

INTEGRATION AND ANNOTATION OF KINETIC DATA OF BIOCHEMICAL REACTIONS IN SABIO-RK

ULRIKE WITTIG^{*}, RENATE KANIA, MARTIN GOLEBIEWSKI,
OLGA KREBS, SAQIB MIR, ANDREAS WEIDEMANN,
HENRIETTE ENGELKEN AND ISABEL ROJAS

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

E-Mail: *ulrike.wittig@eml-r.villa-bosch.de

Received: 30th April 2008 / Published: 20th August 2008

ABSTRACT

SABIO-RK is a curated database for the systems biology community containing biochemical reactions and their kinetic properties, the latter being manually extracted from literature sources. This information is crucial for the quantitative understanding of biological systems. Modellers and wet-lab scientists alike require reliable information about reaction kinetics which is normally contained in publications generated worldwide. In SABIO-RK kinetic data are related to reactions, organisms and biological locations. The type of kinetic mechanism and corresponding rate equations are presented together with their parameters and experimental conditions. In order to enable comprehensive understanding, integration and comparison of data it is necessary to provide annotations and links to community resources, such as external databases and ontologies that augment the content and the semantics of the SABIO-RK database entries. In this short paper we will present SABIO-RK (<http://sabio.villa-bosch.de/SABIORK>) and our approach towards integration and annotation of the kinetic data and their respective biochemical context.

INTRODUCTION

For the understanding of complex biochemical processes, the simulation and modelling of biochemical reactions and complex networks require reliable kinetic information about the individual reaction steps within the process. Information such as the kinetic laws describing the dynamics of the reactions with their respective parameters determined under certain experimental conditions is of paramount importance. These kinetic data are mainly found in the literature and described in many different formats. Within the publications no controlled vocabulary is used for the representation of biochemical reactions, kinetic data and environmental conditions. The integration of these data needs the definition and use of standards for reporting and exchanging.

SABIO-RK extends and supplements the information content of other databases containing kinetic information (e.g. BRENDA [1], BioModels [2], and JWS Online [3]) by storing highly interrelated information about biochemical reactions and their kinetics. It includes reactants and modifying compounds (enzymes, cofactors, inhibitors or activators) of reactions. The kinetic laws with their parameters and information about experimental conditions are connected with the reaction information. The data about biochemical reactions, their rate equations and parameters can be exported in SBML (Systems Biology Markup Language) [4] file format.

The STRENDA [5] (Standards for Reporting Enzymology Data) commission is working on the definition of a standard for reporting on enzyme activity. The standard should contain the minimum amount of information that should accompany any published enzyme activity data. The use of references to controlled vocabularies and ontologies is also of great importance for the implementation of the STRENDA guidelines.

In this paper we describe new implementations in SABIO-RK and in the input interface to match the defined requirements of STRENDA.

DATA INTEGRATION

The data contained in SABIO-RK are extracted from different sources, in order to establish a broad information basis. Most of the reactions, their associations with biochemical pathways and their enzymes are automatically extracted from the KEGG database [6]. Information about chemical compounds is extended additionally by data from ChEBI [7] or PubChem [8].

Kinetic data in SABIO-RK are mainly extracted from literature and inserted manually using a web-based input interface to enter the data into a temporary database. The temporary database is used by biological experts to curate the data and to insert them into the final SABIO-RK database. The main objective of the input interface is to supply a uniform format

for users entering and curating the data found in publications. Apart from this, automatic control mechanisms have been implemented to check for errors and inconsistencies during the integration process. The systems check, for example: whether all reaction participants of a reaction equation have to be defined as compound species; if all parameters in a rate equation have to be defined in the parameter list independent of whether or not the parameter value is known; if the rate equations are mathematically correct; and which (if applicable) parameter type should be related to a compound species (for example, a K_m value always should be related to a compound, however V_{max} or k_{cat} do not need a species assignment).

To avoid errors and inconsistencies within the SABIO-RK database during the first input process, the interface offers lists of controlled vocabularies for the selection of values for the following data:

- Species (compounds search function)
- Species roles (e. g. substrate, product, modifier...)
- Biochemical reactions (search functions)
- Organisms, tissues and cellular locations
- Kinetic law types (e. g. Competitive inhibition or Sequential ordered Bi Bi)
- Parameter types (e. g. K_m , k_{cat} , V_{max} , K_i , K_d , rate constant)
- Parameter units (e. g. mM, μ M, 1/s, nmol/min)

Based on information in the literature, detailed information about the protein catalysing the reaction is inserted. This includes information about a specific isoenzyme or mutations used in the experiments (e.g. wildtype, mutant E540K, wildtype isoenzyme A), UniProt [9] accession numbers and information about the composition of subunits (e.g. (Q6UG02)*4 for a homotetramer).

