

THE KINETICS WIZARD: A DATA CAPTURE TOOL FOR THE SUBMISSION OF ENZYME KINETICS DATA

NEIL SWAINSTON

Manchester Centre for Integrative Systems Biology, University of Manchester,
Manchester, M1 7DN, U.K.

E-Mail: neil.swainston@manchester.ac.uk

Received: 25th June 2008 / Published: 20th August 2008

INTRODUCTION

There are a number of resources containing enzyme kinetics data. Two widely used databases are BRENDA [1] and SABIO-RK [2]. While these databases contain kinetic constants, the key to ensuring that these resources can be usefully employed in a systems biology environment is in the richness of the metadata associated with these values.

Obvious requirements for these metadata include the environmental conditions, such as pH and temperature, under which these constants were measured. However there are other, more subtle, metadata that must also be captured and recorded along with the kinetic parameters to allow the database to be utilised correctly in modelling and simulation studies. This article describes these metadata and also introduces the KineticsWizard, a data capture tool that allows the experimentalist to specify these data in an intuitive manner.

The data capture tool is designed to be used by experimentalists, and as such, it is intended to hide much of the more technical aspects of data management from the user and present an intuitive, biologist-focussed interface from which the necessary metadata can nevertheless be input.

The tool itself is part of a larger system in which kinetic constants are determined through analysis and fitting of spectrophotometric data, and then automatically submitted to a data resource. In addition, the original raw data are also archived, providing the facility to map kinetic parameters back to their original source data, which could then be reanalysed or refitted if necessary.

Enzyme kinetics studies are a subset of the experimental programme of the Manchester Centre for Integrative Systems Biology (MCISB), and will be supplemented by both quantitative metabolomics and proteomics studies. The informatics infrastructure designed by the MCISB is one of a distributed, loosely-coupled system [3], in which a number of independent data resources are populated, and then later queried via web service interfaces in order to parameterise SBML models [4, 5].

The key to the development of such a distributed system is to ensure a consistent means of identifying species, reactions, and parameters across each of these data resources.

PROBLEMS OF CURRENTLY PUBLISHED ENZYME KINETICS DATA

Kummer and Sahle highlighted a number of issues with enzyme kinetics data currently published in existing resources [6]. Some of these problems are summarised below.

Importance of the kinetic equation

In addition to specifying kinetic parameters, such as V_{\max} and K_M , it is also necessary to specify the kinetic equation that is either assumed or was used to determine the constant. This additional information is crucial to ensure that modellers use the constant in the intended manner.

Furthermore, it is insufficient to specify the equation as a textual description, such as “Michaelis-Menten”, as these terms can be used inconsistently and do not unambiguously describe the intended kinetic equation.

Additionally, in cases where the reaction involves species with different stoichiometries, it is important to specify to which participant the parameter applies.

The V_{\max} parameter

V_{\max} parameters are often specified without any indication of the enzyme concentration contained within the term. While many V_{\max} parameters are generated from *in vitro* studies, the enzyme concentration is commonly either unspecified or poorly estimated. This reduces the usability of the parameter in modelling studies, as in these cases the enzyme concentration must be estimated, introducing unnecessary imprecision into the system.

Coherent unit notation

For parameters to be used reliably in modelling and simulation studies, standard units must be specified. Taking the example of enzyme concentration, it is common to see these values described by units such as “enzyme mass per dry weight of protein”, which dramatically reduces the usability of the parameter.

FURTHER ISSUES

In addition to the problems highlighted by Kummer and Sahle, there also remains the issue of inconsistent naming of reaction participants. It has been reported that relying on textual descriptions of small molecules and enzymes can result in inconsistencies, as the naming of such species is largely subjective and can differ greatly from individual to individual [7].

CAPTURE OF ENZYME KINETICS DATA

A submission tool is introduced here which has been designed to provide solutions to the problems specified above. The philosophy of the tool is to provide the facility for experimentalists to submit the required metadata without unduly burdening them with some of the more tedious tasks that this entails.

To provide this, the tool draws heavily on the use of existing data resources which are relevant to the task, and queries these resources via web service interfaces where possible. Exploiting existing data resources greatly reduces the volume of data that the experimentalist must submit. For example, consider the case of specifying a publication. Rather than providing a user with an interface to specify journal name, authors, paper title, etc., the philosophy followed here is to prompt the user for a PubMed identifier (id). Once this id is associated with a given parameter, web services can be queried from which the above fields can be extracted.

