $\beta - mRNA$ will be transported to cytoplasm and translated to protein B. Fig. 6.7 shows the event diagram of this model where the red arrows point to the *implicit events* and black arrows point to the *explicit events*. As observed in the diagram an interesting feature of the events is than an event could belong to both categories and the only segregating check point is the mechanism for triggering an event with respect to qualitative model.

6.5.2.1 Temporal dynamics of the events

The dynamics of the events are captured by their associated stochastic discrete event based models. For chromatin remodeling event we applied the model outlined in chapter 5 with the nucleosome occupancy of 0.3 for the promoter region. For the transcription event we used the model proposed in [36]. This model uses a birth and death markov chain to determine the rate of the transcript production. They have modeled the process based on number of RNA PII that binds to the gene and the elongation rate of RNA PII. We have adapted and calibrated the model to become consistent for

eukaryotic based on the parameters given in [144, 143] (e.g. basal RNA PII elongation rate (40 bases/s), etc.). The splicing event rate is estimated based on the model discussed earlier in this chapter. For the pre-mRNA and mRNA decay events we applied the exponential decay processes with a rates according to to half life of these species reported in [163, 164, 160]. We used a simple stochastic diffusion model proposed in [37] to estimate temporal dynamics of mRNA nuclear transport based on kinetics reported in [155, 165]. For estimation of translation event time, we applied the same markov model in [36] that was used for transcription and calibrate the parameters based on experimental data in [143, 145, 160]. The protein decay event is an exponential decay event with rate reported in [166]. The transcription termination event has a constant time which could be used as a further calibrating point in the simulator.

6.6 Eventology of the insulin signaling in the cardiac myocyte

6.6.1 An overview of insulin signal

Heart pumps the blood at basal rate of 5-6 liter/min and can rise up to 20-30 liter/min during the workload and exercise [167]. Such pumping capacity requires uninterrupted contractile activity and hence, a perpetual level of energy demand. The fulfilment of such energy demand is almost entirely relayed on substrate oxidative pathways in mitochondria and cytoplasm discussed earlier in this chapter. Therefore, cardiac myocytes that are the key tissues of heart contractile activity should have a flexibility in their fuel selection. Although metabolic flux modulation could be regulated at many levels, two of the promising flux modulation in cardiac myocytes are through the control of metabolite uptake and gene expression level [168]. Insulin which is an essential peptide hormone of endocrine system that secretes from β-cells in pancreas is predominantly involved in the fuel selection at both levels. Although the propagation of the insulin signal

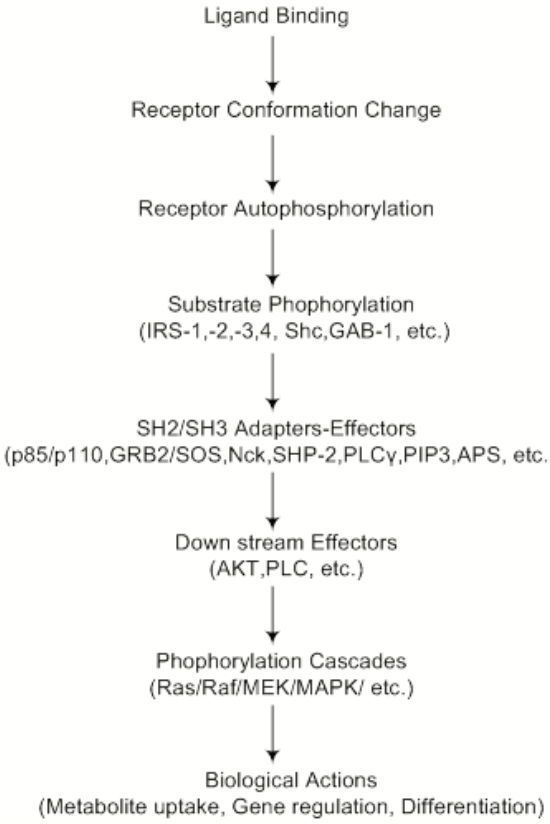


Figure 6.8. The hierarchy of insulin signal propagation in the cell.

within the cell will influence divers cellular functions such as mitogenic, cell growth, etc. however, in this dissertation we focus on the signaling information that culminates on the two modulatory effects. The effect of insulin on the metabolites uptake was discussed earlier therefore, here we draw our attention to the regulation of the metabolic enzyme synthesis.

The insulin signal is sensed by binding the insulin to insulin receptors located on the membrane of cardiac myocytes and belong to the family of ligand-activated tyrosine kinase (RTK) receptors [169]. The information of the insulin signal is propagated within the cell through a non-linear signaling network. The hierarchy of insulin signaling network is depicted in Fig. 6.8

The description regarding the details of insulin signal transduction hierarchy and the function of individual substrate components is available in [170, 169]. Although the effect of an external signal on the fate of a cell could be best analyzed when it is exposed to complex sets of signals in the context of a cell; however, the flip side of the coin in such analysis is to cope with emerging complexity of system. The inherent robustness is the *de-facto* rule of survival for biological systems subjected to natural section. Therefore, most of these systems are robust to the large set of stresses and demonstrate the butterfly effect to substantially smaller sets. Setting this fact *vis-a-vis* the complexity of system we can reduce the complexity by two strategies: i) either by eliminating the components or aggregating their detail to a higher level where it is proven or speculated to have lesser impact on the objective system, ii) exclude a subset of the system from the analysis with the assumption that the rest of the system is in the equilibrium interaction with that subset. With this strategy we have abstracted the insulin signal transduction hierarchy from excessive details and included those components where a consensus exists on their impact on the cell metabolism [167, 168, 171, 108].

The insulin signal transduction network (STN) that has the above property could be found in *KEGG pathway database* [6]. We imported their STN and modified the original version based on data published elsewhere to include some of missing components that were necessary for our work as well as excluded the excessive details. The insulin STN that we stored in the database and is used for our simulation is depicted in Fig. 6.9. There are situations where we need to modify the layout of this STN based on some hypothesis or special experimental condition that will be discussed in place.

6.6.2 Event diagram

The event diagram of Fig. 6.9 is depicted on Fig. 6.10, The color code is used to represent the events with similar physicochemical (e.g phosphorylation, activation,

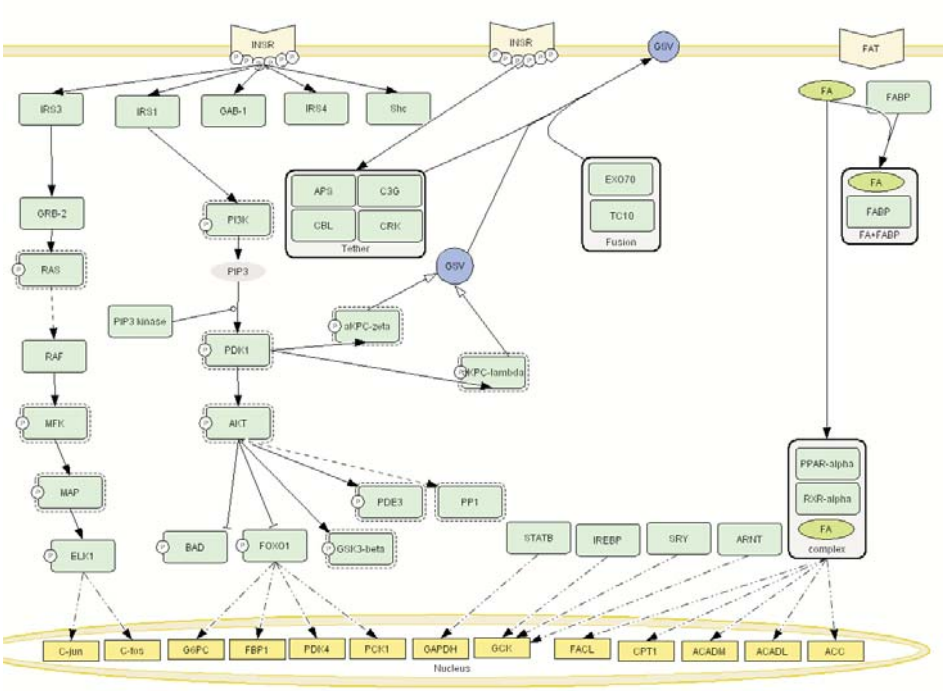


Figure 6.9. The insulin signal transduction network diagram. The information from exogenous signal propagates through a non-liner network.

transport) class. The physicochemical class of the event were explored from the wet lab data available in the literature. An event whose physicochemical class was unidentified is assigned to a biochemical reaction class. We should bring this point into attention of the readers that, since signal transduction and transcription regulatory networks are interrelated we have included a subset TRN that is affected by the insulin in the event diagram. Also Fig. 6.10 does not include the events that pertain to the fatty acid signal which is partially depicted in right corner of Fig. 6.9.

Applying graph therapy analysis on the STN event diagram along with the gene regulatory network (GRN) of a biological system could possibly provide a new realm of results on properties of the system, however we postpone these type analysis for our future works.

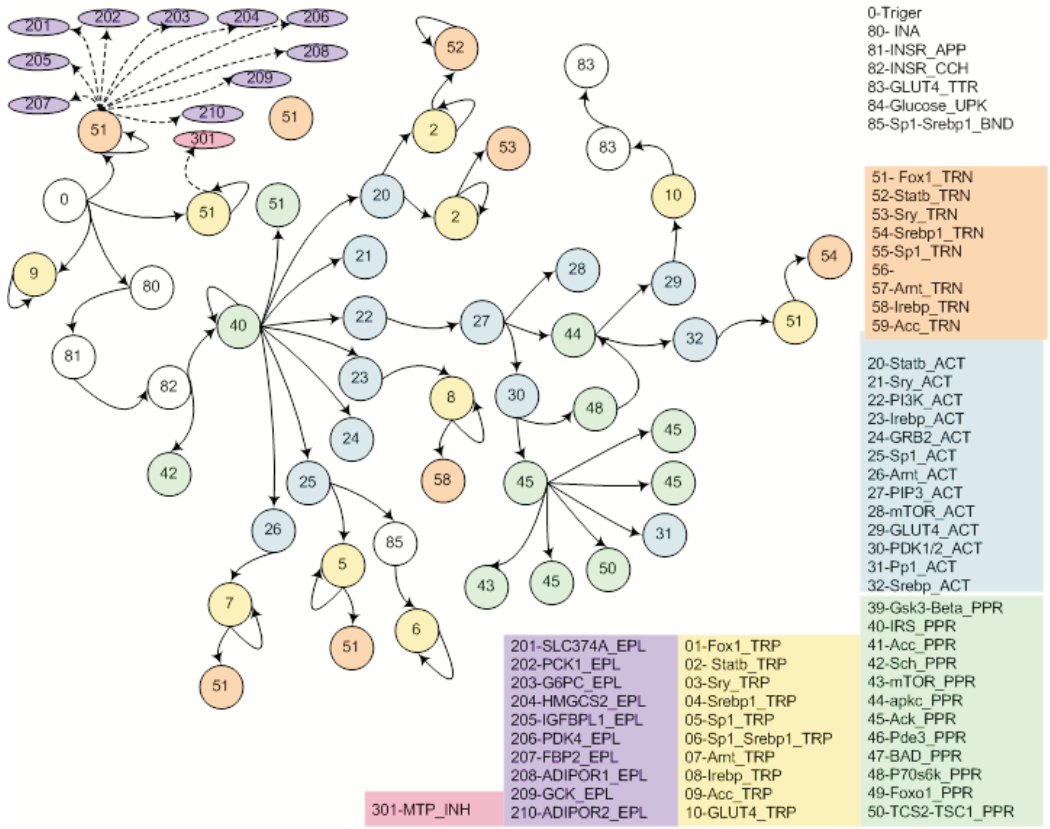


Figure 6.10. The event diagram for insulin signal transduction network in Fig. 6.9. Events with purple color code belong to TRN.

6.7 Transcription regulatory network and in-silico regulatory model

More than 150 genes have been identified that are positively or negatively regulated by the insulin [172]. Amongst genes affected by insulin < 50 genes have been reported, as myocardial genes [167]. Since we are working with two signals i.e. insulin and fatty acid, therefore those gene that are regulated by either of these is keen to our interest in this work.

6.7.1 Regulated genes

Insulin affected genes: We sought to collect as many genes that has been reported and is regulated by insulin signal in the heart muscle cell. Our main sources for tran-

scription regulatory data gathering include but not limited to [173, 13, 170, 172] articles and *KEGG* [6], *HumanCyc* [14], and *Reactome* [7] databases.

Fatty acid affected genes: The regulation of genes by fatty acid is mostly through the activation of peroxisome proliferator-activate receptors (PPAR) [108]. These nuclear receptors are available in three forms: $PPAR\alpha, PPAR\gamma$ and $PPAR\delta$. Fatty acid binds to all three type as ligand and activates them, the activated PPAR then forms an obligate heterodimer with retroid X factor receptor (RXR). The PPAR/RXR is then transported to the nucleolus and binds to the peroxisome proliferator receptor binding element (PPRE) which is located on the promoter site of the genes induced by this transcription factor complex. Details of PPAR/RXR regulatory mechanism is reported in [174, 108]. The set of transcription factors (TFs) and their target genes separated by the type of regulation (up/down) which we included in our in-silico simulation is listed in Fig. 6.11.

6.7.2 In-silico Regulatory Model

To abstract the expression and inhibition of the target gene *'X'*, we attribute the gene with a status flag and a time stamp. The status flag can have any of following three states: being expressed(BE), already inhibited (AI), or no activity (NA). The time stamp indicates the time for the last change in the status flag of the target gene.

6.7.2.1 Gene status transition

Transition of the target gene status from NA to either BE or AI is triggered by the transcription events. To handle the transition form BE or AI to NA a specific Gene Status Check (GSC) event is predicted that is executed periodically and compare the target gene time stamp with current time. If the difference between the two times is greater that a *GENE_HOLD_STATUS* constant then it shifts the status of the gene to NA. Based on current model there is no direct shift between BE and AI states.

	TF	Down regulated	Up regulated
1	Foxo1		SLC374A PCK1 G6PC HMGCS2 IGFBPL1 PDK4 FBP2 ADIPOR1 GCK ADIPOR2 SERPINE1
2	Foxo3		PDK4
3	Srebp1	MTP PCK1	ACACA FAS SCD5 LDLR PFKFB2 PFKM HK2
4	CebpA	PCK1	
5	CebpB	PCK1	
6	P53		ADIPOR2
7	Era	ADIPOR2	
8	Cmyc		ADIPOR2
9	Sp1		ACLY FAS IGF1 APOA1 LDLR INSR CALM1 SERPINE1
10	Sp2		ACLY
11	Srebp2		FAS
12	Usf		FAS
13	JunD		ME2
14	Fra2		ME2
15	Egr1		ME2
16	Creb1		G6PD
17	LxrB		SREBP1
18	Arnt		GLUT1 GLUT3 GAPDH ALDOA PGK1 VEGF
19	Irebp		GAPDH IGFBP
20	Sry		GAPDH
21	Stat5b		GCK
22	Sp3		INSR
23	Pparb		CD36 FABP3 FACL CPT1 ACADM ACADL ACACA MLYCD UCP3 PDK4

Figure 6.11. The transcription factors (TFs) and their target genes that we included in the database as the TRN for our in-silico experiments with *eukaSimBioSys*.

6.7.2.2 Selecting target gene *'X'*

The input to the transcription event is a transcription factor *'T'*. Execution of a transcription event indicates that resource for *'T'* is available. If the non-empty set g includes all the genes that are up/down regulated by transcription factor *'T'*, upon execution of a transcription event one of these genes will be selected for the status change with probability $p = \frac{1}{|g|}$. Based on wether the selected gene *'X'* belongs to up-regulated or down-regulated subset of g, its flag would change accordingly.

6.8 Skim through the *eukaSimBioSys* design issues

The design of the software has a tremendous impact on its functionality and performance. Following we briefly discuss some of the design specifications of the *eukaSimBioSys*.

6.8.1 Persistent Reaction Model

The strategy that we took in projecting the biological function into sequence of event can impose some glitches to system that needs to be resolved. As discussed in chapter 2 each event upon execution reduces the count of each input resource by one and adds one unit to the count of each output resources. Now lets consider a phosphorylation event of protein B which of course is a reaction event. The input resources to this event are *Phosphokinase K*, *Phosphate*, and *protein B*. If we keep up with the simple reaction event after execution of this event the resource count of *Phosphokinase K* will be reduced by one, however we know that a *Phosphokinase* enzyme can potentially participate in many number of reaction during its half life. To cope with this problem we introduce a *Persistent Reaction Model* (PRN) whose length is close to active life of an enzyme. The *holding time* of this event is first manipulated based on the simple reaction model [85], and then dilated over the active half life of the enzyme. Upon the execution of this event the count of enzyme or protein is reduced by one where the count of other input and output resources is altered based on the average number reaction that enzyme (protein) participated in, during its active half life. It should be mentioned that active half life of the *Persistent Reaction Model* is a calibrating point of the reaction.

6.8.2 Fast Reaction Model

The reaction events whose activation energy is quite low ($< 13\ \ k_BT$) occurs very frequently, such occurrence exacerbates during the abundance of reactants. Since the

event queue of the *eukaSimBioSys* forms a heterogenous pool of event types with a broad domain for execution frequency. Therefore, a significant frequency difference among the events time might lead to simulation stiffness, to avoid such undesirable situation we pose a new event that can handle a batch of homogenous reactions. To manipulate the holding time we apply the original reaction model [85], with the difference that the reaction rate will be multiplied by batch size. Also on the resource update the count of input and output resource should be updated according to the batch or the least available resource in the target reaction. There is a trade-off between the batch size and accuracy of the results, to obtain the pseudo-optimum batch size we did an iterative local search process to push the results in vicinity of the experimental data. We observed that for batch size of 1000 reaction the in-silico were in an acceptable agreement with the reported data.

6.8.3 Diffusion and Reaction Bundles

In an effort to increase both the memory and temporal efficiency of the simulator we experimented and observed that if we bundle the reaction of diffusion events into small batches (< 20) the results accuracy remains very similar to before; however, the temporal and memory efficiency of the simulation increase several folds. Hence, we decided to create reaction and diffusion bundles where their current sizes are set to 10 for both bundles.

6.8.4 The remaining models

The protein conformation change and Glut4 Vesicle Transport model are the two models that we use in our event diagrams for which we have not developed any physic-ochemical models as yet , their design will be a part of our future work plan. During in-silico experiments to estimate the holding-time by the two models we use the negative exponential distribution where their rates are potential calibration points of simulation.

6.8.5 The in-silico Experiment Protocol Package (IEPP)

For each in-silico experiment the experiment designer should provide an in-silico experiments protocol package (IEPP). This package compose of three pieces of information which defines the framework of that in-silico experiment, and is mandatory for running any experiment with *eukaSimBioSys*. The three data sets include: event-table, event-map, and molecular resources. Event-table defines the set of signals that exist in each experiment and for each signal lists the events along with their associated inputs, outputs, and stochastic model. Event-map defines the ordered sequences of events triggered in an experiment. On the other hand, event-map describes the event network of the experiment in a machine readable format. The last data set lists the molecular counts of molecular species in the experiment. In current version the IEPP composed of six text files out of which four files are the molecular resources. Fig. 6.12 shows a sample in-silico experiment protocol package.

6.9 *eukaSimBioSys* Implementation Overview

In this section we briefly describe selected specifications and implementation features of *eukaSimBioSys*. The simulator is entirely implemented in JAVA. The total project has 9228 lines of code and consists of 150 classes. This version of software implements 33 different types of events and includes 17 implementations of different physicochemical models. The software has the ability to directly import the data both from Excel 2003 spread sheets and from SBML level 2 version 2 files.

6.9.1 Database

For data repository and database transactions *eukaSimBioSys* makes use of *Caché* for windows version 5.2.3 (evaluation edition) which is an object-oriented database by InterSystems®[175].

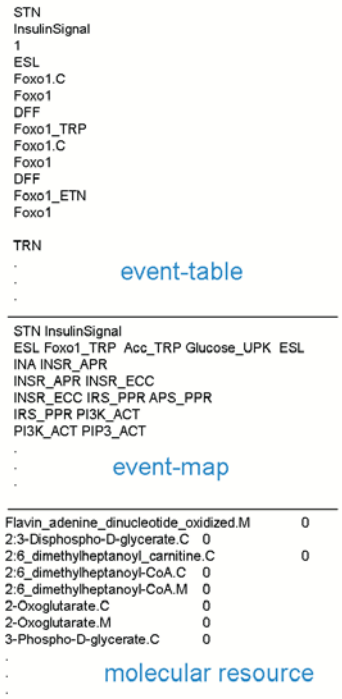


Figure 6.12. Sample in-silico experiment protocol package (IEPP).

6.9.2 *eukaSimBioSys* outputs

eukaSimBioSys has three types of out files:

- Log file: logs the execution detail of every events.
- The molecular resource output file: which reports the count of a specific molecule every t_{report} seconds.
- Systems resource file: reports the utilizations of a predefined set of a system resources every t_{report} seconds.

The current version of *eukaSimBioSys* has a general text based interface. Prior to every in-silico experiment the general data files which supply the information for the objective metabolic, transcription regulatory, and protein-protein interaction networks must be loaded into the *eukaSimBioSys* database.

