PathCase-MAW: An Online Metabolic Network Analysis Workbench

A. Ercument Cicek[1], Xinjian Qi[1], Ali Cakmak[2], Stephen R. Johnson[1], Xu Han[1], Sami Alshalwi[1], Gultekin Ozsoyoglu[1*]

[1]Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106

[2]Department of Computer Science, Istanbul Sehir University, Altunizade, Istanbul, TURKEY 34662

ercument.cicek@case.edu, xinjian.qi@case.edu, alicakmak@sehir.edu.tr, srj15@case.edu, xu.han@case.edu, sma85@case.edu, tekin@case.edu

ABSTRACT

Metabolic networks have become one of the centers of attention in life sciences research with the advancements in the metabolomics field. A vast array of studies analyzes metabolites and their interrelations to seek explanations for various biological questions, and numerous genome-scale metabolic networks have been assembled to serve for this purpose. The increasing focus on this topic comes with the need for software systems that store, query, browse, analyze, and visualize metabolic networks. PathCase Metabolomics Analysis Workbench (PathCase^{MAW}) is built, released, and running on a manually created generic mammalian metabolic network. The PathCase^{MAW} system provides a database-enabled framework and web-based computational tools for browsing, querying, analyzing, and visualizing stored metabolic networks. PathCase^{MAW} editor, with its user-friendly interface, can be used to create a new metabolic network and/or update an existing metabolic network. The network can also be created from an existing genome-scale reconstructed network using the PathCase^{MAW} SBML parser. The metabolic network can be accessed through a web interface or an iPad application. For metabolomics analysis, Steady-State Metabolic Network Dynamics Analysis (SMDA) algorithm is implemented and integrated with the system. SMDA tool is accessible through both the web-based interface and the iPad application for metabolomics analysis based on a metabolic profile. PathCase^{MAW} is a comprehensive system with various data input and data access sub-systems. It is easy to work with by design, and is a promising tool for metabolomics research and for educational purposes.

Categories and Subject Descriptors

H.2.8 [**Database Applications**] *Scientific databases.* H.3.5 [**Online Information Services**] *Web-based services.* H.2.4 [**Systems**] *Rule-based databases.* H.5.2 [**User Interfaces**] *Graphical user interfaces (GUI).*

* Corresponding Author

BCB '13, September 22 - 25, 2013, Washington, DC, USA

General Terms

Algorithms, Design, Experimentation.

Keywords

Metabolic network database, SMDA, SBML.

1. INTRODUCTION

Metabolomics is a relatively new "omics" platform in life sciences research. The advancements in analytical methodology and high throughput rates have led to the collection of large metabolic datasets. Metabolic profiles and genome scale metabolic networks [1] are used in various contexts, such as (1) predicting flux distribution for the metabolic activity over the network (Metabolic Control Analysis [2], Flux balance Analysis [3], and Constraint Based Methods [4]), and (2) drug discovery and disease research [5-7]. The increase in the number and importance of metabolic networks has come with the need for carefully designed databases to store/organize metabolic networks, and efficient online tools to browse/analyze metabolic data.

The goal of PathCase^{MAW} (Metabolic Analysis Workbench) is to provide a metabolic network database and a web- or tablet-based system that enables users to interact with the underlying metabolic network. PathCase^{MAW} provides the following functionalities:

(1) A metabolic network database that captures the metabolic network with a compartment hierarchy and metabolic regulation relationships.

(2) A website that (i) enables users to browse pathways, reactions, metabolites/metabolite pools, and compartments stored in the database, (ii) provides several built-in queries and interactive visualization, and (iii) has the integrated SMDA (Steady state Metabolic Dynamics Analysis) Tool. SMDA tool takes a set of metabolite measurements and a metabolic sub-network of the metabolic network stored in the PathCase^{MAW} database as input. Then, it produces all possible steady-state flow scenarios (called flow graphs) for the selected sub-network as output (that are consistent with the observed metabolite measurements and the underlying biochemistry). Available at http://nashua.case.edu/PathwaysMAW/web/

(3) An iPad application that has all capabilities of the web-based PathCase^{MAW} system with the exception of browsing/querying. Available at Apple AppStore.

(4) An offline metabolic network editor with visualization capabilities that enables users to create their own network in a user-friendly way.
(5) An SBML Parser to parse and store genome-scale reconstructed metabolic networks (e.g., Recon 1 of humans [8]) into the PathCase^{MAW} database.

Currently, the PathCase^{MAW} system works on a manually created (and generic) mammalian metabolic network, which is obtained from the metabolic atlas by Selway et. al. [9]. We also have three genome-scale reconstructed networks hosted and available on the sister PathCase^{RCMN} website [10]. Source codes of the web interface, PathCase^{MAW} editor, SBML Parser, as well as the database schema are available upon request for academic users to create their own networks and to host/access them. User-created networks can also be hosted on the PathCase^{RCMN} website upon request.

2. Implementation

In this section, we summarize the design and implementation details of the PathCase^{MAW} system.

2.1 Architecture

PathCase^{MAW} has a two-tiered client-server software architecture, with a thick client. On the client side, there are four applications.

PathCase^{MAW} Web interface: Via any web browser, users can access, visualize, query, and analyze the data stored in the PathCase^{MAW} database. Application is written in C# language using .NET environment. IIS 7.0 is used as the web server.

PathCase^{MAW} iPad application: Users can visualize pathways stored in the database, and use the SMDA Tool. iPad application can be downloaded for free from the Apple App Store, and is currently functioning in iPad2 and newer models. The application utilizes a buffer for pathway data in its flash memory for fast access, and to reduce data transfer from the servers. The iPad application data is updated whenever the PathCase^{MAW} database is updated. The communication between the iPad application and the server PathCase^{MAW} database is handled through the web services (XML and GML documents are used as the data exchange format). Users are able to use iPad's multi-touch screen to visualize pathways and to run the SMDA tool in a mobile tablet environment.

PathCase^{MAW} Editor: This application is used by data owners to create/update metabolic networks in the PathCase^{MAW} database. This is an offline tool and is implemented in JAVA in order to provide platform independence.

PathCase^{MAW} SBML parser: It is a console application that takes as input an SBML file for a metabolic network, and parses it into the PathCase^{MAW} database as a separate metabolic network. It is written in C# using .NET framework.

