

**Workshop "Storage and Annotation of Reaction Kinetics Data"**  
May 2007, Heidelberg, Germany

## Storing and annotating of kinetic data

**Isabel Rojas, Martin Golebiewski, Renate Kania, Olga Krebs, Saqib Mir,  
Andreas Weidemann and Ulrike Wittig**

Scientific Databases and Visualization Group, EML Research gGmbH, Heidelberg, Germany

\* Corresponding author  
Email: [isabel.rojas@eml-r.villa-bosch.de](mailto:isabel.rojas@eml-r.villa-bosch.de)

Edited by I. Rojas and U. Wittig (guest editors); received and accepted March 21, 2007; published March 29, 2007

---

### Abstract

This paper briefly describes the SABIO-RK database model for the storage of reaction kinetics information and the guidelines followed within the SABIO-RK project to annotate the kinetic data. Such annotations support the definition of cross links to other related databases and augment the semantics of the data stored in the database.

**Keywords:** reaction kinetics, database, systems biology, biochemical networks, kinetic law

---

### Introduction

Setting up mathematical models for the simulation of biochemical networks is a complex task that requires, apart from mathematical and biochemical knowledge, the associated data describing the dynamics of each of the reactions participating. For this reason, information such as the kinetic law defining the rate of a reaction together with its respective parameters and the experimental conditions under which they were determined, are very valuable to modellers. Thus, the success of systems biology projects depends heavily on the exchange of information from experimenters to modellers and *vice versa*. However, while general molecular and pathway information have been compiled in several databases there is currently little experience with databases collecting reaction kinetics data.

Most of the knowledge available about reaction kinetics is found in scientific papers published in journals world wide. Finding the relevant papers and extracting information from them is a very

time-consuming work, which is generally duplicated by different research groups in their everyday work.

Apart from the problem of finding the information, researchers are confronted with the problem of integrating it. Experimenters and model developers hardly ever use non-ambiguous controlled vocabularies to describe the studied objects or the conditions and constraints of their experiments or models. This makes the integration and comparison of the data a very difficult task. Thus, when presenting and exporting data it is very important that we use clearly defined terms, using standards and annotate with references to common vocabularies and ontologies. This allows a better understanding, exchange and integration of the data.

There are currently a small number of databases which contain information about biochemical reaction kinetics. Amongst these we can find: the BRENDA enzyme database [1], that offers a list of kinetic parameters associated with enzyme kinetics; [Kinetikon](http://kinetikon.molgen.mpg.de/) (<http://kinetikon.molgen.mpg.de/>) providing detailed knowledge about biochemical reaction kinetics, however, limited mainly to Yeast; the KDBI database (Kinetic Data of Biomolecular Interactions) [2] which contains a collection of experimentally determined kinetic data of binding or interaction events described in the literature, like protein-protein, protein-RNA, and RNA-ligand. Most of the existing databases containing information about reaction kinetics do not offer mathematical equations describing the reaction rate, connected to their respective parameters and environmental conditions under which these were determined. This connection between mathematical equations, the corresponding parameters and the conditions or constraints is mandatory for the construction of biochemical models. In terms of annotation to external ontologies and controlled vocabularies, this varies from database to database. Most offer links to compound and reaction databases like KEGG (Kyoto Encyclopedia of Genes and Genomes) [3], however, annotations to some very relevant ontologies, such as the System Biology Ontology (SBO, <http://www.ebi.ac.uk/sbo/>), are absent in all cases.

Apart from the databases mentioned above, there are also databases storing complete published models of biochemical reaction networks. These mostly offer mathematical descriptions along with their parameters. However, these data are dependent on the constraints of the models and cannot be regarded as independent experimental data. Examples of such databases are BioModels [4], storing published models of biochemical reactions annotated and linked to relevant data resources (e. g. publications or databases); and JWS [5], a repository of models of biochemical reactions. Both of these contain models of both metabolic as well as signalling networks. In addition, there is DOQCS (Database of Quantitative Cellular Signalling) [6], a database storing models of signalling pathways. In all three cases, there is hardly any information available about the original experimental data for single reactions used to create the models. Since lots of the parameter values are estimations based on related data, it is hard to figure out the underlying experimental data. Furthermore, rarely the constraints for which the model holds or the environmental conditions assumed can be found. Thus, although these databases are useful sources of information for modellers, to uncure, combine and re-use the contained kinetic data of single reactions is very cumbersome and difficult.

