Access to Reaction Kinetics Data: The SABIO-RK Database

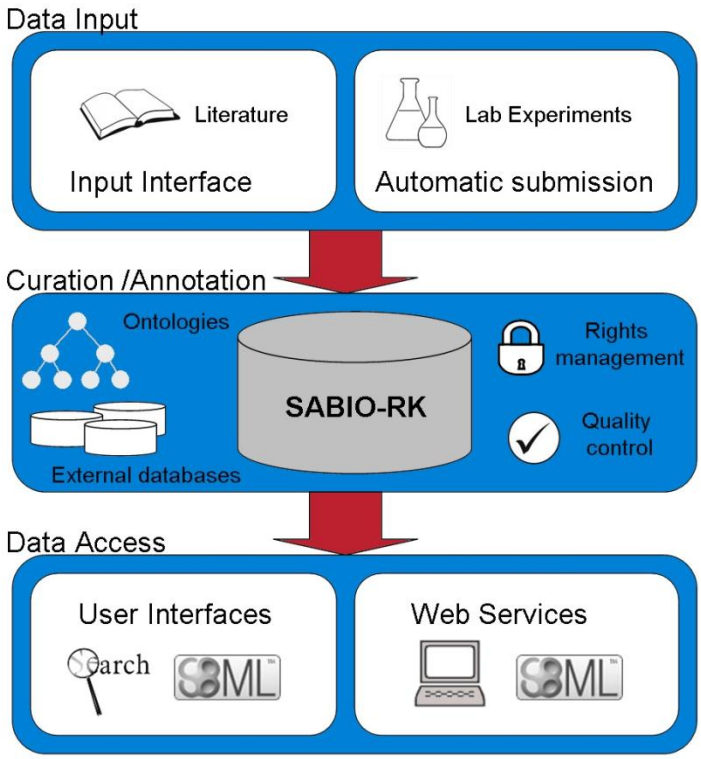
Martin Golebiewski, Ulrike Wittig, Renate Kania, Maja Rey, Enkhjargal Algaa, Meik Bittkowski, Lenneke Jong, Lei Shi, Andreas Weidemann, Elina Wetsch, Isabel Rojas, Wolfgang Müller

Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

Correspondence: martin.golebiewski@h-its.org, HITS, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg

SABIO-RK is a curated database that integrates and stores comprehensive information about biochemical reactions and their kinetic data [1]. It offers data manually extracted from the literature or directly submitted from lab experiments. The captured data is standardized by the use of controlled vocabularies and annotations pointing to other resources and biological ontologies [2]. From currently almost 40000 data entries, more than 7200 refer to reactions in different human tissues.

Reactions and their kinetics are described within their biological and experimental context. Each dataset is shown as single entity with all corresponding information (metadata) extracted from the original source. The kinetic parameters are described together with their corresponding rate equations, as well as kinetic law and parameter types and experimental and environmental conditions (pH, temperature, buffer) under which the data was determined. Detailed information about the biochemical reactions and pathways is also available, including substrates and products of the reaction, modifiers (inhibitors, activators, cofactors), biological location (organism, tissue, cellular location), as well as enzyme information (e.g. variant information, EC number, isoenzyme, UniProt ID, protein complex, molecular weight) for enzyme catalysed reactions.



SABIO-RK data population and access: Biochemical reaction kinetics data is either manually extracted from the literature or directly submitted by experimenters. The data is curated and supplemented with additional information from other databases and ontologies. Confidential data is hidden from the public and can be accessed by authorized users after login.

Literature-based information is inserted by students and biological experts [3] using a web-based input interface that consists of several web-pages with form fields and selection lists for structured data input and correct integration of the data. The same interface is also used for the data curation. It implements a variety of constraints from simple data format checks up to more complex consistency checks to avoid errors, inconsistencies and redundancy. Lists of compounds, reactions, organisms, tissues, cellular locations, kinetic law types, parameter types, and units already existing in the SABIO-RK database are provided as controlled vocabulary via selection lists or can be searched in the input interface. These controlled vocabularies together with annotations to external data resources and ontologies are used to identify and relate the data to the appropriate biological context.

Data from lab experiments can be directly submitted and incorporated in SABIO-RK using a submission interface that accepts data described in the XML-based SabioML format. Data received in

this format is directly inserted into the SABIO-RK database. This helps to automatize the submission process and provides a direct feed of kinetic data from the lab bench to the database, speeding up database population. Together with collaboration partners we have developed a tool for capturing, analysis and submission of data based on this interface [4].

The database can be accessed either manually via a web-based search interface or automatically via web services that allow direct data access by other tools. Both interfaces support the export of the data together with its annotations in SBML (Systems Biology Markup Language) complying with the MIRIAM (Minimal Information Requested In the Annotation of biochemical Models) standard. Recently, a second way of data export was implemented allowing the export of the data also in BioPax format. The web search interface allows simple full-text search in the database, as well as advanced search supported by a query builder. Hierarchical search based on the NCBI organism taxonomy facilitates the specific querying for data in related organisms, e.g. for all 'mammalia (NCBI)'. Likewise corresponding data referring to related tissues or cell types can be queried based on the Brenda Tissue Ontology (e.g. for all 'liver (BTO)' tissues and cell types). Additional filtering can be used to further confine the search results, e.g. for wildtype or mutant enzymes, specified rate equations, data sources, date of insertion, or particular environmental conditions like temperature and pH. In addition to the search interface we have developed web services, following a RESTful (Representational State Transfer) approach, for the programmatic access to SABIO-RK. This interface can be used by computer tools and databases to directly access the data.

SABIO-RK facilitates the exchange of kinetic data between experimentalists and modellers, and thereby supports the setup of quantitative computer models. The existence of web services and the degree of data integration make SABIO-RK well suited for its integration into workflows or applications using or requiring kinetic data, such as systems biology modelling platforms (e.g. CellDesigner, Virtual Cell or SYCAMORE).

Funding

SABIO-RK is supported by the Klaus Tschira Foundation (KTS), the German Federal Ministry of Education and Research (BMBF) through the Virtual Liver Network (VLN) and SysMO-LAB, and the DFG LIS “Integrated Immunoblot Environment”.

http://sabio.h-its.org/

References:

1. Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Algaa E, Weidemann A, Sauer-Danzwith H, Mir S, Krebs O, Bittkowski M, Wetsch E, Rojas I, Müller W (2012): Nucleic Acids Research 40(D1), D790-D796
2. Rojas I, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Wittig U (2007): In silico Biology 7, S37-44
3. Wittig U, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Anstein S, Saric J, Rojas I (2006): Lecture Notes in Bioinformatics 4075, 94-103
4. Swainston N, Golebiewski M, Messiha HL, Malys N, Kania R, Kengne S, Krebs O, Mir S, Sauer-Danzwith H, Smallbone K, Weidemann A, Wittig U, Kell DB, Mendes P, Müller W, Paton NW, Rojas I (2010): FEBS Journal 277, 3769-3779