ONTOLOGY DRIVEN SIMULATION OF BIOCHEMICAL PATHWAYS USING HYBRID PETRI NETS

by

Kishore Nimmagadda

(Under the direction of John A. Miller)

Abstract

Glycans are known to change in the earliest stage of cell development in complex living organisms. Glycoproteins and glycolipids effect the abundance level of cell surface glycans. Thus, the study of production and consumption of glycoproteins and glycolipids are important. In this paper, we present the Glycomics Modeling pathway simulation environment (GlyMpse), an ontology driven simulation model utilizing domain ontologies such as GlycO, EnzyO and ReactO as well as modeling ontology DeMO (to generate simulation models of biochemical pathways). The model provides insight to glycobiologists about the behavior of glycan abundance level on cell surface over time. We also tested GlyMpse with N-Glycan biosynthesis pathway to achieve steady state approximation with kinetic parameters generated by a genetic algorithm, as a starting point.

Index words: Simulation, Ontology, Biochemiacal Pathways, Michaelis-Menton Kinetics

ONTOLOGY DRIVEN SIMULATION OF BIOCHEMICAL PATHWAYS USING HYBRID PETRI NETS

by

Kishore Nimmagadda

B.Tech, Jawaharlal Nehru Technological University, India, 2006

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2008

ONTOLOGY DRIVEN SIMULATION OF BIOCHEMICAL PATHWAYS USING HYBRID PETRI NETS

by

Kishore Nimmagadda

Approved:

Major Professor: John A. Miller

Committee: Krzysztof J. Kochut
William York

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

Acknowledgments

I would like to thank my advisor John A. Miller and Krzysztof J. Kochutfor and William York for being in my committee. I have been lucky enough to work with these three brilliant professors and I would like to thank for all their advice, guidance and support.

I must express the utmost gratitude to my parents for all their support and motivation through my academic career, as well as to my friends for their support and also especailly to Archana Meka for making me smile no matter how tough things can get.

Table of Contents

List of Figures

List of Tables

Chapter 1

Inroduction

Computational systems biology includes both knowledge discovery and simulation-based techniques. The knowledge discovery branch of computational systems biology processes large quantities of experimential data in an effort to find patterns that are difficult to identify without the aid of computational tools, while the simulation-based branch makes use computational modeling techniques to represent the characteristics of biological systems and predict their behavior. The behaviors exhibited by an executable computer model can be validated by comparing them with behaviors observed during experiments. Once computer models have been validated, they can be used to make predictions about the behavior of biological systems.

Biochemical pathways have traditionally been modeled as systems of ordinary differential equations (ODEs). ODEs simulate the time-dependent properties of a pathway well, but as pointed out by Reddy [58], simulation with ODEs relies on parameters produced by experimental data which may be available for single reactions, but may not be available for entire pathways. In the absence of complete data required for parameters in a system of ODEs, many researchers have opted to use Petri Nets [51] for the simulation of pathways [2]. Petri Net models can represent the structure of pathways using an intuitive graph theoretic approach, but standard (discrete) Petri Nets cannot represent the time-dependent properties of a pathway that are necessary for quantitative analysis. The original Petri net is only able to deal with discrete values, but in recent years hybrid Petri net (HPN) [55] and hybrid Dynamic net (HDN) [56] models have been introduced in order to allow Petri net models to handle continuous values, while retaining the properties associated with the original Petri net.

While the extensions introduced in HPNs and HDNs were significant, they were not sufficient to allow for the accurate representation of biochemical pathway via Petri Net models. In an effort to overcome the limitations of HPNs and HDNs, [57] introduced the hybrid functional Petri net (HFPN). Using HFPNs modelers are able to represent the structure of pathways in a intuitive manner and simultaneously use ODEs to represent the time-dependent properties of pathways. The systems described in this paper make use of many of the concepts associated with HFPNs in order to simulation biochemical pathways.

Within the field of computer science an ontology [51] is a formal, machine readable description used to describe and categorize concepts and the relationships among concepts within a particular knowledge domain. The domain of systems biology has seen the development of the Glycomics Ontology (GlycO) [30], the Enzyme Ontology (EnzyO), the reactions ontology (ReactO) and a number of other ontologies such as Gene Ontology (GO) [62], the Sequence Ontology (SO) [63], in recent years. Several ontologies such as the Discrete-event Modeling Ontology (DeMO) [33] and Process Interaction Ontology for Discrete Event Simulation (PIMODES) [52], have also been developed within the domain of simulation and modeling.

The systems described in this paper use ontologies within the domains of systems biology and simulation and modeling to drive the creation of simulation models for biochemical pathways. This is achieved using Ontology Driven Simulation (ODS) [35] where a tool maps concepts from domain ontologies to concepts in the DeMO modeling ontology and then creates instances of DeMO ontology classes to represent a simulation model. We illustrate how ODS can be used to create simulation models of glycan pathways by mapping concepts from the GlycO, EnzyO and ReactO ontologies to concepts in the DeMO modeling ontology and then creating instances of DeMO ontology classes to represent a model. We also show that once the ontology instances representing the model have been created, additional tools can be used to translate the instances into an executable simulation model.

The remainder of this thesis is organized as follows: Chapter 2 discusses systems biology and some of the strategies used in the simulation of biochemical pathways. Chapter 3 summarizes related work. Chapter 4 describes how hybrid Petri nets have been used in the simulation of pathways. Chapter 5 introduces ontology driven simulation of pathways. Chapter 6 describes how elements of simulation models represented as ontology instances are translated into executable simulation models, and Chapter 7 presents some preliminary test results generated by the simulation models described in this paper. Lastly, Chapter 8 discusses conclusions and opportunities for future work.

Chapter 2

Problem Definition and Solution Strategies

2.1 Structure and processes in a cell

The living cell is a biological unit surrounded by a lipid membrane. A cell can be distinguished from smaller biological units (such as virus particles) in that a cell contains all the components required for its replication. Diverse biological processes such as growth, metabolism, stimulus response and replication are carried out by cells, which are categorizes either as prokaryotic (without nuclei) or eukaryotic (with nuclei). Prokaryotic cells are relatively small in size and typically have a rigid cell wall. Eukaryotic cells, which are typically larger than prokaryotic cells, are divided into sub cellular compartments (called organelles), which include nuclei, mitochondria, Golgi, and endoplasmic reticulum (ER). Each organelle has a characteristic cellular function. For example proteins destined for export from the cell are usually synthesized at the surface of the ER.

2.2 Biochemical Pathways

The complex interactions of biomolecules within the cell are often conceptually organized as networks called biochemical pathways, which include metabolic pathways, signal-transduction pathways and gene regulatory pathways.

2.2.1 Metabolic Pathways

A metabolic pathway is an ordered sequence of biochemical (usually enzyme-catalyzed) reactions in which precursor molecules are converted into product molecules. These reactants and

products are called metabolites. Metabolism is responsible for the synthesis and degradation of the molecules that are structural or functional components of cells. Energetics plays a central role in metabolism, which is responsible for the generation, interconversion and utilization of cellular fuels, which include monosaccharides, polysaccharides and nucleotides. Metabolism consists of catabolism (biochemical degradation) and anabolism (biochemical synthesis). Catabolism is the breakdown of complex molecules (carbohydrates, proteins and lipids) into simpler molecules (ammonia, urea and carbon dioxide), often supplying energy for other cellular processes, including anabolism, which is the formation of complex molecules (polysaccharides, glycolipids and proteins) from simpler precursors.

2.2.2 Signaling Pathways

Signal-Transduction pathways are also known as information processing pathways. These pathways allow cells to perceive, process, and respond to outside information. Half of the protein families encoded by the human genome are primarily involved in information processing. Signalling molecules often encode information by interconversion between activated and deactivated states that differentially interact with the cellular machinery, thereby modulating various cellular processes such as overall metabolism and gene expression.

2.2.3 Gene Regulation Pathways

As genes are expressed, they produce RNA and ultimately proteins. Some of these RNA and protein products then bind to promoter sites of other genes, effecting the level at which they express themselves, thus setting a regulatory network or pathway.

