

# Systems biology model repository for macrophage pathway simulation

Masao Nagasaki<sup>1,\*</sup>, Ayumu Saito<sup>1</sup>, André Fujita<sup>2</sup>, Georg Tremmel<sup>1</sup>, Kazuko Ueno<sup>1</sup>, Emi Ikeda<sup>1</sup>, Euna Jeong<sup>1</sup> and Satoru Miyano<sup>1,2</sup>

<sup>1</sup>Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan and <sup>2</sup>Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 05508-090, SP, Brazil

Associate Editor: Alex Bateman

## ABSTRACT

**Summary:** The Macrophage Pathway Knowledgebase (MACPAK) is a computational system that allows biomedical researchers to query and study the dynamic behaviors of macrophage molecular pathways. It integrates the knowledge of 230 reviews that were carefully checked by specialists for their accuracy and then converted to 230 dynamic mathematical pathway models. MACPAK comprises a total of 24 009 entities and 12 774 processes and is described in the Cell System Markup Language (CSML), an XML format that runs on the Cell Illustrator platform and can be visualized with a customized Cytoscape for further analysis.

**Availability:** MACPAK can be accessed via an interactive web site at <http://macpak.csml.org>. The CSML pathway models are available under the Creative Commons license.

**Contact:** masao@hgc.jp

Received on November 12, 2010; revised on March 22, 2011; accepted on March 30, 2011

## 1 INTRODUCTION

The immune system has evolved as a defense mechanism against pathogenic microbes. It consists of both (i) innate immunity, which is evolutionary ancient and (ii) adaptive immunity, which provides specific recognition and immunological memory (Janeway and Medzhitov, 2002). Cellular participants in the immune and inflammatory responses include antigen-presenting cells and phagocytic cells, such as macrophages.

Significant progress has been made recently in understanding the biochemical mechanisms by which lipopolysaccharide (LPS), a bacterial endotoxin which activates immune cells, particularly macrophages (Miller *et al.*, 2005), leads to sepsis, septic shock or systemic inflammatory response syndrome. LPS is a major component of the outer membrane of Gram-negative bacteria and one of the most potent microbial initiators of inflammation (Cohen, 2002). Despite much effort, unfortunately there is no effective therapy for this problem. It is estimated that ~50 000 people die each year only in the USA of septic shock or systemic inflammatory syndrome (Guha and Mackman, 2001). Thus, a more detailed understanding of the molecular mechanisms (pathways) underlying activation and deactivation of macrophages in response to LPS