Furthermore, specifying and storing metadata as external database links greatly enhances the usability of the data resource, providing it with the facility to be queried with standard terms, and greatly enhancing its usability in a distributed, web service-driven infrastructure.

To illustrate this principle, a data submission tool is introduced here to show practical implementations of this approach.

Specifying the reaction components

The tool facilitates the consistent specification of reaction components by providing an interface to the KEGG database web service [8]. The user is presented with the following interface, which allows them to specify an organism and a gene name. Upon specifying these terms, the KEGG web service is queried and all reactions catalysed by the supplied

gene are returned. (Fig. 1). An individual reaction can then be specified, and as this reaction is an entry in the KEGG database, the entry can be queried to harvest a number of terms that would otherwise have to be specified by the user.

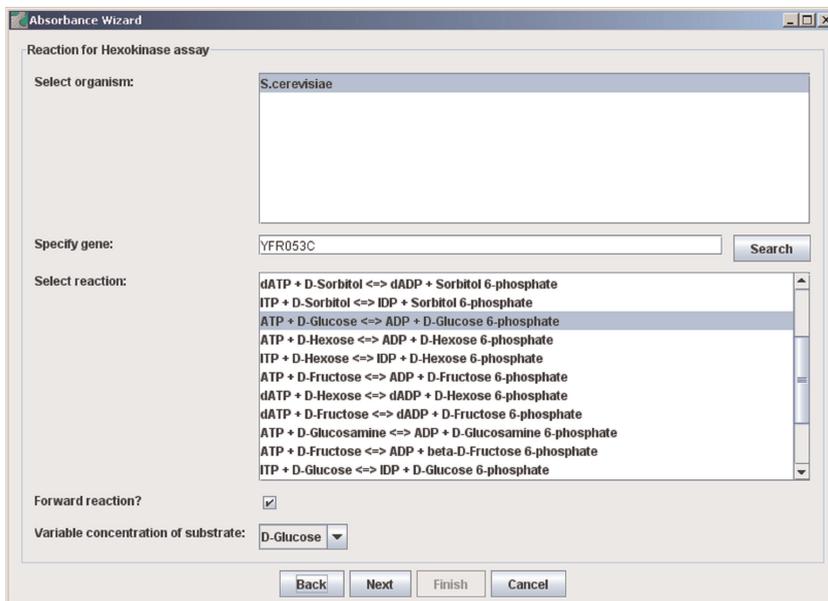


Figure 1. Specifying the reaction components.

The first of these is the enzyme itself. As the user specified a gene name, KEGG can be queried to determine the UniProt id of the enzyme. From this reference, a range of additional information can be subsequently harvested, such as protein name, synonyms, molecular weight, protein sequence, EC number and via external links to other resources, perhaps even also the protein structure. Specifying and ultimately storing metadata as external database references, rather than simply a textual name, can therefore greatly enhance the subsequent usability of this data.

The same principle applies to the reactants and products. By selecting a KEGG reaction, the web service can be queried to determine the reaction participants as entries in either KEGG or the ChEBI database[9], and also the stoichiometry of each of the participants. Consequently the selected reaction, textually presented to the user as “ATP + D-Glucose \leftrightarrow ADP + D-Glucose 6-phosphate”, will be defined computationally as an unambiguous set of reactants (CHEBI:15422 and CHEBI:17634), products (CHEBI:16761 and CHEBI:15954) and enzyme (UniProt:P04806).

It is also important to specify the reactant to which the parameter applies. The interface allows the user to specify this term upon selecting a reaction.

SPECIFYING THE KINETIC EQUATION

The specification of the kinetic equation once again exploits a web service to an existing data resource: in this case, the Systems Biology Ontology (SBO) [10]. SBO consists of a number of vocabularies, one of which – mathematical expression – can be used to describe reaction mechanism.

The user is presented with a tree of these SBO terms, and upon selection of a term, the underlying kinetic equation can be attained. (Fig. 2). This provides two advantages. The first is that the selected SBO term can be used to map to an appropriate data fitting algorithm used in the next step of the data analysis and submission pipeline. The second is that, upon submission, the kinetic parameters are immediately annotated with an unambiguous, standard term that defines the assumed mechanism of reaction and fitting. Rather than defining the reaction mechanism textually as “Michaelis-Menten”, the standard SBO term SBO:0000029 is used. Using such a defined term specifies the mechanism precisely: in this case it indicates that the mechanism is considered to be irreversible. Furthermore, the user is not prompted with the requirement of specifying the kinetic equation directly – a task that would be difficult to standardise and in many cases would be prone to error. This, along with other metadata such as synonyms and a description of the term, can be accessed directly from the SB ontology.