6.10 Summary

In this chapter we tried to wrap up our discussion on preparing and conducing in-silico experiment of a eukaryotic cell by covering a wide range of remaining concepts. The substrate metabolism in heart was described with emphasize on glycolysis and β-Fatty acid oxidation pathways. For either pathways an event based abstraction of the physiology was introduce that could be executed and analyzed in the context of discrete even simulation paradigm. We also elucidate the utilization of the flux balance analysis in estimating the metabolic reaction flux by the metabolic event.

A stochastic discrete event model of constitutive splicing was also proposed and proceeded by the *eventolgoy* of gene expression and protein synthesis in a eukaryotic cell. An abstracted mechanism for the transcription regulatory network and a strategy in regulating the genes underlaying a transcription factor was also discussed. Next we introduced one of the emerging problems due to an event based view on the signal transduction network and proposed the persistent reaction model to solve that. Stiffness and efficiency of the *eukaSimBioSys* was briefly discussed and few strategies were proposed to avoid the former and enhance the latter. Finally we closed our discussion by giving a brief overview on *eukaSimBioSys* implementation.

CHAPTER 7

IN-SILICO EXPERIMENTS AND RESULTS

Now the pieces of the leggo that has been built and painted should be properly assembled to bring about the objective into practice. In this chapter we selected six different reported bench work experiments and tried to utilize the *eukaSimBioSys* to reproduce their empirical results in-silico by mimicking their experimental conditions in the abstract context of stochastic discrete event simulation. The criteria behind choosing these experiments was to demonstrate a unique feature and capacity of our software in simulating the eukaryotic intracellular networks with each experiment. For the reasons discussed earlier, we have selected the insulin signaling and its metabolic effects in cardiac myocytes as the focus of our experiments. Maintaining a coherent focus across the sequence of experiments that we conduct will help the readers to better comprehend the state of art of our approach in modeling biological systems.

In previous chapters we highlighted that, the scale of our modeling and simulation is at the molecular level, where as the biological, biochemical, and physiological experiments are often conducted at tissue or organ level. This scale discrepancy poses another milestone in the road map toward validating our results. We briefly describe our solution to this emerging problem in the proceeding section.

7.1 Re-scaling the results

To eliminate the scale discrepancy among the in-silico and wet lab result we have to re-scale either one of these results. This re-scaling could be done both ways (i.e.

up-scale or down-scale), however there are certain limitations to this approach. The experimentalist normally conduct their experiments at one of following levels:

I. Molecular level (rarely)

II. Cell level

III. Tissue level

IV. Organ level

Despite conducting the experiment at one level it is frequently observed that results have been reported at another level. The first two level of experiments scale are the most desirable and we can simply compare the results. On the case that experimental data is reported at tissue level we can convert the results to the cell level and ultimately the molecular level, should we have a rough approximation on the cell and substrates masses of as well as their dimensions, etc. The most undesirable scenario is once the reported data is at the organ level, normally it is not possible to deduce an approximation for result in cellular or molecular level merely by applying some factors along with conventional mathematical conversions. Therefore, we have avoided to use the experiments whose results subject to the last category.

Following is a list of values for selected parameters we used in our estimations and approximations: Average myocardial cell volume $= 40 \times 10^{-15}\ m^3$ reported in [176, 177], nucleus volume is $\sim$ %10 of the cell volume [58], myocardial cell mitochondria occupies $\sim$ %30 of the cell volume to supply the energy for perpetual contractile function of the myocardium [178], there are $\sim$ 4660 cardiac myocytes per 1mg wet cells (varies among the samples). To convert any molecular counts in heart muscle to nano-Molar concentration we divided the counts by 240.88×10^2. The weight per amino acid is considered 0.11 KDa and an average weight of a eukaryotic cell is $\sim 10^{-9}$ grams.

7.2 Experiment 1: Effect of insulin signal on flux across *6-Phosphofructo-2-kinase reaction*

This experiment aimed to measure the effect of the insulin signal on the uptake rate of glucose and, hence the activity of the *6-Phosphofructo-2-kinase(PFK-2)* enzyme in the Wister rat cardiac myocyte. The signaling, metabolic and transcription regulatory networks that involved in this experiment is depicted in Fig. 7.1. The bench work for this experiment was conducted by Hue *et. al* (1996) and we used their published data for validation of our in-silico experiment. Beside the insulin dependent glucose uptake by GLUT4 transporter, there is a diffusion based mechanism for glucose uptake in the this experiment which works independent of insulin and keeps the glucose content of the cell at the of basal level of < 2 pM. The experiments is done for 0.1 μM of insulin, 0.5 mM of glucose. The concentrations of signaling proteins and unregulated enzymes were set to 2.07 nM and 4 pM, respectively, to be consistent with data reported in [179]. For the number of insulin receptor sites per cell we set 22K sites pre-cell which is consistent with the data reported in [180, 181].

The simulation was set to run for 30 minutes of simulation time which is equivalent to duration of the medium perfusion in the corresponding wet lab experiments. Fig. 7.2 depicts the fold change in glucose uptake rate in presence (solid line) and absence (dotted line) of insulin signal for in-silico (blue) and wet lab (black) experiments. As the diagram in Fig. 7.1 indelicates the increase in cellular content of glucose should pump up the flux across the Glycolysis I and fructose-mannose metabolism pathways. Also downstream the insulin signal, positively regulates the *PFKFB2* gene and, hence, results the increase in the concentration of *PFK-2*. Richer intercellular glucose and *PFK-2* resources lead to flux increase across *PFK-2* reaction which yields more *Fructose 2,6-biphosphate* ($Fru2,6P_2$) product. Fig. 7.3 shows the endogenous concentration of *Fructose 2,6-biphosphate* versus time in presence and absence of the insulin signal. The

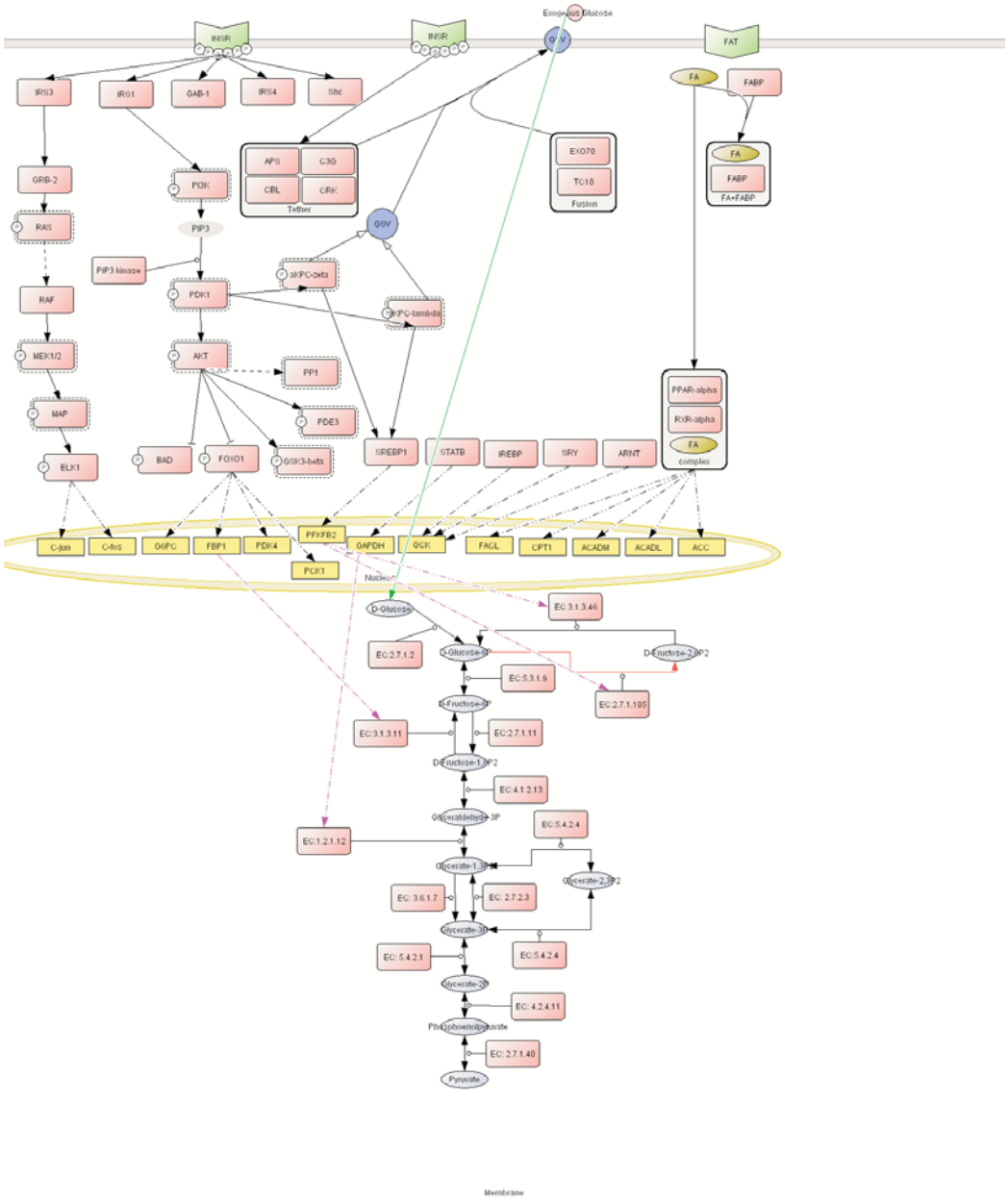


Figure 7.1. Cellular networks involved in experiment 1. The dynamics of reaction colored in red is investigated in experiment 1.

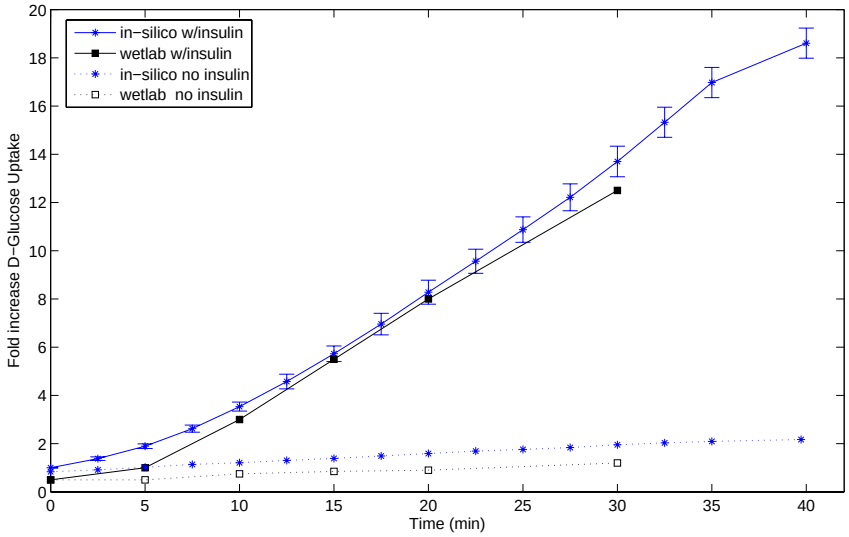


Figure 7.2. Effect of insulin on the rate of $D-Glucose$ in the cardiac myocyte.

fluctuation observed on the in-silico result in the presence of the signal, is due to the interval arrival time of the metabolic events (600 secs). Should the metabolic events occur more frequently this fluctuation will be vanished, with the const of simulation efficiency. The initial positive slope observed in the in-silico concentration of $Fru2,6P_2$ in absence of insulin is due to the initial Glut4 vesicles on the membrane.

As observable both measurements asserts the agreement between the in-silico results and the reported experimental data. Increase in the uptake rate of glucose requires rapid presence of $GLUT4$ transporter on the membrane of the cell. We have measured the membrane $GLUT4$ and the $D\text{-}Glucose\text{-}6P$ concentrations during the evolution of the in-silico experiment. The membrane $GLUT4$ activity is directly proportional to the cellular concentration of $D\text{-}Glucose\text{-}6P$, such dependency is properly demonstrated in Fig. 7.4. The slow decline in picks of membrane $GLUT4$ concentration is the consequence of the gradual decrease in insulin concentration which leads to decline in signal strength.

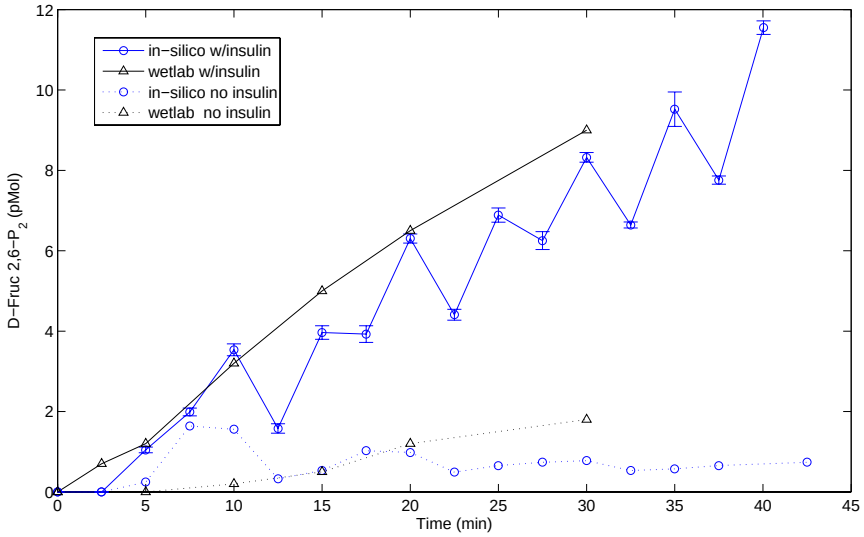


Figure 7.3. Effect of insulin on cellular content of $Fru2,6P_2$.

7.3 Experiment 2: A hypothesis testing on the effect of glucose phosphorylation on the insulin-dependent mTOR signaling in the cardiac myocytes

This experiment targets two objectives: i) how a hypothesis testing experiment could be conducted in-silico, ii) the capability of our software to capture the dynamics of a cross talk between signal transduction and metabolic networks.

An evolutionary conserved kinase which is crucial for nutriment depended regulation of cell growth is the target of rapamycin (TOR). Taegtmeyer and his colleagues in [182] has postulated that metabolism of glucose positively regulated the *mTOR* signaling in the cardiac myocytes.

In order to test this hypothesis in-silico we first added the hypothetical interaction edge between the *D-Glucose-6P* and the *mTOR*. This edge represents a phosphorylation reaction event, on the other hand the availability of the phosphorylated glucose induces the phosphorylation of the *mTOR*. The signaling and metabolic pathways involved in this experiment is shown in Fig. 7.5

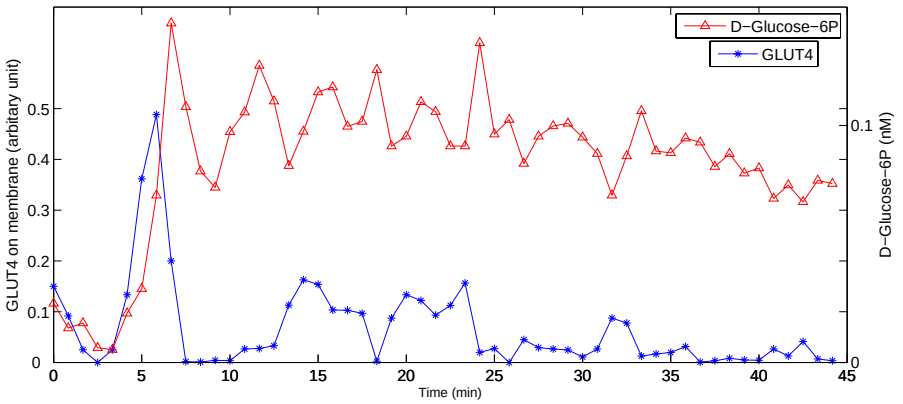


Figure 7.4. In-silico measurements of membrane *GLUT4* and *D-Glucose-6P* concentrations in the presence of insulin .

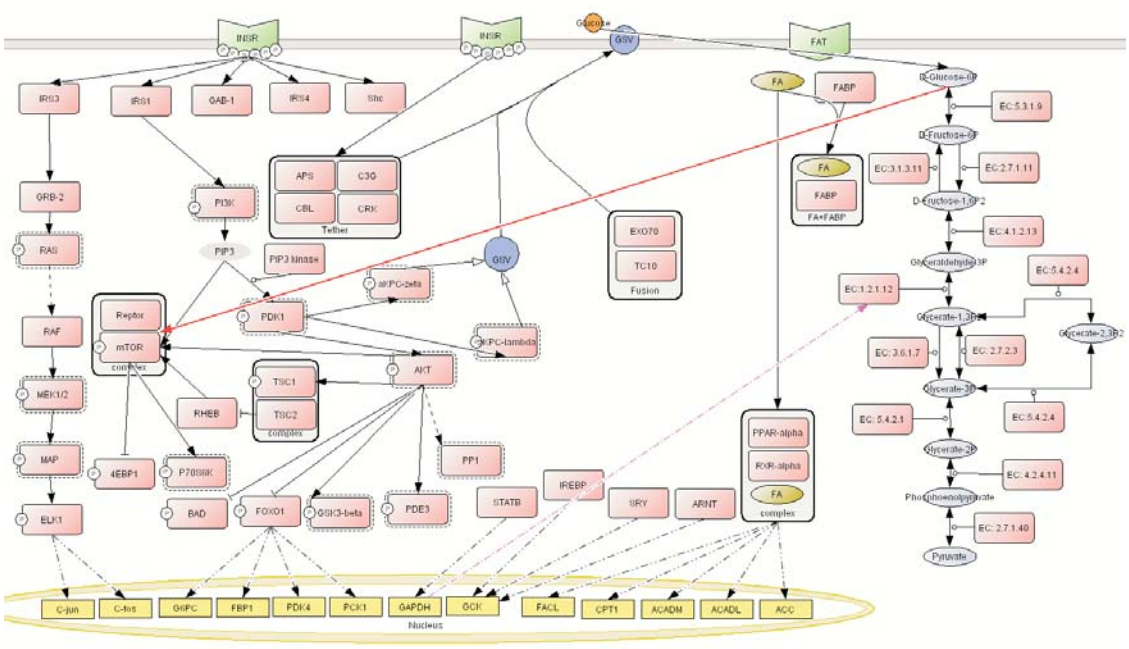


Figure 7.5. The insulin signaling pathways including: target of rapmaycin kinase (mTOR) and downstream effectors: P70s6k and 4ebp1 proteins. The hypothetical cross talk between metabolic and signaling networks is depicted in red.

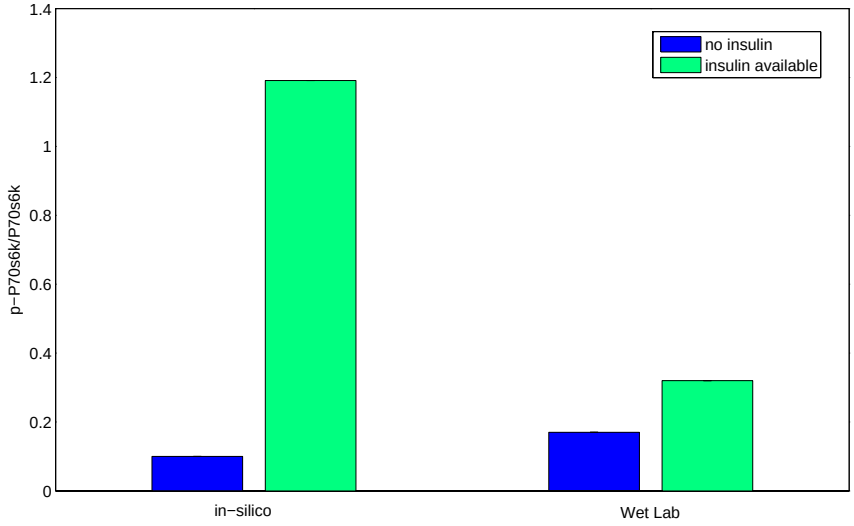


Figure 7.6. The amount of phosphorylated *P70s6k* with insulin as the control.

Molecular resources for this experiment include: 5mM *glucose* , 0.1μ insulin, 0.9 nM of each signaling proteins, and 4 pM of each metabolic enzymes. We use the insulin as the control and first run the simulation with no insulin for 30 minutes and then in presence of insulin for the same period. The basal level for phosphorylated *P70s6k* which is a downstream effector of insulin signal is set to be %18. Fig. 7.6 shows the phosphorylated *P70s6k* after 30 minutes of simulation alongside the experimental data. As we would expect presence of insulin signal strongly induces *P70s6k* phosphorylation and the results are also consistent with the published experimental data. For further validation of our results we measured the ratio of the phosphorylated Akt (p-Akt). Fig. 7.7 shows these ratio for the simulation an wet lab reported data.

To assess the effect of the hypothetical edge, *D-Glucose-6P* enhanced mTOR signaling, we simulated the experiment for both the existing and non-existing edge scenarios and compared the ratio of phosphorylated *mTOR* and *P70s6k* in each case. Fig. 7.9 shows that *D-Glucose-6P* has markedly enhanced both the signal propagation (rate of

phosphorylation) and also strength of the signal (quantity of phosphorylated species). This claim is experimentally confirmed but not quantified in [182].

