Beilstein-Institut

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

Problems of Currently Published Enzyme Kinetic Data for Usage in Modelling and Simulation

Ursula Kummer and Sven Sahle

Bioinformatics and Computational Biochemistry Group, EML Research, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

E-Mail: ursula.kummer@eml-r.villa-bosch.de

Received: 14th August 2006 / Published: 31st August 2007

Abstract

Modelling, simulation and computational analysis have become important tools in modern biochemistry. Moreover, their tight integration with experimental approaches has become an integral part of systems biology which has attracted scientific and political interest all over the world. However, published enzymatic data often does not take a modeller's viewpoint into account, even though in many cases this would only demand minor adjustments and would serve the community a great deal. Supporting users by automating some of the steps in modelling and simulation adds even more requirements. In the following we would like to emphasize a few points that we feel should be further supported or that have been neglected in the discussion about the standardization of enzymatic data, but would be valuable for modellers.

Introduction

Even though computational biochemistry is a quite ancient part of life sciences, its impact and importance for experimental research has not been acknowledged until recently. The recent interest obviously stems from the fact that the sheer complexity of the biochemical network in a living cell (as opposed to simple isolated enzymatic reactions) calls for

computational help. Thus, today systems biology is understood as the tight integration of computational and experimental research in order to understand biochemical systems in their entirety.

In order to set up decent biochemical models we rely heavily on data from experiments and from previous modelling, especially kinetic data. However, the way this data has been published in the past is often lacking information crucial for the set-up of models. This has been recognized recently and discussed at the previous ESCEC meeting. In addition, the more frequent use of modelling techniques and the increasing size and complexity of models has led to the development of software tools that support users in the process of modeling, e. g., Pedro Mendes' group (VBI) and our group have developed COPASI (http://www.copasi.org) which offers a user-friendly, platform independent facility to set-up models, and to simulate and analyse them. In the course of developing the software as well as when performing modelling studies ourselves, we have encountered many problems with the published enzyme kinetic information. Most of that has been thoroughly and extensively discussed during the previous meeting.

However, we feel that some problems still have been neglected or at least are underestimated and we would like to point these out in the following.

Specific Problems

Importance of the kinetic equation

The vast majority of kinetic data published in the literature comprises V_{max} and Km values or other individual rate constants. However, this is only part of the information necessary for modelling the respective system. In many cases the actual kinetic equation which is assumed or even was used to derive the published parameter (often by fitting to the equation) is missing. Without this crucial information, the value of publishing the actual parameter is greatly diminished. It also does not help too much if authors mention the name of the corresponding rate-law in the text, as e. g. Bi–Bi- Ping–Pong, etc.; since these terms are not used in an unambiguous way and therefore can be very misleading. What is actually needed is the explicit notation of the respective equation – nothing else. This would make sure that modellers do not have to guess which equation to use. In addition, wrong use, e. g. using a parameter with the wrong rate law would be avoided. Just to illustrate this obvious point a little bit further we use the following arbitrary example:

We exchange the kinetic term for the hexose transporter in a model for yeast glycolysis by Teusink *et al*. The original term

$$\frac{\frac{Vmax \cdot (A - B)}{Kglc}}{1 + \frac{A + B}{Kglc} + \frac{Ki \cdot A \cdot B}{Kglc^2}}$$

is exchanged against a somewhat simpler Uni–Uni term

$$\frac{Vf \cdot \left(substrate - \frac{product}{Keq}\right)}{substrate + Kms \cdot \left(1 + \frac{product}{Kmp}\right)}$$

in which K_{eq} equals one. Since product and substrate are glucose and the respective K_m values are assumed to be the same, both terms are actually quite similar. We use the same parameters in both cases. The resulting models are analysed w.r.t. their steady state behaviour. This analysis is done using COPASI (http://www.copasi.org). The results are shown in Figs 1 and 2.

