

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

5-2014

Pathway Pioneer: A Web-Based Graphical Tool for the Organization and Flux Analysis of Metabolic Networks

Sumit Kumar Singh
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Singh, Sumit Kumar, "Pathway Pioneer: A Web-Based Graphical Tool for the Organization and Flux Analysis of Metabolic Networks" (2014). *All Graduate Theses and Dissertations*. 2796.

<https://digitalcommons.usu.edu/etd/2796>

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



PATHWAY PIONEER: A WEB-BASED GRAPHICAL TOOL FOR THE
ORGANIZATION AND FLUX ANALYSIS OF METABOLIC NETWORKS

by

Sumit Kumar Singh

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Computer Science

Approved:

Dr. Nicholas.S.Flann
Major Professor

Dr.Xiaojun Qi
Committee Member

Dr.Tung Thanh Nguyen
Committee Member

Dr. Mark R. McLellan
Vice President for Research and
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2014

Copyright © Sumit Kumar Singh 2014

All Rights Reserved

ABSTRACT

Pathway Pioneer: A Web-Based Graphical Tool for the Organization and Flux Analysis of
Metabolic Networks

by

Sumit Kumar Singh, Master of Science

Utah State University, 2014

Major Professor: Dr. Nicholas.S.Flann
Department: Computer Science

As stoichiometric metabolic models increase in complexity and fidelity, design and reconstruction tools are urgently needed to increase the productivity of this time-consuming process. Engineers require software for the exploration, evaluation, and rapid analysis of model alternatives within an intuitive visualization and data management framework. This thesis introduces such a tool: Pathway Pioneer (www.pathwaypioneer.org), a web-based system built as a front-end graphical user interface to the flux balance analysis tool COBRA. Pathway Pioneer adds additional functionality for customized network layout and model-engineering collaboration through shared models and model version control. Pathway Pioneer is a dynamic, clickable, browser-based visualization system for metabolic network models retrieved from databases such as BiGG or developed in-house as SBML or XLS compliant files. The user can customize the network layout to visually organize the metabolites and reactions into functional modules. The tool supports zooming and panning, level-of-detail control, flux visualization, keyword searching, and hierarchical subsystem organization. A reaction may be knocked out, set as an objective, looked up in a database or many other operations by a single click on the visualized network. Following each operation the visualization is refreshed with the new flux values. The system supports model revision

control to manage alternative network configurations and supports sharing of models and layouts to the broader community. By moving the computationally intensive model analysis from the user computer to remote servers, Pathway Pioneer enables the application of high performance cloud-based resources for greater efficiency and scalability. I demonstrate the utility of Pathway Pioneer through application in model reconstruction and analysis of many standard models and also two new models under development: Eukaryotic multi-compartment Chinese Hamster Ovary (Cho) cells and in a large-scale *Escherichia coli* system for bio-manufacturing.

(35 pages)

PUBLIC ABSTRACT

Pathway Pioneer: A Web-Based Graphical Tool for the Organization and Flux Analysis of Metabolic Networks

SUMIT KUMAR SINGH

Every so often the field of genome-scale metabolic networks undergoes radical changes, each bringing its own complexity and challenges. The existing visualization tools, owing to their inability to assist in generating layouts suitable for easy readability and conceptualization, hinder further advancement. The lack of in-depth comprehensible representation of the intricacies of interaction and arrangements within the network topology prevents researchers from being able to advance at a faster speed despite the growing number of networks and systems. The far and wide diverse standards, layout schemas and network formats of the biological realm contribute to the existing challenges of developing visualization tools that can render user-friendly, dynamic, and all-encompassing visualizations of the metabolite reactions. I aim to provide a front-end graphical user-interface to the existing flux balance analysis tool *Cobra-Py*, which incorporates most of the above mentioned desired functionalities. Pathway Pioneer is a dynamic web-based tool based on Web 2.0 technologies, built to provide a complete suite to help researchers analyze, engineer, and customize metabolic network layouts suitably for optimal comprehension. Its graphical user interface makes it user friendly and interactive. The user has maximum flexibility to design layouts per requirements without worrying about details of convention such as joined in and outflow of reactions, disconnected co-factors, and connected metabolites. Furthermore, all layouts can be shared with co-researchers, which makes Pathway Pioneer a collaboration platform for Model Developments and Innovations. The R&D curve for model reconstruction at the genome scale would see positive growth with the aid of this tool. This tool aims to overcome the limitations of existing visualization tools while simultaneously complementing them.

CONTENTS

	Page
ABSTRACT	iii
PUBLIC ABSTRACT	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION	1
2 BACKGROUND	3
2.1 Need For Flux Visualization	3
2.2 Study on Current Visualization Tools	3
3 SYSTEM OVERVIEW	7
3.1 Network Navigation Panel	7
3.2 ONE Click Operations	9
3.3 Zooming and Panning	10
3.4 Import and Export Capabilities	10
3.5 Repository and Versioning Control	11
3.6 Sharing and Cloud Storage : Anytime Anywhere Concept	11
4 SYSTEM IMPLEMENTATION	14
5 SYSTEM DEMONSTRATION	17
6 CONCLUSIONS	21
REFERENCES	23

LIST OF TABLES

Table		Page
2.1	Table showing the availability of standard features in currently popular visualization tools where X-axis representing features and Y-axis representing different tools. Each row characterizes a different tool and each column is a feature. A green tick represents the feature is available in the tool and Red represents the opposite.	5

LIST OF FIGURES

Figure	Page
2.1 Image showing importance of visualization of flux. (1) Different subsystems with a functional objective can be segregated and then their relationship and dependency on others can be studied. (2) Flux flow through the network can be easily identified using thickness of edges and their color values to facilitate an intuitive understanding of network function. (3) Significant pathways can be easily seen on the network compared to tabular values.	4
2.2 Layouts of Escherichia coli generated by [A] FAME and [B] Pathway Pioneer. It is evident from the images that the layout produced by PathwayPioneer is easier to understand compared to that of FAME. Note both images show flux as thickness and color of the graph edges.	5
3.1 Image showing system page of PP with various functionality depicted. (1) System Navigation Panel (2) Search Bar (3) One Click Operation Menu (4) Export Link (5) Modification History Panel (6) Key Legend (7) Edit Layout Link to transfer the control to edit mode	8
3.2 Image showing Repository page of User in PP depicting: (1) Button to create version (2) Versioning of Model (3) SBML/XLS converter and validator plugin	12
3.3 Image illustrating the study of ethanol production in (A) Aerobic and (B) Anaerobic growth condition on Escherichia coli core model. User can save these two different studied model of same parent model in different subversion and study them independently	12
4.1 The Three-Layered Architecture of PP: Client Layer, Communication Layer, and Server Layer. Client Layer receives the user request and send to Communication Layer which is a Web-Server hosted in C#. Server Layer hosts COBRA-PY library where the FBA is performed	15
4.2 Graph showing the time taken (Y-Axis) by dummy users on PP to upload the model using Markov Testing Technique. Fifty parallel users were tested and 87 samples (X-Axis) were collected.	16
5.1 Image from PP displaying (1) the Glycolysis/Gluconeogenesis subsystem and (2) zoomed-in image with reactions that interact with 3-phosphoglycerate .	18

5.2	Image from PP showing Glycine and Serine metabolism using glucose as a carbon source (1) 3-pg is converted into serine (2) serine is converted into glycine through GHMT2r. As a result of growth on glycerol (3) reactions producing serine and (4) glycine have lower flux values than those using glucose as a carbon source	19
5.3	Image from PP showing (1) On updating the glucose exchange reaction lower bound from -10 to 0 and (2) the effect of growth on glycerol by changing glucose exchange reaction	19
5.4	Image from PP displaying (1) with the growth of glycerol, the flux in the glycolysis is much lower and (2) zoomed-in reactions that interact with 3-phosphoglycerate	20