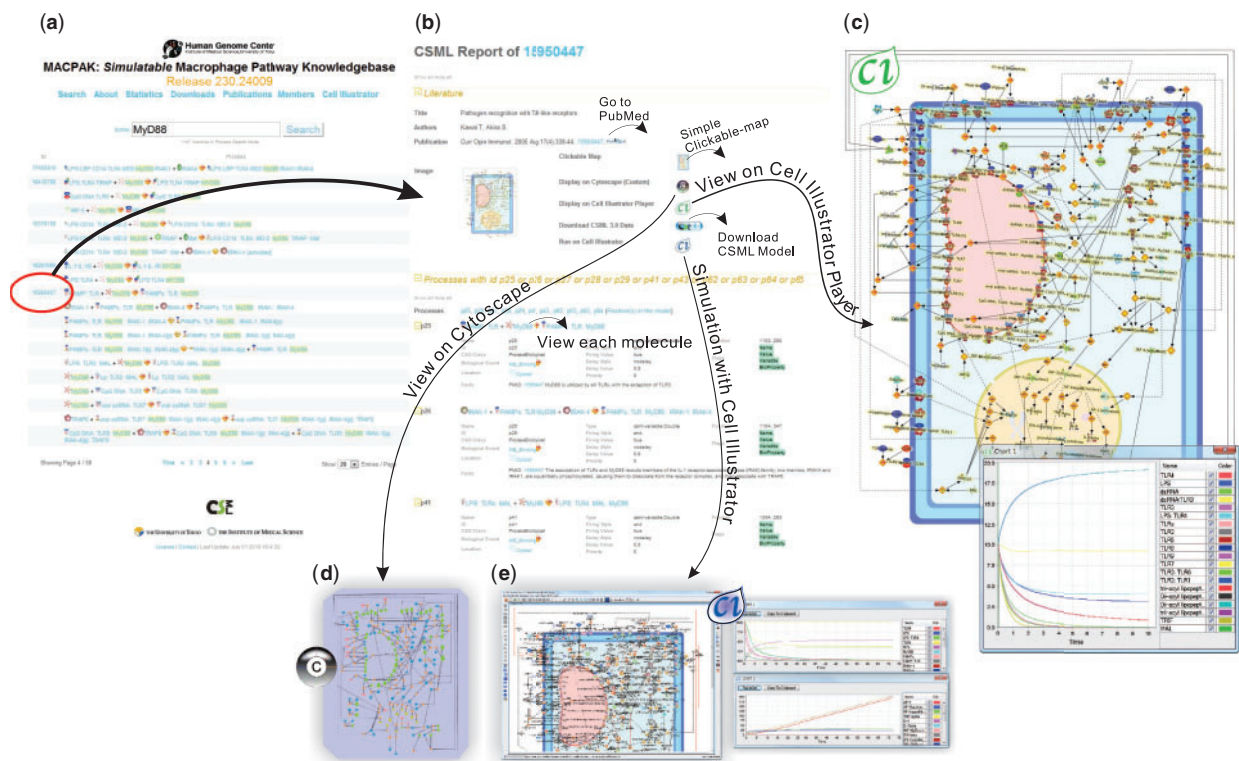
becomes crucial for the development of novel therapies (Palsson-McDermott and O'Neill, 2004).

There are several database resources describing general molecular pathways, such as BioModels Database (Li *et al.*, 2010), CellML model repository (Lloyd *et al.*, 2008), JWS Online (Olivier and Snoep, 2004), Database of Quantitative Cellular Signaling (Sivakumaran *et al.*, 2002) and SABIO-RK (Rojas *et al.*, 2007). The main advantage of these resources is that they are databases which allow simulation of the pathways. On the other hand, some macrophage-specific databases are also proposed, such as (Lynn *et al.*, 2008; Oda *et al.*, 2004) and (Raza *et al.*, 2010). However, to the best of our knowledge, none of them support simulation. In order to aid immunologists quest for an effective septic shock treatment, we present the first knowledgebase that is specifically focused on LPS-induced macrophage signaling pathways and also allows dynamic simulation of manually curated macrophage pathways, namely MACPAK. Here, 'simulation' means that one can perform *in silico* experiments by changing the parameters of the pathway model, for example, knocking down one gene, in order to analyze the results of perturbation before an *in vitro* or *in vivo* analysis. Therefore, experiments can be planned and focused in a problem-driven manner, by predicting the results and consequently, economizing time, human labor and money.

## 2 MACPAK

MACPAK contains information retrieved from 230 reviews which were searched from NCBI PubMed with the expressions *LPS and macrophage*, *PMA and macrophage* (PMA: Phorbol 12-myristate 13-acetate) and sometimes including the word *monocyte*. The complete list of reports used in this work is summarized at <http://macpak.csml.org/statistics/>. Each report was carefully curated and 'translated' to one pathway in an XML format [Cell System Markup Language (CSML); <http://www.csml.org>] on Cell Illustrator. This task was carried out by four researchers for 2 years and several free softwares were used in order to design the CSML models (imagemagick <http://www.imagemagick.org>, subversion <http://subversion.apache.org> and trac <http://trac.edgewall.org>). For more details about the curation criteria and rules, refer to the manual (<http://macpak.csml.org/about/>). During curation, if the quantitative information (initial concentration and maximum value for biological entity, order for priority and speed for biological process) are

\*To whom correspondence should be addressed.