On the server side, we have the following components.

PathCase^{MAW} database: It is stored in Microsoft SqlServer 2008 and is accessed through the Data Access Library.

The Data Access Library: It contains (i) database wrappers that abstract the database schema, (ii) built-in queries, and (iii) an XML Generator that generates a metabolic network/sub-network/pathway visualization file, which is then passed to either (a) JAVA applet of the PathCase^{MAW} client-side browser, (b) PathCase^{MAW} iPad application, or (c) the PathCase^{MAW} editor.

The Web Content UI Manager: It handles synchronous or asynchronous requests sent from the client browser, and responds in HTML format.

The SMDA Tool: It implements the SMDA algorithm [11,12].

The Web Services: This component is used for data access during query processing and for network visualization.

Figure 1 illustrates the communication among the main components of the PathCase^{MAW} system, and Figure 2 shows the architecture of the system. Details regarding the system are given in individual sub-system sections.

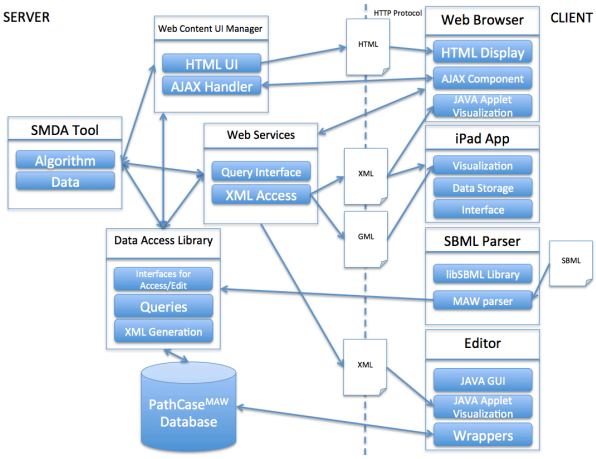

Figure 1. Communication among PathCase^{MAW} components/ sub-components.

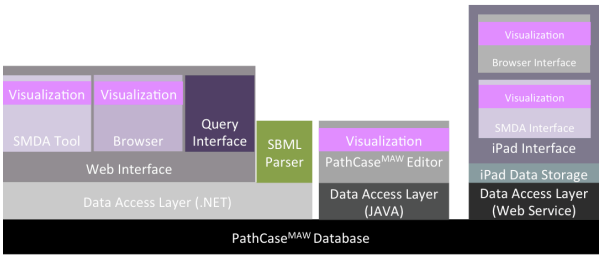

Figure 2. PathCase^{MAW} Software Architecture.

3. Data Access Interfaces

In this section, we present the interfaces provided by PathCase^{MAW} system to browse, query, visualize, and analyze the data. There are two main ways to access the data: (1) a Web browser (e.g., Internet Explorer), and (2) iPad App of PathCase^{MAW}. Next, we explain the capabilities provided in each type of interface.

3.1 Web Browser

PathCase^{MAW} is an online system that can be accessed through an Internet browser. The backend is implemented in C#, using .NET framework. The website is hosted on a machine running Windows Server 2008. We use Internet Information Services (IIS) 7.0 as our web server. It has been tested on four major web browsers with the recent versions of Microsoft Internet Explorer, Mozilla Firefox, Google Chrome, and Apple Safari. The web interface includes four applications, namely, (1) Browser, (2) Visualization Interface, (3) Querying Interface, and (4) SMDA Tool.

3.1.1 *Browser interface*

This interface lets the user to view the data stored in the database from various different starting points. For instance, the user can start from a metabolite to list related reactions, to find the pathways this metabolite plays a role in, or one can start from a pathway to find related reactions and the compartments this reaction takes place.

The browser interface has two parts: The navigation bar on the left hand side and the main frame on the right hand side that shows details of the selected browser item. The left hand side is called the *Browser* page, and the right hand page is called the *Details* page.

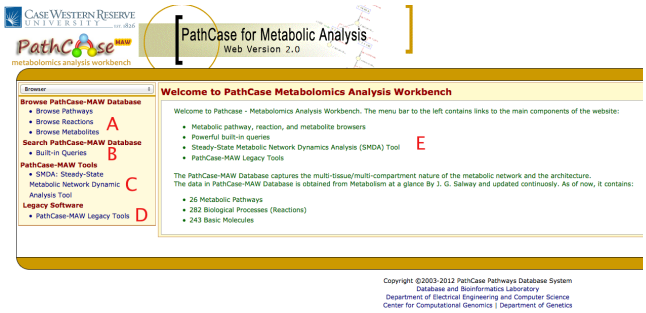

Figure 3. PathCase[MAW] Homepage. A. Browser links to pathways/reactions/metabolites stored in the database. B. Link to the built-in queries. C. Link to the SMDA Tool. D. Link to Legacy Tools. E. Details page reporting summary statistics about the database.

Figure 3 shows the homepage of PathCase[MAW]. The main Browser page has 3 subsections. Part A redirects the user to browse the data starting from (1) pathways, (2) reactions and (3) metabolites stored in the database. The browser page has the following flexible hierarchy that enables user to effectively browse related entities of the network.

> (1) Pathways -> reactions in the selected pathway -> compartments the selected reaction takes place -> metabolites associated with the selected reaction in the selected compartment.
>
> (2) Reactions -> compartments the selected reaction takes place and pathways this reaction is associated with-> (selecting a compartment) metabolites associated with the selected reaction in the selected compartment.
>
> (3) Metabolites -> pathways the selected metabolite is associated with -> compartments the selected pathway takes place in -> reactions of the selected pathway that exist in selected compartment.

Part B opens up the query interface, and Part C links to the implemented tools (currently only SMDA Tool). Part D links the user to our legacy tools, namely, the OMA Tool [13] and MQL [14], which are no longer maintained.