CHAPTER 1

INTRODUCTION

Genome scale modeling and reconstruction has seen a rapid growth in the past decade. Pathway reconstruction and constraint based modeling has gained attention from researchers due to its growing impact in synthetic biological and biomedical research. Oberhardt et al. [1] and Nielsen et al. [2] list five major domains of research and application of these metabolic reconstructions: (1) contextualization of high-throughput data to identify the genotypes and phenotypes of organism, (2) discovery of network properties, (3) identification of important pathways to guide hypothesis-driven discovery in synthetic biology, (4) interrogation of multi-species relationships to gain insight into evolutionary theory, and (5) guiding wet lab experiments in metabolic engineering. In recent years, genome scale modeling has contributed to drug discovery, bio-fuels production and defense systems, to name a few. Market capitalization of Bio-Tech industry has seen a tremendous growth in past 10 years reaching to 477\$ billion in 2012-13 in this industry [3].

This focus on applications in genome scale modeling creates the need for graphic visualization analysis and tools that accelerate the process. In response, many visualization tools have been developed, reviewed in [4]. Most of the listed tools provide visualization of data and its manipulation to some extent, but few provide an integrated suite for network visualization, flux analysis and modeling. A new tool called Pathway Pioneer (PP) has been developed to overcome the limitations of current tools, encompassing their strengths to provide the user with an integrated modeling, analysis and visualization environment. In Pathway Pioneer metabolic models can be retrieved from databases such as BiGG [5], KEGG [6] or can be built in-house *de novo* in SBML or XLS standard formats. Pathway Pioneer is built as a GUI interface for COBRA-Py [7], an established flux balance analysis tool and provides the functionality to calculate and visualize flux with one click operation,

edit the network layout, navigate through the network with an embedded search engine, along with extensive export and import features. Pathway Pioneer is deployed with a semi-automatic layout algorithm where the user has the flexibility of customizing the network layout according to their choice using a graphical edit layout toolbox. The tool also provides the user the ability to save and share the network layout along with their specific metabolic models. PP is developed using Web 2.0 technology as a 3 layer architecture consisting of client, communication and server layer. It runs on web server IIS 7.0 and employs server-hosted parallel and multiprocess environment to facilitate the multiuser environment and reduce the time latency in resource-intensive visualization and flux-balance analysis algorithms.

CHAPTER 2

BACKGROUND

Previous section talks about the necessity of accelerated research and widespread use of genome scale modeling and metabolic engineering. Nielsen [2] talks about various applications of metabolic engineering. The focus on rapid application development in metabolic engineering makes it indispensable for a flux visualization and analysis tool. In this section, we provide a background for the need for visualization, available tools and techniques and their strength and limitations.

2.1 Need For Flux Visualization

Flux balance analysis (FBA) [8] is a well known technique for steady state analysis of biochemical networks specifically genome scale reconstructions. COBRA-Py [7] is a well established tool for FBA, but it is directly usable through an interface that takes commands and produces data in textual form. Visualization of flux analysis results will help to: (1) better understand the interconnection between metabolites and cellular machinery of species as discussed in [9, 10], (2) analyze the flux flow through the metabolic networks that regulate among different pathways, (3) identify important pathways [11] for synthetic biology, species manipulation and industrial applications among others, and (4) significantly increase productivity in comparison to textual analysis, thereby facilitating rapid development of metabolic reconstructions. Figure 2.1 validates the need of visualization in metabolic engineering.

2.2 Study on Current Visualization Tools

Pavlopoulos et al. [4] provide a survey on recently developed tools for analysis and visualization of genome scale networks. These visualization tools face multiple challenges

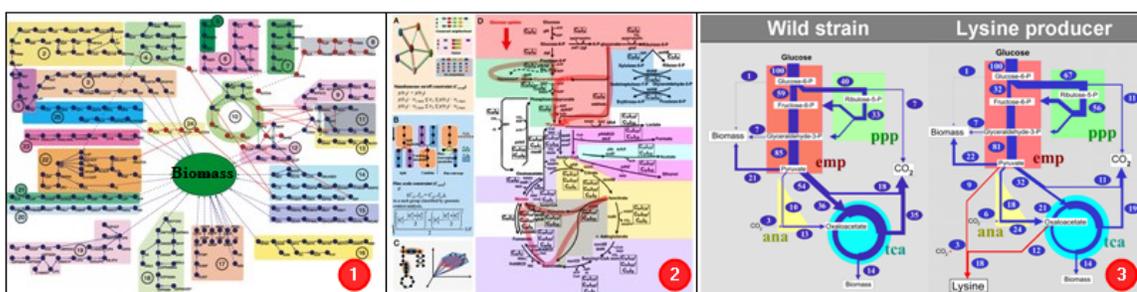


Figure 2.1: Image showing importance of visualization of flux. (1) Different subsystems with a functional objective can be segregated and then their relationship and dependency on others can be studied [12]. (2) Flux flow through the network can be easily identified using thickness of edges and their color values to facilitate an intuitive understanding of network function [13]. (3) Significant pathways can be easily seen on the network compared to tabular values.

including a diversity of standards for encoding the species data, how to manage protein interactions, signalling and regulatory pathways along with metabolic processes in a unified framework. Most importantly, is the lack of tools that produce comprehensible network layouts. Pavlopoulos et al. [4] list a set of features that an effective visualization tool should possess. Table 2.1 provides a matrix of these essential features and an evaluation of the current tools that are used extensively by the modeling community [14–21]. The table provides information about which tool possesses which features, in conjunction with the importance of those features to metabolic visualization tools. To summarize the table, most of the tools provide network visualization of its topological decomposition, allow standard import and export format and are compatible with other tools. Few of them also feature integrated network modeling and simulation, along with network editing and its customization.

Few of the tools provide a platform for network modeling and simulation integrated with dynamic visualization where metabolic flux is visualized on the network itself. Most tools are stand alone systems except Patika [18] and Prometra [21], and none of the existing tools provide a platform for engineering and reconstruction of genome scale models. FAME [22] and AVIS [23] are other web based tools developed for flux analysis and modeling of metabolic networks. FAME lacks the dynamic visualization of flux on the network. For each flux change, a new model has to be uploaded. Moreover, the layout generated by FAME

Table 2.1: Table showing the availability of standard features in currently popular visualization tools where X-axis representing features and Y-axis representing different tools. Each row characterizes a different tool and each column is a feature. A green tick represents the feature is available in the tool and Red represents the opposite.

