Perspectives in Science (2014) 1, 33–40

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/pisc


REVIEW

# Data extraction for the reaction kinetics database SABIO-RK ☆

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

*Scientific Databases and Visualization Group, Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany*

Received 5 March 2013; accepted 4 November 2013; Available online 12 March 2014

**KEYWORDS**
Database;
Reaction kinetics;
Biocuration;
Ontology

**Abstract**
SABIO-RK (http://sabio.h-its.org/) is a web-accessible, manually curated database that has been established as a resource for biochemical reactions and their kinetic properties with a focus on supporting the computational modeling to create models of biochemical reaction networks. SABIO-RK data are mainly extracted from literature but also directly submitted from lab experiments. In most cases the information in the literature is distributed across the whole publication, insufficiently structured and often described without standard terminology. Therefore the manual extraction of knowledge from the literature requires biological experts to understand the paper and interpret the data. The database offers the literature data in a structured format including annotations to controlled vocabularies, ontologies and external databases which supports modellers, as well as experimentalists, in the very time consuming process of collecting information from different publications.
Here we describe the data extraction and curation efforts needed for SABIO-RK and give recommendations for publishing kinetic data in a complete and structured manner.


☆This article is part of a special issue entitled "Reporting Enzymology Data – STRENDA Recommendations and Beyond". Copyright by Beilstein-Institut.

*Corresponding author.
*E-mail address:* Ulrike.Wittig@h-its.org (U. Wittig).

http://dx.doi.org/10.1016/j.pisc.2014.02.004

**Contents**



## Introduction

During the year 2012 about one million scientific papers were published and entered into the literature database Pubmed (Sayers et al., 2011). In the era of bioinformatics and systems biology, it is important to provide intelligent data management systems to support researchers in the retrieval of information from the enormous amount of biological data published every year. In this context, online databases have become important media to afford scientists in accessing and reusing these data. At present 1512 different biological databases are listed in the Molecular Biology Database Collection and partially published in the 2013 database issue of the journal *Nucleic Acid Research* (Fernández-Suárez and Galperin, 2012). Most of these databases are mainly populated with data manually extracted from publications. The main challenge for these databases is to ensure a steady input of new data and to assure a high quality of the data. This requires that experts with biological knowledge have to invest time for data extraction and standardization.

Using SABIO-RK as an example for a biological database, we describe in this chapter the data extraction and curation process and the problems that curators have to overcome in their daily work. SABIO-RK (http://sabio.h-its.org/) (Wittig et al., 2012) is a web-accessible database containing comprehensive information about biochemical reactions and their kinetic properties. The database content includes kinetic data of biochemical reactions, kinetic rate laws and their equations, as well as experimental conditions and the corresponding biological sources. SABIO-RK is not restricted to any organism class and therefore offers all-encompassing organism data. All the data are manually curated and annotated by experts in biology. SABIO-RK can be accessed either via web-based user interfaces or automatically via web services that allow direct data access by other tools.

Although many life-science publications are electronically accessible, the way the information is usually presented is still traditionally scattered randomly across free text, tables and figures. Thus, manual data extraction from the literature is a very time-consuming. Several tools are available to support automatic information extraction (Hirschman et al., 2012) but, as described below in detail, the curation task for SABIO-RK is too complex to be tackled automatically by one of these tools at present. Data extraction for SABIO-RK requires the understanding of the whole paper and the transfer of the relations between the individual data into structured database elements. SABIO-RK database users are mainly biologists who use the data of biochemical reactions and their kinetics to build models of complex biochemical networks to run computer-assisted simulations. Literature search for the required information is a very cumbersome and time consuming task. SABIO-RK offers these data in a structured and standardized format and provides fast and convenient ways for data access.