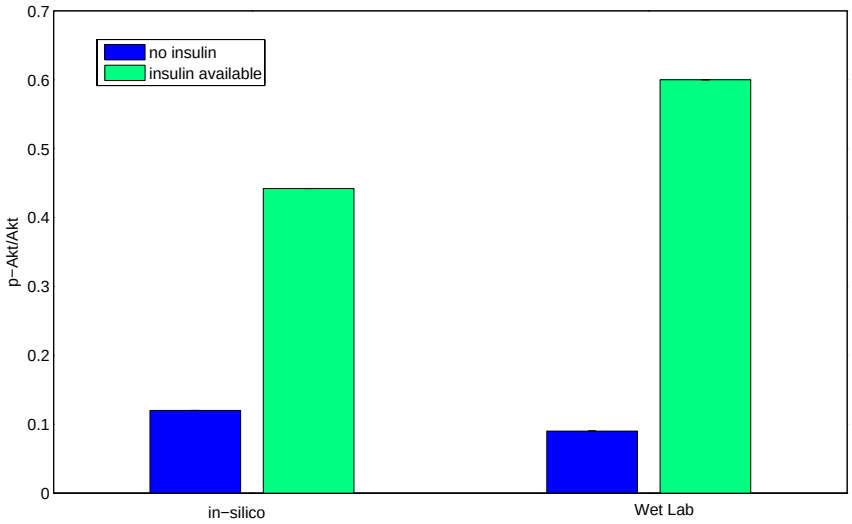


Figure 7.7. Ratio of phosphorylated *Akt* with insulin as the control.

To have a better picture of the interaction of *D-Glucose-6P* and current signaling pathway, in Fig. 7.8 we have provided the average flux across all glycolysis I pathways reactions during the course of this in-silico experiment.

In another effort to asses the contribution of the *D-Glucose-6P* on *mTOR* signaling, researchers in [182] inhibited the glycolysis pathway by mutating the *Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)* with 10 nM of enzyme modulator *iodoacetate*. The authors did not report any quantified results on their experiment, although they have qualitatively (shown on electrophoresis gel images) confirmed the increase on the phosphorylated *mTOR* as expected . We conducted the in-silico experiment for this mutation and measured the effects. We set the in-silico experiment such that the mutation started after 10 minutes of simulation.

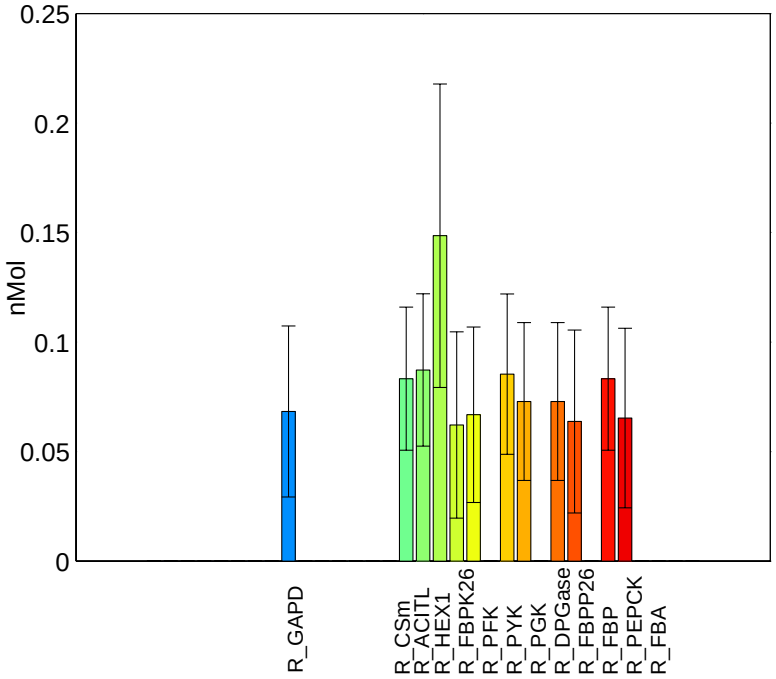


Figure 7.8. Reported flux across glycolysis I pathway reactions during 30 minutes of in-silico experiment.

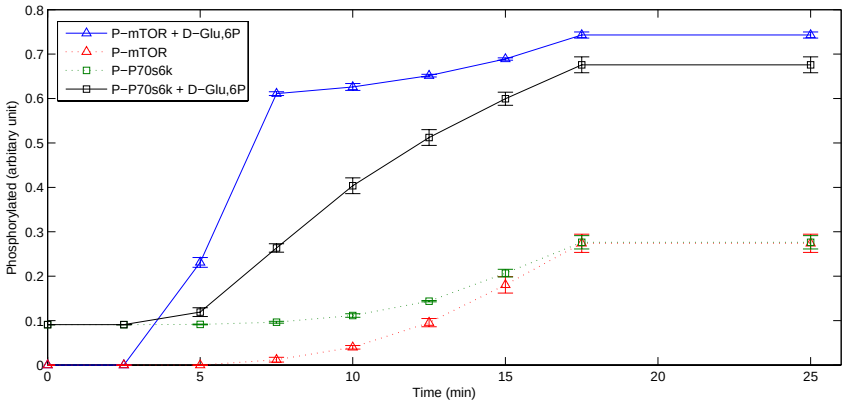


Figure 7.9. Quantified contribution of *D-Glucose-6P* in *mTOR* signaling.

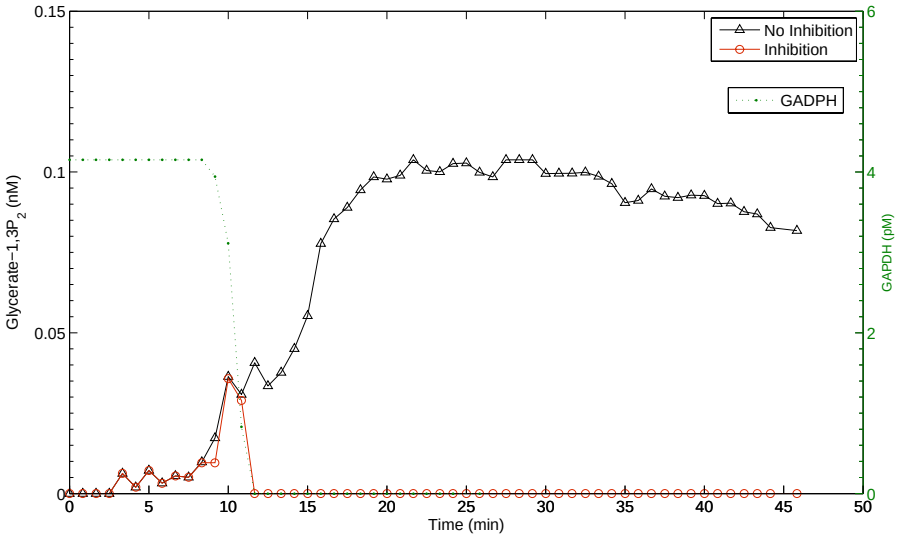
The green line in Fig. 7.10.(a) shows the concentration of wild-type enzyme. As shown in Figs. 7.10(a,b) the inhibition of the glycolysis pathway induces the cellular content of metabolite intermediates upstream the mutation point while diminishes the content of downstream ones. As expected such induction should positively effect the $mTOR$ signaling. Figs. 7.11(a,b) show this effect, the vertical black line in chart specifies the point where the effect of pathway inhibition on the signaling becomes significant ($p < 0.1$).

7.4 Experiment 3: Quantifying the effect of feedback loops on insulin signaling pathways in cardiac myocytes

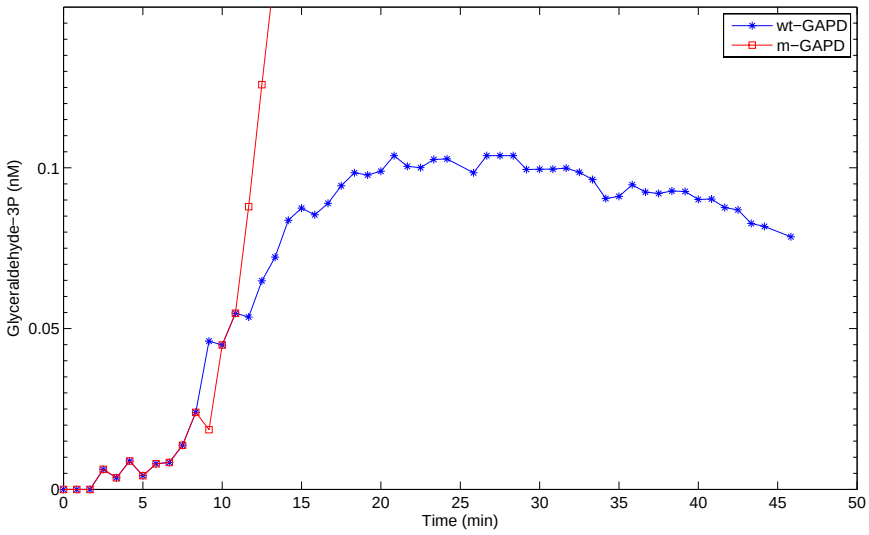
This experiment intended to pursue two objectives: Firstly, to demonstrate the capability of the software to implement and capture the dynamics of negative and positive feedback loops in biological pathways. Secondly, to compare stochastic discrete event based approach with the ordinary differential equation based framework in modeling biological pathways.

Qoun *et. al* (2002) proposed a mathematical model for studying the insulin signaling pathways in cardiac myocytes [183]. In their model they identified all interaction in the insulin signaling pathways of a myocardial cell as biochemical reactions. Then applied the general ordinary differential equation (ODE) method to quantify the system dynamics as follows: i) deferential equations corresponding to the biochemical reactions in the signaling pathway were formed, ii) numerically solved the set of equations to capture the dynamics of the system. We first slightly modified the original insulin signaling diagram to be consistent with the pathways in this experiment. The modified pathway is shown in Fig. 7.12

In the current experiment the insulin receptors ($INSR$) are considered to have three states: 1) non ligand bound $INSR$, 2) Single ligand bound $INSR$, and 3)bi-ligand



(a)



(b)

Figure 7.10. Inhibition of Glycolysis I pathway and its affect on *mTOR* signaling; (a) dynamics of *GAPDH* concentration during inhibition, comparison between the $Glycerate-1,3P_2$ concentration for inhibited and non-inhibited pathways inhibition. (b) compares the *Glyceraldehyde* concentration for inhibited and non-inhibited pathways.

bound *INSR*. Also the phosphorylation of the *IRS* protein could happen on *serine* or *tyrosine* sites. The phosphorylation on the former site is mediated by *aPKC* isoforms and leads to inhibition of *IRS*, where as phosphorylation on latter activates *IRS* and is *INSR* dependent [183].

The PTP1B is a protein tyrosine phosphatase (PTPase) that negatively regulates the insulin receptor and *IRS*. However, it has been identified that this PTPase is the substrate for kinase activity of down stream effector *Akt*. Taking into account the effect of successive negative edges, the ensemble will a the positive feed to the insulin signaling pathway. Both of the feedback loops described are depicted in Fig. 7.13.

To simulate this pathway we considered each reverse reaction as a separate reaction event, the activation energy of the reverse reactions were subject to calibration as well. Also the binding of the second ligand to insulin receptor is also modeled with reaction event. The concentration of species are as follows: 1μM of insulin, 1mM of glucose, 0.9 nM of each signaling proteins, and 4 pM of each metabolic enzymes. We set the simulator such that the insulin signal stoped after 15 minutes of simulation time. We run the simulator for 60 minutes of simulation time once for the pathway including both feedback loops and next for pathway excluding both.

Figs. 7.14 and 7.15 compare the results between ODE approach and our simulation method for the temporal dynamics in activation of *aPKC* and *PIP3*, respectively, by insulin signaling pathway including and excluding feedbacks. In both charts one obvious difference is the lag difference between the maximal activated substrate predicted by two models. One justification for this difference could be that ODE based models can not capture the delays in the biological system due to their modeling nature. However there are possible solutions to this gap (e.g. dummy delay equation) which was not included in model by the authors in [183]. The response to the lack of signal after 15 minutes is immediate in the ODE based model however in our simulation response

appears later. This is also due to the signal propagation delay that is captured by our approach where the general ODE models are insensitive to the delays. A side from difference both approach shows the similar maximal substrate activation ratio in either feedback scenarios.

Fig. 7.16.(a) shows the ratio of the *GLUT4* on the membrane for both approaches and feedback scenarios. The main difference in the results are the decay rates of membrane *GLUT4*, which our simulation shows a significantly higher rate. Also as it is observed in blue plot the reduction in the membrane glucose transporters happened event before the signal removal (15 min), this because once the signal clicks in a considerable number of *GLUT4* move to the membrane and subsequently many of them dissociate from the membrane in a close lag. Considering this with cellular content of *GLUT4* and stochasticity of cell environment will gives a sinusoidal nature to the membrane *GLUT4*-time curve. Hence, should signal persist in this experiment we would be able to detect this behavior. Furthermore, to exhibit the stochasticity of process captured by our simulation we have shown the log ratio of membrane infused *GLUT4* in Fig. 7.16.(b).

In Fig. 7.17 we have plotted the concentration of insulin receptors that were bound with a single or double insulin ligand. As it is observed the abundance of mono-insulin-bound receptors are one order of magnitude higher than the bi-insulin-bound counterpart. Also the dissociation of ligands from the receptors shows slow decay rate where as the ODE results (not shown) has an exponential decay curve of ligand dissociation. The exponential decay curve in ODE approach results from solution to first orderer differential equation which is a negative exponential function. Also, there is one order of magnitude difference in the abundance of activated receptors between the two approaches.

For the same pathway we conducted the dose-response analysis experiment for the insulin signaling. [1] This analysis is for insulin concentration range of 10^{-12}M to 10^{-6}M.

Figs. 7.18 (a) and (b) show the dose-response percentage curves for activated-*INSR* and membrane-*GLUT4*, respectively. The comparison between experimental data, ODE results, and current simulation results reveals that ODE stays in a good agreement with experimental data; however, current simulation demonstrated an acceptable agreement with experimental results for high concentrations of insulin (¿10^{-7}M) but looses its sensitivity for lower concentrations. We did a rigorous study to pin point the root of this sensitivity decline for lower ligand concentration in our simulation. It became apparent that the membrane protein-ligand docking model (from chapter 3) that we used to estimate the time for insulin receptor binding does not provide a precise binding time estimation for ligand concentrations less than 10^{-6}M. Hence, the model proposed in chapter 5 needs a revision and enhancement to provide a more precise estimation for binding time for the low ligand abundance.

In order to characterize the effect of the positive feedback loop on phosphorylation of *IRS1* protein, we have plotted the cellular content of tyrosine phosphorylated *IRS1* with respect to the concentration of phospho-activated downstream effector *p-Akt* in Fig. 7.19 . The concentration of tyrosine phosphorylated *IRS1* exponentially increases with respect to *p-Akt*. However, this exponential increase is the joint effect of positive feedback loop and persisting insulin signal.

Fig. 7.20 shows the bifurcation dynamics of activated *aPKC* and PI3K. These bifurcations are the consequence of insulin signal state change. In the presence of signal the concentration of both substrates characterize a pseudo-monotonic increase and once the

[1] The dose-response is defined as amount of substrate *'x'* that is produced after time t of applying signal *'s'* with magnitude of concentration c.

signal diminished the concentration of both substrates decline monotonically, although with different rates.

Feedback loops affect the *GLUT4* transport activity, hence it affects the rate of exogenous glucose uptake. Fig. 7.21 shows the cell content of *D-Glucose* in either feedback scenarios. Although the initial effect of feedbacks on uptake rate is not significant, however this effect becomes significant in the long run. The rationale for the saw edge shape of the chart as discussed in the earlier experiments, is related to the inter-arrival time of the metabolic events.

One of the major drawbacks of the ODE based modeling approach is there inability to simultaneously incorporate the reactions whose difference in their evolution rates is significant . For instance most of the reactions involved in a signaling network are several orders slower than metabolic network reactions. Therefore, joint analysis of signaling transduction and metabolic networks seems to be a tough nut for ODE modelers. In Fig. 7.22 charts we present the dynamics of fluxes across typical reactions in glucolysis I pathway which is the result of concurrent simulation of signal transduction and metabolic networks in recent experiment. Each flux is for one minute slot time which is manipulated from difference between cell content of the substrate in two successive minutes.

The average flux for reaction R at time t where the metabolic event slot $s = \lfloor \frac{t}{600} \rfloor$, could be obtained from:

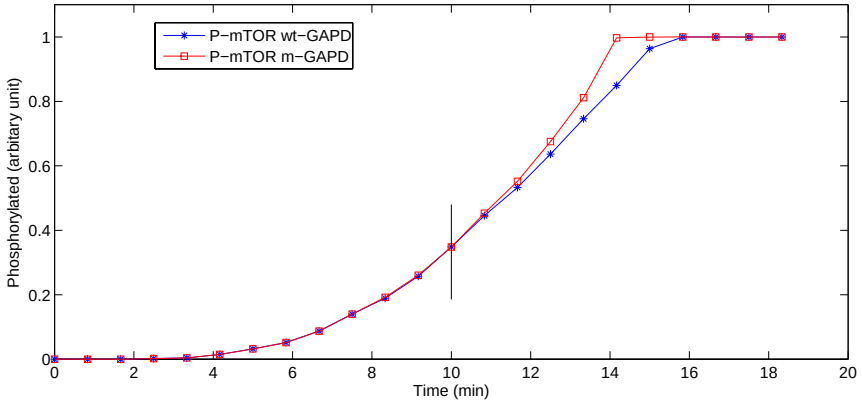
$$\bar{F}^R(s) = \frac{\sum_{t=s\times 600+1}^{(s+1)\times 600} F^R(t)}{600} \tag{7.1}$$

7.5 Experiment 4

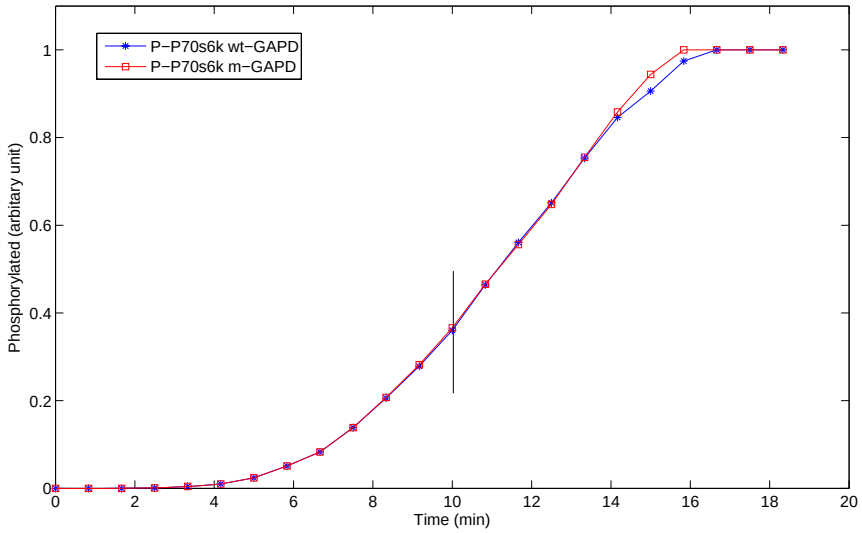
In this experiment we try to simulate the energetics of the cardiac myocytes. This experiment quantifies the contribution of exogenous glucose and fatty acid in the ATP production of the cardiac myocyte in the aerobic respiration. We also analyze the regulatory effect of Acetyl-Coenzyme Carboxylase ACC which is a key regulatory enzymes

in fatty acid oxidation pathway. *ACC* is biotin containing enzyme that catalyzes the carboxylation of cytoplasmic (*Acety-CoA*) to form the *Malonyl-CoA*. This product is the potential inhibitor of carnitine palmytoyltranferase 1 *(CPT-1)* which is the member of the carnitine palmytoyl transport system (CPTS), which is responsible for translocating the fatty acid to mitochondria. *CPT-1* is located on the outer membrane of mitochondria. There have been other pivotal role suggested for *Malonyl-CoA* which is not the target of this study.

This experiment also shows the concurrent effects of insulin and fatty acid signals. An interesting regulatory effect of these signals include the negative regulation of *ACC* by insulin and positive regulation by fatty acid. Fatty acid signal will also positively regulate the pyruvate dehydrogenase kinase *PDK* enzyme which inhibits the pyruvate-dehydrogenase (PDH) localized on the internal membrane of mitochondria. Pyruvate-dehydrogenase is the first component of pyruvate-dehydrogenase-complex (*PDC*) which transforms the pyruvate to *Acety-CoA* through decarboxylation and fuels the citrate acid cycle. Fatty acid also induces the expression of fatty acid ligase *FACL* and mitochondrial precursor for acyl-CoA dehydrogenase (ACADM) which participate in fatty acid synthesis and fatty acid oxidation pathways, respectively. Fig. 7.23 shows the details of signaling and metabolic pathways involved in this experiment.



(a)



(b)

Figure 7.11. Inhibition of Glycolysis I and its affect on *mTOR* signaling; (a) shows the effect of pathway inhibition on *p-mTOR* and (b) shows the same effect on *p-P70s6k*. The *wt-GAPDH* and *m-GAPDH* in the legends refer to the mutated and wild-type enzymes, respectively.

