Beilstein-Institut

ESCEC, March 19th – 23rd, 2006, Rüdesheim/Rhein, Germany

Adding Semantics in Kinetics Models of Biochemical Pathways

Nicolas Le Novere, Melanie Courtot, Camille Laibe

EMBL-EBI, Wellcome-Trust Genome Campus, Hinxton, CB10 1SD
United-Kingdom

E-Mail: lenov@ebi.ac.uk

Received: 24th August 2006 / Published: 31st August 2007

Abstract

The need to exchange and integrate models drove the community to design common data format such as SBML. However, as important as was the definition of a common syntax, we also need to tackle the semantics of the models. The community recently proposed MIRIAM, the Minimal Information Requested in the Annotation of Models, a set of rules for curating quantitative models of biological systems. This standard lists the condition an encoded model has to meet to fully correspond to its reference description, and describe how to annotate each of its components. The Systems Biology Ontology (SBO) aims to strictly index, define and relate terms used in quantitative modelling, and by extension quantitative biochemistry. SBO is currently made up of five different vocabularies: quantitative parameters, participant roles, modelling frameworks, mathematical expressions – that refers to the three previous branches – and events. SBO can be used not only to annotate quantitative models, but also biochemical experiements. It is expected that the adoption of those two semantic layers will favour the reusability of quantitative biochemical descriptions, whether parameters or models.

Introduction

Until very recent times, life science coped extremely well with fuzzy semantics, and that for very sound operational reasons. Except for hard-core biochemists and pharmacologists, concepts like *Kd*, *Kp*, EC_{50} etc. were globally known as inversely proportional to the "affinity", and it was sufficient to have a sensible discussion about biological processes. The Molecular Biology era reinforced even further that trend, by shifting the general interest from quantitative to qualitative description of physiology. Moreover, when quantitation nevertheless sneaked in, the results were quite often expressed as relative rather than absolute values. Typically it was the measurement in a mutant or upon a perturbation, normalised over the same measurement in a wild-type animal, or in unperturbed condition. With the notable exception of the sequence and structure of macromolecules, biologists were not really supposed to store and exchange experimental results, but rather to build on the interpretation of data provided by the authors themselves, who were judged the best suited to analyse it.

Then entered Functional Genomics. The large genome projects had demonstrated that it was possible to produce experimental data on a large scale, with a quality control that far surpassed the standards of isolated academic group [1]. New technologies now promised to allow a similar large-scale generation for more functional types of data. However, those technologies were complex and very expensive. As a consequence, the workforce and degree of expertise needed to put them in application meant that single research groups could not longer run their own data production. And the costs of experiments also meant that each dataset should be produced in a limited number of replicates. Biologists finally had to store and exchange raw data. Despite scale of the datasets, continuous improvements of computing power offered biologists efficient tools to perform the necessary archival/retrieval procedures.

The second shock came from the rise of Systems Biology, which increased the general awareness not only to modelling and simulation of biological processes, but also to quantitative biology in general. As a consequence, what was once the territory of a small population of specialists is now visited by various actors of biomedical research. In parallel, the formal models used in biology are growing, both in size and complexity. A given modeller is therefore less likely to be an expert of all the corners of a quantitative model, whether the biological knowledge or even the mathematical approaches.

This need for quantitative data of high quality calls for a shift of paradigm in the way experimental parameters are exchanged and re-used, but also theoretical concepts handled. There is no point to exchanging quantitative data or models if nobody can understand the meaning of the data and the content of the models beside their initial generators. The community has to define agreed-upon standards for kinetic data generation and curation, so that the experimental measurement can be safely reuse. Controlled vocabularies, where concepts are related one to the other, must be designed for annotating quantitative models

with connections to biological data resource. Finally, one needs to integrate modelling work with the other sources of knowledge, and disseminate the large number of models produced.

To offer a possible answer to those issues, the BioModels.net initiative was launched in 2004 by Michael Hucka, Andrew Finney and the author. BioModels.net (http://biomodels.net) is an international effort to (1) define community standards for model curation, (2) design controlled vocabularies for annotating models with connections to biological data resources, and (3) provide a free, centralised, publicly-accessible database of annotated, computational models in SBML and other structured formats. In this paper we will describe two resources belonging to the first and second classes. The third objective has been tackled with BioModels Database[2] (http://www.ebi.ac.uk/biomodels/)

Minimal Information Requested in the Annotation of Biochemical Models

Searching for existing models relevant to a specific problem, a scientist comes across a model named *Model1*, describing the reactions *rA* and *rB* between the molecular components *X* and *Y*.
What can we make of this model? Where does this model come from? What are exactly the components *X* and *Y* in molecular or cellular terms? It could help a lot to know what biological process is modelled by *rA* and *rB*. Providing one finally elucidates the origin of the model, and the identity of its components, how can we know that when instantiated, this model will provide the correct numerical results?