2.3 Glycan Biosynthetic pathways

Carbohydrates (i.e., simple sugars, oligosaccharides and polysaccharides) play diverse roles in the living cell. Most of the energy harvested from sunlight is stored by synthesis of simple sugars from water and carbon dioxide. In the long-term, this energy is usually stored in

the form of polysaccharides, which are subsequently broken down when energy is required by the organism. Oligosaccharides and polysaccharides are collectively referred to as glycans. Glycoproteins and glycolipids, respectively formed by covalently linking glycans with proteins and lipids, have important roles in molecular recognition and signal transduction. Ultimately, the abundance of each complex glycan in the cell is controlled by regulating the rates of individual reactions within the metabolic network responsible for its biosynthesis. The reaction rates are controlled by varing the abundances of the enzymes that catalyze these reactions and the substrates that take part in these reactions. Analysis of the pathways responsible for glycan biosynthesis involves the detection and quantitation of the glycans themselves and the enzymes that catalyze the reactions. The abundance of each enzyme can be estimated to some extent by quantifying the rate at which the gene encoding the enzyme is transcribed (i.e., transcriptomics) or by quantifying the enzyme itself (i.e., proteomics). The GlycO and ReactO ontologies embody knowledge regarding glycan structures and the metabolic pathways leading to glycan biosynthesis, providing a context for interpretation of glycomics data (i.e, identity and abundance of each glycan in the cell or tissue). Knowledge regarding the enzymes that catalyzes these biosynthetic reactions is contained in the EnzyO ontology. Together these ontologies constitute a robust collection of information required for the description of pathways leading to glycan biosynthesis.

Simulation of the glycan biosynthesis process would be much more robust if it integrates experimental glycomics, transcriptomics and proteomics data. The glycomics group at the University of Georgia (http://glycomics.ccrc.uga.edu) has developed the infrastructure to collect transcriptomic, proteomic and glycomic data. For example, gene expression can be quantified by analysis of mRNA abundances in the cell or tissue by qRT-PCR (quantitative Real Time Polymerase Chain Reaction). Although the abundance levels of cell surface glycans and the enzymes involved in their synthesis can be viewed in the context of GlycO, quantitative relationships between these abundances is highly complex, and meaningful models for the biosynthesis requires simulation of this dynamic process. Such models have the potential

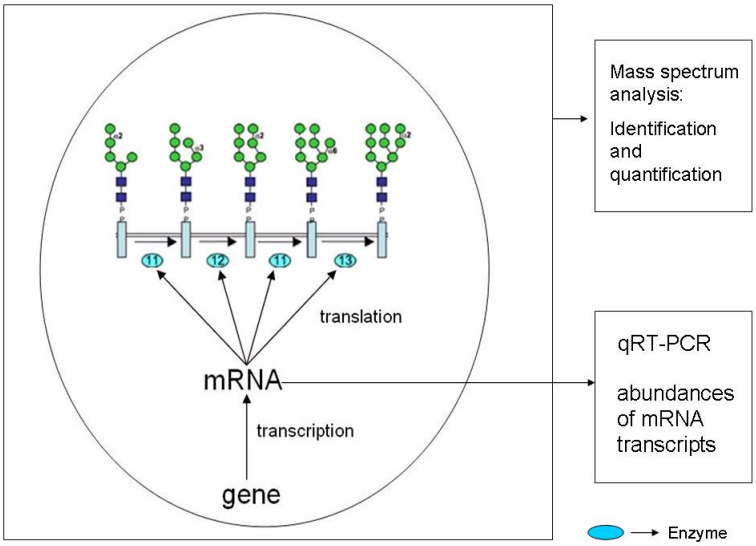

Figure 2.1: Conversion of gene to mRNA to protein

to shed light on the effects of altering gene expression on glycan structure and abundance, which then leads to changes in diverse cellular processes such as cell growth and differentiation and the progression of disease.

GlycoBrowser [43] is an ontologically driven glycan structure building and pathway browser tool. This tool also renders the glycan abundance level and transcriptome expression levels as small bar graphs beneath each glycan and enzyme, respectively as shown in figure 3 [1]. Glycan-protein interaction data collected until now is not sufficiently used for modeling pathways. The main goal of GlyMpse is to model glycan pathways from GlycO, EnzyO and ReactO ontologies. The model will provide insight to glycobiologists about the behavior of glycan abundance level on cell surfaces over time. GlyMpse models provide a means to analyze glycan-enzyme (proteins) interactions using existing information from GlycO, EnzyO and ReactO as a starting point. Data flow from domain ontologies to simulation engine is depicted in Figure 1.1.

[1]This figure is generated using GlycoBrowser [43]

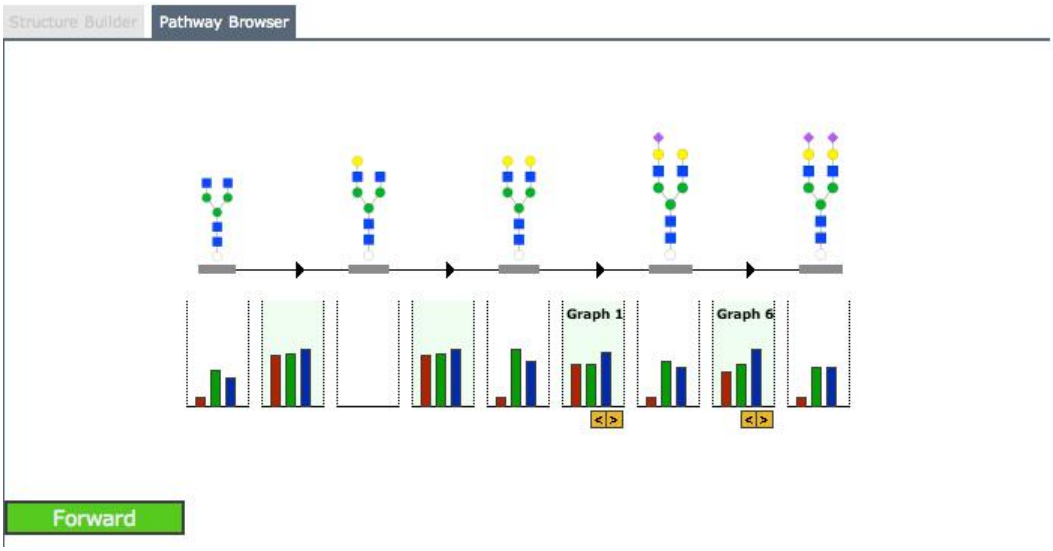

Figure 2.2: Pathway fragment with corresponding abundace level graphs

2.4 System biology approach

The study of systems biology was started by Mihajlo Mesarovic in 1966 with publication of his research paper "Systems Theory and Biology." [44]. However, high quality, high throughput data only became available in 1990's, which also saw significant increase in computing power. With the completion of the Human Genome Project in 2000, systems biology evolved as an important research area in many research institutes. Progress in molecular biology induced a considerable degree of interest in system biology, especially since high-throughput experimental measurements provides complete data sets on cell performance and gains in depth information on the underlying molecular level cellular activities. The analysis of data generated by mathematical and simulation modeling is becoming increasingly important. How a system behaves over time under various conditions can be understood through metabolic analysis. This analysis tracks time-varying changes in the state of the system depending on the concentration of the biochemical factors involved. Due to the large number of parameters

and constraints in biological networks, software infrastructure is becoming an important component for system biology. The concept of ontology (specification of conceptualization) is one of the knowledge representation techniques from computer science used in bioinformatics. Many research groups are using ontologies as a software infrastructure to store biological information. The Systems Biology Mark-up Language (SBML), along with CellML define a standard for an XML-based computer readable models for exchange of data.