The data format defined in the input interface fits most of the current requirements of the STREND A commission for reporting enzymology data. So SABIO-RK offers the structure for inserting and storing data required for a complete description of an experiment (Level 1, List A) which comprise, for example, information about the enzyme (EC number, biological location, isoenzyme information etc.) and the assay conditions. The description of enzyme activity data (Level 1, List B) includes kinetic parameters necessary for enzyme function definition (V_{max} , k_{cat} , K_m etc.), information about enzyme inhibition or activation and the description of the kinetic mechanisms including rate equations.

Another possible way to insert kinetic data into SABIO-RK is by using an XML-based integration tool to import a higher amount of kinetic data automatically.

DATA ANNOTATION

Scientific communication needs standards and a shared vocabulary to avoid misinterpretations. Such standards are especially important when gathering data from different sources. To identify entities or terms unambiguously and to facilitate the search, interpretation and comparison of data, these are standardized to a uniform format and structure. This implies the usage and development of controlled vocabularies. Entities and expressions in SABIO-RK are annotated to other resources and biological ontologies to clarify the biological terms used and to embed the data into its context. This enables users to collect further information through links to external databases and facilitates the integration of different database entries into kinetic models.

Biological ontologies used in SABIO-RK are ChEBI, Gene Ontology (GO) [10], Systems Biology Ontology (SBO) [11], and NCBI taxonomy [12]. ChEBI is a dictionary and ontological classification of small chemical compounds. Gene Ontology comprises controlled vocabularies for molecular functions, biological processes and cellular components of gene products. Systems Biology Ontology defines controlled vocabularies for systems biology, especially in the context of computational modelling. NCBI taxonomy is a controlled vocabulary and complex classification system of organisms. These controlled vocabularies and ontologies are the basis for the definition of the selection lists for compounds, organisms, kinetic law types, parameter types etc. in the SABIO-RK input interface.

Links to external databases in the user interface (Fig. 1) allow the user to connect to other data sources to get further information about the selected entry. Annotations are shown for chemical compounds to KEGG, ChEBI and PubChem. Enzymes are linked through their EC number to ExPASy ENZYME [13], KEGG, IntEnz [14], IUBMB [15] and Reactome [16]. Further links to KEGG are implemented for the reaction equation and the information source is referenced to PubMed [17]. Proteins or protein complexes catalysing the reactions are linked to the specific UniProt accession number(s). These annotations of proteins or protein complexes to UniProt are done manually by biological experts based on information in the publication.

Integration and Annotation of Kinetic Data of Biochemical Reactions in SABIO-RK

Entry Nr. 10580
[⊕][⊖]
Select

Organism:	Homo sapiens
Tissue:	liver
EC Class: 3.5.3.1	Variant: wildtype

Substrates

name	location	comment
L-Arginine	cytosol	-
H2O	cytosol	-

Products

name	location	comment
Urea	cytosol	-
L-Ornithine	cytosol	-

Modifiers

name	location	effect	comment	External References
Co2+	cytosol	Modifier-Activator	-	-
Arginase(Enzyme)	unknown	Modifier-Catalyst	-	Protein: (P05089)*3;

Kinetic Law

$V_{max} * S / (K_m + S)$

Kinetic Law Type: Michaelis-Menten

Parameters

name	species	type	St_value	Deviation	End_value	unit	comment
S	L-Arginine	concentration	0.002	-	0.02	M	-
Vmax		Vmax		-			-
C	Co2+	concentration	2	-		mM	-
Km	L-Arginine	Km	4.4	-		mM	-

Experimental conditions

	St_value	end_value	unit
Temperature	37	-	-
pH	8.7	-	-

Buffer: 50 mM Tris-HCl

General comments: MW: 120 +/-2 kDa; subunit: 34 kDa

PUBMEDID: [7584844](#)

Figure 1. Database entry containing links to external databases and controlled vocabularies.