The aim of MIRIAM [3] is to define processes and schemes that will increase the confidence in model collections and enable the assembly of model collections of high quality. The first part of the guidelines is a standard for reference correspondence, dealing with the syntax and semantics of the model. A second part is an annotation scheme, that specifies the documentation of the model by external knowledge. The scheme for annotation can itself be further subdivided into two sections. The *attribution* covers the minimum information that is required to associate the model with a reference description and an actual encoding process. The *external data resources* covers information required to relate the components of quantitative models to established data resources or controlled vocabularies.

The aim of standard for reference correspondence is to ensure that the model is properly associated with a reference description and is consistent with that reference description. The reference description can be a scientific article, but also any other unique publication, on print or online, that describes precisely the structure of the models, list the quantitative parameters, and described the expected output. In order to be declared MIRIAM-compliant, a quantitative model must fulfil the following rules:

1. The model must be encoded in a public, standardised, machine-readable format such as (but not restricted to) SBML[4] or CellML[5].
2. The model must be clearly related to a single reference description. If a model is derived from several initial reference descriptions, there must still be a unique reference description that references a set of results that one can expect to reproduce when simulating the derived/combined model.
3. The encoded model structure must reflect the biological processes listed in the reference description (this correspondence is not necessarily one-to-one).
4. Quantitative attributes of the model, such as initial conditions and parameters, as well as kinetic expressions for all reactions, have to be defined, in order to allow to instantiate simulations.
5. The model, when instantiated within a suitable simulation environment, must be able to reproduce all results given in the reference description that can readily be simulated.

In order to be confident in re-using an encoded model, one should be able to trace its origin, and the people who were involved in its inception. The following information should always be joined with an encoded model:

- A citation of the reference description with which the model is associated, either as a complete bibliographic record, or as a unique identifier, Digital Object Identifier (http://www.doi.org), PubMed identifier (http://www.pubmed.gov), unambiguous URL pointing to the description itself etc.
- Name and contact information for the creators, that is the persons who actually contributed to the encoding of the model in its present form.
- The date and time of creation, and the date and time of last modification.
- A precise statement about the terms of distribution. The statement can be anything from “freely distributable” to “confidential”. MIRIAM being intended to allow models to be communicated better, terms of distribution are essential for that purpose.

The aim of the external data resources annotation scheme is to link the components of a model to corresponding structures in existing and future open access bioinformatics resources. Such data resources can be, for instance, database or ontologies. This will permit not only the identification of model components and the comparison of components between different models, but also the search for models containing specific components.

This annotation must permit to unambiguously relate a piece of knowledge to a model component. The referenced information should be described using a triplet {“data-type”, “identifier”, “qualifier”}.

- The "data-type" is a unique, controlled, description of the type of data, written as a Unique Resource Identifier. The URIs can be expressed as a Uniform Resource Locator, e. g. http://www.uniprot.org/ or a Uniform Resource Name, e. g. urn:lsid:uniprot.org. They have no physical meaning, and if expressed as an URL, does not have to correspond to an existing website. For instance the URI representing the enzyme classification is http://www.ec-code.org/.
- The "identifier", within the context of the "data-type", points to a specific piece of knowledge. For instance 2.7.11.17 for a calcium/calmodulin regulated protein kinase.
- The optional "qualifier" is a string that serves to refine the relation between the referenced piece of knowledge and the described constituent. Although MIRIAM standard does not impose any restriction on the use of qualifiers, biomodels.net nevertheless provides predefined qualifiers, described below.

Such a triplet can easely be exported later using the Resource Description Framework (http://www.w3.org/RDF/), to ease further automatic treatment. RDF is at the core of what is called "Semantic Web" (http://www.w3.org/2001/sw/), and one of the basic technologies that enables modern data interoperability in life science [9]
The following qualifiers are examples that can be used to characterize model components:

[is] The modelling object represented by the model component is the subject of the referenced resource. For instance, this qualifier might be used to link the encoded model to a database of models.

[isDescribedBy] The modelling object represented by the component of the encoded model is described by the referenced resource. This relation might be used to link a model or a kinetic law to the literature that describes this model or this kinetic law.

The following qualifiers are examples that can be used to characterize the biological entity represented by model components.