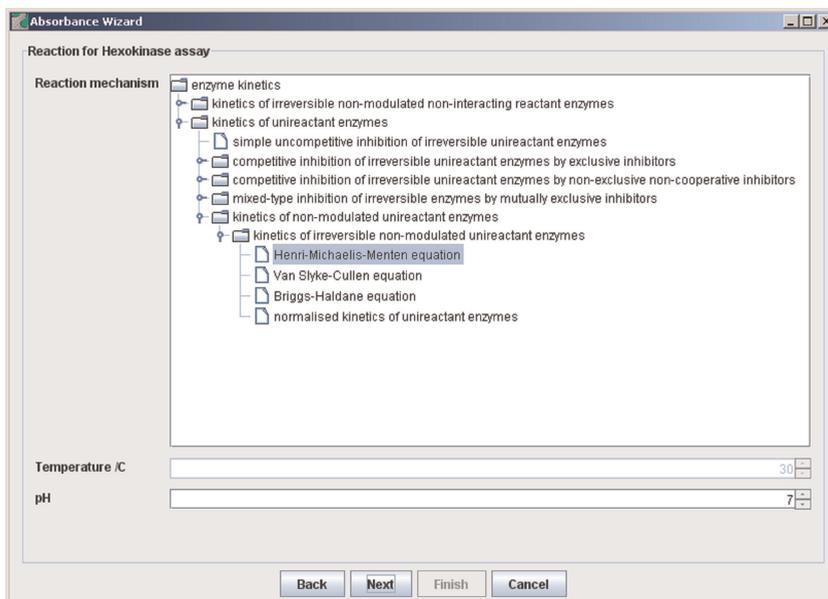


Figure 2. Specifying the kinetic equation.

A further advantage of the use of SBO is that the kinetic equations themselves refer to SBO terms to define the kinetic constants. Therefore, the constants themselves are not referred to in potentially ambiguous terms such as K_M and k_{cat} , but by the standard SBO terms SBO:0000027 and SBO:0000025 respectively.

The V_{max} parameter

Due to the nature of the kinetic assay experiments being performed, the enzyme concentration is known accurately at the time of experiment. As such, the user is prompted for this value and as such, upon data fitting, the k_{cat} value is calculated and submitted to the data resource, rather than the V_{max} .

Coherent unit notation

The user is presented with an interface that prevents the user from supplying units for the required terms. Substrate and enzyme concentrations are assumed to be accurately measured and their input is forced to be in mM and nM respectively. Dictating the units at the point of data submission ensures that these values will be consistent and not contain any “fuzzy”, imprecise or inconsistent terms.

DISCUSSION

The benefit of this approach is that kinetic parameters stored in data resources are associated to standardised terms, which greatly facilitates to querying of such resources. If a user then has an unparameterised but MIRIAM-compliant [11] SBML file, the task of parameterising the SBML file with kinetic parameters is greatly simplified, as both the SBML file and the data resource are annotated with consistent terms for metabolites, enzymes, EC codes, etc. Therefore, the task of automatically parameterising an annotated SBML becomes far easier, as the ambiguity that may exist between mapping metabolites or enzymes in the model with those in the database is removed.

It is the intention of submitting kinetic parameters, and their associated metadata terms, to send them to SABIO-RK automatically upon completion of the submission wizard. SABIO-RK stores kinetic parameters with references to the database terms that are collected through the submission process. Furthermore, SABIO-RK contains a web service interface that allows queries to be formed in terms of these database terms. (Fig. 3).