## SABIO-RK database

SABIO-RK supports scientists in the modelling and understanding of complex biochemical networks by structuring kinetic data and related information from the literature. The focus of the database is a *reaction-oriented* representation of quantitative information on biochemical data. There exist other databases (e.g. BRENDA (Scheer et al., 2011), UniProtKB (The UniProt Consortium, 2011), BioModels (Le Novère et al., 2006), JWS Online (Olivier and Snoep, 2004)) that contain kinetic data, but the focus of these is different. SABIO-RK comprises all available kinetic parameters from a selected publication together with their corresponding rate equations, as well as kinetic laws and parameter types and environmental conditions (pH, temperature, and buffer) under which the kinetic data were measured. Biochemical reactions are defined by their reaction participants (substrates, products), modifiers (inhibitors, activators, cofactors), as well as detailed information about the proteins catalysing the reactions (e.g. EC enzyme classification, UniProtKB accession numbers, protein complex composition of the active enzyme, isozymes, wild-type/mutant information) and their biological source (organism, tissue/cell type, cell location).

A strong feature of the database is that not only standard biochemical reactions are provided but also alternative reactions with partly artificial substrates if they are used for the measurement. Therefore, only about 50% of the reactions in SABIO-RK match the original Kyoto Encyclopedia

of Genes and Genomes (KEGG) (Kanehisa et al., 2010)) reaction identifier. The same holds true for chemical compounds: about 30% of the SABIO-RK compounds are linked to the corresponding Chemical Entities of Biological Interest (ChEBI) (de Matos et al., 2010) identifier and more than 70% to the KEGG compound identifier. The additional storage of alternative reactions containing artificial substrates provides valuable information for the deduction of the enzymatic activity in vivo.

There are two sources for the kinetic data stored in SABIO-RK, scientific articles and wet-lab experiments. Literature-based data are inserted using a web-based, password-protected input interface (Rojas et al., 2007). Students or experts in biology first read the paper and insert the data in a temporary database via this input interface. The interface offers selection lists of controlled vocabularies and search functions for already available data in the database in order to facilitate correct data entries. Furthermore, constraints are implemented for both structuring and controlling the inserted data. To reduce errors and inconsistencies these constraints include data format checking and alignments with regard to the content entered before. After information extraction by student helpers, the same input interface is used by SABIO-RK database curators to validate inserted data and to align them to SABIO-RK data standards.

Data from wet-lab experiments can directly be submitted to SABIO-RK using a XML-based SabioML format (Swainston et al., 2010). It automates the submission process and directly populates the database with kinetic data from the laboratories. Since SABIO-RK always refers to the original source of kinetic data these lab experiment data are linked back to the raw data, like, for example, high-throughput kinetic assay results performed by collaboration partners in Manchester. Within collaboration projects unpublished data can be restricted for public access. Rights can be assigned to nested groups of scientists. During the manual curation process SABIO-RK data are annotated to ontologies, controlled vocabularies and external databases to avoid misinterpretations and to relate information to and exchange data with external sources. Biological ontologies and controlled vocabularies used in SABIO-RK are ChEBI, Systems Biology Ontology (SBO) (Le Novère, 2006), BRENDA Tissue Ontology (BTO) (Gremse et al., 2011), and National Center for Biotechnology Information (NCBI) organism taxonomy (Sayers et al., 2011). Based on these annotations links to other databases and ontologies are included enabling the user to obtain further details, for example about reactions, compounds, enzymes, proteins, tissues, or organisms.