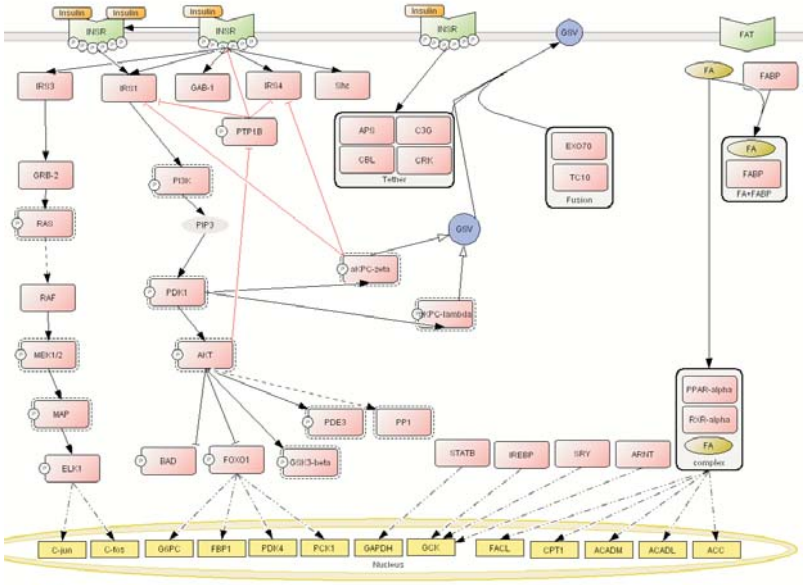


Figure 7.12. Insulin signaling pathway with feedback loops.

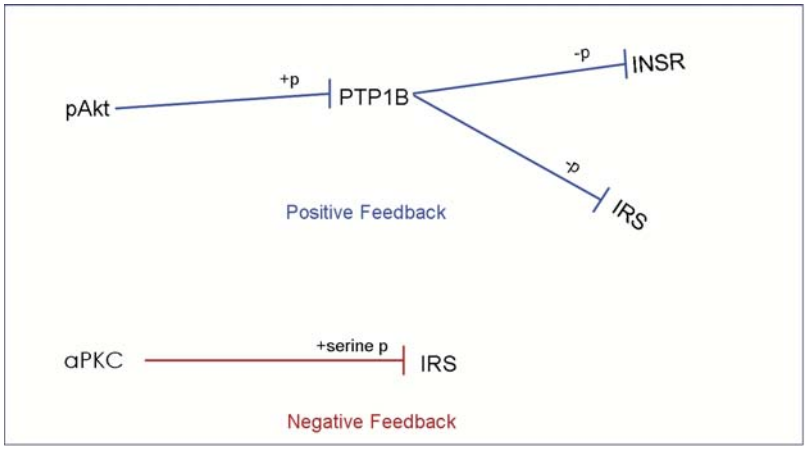


Figure 7.13. The positive and negative feedback loops.

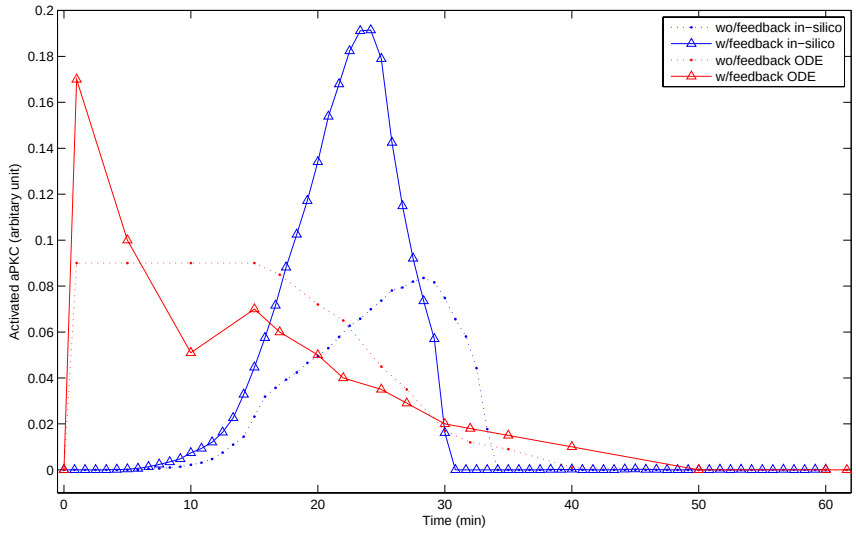


Figure 7.14. Dynamics of *aPKC* activation for in the insulin signaling pathway with feedback (solid line) and without feedback(dotted lines) captured by ODE (red plot) approach and *eukaSimBioSys* (blue plot) simulation.

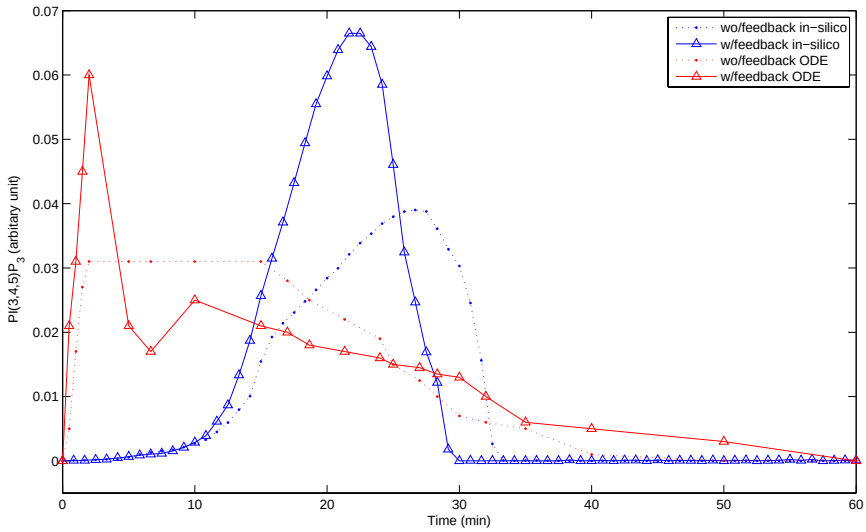
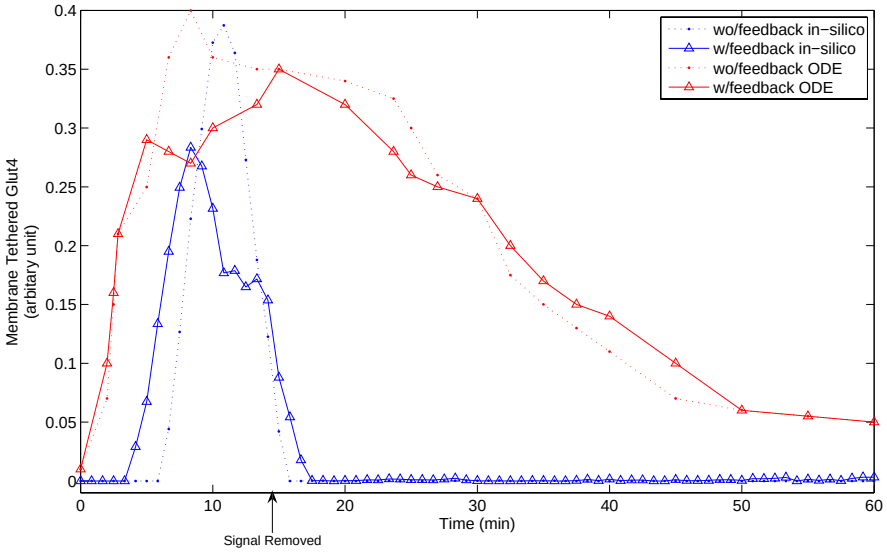
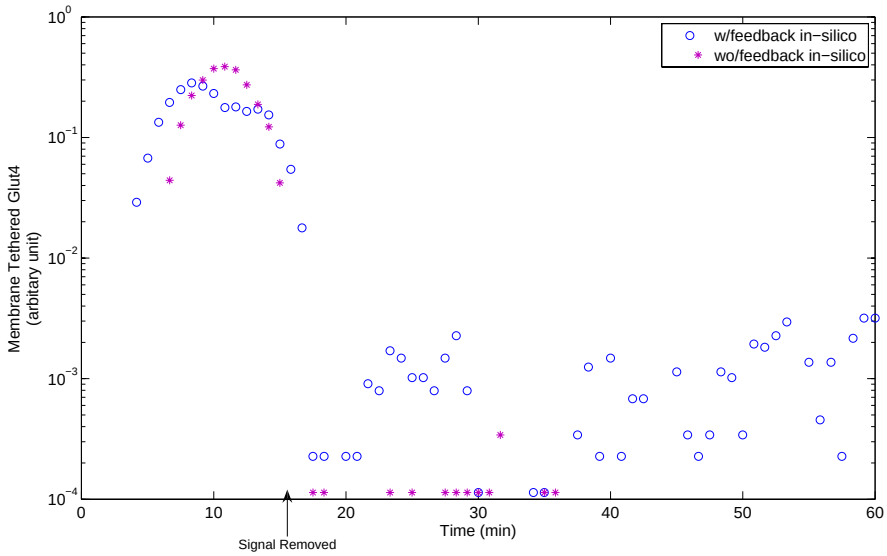


Figure 7.15. Dynamics of *PIP3* activation for in the insulin signaling pathway with feedback (solid line) and without feedback(dotted lines) captured by ODE (red plot) approach and *eukaSimBioSys* (blue plot) simulation.



(a)



(b)

Figure 7.16. Membrane *GLUT4*: (a) Ratio of *GLUT4* tethered to the membrane mediated by the insulin signaling pathway with feedback (solid line) and without feedback(dotted lines) captured by ODE (red plot) approach and *eukaSimBioSys* (blue plot) simulation, (b) Stochasticity in the log ratio of membrane *GLUT4* captured by current simulation of insulin signaling pathway.

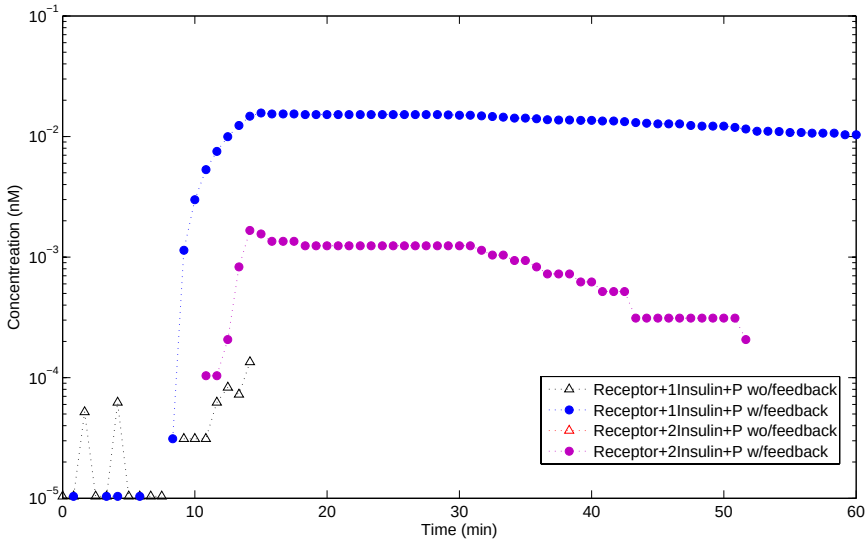
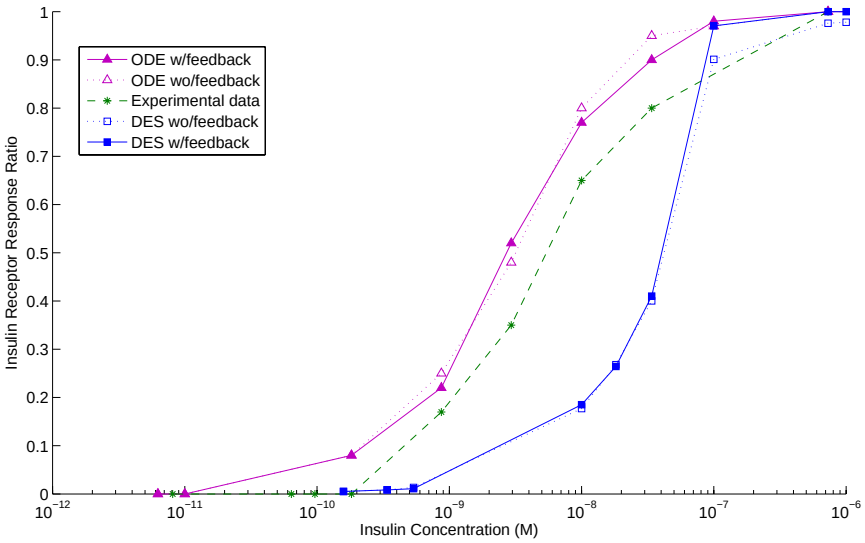
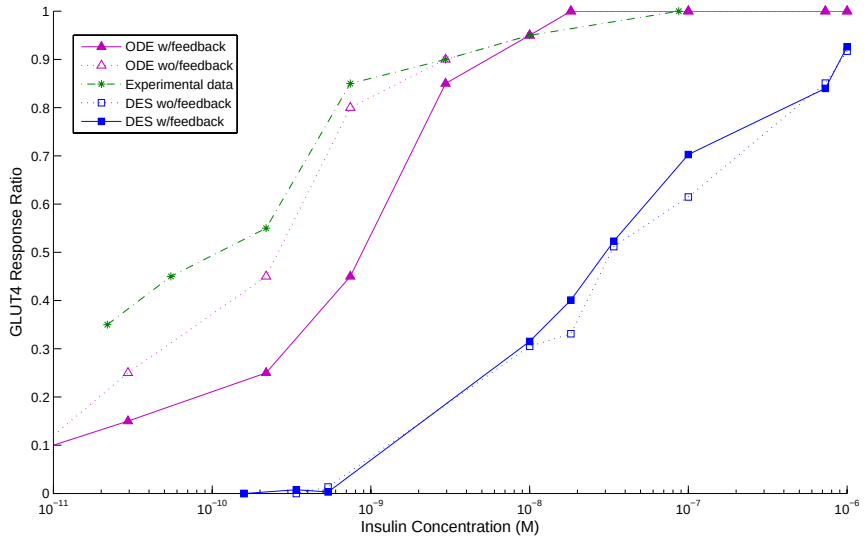


Figure 7.17. Abundance of mono-insulin and bi-insulin activated insulin receptors.



(a)



(b)

Figure 7.18. Percentage of dose-response (PDR): experimental data curve in green, ODE in magenta, and current simulation is plotted in blue. Solid and empty markers denote the variants of pathway for including feedbacks and excluding feedbacks, respectively. (a) insulin receptor PDR,(b) *GLUT4* PDR.

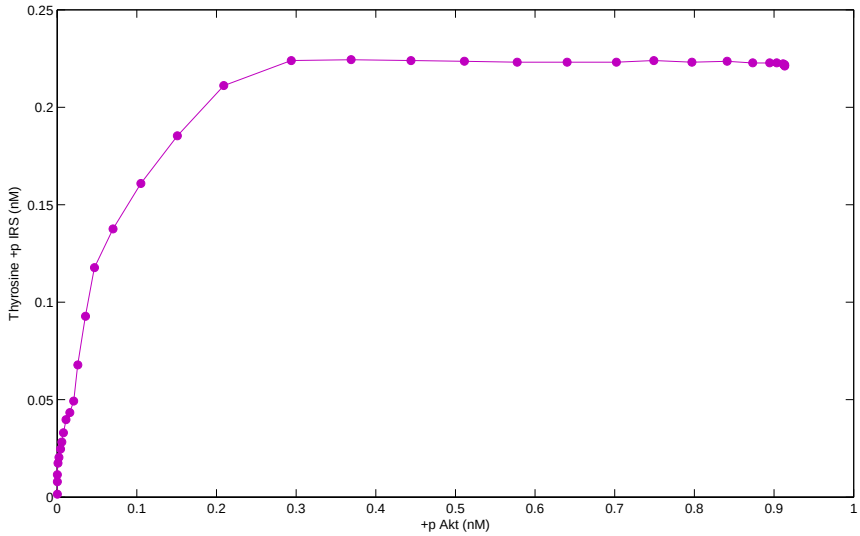


Figure 7.19. Effect of positive feedback loop on phospho-activation of *IRS1*.

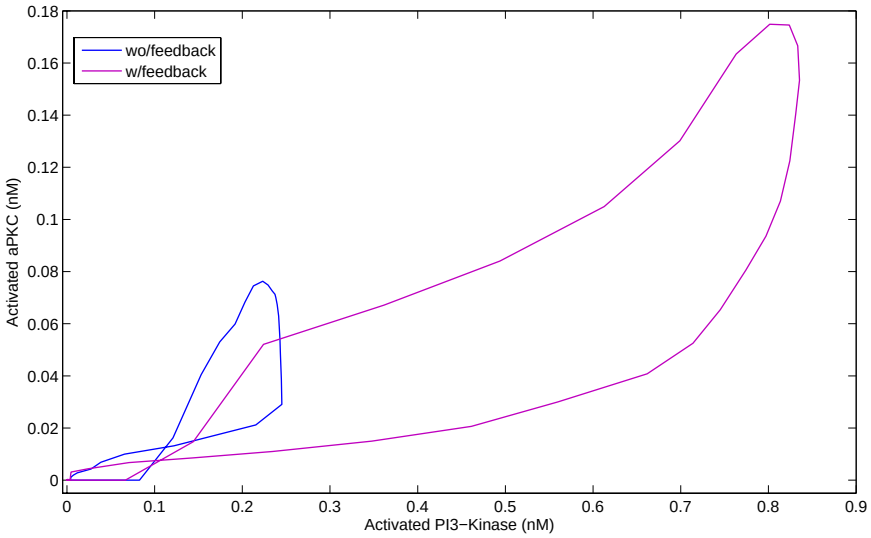


Figure 7.20. Bifurcation dynamics of *aPKC* and PI3K; blue and magenta plots represents the results for pathway with feedback and without feedbacks, respectively.

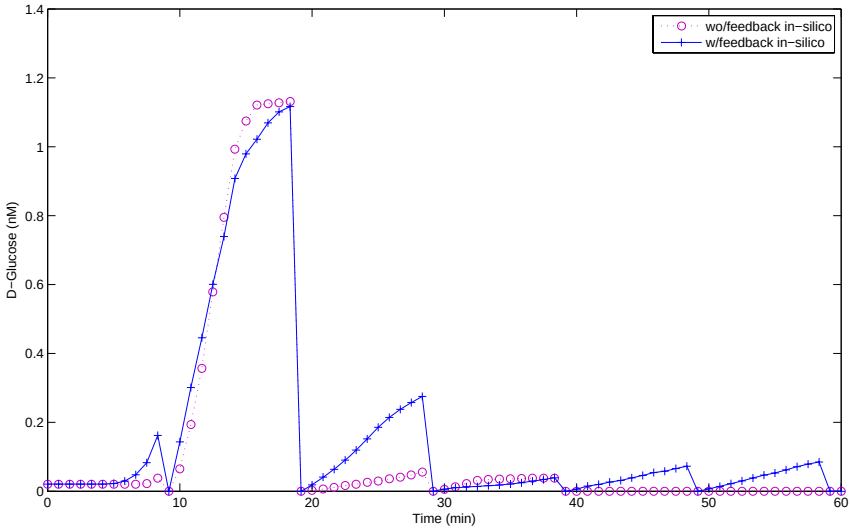
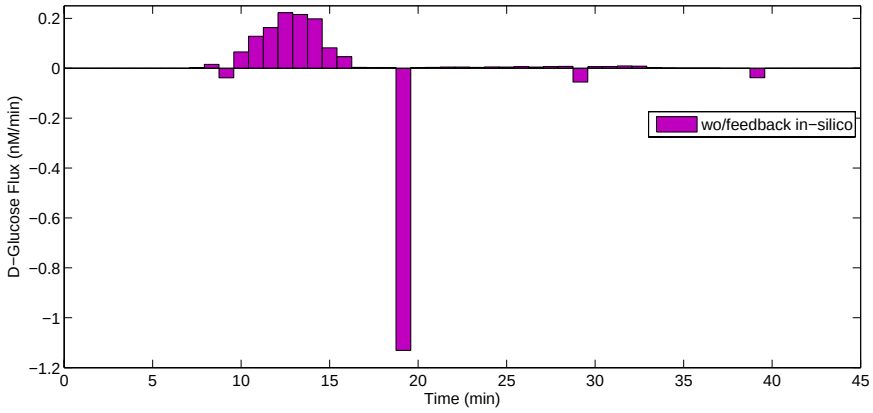
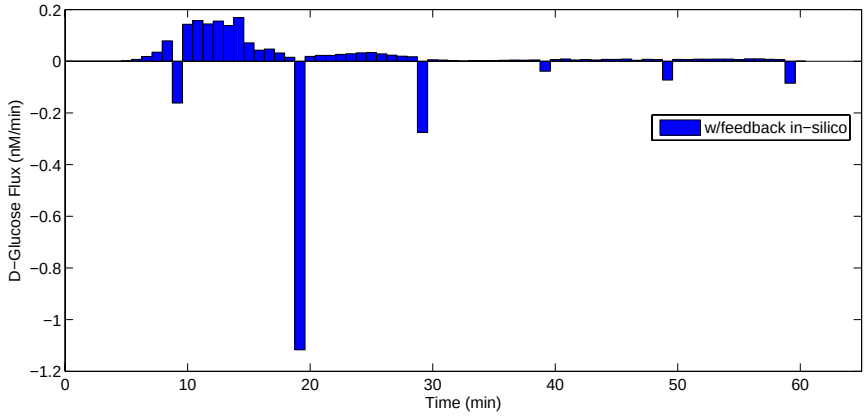


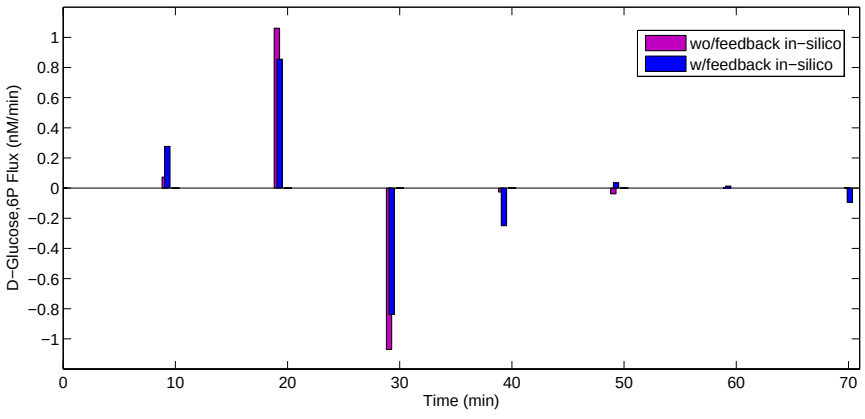
Figure 7.21. Effect of the feedback loops on the glucose uptake rate.