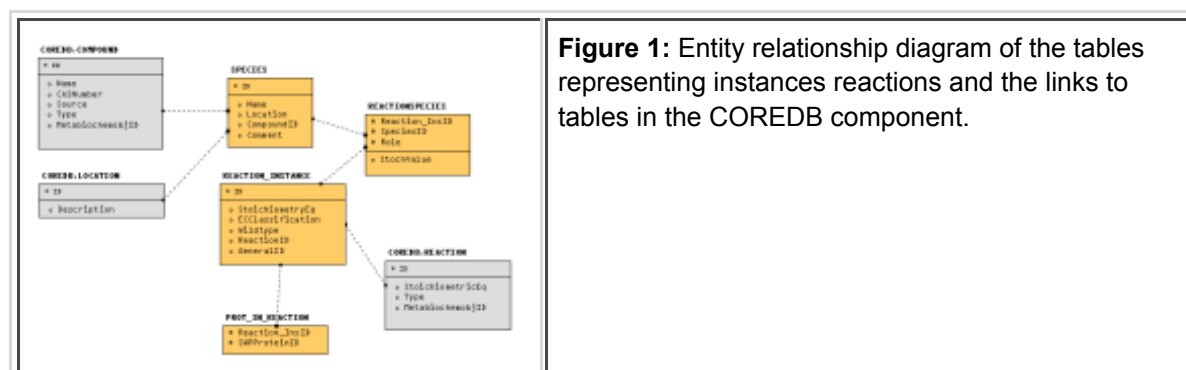
SABIO-RK [7] is a web-accessible curated database offering information about biochemical reactions and their kinetic properties. It integrates information about reactions, such as reactants, effectors, and catalyzing enzymes, with information about organisms, tissues and cellular locations where the reactions take place, and with the kinetic properties of these reactions (type of the kinetic mechanism, modes of inhibition or activation and rate equations together with their parameters and measured values). As kinetic constants highly depend on environmental conditions used for their determination these are given together with the kinetic parameters and the description of the kinetic mechanism. This also facilitates the comparison of data sets based on experiments assayed under similar experimental conditions. The current version of the SABIO-RK web interface allows users to perform

searches for reactions and their corresponding kinetic data by specifying characteristics of the reactions of interest (such as reactants, enzymes or pathways) as well as of the kinetic data searched (e. g. from a particular tissue, determined under certain experimental conditions or only certain parameter types). Data about biochemical reactions and their kinetic parameters (with their respective rate equations) can be exported in SBML (Systems Biology Mark-up Language) [8] file format, allowing its import into simulation and modelling programs supporting SBML.

In this paper we will briefly describe the database model used by SABIO-RK to store the kinetic information. We will also discuss the annotations that have been added to the database entries to support cross links to other related databases and to augment the semantics of the data stored in the database.

## SABIO-RK schema for the storage of kinetic data

The SABIO-RK database is composed of two tightly integrated schemas: one representing the basic (core) data relevant to biochemical reactions and pathways (COREDB), and the other representing the kinetic data (KINDATA). We will mainly refer to tables in KINDATA and whenever referring to tables in COREDB denote it so. SABIO-RK centres on the concept of *biochemical reaction* (COREDB.REACTION) as formed by its elementary **participants**, i. e. substrates and products. This concept encompasses different types of biochemical reaction such as: metabolic, signalling, regulatory and transport. However, the storage of kinetic data for a biochemical reaction is done at the level of instances of a reaction (REACTION\_INSTANCE, see Fig. 1). An instance refers to a reaction (or even probably a particular direction of a reaction) taking place in a particular biochemical context, and reported in a particular article or experiment. Thus, multiple - even reverse - reactions in the REACTION\_INSTANCE table can refer to the same general reaction in the COREDB.REACTION table.



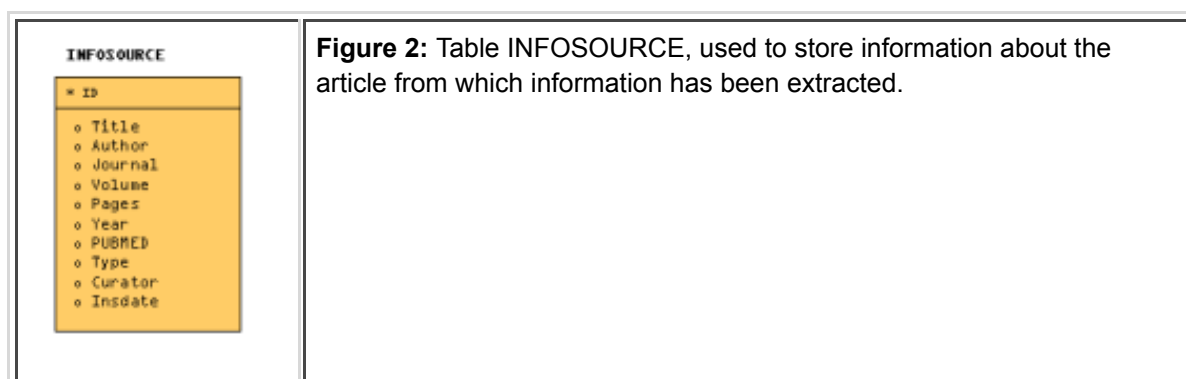
**Figure 1:** Entity relationship diagram of the tables representing instances reactions and the links to tables in the COREDB component.