Data access in SABIO-RK is available through web-based user interfaces and web-services by defining various search criteria. The newly developed and designed web interface offers different search functionalities including full text and advanced search, and beyond that filtering options to restrict the search results. Users can search for reactions and their kinetics by specifying the characteristics of the reactions. Complex queries can be created by specifying reactions defined by their participants (substrates, products, inhibitors, activators, etc.), pathways, enzymes, organisms, tissues or cellular locations, kinetic parameters, environmental conditions or literature sources. To improve and accelerate the database search for the user, the amount of kinetic data entries available in the database is displayed that match the search criteria while entering the search terms and formulating the queries. The list of results can be further sorted by different attributes in the *Entry View* or grouped by biochemical reactions in the *Reaction View*. Additionally a graphical representation of the search result composition in the *Visual Search* also offers the possibility to modify the query by further search criteria. The previous version of the search interface (called "classical") is yet accessible for users who are familiar with it and are interested to use it. The search criteria also comprise SABIO-RK internal identifiers and identifiers from external databases (e.g. UniProtKB, KEGG, ChEBI). For the specific search criteria *organism* and *tissue* different classification levels can be selected based on biological taxonomies or ontologies (Wittig et al., 2011). The search for organisms can be extended by the search for organism classes like the search for all mammals based on the NCBI taxonomy (e.g. search for "Mammalia (NCBI)"). Likewise, the tissue search includes the possibility to extend the search for related BRENDA Tissue Ontology terms to include all ontological children related to the search term (e.g. search for "liver (BTO)").

Web services are implemented using HTTP requests following a Representational State Transfer (REST) approach to allow an easy and direct access to SABIO-RK data (Shi et al., 2011; Richardson and Ruby, 2007). Other tools or databases use the web services in their processes to either link to SABIO-RK (e.g. KEGG, ChEBI) or to integrate SABIO-RK data in modelling platforms like CellDesigner (Funahashi et al., 2007), Virtual Cell (Moraru et al., 2008), or SYCAMORE (Weidemann et al., 2008). ChEBI compounds participating in reactions as substrates or products are linked to SABIO-RK reactions in the cross-references field "Reactions & Pathways". KEGG provides the links to SABIO-RK reactions from KEGG LIGAND reaction pages.

The web interfaces as well as the web services support the export and storage of the retrieved data in different file formats. Standardized and widely-used biological data exchange formats like Systems Biology Markup Language (SBML) (Hucka et al., 2003) or BioPAX/Systems Biology Pathway Exchange (SBPAX) (Ruebenacker et al., 2009) can be selected for data export and subsequent import in modelling tools. Additionally, simple table or text formatted export of data is offered. Kinetic data entry details and corresponding annotations to external databases and ontologies can be exported within SBML, compliant with the Minimum Information Required In the Annotation of Models (MIRIAM) standard (Le Novère et al., 2005). For tracking of the original data source SABIO-RK reaction and kinetic law identifier are themselves listed as MIRIAM data types.

## Data extraction and curation

During the process of data extraction from the literature, curators of the SABIO-RK database encounter issues such as including incomplete or inconsistent information within almost all publications. These data revision challenges are not specific for SABIO-RK but concern all other biological databases that are engaged in information extraction from the literature.

For further evaluation of this obstacle, we decided to examine a set of publications more systematically. As a starting point we selected randomly about 300 articles from the past 50 years which have already been used to extract SABIO-RK relevant data. We are aware that just 300 papers do not reflect the complete spectrum of all published papers from all journals. We make no claim to be complete but want to deliver some insights into the curators' daily work and use the results of the analysis to show problems during data extraction from the literature.

## Scattered data distribution within publications

Most publications of biological experimental data follow the classical rule of ordering the text in an *Introduction*, the description of *Material and Methods*, the experimental *Results* and a *Discussion* or *Summary* at the end. Typically the *Introduction* contains background knowledge and meta-data, e.g. the organism, description of the enzyme and the physiological biochemical reaction which is catalyzed by the enzyme in general. The *Material and Methods* section includes the explanation of the assay procedure and the experimental setup. In many cases the physiological biochemical reaction is not used for the measurement but alternative substrates are included in the experimental setup. The *Results* part describes in detail the measured and analyzed data which are frequently represented in tables and figures. Sometimes this section already contains the *Discussion* of the results which relates and compares the information to data from other experimentalists. The *Discussion* or *Summary* concludes and often repeats parts of the results. This classical paper structure results in a scattering of the relevant data in the paper: Figure 1 shows six pages of a selected full paper containing a color-coded representation of the distribution of different data within a publication. The colors are used to distinguish between different types of information (e.g. protein data, relevant experimental methods, or kinetic data). Figure 1 also represents the same data structured in an SABIO-RK database entry. The data described within the example publication results in 23 different entries in SABIO-RK, each entry having the same structure.