2.5 Reaction laws and kinetics

The rate of a reaction is often found to be proportional to the concentrations of the reactants. For example if the rate of the reaction is proportional to the molar (moles per liter) concentrations of two reactants E and S, in this we can write as

$$V = k[E][S] \qquad \text{where } V \text{ is the rate of the reaction.}$$

The coefficient k is called the rate constant for the reaction. The rate constant is independent of the concentrations (depends on temperature). Formally, a rate a law is an equation that expresses the rate of the reaction as a function of the concentrations of all reactants present in the reaction:

$$V = f([E], [S])$$

The rate of a reaction depends on substrate and enzyme concentration. The Michaelis-Menton kinetic equation relates velocity V to the concentration of substrate $[S]$ and the two parameters V_{max} and K_m. The following equation is used in our simulation to calculate reaction rates:

$$V = \frac{V_{max} \cdot [S]}{K_m + [S]}$$

V_{max} is the maximum velocity of the reaction and K_m is the concentration of substrate that leads to half the maximum velocity V_{max}. The rate is most sensitive to substrate concentration when $[S]$ is less than K_m. However, it is sensitive to $[S]$ at every realistic concentration

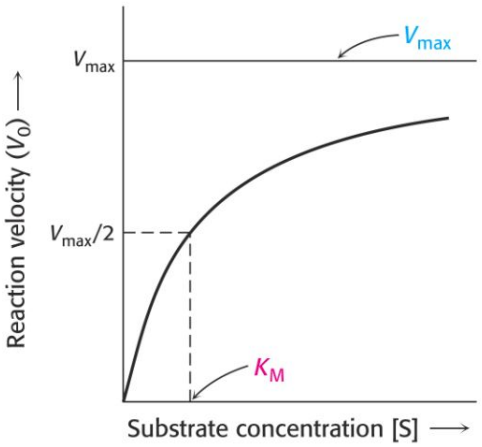

Figure 2.3: Michaelis-Menton rate equation graph

as shown in Figure 2.3. The rate no longer sensitive to [S] when [S] is infinite. Increase in substrate concentration increases the rate of the reaction. The derivation of Michaelis-Menton kinetics are discussed in Appendix A.

2.6 Modeling and simulation of pathways

2.6.1 System differential equations

The rate equations are linear or non-linear ordinary differential equations (ODEs) with concentration levels as variables. ODEs fail to generate accurate output if some of the key parameters are unknown. ODEs are not suitable for animation or visualization and hard to use with discrete events. Discrete events are needed to represent activities such as binding, signaling and continuous values are needed to represent substrate concentrations, these two are needed for simulations of cells at the molecular level, so Hybrid Petri Nets are a well accepted approach.

Complex Pathway Simulator (COPASI) [10, 11] and Biochemical Pathway Simulator (BPS) [5] are examples of the simulation engines developed for biochemical pathways. These simulation engines are limited by their need to have chemical kinetics input by the user. The systems described in this paper differ from these simulation engines in that when a pathway is selected its corresponding kinetics are retrieved from domain ontologies and reactant concentrations from GlycoVault [64]. GlycoVault data include qRT-PCR data as well as glycomics data. Petri Nets which provide structural definitions for the simulation of pathways are not used in any of the previously mentioned simulation engines.

2.6.2 PetriNets

A Petri Net is a formal description for modeling concurrent systems developed by Carl Adam Petri [50] in 1960. A Petri Net is bipartite graph, with place and transition nodes. Figure 2.4 represents a simple Petri Net model with 6 places and 4 transitions. The formal definition of Petri Net is as follows:

Petri Net is a 5 tuple, (P, T, F, W, M_0), where

$P = \{p_1, p_2, ..., p_m\}$, finite set of places

$T = \{t_1, t_2, ..., t_n\}$, finite set of transitions

$F \subseteq (P \times T) \bigcup (T \times P)$, set of arcs

$W : F \rightarrow \{1, 2, ...\}$, weighting function

$M_0 : P \rightarrow \{0, 1, ...\}$, initial marking

M_0 defines the number of tokens per place.

Places are represented by circles and transitions are represented by rectangular boxes. These places and transitions are connected by directed arcs. A Transition can have both input places (with arcs connecting from places to the transition) and output places (with arcs connecting from the transition to places). Generally, arcs are labeled with weights which represent the minimum tokens required by input places to enable the transition. When a transition is fired it will take tokens from each input place equal to the weight of the

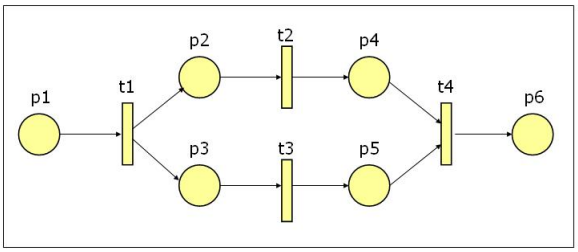

Figure 2.4: Simple Petri Net

Petri Net Type	**Description**
Timed	A timed duration is associated with transitions and places
Colored	Each token can have its own color, gives more descriptive power
Hierarchical	A place can represent another Petri Net
Stochastic	Probability distribution delays are associated with places and transitions
Hybrid	Places can hold integers and real values

Table 2.1: Classification Of Petri Nets

corresponding input arc and create new tokens equals to the weight of the corresponding output arc. To model a metabolic pathway using Petri Nets, we need to take into account both the substrate degradation as well as creation. Therefore, the notion of consumption and production of resources is achieved in Petri Nets. To model biochemical pathways, it is important to understand how timing concepts are handled by Petri Nets. A firing delay is associated with each transition which specifies the time duration for which this particular transition is enabled. Table 2.1 explains classification of Petri Nets:

2.6.3 Advantages of Petri Nets

Petri Nets provide an intuitive graphical approach for modeling biochemical pathways. Since a Petri net is also represented as a network with nodes and edges, mapping the components of a biochemical pathway to a Petri net is somewhat intuitive. When modeling a biochemical network using a Petri net, the amount of a biological molecule/object is represented as a place. The transitions represent the reactions/interactions between the biological molecules/objects, and places are connected to transitions by arcs. Petri Nets are well suited for modeling the logical relationship between components in biochemical pathways. The hybrid modeling methodology most often used for biochemical pathway simulation is the Hybrid Functional Petri Net (HFPN) [57]. The HFPN retains the advantages of the Petri Net, but also allows the continuous features of a biochemical pathway to be accurately represented. Besides these advantages, graphical patterns of pathways can be modeled with HFPN which can be easily understood by a biologist without any mathematical equations.

2.7 Quantitative analysis

The Petri Nets outlined above have recently been applied to quantitative analyses in many areas of systems biology. An extension of the Hybrid Petri Net has been used in the representation and simulation of gene regulatory networks and signal transduction pathways [7]. Simulation software packages based on Petri Nets allow the user to study the time evolution of concentrations of substrates within the system based on the Michaelis-Menton kinetic equations.

These applications are good illustrations of the suitability of Petri Nets for the quantitative analysis of complex biological systems. It seems likely that they will prove useful in the future as bioinformaticians develop more integrated representations for biological networks and also when more pathway related data is collected from experiments.

2.8 Related Work

General Pathway Simulator (GEPASI) [10, 11] simulates the steady-state and time-course behavior of reactions. In this tool, users supplies input to the program with information about the stoichiometric structure of the pathway, kinetics of each reaction, volumes of the compartments and initial concentration of all chemical species. GEPASI then builds the differential equations that are required by the system and solves them.

Similarly, E-Cell [14] is a software environment where user can even define functions of proteins, protein-protein interactions, protein-DNA interactions, regulation of gene expression and other features of cellular metabolism, as a set of reaction rules. In these reaction rules a system of differential equations is implicitly defined. E-Cell then solves these differential equations to simulate the model behavior. The simulator KINSOLVER [15] solves chemical reaction networks based on mass balance kinetics. This tool uses five standard methods (Euler, Modified Euler, Runge Kutta (RK), Adaptive RK-Fehlberg, and LSODES).

Biochemical Pathway Simulator (BPS) [5] is based on programmed cell death (apoptosis) and growth factor activated kinase (MAPK) in a biochemical network. Cell Illustrator is a commercial software tool that allows users to model, simulate and visualize biological pathways using Cell System Markup Language (CSML). ScrumPy [45] is another analysis tool for metabolic models. This software package has user interface developed in Python. A Discrete-Event Simulator of Metabolic Networks (DiMSim) [47] models a pathway as bipartite graph with reactions and reactants. DiMsim also supports Michaelis-Menten kinetics. Cell Illustrator [46] is a software tool supporting Cell System Markup Language (CSML). The design of this software is based on Petri Nets. Standard XML formats like SBML and CellML can also be converted to CSML using this software.

Chapter 3

Simulation using Hybrid PetriNets

3.1 Hybrid PetriNets in Simulation

Hoestadt and Thelen (1998) extended the ideas introduced in Reddy, Liebman, and Mavorououniotis (1993) by using the Hybrid Petri net/Hybrid Dynamic net model (HPN/HDN) to model both qualitative and quantitative features of biochemical pathways. The HPN/HDN model was further extended by Matsuno [57] to create the HFPN. The formal definition of HFPN is as follows:

HFPN is a 6 tuple, $(P, T, h, Pre, Post, M_0)$, where

$P = \{p_1, p_2, ..., p_m\}$, finite set of places.