	Network Editing	Visualize Network Topology	Network Modeling & Simulation	Web Based	Tools Compatability	Model Engineering	Network Customization & Annotation	Standard Format
Arcadia	✓	✓	✗	✗	✓	✗	✗	✓
Cytoscape	✓	✓	✓	✗	✓	✗	✓	✓
Paint4Net	✗	✓	✓	✗	✓	✗	✗	✓
Proviz	✗	✓	✗	✗	✓	✗	✓	✓
Patika	✓	✓	✗	✓	✓	✗	✗	✓
Pajek	✗	✓	✗	✗	✗	✗	✓	✗
Cell Designer	✓	✓	✓	✗	✓	✗	✓	✓
Prometra	✗	✓	✗	✓	✓	✗	✗	✓

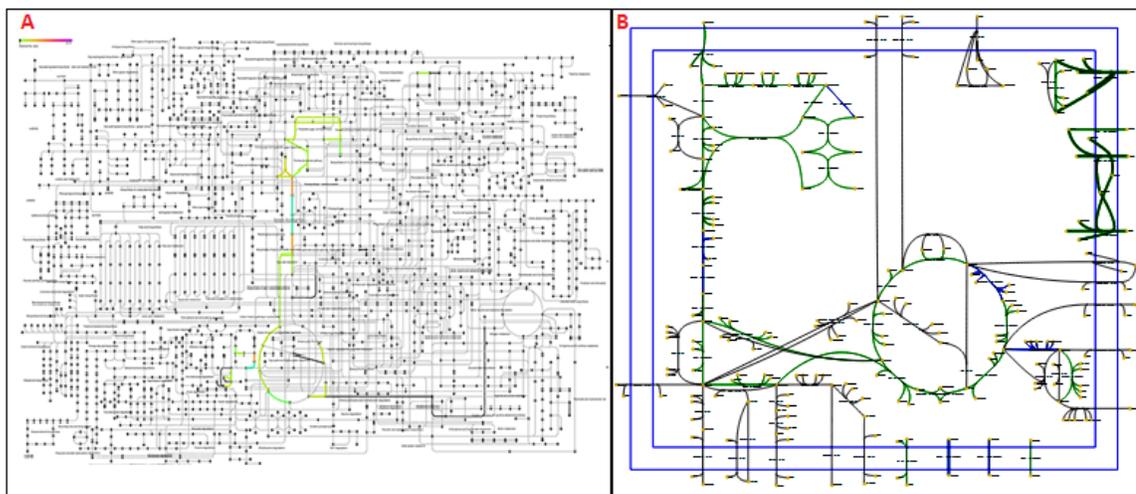


Figure 2.2: Layouts of Escherichia coli generated by [A] FAME and [B] Pathway Pioneer. It is evident from the images that the layout produced by PathwayPioneer is easier to understand compared to that of FAME. Note both images show flux as thickness and color of the graph edges.

is complex and unintuitive, reducing its utility value in terms of analysis and knowledge inference compared to those produced in Pathway Pioneer (see Figure 2.2). AVIS provides only visualization of cell signaling networks and does not support modeling and dynamic analysis. In summary, most of the tools concentrate on visualization or analysis of networks

but none support both. Most tools produce a complex and unintuitive layouts that does not scale to large scale metabolic networks.

CHAPTER 3

SYSTEM OVERVIEW

In this section we provide a brief overview of all the capabilities of Pathway Pioneer (PP) for metabolic model building, flux analysis, network customization, annotation and its visualization. The web page layout of PP is divided into three panels: left, center and right panels respectively as shown in Figure 3.1. The center panel visualizes the network and flux. The left and right panels enable navigation analysis functions and provide model information respectively. The two side panels can be collapsed or expanded as needed. The left-panel (Figure 3.1 (1)) provides for a textual based menu that lists the subsystems, reactions, metabolites hierarchically along with specific information once the model has been uploaded Figure 3.1 (2) shows the embedded search engine that provides search-and-navigate capability to a particular reaction or metabolite in the model. Additional analysis tool available on the left panel are described later. The right-panel (Figure 3.1 (6)) is the Key legend that provides high level statistical and summary data of the model such as current growth rate, objective reaction name and counts of reactions and metabolites along with the key for blue, green and grey flux annotations. The central-panel of the page is a graphical interface that displays the metabolic network and is visually synchronized with the textual panel information. The textual and graphical interfaces are functionally equivalent in that they provide identical access to information and commands, either visually or textually. Figure 3.1 illustrates the model display page of PP for a small *Escherichia coli* model. The next section describes each section in detail.

3.1 Network Navigation Panel

To manage the complex nature of the network hypergraph, it is divided into topological subgraphs based on functional modules of biochemical pathways [24]. The left-navigation

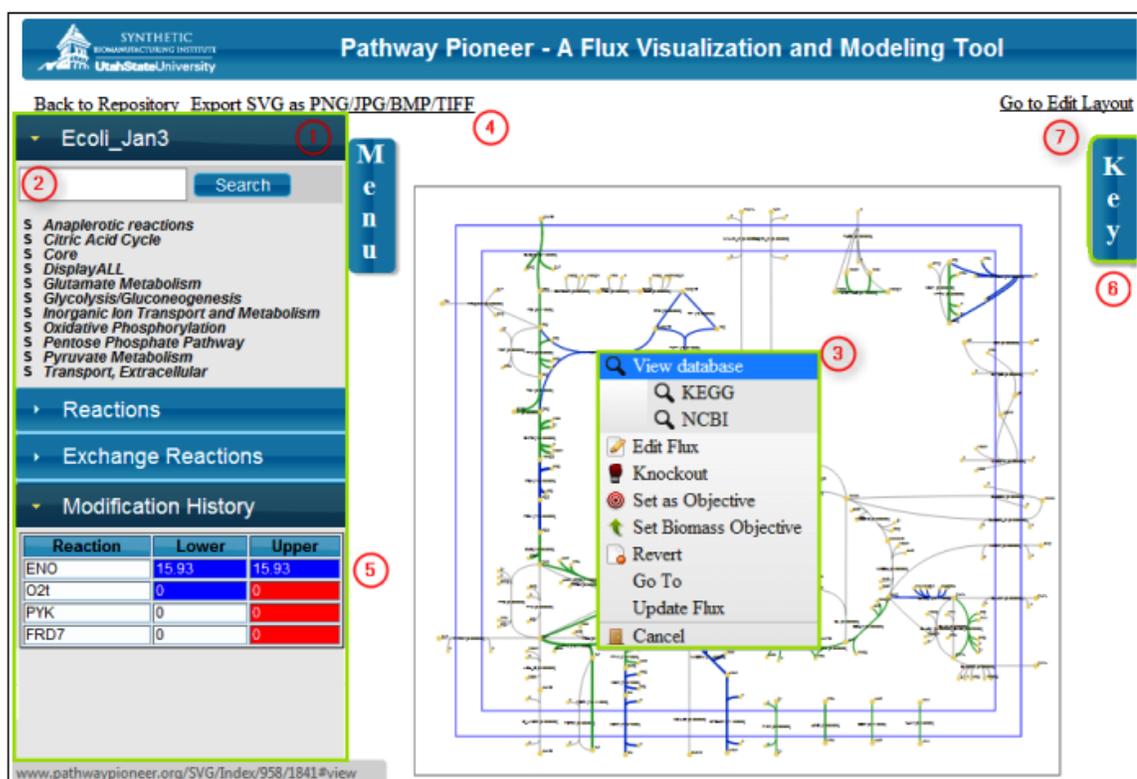


Figure 3.1: Image showing system page of PP with various functionality depicted. (1) System Navigation Panel (2) Search Bar (3) One Click Operation Menu (4) Export Link (5) Modification History Panel (6) Key Legend (7) Edit Layout Link to transfer the control to edit mode

toolbox contains the SubSystem menu containing the list of all subsystems present in the network model. Clicking on any subsystem will cause a navigation to the selected subsystem, expand the text menu to show contained reactions and display the reactions contained in that subsystem on the central panel. Further clicking on any reaction either on the left or center panel will cause that reaction to be visually centered in the display (center) panel. There follows a list of the function tabs in the left navigation panel.

- **Reaction Tab** contains a list of all the reactions in the network. Along with each reaction two columns are provided that give the current state of flux for that reaction. Right clicking on any of the reactions will pop up a sub menu which lists options to perform FBA on the network (described in the one-click operation in section 3.2).