Part E shows the details page where details about the selected browser item are loaded as separate collapsible panels. In the homepage, statistics about the database is shown along with a quick introduction. Users are able to click to any entity on the browser page (as described above), to open up the details page. The following information is displayed per entity:

- **Pathway**: Panels for; (i) loading the applet that visualizes the pathway, (ii) listing reactions taking place in the pathway, (iii) listing connected pathways in the complete network, and finally (iv) listing all available pathway related built-in queries.
- **Reaction**: Panels for; (i) listing involved metabolites, (ii) listing related pathways, and finally (iii) listing all available reaction related built-in queries.
- **Metabolite**: Panels for; (i) listing related pathways, (ii) listing synonyms stored in the database for this metabolite (e.g., Fructose 1,6 bisphosphate; F16P; F1,6P), and finally (iii) listing all available metabolite related built-in queries.
- **Compartment**: A panel to list all reactions that take place in this compartment, and a panel to list all pathways that take place in the selected compartment.

Figure 4 shows browser and detail pages for *Urea Cycle* pathway. Part A shows the hierarchy underneath the pathway. There are seven reactions in the pathway; selecting *ornithine transcarbamolyse* shows that it only takes place in mitochondrion in cytosol of a liver cell. Underneath mitochondrion, we find all four metabolite pools that are associated with *ornithine transcarbamolyse*. Clicking *Urea Cycle* on the browser page, loads up the collapsible panels to the details page in the collapsed form.

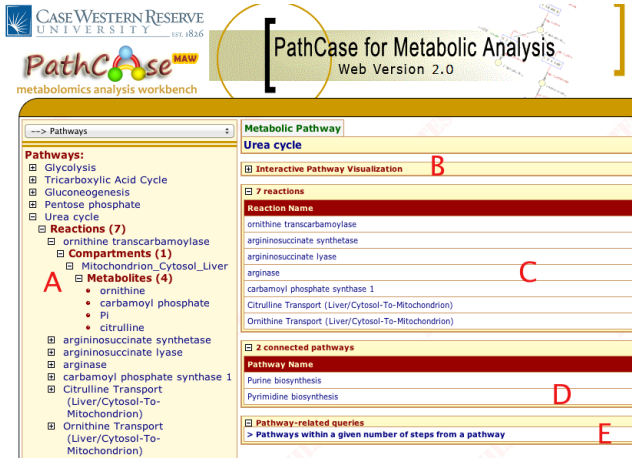

Figure 4. PathCase[MAW] Browser and Detail pages for the pathway Urea Cycle.

Panel B downloads the applet once clicked (only at the first usage) and loads up the XML data to visualize the pathway. A screen shot for the visualization of *Urea Cycle* is shown in Figure 5 and details about the visualization interface is given in *Web/Editor Visualization* Section. Panel C shows the reactions associated with *Urea Cycle* and clicking each one of them would take the user to the details page regarding that reaction. Panel D shows there are two connected pathways to *Urea Cycle* in the database through shared metabolite pool(s): *Purine Biosynthesis* and *Pyrimidine Biosynthesis*. Finally, Panel E takes user to the page to run the query for finding pathways with in a number of steps away from *Urea Cycle*.

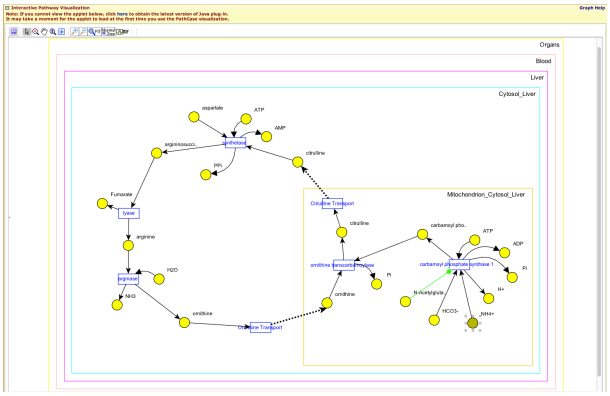

Figure 5. Visualization of the Urea Cycle pathway.

The details page accesses the data using AJAX calls. Therefore, once an item is clicked, the collapsible panels are loaded. Collapsible panels make calls to the server after they are displayed, so the data is accessed asynchronously. If a collapsible panel is clicked, while the data is still being fetched, user is informed that it is being loaded. This brings a huge advantage in terms of user experience, when pathways with a large number of reactions and metabolites are accessed. In the synchronous case, users would have to wait for all panels to load before they see the page. Users may also interact with the loaded panels while some panels are still loading data in the background.

3.1.2 Web-based SMDA Tool

SMDA (Steady-State Metabolic Network Dynamics Analysis) [11, 12] uses an algorithmic approach to analyze metabolomics data in terms of the dynamic behavior of the metabolic network. Given a set of metabolite measurements, it identifies the metabolic activity over the network that lead to the observed concentrations of given metabolite pools. The input data consists of one or more pathways; zero or more reactions, which constitute the selected metabolic sub-network to be analyzed. Also, measured concentrations of specific metabolites within the selected metabolic sub-network are provided as input. The observed metabolites are compared with the normal/abnormal values of those metabolites reported in HMDB [15] and assigned labels: Unavailable, Available, Accumulated and Severely Accumulated. Based on the biochemical principles and the input metabolite levels captured by the algorithm, it is determined if a reaction is active or not. The algorithm results in possibly several scenarios that label each reaction as active or not. It has been shown as an effective method to analyze sub-network activation scenarios for Cystic Fibrosis research [11].

SMDA Tool implements the SMDA Algorithm on server side and can be accessed via the web interface. The tool is implemented using C# language in .NET environment. Next we introduce the capabilities of the web interface for the SMDA tool over an example, shown in Figure 6.

First, there are two options to input data. (1) Users can pick one of eight provided sample input data shown in Part A. Clicking the button loads up the sub-network selected and corresponding observations. They also can upload an XML document as input. (2) Users can input data manually. For the second case, user first needs to select the metabolic sub-network of interest. In Part B, user picks a pathway and a compartment (e.g., organ) that the pathway is taking place in (note that a pathway can exist in more than one compartment). In this example, *TCA Cycle* in liver cell is selected. Next, in Part C, users select stand-alone reactions (which are not part of a pathway). They can select transport reactions or regular reactions. Transport reactions are unique so they do not require input of compartment, whereas regular reactions require the compartment to be specified. Users are also able to simplify pathways into a single reaction as done in many works in the literature. Then in Part D, *Glycolysis* pathway is added as a single reaction that has *D-Glucose* as input and *Pyruvate* as output in again liver cell. Part E lets the user to pick metabolite pools and specify the concentrations. In this example the only observation is *D-Glucose* in the cytosol of liver cell. Finally, Part F concludes data input part by inputting observed enzymes for reactions. In Figure 6, *Citrate Synthase* is observed as Unavailable. This observation deactivates the reaction.