$T = \{t_1, t_2, ..., t_n\}$, finite set of transitions.

$h = P \cup T \rightarrow \{D, C\}$, every place or transition whether it is a discrete or continuous one.

$Pre(P_i, T_j)$ is an arc from place P_i to transition T_j which has weight of non-negative integer.

$Post(P_i, T_j)$ is an arc from place P_i to transition T_j which has weight of non-negative real numer.

M_0 defines the number of tokens per place.

The HPFN has discrete places and transitions, which can be used to model discrete features of biochemical pathways, and it has continuous places and transitions which can be used to model the continuous features of the pathways. The continuous places can hold real numbers to represent concentrations of various entities (molecules, etc.) in the model and continuous transitions that can fire continuously with their firing speed determined by a function that uses the values from the continuous places. The HFPN also employs different kinds of arcs listed below:

1. Discrete Input Arc: directed from a discrete place to a discrete transition
2. Continuous Input Arc: directed from a continuous place to a continuous transition
3. Discrete Output Arc: directed from a discrete transition to a place of any kind
4. Continuous Output Arc: directed from a continuous transition to a continuous place
5. Test Input Arc: can be directed from a discrete place to a discrete transition or from a continuous place to a continuous transition

A discrete transition will fire only if the place on the source end of the input arc contains a value greater than the weight of the arc, and when the transition fires, content from the source place is consumed. A discrete transition may also contain a delay time specified by a non-negative integer. A continuous transition fires continuously at the speed specified by its transaction function. The input arc of a continuous transition may also contain a weight that determines whether the transition fires. Transitions that have test input arcs fire in the same way as transitions that have normal input arcs, but the content of the source place of a test input arc is not consumed when the transition fires.

From a schema of the pathway, systematically each reaction (with assumptions) is replace with its hybrid Petri Net counterpart. Place nodes are usually used to represent biochemical entities (compounds, enzymes, etc.) and transition nodes are used to represent reactions. Enzymatic reactions are possible when sufficient enzyme is available, and have no impact on the amount of enzyme although the current concentration of the enzyme will affect the rate of reaction. Figure 3.1 shows a Petri Net model of enzyme reaction where Glycan A is converted to Glycan B in presence of an active enzyme.

3.2 Simulation Databases and Ontologies

Biological pathways are represented as a network of nodes representing some biological molecules and connections between nodes representing reactions/transfers between them

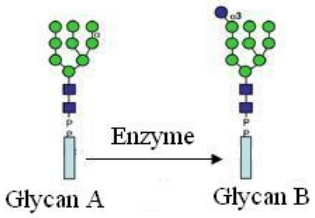

Figure 3.1: Enzyme catalyzed reaction

together with links to biological annotations. Pathway databases sucha as KEGG [22], EcoCyc and MetaCyc [23], TRANSPATH [24], Biocyc [25] and Genome [26] uses these type of representation. BRENDA [59] is relational database containing enzymatic and kinetic information mainly extracted from literature. System for the Analysis of Biochemical Pathways - Reaction Kinetics (SABIO-RK)[60] is another curated database holding kinetics information. However, BRENDA and SABIO-RK do not have complete set of kinetics for enzymes participating in glycan pathways. The biggest challenge for pathway simulation when using methods like Michaelis-Menton is getting kinetic data. In these pathway databases, information about Michaelis kinetics data is often not available. Thus, an important challenge is to create an information system which holds kinetic information for pathway simulation. The Reactions ontology (ReactO) has been developed with the idea of holding pathway kinetics.

The Glycomics ontology (GlycO) defines glycan structures and their synthesis from biochemical pathways [29]. GlycO embodies knowledge regarding the structures, biosynthesis, and biological functions of complex glycans. GlycO contains knowledge regarding other aspects of glycobiology which includes functional and structural relationships between different glycans. Graphical representation of GlycO knowledge can be accessed using GlycoBrowser [43]. The Enzyme ontology (EnzyO) describes enzymes that catalyzes reactions in the pathways. Input to this simulation engine is based on GlycO, EnzyO and ReactO.

Chapter 4

Ontology Driven Simulation Of Pathways

4.1 Domain Ontologies

Ontologies [51] are formal descriptions used to describe and categorize concepts and the relationships among concepts within a particular knowledge domain. They can be processed by machines or read by domain experts. Ontologies may be used to share a common understanding of the structure of information, enable reuse of domain knowledge, make domain assumptions explicit, separate domain knowledge from operational knowledge, and analyze domain knowledge [53]. They allow researchers, domain experts, and software agents to share a common understanding of the concepts and relationships of a domain. The past few years have seen the publication of ontologies for a large number of domains. For example, in the domain of biology the Glycomics ontology (GlycO) (in the biology domain) defines glycan (carbohydrate) structures and the items in the biosynthetic pathways that create them [29], and the Enzyme Ontology (EnzyO) models the structure and function of enzymes [28]. The modeling and simulation community is beginning to see potential for using these ontologies in the modeling process.

Simulation models are used to simulate processes in a variety of domains. Traditionally, during the development of a model, the simulation modeler has relied on the information provided to him/her by domain experts. Recently another source of domain knowledge, known as the domain ontology, has become available. While domain ontologies cannot replace domain experts, they are useful to simulation modelers because they provide formal methods for describing the concepts, categorized, and relationships within a domain. An added advantage of ontologies written in languages such as the Web Ontology Language (OWL) [54]

is that they can be processed by machines. Domain ontologies can be of particular use to simulation modelers because it is possible to use them to communicate domain information to simulation and modeling tools with limited human intervention.

4.2 Modeling Ontologies

Efforts to integrate ontologies into modeling and simulation frameworks and to use them in the development of formal methods for simulation and modeling have been undertaken by several groups. Lacy [52] developed the Process Interaction Modeling Ontology for Discrete Event Simulations (PIMODES) to focus specifically on the process interaction world view and support the interchange of simulation models between commercial simulation packages. Benjamin [31] presents an ontology driven framework for process oriented applications, including process interaction simulations. The framework is an attempt to overcome the problem of semantic inaccessibility where the semantic intentions of the application developers and the semantic rules of the application are not available to other entities within an organization. Turnitsa and Tolk [65] describe what is required for an ontology to be sufficiently complete to serve as a reference for simulation interoperability and they also propose a method for evaluating an ontology to determine it completeness.

The Discrete-event Model Ontology (DeMO) [32] is used as a foundational component for the ontology driven simulation system described in this paper. DeMO attempts to provide a comprehensive ontology for discrete-event simulation with parts for state oriented, activity oriented, event oriented, and process oriented models [33]. It was defined using the Web Ontology Language (OWL). The principal component of DeMO, the DeModel class, is divided into four subclasses, StateOrientedModel, ActivityOrientedModel, EventOrientedModel, and ProcessOrientedModel which form the foundation for all of the modeling techniques supported in DeMO. The systems described in this paper make use of Time Hybrid Petri Nets to model and simulation biochemical pathways. DeMO's TimedHybridPetriNet

class, a subclass of ActivityOrientedModel, is used as the foundational component for representing and generating the executable Petri Net models of biochemical pathways described in this paper.

Chapter 5

GlyMpse

5.1 Architecture

As shown in Figure 5.1, GlyMpse has four important components: domain ontologies, ontology mapping tool (DeMO), simulation engine and animation engine. Initial input data for this simulator is obtained from domain ontologies GlycO, EnzyO, ReactO and GlycoVault which contains qRT-PCR and glycomics data (discussed in section 2). DeMO provides a comprehensive ontology for simulation modeling based on domain ontologies. This DeMO ontology also facilitates integration of retrieved information and supports code generation for simulation engine. TimedHybridPetriNet class (discussed in section 4) in DeMO contains information about places, transitions, arcs and their initial markings for a pathway model. A Petri Net model is drawn by the animation engine based on the model in the simulation engine. Any change in the concentration level of a substrate along with time is sent to animation engine to display the corresponding changes in the animator.