(a)



(b)



(c)

Figure 7.22. (a) *D-glucose* flux when signaling pathway has no feedback loop, (b) *D-glucose* feedback while the feedback loops are present, (c) *D-glucose-6P* flux for both variants of signaling pathway (blue: no feedback loop, magenta feedback loops exist).

For the in-silico experiment we used 11 mM of the Octadecaontate_(n-18:0) which is a saturated stearic fatty acid as the exogenous fatty acid resource. The same concentration of glucose was also available to the cell where the concentration of the other signaling substrates and metabolic enzyme were 0.4 nM, and 4 pM, respectively. The bench experiment was conduct by Lupaschuk *et. al* (1993) [184] at the organ level on Spare-Wawley rat heart and the results reported at the tissue level. We approximated the results to the cell level values with method mentioned earlier in the chapter. Also the authors used hexadecanoic which is a palmitic (16:0) saturated fatty acid as the exogenous fatty acid substrate. The choice of the stearic or palmitic fatty acid would not skew the results since the stearic acid will be converted to a palmitic acid by metabolic reaction *R_FAOXC180* which is an oxidation-reduction reaction in β-fatty acid oxidation pathway. We constructed the event table and event diagram for this experiment and run the *eukaSimBioSys* for one hour of simulation time.

7.5.1 ATP Manipulation

The aerobic glucose oxidation produces eight ATP molecules but is also consumes two. Also each pyruvate molecule resulted from glycolysis will be used by TCA cycle to produce three $NADH$, one $FADH_2$ and one GTP meanwhile uses one ATP to convert pyruvate to *Acetyl-CoA*. Furthermore $NADH$ and $FADH_2$ will use electron transport chain on the mitochondria membrane to produce three and two *ATPs*, respectively. Therefore, the net ATP produced from oxidation of a single glucose molecule would be six ATP molecules. Noting that the complete oxidation of one glucose molecule including the aerobic and anaerobic pathway and pyruvic acid metabolism in TCA cycle will result 36 total ATP molecules.

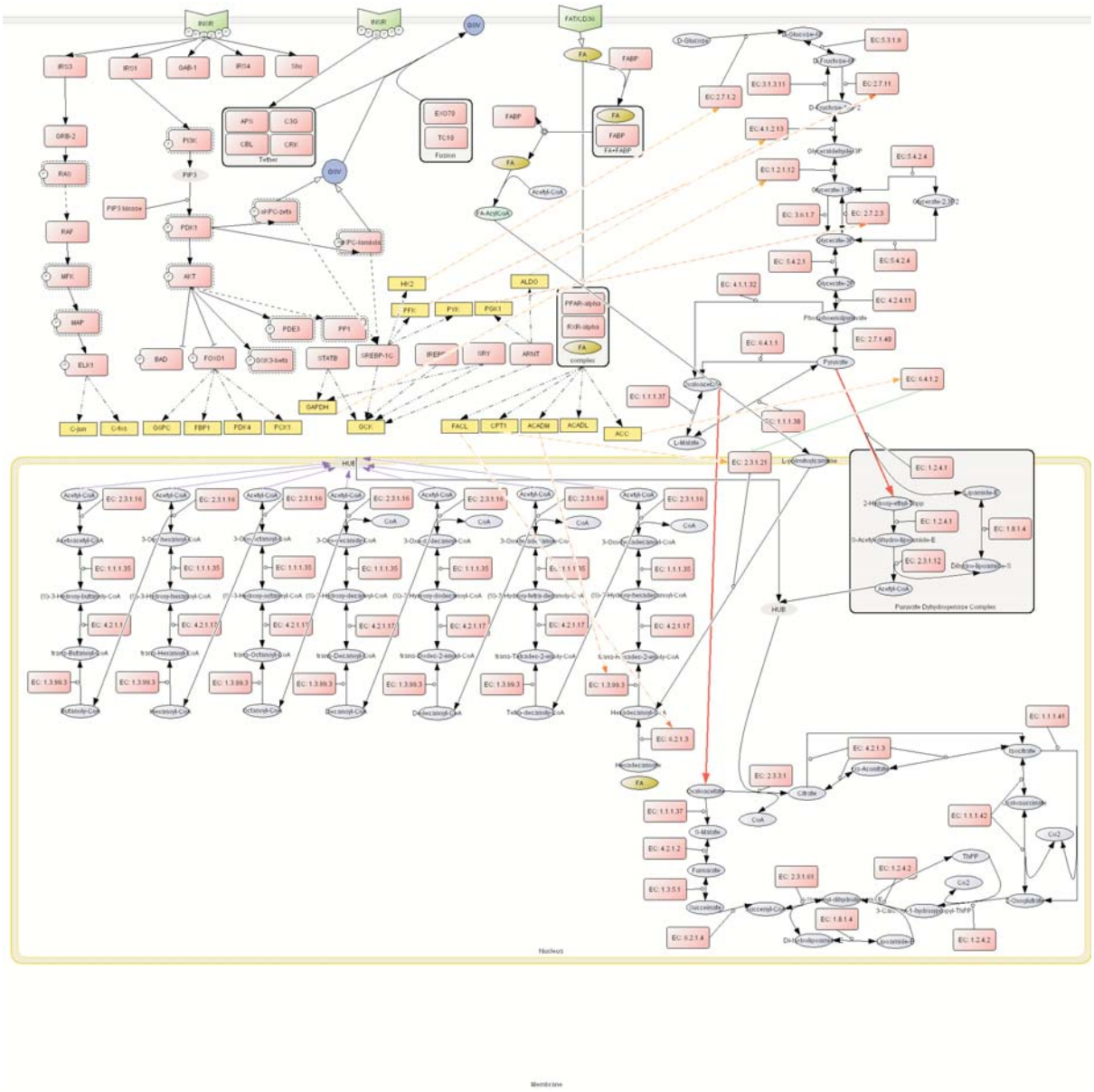


Figure 7.23. This diagram shows insulin and fatty acid fatty acid signaling pathways along with Glycolysis I, Citrate acid cycle, pyruvate metabolism and β-fatty acid oxidation metabolic pathways. Selected genes from transcription regulatory network effected by the downstream signal effectors are depicted in yellow.

Per stearic acid oxidation twenty seven $NADH_2s$ and nine $FADHs$ would be produces meanwhile two $GTPs$ would consume, hence the net production would be ninety seven ATP molecules.

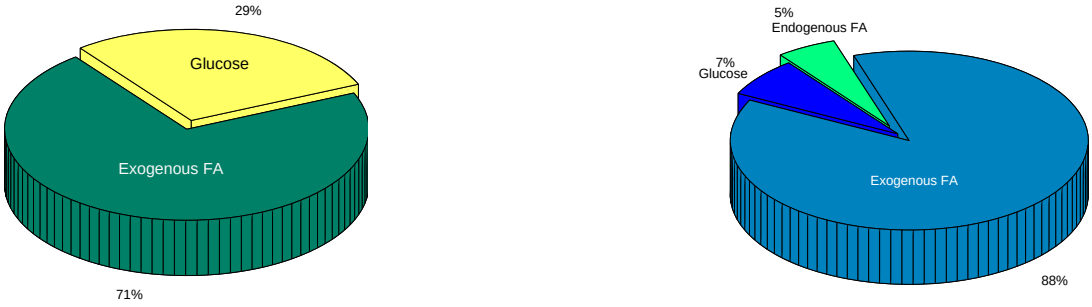


Figure 7.24. The share in ATP production between glycose and fatty acid sources in myocardial cell: in-silico result without endogenous fatty acid resource (left), data from wet lab experiment (right).

7.5.2 Results on cardiac cell energetics

Taking into the account that we had no endogenous fatty acid resources for the in-silico version, the contributions of glucose and fatty acid to the ATP provision in a cardiac myocyte is shown in Fig. 7.24 for in-silico and wet lab experiments.

We also have measured the oxidation rate of both glucose and fatty acid. The rate for fatty acid was about six fold higher than the glucose counter part (Fig. 7.25). These results which is endorsed by the experimental data indicates that our model for fatty acid uptake, transport and oxidation works fairly well.

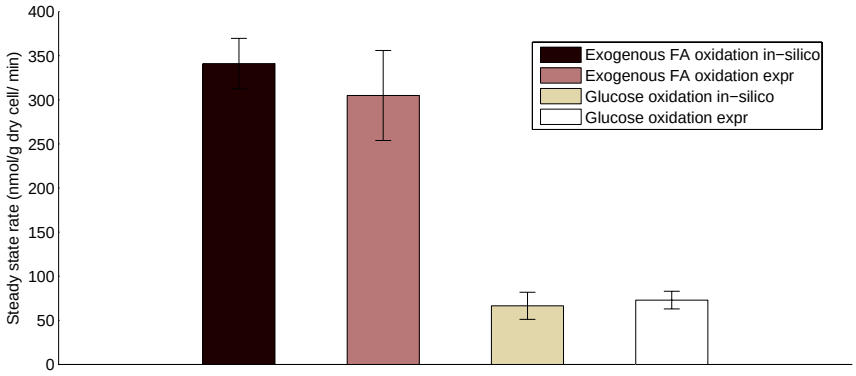


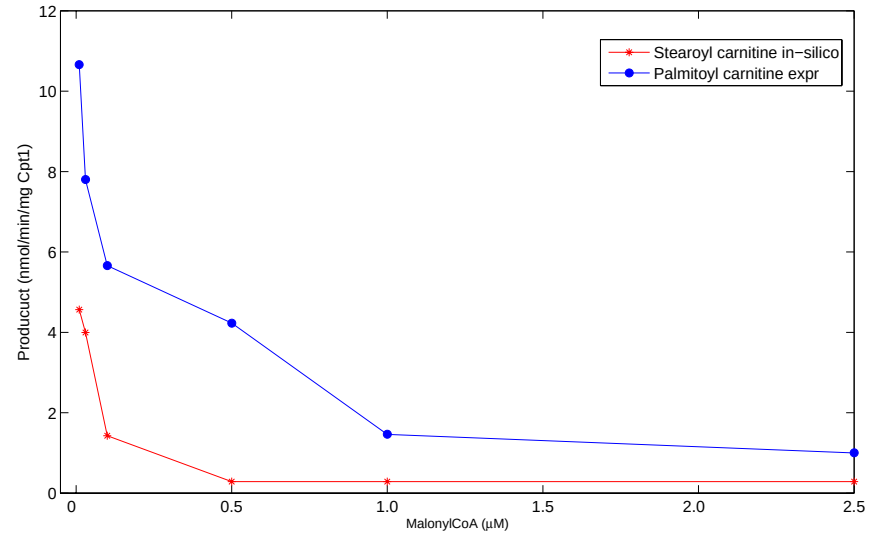
Figure 7.25. Fatty Acid and Glucose oxidation rates; we have up-scaled our results from cell level to the tissue level.

7.5.3 Analysis of the *Malonyl-CoA* regulatory effect

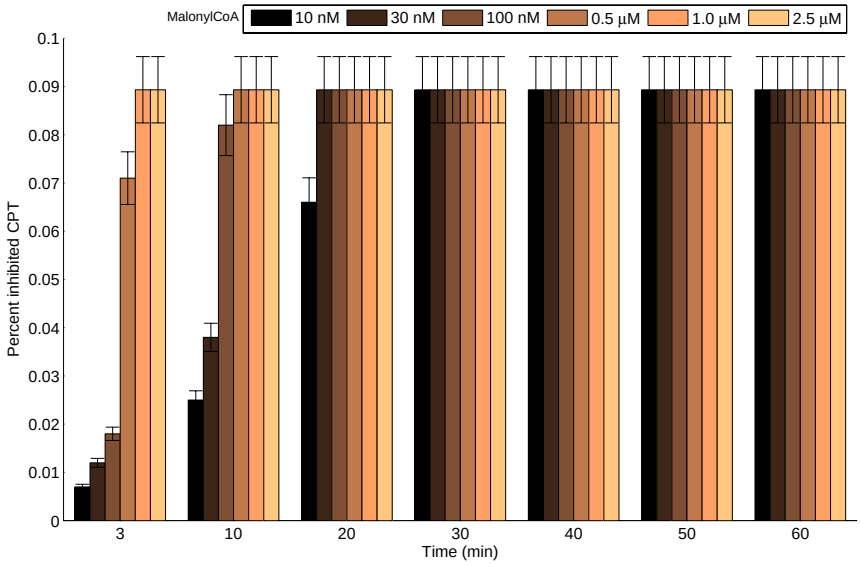
To measure the inhibitory effect of *Malonyl-CoA* on CPT1, we redesigned the above simulation this time by adding different initial concentration for *Malonyl-CoA* and periodically measured the Stearoylcarnitine which is the fatty acid containing complex trafficked across mitochondrial membrane by carnitine palmytolyl transport system (CPTS). Fig. 7.26.(a) shows the abundance of *Malonyl-CoA* significantly reduces the activity of CPTS and consequently inhibits the mitochondrial pathway for fatty acid oxidation. The same chart also compares our in-silico results and the empirical data reported in [184]. Fig. 7.26.(b) depicts the effect of *Malonyl-CoA* concentration on CPT1 inhibition rate.

7.5.4 Analysis of Pyruvate Dehydrogenase Kinase (PDK) activity

As mentioned earlier fatty acid impinges its transcription regulatory effect by activating the proxisome proliferator-activated receptors (PPARs) [171]. The *PDK* is amongst genes up-regulated by PPARS transcription factor and phosphos-deactivates the *PDH*, which results in flux reduction across *R_PDHm* reaction in mitochondria. To quantify this effect we redesigned the experiment 4 with two scenarios: i) mutating the



(a)



(b)

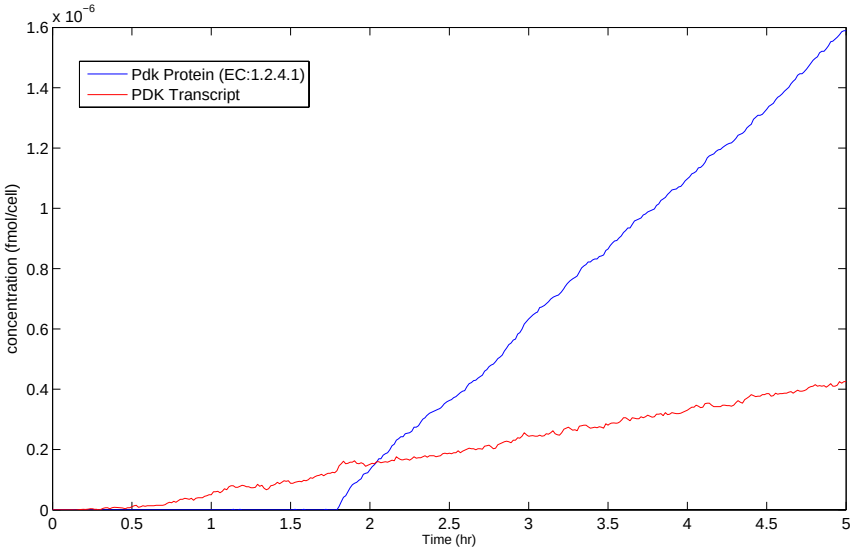
Figure 7.26. (a) Effect of the *Malonyl-CoA* on formation of Stearoylcarnitine complex, for in-silico and wet lab experiments;(b) Dynamics of CPT1 inhibition rate with respect to *Malonyl-CoA* concentration and time.

PDK enzyme so no PDH phospho-inhibition could happen, ii) with wild-type *PDK* and initial *PDK=0*. The evolution of the simulation was continued uninterruptedly until we could observe a noticeable effect of *PDK* on the mitochondrial metabolic reaction for pyruvate decarboxylation.

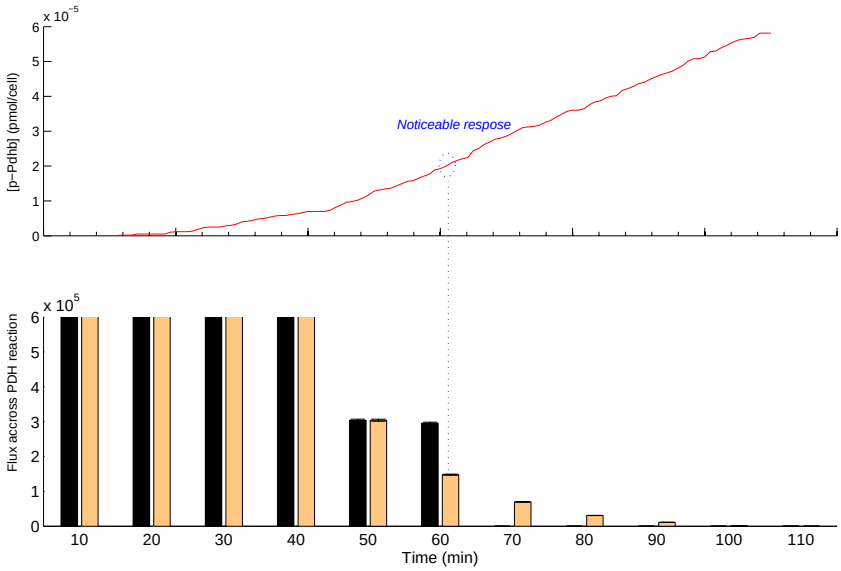
Fig. 7.27.(a) shows the concentrations of *PDH* transcript and *Pdh* linearly increase with time due to the transcriptional activation effect of fatty acid. The effect of pyruvate-dehydrogenase-kinase on the mitochondrial metabolic flux was observed after approximately one hour of simulation (Fig. 7.27.(b)). The decline of flux on pyruvate decarboxylation reaction (*R_PDHm*) which is due to paralyzed PDC would lead to reduction in mitochondrial *Acetyl-CoA* and *citrate* concatenations (TCA cycle precursor substrate and intermediate product) compared to the case where *PDH* is mutated and could no longer inhibit the PDC (Fig. 7.28).

7.6 Experiment 5: Metabolic plasticity of the cardiac myocyte

In this experiment we intend to validate the *eukaSimBioSys* in modeling and simulation of the metabolic plasticity of the cardiac myocyte. Metabolic plasticity is the capacity of a cell to adopt to the available metabolic substrate as the source for its energy. The track of plasticity is best observed by looking into the expression profile of genes contributing a specific substrate metabolism. On the other hand, upon the abundance of a specific metabolic resource the availability of its transport proteins as well as enzymes for its metabolism should be promising. Hence, a higher gene expression profile and consequently transcript concentration is expected for the underlaying genes. Van Bilsen and his colleagues [185] conducted such experiment for the rat heart and identified the expression patterns for some of the genes contributing for glucose and fatty acid oxidations in the rat heart.



(a)



(b)

Figure 7.27. (a)Transcriptional effect of fatty acid on *PDH* transcript and *Pdh* protein synthesis, (b) the *PDH* effect on flux across *R_PDHm* reaction.

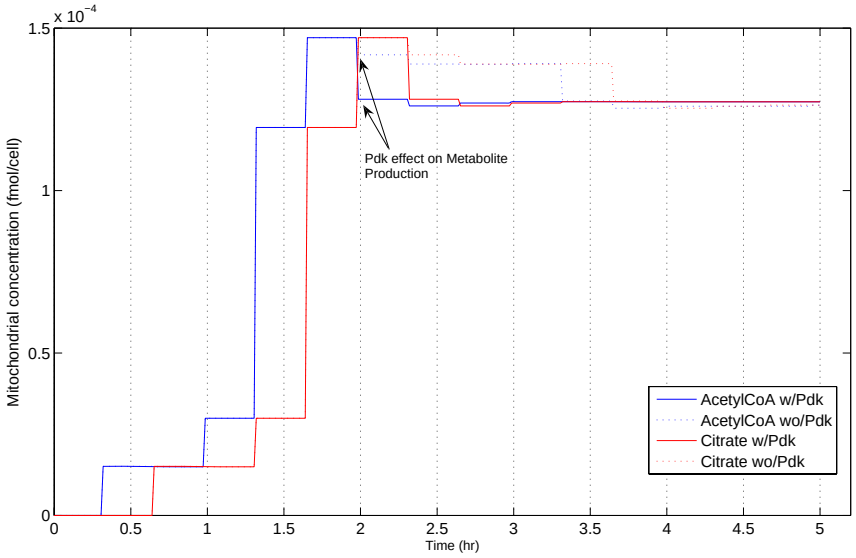


Figure 7.28. Effect of *PDH* on the mitochondrial production of *Acetyl-CoA* and *Citrate*.