**Fig. 1.** (a) Example of search result obtained for the keyword *MyD88*. Each ID represents one pathway; (b) by clicking on the ID 15950447, the description of the pathway is displayed. In this page, the user can find the paper information (from where the information about this pathway was retrieved), links to PubMed or details about the processes present in the model. Each process is represented by a unique icon. Detailed descriptions of the used icons can be found at <http://macpak.csml.org/statistics/>. There are also links to several visualization/simulation platforms such as Cytoscape (Smoot *et al.*, 2011), Cell Illustrator Online and Cell Illustrator Player (Nagasaki *et al.*, 2010). Users can also download the CSML model from this page; (c) the simulatable CSML pathway model on Cell Illustrator Player. Here is possible to run the simulation and observe what happens in the biological system; (d) the pathway can be visualized on Cytoscape and then saved in the numerous file formats available by this platform. For example, after uploading on Cytoscape, the user can export the CSML model in the XGMML, Vizmap, SIF and GML formats and use this file for further downstream analysis in other platforms; (e) the simulation of the pathway on Cell Illustrator Online. By using CIO, one can change the parameters of the model and visualize the results under different perturbations.

available then the provided value is used. Otherwise, the quantity is set to a default value automatically by Cell Illustrator.

Currently, the knowledgebase contains 230 pathways, one for each review, summing a total of 24 009 entities (i.e. mRNA, protein, modified protein and complex), 12 774 processes (binding, phosphorylation, activation, ubiquitination, for instance) and 39 823 connectors (reactant process and process–product relationship). Each reaction in this repository is annotated with one or more facts derived from the original literature as evidence.

Since MACPAK is manually curated (by four researchers that built, checked and double-checked the CSML models), the quotations from the original publication can be used as a benchmark to test and evaluate the performance of text mining tools that try to reconstruct pathways by using literature information. For example, consider the following process: PMID: 11257452. In the ‘Facts’ section, there is the sentence ‘Adaptor protein called TAB2 that mediates activation of TAK1 by linking TAK1 to TRAF6’. This sentence was copied and pasted from the corresponding paper and used by the annotators to construct the CSML model. Therefore, text mining researchers can evaluate their methods using these ‘Facts’ (texts) as inputs and evaluate the performance by comparing

the output of their methods with the CSML models that can be downloaded for non-commercial use under the Creative Commons license.

In order to access each macrophage pathway, the user only needs to input the name of the gene/protein of interest or click on the suggested links in the web page (<http://macpak.csml.org>). Then, a list of pathways that contains the gene/protein of interest is shown (Fig. 1a). After selecting the pathway to be analyzed, the web page (Fig. 1b) containing the information about the process is shown, which has links to external databases like Entrez Gene, InterPro and PubMed and an executable file that can be simulated under Cytoscape, Cell Illustrator Online and Cell Illustrator Player platforms. Figure 1c shows an example of model that was obtained by analyzing the *MyD88* pathway using Cell Illustrator Player. Figure 1d is the output of Cytoscape and Figure 1e is the simulation result using Cell Illustrator Online.

### 3 RESULTS AND DISCUSSIONS

Although there are other important databases that provide macrophage information, such as proposed by (Oda *et al.*, 2004),

(Lynn *et al.*, 2008) and (Raza *et al.*, 2010), and to the best of our knowledge, MACPAK is the first macrophage-specific database which integrates manually curated network (pathway) information to experimental quantitative results and also allows researchers to simulate and study the pathways in a dynamic and interactive manner. If observed time-course biological data are available (microarray data, for instance), it can be used to find the best set of parameters that fit to the model by using Data Assimilation techniques (Koh *et al.*, 2010). Therefore, by combining the pathway models and actual biological data, biomedical researchers can obtain hints of what is happening and what will happen when the biological system is perturbed, for example, by inhibiting one pathway using a drug.