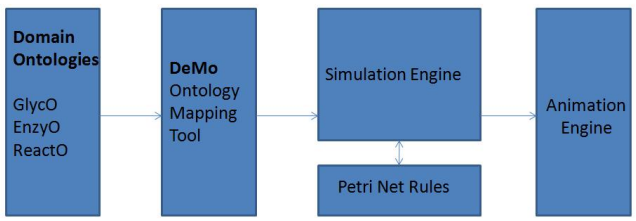

Figure 5.1: System Architecture

5.2 Implementation

GlyMpse is a platform-independent environment developed in Java based on the JSIM [61] simulation library. JSIM is a Java-based simulation and animation environment. The JSIM library includes many Java classes to make developing simulation models easy. The principal components of GlyMpse simulation environment are the PetriNet and Animator classes. The PetriNet class provides a simulation engine for Hybrid Colored Petri Nets. This class also contains subclasses DPlace, CPlace, DArc, CArc and Transition to represent components of Petri Net discrete place, continuous place, discrete arc, continuous arc and transition respectively. A pathway model is created by extending PetriNet class with pathway specific places, arcs, transitions and kinetic parameters. Animator is an interface that specifies the commands to create, destroy, move and scale components in an animated canvas. Besides these classes, Graph class displays graphical results of substrate concentrations over time.

5.3 Firing Rules

Simulation of a pathway model is driven by Petri Nets rules. PetriNetRules class is used to define firing rules for the PetriNet class. These rules determine wheather a transition is ready to fire or not. To determine the amount of substrate converted to product when a transition is fired is defined by the Michaelis-Menton rate law and Runge Kutta differential equations integrator. The fourth order Runge Kutta method does four function evaluations per step to give a method with fourth order accuracy. The actual amount of substrate converted into product in a specified time period is determined by the step size used in Runge Kutta method. Calculations are made at the initial time, two at half the step size beyond the initial time and at the final time as shown in the Figure 5.2.

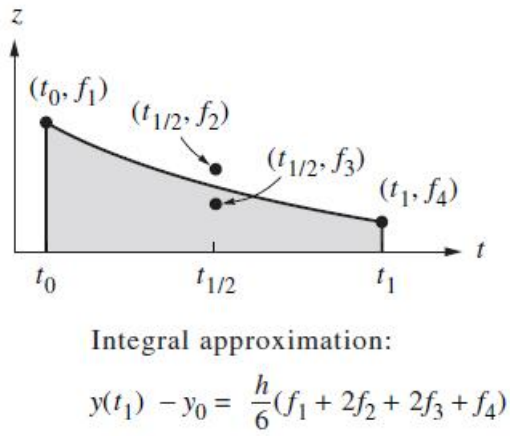

Figure 5.2: Ranga Kutta 4th order integral approximation

5.4 Parameter Estimation

Enzymatic reaction kinetics plays vital role in biochemical pathway simulation. Since complete data is not available from domain ontologies we developed a Genetic Algorithm (GA) which approximates kinetic parameters. For the given concentration of source and sink (product) in a period of time, GA estimates the parameters based on fitness function. In this simulation, fitness is calculated based on how close the product concentration is with GA kinetic parameters. A kinetics set, K, can be defined as the set of V_{max} and K_m values of all enzymes particiapting in a model. Crossover and mutation are two basic operators of GA. Mutation is a genetic operator that alters one ore more values in the initial kinetic set. Crossover is a genetic operator that combines two values to produce a new kinetic value. The following algorithm has been developed for parameter estimation using GA:

```
1. Generate random kinetics set K
2. Run simulation with these kinetic
    parameters, return product concentration
3. Calculate the fitness function with K
4. Apply mutation, crossover to K
```

```
5. Repeat steps 2, 3 and 4 until desired
    fitness values is reached
```

Even though this algorithm evolved from initial random kinetics, this GA can generate pathway kinetics which gives the desired product concentration. This algorithm can get more accurate results if at least some of the parameters (for N-Glycan biosynthesis pathway) are known because based on known parameters, unknown parameters can be estimated.

5.5 Model Execution

Simulation data is retrieved from domain ontologies (GlycO, EnzyO, ReactO and GlycoVault). The DeMO ontology facilitates integration of retrieved information and supports code generation for the simulation engine. Finally, a model is generated from the DeMO ontology with all input parameters such as initial substrate concentrations and kinetic parameters which are required for pathway simulation. When a simulation starts running, all the transitions are checked to see if are ready to fire based on the PetriNet rules and substrate concentration levels. Once a transition is ready to fire it is placed in a time based priority queue. A transition in the queue fires when the simulation clock reaches the transition firing time. This process continues until all the transitions in the queue have fired or the simulation stop time is reached. To determine the amount of substrate converted to product when a transition fire is defined by the Michaelis Menton rate law and Runge Kutta differential equations integrator (discussed in previous section).

A simulation model begins with the construction of a Petrt Net from a biological pathway. Figure 5.4 [1] shows the relationships of reactants, products, the reactions in N-Glycan biosynthesis pathway. For simplicity, we will explain the construction of a model with participating substrates, enzymes and kinetics within the context of the model depicted in Figure 5.5. This is a fragmented N-Glycan biosynthesis pathway with 5 reactants, 4 enzymes. In this example

[1] This figure is from http://www.ccrc.uga.edu/ moremen/glycomics/

2-GlcNAc,5-Mannose is converted to 2-GlcNAc,9-Mannose,1-Glucose with a sequence of reactions.

When a Petri Net model is generated the following algorithm is executed to simulate the model:

```
while clock ≤ simEndTime do
    foreach Transition t do
        if t has atleast minFluids then
            add to PriorityQueue pq
        end
    end
    if pq ≠ empty then
        pull Transition t from pq
        fire(Transition t);
        clock = t firingTime
    end
end
```

Algorithm 1: Petri Net pathway simulation algorithm

In this simulation process, concentrations of all the substrates are collected and written to an Excel file. The time vs. concentration graph obtained from this graph is used for further analysis of the pathway. Figure 5.3 shows a model generated with our animation engine. Each substrate has a distinct color and when a user clicks on any place holding a substrate its corresponding structure is displayed. The vertical bar in each place represents its substrate concentration level. A clock is displayed to keep track of simulation time.