[is] The biological entity represented by the model component is the subject of the referenced resource. This relation might be used to link a reaction to its exact counterpart in KEGG or Reactome for instance.

[hasPart] The biological entity represented by the model component includes the subject of the referenced resource, either physically or logically. This relation might be used to link a complex to the description of its components.

[isPartOf] The biological entity represented by the model component is a physical or logical part of the subject of the referenced resource. This relation might be used to link a component to the description of the complex is belongs to.

[isVersionOf] The biological entity represented by the model component is a version or an instance of the subject of the referenced resource.

[hasVersion] The subject of the referenced resource is a version or an instance of the biological entity represented by the model component.

[isHomologTo] The biological entity represented by the model component is homolog, to the subject of the referenced resource, i.e. they share a common ancestor.

[isDescribedBy] The biological entity represented by the model component is described by the referenced resource. This relation should be used for instance to link a species or a parameter to the literature that describes the concentration of the species or the value of the parameter.

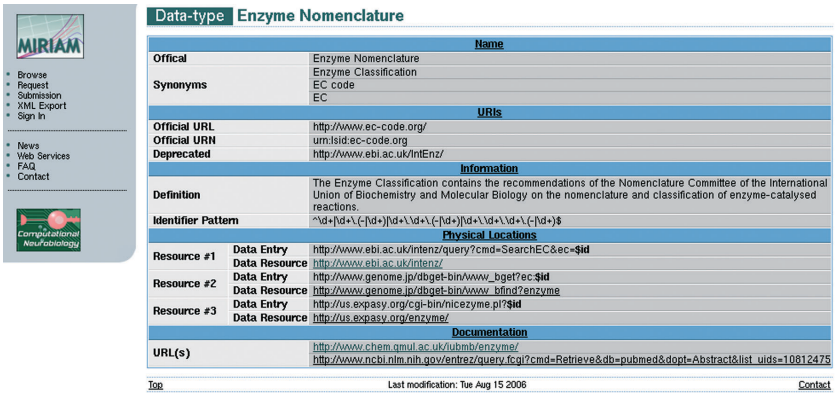

Figure 1: Example of the entry in MIRIAM database describing the Enzyme Classification.

To enable interoperability, a set of standard valid URIs has to be maintained, and tool provided to automatically retrieve valid URL(s) corresponding to a given URI. This is the purpose of MIRIAM Database and the associated Web Services (http://www.ebi.ac.uk/compneur-srv/miriam/).

MIRIAM is maintained at the EBI using an open relational database management system (MySQL) and a web application using a free implementation of Java Server Page (servlet container Apache Tomcat http://tomcat.apache.org/). Each entry of the database contains a diverse set of details about a given data-type: official name and synonyms, the URIs (URL and/or URN forms), patterns of identifiers, links to documentation documentation etc. In addition, each data-type can be associated with several physical locations. An example is shown on figure.

Users are able to perform queries such as retrieving valid physical locations (URLs) corresponding to a given URI (physical location of a generic data-type or of a precise piece of knowledge), retrieving all the information stored about a data-type (such as its name, its synonyms, links to some documentation etc.). Moreover, a programmatic access through Web Services http://www.w3.org/2002/ws/ (based on Apache Axis (http://ws.apache.org/axis/) and SOAP messages [10]) allows a software not only to resolved model annotations, but also to generate the correct URIs based on resource name and accession numbers.

For instance, a software generating content needs to annotate an enzymatic activity: The query getURI("enzyme nomenclature", "2.7.11.17") returns the result http://www.ec-code.org/#2.7.11.17. Conversely, an interface to a database of enzymatic activity needs to generate an hyperlink: The query `getDataEntries(\)http://www.ec-code.org/#2.7.11.17\))` returns the result (at the time of redaction of this chapter) `http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=2.7.11.17`, `http://www.genome.jp/dbget-bin/www_bget?ec:2.7.11.17`,`http://us.expasy.org/cgi-bin/nicezyme.pl?2.7.11.17`.

Dozens of other methods are available to interact with the database and make the most of annotations.

Systems Biology Ontology

Whilst many controlled vocabularies exist that can directly be used to relate quantitative models to biological knowledge, there were no classification of the concepts themselves used in quantitative modelling. BioModels.net partners recognized that several additional small controlled vocabularies were required to enable the systematic capture of information in those models started to develop their own ontology.