The segregation of related data within a paper makes automatic information extraction very difficult. Without understanding of the complete paper, it is almost impossible to collect and restructure the data in a correct way. Therefore the available tools for automatic information extraction are not suitable for the full extraction process. For example, if there is a description of a kinetic law equation used for the determination of kinetic parameters all values given in the equation should be extracted and inserted in the database entry. For the example paper in Figure 1 passages in the text containing kinetic parameters and data about the mathematical equation are highlighted

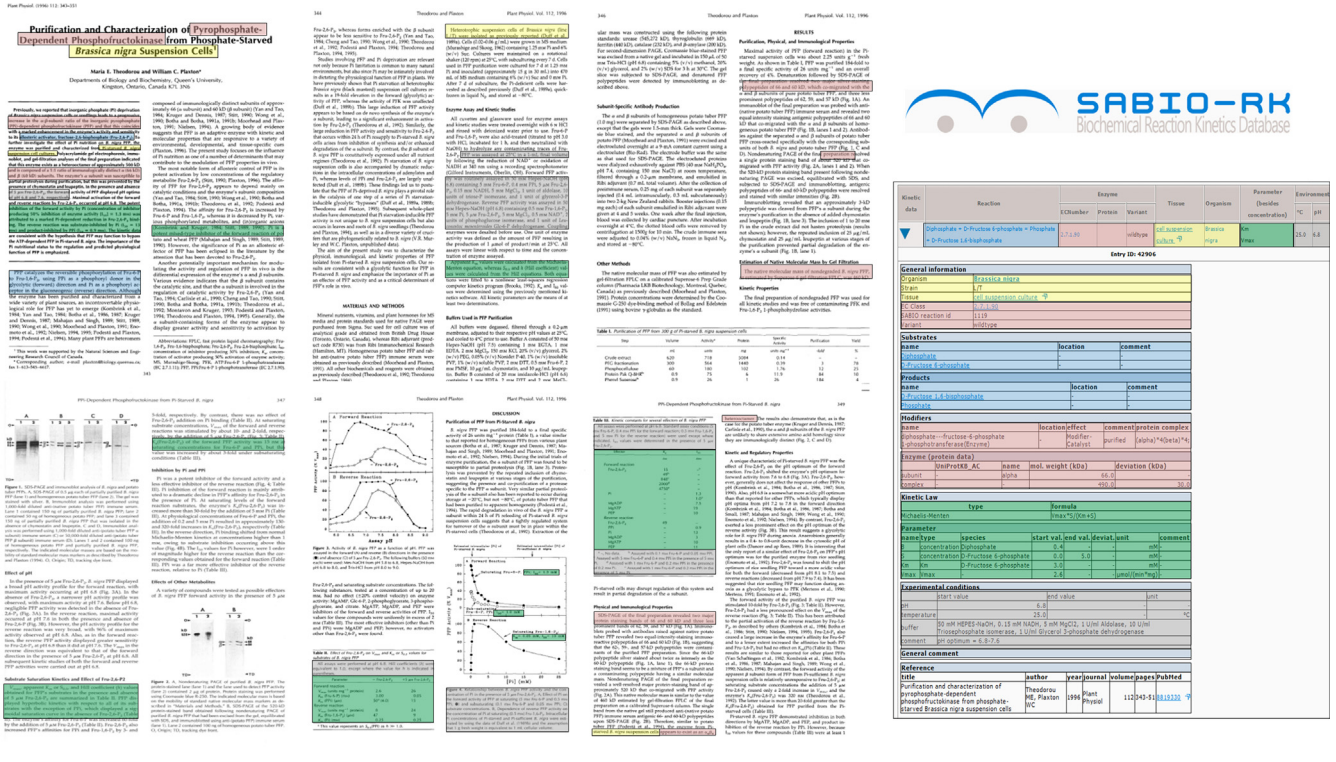


**Figure 1** Color-coded representation of data in a publication and in a SABIO-RK database entry.

in green showing that the data are distributed in the text and also written in tables and displayed in figures. This is a typical way of writing it in a paper.