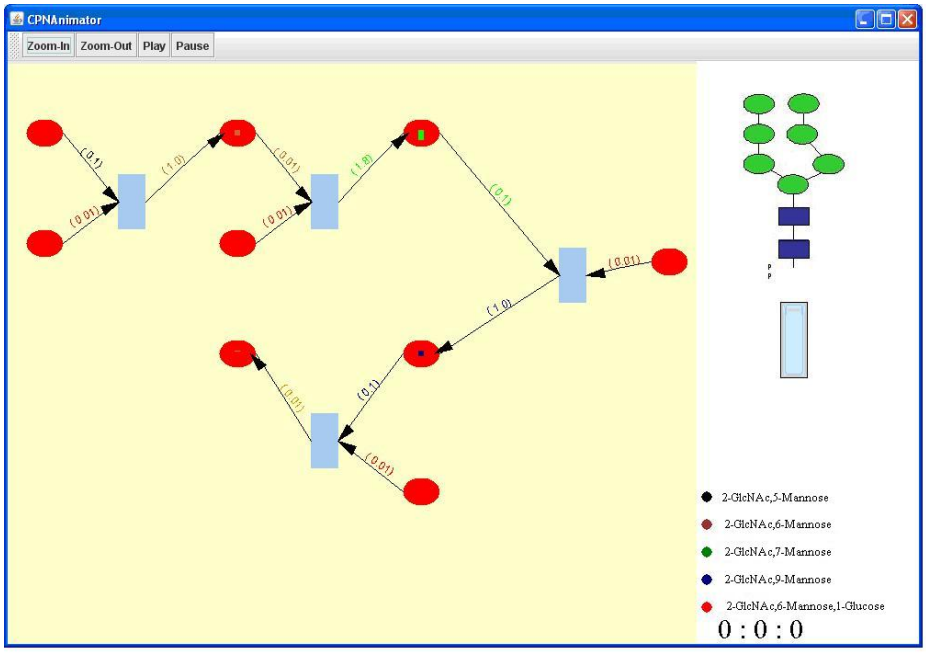

Figure 5.3: Model generated by GlyMpse

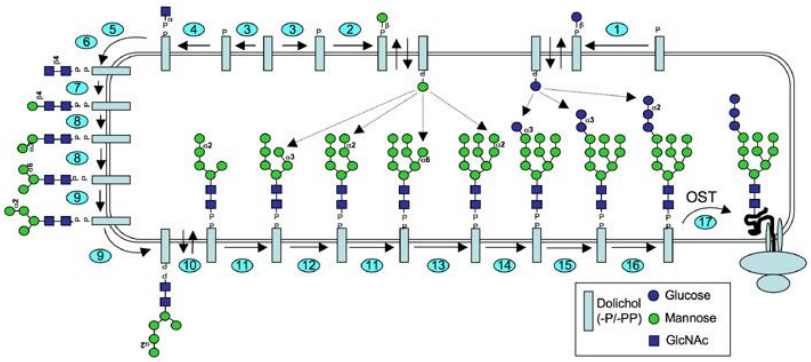

Figure 5.4: N-Glycan Biosynthesis - Lipid-Linked Precursor

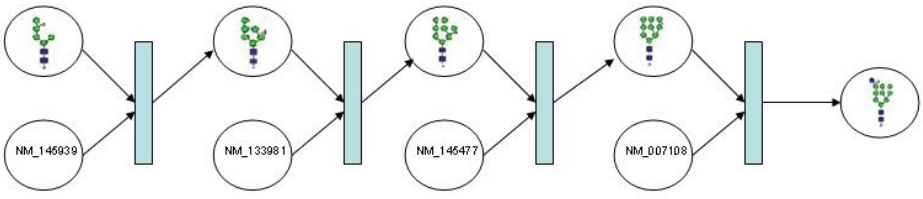

Figure 5.5: Petri Net Model of fragmented Lipid-Linked Precursor

Chapter 6

Testing the GlyMpse Simulation Engine

To study the system behavior, we studied simulation of N-Glycan biosynthesis (part of) pathway. The system considered five substrates: 2-GlcNAc-5-Mannose, 2-GlcNAc-6-Mannose, 2-GlcNAc-7-Mannose, 2-GlcNAc-9-Mannose, and 2-GlcNAc-9-Mannose-1-Glucose which are respectively catalyzed by the four enzymes: NM_145939, NM_133981, NM_145477, NM_007108 (referenced by their RefSeq number). The parameters used in this test case are listed in Table 6.1.

The detailed behavior of each place (substrate, enzyme or product) can be observed during the simulation as the simulation of the system is done stepwise by following the occurrence rule. In addition, modeling capability is provided as the system can be easily modified by altering the kinetic parameters.

It also allows direct observation of the effects of the concentration changes on various steps in the pathways. This is possible because whenever a transition is fired according to the Petri Net rule during the simulation the changes in the concentrations of all molecules can

Enzyme	$K_m \ \frac{mmol}{mL}$	V_{max}
NM_145939	4.1	1.3
NM_133981	1.3	1.18
NM_145477	1.9	2.15
NM_007108	1.2	1.18

Table 6.1: Michaelis-Menton kinetic constants

be displayed in the corresponding places. These characteristics will be useful for the system with uncertain kinetic parameters, which need to be estimated for the simulation.

The steady state approximation assumes that, after an intial induction period, an interval during which the concentration of inrtermediates, rise from zero, and during the major part of the reaction, the rates of change of concentrations of all reaction intermediates are negligibly small. For example let:

$$S \rightarrow I_1 \rightarrow I_2 \rightarrow P$$

be consecutive reactions with reaction constants k_{i1}, k_{i2}, k_b

$$\frac{d[P]}{dt} = k_b[I]$$

After passing the induction period, the rate of change of concentration of all reaction intermediates will approach to zero.

$$\frac{d[I_1]}{dt} \approx 0 \text{ and } \frac{d[I_2]}{dt} \approx 0$$

Figures 6.1 and 6.2 exactly depict the equilibrium state after a certain point of time. Figure 6.1 is the result after running simulation for 800 milliseconds with source substrate concentration kept at 20 $\frac{mmol}{mL}$. Similarly, Figure 6.2 is the result after running simulation for 800 milliseconds with source substrate concentration kept at 40 $\frac{mmol}{mL}$. One can easily see that the steady state is approached in both figures after induction period. The straight line shown in both figures shows source substrate concentration which is kept constant. Intermediate products are produced at high rate before induction period and then approach a steady state. The product concentration increases through out the simulation running time.

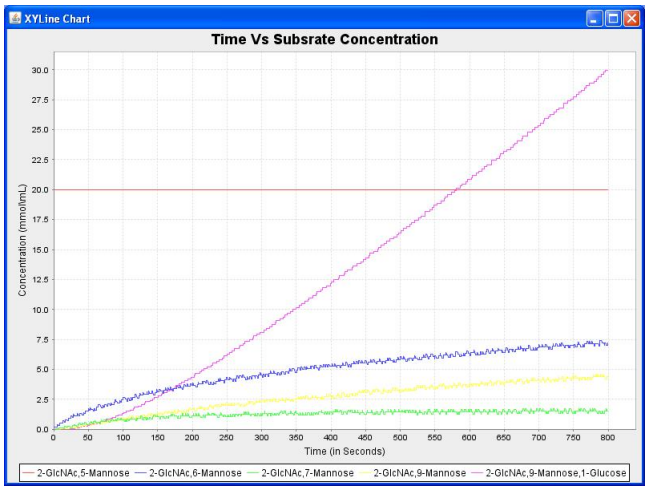

Figure 6.1: Simulation time: 800, $[S]_0 = 20\frac{mmol}{mL}$

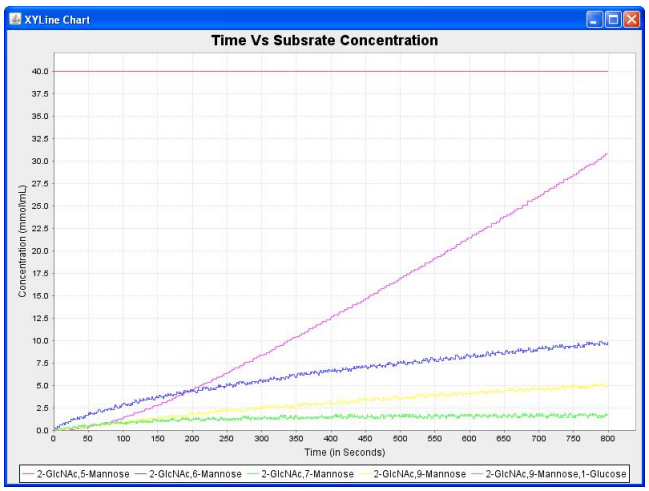

Figure 6.2: Simulation time: 800, $[S]_0 = 40\frac{mmol}{mL}$

Chapter 7

Conclusions and Future Work

GlyMpse, ontology driven simulator using Hybrid Petri Nets, is particularly developed to model pathways from domain ontologies GlycO, EnzyO and ReactO. Since complete data is not available in domain ontologies, the simulation results have not not yet been validated. In the near future, we are expecting more time based experimental results from the Glycomics project (http://glycomics.ccrc.uga.edu). Once these results are available, this software can be used rigorously to model pathways from our domain ontologies.

Ontology driven simulation [31] allows simulation models to be stored at high-level and translated on demand into executable models. The Discrete-event Modeling Ontology called DeMO [32, 33] is able to represent many of the commonly used types of simulation models. The Model types are classified into four basic types: state-oriented, event-oriented, process-oriented and activity-oriented. Some hybrid models that include smooth transitions on continuous variables are also included.

We are currently applying Ontology Driven Simulation to the modeling of biochemical pathways [34] involved in glycan biosynthesis. Our current work focuses on modeling pathways using Hybrid Functional Petri Nets [35]. Transitions involving continuous places utilize ordinary differential equations to determine the amount of substrate that is converted into product by a reaction. Information is extracted from the following domain or application ontologies (Glyco [glycans], Enzyo [enzymes] and Reacto [reactions]) to customize and parameterize the Petri net models. The DeMO ontology facilitates integration of this information and supports code generation for multiple simulation engines. In a similar vein,

for multiscale modeling of human physiology, [36] developed an ontology to facilitate the integration of multiple types of models.