The word ontology is defined here in its information science meaning, as a hierarchical structuring of knowledge. In our case, it is a set of relational vocabularies, that is a set of terms linked together. Each term has a definition and a unique identifier. Those ontologies have seen their role in structuring our knowledge growing steadily in life science over the last few years. The most famous ontology in life-science is Gene Ontology (GO) [11]. They have actually been used by life scientists for a while, also not recognised as such. The most obvious examples are the various taxonomies, of organisms, of sequences families or of protein domains. Less apparent is the fact that other biochemical knowledge management frameworks, such as the Enzyme Classification, also fulfill many of the criteria necessary to qualify as an ontology.

One of the goals of the Systems Biology Ontology (SBO, http://www.ebi.ac.uk/sbo/) is to facilitate the immediate identification of the relation between a model component and the model structure. SBO is currently made up of five different vocabularies (Figure 2). Within a vocabulary, the terms are related by "is a" inheritances, which represent sub-classing.

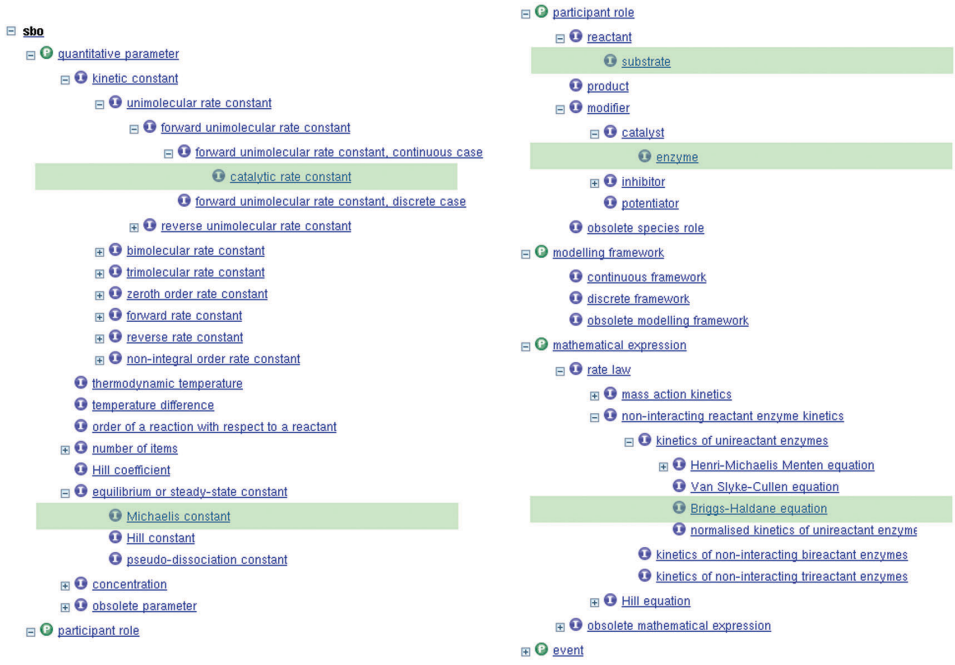

Figure 2: Partially unfolded view of SBO tree. Highlighted are the terms we use in the example described in the text.

1. A controlled vocabulary for parameter roles in quantitative models. This CV includes terms such as "forward unimolecular rate constant", "Hill coefficient", "Michaelis constant" etc.
2. A taxonomy of the roles of reaction participants, including the following terms: "catalyst",
3. "substrate", "competitive inhibitor" etc.
4. A list of modelling framework, that precises how to interpret a mathematical expression, such as "deterministic", "stochastic", "boolean" etc.
5. A classification of mathematical expression used in biochemical modelling. In particular this controlled vocabulary contains a taxonomy of kinetic rate equations. Examples of terms are "mass action kinetic", "Henri-Michaelis-Menten kinetics", "Hill equation" etc.

6. A branch containing the classification of events represented by biochemical models, such as “binding”, “transport” or “degradation”.

Each SBO term is made up of a stable identifier, a name, a definition, synonyms, a list of parentages, comments, and, for the mathematical expression branch, an equation. The identifier is a unique string that is never deleted once it is created. If a term needs to be suppressed, it is made child of the “obsolete” branch of the corresponding vocabulary. The name is unique in the ontology, but can change over time. The parentages are of two types, a subclassing (or subsumption or hyponymy) “Is A”, and a dissection (or meronymy) “Part Of”. Contrarily to other ontologies such as Gene Ontology, the latter is used only to link direct children of the root (the five vocabularies).

As an example, the term describing Briggs-Haldane kinetics is described on figure 3.