In such experiment the cardiac myocytes will be forced to subscribe to metabolism of a certain metabolic substrate i.e. glucose and fatty acid. The former is the result of *feeding* the model animal with glucose rich food and the latter is by letting the species starve (*fasting*) for 40 hours. The starvation will force the body to release the fat stored in adipocytes into the blood. This, would let the other cells (e.g. cardiac myocyte) to uptake and oxidize the fatty acids for their functions which obligates activating the fatty acids-dependent uptake and oxidation pathways.

Conducting this experiment demonstrates the unique capacity of *eukaSimBioSys* in simulating the dynamics of a eukaryotic cell at the system level for 48 hours. To date of this dissertation no in-silico simulation tool with such capacity has been reported elsewhere.

To design the experiment we supplied 1.4 nM of each signaling proteins, 1.4 nM of each metabolic enzymes, and a basal level of transcript for each of the genes listed in Table 6.11. We simulated two scenarios for the myocyte: *normal* and *fasting*. For the

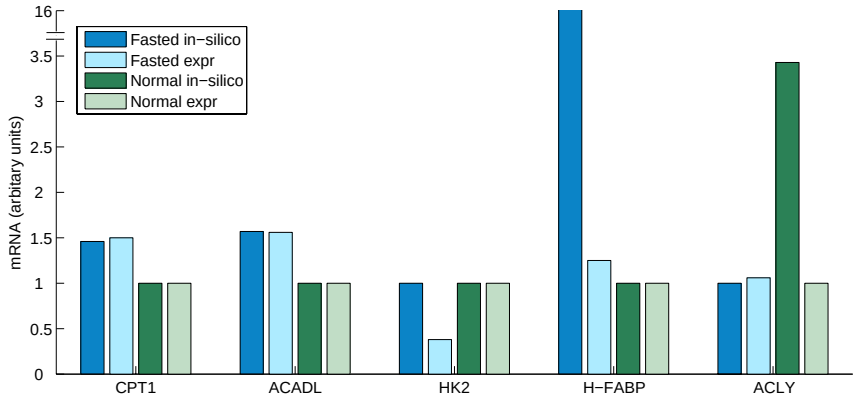


Figure 7.29. Change in the expression profile of selected myocardial genes for *Normal* feeding (after 8 hours) and *Fasting* cells after 48 hours. In-silico result and empirical data are shown in blues and greens, respectively.

former, cell is supplied with resources such that in the course of experiment [Glucose] >> [Fatty acid] and also sufficient source for insulin was available where in the latter case [Fatty acid] >> [Glucose] throughout the experiment.

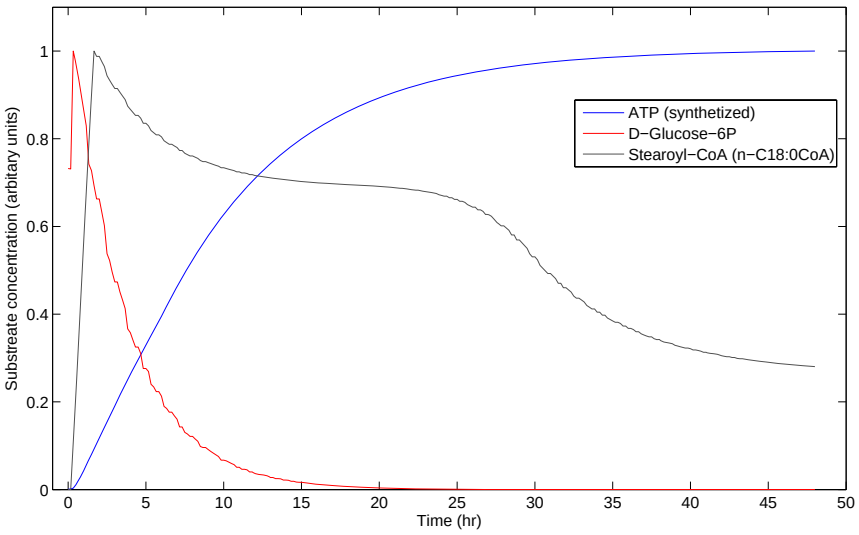
The fold change in concentration of the transcript for those gene whose data could be validated with published data is depicted in Fig. 7.29. These measurement is after 8 and 48 hours of simulation (blue color domain) and bench data (green color domain). As it is observed the *CPT1* which a member of CPTS is increased during the fasting and *ACADL* which is Long-chain specific acyl-CoA dehydrogenase is also induced in the fasting period. The in-silico results shows that HK2 (hexokinase 2) which is a glycolysis pathway metabolic enzyme remains constant during fasting where as the empirical data suggested reduction by half fold for the same period. This can suggest a possible inhibitory regulation which is non-included in our simulation. Also although both results agree on the increase for Fatty acid-binding protein (*FABP*) during *fasting*, however in-silico results show a significantly higher fold in the increase which demands for further

regulatory mechanism not properly handled by the simulation. This proposition stays valid for ATP-citrate synthase ($ACLY$) too, but this time during the *normal* feeding.

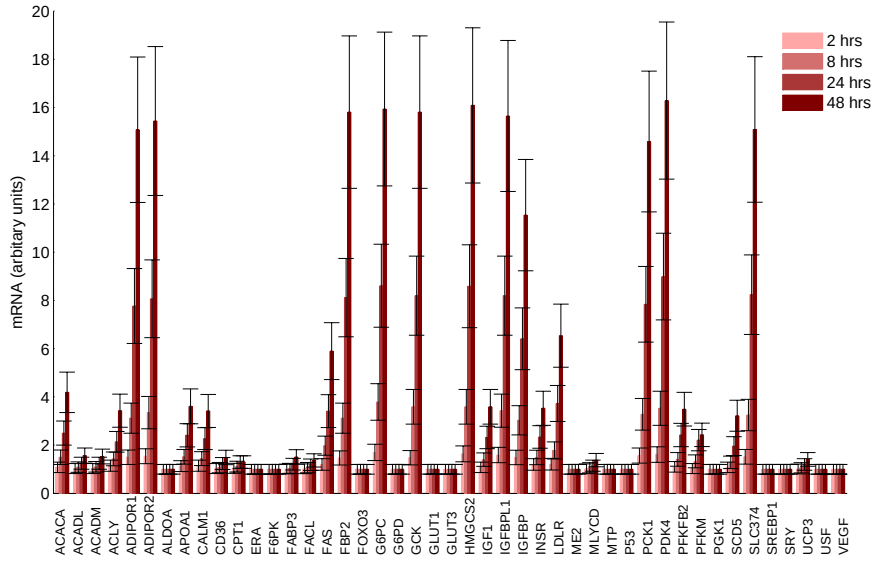
To further observe the plasticity of myocardial cell we conducted another scenario where an 8 hours period of normal feeding followed by 40 hours of fasting. In this in-silico experiment in addition to the gene expression profile we further looked into the metabolic flux, ATP synthesis and some substrates concentration profile during the course of experiment. The concentration of glucose was set such that would last for $\sim$ 8 hours, also the concentration of fatty acid was not unlimited and would gradually decrease.

Fig. 7.30.(a) shows that at beginning both exogenous substrate and fatty acid were highly utilized in energy production of the cell and *ATP* concentration grows exponentially. The decay rate of *D-Glucose-6P* follows an exponential decay which indicates a very high utilization of glucose in cell. After initial raise in the concentration of metabolite substrates which is the result of initial signal impinge on the cell for the hours between 6 to 20 we observe a decline in the slope of *Stearoyl-CoA(18:0CoA)* decay. This smoother slope is the result of negative regulation of *CPT1* by *Malonyl-CoA* as well as marginal inhibitory effect of insulin signal on Fatty acid transport system [184] which we have incorporated in the event network as a slow reaction event on *FAT/CD36*. Once the cellular content of *Malonyl-CoA* reduces, the activity of CPT1 increases and one can see the increase in rate of fatty acid oxidation after the first day (24 hrs).

In Fig. 7.30.(b) we have shown the transcription regulatory effect of current feeding scenario. The genes that induced by *FOXO1* have exponential increase in the expression profile after 5 hours of simulation because the insulin signal which negatively regulates theses genes gradually ceased to exist; although the increase in their expression level was expected where level-fold of their increase subject to further validation with empirical data. Many of the genes involved in fatty acid transport and oxidation pathway show



(a)



(b)

Figure 7.30. Effect of 8 hours period of normal feeding followed by 40 hours of fasting on:(a)concentration of *D-Glucose-6P*(red), *Stearoyl-CoA(18:0CoA)* (gray), and ATP in blue. (b) gene expression profile for all the genes in transcription regulatory network underlaying the current in-silico experiment.

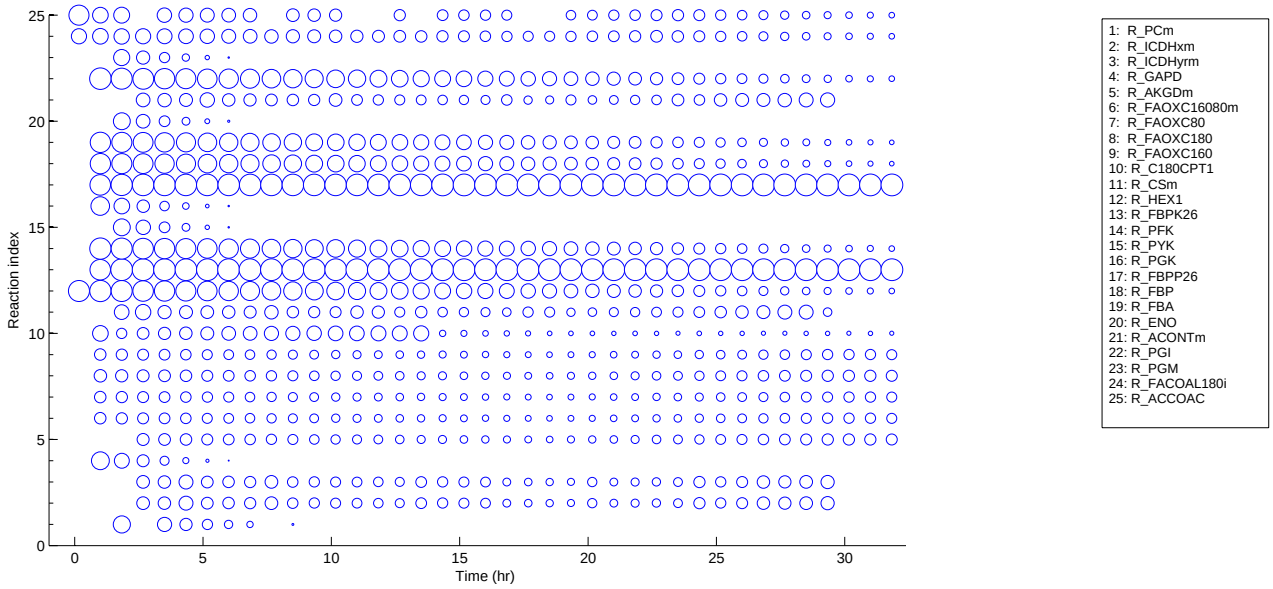


Figure 7.31. Reaction fluxes for 25 active reactions: radius of each circle represents the log value of the flux per reaction, numbers on the y axis correspond to the reaction indices on the list to right of the chart.

a one to two folds increase which is in agreement with the experimental data reported elsewhere. Since the data on negative regulatory effect of transcription factors was very limited, hence we were not successful to show negative regulatory effect on the gene expression profiles.

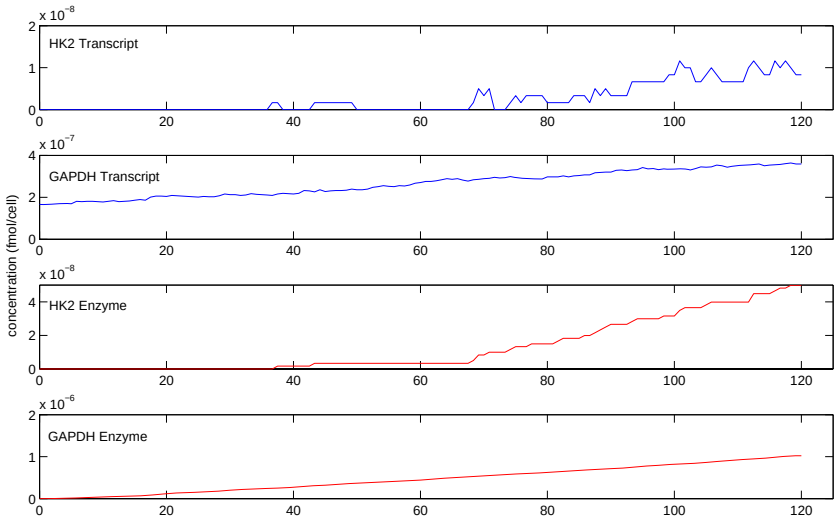
In Fig. 7.31 shows the flux across all active metabolic reaction in Glycolysis I, TCA cycle, pyruvate metabolism, and β-fatty acid oxidation pathways during the course of this experiment. The radius of circles are *log(flux)* for the every 45 minutes, x and y axis are the time and reaction index, respectively. From here we can conclude that the were no flux across glycolysis pathways after 5 hours, where as the flux across fatty acid reactions fluxes sustained during for the entire length of experiment.

7.7 Experiment 6: Null mutation experiment on *HK2* and *GAPDH* genes

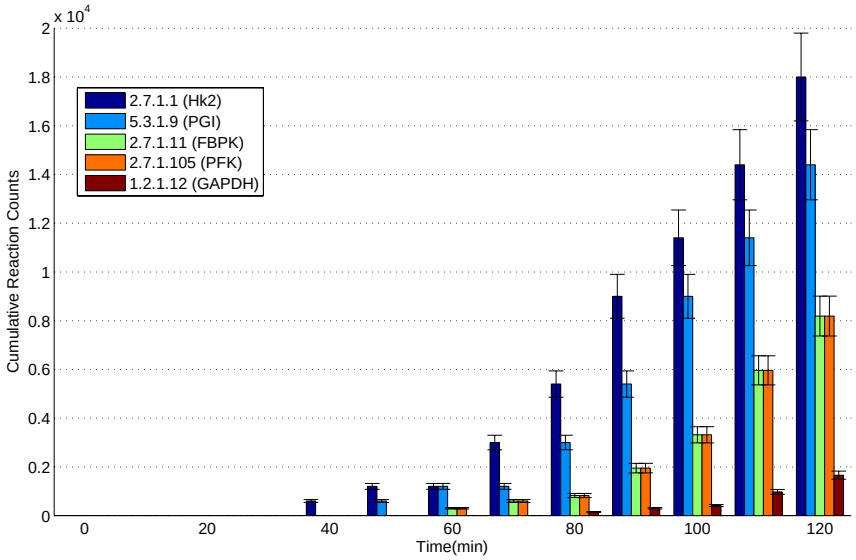
This experiment is only conducted in-silico to demonstrate the capability of the *eukaSimBioSys* in conducting gene null-mutation experiment. For this purpose we designed the experiment such that only the insulin signal and exogenous glucose source exist. The molecular resources supplied are briefly listed as follows: 1.4 nM of each signaling proteins, 1.4 nM of each metabolic enzymes except Hexokinase 2 (*HK2*) and Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) that were set to 0, the initial transcript resources were 0 for all proteins except for $GAPDH$=100, glucose concentration is kept at 5 mM at all times. The experiment was conducted with three different gene regulatory networks (GRN): i)both *HK2* and *GAPDH* genes were included in the GRN , ii) HK2 was removed (nullified) from the GRN, and iii) *GAPDH* gene was removed (nullified) from the GRN.

Removing a gene from GRN avoids that to become the target of any transcription factor, therefore the gene would never be expressed. Furthermore, since we have no initial content for the enzymes encoded by these genes, therefore failure in their expression would halt the glycolytic flux across the reaction catalyzed by that enzyme. We monitored the cellular content of *D-Glucose-6P, Glyceraldehyde*, and $Glycerate - 1,3P_2$ during the 2 hours of the simulation with three different GRNs to learn the gene nullification effect on the metabolic fluxes.

Insulin signal activates the *Srbp1* and *Arnt* which induce the expression of *HK2* and *GAPDH* genes, respectively. The three sub-charts in Fig. 7.32.(a) show the concentrations of HK2-mRNA, GAPDH-mRNA, *Hk2* protein, and *Gapdh* protein versus time. The basal level of *Gapdh* transcript led to faster growth in the cellular content of the enzyme. In Fig. 7.32.(b) we have depicted the cumulative flux for reactions catalyzed by Hexokinase 2 (HK2), Phosphoglucose isomerase (*PGI*), 6-phosphofructo-2-kinase/fructose-2,6-



(a)



(b)

Figure 7.32. (a) Insulin dependent increase in concentrations of *HK2* and *GAPDH* transcripts and proteins,(b) metabolic flux across selected reactions from Glycolysis I pathway with no mutations.

biphosphatase (*PFKFB1*), Phosphofructokinase (*PFK*) and Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*).

To observed the effects of gene nullification, first *HK2* was removed from the GRN and then *GAPDH*, in each case we have monitored the concentrations of the three product substrates of the above three reactions during individual simulation runs. Fig. 7.33 shows the cellular content for three substrates, the name of the gene that was null-mutated is given next to its metabolic reaction product substrate, if the null-mutation occurred. The delay that observed for initial flux across Hexokinase 2 reactions in Figs. 7.32.(b) and 7.33 were due to time required for insulin signal-dependent protein synthesis to produce the enzyme. As expected when the HK2 mutated none of the three product were produced, where as once the *GAPDH* got mutated only the portion of the pathway downstream the point of mutation got halted.

7.8 Summary

In this chapter we initially described the approach that we used for re-scaling the experimental results and listed some the cellular dimensions we used in our manipulations. The reset of the chapter was dedicated to the in-silico experiments for validating the software results with experimental data and also demonstrate its capacity for simulating complex biological networks of a eukaryotic cell. The first experiment showed the metabolic effect of insulin signal on inducting the flux across a metabolic reaction by both increasing the rate of glucose uptake and up-regulating the genes pertaining to its metabolism. In the second experiment a hypothesis testing for the positive role of glucose phosphorylation on the insulin-dependent mTOR signaling in the cardiac myocytes was assessed. Implementation of the feedback loops in a biological system with *eukaSimBioSys* and also a comparison between the results from our modeling approach and ODE based modeling was studied in experiment four. The fifth experiment focused

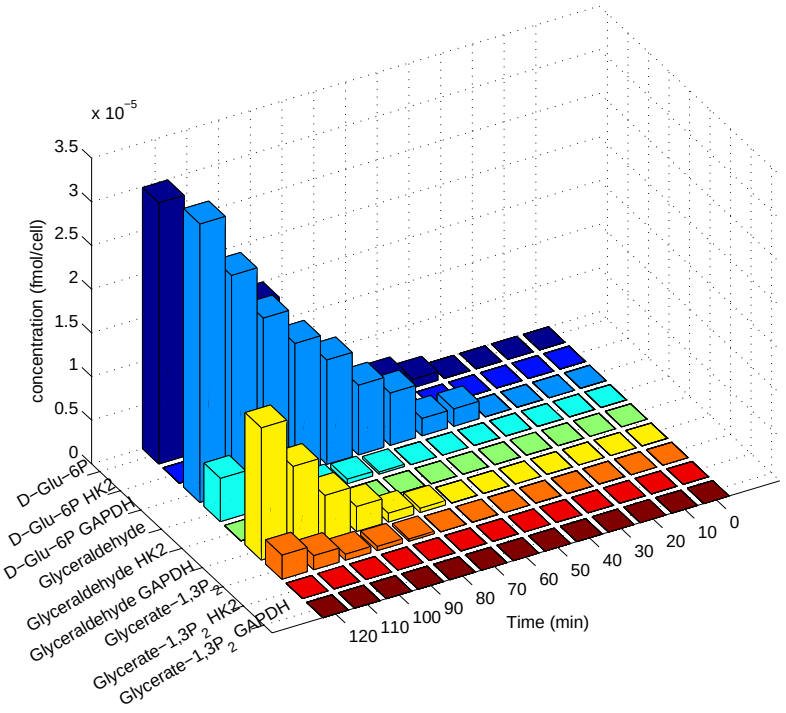


Figure 7.33. Cellular content for *D-Glucose-6P, Glyceraldehyde*, and $Glycerate - 1,3P_2$ for three versions of GRN: none mutated GRN (first set), *HK2* mutated GRN (second set), *GADPH* mutated GRN (third set).

on contributions of glucose and fatty acid in myocardial energetics. The plasticity of cardiac myocytes along with the transcriptional regulatory effects of insulin was studied in experiment six. Meanwhile we also showed the unique capacity of our software in simulating the cardiac myocytes biological networks for a prolong period of time (48 hours). Finally in experiment six we showed how *eukaSimBioSys* could be utilized for the null-mutation studies in biological systems. For the first five experiment in each case we validated the in-silico results with the published empirical data.

CHAPTER 8

CONCLUSION AND FUTURE DIRECTIONS

Discrete event simulation is a well-known discipline in modeling and simulation of complex systems and has been extensively applied in numerous modeling and simulation projects from manufacturing systems, tactical battlefield management, internet traffic engineering, bulk handling supply chains and many more. In this dissertation we endeavored to design a software tool based on this methodology that could be applied in simulating the interactions among biological networks within a eukaryotic cell. The evolution of such platform is stochastic in time and space and is capable to capture the dynamics of a eukaryotic cell at the molecular level. Also since the heart diseases are still ranked number one among all fatal disease worldwide we chose to focus our work on cardiac myocytes. The heart muscles are most promising component in perpetual contracting function of the heart and specifically very well studied tissues. Furthermore, we decided to focus on insulin signaling pathway, since it is critical to metabolism of cardiac myocytes.