Bibliography

[1] Reddy, V. N., Mavrovouniotis, M. L. and Liebman, M. N. (1993). Petri net representations in metabolic pathways. Proc. Int. Conf. Intell. Syst. Mol. Biol. 1, 328-336

[2] Murata, T. Petri nets: Properties, analysis and applications. Proceedings of the IEEE Volume 77, Issue 4, Apr 1989 Page(s):541 - 580

[3] Gregory A. Silver, Osama Al-Haj Hassan, John A. Miller 2007. FROM DOMAIN ONTOLOGIES TO MODELING ONTOLOGIES TO EXECUTABLE SIMULATION MODELS. Proceedings of the 2007 Winter Simulation Conference

[4] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. and Kummer, U. (2006) COPASI - a COmplex PAthway SImulator. Bioinformatics, Vol 22, No. 24, pp.3067-3074.

[5] Muffy Calder, David Gilbert, Walter Kolch, Keith van Rijsbergen, Brian Ross. BPS - A Software Tool for Simulation and Analysis of Biochemical Networks: University of Glasgow, 2002

[6] David R, Alla H and Bail J, Proc. 1st Int. ECC European Control Conference, Grenoble, 1472-1477 (1991)

[7] Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M. and Miyano, S. (2003) In Silico Biol. 3, in the press

[8] Genrich, H., Ku ffner, R. and Voss, K. (2001) Int. J. STTT 3, 394-404

[9] Cornish-Bowden, A. and Hofmeyer, J.H, 1991. MetaModel: a program for modling and control analysis of metabolic pathways on the IBM PC and compatibles. Comput Applic. Biosci. 7, 89-93

[10] Mendes, P., 1993 GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. Comput. Applic. Biosci. 9, 563-571

[11] Mendes, P., 1997 Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. Trends Biochem. Sci. 22, 361-363.

[12] Sauro, H.M., 1993. SCAMP: a general-purpose simulator and metabolic control analysis program. Comput. Applic. Biosci. 9, 441-450

[13] Barshop, B.A., Wrenn, R.F., and Frieden, C., 1983. Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM-a flexible, portable system. Anal. Biochem. 130, 134-145

[14] Ehlde, M. and Zacchi, G., 1995. MIST: a user-friendly metabolic simulator. Comput. Applic. Biosci. 11, 201-207

[15] Tomita, M., Hashmoto, K, Takahasi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C.A., 1999. E-CELL: software environment for whole-cell simulation. Bioinformatics 15, 72-84

[16] Boanerges Aleman-Meza1, Yihai Yu1, H.-B. Schttler2, Jonathan Arnold3, Thiab R. T. KINSOLVER: A simulator for computing large ensembles of biochemical and gene regulatory networks.

[23] Murray, J. D., 1993. Mathematical Biology. Springer-Verlag, NY

[17] Weaver, D.C., Workman, C.T. and Stormo, G.D., 1999. Modeling regulatory networks with weight matrices. Pac. Symp. Biocomput. 4, 112-123

[18] Murphy, K. and Mian, S., 1999. Modeling gene expression data using dynamic Bayesian networks. Technical Report. Berkeley.

[19] Huang, S., 1999. Gene expression profiling, genetic networks and cellular states: an integrating concept for tumorigenesis and drug discovery. J. Mol. Med 77, 469-480

[20] Shmulevich, I., Yli Harja, O. and Astola, J., 2001. Inference on genetic regulatory networks under the best-fit extension paradigm. In Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP-01), June 3-6, Maryland, Baltimore.

[21] Shmulevich, I, Dougherty, E.R., Kim, S., and Zhang, W., 2002. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics 18, 261-274

[22] Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1):27-30, 2000.

[23] P. D. Karp, M. Riley, M. Saier, I. T. Paulsen, I.T., S. M. Paley and A. Pellegrini-Toole, The EcoCyc and MetaCyc databases. Nucleic Acids Res. 28 (2000) 56.

[24] M. Krull, N. Voss, C. Choi, S. Pistor, A. Potapov, and E. Wingender, TRANSPATH: an integrated database on signal transduction and a tool for array analysis. Nucleic Acids Res. 31 (2003) 97.

[25] http://biocyc.org/

[26] http://www.genome.jp/kegg/

[27] http://www.ingenuity.com/

[28] W.S. York, K. Kochut, J.A. Miller, M. Eavenson. unpublished results

[29] Thomas, C., J., A. P Sheth,, and W. S. York. 2006. Modular Ontology Design Using Canonical Building Blocks in the Biochemistry Domain. In Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS). Fairfax, Virginia: IOS Press.

[30] Gregory A. Silver, Osama Al-Haj Hassan, John A. Miller 2007. FROM DOMAIN NTOLOGIES TO MODELING ONTOLOGIES TO EXECUTABLE SIMULATION MODELS. Proceedings of the 2007 Winter Simulation Conference

[31] Benjamin, P., Akella, K.V., Malek, K., and Fernandes, R. 2005. An Ontology-Driven Framework for Process-Oriented Applications of the 37th Winter Simulation Conference.

[32] Fishwick, P.A., Miller, J.A. 2004. Ontologies for Modeling and Simulation: Issues and Approaches. Proceedings of the 36th Winter Simulation Conference.

[33] Miller, J.A., Baramidze, G.T., Sheth, A.P., and Fishwick, P.A. 2004. Investigating Ontologies for simulation Modeling. Proceedings of the 37th Annual Simulation Symposium.

[34] Silver, G.A., Hassan, O.A., and Miller, J.A. 2007. From Domain Ontologies to Modeling Ontologies to Executable Simulation Models. Proceedings of the 39th Winter Simulation Conference.

[35] Matsuno, H., Doi, A., fujita, S., Nagasaki, M., and Miyano, S. 2004. Constructing Biological pathway Models with Hybrif Functional petri Nets. In silico Biology.

[36] Erson, E.Z., Cavusoglu, M.C. 2007. Ontology based Design for integrative Simulation of Human physiology. Proceedings of the International Symposium of Health Informatics and Bioinforamtics.

[37] Agarwal, P.K. and Alam, S.R. 2007. On the path to Enable Multi-Scale Biomolecular Simulations on PeraFLOPS Supercomputer with multi-core Processors. Proceedings of HICOMB 2007.

[38] Stein, M., Gabdoulline, R.r. and Wade, R.C 2007. Bridging from Molecular Simulation to Biochemical Networks. Current Opinion in Structural Biology, Vol. 17, No. 2

[39] Uhrmacher, A.M., Degenring. D. and Zeigler, B.P. 2005. Discrete Event Multi-Level Models for Systems Biology. Proceedings of the 37th Winter Simulation Conference.

[40] Hybinette, M., Kraemer, E., Xiong, Y., Matthews, G., and Ahmed, J. 2006. SASSY: A Design for a scalable Agent-based Simulation System using a Distributed Discrete Event Infrastructure. Proceedings of the 38th Winter Simulation Conference.

[41] Nicol, D.M., and Roy , S. 1991. Parallel Simulation of Timed Petri-Nets. Proceedings of the 23rd Winter Simulation Conference.

[42] Park, S., hunt, A., Zeigler, B.P. 2006. Partitioning for Distributed and Parallel Simulation of Decomposable Constructive Models, Simulation, Vol. 82, and No.12

[43] Matthew Eavenson, Maciej Janik, Shravya Nimmagadda, John A. Miller, Krys J. Kochut, and William S. York. GlycoBrowser: A Tool for Contextual Visualization of Biological Data and Pathways Using Ontologies.Bioinformatics Research and Applications (2008), pp. 305-316.

[44] Mesarovic, M. D. (1968). Systems Theory and Biology. Springer-Verlag.

[45] Poolman, M. G. (2006) ScrumPy: Metabolic modelling with Python. IEE Proc. Syst. Biol. 153, 375378

[46] M. Nagasaki, A. Doi, H. Matsuno, S. Miyano, A versatile Petri net based architecture for modeling and simulation of complex biological processes, Genome Informatics 15(1): 180-197 (2004).