The annotation of quantitative model components with SBO terms will be an essential step to reach MIRIAM-compliance. Such an annotation will add the layer of semantics necessary to link mathematical representations of biochemical models encoded in SBML or CellML with graphical notations such as the Systems Biology Graphical Notation (http://www.sbgn.org/), or semantically enriched computing formats to represent biochemical knowledge such as BioPAX [12] (http://www.biopax.org). SBO will enhance our capacity to understand and to programmatically analyse models. Finally, SBO will also power the search strategies used by the databases of models and kinetics. In the following we present some examples of SBO use.

SBO to Discriminate between Implicit Hypothesis

The conversion between a continuous and a discrete modelling framework sometimes requires the transformation of a unique complex rate-equation into the description of several elementary reactions. The complex rate-equation has been generally derived using hypothesis that most often are not explicit from the equation itself. As an example, let's consider the case of a simple irreversible unireactant enzyme catalysis. The transformation of a substrate *S* into a product *P*
by an enzyme *E* as been formalised by Victor Henri in 1903 [13] and later by Leonor Michaelis and Maud Menten in 1913 [14] as following the kinetic law:

$$v = [E] \times \frac{k_{cat} \times [S]}{K_S + [S]}$$

[Term]
id: SBO:0000031
name: Briggs-Haldane equation

def: ”Rate-law presented in ”G.E. Briggs and J.B.S. Haldane (1925) A note on the kinetics of enzyme action, *Biochem. J.*, **19**: 339 – 339”. It is a general rate equation that does not require the restriction of equilibrium of Henri-Michaelis-Menten or irreversible reactions of Van Slyke, but instead make the hypothesis that the complex enzyme-substrate is in quasi-steady-state. Although of the same form than the Henri-Michaelis-Menten equation, it is semantically different since Km now represents a psudoequilibrium constant, and is equal to the ratio between the rate of consumption of the complex (sum of dissociation of substrate and generation of product) and the association rate of the enzyme and the substrate.
mathml:

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <semantics definitionURL="http://biomodels.net/SBO/\#SBO:0000062">
    <lambda>
      <bvar><ci definitionURL="http://biomodels.net/SBO/\#SBO:0000025">kcat</ci></bvar>
      <bvar><ci definitionURL="http://biomodels.net/SBO/\#SBO:0000014">Et</ci></bvar>
      <bvar><ci definitionURL="http://biomodels.net/SBO/\#SBO:0000015">S</ci></bvar>
      <bvar><ci definitionURL="http://biomodels.net/SBO/\#SBO:0000027">Km</ci></bvar>
      <apply>
        <divide/>
        <apply>
          <times/>
          <ci>kcat</ci>
          <ci>Et</ci>
          <ci>S</ci>
        </apply>
        <apply>
          <plus/>
          <ci>Km</ci>
          <ci>S</ci>
        </apply>
      </apply>
    </lambda>
  </semantics>
</math>

is_a: SBO:0000028 ! kinetics of unireactant enzymes
```

Figure 3: SBO term describing Briggs-Haldane kinetics using the OBO flat format

k_{cat}x[E] is equal to the experimental maximal velocity, and K_S corresponds to the experimental substrate concentration required to reach half-maximal velocity. Henri-Michaelis-Menten mechanism assumed an underlying set of three elementary reactions:

$$E + S \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} ES \xrightarrow{k_{\mathrm{cat}}} E + P$$

In addition, those authors supposed a fast equilibrium between the enzyme/substrate complex and the free enzyme and substrate. As a consequence, K_S=k_{off}/k_{on}
One year later, Donald Van Slyke and Glenn Cullen [15] proposed another explanation based on two successive and irreversible reactions:

$$E + S \xrightarrow{k_{\mathrm{on}}} ES \xrightarrow{k_{\mathrm{cat}}} E + P$$

Although the microscopic mechanism is different, the general form of the velocity is equivalent. However, the constant, equivalent to K_S, is now $K=k_{cat}/k_{on}$.
Finally, in 1924, George Edward Briggs and John Burdon Sanderson Haldane generalised the mechanism than Michaelis and Menten described, releasing the hypothesis of fast equilibrium. Instead they replaced it with the famous quasi-steady-state approximation for the enzyme/substrate complex. The velocity follows yet again the same rate-law.

However, $K_m = \frac{k_{off} + k_{cat}}{k_{on}}$.

Now let's say we come accross a model describing a reaction using the Henri-Michaelis-Menten equation. Here is the SBML description of the reaction:

[SBML code]