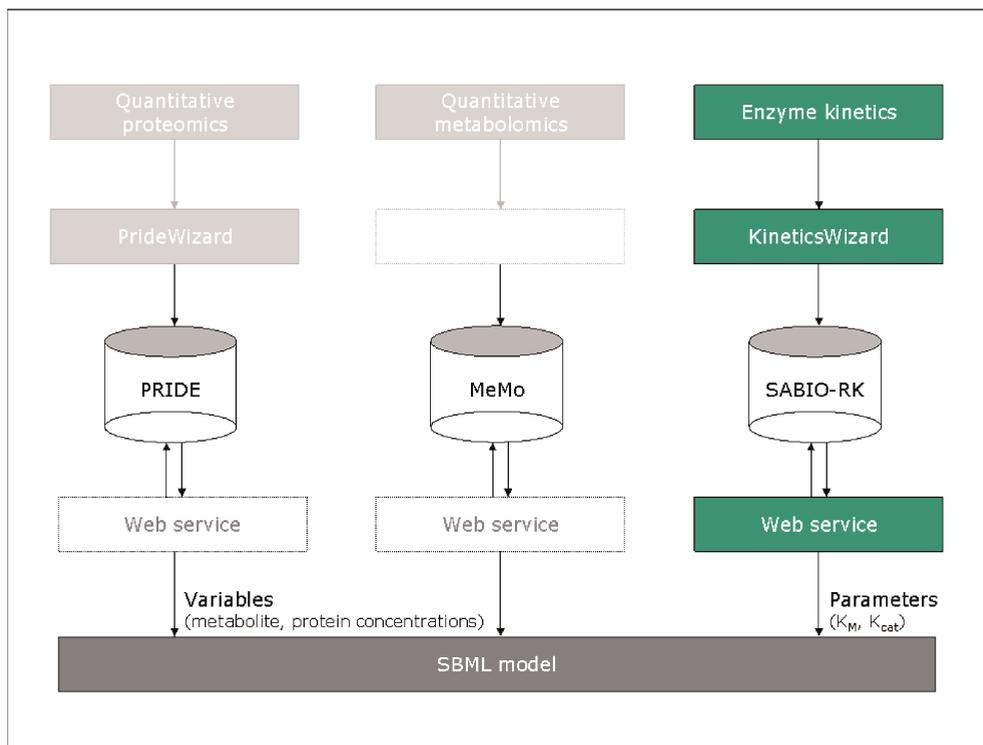


Figure 3. Overview of the MCISB informatics infrastructure. (MeMo is the MCISB metabolomics database [12]).

Whilst the importance of associating kinetics parameters with standardised database terms has been discussed, it is also necessary to ensure that this can be performed in a manner that is intuitive to the experimentalist. In order to facilitate this, existing web services are exploited such that options can be presented to users in human-readable format, while behind the scenes, many database references are gathered which allows the concept to be unambiguously specified using standardised terms.

ACKNOWLEDGEMENTS

I thank the EPSRC and BBSRC for their funding of the Manchester Centre for Integrative Systems Biology (<http://www.mcisb.org/>). I am also very grateful for the help and collaboration of colleagues at EML Research, specifically Isabel Rojas, Martin Golebiewski, Renate Kania, Olga Krebs, Saqib Mir and Ulrike Wittig.

REFERENCES

- [1] Schomburg, I., Hofmann, O., Baensch, C., Chang, A., Schomburg, D. (2000) Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. *Gene Funct. Dis.* **3–4**:109–18.
- [2] Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J., 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.
- [3] Kell, D.B. (2006) Metabolomics, modelling and machine learning in systems biology: towards an understanding of the languages of cells. The 2005 Theodor Bücher lecture. *FEBS J.* **273**:873–894.
- [4] Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., and the rest of the SBML Forum: 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 J. Wang (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**:524–531.
- [5] Li, P., Oinn, T., Stoiland, S., Kell, D.B. (2008) Automated manipulation of systems biology models using libSBML within Taverna workflows. *Bioinformatics* **24**:287–289.
- [6] Kummer, U., Sahle, S. (2006) Problems of currently published enzyme kinetic data for usage in modelling and simulation. *In: Proceedings of the 2nd International Beilstein Workshop on Experimental Standard Conditions of Enzyme Characterizations, Beilstein-Institut. Logos Verlag Berlin*, pp 129–136.
- [7] Herrgård, M., Swainston, N, *et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* Submitted.
- [8] Kanehisa, M., Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**:27–30.
-

- [9] Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M. (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* **36**:D344 – D350.
- [10] Le Novère, N. (2006) Model storage, exchange and integration. *BMC Neurosci.* Oct 30; 7 Suppl 1:S11.
- [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., Wanner, B.L. (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **12**:1509 – 15.
- [12] Spasiæ, I., Dunn, W.B., Velarde, G., Tseng, A., Jenkins, H., Hardy, N., Oliver, S.G, Kell, D.B. (2006) MeMo: a hybrid SQL/XML approach to metabolomic data management for functional genomics. *BMC Bioinformatics* **7**:281.
-