The **species** (SPECIES) of a reaction instance are represented in the table REACTIONSPECIES. A species corresponds to a **compound** in a given **location**. The species of a reaction instance will not only include substrates and products, but the **enzyme** and other **modifiers** (activators, inhibitors or cofactors) of the reaction instance. The direction of the reaction instance is specified by the assignment of roles (substrates and products) to its species.

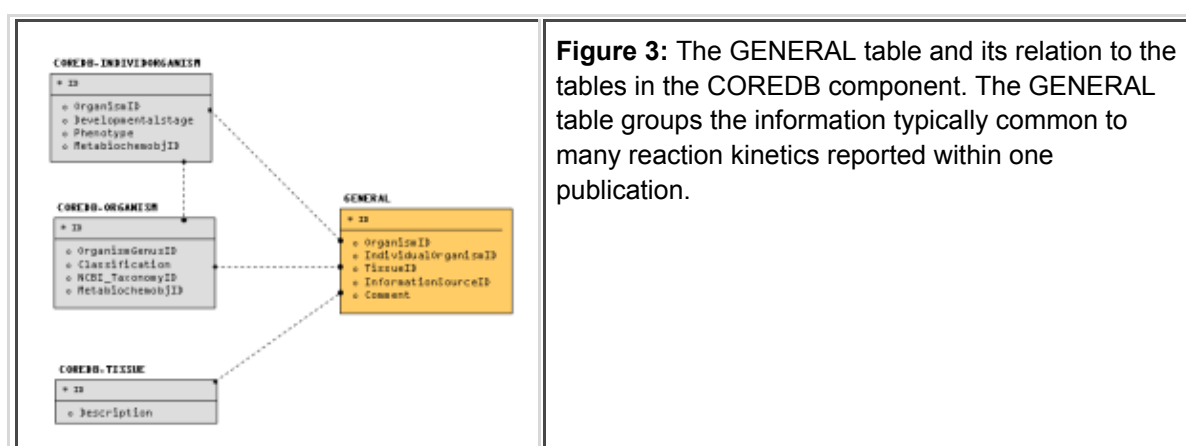
For metabolic reactions we store the enzyme classification (EC) number associated with the reaction instance for which the kinetic data was measured/estimated, together with detailed information on the type of the protein having the enzymatic activity (wildtype, mutant, isozyme information etc. in field

wildtype) and – when available – the composition of the protein represented by one or more UniProtIDs (*PROT\_IN\_REACTION*). Additionally, in the COREDB schema the relation REACTION-ECClassID is represented in a table, allowing the association of one reaction with many EC classifications.

The kinetic information included in the SABIO-RK database is currently mainly coming from literature. The information about the literature **source** is contained in the table *INFOSOURCE* (see Fig. 2). In this table we also document who was the curator that worked on the referred paper and the insertion date.

