Social pathway annotation: extensions of the systems biology metabolic modelling assistant

Ismael Navas-Delgado*, Alejandro Real-Chicharro*, Miguel Ángel Medina, Francisca Sánchez-Jiménez and José F. Aldana-Montes

Submitted: 4th May 2010; Received (in revised form): 1st August 2010

Abstract

High-throughput experiments have produced large amounts of heterogeneous data in the life sciences. These data are usually represented in different formats (and sometimes in technical documents) on the Web. Inevitably, life science researchers have to deal with all these data and different formats to perform their daily research, but it is simply not possible for a single human mind to analyse all these data. The integration of data in the life sciences is a key component in the analysis of biological processes. These data may contain errors, but the curation of the vast amount of data generated in the ‘omic’ era cannot be done by individual researchers. To address this problem, community-driven tools could be used to assist with data analysis. In this article, we focus on a tool with social networking capabilities built on top of the SBMM (Systems Biology Metabolic Modelling) Assistant to enable the collaborative improvement of metabolic pathway models (the application is freely available at http://sbmm.uma.es/SPA).

Keywords: social networks; metabolic modeling; systems biology

INTRODUCTION

Almost a century of metabolic research has yielded a great amount of diverse information. During the last decade, the incorporation of high-throughput experimental procedures has accelerated the velocity of acquisition of new biological data of metabolic relevance. New systemic approaches are required to gain a comprehensive view of metabolism as a whole [1]. Because hypothesis-driven systems biology approaches involve iterative rounds of model building, prediction, experimentation, model refinement and development, an increasing number of metabolic models are available in the literature and accessible from several repositories and databases.

The System Biology Metabolic Modeling Assistant (http://sbmm.uma.es) [2] is a tool developed to search, visualise, manipulate and annotate both identity data and kinetic data. The possible inputs for searching metabolic pathways are the pathway’s name or KEGG (Kyoto Encyclopedia of Genes and Genomes) code, a set of enzymes or a correctly annotated model (user-defined or not). Users can retrieve online data about metabolic pathways and then edit them. This tool allows users to
export the metabolic pathway to SBML format, enriched automatically without any previous configuration using Minimum Information Requested In the Annotation of biochemical Models (MIRIAM) [3] and CellDesigner 4.0 annotations [4]. This facilitates good presentation of the data based on the Systems Biology Graphical Notation (SBGN annotations) [5], which are visual notations for network diagrams in systems biology.

Individual users can make single improvements of different metabolic pathways based on their experience. Collectively, a community of scientific users could provide highly curated metabolic pathways based on the community’s experience. In this sense, social networking is a new issue that is attracting a lot of interest in many domains. In the case of biology, research networks have usually been closed ones, in which the most prominent researchers decided the roadmap of this research. Currently, new technologies are opening up new opportunities for knowledge exchange by means of social networks, creating new knowledge-driven digital communities.

Metabolic Modelling is attracting a lot of interest due to its potential to study biological processes from a systems biology point of view. In this context, several tools provide different capabilities for metabolic modelling, such as Payao [6], which is based on CellDesigner, ByoDyn [7], BioPP [8], WikiPathways [9], COPASI [10] and Sycamore [11]. Additionally, these tools are starting to include some social networking capabilities, which partly fulfil the needs of community-based metabolic pathway curation. In this article, we present Social Pathway Annotation (SPA, http://sbmm.uma.es/SPA), an extension of the SBMM Assistant, as community-based approach to the curation process. This extension includes the following relevant tools for this community of users:

- a bibliographic tool that allows users to discover scientific networks through the links between papers related to a metabolic pathway, reaction or biological component;
- a curation tool that enables the editing of metabolic pathway elements (metabolites, reactions and their kinetics) by individual users; and
- a social network tool that provides users a way to store metabolic pathway models and collaboratively curate them.