```
<reaction>
  <listOfReactants>
    <speciesReference species="S" />
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="P" />
  </listOfProducts>
  <listOfModifiers>
    <speciesReference species="E" />
  </listOfModifiers>
  <kineticLaw>
    <listOfParameters>
      <parameter id="Km"/>
      <parameter id="kp"/>
    </listOfParameters>
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <apply>
        <times/>
        <ci>compartment</ci>
        <apply>
          <divide/>
          <apply><times/><ci>E</ci><ci>kp</ci><ci>S</ci></apply>
          <apply><plus/><ci>Km</ci><ci>S</ci></apply>
        </apply>
      </apply>
    </math>
  </kineticLaw>
</reaction>
```

There are several situations where we have to develop an elementary step equivalent, instead of using directly the combined non-linear version. For instance, we cannot use such a rate-law when the condition of substrate excess is not met or the total concentration of enzyme varies significantly. Another situation is the use of stochastic simulation tools. First, we have to create first an extra species *ES*. Then we have three possibilities for the reaction scheme.

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES}$$
$$\mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P}$$

With $k_{off}=K_m \times k_{on}$.

$$\mathrm{E} + \mathrm{S} \xrightarrow{k_{\mathrm{on}}} \mathrm{ES}$$
$$\mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P}$$

With $k_{on} = \frac{K_m}{k_{cat}}$.

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES}$$
$$\mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P}$$

With $k_{off} = K_m \times k_{on} - k_{cat}$

The second case is determined. In the first and third cases, one of the parameters has to be estimated, either from external knowledge or using a parameter estimation procedure[1].

If the model description is provided without additional information, there is no way to choose between the three alternatives. On the contrary, the annotation of the model is sufficient, not only to help us to decide between the three alternatives, but also to automatically convert the parameters and the rate-laws. Note the absence of MathML description of the rate-law in the kineticLaw element, unecessary in this case. Since all the parameters are local, a software can reconstruct the the rate-law by matching the SBO term reference on species, parameters and kineticLaw with the ones on the bvar of the SBO term MathML:

[1] k_{on} was choosen as the unknown, because it does not truly depend on the characteristics of the enzymatic reaction. Instead, it depends only on the environment (molecular crowding, viscosity) and scales with the square-root of the mass.

[MathMl code]

```
<reaction>
  <listOfReactants>
    <speciesReference species="A" definitionURL="http://www.biomodels.net/SBO/#SBO:0000015" />
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="B" definitionURL="http://www.biomodels.net/SBO/#SBO:0000011" />
  </listOfProducts>
  <listOfModifiers>
    <speciesReference species="C" definitionURL="http://www.biomodels.net/SBO/#SBO:0000014" />
  </listOfModifiers>
  <kineticLaw definitionURL="http://www.biomodels.net/SBO/#SBO:0000031" >
    <listOfParameters>
      <parameter id="U" definitionURL="http://www.biomodels.net/SBO/#SBO:0000027" />
      <parameter id="V" definitionURL="http://www.biomodels.net/SBO/#SBO:0000025" />
    </listOfParameters>
  </kineticLaw>
</reaction>
```

An SBO-aware software will have access to the vocabularies of SBO, either as a local copy, or using a programmatic access to the master copy. It will recognize that the kineticLaw represents a Briggs-Haldane kinetics and transform the description of the enzymatic reaction into the following elementary steps:

```
<listOfParameters>
  <parameter id="kon" definitionURL="http://www.biomodels.net/SBO/#SBO:0000036" constant="false" />
  <parameter id="koff" definitionURL="http://www.biomodels.net/SBO/#SBO:0000038" />
  <parameter id="U" definitionURL="http://www.biomodels.net/SBO/#SBO:0000027" />
  <parameter id="V" definitionURL="http://www.biomodels.net/SBO/#SBO:0000025"/>
</listOfParameters>
<listOfRules>
  <assignmentRule variable="kon">