- **Exchange Reaction Tab** is specialized to maintain only the exchange, demand and sink reactions. It behaves identically to the reaction tab described above.
- **Modification History Tab** shown in Figure 3.1 allows the user to inspect the analysis state of the current model by recording the edit flux and knockout reaction operations performed on the metabolic network. A user has the capability to undo the already committed transactions to their previous state. In this way, the user can move through design alternatives in network configurations, making and undoing changes.
- **Biomass Tab** is specialized for the most significant reaction in any bioproduct: the biomass reaction. It contains a list of all the metabolites present in the biomass reaction and their flux.
- **Metabolite Tab** contains the information of all metabolites in the network in a tabular form along with the chemical formula of the metabolites.

3.2 ONE Click Operations

The core functionality of PP provides the user with capabilities for rapid exploration and flux analysis on metabolic networks with *one click* operations. A user can apply the following functionality by right clicking on any reaction name in either the textual (left panel) or graphical (center panel) in PP. The popup menu is as shown in Figure 3.1 (3):

- **View Database** provides the user with links to different standard databases such as KEGG and NCBI where a user obtains in-depth information of the specific clicked reaction.
- **Edit Flux** This option will pop up a dialog box where a user can edit the bounds on the flux passing through that single reaction. This function calls the COBRA ToolBox `changeRxnBounds` function running on the server to perform FBA and then update flux values on the network. These updated flux values are illustrated directly on the network visualization as the color and thickness of the reaction edges. The thickness indicates the amount of flux, green, blue and grey represents positive, negative and

no flux respectively. On top of each reaction edge, the flux value is shown up to 5 digits of precision.

PP also provides a user with the option to edit the flux bounds of multiple reactions at the same time by selecting each reaction, changing their flux and then clicking on **Update Flux**.

- **Knockout** This option removes the reaction from the network by setting its upper and lower bound flux to 0 and then calling the standard COBRA `changeRxnBound` function. It is used for impact analysis on the network when some particular reaction is disabled from the metabolic network.
- **Set Objective** This option allows the user to select a reaction to be the objective function. When clicked, `setObjective` function of COBRA-Py is instantiated with the selected reaction as an argument. The flux through that reaction is then maximized during FBA. The default objective function is the biomass reaction. The flux of this objective reaction is displayed in the right panel summary tab.

3.3 Zooming and Panning

PP's visualization framework utilizes the standards of SVG (Scalable Vector Graphics) that is a vector format and hence can be zoomed up to any level with no diminishment of image quality. The user can also click and translate the network panning to change the focus of attention within the central panel.

3.4 Import and Export Capabilities

Import and export capabilities are one of the most important features of any visualization tool. Pavlopoulos et al. [4] describe the diversity of file formats used in network data visualization: SBML [25], CellML [26], PSI-MI [27], CML [28]. SBML standards are the most adopted with approximately 89 software tools for genome scale modeling and visualization. Excel sheet (XLS) formats are also used for de-novo synthesis and reconstruction of

models because they are easy to understand, annotate and are easy to incrementally extend by simply adding rows as new reactions are determined.

Pathway Pioneer supports both the SBML and XLS model format. To strengthen the import facility, PP includes a plug-in: an XLS-SBML converter that decodes the information from XLS file to generate a SBML file compatible to COBRA-Py. This plug-in incorporates extensive error checking on the XLS file format to identify and help correct data entry errors.

Extensive export compatibility of Pathway Pioneer allows a user to download the network and flux visualization in various formats such as SVG, PNG and GIF. In addition to this, a user can download the SBML file of a model along with the SVG and its corresponding layout file called a map file. Map files were first used by the COBRA ToolBox to encode the layout information of the network, but were discontinued.

3.5 Repository and Versioning Control

To facilitate independent, parallel yet incremental research and development, PP provides users with their own secured account. Each account has a personal repository space where the user can store any number of metabolic models, as shown in Figure 3.2. PP also supports versioning control by which a user can create any number of versions of specific models to save different analysis states and facilitating independent analysis and design. It also enables backward compatibility and incremental design modifications on any model. For instance, if a user wants to study ethanol production, multiple versions of the same model can be created to compare the ethanol production under different growth conditions, such as aerobic and anaerobic, as shown in Figure 3.3. In this case, studying the two versions shows that growth in anaerobic conditions produces more ethanol.

3.6 Sharing and Cloud Storage : Anytime Anywhere Concept

As elucidated in the previous section, heterogeneity in network visualization formats is both a bottleneck and a challenge for the tool developers. Pathway Pioneer helps to solve this problem by providing a platform to publish the layout map file so as it can be

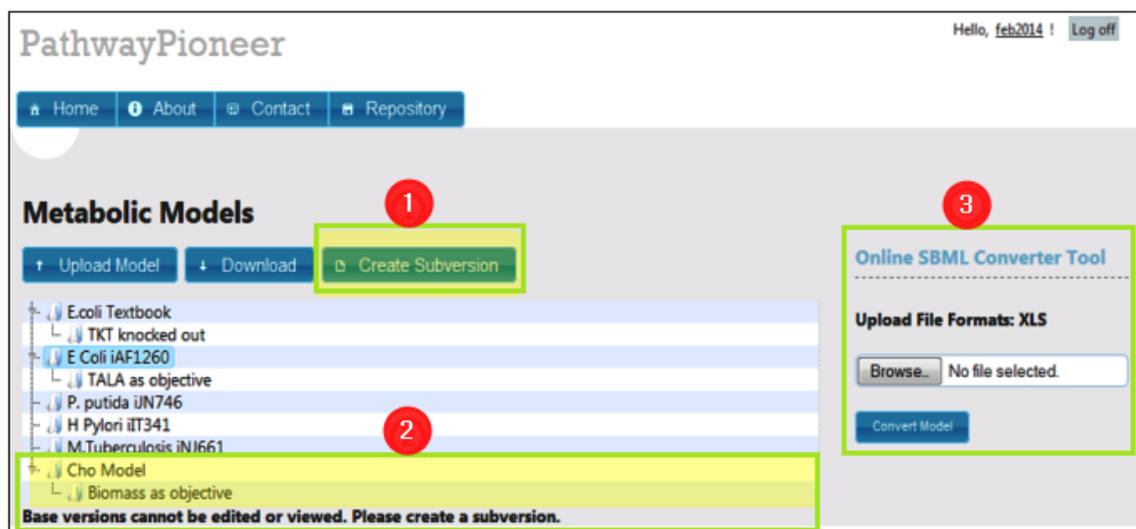


Figure 3.2: Image showing Repository page of User in PP depicting: (1) Button to create version (2) Versioning of Model (3) SBML/XLS converter and validator plugin