[47] Xiao-Qin Xia and Michael J. Wis, DiMSim: A Discrete-Event Simulator of Metabolic Networks, J. Chem. Inf. Comput. Sci., 2003, 43 (3), pp 10111019

[48] Atkins,P.W - Physical Chemisty/P.W.Atkins.- 6th edition (ISBN 0-7167-2871-0)

[49] Berg, J. M.; Tymoczko, J. L.; Stryer, L.: Biochemistry. 5th ed., Freeman, New York 2002

[50] Petri, C.A. (1962) PhD Thesis, Institut fu r Instrumentelle Mathematik, Bonn

[51] Gruber, G.R. 1993. Toward Principles for the Design of Ontologies Used in Knowledge Sharing. International Journal of Human Computer Studies. Vol. 45: 907- 928

[52] Lacy, L. W. 2006. Interchanging Discrete-Event Simulation Process-Interaction Models using the Web Ontology Language OWL. Unpublished PhD Dissertation, Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, Florida.

[53] Noy, N. F. and D. L. McGuinness. 2001. Technical Report KSL-01-05, Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory, Stanford California.

[54] McGuinness, D. L. and F. Harmelen. 2004. OWL Web Ontology Language Overview. Available at http://www.w3.org/TR/owl-features/ [Accessed November 23, 2008].

[55] Alla, H., R. David, 1998. Continuous and Hybrid Petri Nets, Journal of Circuits Systems Computers, Vol. 8, No. 1, pp. 159-88

[56] Drath R., 1998. Hybrid object nets: an object oriented concept for modeling complex hybrid systems. Proceedings of the Hybrid Dynamical Systems, 3rd International Conference on Automation of Mixed Processes, ADPM'98. 437–42.

[57] Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M. and Miyano, S. (2003) In Silico Biol. 3, in the press

[58] Reddy, N, N Liebman, and M Mavrovouniotis. 1996. Qualitative Analysis of Biochemical Reaction Systems. Computers in Biology and Medicine, Vol. 26, Issue 1, pp. 9-24.

[59] I. Schomburg, A. Chang, and D. Schomburg, "Brenda, enzyme data and metabolic information." Nucleic Acids Res, vol. 30, no. 1, pp. 47-49, January 2002.

[60] Isabel Rojas, Martin Golebiewski, Renate Kania, Olga Krebs, Saqib Mir, Andreas Weidemann, Ulrike Wittig, SABIO-RK: a database for biochemical reactions and their kinetics, BMC Systems Biology 2007, 1(Suppl 1):S6 (8 May 2007)

[61] John A. Miller, Rajesh Nair, Zhiwei Zhang and Hongwei Zhao, "JSIM: A Java-Based Simulation and Animation Environment," Proceedings of the 30th Annual Simulation Symposium (ANSS'97), Atlanta, Georgia (April 1997) pp. 31-42.

[62] http://www.geneontology.org/

[63] Karen Eilbeck, Suzanna E Lewis, Christopher J Mungall, Mark Yandell, Lincoln Stein, Richard Durbin and Michael Ashburner. The Sequence Ontology: a tool for the unification of genome annotations. Genome Biology 2005, 6:R44 (doi:10.1186/gb-2005-6-5-r44)

[64] Shravya Nimmagadda, A. Basu, M. Eavenson, J. Han, M. Janik, R. Narra, K. Nimmagadda, A. Sharma, K.J. Kochut, J.A. Miller and W.S. York, "GlycoVault: A Bioinformatics Infrastructure for Glycan Pathway Visualization, Analysis and Modeling," Proceedings of the 5-th International Conference on Information Technology: New Generations (ITNG'08), Las Vegas, Nevada (April 2008)

[65] Turnitsa, C. and A. Tolk. 2005. Evaluation of the C2IEDM as an Interoperability-Enabling Ontology. In Proceedings of the European Simulation Interoperability Workshop. Toulouse, France.

Appendix A

Michaelis-Menton mechanism

The rate of a enzyme-catalyzed reaction in which a substrate S is converted into product P is dependent on the concentration of enzyme E (though the enzyme undergo no net change) [48].

This (depicted in Figure A1) can be represented as the following equation:

$$E + S \rightleftharpoons ES \rightarrow E + P \qquad k'_a, k_a, k_b$$

In the above equation ES denotes an enzyme-substrate complex. Hence [ES] can be represented as

$$[ES] = \frac{k_a[E][S]}{k'_a + k_b}$$

$[E]$ and $[S]$ are the concentrations of the free enzyme and free substrate. if $[E]_0$ is the total concentration of the enzyme then

$$[E] + [ES] = [E]_0$$

Therefore,

$$[ES] = \frac{k_a([E]_0 - [ES])[S]}{k'_a + k_b}$$

which rearranges to

$$[ES] = \frac{k_a[E]_0[S]}{k'_a + k_b + k_a[S]}$$

it follows that the rate of the formation of product is

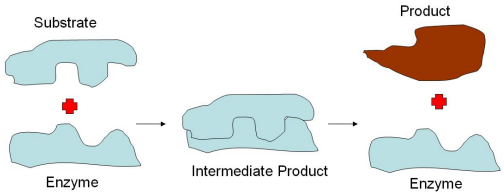

Figure A.1: The basis of the Michaelis-Menten mechanism of enzyme action.

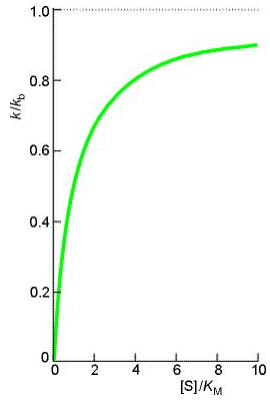

Figure A.2: The variation of the effective rate constant k with substrate concentration according to the Michaelis-Menten mechanism.

$$\frac{d[P]}{dt} = k[E]_0 \qquad k = \frac{k_b[S]}{K_M + [S]}$$

Here the Micaelis constant, K_M is

$$K_M = \frac{k'_a + k_b}{k_a}$$

Based on the above equations, the rate at which substrate is converted to product varies linearly with enzyme concentration. However, in more complicated manner with the concentration of substrate in Figure A2.

When $[S] >> K_M$, the rate law reduces to

$$\frac{d[P]}{dt} = k_b[E]_0$$

This results means that under these conditions, the rate is constant. There is so much $[S]$ present that it remains effectively the same concentration even though products are being formed. Moreover, the rate of formation of product is maximum, and $k_0[E]_0$ is called maximum velocity.

Appendix B

GlyMpse Implementation Details

GlyMpse software is developed based on JSIM Library. JSIM is a simualtion library which provied support for variety of model environments. The principal components of GlyMpse simulation environment are the PetriNet and Animator classes.

The PetriNet class provides a simulation engine for Hybrid Colored Petri Nets. This class also contains subclasses DPlace, CPlace, DArc, CArc and Transition to represent components of Petri Net discrete place, continuous place, discrete arc, continuous arc and transition respectively. When a transition in the priority queue is fire, deltaS which represents the amount of substrate converted to product, is calculated using Michaelis rate equation and Ranga Kutta differentail equation integrator.

A pathway model is created by extending PetriNet class with pathway specific places, arcs, transitions and kinetic parameters. Animator is an interface that specifies the commands to create, destroy, move and scale components in an animated canvas. Besides these classes, Graph class displays graphical results of substrate concentrations over time.

After runnig a model using GlyMpse, an Excel file, a time vs. concentration and an animator ouputs are generated. The spread-sheet file contains the time series concentration of each substrate participating in the model. This data can be used for futher analysis.

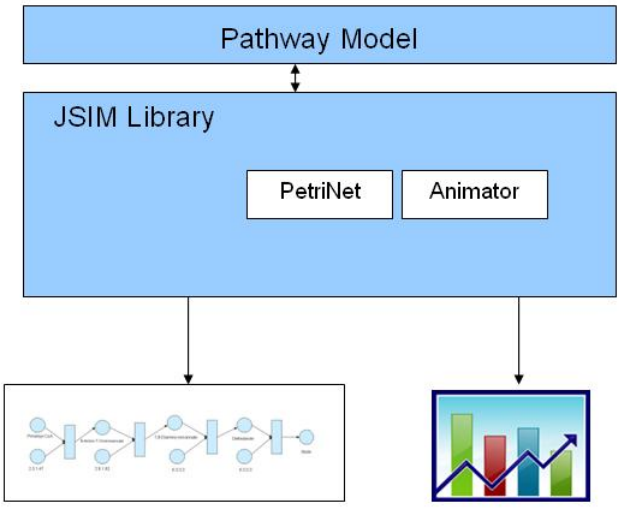

Figure B.1: GlyMpse implementation architecture