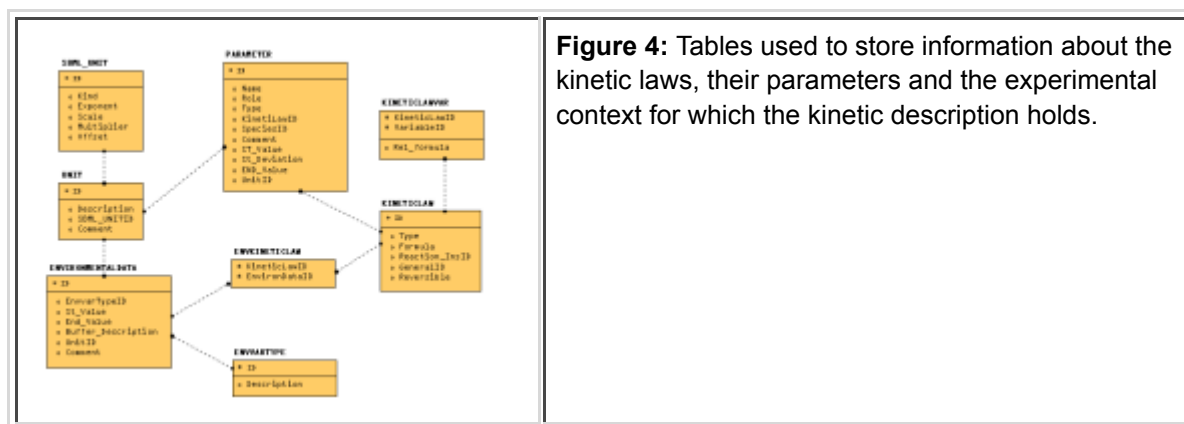


An article can report on kinetic data of reactions measured under different organisms, strains of the organism, or tissues. However, these tend to be common parameters for several measurements/estimations reported in a paper. Thus, we group this information in a so called GENERAL table (see Fig. 3), allowing a paper to have more than one entry in this table. An entry in the GENERAL table refers to descriptions of the **organisms**, their strains (**individual organisms**) and **tissues** in the corresponding tables of the COREDB schema. The descriptions of the particular instances of a reaction and their respective kinetics refer to an entry in the GENERAL table, describing parts of the biochemical context.



For the description of a **kinetic law** of a reaction instance (represented in the table *KINLAW*) we provide fields for the storage of the kinetic law's formula (mathematical formula), the kinetic law's type and for the specification if the law is only for one direction (from substrate to product) or reversible.

A kinetic law may have many associated **parameters** (PARAMETER) determined under particular environmental conditions (relation EnvironmentKineticLaw). For each parameter we store its name, role (constant or variable), type ( $V_{max}$ ,  $K_m$ ,  $k_{cat}$ , etc.), the range of its values (in the form start to end, or start +/- deviation), with its corresponding unit. If the parameter is associated to a species this is given by means of a link to the SPECIES table. The units in the UNITS table find a corresponding entry in the SBML\_UNIT table, which contains the description of the units in the SBML unit definition format. The description of the environmental (experimental) conditions under which the kinetics of a reaction was determined is described by entries in the table ENVIRONMENTDATA. It is possible to store information about multiple conditions (types) describing the environment; additionally there is the possibility to describe the buffer used (Fig. 4).



## Annotations and usage of controlled vocabularies

In order to facilitate the interpretation and integration of the data stored in SABIO-RK, we employ controlled vocabularies and notation standards for several attributes. Some of these standards we develop ourselves, e. g. to differentiate between kinetic law types or parameter types, whereas some others reflect already established controlled vocabularies, as for example organism taxonomy or descriptions of tissues and cellular locations. When new data is manually entered in SABIO-RK, expressions and controlled vocabularies already stored are offered by selective lists to assure the internal consistency of the data. This procedure eventually helps users in finding related data.

SABIO-RK not only uses controlled vocabularies, but also contains synonymic notations for chemical compounds and enzymes in order to unambiguously identify these entities. This is important to relate database entries referring to identical molecules, since a compound or enzyme can have many different names. In order to generate lists of synonyms, we integrate synonymic names from other resources like KEGG, and supplement and curate them manually. To support the identification and comparison of synonymic notations for chemical compounds, we are developing linguistic tools for the normalization of compound names and bioinformatics tools for deriving the chemical structure of a compound from its name.

In order to appropriately export the data stored in SABIO-RK to for example other programs like simulation tools, global standards for the description of entities and notations are applied. These include annotations of entities and notations to external databases (e. g. UniProt/Swiss-Prot protein database [9] and KEGG) as well as to domain ontologies like SBO and ChEBI (Chemical Entities of Biological Interest [10]). Such annotations assist users (or programs accessing the data) in identifying the entities and expressions stored in SABIO-RK, enable links to external resources, and also

facilitate the embedding of SABIO-RK data into workflows involving reaction kinetics data. For instance, metabolic reactions in SABIO-RK are annotated to one or many enzymatic activity classes (EC numbers) as defined by the enzyme nomenclature of the International Union of Biochemistry and Molecular Biology ([IUBMB](http://www.chem.qmul.ac.uk/iubmb/enzyme/)) (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). This EC number can be used to query several enzyme databases or even Protein databases like UniProt.