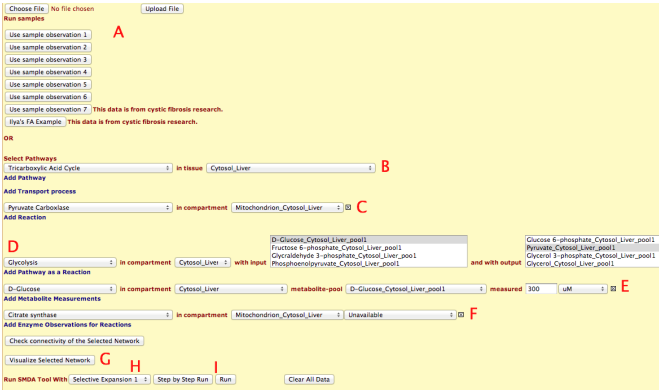

Figure 6. Web interface for SMDA Tool.

After inputting data, users can check if the selected sub-network is connected or not and can visualize their network, using the buttons on Part G. Finally, user can run the algorithm till it completes, or he/she can see how the algorithm proceeds (e.g., assigns labels to next reaction eliminate options) step by step by selecting one of the buttons shown on Part I. Part H lets the user to pick one of the iteration algorithms as explained in [11]. Note that SMDA tool is also equipped with AJAX calls. For instance, selecting a pathway opens up the compartment selection combo-box, which loads only the compartments this pathway exists in.

The results are passed back to the client in XML format and directly visualized. SMDA Visualization uses the same system used to visualize pathways. The only differences are (1) there are possibly more than one resulting graphs, each representing a possible scenario, and (2) thin edges represent inactive reactions and thick edges represent active reactions.

3.1.3 Web Querying Interface

PathCase^{MAW} currently has five built-in queries to enable users find related pathways, reactions, metabolites with respect another entity. The currently available queries are:

1. *Pathways within a given number of steps from a pathway*: This query searches for pathways that are *n* reactions away from a given pathway.
2. *Reactions within a given number of steps from a reaction in a pathway*: This query searches for reactions that are *n* reactions away from a reaction in a pathway.
3. *Reactions within a given number of steps from a reaction in the metabolic network:* This query searches

for reactions that are *n* reactions away from a reaction in the network.

4. Reactions involving a metabolite in a pathway: This query searches for reactions that involve a given metabolite in a pathway.
5. Metabolites within a given number of steps from another metabolite in the metabolic network: This query searches for metabolites that are *n* reactions away from a metabolite in the network.

The system is equipped with AJAX calls that prunes out irrelevant selections without reloading the page. For example, for the second query, the user is asked to pick a pathway, then a reaction. Once the user picks the pathway, the next combo box to pick the reaction in that pathway is loaded with only the reactions in that pathway, and the rest is discarded. This prevents the user from selecting illegal items (e.g., choosing a reaction not existing in the already selected pathway) and contributes to a positive user experience.

3.1.4 Web/Editor Visualization Interface

PathCase^{MAW} visualization interface is responsible for displaying interactive and user-friendly depictions of metabolic sub-networks (e.g., pathways or a collection of pathways and reactions). It is used in three sub-systems of PathCase^{MAW}, namely, (1) Within a collapsible panel for pathway visualization in the webpage, (2) As a tab for pathway visualization in the editor, and (3) In the SMDA Tool, as a separate window for visualizing the selected sub-network (any connected set of pathways and reactions) and visualizing the resulting graphs. In the first two cases, the applet displays only the selected pathway. It uses a JAVA applet that runs on the client machine, and with no server-side intervention or communication other than an XML document exchange, through the web service. This allows PathCase^{MAW} to scale for a large number of users. It is run within the Internet browser. The visualization interface downloads the applet itself only at the time of the first use, and comes embedded into the corresponding pathway pages of the SMDA tool interface. Therefore, (i) the visualization does not require a separate installation effort (a manageability convenience for users), and (ii) provides a platform-independent access regardless of the operating system or browser that is run on the user machine. For the editor, it comes embedded within the jar file, and displays all selected pathways and reactions.

The interactive graph displays metabolites (circles), reactions (small rectangles), compartments (rectangles that contain metabolites and reactions and sub-compartments) and the relations between them in the model (as lines). The directions of the lines set producer/consumer relation for a reaction (two-headed lines indicate reversible reactions). Red/green line indicates inhibitor/activator relationship. Interface captures the compartment hierarchy. Metabolite pools and reactions are shown in the corresponding compartments.

Users are able to interact with the visualization; they can (1) move all items displayed, (2) zoom in/out in different ways (e.g., fit-to-screen, magnifier), (3) save the modified layout so that it appears the same on the next load for that pathway/sub-network (password protected for data-owner).

Metabolites that participate in many reactions (e.g., NAD, ATP, O_2 etc.) have high connectivity and tend to result in unclear visualizations. Such metabolites are called *common metabolites*. Unlike all metabolite pools, which are displayed once, these metabolites are displayed per participated reaction to have better results.

Applet provides tooltips for the user, that is, when the cursor is moved over an item the name (e.g., reaction or metabolite name) is displayed in a little box that pops up and disappears when cursor is moved out. This lets the applet to keep truncated forms of the long names and keep the image tidy, while letting the user see the names of interest in advance. Another useful functionality is the map, which is crucial for large pathways/sub-networks. Picking this option opens up a small and zoomed-out snapshot of the complete visualization. This way, user is able to navigate through the visualization without a need to zoom-out and then zoom-in again.