    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <apply>
        <divide/>
        <apply><plus/><ci>koff</ci><ci>V</ci></apply>
        <ci>U</ci>
      </apply>
    </math>
  </assignmentRule>
</listOfRules>
<reaction id="v1" reversible="true">
  <listOfReactants>
    <speciesReference species="A" definitionURL="http://www.biomodels.net/SBO/#SBO:0000015" />
    <speciesReference species="B" definitionURL="http://www.biomodels.net/SBO/#SBO:0000014" />
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="AB" definitionURL="http://www.biomodels.net/SBO/#SBO:0000011" />
  </listOfProducts>
  <kineticLaw definitionURL="http://www.biomodels.net/SBO/#SBO:0000101" />
</reaction>
<reaction id="v2" reversible="false">
  <listOfReactants>
    <speciesReference species="AB" definitionURL="http://www.biomodels.net/SBO/#SBO:0000010" />
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="C" definitionURL="http://www.biomodels.net/SBO/#SBO:0000011" />
  </listOfProducts>
  <kineticLaw  definitionURL="http://www.biomodels.net/SBO/#SBO:0000049" />
</reaction>
```

The kineticLaw of the reaction v1 is annotated with the SBO term "second order forward with two reactants, first order reverse, reversible mass action kinetics, continuous scheme", while the kineticLaw of the reaction v2 is annotated with the SBO term "first order irreversible mass action kinetics, continuous scheme". Note that the species AB is associated with different SBO terms according to the reaction. The procedure would be exactly the same if instead of continuous descriptions for the elementary reactions, one wanted to use discrete rate-laws. The only changes would be the SBO terms on the kineticLaw elements.

Use of SBO to annotate experimental measurements

Precise annotation is not only necessary for theoreticians, but also for experimentalists. It is unfortunately all too frequent to come across confusions between V_{max} and k_{cat}, or K_p, K_d and IC_{50}. Similarly, the rate-law used to fit the experimental data-point and extract parameters is sometimes omitted. This potentially results in incorrect interpretations. This confusion reduces much the reuse of quantitative information in biochemistry, or even worse, lead to false interpretations.

A careful annotation of both rate-laws and parameters with the relevant SBO terms would increase the amount of information transferred from the data generation step to the data analysis one, minimising the risk of confusion, maximising the value for money of biochemical experimentation, and finally avoiding the continuous reiteration of the same data generation for different purposes.

Such an annotation could be directly reused by databases of quantitative biochemistry such as BRENDA [16] or SABIO-RK [17]. SBO terms could serve as a glue between various part of those knowledge management systems, but could also be used to query the resources, searching for a given parameter or a type of kinetics.

In addition, SBO annotation could help automatically generating part of the resources. For instance, using a mathematical expression term, one can directly create the adequate forms to enter the concentrations and parameters, as well as the corresponding structures in RDBMS tables.

SBO development and export

The Systems Biology Ontology is now listed as part of the Open Biomedical Ontologies (OBO). OBO is an umbrella for well-structured controlled vocabularies for shared use across different biological and medical domains. OBO seek to enforce some criteria of quality, orthogonality and stability among its ontologies. In addition, OBO ontologies share common formats and processing tools. As other OBO ontologies, SBO is an open-resource, developed and maintained by the scientific community, and reusable under the terms of the artistic license (http://www.opensource.org/licenses/artistic-license.php).

Everybody can submit request for new terms or suggestions to modify the structure of the ontology, or of the associated services through the Sourceforge project (http://sourceforge.net/projects/sbo/)

To curate and maintain SBO, we developed a dedicated resource (http://www.ebi.ac.uk/sbo/). A relational database management system (MySQL) at the back-end is accessed through a web interface based on JSP and JavaBeans. Its content is encoded in UTF8, therefore supporting a large set of characters in the definitions of terms. Distributed curation is made possible by using a tailored locking system allowing concurrent access. This system allows a continuous update of the ontology with immediate availability and suppress merging problems.

At the time we are writing this chapter, SBO is exported in the OBO flat format (http://www.godatabase.org/dev/doc/obo_format_spec.html). This format is rather unstructured, easely human-readable and shared by the majority of OBO ontologies. However, these qualities make the format a rather poor substrate for automated treatments, particularly in our case, where a portion of the content in in a highly structured form (MathML). At the time the chapter will be published however, it is likely that SBO will alsop be exported in OBO-XML and OBO-OWL. OBO-XML contains the same information that OBO-flat, but is expressed in eXtensible Markup Language [18], that permits extensive computing treatment. In addition it will make the incorporation of the MathML component of the mathematical expression branch trivial.