Currently, SABIO-RK supports the annotation of following entities to the termed external resource identifiers:

- Chemical compounds to KEGG compound IDs, KEGG drug IDs and ChEBI-IDs
- Protein complexes (assigned to an enzymatic activity) to UniProt/Swiss-Prot-IDs
- Reactions to KEGG reaction IDs
- EC numbers to query enzyme databases like [ExpASY](http://www.expasy.org/enzyme/) (Expert Protein Analysis System)-Enzyme ( <http://www.expasy.org/enzyme/>), [KEGG](http://www.genome.ad.jp/kegg/) enzyme (<http://www.genome.ad.jp/kegg/>), [IntEnz](http://www.ebi.ac.uk/intenz/) (Integrated relational Enzyme database) (<http://www.ebi.ac.uk/intenz/>), [IUBMB](http://www.chem.qmul.ac.uk/iubmb/enzyme/) enzyme nomenclature or [Reactome](http://www.reactome.org/) (<http://www.reactome.org/>)
- Information sources (publications) to [NCBI PubMed](http://www.ncbi.nlm.nih.gov/entrez/) IDs (<http://www.ncbi.nlm.nih.gov/entrez/>)
- Organisms to [NCBI Taxonomy](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy) (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy>)
- Types of kinetic laws (rate equations), kinetic parameter types (e. g.  $K_m$ ,  $k_{cat}$ , rate constants, etc.) and roles of species in reactions to SBO terms

The development of SABIO-RK contributes to the development of the SBO by suggesting the inclusion of new entries for categories like kinetic parameter types or reaction mechanism types. The close cooperation between the development of SBO and SABIO-RK increases the quality of the annotations and thus the understanding and exchange of the data.

Annotations in the SBML file are being included following the MIRIAM (Minimum information requested in the annotation of biochemical models) [11] standard. Although the current version of the SBML standard (level 2, version 2) contemplates the identification of elements with SBO terms, the SBML currently generated by SABIO-RK omits the inclusion of SBO terms given that this modifications are still not available in the libSBML library (but should be included soon), and thus it makes it very cumbersome to add the SBO attributes as part of the element tags. The structure of SABIO-RK is independent from the SBML schema; there is much information in the database that cannot be yet included into the SBML file. However, we hope to contribute to the development of SBML so that in the future most of the information can be represented in SBML. Currently, the only point where the SBML structure directly affects the content of the database is in the definition of the SBML units, given that we have an additional table to store the representation of units in SBML format.

---

## Conclusions and future work

In this paper we have briefly introduced the database model of SABIO-RK for the storage of kinetic data of biochemical reactions. We also presented a description of the annotations that we perform over the data in order to facilitate the exchange, integration and understanding of the data stored in and exported from the SABIO-RK database. In the brief description of the model we have left out some recent extensions done to support the storage of the reactions' mechanisms, expressed in terms of elementary reactions, and the kinetics of the elementary reactions. This is still in test phase, and we start populating the database with this data very soon.

Currently SABIO-RK mainly contains kinetic data about metabolic reactions. However, we are already working on the extensions to the database and its interface in order to efficiently support the storage and querying of kinetic data for signalling and regulation reactions. The incorporation of new type of data will also result in the incorporation of new annotation types to the data.

Other extensions planned for the database involve the support for the storage of typical *in-vivo* concentrations for metabolites and for the storage of experiment or simulation results expressed as time series.

---

## Acknowledgements

We would like to thank the Klaus Tschira Foundation as well as the German Research Council (BMBF) for their funding. We would also like to thank the members of the Bioinformatics and Computational Biochemistry and the Molecular and Cellular Modelling Groups of EML Research for their helpful discussions and comments. Last but not least, we thank all the student helpers, who have contributed to the population of the database.

---

## References

1. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* **32**, D431-D433.
2. Ji, Z. L., Chen, X., Zhen, C. J., Yao, L. X., Han, L. Y., Yeo, W. K., Chung, P. C., Puy, H. S., Tay, Y. T., Muhammad, A. and Chen, Y. Z. (2003). KDBI: Kinetic Data of Bio-molecular Interactions database. *Nucleic Acids Res.* **31**, 255-257.
3. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**, D354-D357.
4. Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J. L. and 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.
5. Olivier, B. G. and Snoep, J. L. (2004). Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**, 2143-2144.
6. Sivakumaran, S., Hariharaputran, S., Mishra, J. and Bhalla, U. S. (2003). The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **19**, 408-415.
7. Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J. and Rojas, I. (2006). SABIO-RK: Integration and Curation of Reaction Kinetics Data. *In*: proceedings of the 3rd International workshop on Data Integration in the Life Sciences 2006 (DILS'06). Hinxton, UK. *Lecture Notes in Bioinformatics* **4075**, 94-103.
8. 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., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novère, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J. and Wang, J. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524-531.
- 
9. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003). The Swiss-Prot Protein Knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365-370.
- 
10. de Matos, P., Ennis, M., Darsow, M., Guedj, M., Degtyarenko, K. and Apweiler, R. (2006). ChEBI - Chemical Entities of Biological Interest. *Nucleic Acids Res. Database Summary Paper* 646.
- 
11. Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J. L., Spence, H. D. and Wanner, B. L. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509-1515.