3.1.5 iPad Interfaces

iPad has been a revolutionizing mobile equipment with the multi-touch based interface that have sold 9.25 million only in the fiscal 2011 third quarter [16]. Other tablet companies have followed, and it has been a widespread technology. More widespread the tablets are, more attention they attract for educational purposes. For instance, Apple has launched an effort to revolutionize textbooks [17], and countries (such as Turkey) distribute tablets to students [18]. Our goal in building an iPad app for PathCase^{MAW} is to provide educational tools for life scientist in biochemistry teaching and for research. Users are able to have nice visual depictions of pathways and run SMDA Tool using the iPad interface. The system is written in Objective-C language using Cocoa framework. While the system works on iPad 1, currently, the system is optimized for efficient use on iPad 2 (which is called iPad from now on). The system has three main interfaces: Browsing, Visualization, and SMDA Tool interfaces. Next we give details about each interface.

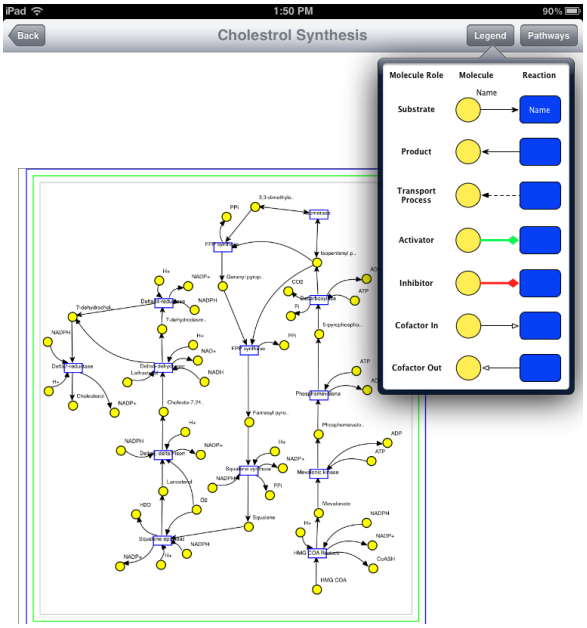

Figure 7. Visualization of Cholesterol Synthesis Pathway on iPad is shown along with the legend.

3.1.5.1 iPad Browser Interface

The browser of the application consists of only pathway browsing that acts a link to the visualization of the pathway. At the current stage, reaction, metabolite and compartment browsing is not supported. Pathway browser can be accessed through the home page, where pathways are listed on the right frame or, from within

a visualization of any pathway by tapping the "Pathway" button in the tool bar as seen on the top right of Figure 7.

3.1.5.2 iPad Visualization Interface

The visualization interface employed for iPad looks very similar to the web interface's visualization; however, the implementation is quite different. The application employs Core Animation Framework to render objects. The data is obtained in the form of an XML + GML (Graph Markup Language) document via web services. Due to the hierarchical and interconnected nature of the data model, it is not practical to dynamically invoke the web services at the time their data is needed. Instead, an internal representation is updated in bulk at the user's request. When the user taps on the "Update Local Data" button in the SMDA input view, a series of asynchronously executed Objective-C blocks are invoked. This design choice cuts down on the data quote usage of the device, and provides fast response time for the user. Figure 7 shows the visualization for *Cholesterol Synthesis* pathway along with the legend.

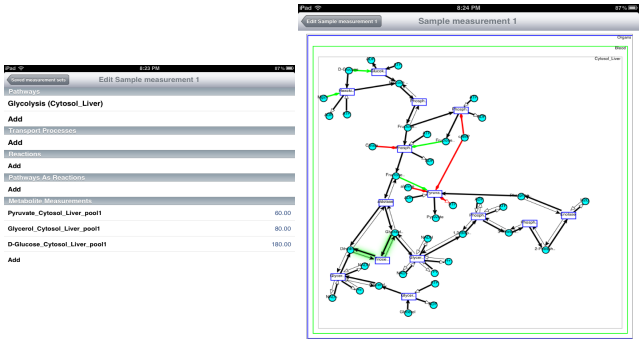

Figure 8. Running SMDA on iPad for Glycolysis Pathway.

3.1.5.3 iPad SMDA Tool Interface

SMDA Tool Interface uses the same server side component that implements the SMDA Algorithm. It provides a similar interface as the web interface, which lets user to pick the sub-network, and the observations for metabolites. Results are visualized in the same manner (thick edges represent active reactions and thin edges represent inactive reactions). As stated before, there are possibly more than one resulting visualizations. The iPad Visualization interface implements a nice feature to mark the differences between the current image and the previous one.

Figure 8 displays a run of the SMDA Tool on *Glycolysis* pathway. On the left the input is shown and on the right one of the four results is displayed. The glowing green edge shows the difference between the previous possible scenario and this one. Users can switch between results by a one finger swipe to right and left.

4. Data Input Tools

In this section, we describe two tools used in order to manage the data stored in PathCase^{MAW} database.

4.1 PathCase^{MAW} Editor

Entering metabolomics data manually into a relational database is a complicated and error-prone process. Consider a user entering a new reaction (with a new substrate and a new product). First, she needs to enter two new metabolites into the *Metabolite* table and then two new entries into the *MetabolitePool* table keeping the ids of the metabolites in mind. Next, she needs to add the reaction to the *Reaction* table. Finally she needs to add two entries into *ReactionMetabolite* table coupling the ids of the metabolite pools and the reaction along with the role ids of the metabolite pools. Thus, even entering a single reaction is a quite complicated and an error-prone task. In order to enable users to manage the data easily, we have developed an easy-to-use editor that houses tools for (a) populating, (b) editing, and (c) visualizing the data stored in PathCase^{MAW} database. The editor is a JAVA application and therefore it is platform independent. It has a user-friendly graphical user interface.

PathCase^{MAW} Editor has the following functionalities/ advantages:

1. The system hides from users the internal object IDs. All relationships are shown via object names in the GUI, and the user cannot modify object's ID directly; this effectively eliminates much of the data inconsistency possibilities.
2. PathCase^{MAW} Editor is designed to prevent entering illegal data by providing a drop-down box for a field with fixed set of allowed values, rather than providing a free-text-box. In cases where the first measure fail, it explicitly checks the data against certain constraints.
3. Users can insert new data in two ways: (i) They can insert a new entity (i.e., from scratch), and (ii) For some object types (i.e. compartment), they can copy an existing object, and insert it into the database as a new record.
4. To input data efficiently, PathCase^{MAW} Editor provides an auto-complete feature for text fields. Given the complexity of the biological terminology, this feature saves unnecessary keystrokes and prevents illegal data insertion to the database.
5. Changes made on the data are not reflected to the database right away. Users need to approve their changes by "saving" their work to avoid updating the database by mistake.
6. The integrated visualization applet enables user to see results of the update on the fly, which acts as an efficient visual debugger.