Finally, we seek to export SBO also using the Web Ontology Language (http://www.w3.org/2004/OWL/). OWL builds on RDF http://www.w3.org/RDF/ and URIs [7], and adds more vocabulary for describing properties and classes, thus improving the semantics of the format and facilitating automated interpretation.

We are also developing WebServices that will allow software to process SBO, or use it either to annotate dataset, or to interpret their annotation.

Conclusion and Perspectives

The need to exchange and integrate models drove the community to design common data format such as SBML. However, as important as was the definition of a common syntax, we also need to tackle the semantics of the models. It is expected that the adoption of MIRIAM and the Systems Biology Ontology will enhance the semantic content of quantitative biochemical description and favour their reusability.

Acknowledgements

BioModels.net and the associated projects has been initiated by Michael Hucka, Andrew Finney and NLN. We are grateful to all the people from the community who gave us their feedback. MC and CL are supported by a grant from the National Institute of General Medical Sciences (USA).

References

[1] A Felsenfeld, J Peterson, J Schloss, and M Guyer. Assessing the quality of the dna sequence from the human genome project. *Genome Res,* **9**:1–4, 1999.

[2] Nicolas Le Nov'ere, Benjamin Bornstein, Alexander Broicher, M'elanie Courtot, Marco Donizelli, Harish Dharuri, Lu Li, Herbert Sauro, Maria Schilstra, Bruce Shapiro, Jacky L Snoep, and Michael Hucka. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, **34** (Database issue): D 689 –D 691, 2006.

[3] Nicolas Le Nov'ere, Andrew Finney, Michael Hucka, Upinder S. Bhalla, Fabien Campagne, Julio Collado-Vides, Edmund J. Crampin, Matt Halstead, Edda Klipp, Pedro Mendes, Poul Nielsen, Herbert Sauro, Bruce Shapiro, Jacky L. Snope, Hugh D. Spence, and Barry L. Wanner. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology*, **23(12)**:1509–1515, 2005.

[4] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, J.C. Doyle, A.P. Arkin, B.J. Bornstein, D. Bray, A. Cornish-Bowden, A.A. Cuellar, S. Dronov, E.D. Gilles, M. Ginkel, V. Gor, I.I. Goryanin and W.J. Hedley, T.C. Hodgman, J.-H. Hofmeyr, P.J. Hunter, N.S. Juty, J.L. Kasberger, A. Kremling, U. Kummer, N. Le Nov'ere, L.M. Loew, D. Lucio and P. Mendes, E. Minch, E.D. Mjolsness, Y. Nakayama, M.R. Nelson, P.F. Nielsen, T. Sakurada, J.C. Schaff, B.E. Shapiro, T.S. Shimizu, H.D.Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics,* **19(4)**:524–531, 2003.

[5] CM Lloyd, MD Halstead, and PF Nielsen. CellML: its future, present and past. *Prog. Biophys. Mol. Biol.*, **85**:433–450, 2004.

[6] T Berners-Lee. Uniform resource locators (URL). a syntax for the expression of access information of objects on the network. Available via the World Wide Web at http://www.w3.org/Addressing/URL/url-spec.txt.

[7] T Berners-Lee, R Fielding, and L Masinter. Uniform resource identifier (URI): Generic syntax. Available via the World Wide Web at http://www.ietf.org/rfc/rfc3986.txt.

[8] R Moats. URN syntax. Available via the World Wide Web at http://www.ietf.org/rfc/rfc2141.txt.

[9] Xiaoshu Wang, Robert Gorlitsky, and Jonas S Almeida. From XML to RDF: how semantic web technologies will change the design of 'omic' standards. *Nat Biotechnol*, **23(9)**:1099–1103, Sep 2005.

[10] N Mitra. Soap version 1.2 part 0: Primer. Available via the World Wide Web at http://www.w3.org/TR/soap12-part0/.

[11] M Ashburner, CA Ball, JA Blake, D Botstein, H Butler, JM Cherry, AP Davis, K Dolinski, SS Dwight, JT Eppig, MA Harris, DP Hill, L Issel-Tarver, A Kasarskis, S Lewis, JC Matese, JE Richardson, M Ringwald, GM Rubin, and G Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.*, **25**:25–29, 2000.

[12] JS Luciano. Pax of mind for pathway researchers. *Drug Discov Today*, **10**:937–942, 2005.

[13] V Henri. Lois G'en'erales de l'Action des Diastases, pages 85–93. Herman, 1903.

[14] L Michaelis and ML Menten. Die kinetik der invertinwirkung. *Biochem. Z.*, **49**:333–369, 1913.

[15] Donald D Van Slyke and Glenn E Cullen. The mode of action of urease and of enzymes in general. *J. Biol. Chem.*, **19**:141–180, 1914.

[16] Ida Schomburg, Antje Chang, and Dietmar Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res*, **30(1)**:47–49, Jan 2002.

[17] U Wittig, M Golebiewski, R Kania, O Krebs, S Mir, A Weidemann, S Anstein, J Saric, and I Rojas. SABIO-RK: integration and curation of reaction kinetics data. *Lecture Notes in Bioinformatics*, **4075**:94–103, 2006.

[18] T Bray, J Paoli, CM Sperberg-McQueen, E Maler, F Yergeau, and J Cowan. Extensible markup language (xml) 1.1 (second edition), 2006.