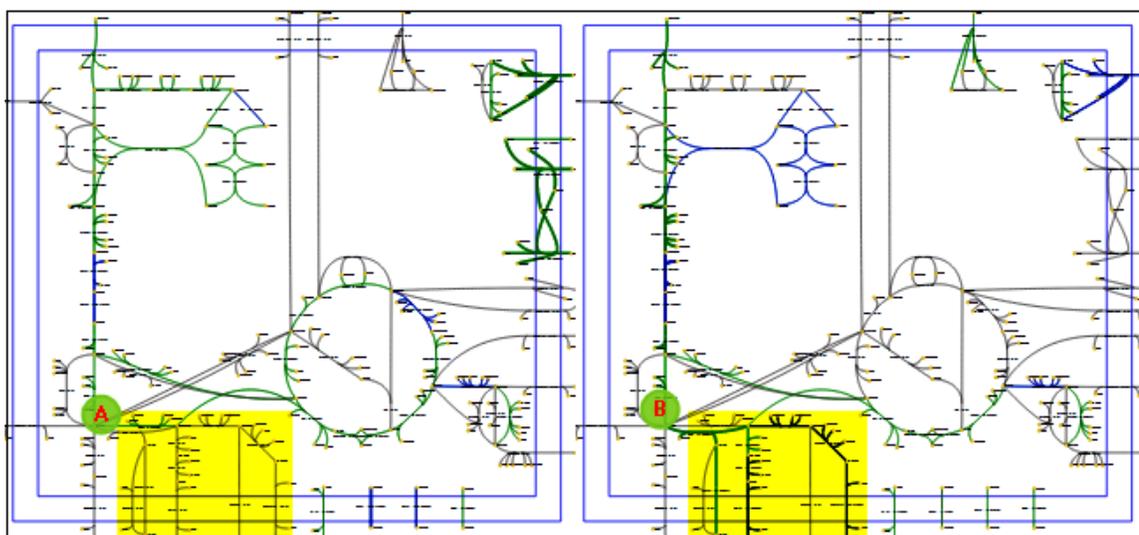


Figure 3.3: Image illustrating the study of ethanol production in (A) Aerobic and (B) Anaerobic growth condition on Escherichia coli [29] core model. User can save these two different studied model of same parent model in different subversion and study them independently

utilized by any user to layout their models. This functionality helps in simultaneous development and collaboration, saving resources and time; ultimately accelerating the research and development in the biological modeling community.

CHAPTER 4

SYSTEM IMPLEMENTATION

Pathway pioneer is a web based tool and runs on COBRA-Py in the back end.

Being a web-based system it deploys a three-layer architecture: a client layer, a communication layer and a server layer. The current system runs on windows webserver IIS 7.0. The server side programming is written in C#.net 4.0 ASP.net. The website is developed in MVC (Model/View/Controller) architectural pattern. Fig 4.1 displays the layered architecture with detailed information of data and process flow in PP.

4.0.1 Client Layer

Client interface works in web 2.0 technology and encompasses dynamic HTML5 pages generated by IIS server along with CSS, strong binding of pure JavaScript and its libraries including SVG [30], for qtip, zoom-and-pan and context menu functionality. All the client service requests are handled asynchronously using an ajax-request and when the response is returned, the user interface is dynamically updated. Asynchronous handling allows for lazy display updates, shifting the extensive computation to the back-end and providing the results to user with very less turn around time.

4.0.2 Communication Layer

The communication layer is developed using C#.net 4.0 ASP.net and functions as a request handler for the clients. It creates a multiuser and multiprocess scalable environment for minimizing the time latency in processing the extensive transactions of complex graph visualization and its analysis. It is capable of making parallel requests to the python engine where the COBRA-Py library exists. Multiprocessing allows parallelism by using all the cores available on the system hardware using a process-pool, work-queue and user-map.

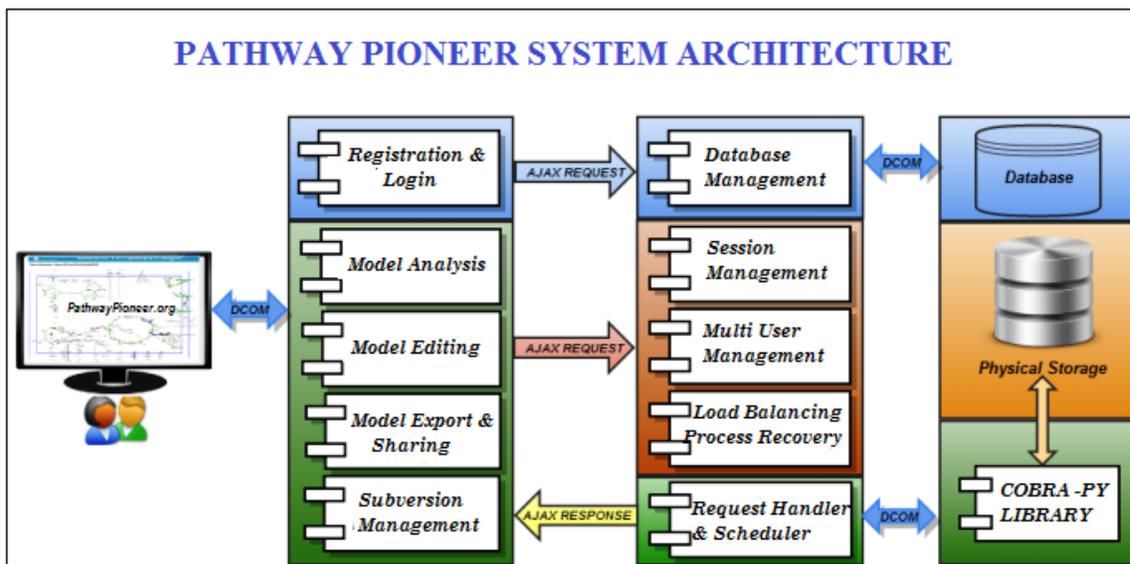


Figure 4.1: The Three-Layered Architecture of PP: Client Layer, Communication Layer, and Server Layer. Client Layer receives the user request and send to Communication Layer which is a Web-Server hosted in C#. Server Layer hosts COBRA-PY library where the FBA is performed

This layer is responsible for performing lazy updates by sending only the minimum and necessary information to the client side for updating the client window and keeping the relevant metadata information in the user repository.

To analyze the performance of multi-user and parallel environment in Pathway Pioneer, we tested the system with 50 parallel users who performed sequences of upload model and flux analysis operations (knock out reaction, edit flux, set objective) on PP. Each user is modeled as a simple Markov process calibrated to mimic sample PP users. Figure 4.2 shows the graph of time-performance of 87 update model operations only, since they represent the most time consuming steps using PP. It is important to note that during this test many flux analysis operations were simultaneously executed. In the beginning of every sample user interaction, each dummy user first uploaded the model and then performed random operations either on the model or uploaded a new model once again. We ran this test for 50 parallel users and observed that the upload model operations were much faster for the realistic Markov model user than when all 50 users simultaneously upload models. The average time to upload a model in the Markov scenario was 8.85 seconds compared to 14.56