Figure 9 shows a screen shot of the editor mainframe. Part A on the left hand side shows the navigation hierarchy for each entity (e.g., Pathways, Reactions, Compartments, etc.). In this example, the pathway *Glycolysis* is chosen, and the reactions belonging to the pathway are shown. Similarly, one can go one more level down to see the metabolite pools associated with these reactions. The same logic applies to other nodes shown on the tree. For instance, toggling "Metabolites" group would list all metabolites, and going down one more step would show the pathways, in which this metabolite participates, etc. Part B on the top right hand-side shows the reactions associated with *Glycolysis* pathway in a grid form. In Figure 9, reaction *Hexokinase* is clicked, which opens up a dropdown box listing all reactions available in the database. User can replace *Hexokinase* using this dropdown box. This feature prevents users from entering illegal entries to the database. Part C shows the input/output metabolites for *Glycolysis*. Right below Part C, there are two tabs. Selecting "Visualization" tab shows the visualization of *Glycolysis* pathway for Liver and Adipose tissues. The visualization applet is shared by the web interface and details about the visualization interface are discussed in *Web/Editor Visualization Interface* Section.

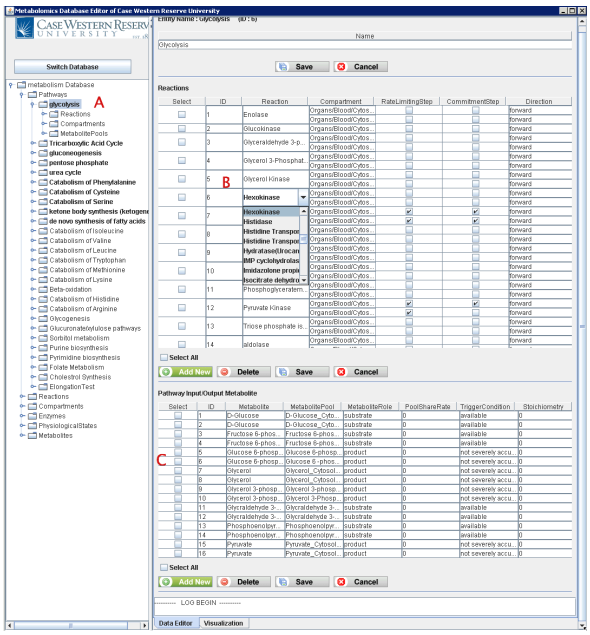

Figure 9. PathCase^{MAW} Editor showing data for Glycolysis Pathway

Although we have presented a few examples of updating pathways, reactions or metabolites; the editor has a large number of different convenient ways for users to interact with the database, which are not shown here.

5. PathCase^{MAW} SBML Parser

SBML [19] is one of the standards to create and distribute systems biology models. Although it has been used to describe many kinds of systems, in the context of PathCase^{MAW}, we are dealing with metabolic networks that have been described in SBML format. PathCase^{MAW} SBML Parser is a descendent of PathCase^{SB} SBML Parser [20], a subsystem of PathCase^{SB} project. PathCase^{SB} [21-23] is a system that stores kinetic models of pathways provided by BioModels Database [24], and offers extensive functionalities over the model data by linking multiple resources. PathCase^{SB} SBML parser is used to parse SMBL, and to populate the PathCase^{SB} database [22]. Schemas and functionalities of PathCase^{SB} and PathCase^{MAW} databases are quite different. PathCase^{SB} database focuses on integrating multiple resources, whereas PathCase^{MAW} database is intended to effectively store metabolic networks. PathCase^{SB} SBML parser's data access layer have been completely changed to suit PathCase^{MAW} database schema, and the XML processing component is modified to handle differences between description of kinetic models for pathways and description of metabolic networks. The challenges and problems with SBML format misuse have been discussed in *Results and Discussion* section in detail.

The parser is written in C# language using .NET framework. It relies on the libSBML library (currently Version 1 of Level 3) [25], which is a public library that provides a framework for parser to access elements of the SBML model. As the model is parsed, the database is accessed using the Data Access Library that contains the object-oriented wrapper classes and their functions. Wrapper classes are shared by the Browser code, so any change made on the database schema can be handled easily by changing the related wrapper classes. These wrapper classes are different than Editor's wrappers as they are implemented in JAVA language. System parses the file stored in a given directory and updates the database accordingly.

6. Results and Discussion

All presented sub-systems of PathCase^{MAW} have been implemented and released to the research community. Currently, the system runs on a manually created generic mammalian database that consists of 26 pathways, 282 reactions, and 243 metabolites. We have also parsed and released three genome-scale reconstructed metabolic networks (2 for Mus Musculus and 1 for T. Cruzi) in the sister PathCase^{RCMN} site. Next, we provide a comparison of the existing systems with PathCase^{MAW} and then the challenges/shortcomings of the current system.

6.1 Comparison of PathCase^{MAW} with Existing Systems

KEGG [26-29] has been a major source of metabolic pathways, which provides an application programming interface and data download options. Unlike PathCase^{MAW}, KEGG (i) provides limited visualization and browsing capabilities, (ii) does not capture compartment information for metabolites, (iii) provides limited functionality over the data, and (iv) ignores metabolic regulation such as covalent activation/inhibition or metabolite ratios (e.g., high NAD/NADH ratio activates Alpha-ketogutarate Dehydrogenase).

PathCase^{SB} and PathCase^{KEGG} [30] have been released to provide additional functionalities such as querying and interactively visualizing the data. PathCase^{KEGG} hosts KEGG data in its own schema and provides similar browsing and visualizing capabilities as PathCase^{MAW} system. PathCase^{SB} stores kinetic models of pathways provided by Biomodels Database [18]. It integrates KEGG data with the stored models, enables users to simulate and compose models. However, neither of these systems have metabolomics analysis goals or tools. They only work on data provided by well-known third parties (KEGG and Biomodels Database). Conversely, the goal of PathCase^{MAW} is to provide a system, on which, users can base their own data. Finally, among the three above-mentioned systems, only PathCase^{KEGG} has a mobile interface (iPad application).