For this purpose in chapter 1 we projected all the intracellular interactions into a framework that composed of three interacting networks: Signal Transduction Network, Transcription Regulatory Network and Metabolic Network. We discussed that a cell performs it's functions through the collaboration among one or more of the pathways within theses networks. Each of the cellular pathways handle one or more cellular processes. We identified these cellular processes as bioevents with random execution times. For each bioevent type a physicochemical model is required that can estimate the holding-time (execution) of the underling bioevent. A proper sequence of these events could form a

cellular function. Furthermore, the convolution of holding-times for bioevents pertaining to a cellular function along with changes to underlaying molecular resources counts would picture the dynamics of the cellular function. In the same chapter we reviewed some of the prevalent modeling and simulation techniques in studying biological pathways. We argued that the main reason for their failure in simulating the biological networks at system level is the several orders of difference that exists among the holding-times for various biological processes.

The evolution of the discrete event simulation is made possible through the execution of the individual event. Hence, proper identification of the events is crucial for functionality of the system. In chapter 2 we described the approach for identifying the events in biological network. We proposed the architecture for the *eukaSimBioSys* and discussed the details for stochastic discrete event simulation algorithm which works as the engine of the software. The important concept of compartmentalization and its design issues such as *Metabolic Squad Event* was also discussed in details. We also listed the physicochemical models that were essential for the functionality of *eukaSimBioSys*.

In chapter 3 we focused on building a stochastic parametric model for estimating the temporal behavior of ligand docking to the receptors on the membrane of the cell, the proposed model could be used for both eukaryotic and prokaryotic cells, should the model coupled with the proper parameters set. The proposed model is one of the essential models for *eukaSimBioSys*. This model was used in later chapters to estimate the insulin signal timing or manipulate the glucose and fatty acid uptake rates.

In eukaryotes nucleosomes modulate the access to DNA functional sites. Binding the transcription factors to these sites is precursor to gene expression which is one of the core models of our simulation framework. To model the access to an occluded DNA sites we studied the dynamics of the nucleosome in chapter 4 and proposed a model for accessing the DNA through passive pathway and showed that is an exhaustive process.

A biologically piratical pathway to accesses a nucleosome-blocked DNA site is through the chromatin remodeling process, for which we have proposed a stochastic model in chapter 5. In the same chapter we showed the probability distribution function of the time for chromatin remodeling process follows a gaussian curve. These model were all incorporated as an integral part of *eukaSimBioSys* model repository.

In chapter 6 we covered the remaining materials necessary for completing *eukaSimBioSys* and setting up in-silico experiments. The *'eventology'* of protein synthesis which is composed of serval sub-models including chromatin remodeling, splicing, transcription, translation and nuclear transport was discussed in detail in chapter 6. In this chapter we also abstracted the constitutive splicing as sequence of stochastic miscreants and proposed a mathematical model for its temporal behavior. Also we proposed an abstract mechanism for the set of genes regulated by the same transcription factor, this mechanism gives an even chance to all the genes in underlaying gene-set of a transcription factor to be regulated at each onset of transcription for a fixed period of time.

To determine the dynamics of metabolic reactions and their influences on the molecular resources we adpoted a flux balance analysis approach that captures these dynamics under the umbrella of metabolic event. The insulin signaling pathways and fatty acid signaling pathways were discussed and their *'eventology'* was addressed. The metabolic pathways that involved in this study including glycolysis I, tricarboxylic acid cycle, β-fatty acid oxidation, fructose and mannose metabolism were reviewed in chapter 6. To avoid the stiffness we suggested several design strategies . Persistent reaction model was designed to overcome the resource inconsistency problems that would emerge due to event based view on the signaling networks. Lastly in chapter 6, we described an important step that involved in designing an in-silico experiment with *eukaSimBioSys* and that was defining the event-table, event-map and molecular resources of the experiment based on the equivalent experiment conducted in wet lab.

Validation of the results generated by *eukaSimBioSys* is a vital step for both proofing of our concept in modeling the biological networks as well as the usefulness of the software. We conducted 5 in-silico experiments including: energetics of cardiac myocytes, hypothesis testing for the effect of glucose on mTOR insulin-dependent signaling, cardiac myocyte placticty, analysis of feedback loops within insulin signaling pathways. In all of these experiment we showed that our results stayed in a good agreement with the published data. During one of these experiment we successfully simulated the indicated set cardiac myocytes for 48 hours. A sperate in-silico experiment was conducted to demonstrate the utilization of *eukaSimBioSys* in null-mutation analysis of the genes in the cardiac myocyte.

8.1 Future direction of the research

The interdisciplinary nature of the current research envisions multiple dementiaes for potential extensions in the future. Possible extensions to the current software include, but not limited to, the following topics:

8.1.1 Developing new physicochemical models

Since the physicochemical models of physiological processes form the brain of this framework, therefore having precise efficient models in the model library would guarantee better results. Also in order to utilize this software to simulate the other cellular functions such as cell growth, cell differentiation, cell apotheosis, etc. we need to develop new stochastic event based models that can capture the temporal behavior of the processes involved in these cellular functions.

8.1.2 Genome-scale simulation

Current software has the potential to be upgraded for a genome-scale simulation. However, to have a genome-scale simulation we need a far better and comprehensive regulatory mechanism abstraction that include a reach set of transcription regulation parameters such as upstream/downstream regulator of the genes, loci, chromosomes, second and third level transcription regulatory networks, etc. all of which demands for substantial amount of effort.

8.1.3 Distributed computing capability

Biological networks are highly complex, therefore as the simulation domain grows the computation complexity and memory demands of the software would not necessarily follow a linear growth function. Therefore, it is not only desirable but also necessary to have the simulation run on a grid of processors and benefit form a large pool of memory. This requirement demand for an enhancement to current architecture that grants the above features to the software and promises a high efficiency for the simulation all the times.

8.1.4 High-Throughput and interactive database

eukaSimBioSys has a very high rate of database transactions, therefore an efficient database design along with a high-throughput database engine can significantly improve the performance of the software. Also the underlaying database should have an interactive interface with other databases elsewhere in the world to update its knowledge base with most recent scientific findings that could potentially increase its accuracy.

REFERENCES

[1] C. Bustamanate, C. L. Peterson, B. R. Cairns, S. B. Smith, S. Mihardja, S. W. Grill, A. Saha, C. L. Smith, and Y. Zhang, "Dna transclocation and loop formation mechanism of charomatin remodeling by swi/snf and rsc," *Mol. Cell*, vol. 24, pp. 559–568, 2006.

[2] F. Mohammad-Rafiee, I. M., and H. K. Schiessel, "Theory of nucleosome corkscrew sliding in the presence of synthetic dna ligands," *J. Mol. Biol.*, vol. 344, pp. 47–58, 2004.

[3] (2007) List of sequenced genomes in genome news network. [Online]. Available: http://www.genomenewsnetwork.org/resources/sequenced_genomes/

[4] H. Kitano, "Computational systems biology," *Nature*, vol. 420, pp. 206–210, Nov. 2002.

[5] R. Brent and J. Bruck, "Can computer help to explain biology," *Nature*, vol. 440, pp. 416–417, Mar. 2006.

[6] M. Kanehisa, "Kegg data base," *Novartis found Sym.*, vol. 247. [Online]. Available: http://www.genome.ad.jp/kegg/

[7] L. .Matthews, P. D'Eustachio, M. Gillespie, D. Croft, B. de Bono, G. Gopinath, B. Jassal, S. Lewis, E. Schmidt, I. Vastrik, G. Wu, E. Birney, and L. Stein, "An introduction to the reactome knowledgebase of human biological pathways and processes," *Bioinformatics Primer, NCI/Nature Pathway Interaction Database*, vol. doi:10.1038/pid.2007.3. [Online]. Available: http://www.reactome.org/

[8] M. .Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis, and E. Wingender,

"Transpath: An information resource for storing and visualizing signaling pathways and their pathological aberrations," *Nucleic Acid Reseach*, vol. 34.

[9] P. F. F. R. Driel and P. J. Verschure, "The eukaryotic genome: a system regulated at different hierarchical levels," *Cell Science*, vol. 116, pp. 4067–4075, Sept. 2003.

[10] A. J. Griffiths, W. M. Gelbart, R. C. Lewontin, and J. H. Miller, *Modern Genetic Analysis: Integrating Genes and Genomes*. W. H. Freeman.

[11] V. Matys, O. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. Kel, and E. Wingender, "Transfac and its module transcompel: transcriptional gene regulation in eukaryotes," *Nucleic Acid Reseach*, vol. 34.

[12] T. Ryu, J. Jung, S. Lee, H. J. Nam, S. W. Hong, J. W. Yoo, D. ki Lee, and D. Lee, "bzipdb : A database of regulatory information for human bzip transcription factors," *BMC Genomics*, vol. 8, May 2007. [Online]. Available: http://creativecommons.org/licenses/by/2.0

[13] K. Tuncay, L. E. amd A. A. Haidar, F. Stanley, M. Trelinski, and P. Ortoleva, "Transcriptional regulatory networks via gene ontology and expression data," *in-silico Biology*, vol. 7, Dec. 2006. [Online]. Available: systemsbiology.indiana.edu/trndresults

[14] P. Romero, J. Wagg, M. L. Green, D. Kaiser, M. Krummenacker, and P. D. Karp, "Computational prediction of human metabolic pathways from the complete human genome," *Genome Biology*, vol. 6:R2. [Online]. Available: http://humancyc.org/

[15] B. D. Ventura, C. Lemerle, K. Michalodimitrakis, and L. Serrano, "From in vivo to in silico biology and back," *Nature Reviews*, vol. 433, pp. 527–533, Oct. 2006.

[16] P. Mendes, "Gepasi: A software package for modeling the dynamics, steady states and control of biochemical and other systems," *Comput. Applic. Biosci*, vol. 9.

[17] H. M. Sauro, "Jarnac: a system for interactive metabolic analysis. animating the cellular map," *9th International BioThermoKinetics Meeting (eds: J. H. S. Hofmeyr , J. M. Rohwer, J. L. Snoep) Stellenbosch University Press*, vol. 9.

[18] M. S. Samoilov and A. P. Arkin, "Deviant effects in mulecular reaction," *Nature Computational Biology*, vol. 24, pp. 1235–1240, Oct. 2006.

[19] D. Longo and J. Hasty, "Imaging gene expression: tiny signals make a big noise," *Nat. Chem. Biol.*, vol. 2(4), pp. 181–182, 2006.

[20] K. Wiesenfeld and F. Jaramillo, "Minireview of stochastic c resonance," *Chaos*, vol. 8.

[21] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry.* Amersterdam: North Holland, 2nd edition, 1992.

[22] D. T. Gillespie, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Chem. Phys.*, vol. 115(4).

[23] C. V. Rao and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steadystate assumption: Application to the gillespie algorithm," *J. Chem. Phys.*, vol. 118(11).

[24] K. Burrage, T. Tian, and P. Burrage, "A multi-scaled approach for simulating chemical reaction systems," *Progress in Biophysics and Molecular Biology*, vol. 85.

[25] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Phys. Chem.*, vol. 104.

[26] N. L. Novre and T. S. Shimizu, "Stochsim: modeling of stochastic biomolecular processes," *Bioinformatics*, vol. 17.

[27] H. Salis and Y. Kaznessis, "Accurate hybrid simulation of a system of coupled chemical or biochemical reactions," *J. Chem. Phys.*, vol. 122.

[28] E. L. Haseltine and J. B. Rawlings, "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics," *J. Chem. Phys.*, vol. 117(15).

[29] E. C. A. Alfonsi, G. Turinici, B. D. Ventura, and W. Huisinga, "Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems," *ESAIM Proceeding*, vol. 14.

[30] C. Gadgil, C. H. Lee, and H. G. Othmer, "A stochastic analysis of first-order reaction networks," *Bull. Math. Biol.*, vol. 67.

[31] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *J. Chem. Phys.*, vol. 121.

[32] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22.

[33] ——, "Concerning the validity of the stochastic approach of chemical kinetics," *J. Stat. Phys.*, vol. 16.

[34] A. Chatterjee, K. Mayawala, J. Edwards, and D. G. Vlachos, "Time accelerated monte carlo simulations of biological networks using the binomial tau-leap method," *Bioinformatics*, vol. 21(9).

[35] P. Ghosh, S. Ghosh, K. Basu, and S. Das, "Parametric modeling of protein-dna binding kinetics: A discrete event based simulation approach," *Elsevier Journal on Discrete Applied Mathematics (DAM)*, 2007.

[36] S. Ghosh, P. Ghosh, K. Basu, and S. Das, "Modeling the stochastic dynamics of gene expression in single cells: a birth and death markov chain analysis," *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 308–316, 2007.

[37] P. Ghosh, S. Ghosh, K. Basu, and S. Das, "A diffusion model to estimate the inter-arrival time of charged molecules in stochastic event based modeling of complex biological networks," *IEEE Comp. Systems Biol. Conf*, 2005.

[38] B. Gains, "General systems research: Quo vadis?" *General Systems Yearbook*, vol. 24.

[39] B. Palsson, *Systems biology: properties of reconstructed networks.* Cambridge University Press, 2006.

[40] A. Guffanti, "Modeling molecular networks: a systems biology approach to gene function," *Genome Biology*, vol. 3(10), p. 4031.14031.3, 2002.

[41] N. Price and I. Shmulevich, "Biochemical and statistical network models for systems biology," *Curr. Opin. Biotechnol*, vol. 18(4), p. 365370, 2007.

[42] R. Schwartz, *Biological modeling and simulation: a survey of practical models, algorithms, and numerical methods.* MIT Press, 2008.

[43] S. Ramsey, D. Orell, and H. Bolouri, "Dizzy: Stochastic simulation of large scale genetic regulatory networks," *J. Comput. Biol.*, vol. 3(2), pp. 415–36, Apr. 2005.

[44] C. M. Firth, *Stochastic Simulation of Cell Signalling Pathways.* PhD thesis, University of Cambridge.

[45] D. Ridgway, G. Broderick, and M. J. Ellison, "Accommodating space, time and randomness in network simulation," *Curr. Opin. Biotech.*, vol. 17(5), pp. 493–498, Oct. 2005.

[46] S. Efroni, D. Harel, and I. Cohen, "Efroni, s., d. harel, and i. cohen. 2005. reactive animation: Realistic modeling of complex dynamic systems," *IEEE Computer magazine*, vol. 38(1).

[47] M. Ginkel, A. Kremling, T. Nutsch, R. Rehner, and E. D. Gilles, "Modular modeling of cellular systems with promot/diva," *Bioinformatics*, vol. 19, pp. 1169–1176, Oct. 2003.

[48] C. van Gend and U. Kummer, "Stode automatic stochastic simulation of systems described by differential equations," *Int. Conf. Sys. Biol.*

[49] H. de Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *J. Comp. Biol.*, vol. 9(1), pp. 67–103, Oct. 2002.

[50] P. J. Goss and J. Peccoud, "Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets," *Proc. Natl. Acad. Sci.*, vol. 95(12), pp. 6750–6755, Oct. 1998.

[51] M. Calder, S. Gilmore, and J. Hillston, "Modelling the influence of rkip on the erk signalling pathway using the stochastic process algebra pepa," *Trans. Comp. Sys. Biol.*, vol. 4230, pp. 1–23, 2006.

[52] C. Kuttler, "Simulating bacterial transcription and translation in a stochastic pi calculus," *Trans. Comp. Sys. Biol.*, vol. 4220, pp. 113–149, 2006.

[53] A. M. Uhrmacher and C. Priami, "Discrete event systems specification in systems biology - a discussion of stochastic pi calculus and devs," *Proc. Winter Sim. Conf.*

[54] ——, "Concepts of object and agent oriented simulation," *Trans. on SCS*, vol. 14(2), pp. 59–67, 1997.

[55] T. Emonet, C. Macal, M. North, C. Wickersham, and P. Cluzel, "Agentcell: A digital single-cell assay for bacterial chemotaxis," *bioinformatics*, vol. 21(11), pp. 2714–2721, 2005.

[56] B. P. Zeigler, T. G. Kim, and H. Praehofer, *Theory of Modeling and Simulation.* Academic Press, 2_{nd} Ed.

[57] S. Ghosh, K. Basu, S. Daefler, and S. Das, "isimbiosys: A discrete event simulation platform for 'in silico' study of biological systems," *Proc. 39th IEEE Ann. Symp. on Sim.*, pp. 204–213, 2006.

[58] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell.* US: Garland Publishing Inc; 4Rev Ed edition, 2002.

[59] H. P. Leeuwen and W. Koster, *Physicochemical Kinetics and Transport at Bioinformatics.* England: John Wiley, 2004.

[60] H. C. Berg, *Random Walks in Biology.* NJ: Princeton Univ. Press, 1983.

[61] (2004) From mathworld–a wolfram web resource,. [Online]. Available: http://mathworld.wolfram.com/SphericalCap.html

[62] C. Fall, E. Marland, J. Wagner, and J. Tyson, *Interdisciplinary Applied Mathematics.* NY: Springer Verlag, 2002.

[63] G. Gottschalk, *Bacterial Metabolism Second Edition.* NY: Springer Verlag, 1986.

[64] F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbargar, *Escherichia Coli and Salmonella Tyhpimutium Cellular and Molecular Biology Vol. I.* DC: American Socienty of Mirobiology, 1987.

[65] P. D. Karp, M. Riley, S. M. Paley, and M. K. A. Pellegrini-Toole, "Ecocyc: Encyclopedia of escherichia coli genes and metabolism," *Nucliec Acids Res.*, vol. 26, pp. 50–55, 2006. [Online]. Available: http://ecocyc.org/

[66] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D. S. Wishart, "The cybercell database (ccdb): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of escherichia coli," *Nucleic Acids Res.*, vol. 32, pp. 293–295, Jan. 2004. [Online]. Available: http://redpoll.pharmacy.ualberta.ca/CCDB/

[67] A. H. Romano, S. J. Eberhard, S. L. Dingle, and T. D. McDowell, "Distribution of the phosphoenolpyruvate: Glucose phosphotransferase system in bacteria," *Biotechnology*, vol. 104, pp. 808–813, Nov. 1970.

[68] J. M. Rohwer, N. D. Meadowi, S. Rosemani, H. V. Westerhoff, and P. W. Postma, "Understanding glucose transport by the bacterial phosphoenolpyruvate: Glycose phosphotransferase system on the basis of kinetic measurements in-vitro," *Biol. Chemistry*, vol. 275(45), p. 3490934921, Nov. 2000.

[69] T. Richmond and C. A. Davey, "The structure of dna in nucleosome core," *Nature*, vol. 423, pp. 145–150, 2003.

[70] G. Li, M. Levitus, C. Bustamante, and J. Widom, "Rapid spontaneous accessibility of nucleosomal dna," *Nature Stru. & Mol. Biol.*, vol. 12, pp. 46–53, 2005.

[71] D. A. Beard and T. Schlick, "Computational modeling predicts the structure and dynamics of chromatin fiber," *Elsevier J. Strcu.*, vol. 9, pp. 105–114, 2001.

[72] B. Levin, *Genes VIII*. NJ: Pearson Prentice Hall, 2004.

[73] K. Luger, A. W. Mader, , R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 resolution," *Nature*, vol. 389, pp. 251–260, 1997.

[74] A. Flaus and T. Owen-Hughes, "Mechanism for nucleosome mobilization," *J. Biopolymers*, vol. 66, pp. 563–578, 2003.

[75] M. Kobor, S. Venkatasubrahmanyam, M. D. Meneghini, J. W. Gin, J. L. Jennings, A. J. Link, H. D. Madhani, and J. Rine, "A protein complex containing the conserved swi2/snf2-related atpase swr1p deposits histone variant h2a.z into euchromatin," *PLoS Biol.*, vol. 2E131, pp. 597–599, 2004.

[76] A. Saha, J. Wittmeyer, and B. R. Cairns, "Cairns chromatin remodelling: the industrial revolution of dna around histones," *Nature Rev. Mol. Cell Biol.*, vol. 7, pp. 437–447, 2004.

[77] X. Guo, K. Tatsuoka, and A. R. Liu, "Histone acetylation and transcriptional regulation in the genome of saccharomyces cerevisiae," *Bioinformatics*, vol. 22, pp. 392–399, 2006.

[78] G. J. Narlikar, H. Fan, and R. E. Kingston, "Cooperation between complexes that regulate chromatin structure and transcription," *Cell*, vol. 108, pp. 475–487, 2002.

[79] S. Pennings, G. Meersseman, and E. M. Bradbury, "Mobility of positioned nucleosomes on 5 s rdna," *J. Mol. Biol.*, vol. 220(1), pp. 101–110, 1991.

[80] P. Beard, "Mobility of histones on the chromosome of simian virus 40," *Cell*, vol. 15(3), pp. 955–967, 1978.

[81] J. Mellor, "The dynamics of chromatin remodeling at promoters," *J. Molecular Cell*, vol. 19, pp. 147–157, 2005.

[82] L. V. Yakushevich, *Nonlinear Physics of DNA*. Germany: Wiley-VCH, 2004.

[83] H. Schiessel, "The physics of chromatin," *Max-Planck-Institut fur Polymerforschung, Theory Group*, 2003.