### SPA System Description

The SPA extensions (Figure 1) have been developed to complement the SBMM Assistant so that the community of users can annotate pathway models. Thus, these extensions will allow users to collaboratively curate metabolic pathway models and improve upon the existing annotation of the interaction among scientific networks. Users can take advantage of the following main characteristics of this tool:

- online access to external databases (UNIPROT [12], KEGG [13], CHEBI [14], BRENDA [15] and SABIO-RK [16]) to retrieve information related to metabolic pathways, using a mediator-based solution (KOMF [17]);
- graphical representation of the metabolic pathways using CellDesigner tags inspired by SBGN annotations; and
- export option for metabolic pathways to SBML [18].

SPA extensions are divided into two main elements: a metabolic pathway model repository (to enable the social annotation of metabolic pathways stored in this repository) and a bibliographic analysis tool (to understand the context of the research studies developed in relation to the selected metabolic pathway).

### Metabolic Pathway Model Repository

Using SBMM Assistant capabilities, users can retrieve complete metabolic pathways containing information from different data repositories. The first step for metabolic pathway annotation and curation involves saving the model of interest locally. Later, the model is uploaded to the metabolic pathway model repository by specifying the organism, a description (as complete as possible) and a model name (the owner and the creation date are implicit and automatically added to the model).

The social curation process includes several steps (Figure 2). First step allows users to comment and vote on the whole metabolic pathway. These votes are categorized (Correct, Correct with minor mistakes, Correct with major mistakes and Incorrect) and a chart shows the proportion of votes introduced by the users (Figure 3). Comments and votes are emerging semantics on the metabolic pathway quality and using them a user can produce a new version of the metabolic pathway, solving the mistakes...
Figure 1: System architecture.
detected on the original version. The versioning tool orders the metabolic pathway versions in a tree, where root is the model uploaded by a user, leaves are curated versions of the metabolic pathway and the nodes are intermediate versions of the metabolic pathway partly curated.

The metabolic pathway model repository allows users to save models, share them with other users and create improved versions of existing models. At this point, we have introduced the ownerships of the models. Thus, a model (or version of a model) can be modified by his/her owner, and only new versions or commentaries can be added by other users. When a user begins to edit a model, the downloading of data from different databases is automatically started, and the user can start working on this metabolic pathway. Newly created and existing metabolic models can be curated and locally stored as SBML files. The tool allows users to perform different tasks in the curation process, including the modification of reaction data (concentrations, reaction parameters, kinetic laws, stoichiometric equations, etc.) and the addition of new components (substrate, product, activator, inhibitor, etc.) in the metabolic pathway (Figure 4).

The metabolic pathway model repository enables navigation in the repository to discover curated metabolic pathways. This navigation panel includes not only user pathways but also versions of these models. Thus, different users can produce different kinds of annotations, such as new versions of metabolic pathways, votations and comments on other user pathways (Figure 3). When a user displays a pathway model and detects errors, inconsistencies or mistakes, then he/she can vote with one of four options: Correct, Correct with minor mistakes, Correct with major mistakes and Incorrect. The vote should be accompanied by comments about the model errors. These SPA extensions facilitate communication between researchers, enabling them to interact and participate in discussions concerning experimental data.

### Bibliographic analysis tool

Retrieved metabolic pathways contain information from different databases, which can contain bibliographic references concerning the data they store. To take advantage of this information, the SBMM Assistant has been extended in SPA to cover the relationships between bibliographic references and the different pathway components. This extension provides a way to analyse the networks of publications related to an enzyme or a complete metabolic pathway. This tool enables users to locate those researchers working in a given field. Thus, when a user retrieves information about a metabolic pathway in which he/she is interested, the user will be presented with a graph that contains two types of nodes (researcher and publication nodes) and two types of connections (the author connection and the co-author connection).

This tool provides different icons in the bibliographic network of a metabolic pathway to show the differences between nodes (publications and
authors). It is useful to follow the relationship between an interesting publication and the author’s ‘neighbourhood’ (the authors working in the same research area), to query and visualise related publications in Pubmed, and to look for relevant background information on the author (Figure 5).