Many annotations to external databases and controlled vocabularies defined in the user interface are also stored in the SMBL file for exporting the information (Fig. 2).

```

</species>
- <species metaid="meta_spc_3" id="spc_3" name="Urea" compartment="compart_1">
- <annotation>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_spc_3">
- <bqbiol:is>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/#CHEBI:16199" />
- <rdf:li rdf:resource="http://www.genome.jp/kegg/compound/#C00086" />
- </rdf:Bag>
- </bqbiol:is>
- </rdf:Description>
- </rdf:RDF>
- </annotation>
</species>
- <species metaid="meta_spc_4" id="spc_4" name="L-Ornithine" compartment="compart_1">
- <annotation>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_spc_4">
- <bqbiol:is>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/#CHEBI:15729" />
- <rdf:li rdf:resource="http://www.genome.jp/kegg/compound/#C00077" />
- </rdf:Bag>
- </bqbiol:is>
- </rdf:Description>
- </rdf:RDF>
- </annotation>
</species>
- <species metaid="meta_enz_1" id="enz_1" name="Arginase(Enzyme)" compartment="compart_1" />
</listOfSpecies>
- <listOfReactions>
- <reaction metaid="meta_reac_0" id="reac_0">
- <annotation>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_reac_0">
- <bqbiol:isVersionOf>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.ec-code.org/#3.5.3.1" />
- </rdf:Bag>
- </bqbiol:isVersionOf>
- <bqbiol:is>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.genome.jp/kegg/reaction/#R00551" />
- </rdf:Bag>
- </bqbiol:is>
- </rdf:Description>
- </rdf:RDF>
- </annotation>
</reaction>
</listOfReactions>

```

Figure 2. SBML file containing annotations to external databases and controlled vocabularies.

FUTURE DIRECTIONS

In order to offer the users more references to additional information we are working on the cross-linking and annotation of the database content to more database resources and ontologies.

One of the next steps for further SABIO-RK developments is the incorporation of detailed information about the mechanisms of the reaction to allow the user to obtain information about the kinetic properties of sub-reactions or binding mechanisms of enzymes and substrates. This includes the visualization of reaction mechanism information in the user-interface.

In the future the input interface will be further adapted to the current and changing requirements of the STRENDA commission which could comprise, for example, more detailed and structured information about experimental conditions and the methodology of measurement, purification etc. SABIO-RK will participate in further discussions about STRENDA requirements and work on their implementation. SABIO-RK could serve as the basis for a general data input and storage system for experimentalists and modellers.

SUMMARY

SABIO-RK is a curated data resource for modellers of biochemical networks to assemble information about reactions and their kinetics. It also offers experimentalists the opportunity to obtain information about biochemical reactions and their kinetics, within the context of cellular locations, tissues and organisms. The data are extracted automatically from different databases and kinetic information is manually extracted from literature. The database uses controlled vocabularies and links to other ontologies or external databases to allow the comparison of data and to extract additional information from other sources.

ACKNOWLEDGEMENTS

We would like to thank the Klaus Tschira Foundation as well as the German Research Council (BMBF) for their funding. 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., Schomburg, D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* **32**:D431-D433.
 - [2] 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.
 - [3] Olivier, B.G., Snoep, J.L. (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**:2143–2144.
 - [4] 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., Wang, J. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**:524–531.
 - [5] STRENDA: <http://www.strenda.org>
-

- [6] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**:D354–D735.
- [7] ChEBI: <http://www.ebi.ac.uk/chebi/>
- [8] PubChem: <http://www.ncbi.nlm.nih.gov/sites/entrez? db=pccompound>
- [9] The UniProt Consortium. The Universal Protein Resource (UniProt). (2008) *Nucleic Acids Res.* **36**:D190–D195.
- [10] Gene Ontology: <http://www.geneontology.org/>
- [11] Systems Biology Ontology: <http://www.ebi.ac.uk/sbo/>
- [12] NCBI taxonomy: <http://www.ncbi.nlm.nih.gov/sites/entrez? db=taxonomy>
- [13] ExPASy ENZYME: <http://www.expasy.org/enzyme/>
- [14] IntEnz: <http://www.ebi.ac.uk/intenz/>
- [15] IUBMB: <http://www.chem.qmul.ac.uk/iubmb/enzyme/>
- [16] Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E., Stein, L. (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**:Database Issue:D428–D432.
- [17] PubMed: <http://www.pubmed.gov>
-