[84] I. Kulic and H. Schiessel, "Chromatin dynamics: nucleosome go mobile through twist defects," *Phy. rev. lett.*, vol. 91(14), 2003.

[85] P. Ghosh, S. Ghosh, K. Basu, and S. K. Das, "Holding time estimation for reactions in stochastic event-based simulation of complex biological systems," *Elsevier Simulation Modelling Practice and Theory*.

[86] I. Kulic and H. Schiessel, "Nucleosome repositioning via loop formation," *J. Biophys.*, vol. 84, pp. 3197–3211, 2003.

[87] L. Kleinrock, *Queueing Systems, Vol. I: Theory*. NY: Wiley, 1975.

[88] Y. S. Kivshar, O. M. Braun, and J. S. Kivsar, *The Frenkel-Kontorova Model: Concepts, Methods, and Applications*. Springer, 2004.

[89] K. Murthy and K. W. Kehr, "Mean first-passage time of random walks on a random lattice," *Phys. rev. A.*, vol. 40, p. 2082, 1989.

[90] H. Schiessel, J. Widom, R. F. Bruinsma, and W. M. Gelbart, "Polymer reptation and nucleosome repositioning," *Phy. rev. lett.*, vol. 86(19), pp. 4414–4417, 2001.

[91] S. Aoyagi, G. J. Narlikar, C. Zheng, S. Sif, R. E. Kingston, and J. J. J. J. Hayes, "Nucleosome remodeling by humans swi/snf complex requires transient global disruption of histone-dna interaction," *Mol. Cell Biol.*, vol. 22, pp. 3653–3662, 2002.

[92] T. Oven-Haghes, D. M. J. Lilley, P. A. Wade, R. E. Kingston, M. Phelan, and K. Havas, "Generation of superhelical torsion by atp-dependent chromatin remodeling activities," *Cell*, vol. 103, pp. 1133–1142, 2000.

[93] A. Saha, J. Wittmeyer, and B. Cairns, "Chomatin remodeling the industrial revolution of dna arround histones," *Nature Rew. Mol. Cell Biol.*, vol. 7, pp. 437–447, 2006.

[94] J. L. Workman, K. E. Neely, A. H. Hassan, and M. Vignali, "Atp-dependent chromatin-remodeling complexes," *Mol. and Cell Biol.*, vol. 20, pp. 1899–1910, 2000.

[95] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang, and J. Widom, "A genomic code for nucleosome positioning," *Nature*, vol. 442, pp. 772–778, 2006.

[96] G. R. Schenitzeler and N. P. N. P. Ulyanova, "Human swi/snf generates abundant, structuraly altered dinucleosome on poly nucleosomal templates," *Mol. and Cell Biol.*, vol. 25, pp. 11 156–11 170, 2005.

[97] C. L. Peterson and C. L. C. L. Smith, "A conserved swi2/snf2 atpase motif couples atp hydrolysis to chromatin remodeling," *Mol. and cell Biol.*, vol. 25, pp. 5880–5892, 2005.

[98] P. Wang, L. Wei, S. Dou, and P. Xie, "Brownian dynamics simulation of directional sliding of histone octamers caused by dna bending," *Phys. Rev.*, vol. 73, pp. 051 909:1–7, 2006.

[99] A. E. Gorbalenya and E. V. Koonin, "Helicase: amin acid sequence comparisons and structure-function relationships," *Curr. Opin. Struct. Biol.*, vol. 3, pp. 419–429, 1993.

[100] B. Bartholomew, S. R. Kassabov, B. Zheng, and J. Persinger, "Swi/snf unwraps, slides and rewraps the nucleosome," *Moll Cell*, vol. 11, pp. 391–403, 2003.

[101] A. Saha, J. Wittmeyer, and B. Cairns, "Chromatin remodeling through directional dna translocation from an internal nucleosomal site," *Nature Struc. and Mol. Biol.*, vol. 12, pp. 744–757, 2005.

[102] B. Bartholomew, S. R. Kassabov, B. Zheng, J. Persinger, and M. Zofall, "Chromatin remodeling by iswi2 and swi/snf requires dna translocation inside the nucleosome," *Nature Struc. and Mol. biol.*, vol. 13(4), pp. 339–346, 2006.

[103] M. M. Grohima, M. G. Munteanu, A. Garielian, and S. Pongor, "Anisotropic elastic bending model of dna," *J. Biol. Phys*, vol. 22, pp. 227–243, 1996.

[104] H. Schott and H. Eckstein, "Studies on interactions between immobilized lysine residues and oligomers of thymidylic and deoxyadenylic acids," *Eur. J. Biochem.*, vol. 104, pp. 79–84, 1980.

[105] D. J. Wilkinson, *Stochastic Modeling for Systems Biology.* NY: Chapman and Hall/CRC, 2006.

[106] B. Bernstein, C. Liu, E. Humphrey, E. Perlstein, and S. Schreiber, "Global nucleosome occupancy in yeast," *Genome biology*, vol. 5(9), pp. 62.1–62.11, Aug. 2004.

[107] S. L. Smith, "Functional and structural analysis of the yeast swi/snf complex," *Ph.D. Dissertation, Univ. of Mass. Med. School*, 2004.

[108] I. Kodde, J. van der Stok, R. Smolenski, and J. de Jong, "Metabolic and genetic regulation of cardiac energy substrate preference," *Comparative Biochemistary and Physiology*, vol. 146, pp. 26–39, Oct. 2006.

[109] G. Gould and G. Holman, "The glucose transporter family: structure, function and tissue specific expression," *Biochemistry*, vol. 295(2), pp. 329–341, 1993.

[110] A. He, X. Liu, L. Liu, Y. Chang, and F. Fang, "How many signals impinge on glut4 activation by insulin," *Cellular Signalling*, vol. 19, pp. 1–7, June 2007.

[111] L. Chang, S. Chiang, and A. Saltiel, "Insulin signaling the regulation of glucose transport," *Molecular Medicine*, vol. 10, pp. 7–12, July 2004.

[112] R. Watson and J. Pessin, "Glut4 translocation: The last 200 nanomters," *Cellular Signalling*, vol. 19, pp. 2209–2217, July 2007.

[113] G. Richieri, A. Anel, and A. Kleinfeld, "Interaction of long chian fatty acids and albumin: determination of free fatty acid levels using the fluorescent probe adifab," *Biochemistry*, vol. 329, pp. 7574–7580, 1993.

[114] D. Koonen, J. Glatz, A. Bonen, and J. Luiken, "Long chain fatty acid uptake and fat/cd36 trasnlocation in heart and skeletal muscle," *Biochim Biophys Acta.*, vol. 1736(3), pp. 163–180, October 2005.

[115] G. van der Vusse, M. van Bilsen, J. Glatz, D. Hasselbaink, and J. Luiken, "Critical steps in fatty acid uptake and utilization," *Mol. Cell. Biochem.*, vol. 254, pp. 311–318, 2002.

[116] J. Bonnefont, F. Demaugre, C. Prip-Buus, J. Saudubray, M. Brevit, N. Abadi, and L. Thuillier, "Carnitine palmitoyltransferase deficiencies," *Mol. Genet. Metab.*, vol. 68, pp. 424–440, 1999.

[117] V. Zammit, "Carnitine acyltransferase functional signifigance of subcellular distribution and membrane topology," *Prog. Lipid Res.*, vol. 38, pp. 199–224, 1999.

[118] (2007) The bigg database. [Online]. Available: http://bigg.ucsd.edu

[119] N. Duarte, S. Becker, N. Jamshidi, I. Thiele, M. Mo, T. Vo, R. Srivas, and B. Palsson, "Global reconstruction of the human metabolic network based on genomic and bibliomic data," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104(6), pp. 1777–1782, 2007.

[120] F. Gunstone, *The lipid hand book.* CRC 3rd edition.

[121] ——, *Fatty Acid and Lipid Chemistry.* Springer Verlag.

[122] H. Schulz, "Regulation of fatty acid oxidation in heart," *Nutrition*, pp. 165–171, 1994.

[123] P. Rinaldo and D. Matern, "Fatty acid oxidation disorders," *Annu. Rev. Physiol.*, vol. 64, p. 477502, 2002.

[124] K. Kauffman, P. Prakash, and J. Edwards, "Advances in flux balance analysis," *Elsevier J. Biochemical Engineering*, vol. 14, pp. 491–496, 2003.

[125] M. Covert, C. Schilling, and B. Palsson, "Regulation of gene expression in flux balance models of metabolism," *Theoritical Biology*, vol. 213, pp. 73–88, 2001.

[126] M. Covert and B. Palsson, "Transcriptional regulation in constraints-based metabolic models of escherichia coli," *Biological Chemistry*, vol. 277, pp. 28 058–28 064, May 2002.

[127] ——, "Constraints-based models: regulation of gene expression reduces the steady-state solution space," *Theoritical Biology*, vol. 221, pp. 309–325, 2003.

[128] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, pp. 7695–7700, 2003.

[129] M. Covert and B. Palsson, "Integrating high-throughput and computational data elucidates bacterial networks," *Nature*, vol. 429, pp. 92–96, 2004.

[130] N. Hernandez and W. Keller, "Splicing of in vitro synthesized messenger rna precursors in hela cell extracts," *Cell*, vol. 35, pp. 89–99, 1983.

[131] A. Krainer, T. Maniatis, B. Ruskin, and M. Green, "Normal and mutant human β-globin pre-mrnas are faithfully and efficiently spliced in vitro," *Cell*, vol. 36, pp. 993–1005, 1984.

[132] R. Lin, A. Newman, S. Cheng, and J. Abelson, "Yeast mrnasp licing in vitro," *Biological Chemistry*, vol. 260, pp. 14 780–14 792, 1985.

[133] M. Jurica and M. Moore, "Pre-mrna splicing: awash in a sea of proteins," *Biological Chemistry*, vol. 12(1), pp. 5–14, 2003.

[134] K. Neugebauer, "On the importance of being co-transcriptional," *Cell*, vol. 115, pp. 3865–3871, 2002.

[135] P. Lopez and B. Sraphin, "Yidb: the yeast intron database," *Nucleic Acids Research*, vol. 28(1), pp. 85–86, 2000.

[136] A. Zaug, P. Grabowski, and T. Cech, "Autocatalytic cyclization of an excised intervening sequence rna is a cleavage-ligation reaction," *Nature*, vol. 301, pp. 578–583, 1983.

[137] J. Sanford and J. Caceres, "Pre-mrna splicing: life at the centre of the central dogma," *J. Cell Science*, vol. 117, pp. 6261–6263, 2004.

[138] B. Modrek and C. Lee, "A genomic view of alternative splicing," *Nature Gentics*, vol. 30, pp. 13–19, 2002.

[139] M. Green, "Biochemical mechanisms of constitutive and regulated pre-mrna splicing," *Annu. Rev. Cell Biol.*, vol. 7, pp. 559–599, 1991.

[140] ——, "pre-mrna splicing," *Annu. Rev. Genetics*, vol. 20, pp. 671–708, 1986.

[141] M. Slutsky and L. Mirny, "Kinetics of protein-dna interaction: Facilitated target location in sequence-dependent potential," *Biophys. J.*, vol. 87, pp. 4021–4035, 2004.

[142] S. Mohr, M. Matsuura, P. Perlman, and A. Lambowitz, "A dead-box protein alone promotes group ii intron splicing and reverse splicing by acting as an rna chaperone," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 23(10), p. 35693574, 2006.

[143] H. Kim and J. Yin, "Effects of rna splicing and post-transcriptional regulation on hiv-1 growth: a quantitative and integrated perspective," *IEE Proc. Syst. Biol.*, vol. 152(3), pp. 138–152, 2005.

[144] S. Singh, H. Yang, M. Chena, and S. Yu, "A kinetic-dynamic model for regulatory rna processing," *Journal of Biotechnology*, vol. 127, pp. 488–495, 2006.

[145] A. Audibert, D. Weil, and F. Dautry, "In vivo kinetics of mrna splicing and transport in mammalian cells," *Molc. and Cell. Biol.*, vol. 22(19), p. 67066718, 2002.

[146] B. Alberts, D. B. J. Lewis, K. R. M. Raff, and J. D. Watson, *Molecular biology of the Cell third edition.* New York: Garland Publishing Inc, 1994.

[147] O. Flores, H. Lu, and D. Reinberg, "Factors involved in specific transcription by mammalian rna polymerase ii," *Biol. Chemistry*, vol. 267(4), pp. 2786–2793, 1992.

[148] D. Bushnell, K. Westover, R. Davis, and R. Kornberg, "Structural basis of transcription: an rna polymerase ii-tfiib cocrystal at 4.5 angstroms," *Science*, vol. 303, pp. 983–988, 2004.

[149] H. Boeger, D. Bushnell, R. Davis, J. Griesenbeck, Y. Lorch, J. Strattan, K. Westover, and R. Kornberg, "Structural basis of eukaryotic gene transcription," *FEBS Letters*, vol. 579, pp. 899–903, 2005.

[150] R. J. S. III, R. Belotserkovskaya, and D. Reinberg, "Elongation by rna polymerase ii:the short and long of it," *Genes & Dev.*, vol. 18, pp. 2437–2468, 2004.

[151] T. Lee and R. Young, "Regulation of gene expression by tbp-associated proteins," *Genes & Dev.*, vol. 12, pp. 1398–1408, 1998.

[152] J. Perez-Ortin, P. Alepuz, and J. Moreno, "Genomics and gene transcription kinetics in yeast," *TRENDS in Genetics*, vol. 23(5), pp. 250–257, 2007.

[153] A. Jarmolowski, W. Boelens, E. Izaurralde, and I. Mattaj, "Nuclear export of different classes of rna is mediated by specific factors," *Cell Biology*, vol. 124, pp. 627–635, 1994.

[154] C. Zhang, K. Zobeck, and Z. Burton, "Human rna polymerase ii elongation in slow motion: role of the tfiif rap74α1 helix in nucleoside triphosphate-driven translocation," *Molec. and Cell. Biology*, vol. 25(9), pp. 3583–3595, 2005.

[155] U. Kubitscheck, D. Grnwald, A. Hoekstra, D. Rohleder, T. Kues, J. Siebrasse, and R. Peters, "Nuclear transport of single molecules: dwell times at the nuclear pore complex," *Cell Biology*, vol. 168(2), pp. 233–243, 2005.

[156] R. Jackson, "Alternative mechanisms of initiating translation of mammalian mrnas," *Biochemical Society Lectures*, pp. 1231–1241, 2005.

[157] T. V. Pestova, V. Kolupaeva, I. Lomakin, E. Pilipenko, I. Shatsky, V. Agol, and C. Hellen, "Molecular mechanisms of translation initiation in eukaryotes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98(13), pp. 7029–7036, 2001.

[158] C. Hellen and T. V. Pestova, "Translation initiation: Molecular mechanisms in eukaryotes," *Encyclopedia of Life Sciences*, 2006.

[159] P. Hilleren and R. Parker, "mrna surveillance in eukaryotes: kinetic proofreading of proper translation termination as assessed by mrnp domain organization," *RNA Journal*, vol. 5, pp. 711–719, 1999.

[160] J. Lorsch and D. Herschlag, "Kinetic dissection of fundamental processes of eukaryotic translation initiation in vitro," *The EMBO Journal*, vol. 18(23), pp. 6705–6717, 1999.

[161] J. Fernandez, I.Yaman, C. Huang, H. Liu, A. Lopez, A. Komar, M. Caprara, W. Merrick, M. Snider, R. Kaufman, W. Lamers, and M. Hatzoglou, "Ribosome stalling regulates ires-mediated translation in eukaryotes, a parallel to prokaryotic attenuation," *Molecular Cell*, vol. 17, pp. 405–416, 2005.

[162] P. Muller and H. Trachsel, "Translation and regulation of translation in the yeast saccharomyces cerevisiae," *Eur. J. Biochem*, vol. 191, pp. 257–261, 1990.

[163] D. Elliott and M. Rosbash, "Yeast pre-mrna is composed of two populations with distinct kinetic properties," *Eeperimental cell Research*, vol. 229, pp. 181–188, 1996.

[164] D. Cao and R. Parker, "Kinetic dissection of fundamental processes of eukaryotic translation initiation in vitro," *RNA Jpurnal*, vol. 7, p. 11921212, 2001.

[165] N. Shulga, P. Roberts, Z. Gu, L. Spitz, M. Tabb, M. Nomura, and D. Goldfarb, "In vivo nuclear transport kinetics in saccharomyces cerevisiae: a role for heat shock protein 70 during targeting and translocation," *Cell Biology*, vol. 135(2), pp. 329–339, 1996.

[166] A. Belle, A. Tanay, L. Bitincka, R. Shamir, , and E. OShea, "Quantification of protein half-lives in the budding yeast proteome," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103(35), p. 1300413009, 2006.

[167] R. Brownsey, A. Boone, and M. Allard, "Action of insulin on mammalian heart: metabolism, pathology and biochemical mechanism," *Cardiovascular Research*, vol. 34, p. 324, 1997.

[168] J. Huss and D. Kelly, "Nuclear receptor signaling and cardiac energetics," *Circulation Research*, vol. 95, pp. 568–578, 2004.

[169] F. Nystrom and M. Quon, "Insulin signaling: metabolic pathways and mechanisms for specificity," *Cell Signal*, vol. 11(8), pp. 563–574, 1999.

[170] A. Saltiel and R. Kahn, "Insulin signaling and the regulation of glucose and lipid metabolism," *Nature*, vol. 414, pp. 799–806, 2001.

[171] W. Stanley, F. Recchia, and G. Lopaschuk, "Mycardial substrate metabolism in the normal and failing heart," *Physio. Rev.*, vol. 85, pp. 1093–1129, 2005.

[172] C. Mounier and B. Poser, "Transcriptional regulation by insulin: from receptor to the gene," *Physiol. Pharmacol*, vol. 84, pp. 713–724, 2006.

[173] C. Montessuit, N. Rosenblatt-Velin, I. Papageorgious, L. Campos, C. Pellieux, T. Palma, and R. Lerch, "Regulation fo glucose transporter expression in cardiac myocytes: p38 mapk is strong inducer of glut4," *Cardiovascular Research*, vol. 64, pp. 94–104, 2004.

[174] B. Desvergne, L. Michalik, and W. Wahli, "Transcriptional regulation of metabolism," *Physiol. Rev*, vol. 86, pp. 465–514, 2006.

[175] The intersystem website. [Online]. Available: http://www.intersystems.com/cache/index.html

[176] B. Maisch, "Enrichment of vital adult cardiac muscle cells by continuous silica sol gradient centrifugation," *Basic Res. Cardiol.*, vol. 76, pp. 622–629, 1981.

[177] B. Oviatt, "Cardiac muscle cells in man and certain other mammals," *Proc. of the American Society of Microscopists*, vol. 76, pp. 283–298, 1887.

[178] A. Drake-Holland and M. Noble, *Cardiac metabolism.* New York: John Wiley and Sons, 1983.

[179] H. Miranda, A. Ferreira, A. Quintas, and C. Cordeiro, "Measuring intracellular enzyme concentrations," *Biochemistry and Molecullar Biology Education*, vol. 36(2), pp. 135–138, 2007.

[180] R. Bar, P. Gorden, J. Roth, R. Kahn, and P. D. Meyts, "Fluctuation in the affinity and concentration of insulin receptors on circulating monocytes of obese patients: Effect of starvation, refeeding, and dieting," *Clinical Investigation*, vol. 58, pp. 1123–1135, 1976.

[181] A. Thorsson and R. Hintz, "Insulin receptors in the newborn: Increase in receptor affinity and number," *New England Journal of Medicine*, vol. 297(17), pp. 908–912, 1977.

[182] S. Sharma, P. Guthrie, S. Chan, S. Haq, and H. Taegtmeyer, "Glucose phosphorylation is required for insulin-dependent mtor signalling in heart," *Cardiovascular Research*, vol. 76, pp. 71–80, 2007.

[183] A. Sedaghat, A. Sherman, and M. Quon, "A mathematical model of metabolic insulin signaling pathways," *Physiol. Endocrinal Metab*, vol. 283, pp. E1084–E1101, 2002.

[184] M. Saddik, J. Gamble, L. Witter, and G. Lopaschuk, "Acetyl-coa carboxylase resulation of fatty acid oxidation in the heart," *J. of Biological Chemistry*, vol. 268(34), pp. 25 836–25 845, 1993.

[185] K. V. der Lee, P. Willemsen, S. Samec, J. Seydoux, A. Sulloo, M. Pelsers, J. Glatz, G. V. der Vusse, and M. V. Bilsen, "Fasting-induced changes in the expression

of genes controling metabolism in rat heart," *J. of Lipid Research*, vol. 42, pp. 1752–1758, 2001.

BIOGRAPHICAL STATEMENT

Amin was born in 1975 in Shiraz, Iran. In 1998 he received his B.S. degree from the Azad University of Tehran, Iran, in Computer Software Engineering. In 2003 he received his M.S. in Computer Science and Engineering from University of Texas at Arlington, USA. During his Master he joined the Center for Research in Wireless Mobility and Networking (CReWMaN) in computer science and engineering department at UTA. His master research was on multimedia streaming over wireless networks and data transport and switching in core optical networks. For his doctoral research he was inspired by his mentor to switch gear and move to the interdisciplinary area of Systems Biology. He received his Ph.D. in Computer Science and Engineering from the University of Texas at Arlington in August 2008. His current research interests focus on modeling and simulation of complex biological networks in eukaryotic cells.