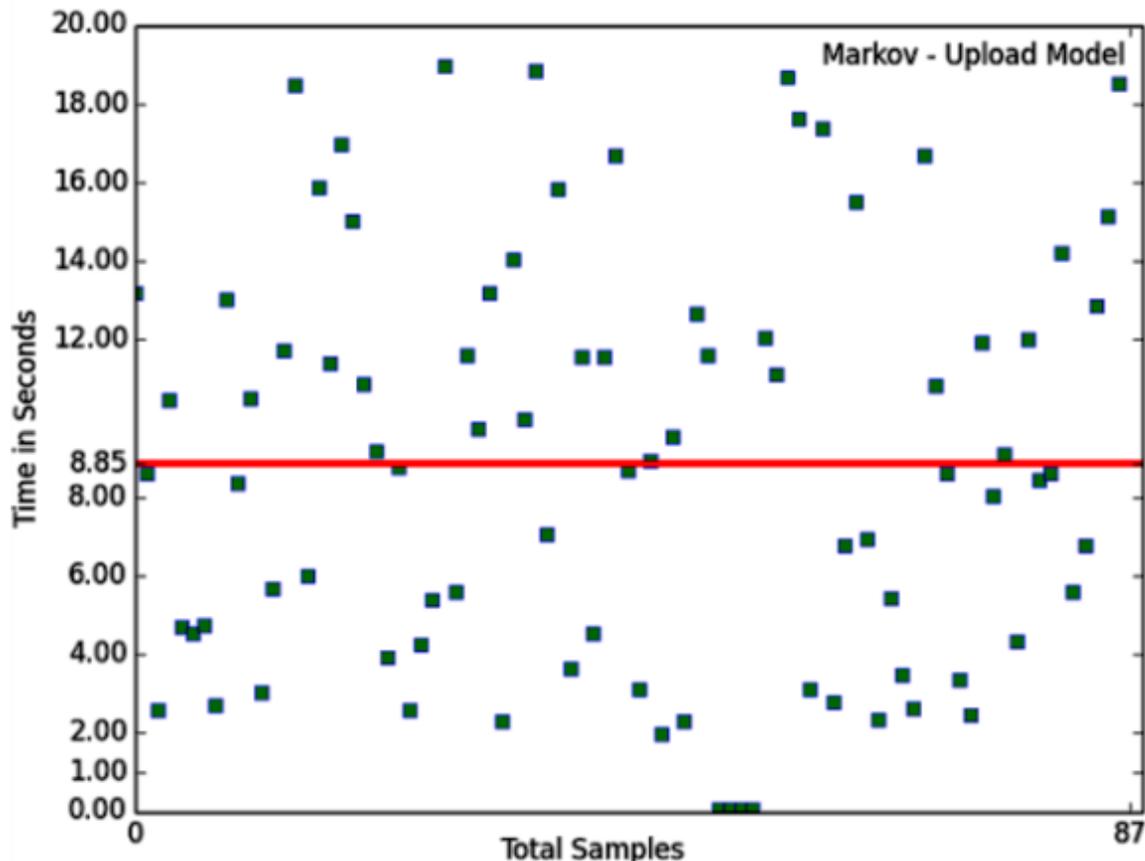


Figure 4.2: Graph showing the time taken (Y-Axis) by dummy users on PP to upload the model using Markov Testing Technique. Fifty parallel users were tested and 87 samples (X-Axis) were collected.

seconds when conducting the simultaneous model uploading.

4.0.3 Server layer

PP analysis is built on top of COBRA-Py [31], a tool for modeling and flux balance analysis running in Python 2.7. The server layer receives the request from the communication layer and performs FBA [8] on the model. It also annotates the model information with the reactions' and metabolites' color, size and edge thickness and layout information. The created SVG (the visualized layout representation) is then stored in the user repository and simultaneously sent back to the communication layer for updating the results on the client side.

CHAPTER 5

SYSTEM DEMONSTRATION

Glycine is one of the major components of spider silk protein. 3-phosphoglycerate, created during glycolysis, is the central metabolism precursor to glycine. Figure 5.1 (1) shows the Glycolysis/Gluconeogenesis subsystem when glucose is the sole carbon source. Figure 5.1 (2) shows the two reactions that directly interact with 3-phosphoglycerate: PGK, which is phosphoglycerate kinase, and PGM, phosphoglyceratemuase.

In the Glycine and Serine Metabolism subsystem, 3-phosphoglycerate is converted into serine through a series of reactions as shown in Figure 5.2 (1), and then is converted into glycine through the GHMT2r, glycine hydroxymethyltransferase, as shown in Figure 5.2 (2).

To explore how different carbon sources affect the production of 3-phosphoglycerate and ultimately glycine, the lower bound of the glucose exchange reaction is changed from -10 to 0 $\text{mmolgDW}^{-1}\text{hr}^{-1}$, and the lower bound of the exchange reaction for the other carbon source is changed from zero to a negative number. To study the effects of growth on glycerol, the glucose exchange reaction was changed from -10 to 0 $\text{mmolgDW}^{-1}\text{hr}^{-1}$, and the glycerol exchange reaction was changed from 0 to -10 $\text{mmolgDW}^{-1}\text{hr}^{-1}$, as shown in Figure 5.3 (1) and Figure 5.3 (2).

As a result of growth on glycerol, the fluxes through the reactions interacting with 3-phosphoglycerate are much lower, as shown in Figure 5.4 (1) and Figure 5.4 (2).

The fluxes through the reactions creating serine and glycine are much lower, as shown in Figure 5.2 (3) and Figure 5.2 (4).

The results show that glucose is a more suitable carbon source than glycerol to optimize the production of glycine.

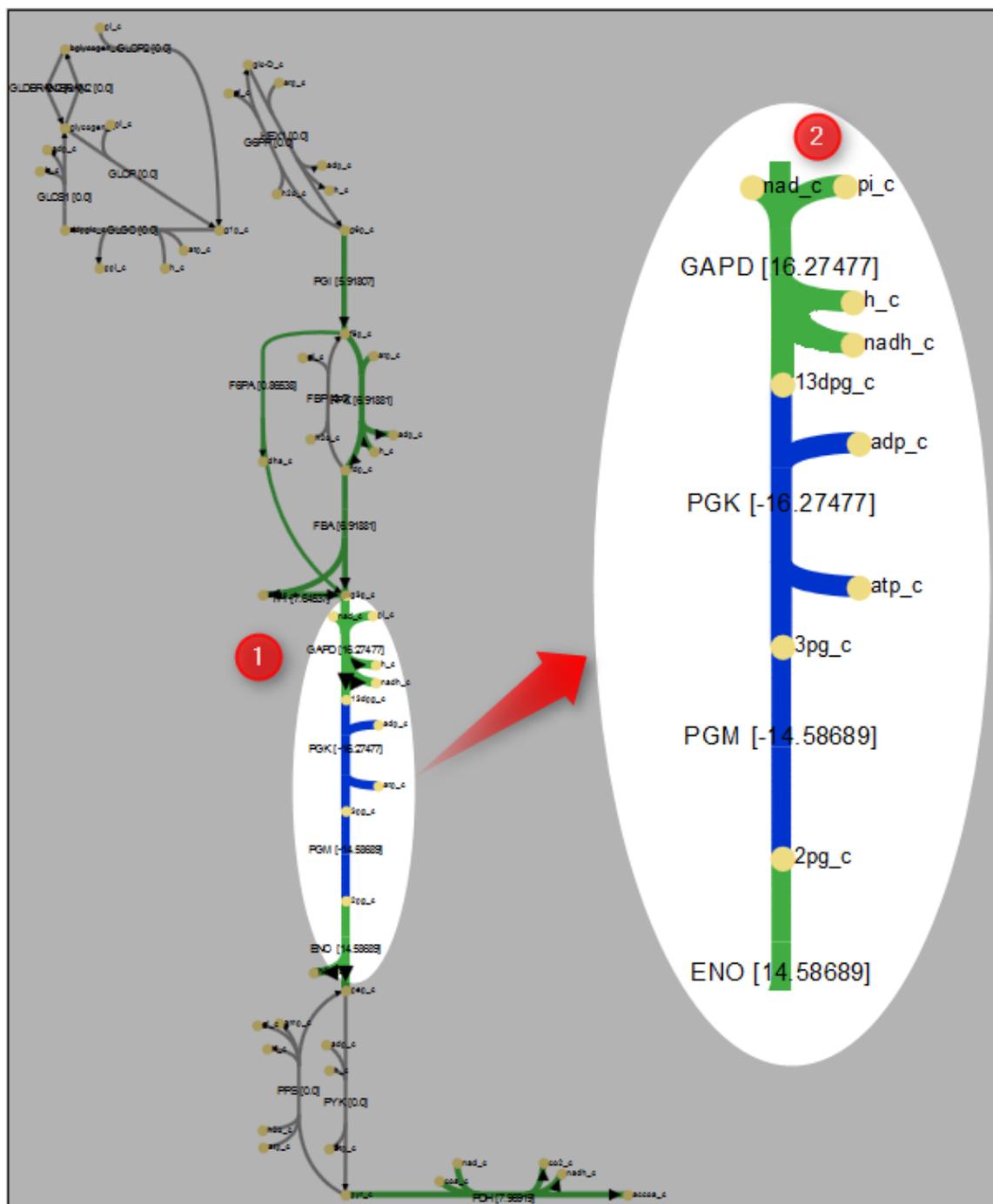


Figure 5.1: Image from PP displaying (1) the Glycolysis/Gluconeogenesis subsystem and (2) zoomed-in image with reactions that interact with 3-phosphoglycerate