**USE CASES**

As use cases, we asked a group of users to deal with the task of curating a malformed metabolic pathway using the SPA extension. The first user group consisted of biologists with basic notions of metabolic modelling. This kind of user can detect basic mistakes and perform simple curation tasks with the tool. The second group consisted of a set of biologists, who had an average-to-high level of comprehension of metabolic modelling but a low level of expertise in the use of modelling tools. This kind of user can perform more complex curation tasks. Neither group had particularly good knowledge of IT tools because we aimed to discover the ‘assistant’ capability of the tool. To perform the test, we had developed two metabolic pathway models that were a malformed version of an...
already curated metabolic model, the Teusink model of glycolysis in *Saccharomyces cerevisiae* [19] in SBML format (Figure 6). The test group had to curate it by building a *Homo sapiens* glycolysis model. The test model contained some errors, including the nonexistence of some species, some malformed kinetic equations and some nonexistent reactions.

The selected users were able to deal with the curation of the proposed models. In the case of the less-experienced users, the curation annotation was focussed on indicating some errors, but these users did not deal with the modification of the model because of their inexperience with modelling tasks.

In contrast, the most-experienced users were able to discover and correct the introduced errors, even with their low knowledge of IT tools. In this case, users created new versions of the proposed model by pruning the introduced errors. Thus, the more-experienced users introduced the correct annotations to perform the following functions:

- to add new compounds (such as dihydroxyacetone-phosphate);
- to add new reactions (such as glucose transport);
- to change the compartment of some compounds; and
- to interchange compounds (for example, to substitute the less specific term Triose-phosphate with *Glyceraldehyde-3-phosphate*, the product of *Aldolase* and substrate of *Glyceraldehyde-3-phosphate dehydrogenase*).

Additionally, users could change the reactions themselves by changing not only products and substrates but also activators, inhibitors, kinetic laws and kinetic parameters. However, in the described use cases, users did not perform this task because this would have required a more specific knowledge of

---

**Figure 4:** Reaction Update Tool. This figure shows how a reaction can be completed by adding new substrates and products.
RELATED WORK

We have analysed some of the more well-known applications (Table 1) of Metabolic Modelling (COPASI [10], Payao [6] based on CellDesigner [4], ByoDyn [7], BioPP [8], WikiPathways [9] and Sycamore [11]) and their social networking capabilities. Also, we have found social curation improvements in a more specific area like kinetic literature curation, such as the SABIO-RK [20] database. We have taken into account some characteristics such as their capability to offer a powerful simulation service, a kinetic service to provide stored data from literature or databases and bibliographic tools. Pathway model management has also been taken into account because this is the main requirement to create a social platform. Related with the model management, the analysis of model-versioning capabilities has been checked to detect if they include improvement and correction management for shared models. Pathway model versioning is needed for social metabolic pathway curation as a way to evaluate multiple ways of curating the same model, leaving the community to choose the best solution among multiple ones. Thus, the community knowledge will be used

Figure 5: Bibliographic Network example for the enzyme 1.2.4.1 (pyruvate dehydrogenase). Main authors are linked with red arrows, publications are the green boxes and the authors are those in the blue boxes.
to prune invalid pathway models, by introducing model annotations.

Foremost in 2003, CellDesigner [4] emerges as one of the most-used process diagram editors for biochemical networks. It had a nice and useful graphical user interface (SBML-compatible). This interface provided the visualization of the metabolic models using a draft of the SBGN standard. A better and more complete simulator for the models should be improved. Currently, CellDesigner provides an API for plugins that allows users to extend the functionalities of the application and interconnection to SBW-powered [21] simulators. Additionally, this tool provides a new feature for searching over SABIO-RK [20]. It has a free-use license and is a Java program downloadable from http://www.celdesigner.org. The BioModels database [22] was published in January 2006 as the first aim to centralize correctly curated kinetic models. The way BioModels works can be summarized as follows: (i) The owner of a model sends it to BioModels. (ii) Persons in care of BioModels make a curation process (that implies a consistency check, curation and simulation). (iii) Each part of the model is submitted to an annotation process. (iv) And, finally, the model is published, open and accessible to third users.