Based on these findings we investigated the distribution and representation format of kinetic parameters within the above mentioned list of about 300 articles. Kinetic parameters (e.g. $K_m$, $K_i$, $k_{cat}$, $V_{max}$) which are important for the description of enzyme and reaction characteristics and comprise the key data of the SABIO-RK database can be found in three types of representations, in (i) free text, (ii) tables and (iii) figures. Such an inconsistent representation makes it hard to use or develop automatic information-extraction methods. Parameters are described in free text in 80% of the analyzed articles, displayed in tables in about 65% and in figures in about 8%. In 31.8% of the publications parameters are only within free text and in 18.2% only in tables. About 42% of the papers have parameters both in text and tables. In some of the papers, there are even conflicts between text and tables information.

The selected list of publications was also analyzed according to the distribution of the assay information and checked for different formats in which these data are represented in the publications. In more than 90% of the papers the assay conditions are described in free text, mainly within the *Material and Methods* section. But about 50% of the publications also represent assay conditions in the legends of tables or figures. And a similar amount includes compound concentrations as part of the assay conditions within figures so that concentrations have to be extracted from graph axes. In some cases there are conflicts between information written in the free text of the *Material and Methods* section and assay conditions represented in the legends of tables or figures. Within the set of analyzed articles we found two papers containing such conflicts. To solve these problems curators try to contact the authors where possible. Often the *Material and Methods* section contains a general description of the assay method and the legends contain more detailed or modified information about the experimental conditions for the measurement of the parameters displayed in the table or figure.

## Missing unique identifiers

One of our main interests in the paper analysis was the question how exact the entities (e.g. proteins, enzymes) can be identified within an article. The outcome was very surprising.

We know that some older papers have incomplete data due to the lack of the state of the art at the time. For example, a definite identification of isozymes is often missing in old publications because it was simply not known at that time point that different isozymes exist. In the 1980s three main data resources were available and evolved as standard repositories for nucleotides and proteins: the Protein Data Bank (PDB) (Berman, 2008), SwissProt/UniProtKB (The UniProt Consortium, 2011) and the International Nucleotide Sequence Database Collection (INSDC) comprised of the three databases DDBJ/EMBL/GenBank (Nakamura et al., 2013). Based on the availability of such standard protein and gene databases authors now have the possibility to exactly assign proteins to specific known isozymes by using database accession numbers. Additionally, starting in the 1990s, online repositories for ontologies and controlled vocabularies were developed to establish a universal standard terminology in biology e.g. Gene Ontology (The Gene Ontology Consortium, 2000) or NCBI organism taxonomy. A defined vocabulary is important to avoid misinterpretations and helps to exchange data between resources correctly. Ontologies and hierarchical classifications structure the data of a specific domain, describe the objects and define relationships between these objects. The usage of unique identifiers given by ontologies, controlled vocabularies and databases is essential for a definite data assignment.

Based on our past experience and problems with the correct assignment of entities during the data extraction process for SABIO-RK we were interested in quantitative information on the frequency of the usage of database accession numbers and identifiers of ontologies and controlled vocabularies in the randomly selected publications. Most of the big standard databases for genes and proteins were already developed and established as standard resources at the end of the 1980s. So we decided to start the analysis according to accession numbers with articles published in the mid-1990s. Figure 2 illustrates the fraction of database identifiers used in articles published in the given year. The number of analyzed papers per year is in the range between 10 and 20. Although this is just a starting point for a more comprehensive analysis of more publications, Figure 2 shows that there is no tendency for an increase of the usage of database identifiers dependent on the duration of database online availability. We expected an increase of protein or gene identifiers usage over the past years but this was not observed. In summary we conclude that exact names for proteins or genes are mainly used for description but no identifiers. Many times parts of sequences or sequence comparisons are represented in the paper but no corresponding gene or protein identifiers are displayed.