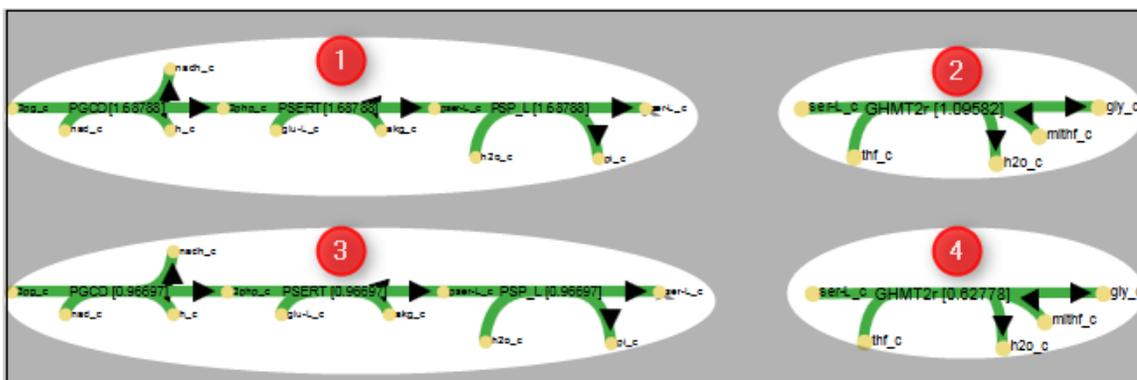


Figure 5.2: Image from PP showing Glycine and Serine metabolism using glucose as a carbon source (1) 3-pg is converted into serine (2) serine is converted into glycine through GHMT2r. As a result of growth on glycerol (3) reactions producing serine and (4) glycine have lower flux values than those using glucose as a carbon source

Figure 5.3 shows the software interface for updating the glucose exchange reaction. The 'Exchange Reactions' menu is open, and the 'Update Flux' option is selected. The 'Modification History' table shows the following data:

Reaction	Lower	Upper
EX_glc(e)	0	1000
EX_glyc(e)	-10	1000

Figure 5.3: Image from PP showing (1) On updating the glucose exchange reaction lower bound from -10 to 0 and (2) the effect of growth on glycerol by changing glucose exchange reaction

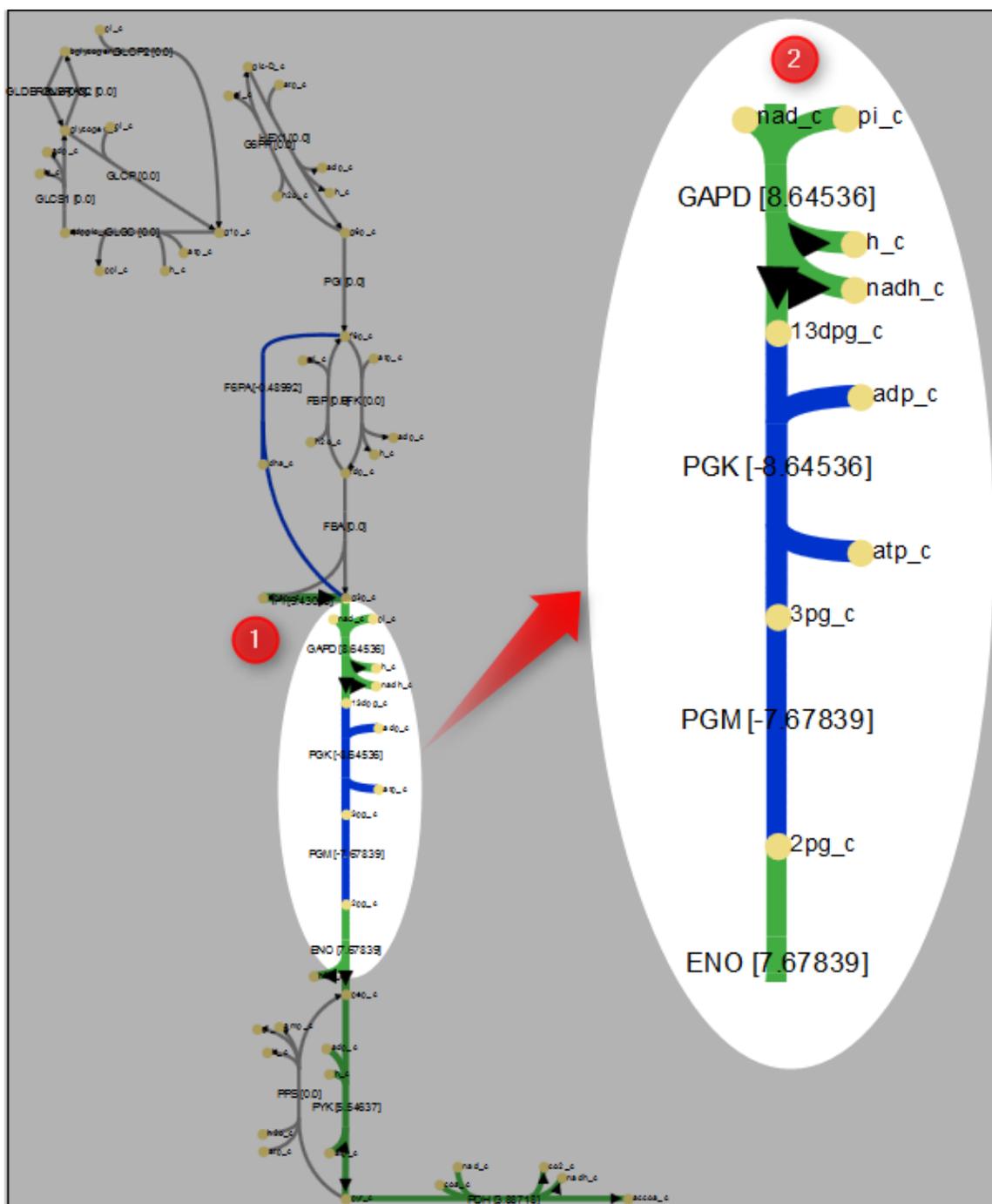


Figure 5.4: Image from PP displaying (1) with the growth of glycerol, the flux in the glycolysis is much lower and (2) zoomed-in reactions that interact with 3-phosphoglycerate