BioCyc [32] is another major pathway/genome database, which host numerous different organism databases at various curation levels. They provide tools like PathoLogic [33] to guess the metabolic network from the genome and SRI's pathways tools to adopt and curate networks. For an organism, the database contains the following information: genome, gene products, metabolic network, regulatory network and the transporter complement of the organism. Although the data is comprehensive (e.g., PathCase^{MAW} is only dealing with metabolic networks), browsing and visualizing capabilities are limited on their website. BioCyc provides a SVG-based, static visualization of the complete metabolic network, and links to related sources per item; but there is no other interaction, unlike the PathCase^{MAW} Visualization Interface. *Pathway Tools* [34] provide such features; however, it is an offline system and can be compared to the PathCase^{MAW} Editor. Computational tools provided by Pathway Tools focus on prediction of metabolic features such as pathways, choke points or operons for metabolic networks. On the other hand, SMDA Tool is concerned with analyzing metabolic profiles and predicting metabolic activity on metabolic networks. They also provide a Flux Balance Analysis tool, but only in the desktop version. BioCyc does have an iPhone application for only EcoCyc [35] and there is no application for iPad. In short, BioCyc metabolic network creation from the genome is a big (and, as they

also state, difficult) and different task than PathCase^{MAW} 's task, which deals with already created networks such as those from already reconstructed genome-scale networks. PathCase^{MAW} clearly provides a much smaller subset of the information provided by BioCyc; however, for metabolic networks, it provides a user-friendly, manageable and online system with nice browsing and visualizing capabilities.

Reactome [36-39] is a pathway database curated and updated by life-scientists. Although the visualization is static (e.g., no moving around of items); it provides beautiful images of selected pathways with zoom in/out capabilities and links to related resources (based on Systems Biology Graphical Notation). Reactome provides tools such as pathway analysis (e.g., overrepresentation analysis, mapping provided gene/protein ids to the pathways, comparing pathways of species). They also provide a built-in query interface similar to PathCase^{MAW}. There is no mobile interface provided. Reactome's goal is to have an all-inclusive one-stop-shop type of curated metabolic pathway database, whereas PathCase^{MAW}'s goal is to provide tools and schema for researchers to move their own data to PathCase^{MAW} schema to have the interfaces available.

MEMOSys [40] is a system that provides a platform for researchers to collaborate and create reconstructed metabolic networks. The system has a version control mechanism to keep track of the history of the models. The networks that are already created are stored in a repository. They can be browsed, compared and exported via the user-friendly web interface. They provide PDF based maps for some models. However, the system does not offer metabolomics analysis tools, query interface, mobile interface or interactive visualization.

In summary, PathCase^{MAW} is a unique system with its goal and capabilities. It has similarities and differences, compared to the existing systems. However it is not a stand-alone advancement over an existing system.

6.2 Challenges and Shortcomings

One of the biggest challenges of the PathCase^{MAW} system is with parsing reconstructed metabolic networks. And, the reason is the misuse of the SBML format for creating networks by researchers. For instance, Edinburgh metabolic network for humans [41] does not specify *compartment* attribute for species, as they do not distinguish between compartments, and assume all information is in the same single compartment *cell*. Similarly, many models include information (e.g., associated pathway/subsystem for a reaction) in the <notes> tag, which is simply plain text from the parser's point of view, and creates parsing problems for all SBML parsers. These challenges are explained in detail in Alshalwi, 2011 [42]. The second challenge is with the exponential complexity of the SMDA Tool, which is also explained in detail in Cakmak et al, 2012 [10]. For networks with high numbers of reactions and in those cases with a small number of metabolite observations, the SMDA system may have an unacceptably long response time, and may even run out of memory [11]. That said, note that this is a problem shared by all Constraint Based Methods [4] with similar goals and assumptions.

7. Conclusions and Future Work

PathCase^{MAW} is an online multi-part system to effectively create, store, browse, query, visualize and analyze metabolic networks. The system provides an SBML parser to create networks from a an SBML document, and a user-friendly interface that lets users to update the database. Stored information can be accessed online either through the web interface or the iPad application. PathCase^{MAW} is a useful system to access user-created reconstructed networks online for researchers. An analysis made online on the SMDA Tool has been published [11] as a proof of concept.

As future work, our goal is to link the data stored in the database to outside sources such as KEGG for reaction references and HMDB for metabolite references, whenever the information is provided in the input model. Next, we plan to have a collective site that includes all of the released models for the use of research community.

8. ACKNOWLEDGMENTS

Authors acknowledge the PathCase design and development contributions of many graduate students, This research has been supported by the National Science Foundation grants DBI 0743705, DBI 0849956, CRI 0551603 and by the National Institute of Health grant GM088823.

9. REFERENCES

[1] Fell, D. A. 1997. *Understanding the control of metabolism*. Portland Press, London, UK.

[2] Schilling, C.H., Schuster, S., Palsson, B.O., Heinrich, R. 1999. Metabolic Pathway Analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog*. 15, 296-303.

[3] Feist, A.M., Palsson, B.O. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat Biotechnol* 26, 6, 659-667.

[4] Zelezniak, A., Pers, T.H., Soares, S., Patti, M.E., Patil, K.R. 2010**.** Metabolic Network Topology Reveals Transcriptional Regulatory Signatures of Type 2 Diabetes. *PLoS Comput Biol*, 6, 4, e1000729.

[5] Patil, K.R., Nielsen, J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 102, 2685–2689.

[6] David, H., Hofmann, G., Oliveira, A.P., Jarmer, H., Nielsen, J. 2006. Metabolic network driven analysis of genome-wide transcription data from Aspergillus nidulans. *Genome Biol* 7, R108.

[7] Cicek, A.E., Bederman, I., Henderson, L., Drumm, M.L., Ozsoyoglu, G. 2013. ADEMA: An Algorithm to Determine Expected Metabolite Level Alterations Using Mutual Information *PLoS Comput Biol*, 9, 1, e1002859. doi:10.1371/journal.pcbi.1002859.

[8] Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.O. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS* 104, 6,1777-1782.