Data in SABIO-RK are linked to UniProtKB and accordingly to the IUBMB (International Union of Biochemistry and Molecular Biology, http://www.chem.qmul.ac.uk/iubmb/enzyme) and several enzyme databases via EC number. But about 25% of the analyzed articles of the time period between 1995 and 2009 neither contain any protein (Swis sProt/UniProtKB, PDB) or gene (DDBJ/EMBL/GenBank) iden tifier nor an EC number. The lack of the description of the entities with correct and unambiguous database identifiers may result in wrong assignments even for experienced database curators.

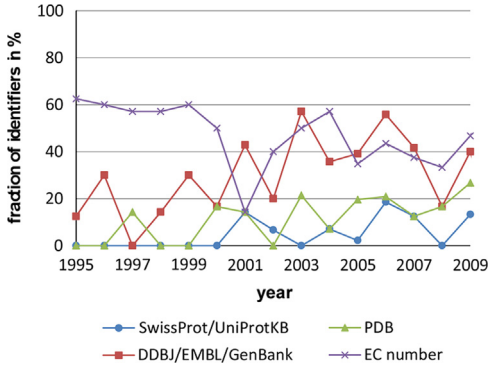


**Figure 2** Usage of database identifiers in publications.

Furthermore, 25% of the papers contain only an EC number for the enzyme classification but no additional protein or gene identifier. EC numbers were established in the 1960s and should be used as a standard enzyme annotation. But the rate of usage of EC numbers in publications is not increasing over time. Figure 2 illustrates that the assignment of EC numbers in the articles is on average only about 45%. Analyzed publications of the time period between 1961 and 1994 show 66% EC number assignment, which implies a more inattentive usage of these identifiers in newer articles. Authors always use enzyme names and maybe assume that the reader of the article knows or can deduce the EC number, especially for very well-studied enzymes like *pyruvate kinase*.

In the whole sample not a single paper contains any identifiers for organism, tissue, cellular process, protein function, cell location, reaction or compound. There are standard taxonomies, ontologies and controlled vocabularies as well as databases available which could be used for an exact identification of entities in the publication but data of the above categories are assigned not at all.

## Missing and inconsistent data

Since SABIO-RK stores information about reactions and their kinetic properties and in addition experimental conditions under which kinetic parameters were measured we also had a closer look at the correctness and completeness of the assay conditions because temperature, pH-value and the buffer composition are essential for the interpretation of experimental results. About 10% of the analyzed papers contain no information about the temperature used in the experiments. About 3% of the papers only give the imprecise information that the experiments were done at “room temperature”. In about 10% of publications the authors refer to another paper for the experimental method used for the measurement which causes a time-consuming search for the correct method in a reference paper. Sometimes the reference paper again refers to another paper for the method description. 20% of the publications describe the buffer composition and the compound concentrations not in standard units but use an indication of weight per assay volume which has to be manually converted to a standard unit.

A biochemical reaction is defined by the chemical compounds as reaction participants in particular substrates, products, enzymes and reaction modifiers like inhibitors and activators. About 25% of the publications used for insertion in SABIO-RK only contain incomplete reaction descriptions. For example in many cases the corresponding product for a substrate used in the experimental assay is missing. For data insertion in SABIO-RK biochemical reactions have to be complete containing all substrates and products. If the corresponding product information is missing in the publication SABIO-RK curators have to deduce the product (s) manually or if not possible include *Unknown* as compound.

## Comparison of data between publications

Frequently kinetic parameters in a paper were compared with values from other publications and were represented together in one table. Then the legend of the table or some phrases in the free text refers to the original source. Our analysis shows that there are no standard guidelines for authors how to refer to referenced values. The challenge for data extraction is here to filter the reference values from the original paper values. In SABIO-RK the parameter values are always only linked to the original source.

## Recommendations

The examples for the challenges of correct data extraction from the literature as mentioned above illustrate that a large amount of manual work by experts in biology is still needed. Natural language processing tools for automatic data extraction and text understanding are far away from being suitable for our application.