Figure 6: Original Teusink model with mistakes on purpose added.
Currently, the database has been improved [23] and it is still growing, counting with >450 reactions and models. To provide a feature allowing users to collaborate in the curation process would be very interesting to increase the quality of models available in this repository. It is SBML-compatible and its models are available in http://www.ebi.ac.uk/biomodels-main. COPASI [10], firstly published in December 2006, is surely the most complete SBML-compatible simulator for biochemical networks. However, COPASI lacks of collaborative features to provide social curation/simulation capabilities. It is available at http://www.copasi.org as both a free version and a commercial one. Launched also in 2006, ByoDyn [7] started as a promising SBML-compatible editor and simulator for metabolic pathways, providing a remote manager and a repository to manage the metabolic models generated with the application. It is a Web-based solution available on http://cbbl.imim.es:8080/ByoDyn.

After a time, in May 2007 BioPP [8] was published, allowing users to export a SBML model into HTML for their own purposes or to allow the scientific community to access it, enabling hyperlinks on model elements to related data repositories. It is a good application for deploying final solutions for fully detailed information searches, but it does not allow the tagging, annotation or versioning of the models, which can provide necessary mechanisms to efficiently curate the models in a medium–large community of curators. Thus, it is currently only useful for small communities. In summary, by 2007 we could find two groups of metabolic pathway applications: on one side, stand-alone editors and simulators for metabolic models; and on the other side, applications introducing features to curate metabolic pathways on a collaborative way. However, the integration of most functionalities was still not enough for social curation, not enough to avoid a growing group to curate their own metabolic models. A point of no return was the publication of WikiPathways [9] in April 2008. This wiki web is the actual first application able to provide the essential features for social curation. It allows managing, editing and versioning of pathways with a great efficiency and ease. It does not support exportation of pathways to SBML [18], BioPAX [24], CellML [25] or PSI-MI [26] formats, which would allow users to correctly exchange, simulate and annotate standard identifiers for each reaction or compound, and allows the use of the model with tens of standard compatible applications (i.e. SBML). Currently, it has perhaps the biggest user community (>1000 users). It is available at http://www.wikipathways.org.

With the publication of Sycamore (June 2008) and SBMM Assistant (January 2009), the concept of assistance was implemented allowing users to curate the metabolic models. Both approaches provide automatic kinetic searching. Sycamore is a web-browser application able to construct, simulate and analyse metabolic models. It is useful for building annotated kinetic models because it provides resources for extracting data to SABIO-RK. It does not provide capabilities for social curation, but it provides tools for users to manage the models. SBMM Assistant integrates curated kinetic databases as SABIO-RK, Brenda, ChEBI, KEGG and UniProt managed by the KOMF [27] mediator and using the AMMO ontology [28]. Sycamore is a SBML-compatible web-based application available in http://sycamore.eml.org/sycamore under free academic use license, and SBMM Assistant is a SBML-compatible Java with an application available in http://www.sbmm.uma.es under Creative Commons license.

Table 1: Social and distributed capabilities comparison between common modelling applications

<table>
<thead>
<tr>
<th>Payao</th>
<th>Byodyn</th>
<th>BioPP</th>
<th>WikiPathways</th>
<th>Copasi</th>
<th>Sycamore</th>
<th>SBMM SPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>BT</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>MM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>MV</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>MTC</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

SS, simulation service; KS, kinetic searcher; BT, bibliography tools; MM, modeling managing; MV, model versioning; MTC, model tagging and commenting; X, functional capability.
Finally in April 2010 appears Payao, which uses CellDesigner to show SBML models. It includes capabilities for privilege levels, adding tags to some targets (as reactions) and commenting upon these tags, all in real time and concurrently. Model management can occur in three ways (all, favorites, own). Currently, Payao does not allow the versioning of models and eventually, it will allow models to be updated, but it is in the right way to create a useful community of users to curate metabolic pathways. This is a SBML-compatible web-application available in http://sblab.celldesigner.org/Payao10/bin.

All the previously commented tools fall short of providing a complete set of tools for enabling the community curation of metabolic pathways such as searching capabilities, a metabolic pathway repository, annotation tools and metabolic pathway versioning. Thus, the problem is partly solved by existing tools, but SPA tends towards providing a complete solution for metabolic pathway curation based on community knowledge. The versioning capability is an important characteristic for community curation because it is a means of producing improved versions of metabolic pathways. This characteristic is shared with WikiPathways. There are also social curation capabilities in Payao, BioPP and WikiPathways. Additionally, SPA provides a bibliographic tool, a unique feature that enables the analysis of the context of a given metabolic pathway.