CHAPTER 6

CONCLUSIONS

Despite much future work left to be researched, this thesis demonstrates that Pathway Pioneer is capable of accelerating the research and development in genome scale modeling. It is a complete suite for analysis, engineer, de-novo synthesis, visualization and customization of large scale metabolic networks. Genome-Scale metabolic networks are intricate, which render them difficult to interpret and create. The extent of information to be represented is extensively diverse. A single network constitutes within itself functional modules, compartmentalized linear / cyclic / branched schema, list of metabolites and their respective formulae, number of reactions, subsystems, growth rate, current objective with its corresponding flux value and more. The complexity of these networks combined with their increasing numbers highlights the significance of a visualization tool which can enable representing such complexities in a comprehensible and decipherable manner. Currently, most existing tools are not all-encompassing. While most tools offer the basic functionalities of network topology visualization, none of them offer advanced features such as Model Engineering or Network customization and annotation. Pathway Pioneer a web-based network visualization tool is aimed to provide maximum features to the research community to ensure their progress is aided and helps improve the quality of hypergraphs which are used to study these intricate networks. In comparison to all the existing tools, Pathway Pioneer distinguishes itself by enabling a user to not only allow network editing and topology visualization but also provide a platform to model and simulate networks, be compatible with other tools and allow customization and annotation within the network. The most poignant feature of this tool is its ability to provide for Network Engineering. No other existing visualization tool supports this feature as of date. This web-based tool is developed to support the research and development community in the biological domain. The tool is applicable

over multiple models simultaneously and is visually centered. With a GUI interface and friendly intuitive usability, working with the tool becomes an easy, hassle-free undertaking. Furthermore, a user has the option for rapid explorations and even share the designs and results of this exploration with co-researchers. Additional features of the tool include editing layout, support for reaction translation and rotation and adjusting curvatures for predictions. The software suite is developed to create, customize, filter and analyze metabolic networks. The analysis of the networks is enabled by the presence of a repository with a versioning system feature. The ability to provide version numbers helps creating and saving multiple network diagrams, each of them providing a feasibility analysis of combination of different reactions. An embedded search engine within the tool makes it possible to look for specific graphs pertaining to the requirement. Mouse-over facility for each node, color spectrum visualization, smooth navigation and one-click operations facility help Pathway Pioneer be an stand-out tool for accelerating advancements in the field.

REFERENCES

- [1] M. A. Oberhardt, B. Ø. Palsson, and J. A. Papin, “Applications of genome-scale metabolic reconstructions.,” *Molecular Systems Biology*, vol. 5, Nov. 2009.
- [2] J. Nielsen, “Metabolic engineering,” *Applied Microbiology and Biotechnology*, vol. 55, no. 3, pp. 263–283, 2001.
- [3] J. Giovannetti, “Beyond borders, matters of evidence: Biotechnology industry report 2013.” <http://www.ey.com/GL/en/Industries/Life-Sciences/Beyond-borders-Matters-of-evidence-biotechnology-industry-report-2013>—Financial-performance-measures-that-matter, June 2013.
- [4] A. Pavlopoulos, A.-L. Wegener, and R. Schneider, “A survey of visualization tools for biological network analysis,” *BioData Mining*, vol. 1, pp. 1–11, Nov. 2008.
- [5] J. Schellenberger, J. Park, T. Conrad, and B. Palsson, “BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions,” *BMC Bioinformatics*, vol. 11, pp. 213+, Apr. 2010.
- [6] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, “The KEGG resource for deciphering the genome,” *Nucleic Acids Research*, vol. 32, pp. D277–D280, Jan. 2004.
- [7] J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Ø. Palsson, “Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.,” *Nature protocols*, vol. 6, pp. 1290–1307, Sept. 2011.

- [8] J. D. Orth, I. Thiele, and B. O. Palsson, “What is flux balance analysis?,” *Nature Biotechnology*, vol. 28, pp. 245–248, Mar. 2010.
- [9] R. Sharan and T. Ideker, “Modeling cellular machinery through biological network comparison,” *Nature Biotechnology*, vol. 24, pp. 427–433, Apr. 2006.
- [10] V. Hatzimanikatis, C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski, and L. J. Broadbelt, “Exploring the diversity of complex metabolic networks,” *Bioinformatics*, vol. 21, pp. 1603–1609, Apr. 2005.
- [11] S. Schuster, D. A. Fell, and T. Dandekar, “A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks,” *Nature Biotechnology*, vol. 18, pp. 326–332, Mar. 2000.
- [12] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, and A. L. Barabasi, “Global organization of metabolic fluxes in the bacterium *Escherichia coli*,” *Nature*, vol. 427, pp. 839–843, Feb. 2004.
- [13] J. M. M. Park, T. Y. Y. Kim, and S. Y. Y. Lee, “Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 14931–14936, Aug. 2010.
- [14] A. C. Villéger, S. R. Pettifer, and D. B. Kell, “Arcadia: a visualization tool for metabolic pathways,” *Bioinformatics*, vol. 26, pp. 1470–1471, June 2010.
- [15] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks,” *Genome Research*, vol. 13, pp. 2498–2504, Nov. 2003.
- [16] A. Kostromins and E. Stalidzans, “Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism,” *Bio Systems*, vol. 109, pp. 233–239, Aug. 2012.

- [17] F. Iragne, M. Nikolski, B. Mathieu, D. Auber, and D. Sherman, “ProViz: protein interaction visualization and exploration,” *Bioinformatics*, vol. 21, pp. 272–274, Jan. 2005.
- [18] E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk, “Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways,” *Bioinformatics*, vol. 18, pp. 996–1003, July 2002.
- [19] V. Batagelj and A. Mrvar, “Pajek Analysis and Visualization of Large Networks,” in *Graph Drawing* (P. Mutzel, M. Jünger, and S. Leipert, eds.), vol. 2265 of *Lecture Notes in Computer Science*, pp. 477–478, Springer Berlin Heidelberg, 2002.
- [20] A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, and H. Kitano, “CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks,” *Proceedings of the IEEE*, vol. 96, pp. 1254–1265, Aug. 2008.
- [21] H. Neuweger, M. Persicke, S. Albaum, T. Bekel, M. Dondrup, A. Huser, J. Winnebold, J. Schneider, J. Kalinowski, and A. Goesmann, “Visualizing post genomics data-sets on customized pathway maps by ProMeTra - aeration-dependent gene expression and metabolism of *Corynebacterium glutamicum* as an example,” *BMC Systems Biology*, vol. 3, no. 1, pp. 82+, 2009.
- [22] J. Boele, B. Olivier, and B. Teusink, “FAME, the Flux Analysis and Modeling Environment,” *BMC Systems Biology*, vol. 6, pp. 8+, Jan. 2012.
- [23] S. I. Berger, R. Iyengar, and A. Ma’ayan, “AVIS: AJAX viewer of interactive signaling networks,” *Bioinformatics*, vol. 23, pp. 2803–2805, Oct. 2007.
- [24] C. H. Schilling and B. O. Palsson, “Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis,” *Journal of Theoretical Biology*, vol. 203, pp. 249–283, Apr. 2000.
- [25] A. Finney and M. Hucka, “Systems biology markup language: Level 2 and beyond,” *Biochemical Society transactions*, vol. 31, pp. 1472–1473, Dec. 2003.

- [26] C. M. Lloyd, M. D. Halstead, and P. F. Nielsen, “CellML: its future, present and past.,” *Progress in Biophysics and Molecular Biology*, vol. 85, pp. 433–450, July 2004.
- [27] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, S. G. Grant, C. Sander, P. Bork, W. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, I. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, and R. Apweiler, “The HUPO PSI’s molecular interaction format—a community standard for the representation of protein interaction data.,” *Nature Biotechnology*, vol. 22, pp. 177–183, Feb. 2004.
- [28] P. Murray-Rust and H. S. Rzepa, “Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles,” *J. Chem. Inf. Comput. Sci.*, vol. 39, pp. 928–942, Sept. 1999.
- [29] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson, “A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.,” *Molecular Systems Biology*, vol. 3, June 2007.
- [30] J. Ferraiolo, F. Jun, and D. Jackson, *Scalable vector graphics (SVG) 1.0 specification*. iuniverse, 2000.
- [31] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. O. Palsson, and M. J. Hergard, “Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox,” *Nature Protocols*, vol. 2, pp. 727–738, Mar. 2007.