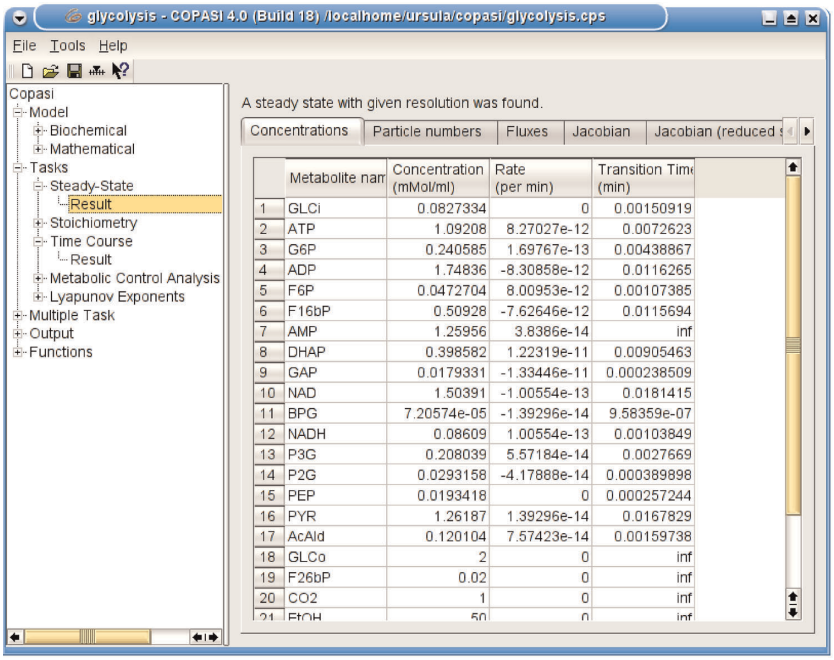

Figure 1. Steady-state concentrations as computed by COPASI using the glycolysis model of Teusink *et al*.

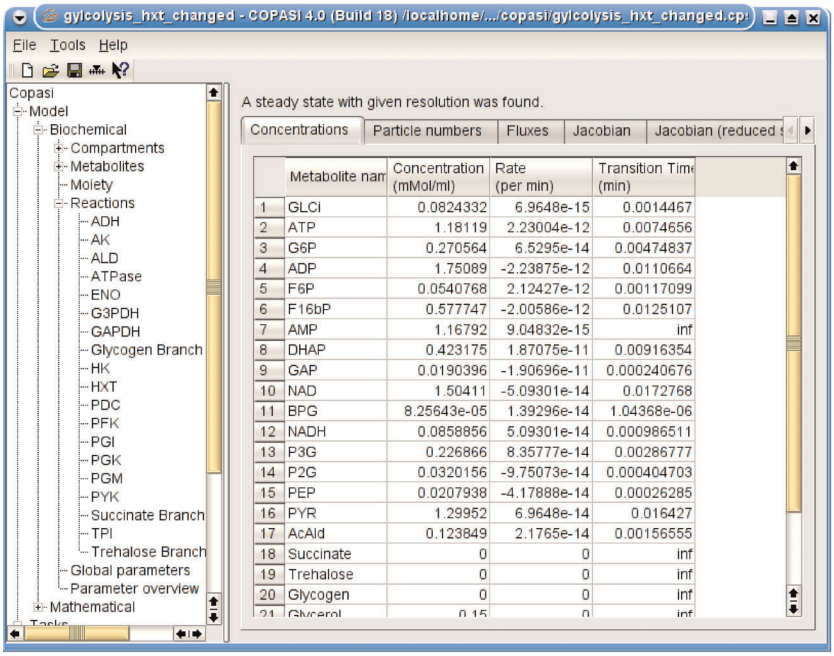

Figure 2. Steady-state concentrations of the same model as Fig. 1 with the term for the hexose transporter changed as explained in the text.

It is easy to see that the steady-state concentration of most variables differs by more that ten percent. Thus, the systems behaviour is significantly changed by this minor change in kinetics with the same parameters used. This trivial example illustrates the above said and calls for the inclusion of the notation of the kinetic equation in the standardization of published enzymatic data.

Finally, if a reaction involves participating species with different stoichiometries, it should be stated clearly to which participant (substrate or product) the rate law applies (as is often, but not always, done in literature). Preferably the rate law should be stated for a species with unity stiochiometry.

The V_{max} parameter

Another apparent (and recognized) problem is the publication of the V_{max} values. Since most studies are done *in vitro* the enzyme concentration contained in the V_{max} is the one in the test tube. However, modellers are usually interested in the enzyme concentration in the living cell instead. Even though the enzyme of interest has been isolated from cellular material in most cases, there is often not even an estimate of the amount present in the respective life material. An estimation of the original amount often is also not possible by calculating backwards since the results of the purification steps are not reported in sufficient detail.

In addition, instead of simply reporting the components of V_{max}, namely the enzyme concentration and the rate constant, many authors hamper the calculation of the individual rate constant by not explicitly writing down the respective enzyme concentration in the test tube, but rather giving the activity of the enzyme without giving amounts etc. (see unit notation below).

All in all, this effectively turns V_{max} into an unknown variable in most cases, introducing a lot of fuzziness into the system. Of course, in many if not most cases, there can be no exact quantificatation of the enzyme of interest in a specific cell type. This implies that parameter estimation techniques have to be used at some point in time. However, this procedure is obviously more reliable and much faster if the initial values are good guesses. These estimates could be very well provided in the primary literature.

Reversible rate laws

The notation of reversible rate laws is another, albeit less severe problem. Reversible rate laws do not pose any problem when models are written down using ordinary differential equations (ODEs). Forward and backward flows of a reversible reaction can cancel each other out so that the overall rate can be given as a single expression. Depending on the concentrations of the substrates and products the rate can be positive or negative, it is zero if the reaction is in equilibrium.