Ideally journal editors should ask the authors for complete, standardized and structured data in their future articles. Collaborations between the publisher and the database site to develop common standards and data format are preferable. First steps are done in this direction by the Standards for Reporting Enzymology Data (STRENDA, http://www.strenda.org) Initiative (Tipton et al., 2014) which created recommendation for the publication of enzyme data including minimum information for the description of enzymes and related data. These STRENDA recommendations are already accepted by some biological journals and inserted in the author’s guidelines of these journals. Within the biocuration community which was recently enforced by the foundation of the International Society for Biocuration (http://biocurator.org) there are also initiatives to improve the collaboration between database curators and publishers. The adaption of publica tions to the needs of the database developers will increase the quality and re-usability of published data. The hope from the database curators’ point of view for future papers would be, for example, the consistent usage of identifiers from standard databases, ontologies and controlled voca bularies for a correct identification of entities of interest. Of course, this would only hold for future publications. The extraction of data from already existing papers will be still a big challenge, including time-consuming manual curation work. Currently there are no software tools to automati cally support the identification of missing or inconsistent data. Another challenge for the extraction of data for a reaction kinetics database like SABIO-RK is the spreading of data through the whole text of the publication. In addition, different formats for the representation of data within the paper (e.g. kinetic parameters in tables, figures or text) are difficult to handle with automatic extraction methods.

To follow up our findings we are planning to start a more comprehensive analysis of publications. In addition, we are considering the labeling of the part of information in the database that was missing from the publication, but has been investigated and added manually by the curators.

## Summary

We have described the biochemical reaction kinetics database SABIO-RK and the data extraction and curation process

used to maintain it. SABIO-RK is a manually curated database containing biochemical reactions and their kinetic properties. The database is established as a data resource for both experimentalists and modellers. Data in SABIO-RK are mainly extracted manually from the literature and stored in a structured and standardized format. The database content comprises the relevant data which are essential to describe the characteristics of biochemical reactions, the corresponding biological source, kinetic properties and experimental conditions. Annotations to controlled vocabularies, ontologies, and external databases allow the comparison and exchange of data. For a high quality data in a database the original source should be comprehensive and complete. Based on our experience, and confirmed by our analysis of a set of SABIO-RK relevant publications, we suggest improvement opportunities for publishing experimental data. We would recommend guidelines for authors, reviewers and publishers to improve the re-usability of their papers by checking for missing data, missing identifiers, inconsistent information within the paper e.g. causing conflicts between data in text and tables, usage of standard formats and names, and defined usage of referenced values and experimental methods.

## Conflict of interest statement

None of the authors have any conflict of interest.

## Acknowledgements

The SABIO-RK project is financed by the Klaus Tschira Foundation (http://www.klaus-tschira-stiftung.de/), the German Federal Ministry of Education and Research (http://www.bmbf.de/) through Virtual Liver and SysMO-LAB (Systems Biology of Microorganisms), and the DFG LIS (http://www.dfg.de/) as part of the project *Integrierte Immunoblot Umgebung*.

## References

Berman, H.M., 2008. The protein data bank: a historical perspective. Acta Crystallogr. A 64, 88–95.

de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S., Steinbeck, C., 2010. Chemical entities of biological interest: an update. Nucleic Acids Res. 38, D249–D254.

Fernández-Suárez, X.M., Galperin, M.Y., 2012. The 2013 nucleic acids research database issue and the online molecular biology database collection. Nucleic Acids Res. 41 (Database issue), D1–D7.

Funahashi, A., Jouraku, A., Matsuoka, Y., Kitano, H., 2007. Integration of cell designer and SABIO-RK. In Silico Biol. 7, S81–S90.

Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Ebeling, C., Schomburg, D., 2011. The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucleic Acids Res. 39, D507–D513.