DISCUSSION AND CONCLUSIONS
Curation tools must evolve towards social curation [29], due to the impossibility (both in terms of people and resources) of managing the huge quantity of constantly increasing biological information provided by experimental methodologies.

In the future, all new tools, upgrades of old ones and collaborations must have as basis the use of standards like the interchange format SBML, the visual interchange format SBGN and the standard for minimum information into a biochemical model MIRIAM. This will provide the main point for common work between multiple user and tools.

We have developed a novel tool that tests some basic theoretical aspects of the future of data curation [30], based on metabolic pathway curation. The advantages of this solution in comparison to a conventional metabolic pathway curation are as follows:

- many eyes are able to look for model inconsistencies, whereas in conventional curation only a few eyes look for those inconsistencies. This capability allows better quality of results to be obtained because more controlled discussions are developed to choose the best options. This capability also allows us to obtain faster curations because bottlenecks could be prevented by using the cumulative experience of many users;
- the controlled creation of models and versions allows the production of new ordered knowledge, which would not be the case if a lot of files were shared as an unordered amalgam of names;
- the use of a controlled vocabulary allows errors to be kept to a minimum in curation and also helps with the organisation, maintenance and querying of the constantly incoming metabolic pathway models; and
- the extension over an easy-to-use, albeit powerful, and graphical user interface facilitates the curation task for inexperienced users.

This tool has been tested with two user groups in a short-scale experiment. Because these results are limited by the small set of users involved in these use cases, future work will expand these use cases to include enough users to ensure statistical significance; these future experiments will allow us to ascertain SPA usability and the usefulness of the collaborative curation of metabolic pathways. Looking forward, upgrades will be incorporated in the near future:

- Currently, the annotations of changes are manually provided by the user. The idea is to provide an automatic controller of changes, which controls the changes between a model/version and other versions.
Concerning the bibliography tool, relationships that could be established between an article and its citations would improve its current potential.

- Social curation allows users to annotate inconsistency within a metabolic pathway. However, parallel curation tasks can be opened for the same metabolic pathway. In our future work we plan to detect these situations to enable ways of combining similar efforts in the same pathway. A mechanism to find common resources between models must be implemented to allow users to propose the suppression of repeated pathways or the migration of a model as a version of another model.
- The owner of a kinetic model will be able to establish himself as moderator of his model, to delegate this task on another or allow the community to assume the task of moderator.
- We believe that, in the near future, a necessary improvement for social curation will be the management of kinetic model results (e.g. simulation results). This capability will enable a deep analysis of model quality and usefulness.
- Social networking can be used in teaching activity, and the tool could prove to be effective in improving student ability in the field. Thus, the use of this tool could be included in the courses in which pathway modelling is taught.

**Key points**

- A bibliographic tool that allows users to discover scientific networks through the links between papers related to a metabolic pathway, reaction or biological component.
- A curation tool that enables the edition of metabolic pathway elements (metabolites, reactions and their kinetics) by individual users.
- A social network tool that provides users a way to store metabolic pathway models and collaboratively curate them.

**Acknowledgements**

We would like to thank Daniel Pastor for his constant testing of the tool and advice, Amine Kerzazi for his continuous maintenance of the data wrappers and Ian Morilla for his invaluable help in the Tutorial composition. We would also like to mention the undergraduate students enrolled in a Metabolic Biochemistry Group and the graduate students enrolled in a master course on ‘Analysis and Modelling of Complex Biological Systems’ for testing the system in the use cases described in this manuscript.

**FUNDING**

Plan Andaluz de Investigación (BIO-267, P07-CVI-02999, P07-TIC-02978, TIC-136), the Spanish Ministry of Sciences and Innovation (TIN2008-04844, SAF2008-02522, PS09/02216) and Fundación Ramón Areces. The ‘CIBER de Enfermedades Raras’ is an initiative from the ISCIII (Spain).

**References**


