SABIO-RK: Integration and Curation of Reaction Kinetics Data

Ulrike Wittig, Martin Golebiewski, Renate Kania, Olga Krebs, Saqib Mir, Andreas Weidemann, Stefanie Anstein, Jasmin Saric, and Isabel Rojas

Scientific Databases and Visualization Group, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany
Ulrike.Wittig@eml-r.villa-bosch.de
http://sabiork.villa-bosch.de/

Abstract. Simulating networks of biochemical reactions require reliable kinetic data. In order to facilitate the access to such kinetic data we have developed SABIO-RK, a curated database with information about biochemical reactions and their kinetic properties. The data are manually extracted from literature and verified by curators, concerning standards, formats and controlled vocabularies. This process is supported by tools in a semi-automatic manner. SABIO-RK contains and merges information about reactions such as reactants and modifiers, organism, tissue and cellular location, as well as the kinetic properties of the reactions. The type of the kinetic mechanism, modes of inhibition or activation, and corresponding rate equations are presented together with their parameters and measured values, specifying the experimental conditions under which these were determined. Links to other databases enable the user to gather further information and to refer to the original publication. Information about reactions and their kinetic data can be exported to an SBML file, allowing users to employ the information as the basis for their simulation models.

1 Introduction

The biosciences have undergone some dramatic changes in the last few years. Novel lab approaches like high-throughput methods enable scientists to rapidly produce an enormous amount of data. For researchers this poses problems connected with retaining an overview of these data and accessing them. Thus one of the biggest challenges in biological science at present is to achieve data comparability and ease of access for the scientific community. To attain this goal, experimental data from different sources need to be standardized and integrated into databases.

At the moment only a small number of databases exist which contain information about biochemical reaction kinetics. The BRENDA enzyme database [1] offers a comprehensive list of kinetic parameters based on literature information. UniProt [2] started to include kinetic parameters as comments related to biophysicochemical properties, also manually extracted from publications. The BioModels database [3] stores published mathematical models of biological interest annotated and linked to relevant data resources (e.g. publications or databases). The models include kinetic

U. Leser, F. Naumann, and B. Eckman (Eds.): DILS 2006, LNBI 4075, pp. 94 – 103, 2006.

laws and their parameters represented in SBML (Systems Biology Mark-up Language) format [4] and can be used for simulations of biochemical reactions or networks.

In order to compare kinetic data and develop biochemical network models, kinetic parameters need to be consistently described and related to kinetic mechanisms, equations representing the kinetic laws and environmental conditions. The known mechanisms of biochemical reactions should be reflected in mathematical formulas, which have to be linked to the corresponding parameters, such as kinetic constants and concentrations of each reaction participant. As kinetic constants highly depend on environmental conditions, they only can be specified completely by describing these conditions used for determination. Data sets based on an experiment assayed under similar experimental conditions should be associated to each other to facilitate the comparison. Therefore, users interested in information about reaction kinetics require databases that merge and structure all these data.

The SABIO-RK (**S**ystem for the **A**nalysis of **Bio**chemical Pathways - **R**eaction **K**inetics) database is designed to meet these requirements and to support researchers interested in information about biochemical reactions and their kinetics. This report will mainly focus on the database content, integration and curation processes. Modelling of the database and the retrieval of data by database searching will be briefly discussed.

2 Data Integration

SABIO-RK represents an extension of the SABIO biochemical pathway database also developed at EML Research [5]. Figure 1 represents a simplified schema of the main database objects and their relations in SABIO and SABIO-RK. SABIO contains information about biochemical pathways, reactions and their participants (enzymes, reactants etc.). These data are connected with specific protein information, organisms or cellular locations. SABIO-RK combines the general data about biochemical reactions stored in SABIO with information about their kinetic properties. The type of kinetic law and its representation in a formula is given if provided in the literature. This also includes effectors (e.g. cofactors, activators or inhibitors) of the reactions and their type of interaction (e.g. competitive or non-competitive inhibition). The kinetic laws are represented with their parameters, including their measured values. Since many of the publications only contain kinetic constants (e.g. Km, kcat or Vmax) but have no description of the kinetic law type, these parameter values are also inserted independent from a kinetic law type.

Additionally the database contains descriptions of the experimental conditions (e.g. pH, temperature, and buffer) for the measured parameter values. In the buffer description all components of the assay are represented including coupled enzyme assays.

In order to establish a broad information basis, data from different sources are integrated into SABIO-RK (Figure 1). Most of the reactions, their associations with biochemical pathways and their enzymatic classifications (EC classifications [6]) are downloaded from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [7] and stored in SABIO. In contrast, the kinetic data contained in SABIO-RK are