Hirschman, L., Burns, G.A., Krallinger, M., Arighi, C., Cohen, K.B., Valencia, A., Wu, C.H., Chatr-Aryamontri, Dowell, K.G., Huala, E., Lourenço, A., Nash, R., Veuthey, A.L., Wiegers, T., Winter, A.G., 2012. Text mining for the biocuration workflow. Database 2012: bas020.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., et al., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531.

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.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., et al., 2005. Minimum Information Required In the Annotation of Models (MIRIAM). Nat. Biotechnol. 23, 1509–1515.

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 Res. 34, D689–D691.

Le Novère, N., 2006. Model storage, exchange and integration. BMC Neurosci. 7 (Suppl. 1), S11.

Moraru, I.I., Schaff, J.C., Slepchenko, B.M., Blinov, M.L., Morgan, F., Lakshminarayana, A., Gao, F., Li, Y., Loew, L.M., 2008. Virtual cell modelling and simulation software environment. IET Syst. Biol. 2 (5), 352–362.

Nakamura, Y., Cochrane, G., Karsch-Mizrachi, I., 2013. The international nucleotide sequence database collaboration. Nucleic Acids Res. 41, D21–D24.

Olivier, B.G., Snoep, J.L., 2004. Web-based kinetic modelling using JWS Online. Bioinformatics 20, 2143–2144.

Richardson, L., Ruby, S., 2007. Restful web services, first ed. O'Reilly Media, USA.

Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Wittig, U., 2007. System for the Analysis of Biochemical Pathways Reaction Kinetics (SABIO-RK). In: Proceedings of the 2nd International Symposium on "Experimental Standard Conditions of Enzyme Characterizations", Ruedesheim am Rhein, Germany.

Ruebenacker, O., Moraru, I.I., Schaff, J.C., Blinov, M.L., 2009. Integrating BioPAX pathway knowledge with SBML. IET Syst. Biol. Models 3 (5), 317–328.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., et al., 2011. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 39, D38–D51.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676.

Shi, L., Jong, L., Mueller, W., Algaa, E., 2011. 50x Faster: Speeding up an SQL-based legacy system with few changes. In: Lecture Notes in Informatics (LNI) – Proceedings, Series of the Gesellschaft für Informatik (GI), pp. 192.

Swainston, N., Golebiewski, M., Messiha, M.L., Malys, N., Kania, R., Kengne, S., Krebs, O., Mir, S., Sauer-Danzwith, H., Smallbone, K., et al., 2010. Enzyme kinetics informatics: from instrument to browser. FEBS J. 277 (18), 3769–3779.

The Gene Ontology Consortium, 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25 (1), 25–29.

The UniProt Consortium, 2011. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 39, D214–D219.

Keith, Tipton, Richard N. Armstrong, Barbara, Bakker, Amos, Bairoch, Athel, Cornish-Bowden, Peter, Halling, Jan-Hendrik, Hofmeyr, Thomas S. Leyh, Carsten, Kettner, Frank M. Raushel, Johann, Rohwer, Dietmar, Schomburg, Christoph, Steinbeck, 2014. Standards for Reporting Enzyme Data: the STRENDA Consortium: What it Aims to Do and Why it Should be Helpful. Perspectives in Science 1, 110–120.

Weidemann, A., Richter, S., Stein, M., Sahle, S., Gauges, R., Gabdoulline, R., Surovtsova, I., Semmelrock, N., Besson, B., Rojas, I., Wade, R., Kummer, U., 2008. SYCAMORE – a systems biology computational analysis and modeling research environment. Bioinformatics 24, 1463–1464.

Wittig, U., Algaa, E., Weidemann, A., Kania, R., Rey, M., Golebiewski, M., Shi, L., Jong, L., Müller, W., 2011. Ontology-based search in SABIO-RK. In: Proceedings of the 5th International Symposium on "Experimental Standard Conditions of Enzyme Characterizations", Ruedesheim am Rhein, Germany.

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. SABIO-RK – database for biochemical reaction kinetics. Nucleic Acids Res. 40 (D1), D790–D796.