A local installation of MACPAK is not necessary because all the applications including Cell Illustrator Online, Cell Illustrator Player and a customized Cytoscape run under the Java Web Start technology. The customized Cytoscape has a CSML plugin enabling Cytoscape to import CSML models developed by our members. All the pathways are stored in an XML format, shown in an HTML file in the web browser, while the simulations run via Internet using Cell Illustrator Online. Once a CSML file is imported to the customized Cytoscape (Smoot *et al.*, 2011), our model can be exported to various formats, e.g. XGMML, Vizmap, SIF and GML (Graph Modeling Language), which allow further visualization and analysis by other platforms such as eXpanda (Negishi *et al.*, 2007) and MAVisto (Schreiber and Schwobbermeyer, 2005). Moreover, ontology information (Cell System Ontology) of each entity and/or process is also provided. Thus, the curated pathway can also be validated using the ontology-based approach proposed by (Jeong *et al.*, 2011).

In order to expand the usability of MACPAK, several functionalities are continuously being developed, such as the software to convert CSML 3.0 to the well-known SBML level 2 version 4 (Systems Biology Markup Language) (Hucka *et al.*, 2003). With the customized Cytoscape available at the web page, it is possible to save the CSML model in any format supported by Cytoscape and continue the downstream analysis in other platforms that accept Cytoscape output files. In parallel, a more interactive interface based on emerging web standards and a graphical representation following the Systems Biology Graphical Notation (SBGN) are also being developed (Le Novère *et al.*, 2009).

*Conflict of Interest:* none declared.

## REFERENCES

- Cohen, J. (2002) The immunopathogenesis of sepsis. *Nature*, **420**, 885–891.
- Guha, M. and Mackman, N. (2001) LPS induction of gene expression in human monocytes. *Cell. Signal.*, **13**, 85–94.
- Hucka, M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **24**, 524–531.
- Janeway, C.A. and Medzhitov, R. (2002) Innate immune recognition. *Annu. Rev. Immunol.*, **20**, 197–216.
- Jeong, E. *et al.* (2011) Ontology-based instance data validation for high-quality curated biological pathways. *BMC Bioinformatics*, **12** (Suppl. 1).
- Koh, C.H. *et al.* (2010) DA 1.0: parameter estimation of biological pathways using data assimilation approach. *Bioinformatics*, **26**, 1794–1796.
- Li, C. *et al.* (2010) BioModels Database: an enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.*, **4**, 92.
- Le Novère, N. *et al.* (2009) The systems biology graphical notation. *Nat. Biotechnol.*, **27**, 735–741.
- Lloyd, C.M. *et al.* (2008) The Cell model repository. *Bioinformatics*, **24**, 2122–2123.
- Lynn, D.J. *et al.* (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.*, **4**, 218.
- Miller, S.I. *et al.* (2005) LPS, TLR4 and infectious disease diversity. *Nat. Microbiol.*, **3**, 36–46.
- Nagasaki, M. *et al.* (2010) Cell Illustrator 4.0: a computational platform for systems biology. *In Silico Biol.*, **10**.
- Negishi, Y. *et al.* (2007) eXpanda: an integrated platform for network analysis and visualization. *In Silico Biol.*, **7**, 141–144.
- Oda, K. *et al.* (2004) Molecular interaction map of macrophage. *AfCS Res. Reports*, **2**, 1–12.
- Olivier, B.G. and Snoep, J.L. (2004) Web-based kinetic modelling using JWS online. *Bioinformatics*, **20**, 2143–2144.
- Palsson-McDermott, E.M. and O'Neill, L.A.J. (2004) Signal transduction by the lipopolysaccharide receptor, Toll-like-4. *Immunology*, **113**, 153–162.
- Raza, S. *et al.* (2010) Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. *BMC Syst. Biol.*, **4**, 63.
- Schreiber, F. and Schwobbermeyer, H. (2005) MAVisto: a tool for exploration of network motifs. *Bioinformatics*, **21**, 3572–3574.
- Sivakumaran, S. *et al.* (2002) The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, **19**, 408–415.
- Smoot, M.E. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
- Rojas, I. *et al.* (2007) SABIO-RK: a database for biochemical reactions and their kinetics. *BMC Syst. Biol.*, **1** (Suppl. 1S6).