[9] Salway, J.G. 2006. *Metabolism at a Glance Third Edition*. Blackwell Publishing, Malden, MI.

[10] PathCase-RCMN **[**http://nashua.case.edu/PathCaseRCMN/Web/**]**

[11] Cakmak, A., Qi, X., Cicek, A.E., Bederman, I., Henderson, L., Drumm, M.L., Ozsoyoglu, G. 2012. A New Metabolomics Analysis Technique: Steady-State Metabolic Network Dynamics Analysis. *J Bioinform Comput Biol* 10, 1**,** 36.

[12] Cicek, A.E., Ozsoyoglu, G. 2012. Observation Conflict Resolution in Steady-State Metabolic Network Dynamics Analysis**.** *J Bioinform Comput Biol* 10, 1**,** 25.

[13] Cakmak, A., Ozsoyoglu, G., Hanson, R.W. 2012. Querying Metabolism under Different Physiological Constraints*. J Bioinform Comput Biol,* 8, 2, 247-293**.**

[14] Cakmak, A., Ozsoyoglu, G., Hanson, R.W. 2012. Managing and Querying Mammalian Metabolic Networks: A Metabolism Query Language and Its Query Processing. In *Proceedings of 8th International Conference on Computational Systems Bioinformatics* (Stanford, CA, August 10-13, 2009)*.* CSB '09.

[15] Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., et. al. 2009. HMDB: a knowledgebase for the human metabolome*. Nucleic Acids Res 37(Database issue), D603-610.*

[16] Apple Press Info **[**http://www.apple.com/pr/library/2011/07/19Apple-Reports-Third-Quarter-Results.html]

[17] Apple Education **[**http://www.apple.com/education/ipad/**]**

[18] Sabah Newspaper: Turkish Ministry of National Education distributes tablets in schools **[**http://english.sabah.com.tr/National/2012/02/01/distribution-of-free-tablets-to-students-has-begun**]**

[19] Hucka, M., Finney, A.M., Hoops, S., Keating, S.M., Le Novere, N. 2007. Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model definitions. Sept. 26, 2007.

[20] Das, M. 2009. Parsing BioModels and CellML Models and insertion of data. Masters Project*.* Case Western Reserve University, Electrical Engineering and Computer Science Department.

[21] PathCase-Systems Biology Web Site [http://nashua.case.edu/PathwaysSB/Web]

[22] Cakmak, A., Qi, X., Coskun, S.A., Das, M., Cheng, E., Cicek AE, Lai N, Ozsoyoglu ZM, Ozsoyoglu G: PathCase-SB Architecture and Database Design. *BMC Systems Biology* 2011, 5:188.

[23] Elliott, B., Kirac, M., Cakmak, A., Yavas, G., Mayes, S., Cheng, E., Wang, Y., Gupta, C., Ozsoyoglu, G., Ozsoyoglu, Z.M. 2006 PathCase: pathways database system. *Bioinformatics*, 24**:** 2526–2533.

[24] Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L., Hucka, M. 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research,* 34(Database-Issue), 689-691.

[25] Bornstein, B.J., Keating, S.M., Jouraku, A., Hucka, M. 2008 LibSBML: an API Library for SBML. *Bioinformatics,* 24, 6, 880-881**.**

[26] KEGG (Kyoto Encyplopedia of Genes and Genomes) Pathways **[**http://www.genome.jp/KEGG/pathway.html**]**

[27] Kanehisa, M., Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28, 27-30**.**

[28] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. 2006. From genomics to chemical genomics: new developments in KEGG**.** *Nucleic Acids Res,* 34, D354-357**.**

[29] Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res,* 38, D355-D360**.**

[30] PathCase-KEGG **[**http://nashua.case.edu/PathwaysKegg/Web/

[31] Johnson, S.R., Qi, X., Cicek, A.E., Ozsoyoglu, G. 2013 iPathCaseKEGG: an iPad interface for KEGG metabolic pathways. Health Inf Sci Syst, 1, 4.

[32] Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Paley, S., Popescu, L., Pujar, A., Shearer, A.G., Zhang, P., Karp, P.D. 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research,* 38, D473-9.

[33] Paley, S.M., Karp, P.D. 2002. Evaluation of computational metabolic-pathway predictions for H. pylori. *Bioinformatics* 18, 715-24.

[34] Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R. 2010. Pathway Tools version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *Briefings in Bioinformatics*, 11, 40-79.

[35] Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A.G., Mackie, A., Paulsen, I., Gunsalus, R.P., Karp, P.D. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology**.** *Nucleic Acids Research*, 39, D583-590.

[36] Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., Kanapin, A., Lewis, S., Mahajan, S., May, B., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L., D'Eustachio, P. 2008. Reactome knowledgebase of biological pathways and processes. *Nucleic Acids Res*, 37(Database issue), D619-22.

[37] Matthews, L., D'Eustachio, P., Gillespie, M., Croft, D., de Bono, B., Gopinath, G., Jassal, B., Lewis, S., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L. 2007. An Introduction to the Reactome Knowledgebase of Human Biological Pathways and Processes. *Bioinformatics Primer, NCI/Nature Pathway Interaction Database*. doi:10.1038/pid.2007.3

[38] Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., Wu, G., Birney, E., Stein, L. 2007. Reactome: a knowledge base of biologic pathways and processes. *Genome Biology*, 8, R39.

[39] Joshi-Tope, G., Vastrik, I., Gopinathrao, G., Matthews, L., Schmidt, E., Gillespie, M., D'Eustachio, P., Jassal, B., Lewis, S., Wu, G., Birney, E., Stein, L. 2003. The Genome Knowledgebase: A Resource for Biologists and Bioinformaticists. *Cold Spring Harb Symp Quant Biol.*, 68, 237-43.

[40] Pabinger, S., Rader, R., Agren, R., Nielsen, J., Trajanoski, Z. 2011 MEMOSys: Bioinformatics platform for genome-scale metabolic models. *BMC Syst Biol*, 5, 20.

[41] Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., Goryanin, I. 2007. The edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*, 3, 135.

[42] Alshalwi S 2012. PathCase-RCMN: ReconstruCted Metabolic Networks of organisms. Masters Project. Case Western Reserve University, Electrical Engineering and Computer Science Department.