However, when modelling biochemical systems containing only relatively low numbers of the participating compounds, e.g. because of volume limitations (e.g. in vesicles) or because of functional necessity (e.g. signalling), we often have to refer to stochastic methods on discrete particle basis [1]. In the stochastic modelling and simulation framework each reaction is characterized by a reaction probability (instead of a reaction rate). A stochastical simulation works as follows: first the probabilities of all reactions are calculated. These depend on the concentrations of the species that take part in the reactions. Then, taking into account the probabilities of all the reactions, it is determined which reaction will take place next and at which point of time this will happen. This is done by drawing random numbers from a random number generator. The chosen reaction is then "executed" by increasing the particle numbers of the corresponding product species and decreasing the particle numbers of the substrates. So far one single reaction step was simulated. The whole process is repeated.

This stochastic simulation process ensures that the effects of discreteness (the fact that particle numbers are always integers) and the effects of stochasticity (the single reaction events happen at random points of time) are considered.

Concerning the relation between reaction rates and reaction probabilities it is clear that reaction rates can also be expressed as an average number of reaction events happening in a unit of time. This in turn can easily be translated into a reaction probability. Thus in many cases (and under certain conditions) the traditional rate laws and kinetic parameters can be utilized for stochastic simulations. A problem occurs, however, if the rate law describes a reversible reaction. Consider for example a reversible reaction in equilibrium. The net rate

is zero, which means that substrate and product concentrations do not change due to this reaction. It does not matter that in reality many reaction events in both direction take place. In the stochastic simulation however every single (forward and backward) reaction event needs to be simulated. Since the reactions are random, this leads to fluctuations around the equilibrium. For some short time more forward reaction events may happen, after that more backward reaction events occur. Only as an average over some time the reaction rate is zero. Therefore, separate rate laws for the forward and backward part of the reactions need to be available.

Thus, if rate laws are given for reversible reactions these terms have to be dismantled which is of course possible to do manually. Due to the increasing size of biochemical systems modelled and the reuse of parts of a model in other models, an automatization of this process however would be useful. Thus, e. g. COPASI contains a preliminary tool which is able to dissect reversible reactions automatically into forward and backward reactions (Fig. 3), but right now (apart from the trivial mass action case) the respective kinetics have to be adjusted by the user.

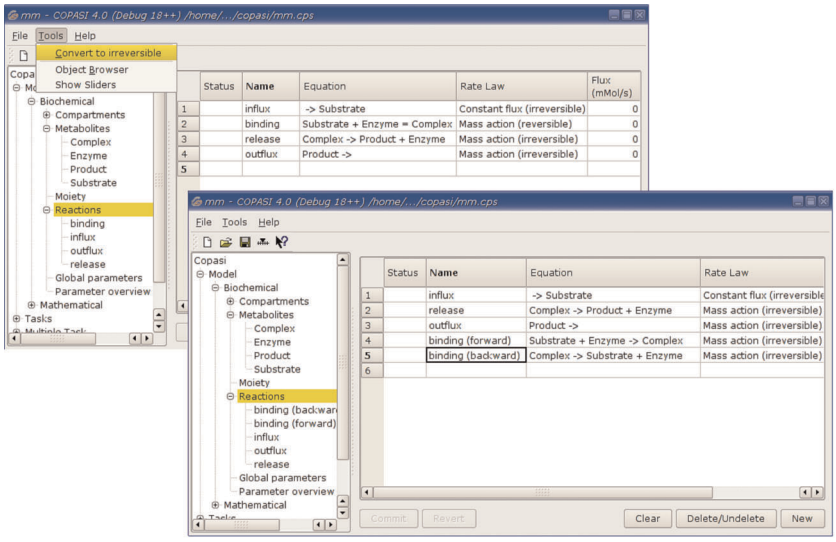

Figure 3. Screenshot of COPASI demonstrating the tool that renders reversible reactions into two irreversible reactions. The two windows show the list of reactions before and after the conversion.

Simultaneous and combined use of different simulation methods would be facilitated if rate laws were either written down individually (for forward and backward reactions) or written down in such a way that they can easily be dismantled automatically by computer programs. Thus, if the forward reaction rate is simply the first term of the numerator divided

by the denominator and the backward reaction rate is the second term of the numerator divided by the denominator as in the following example, an automatic dismantling is relatively simple, irrespective of e. g. brackets in this term.

An example taken from Holzhütter *et al*. [2] as stored in JWS online [3]:

$$v11[ENO] = \frac{Vmaxv11\left(Gri2P[t] - \frac{PEP[t]}{Keqv11}\right)}{Gri2P[t] + K2PGv11\left(1 + \frac{PEP[t]}{KPEPv11}\right)}$$

However, another example from the same paper as stored in the database shows a case where this is not as simple:

$$v9[BPGP] = \frac{Vmaxv9\left(Gri23P2f[t] - \frac{Gri3P[t]}{Keqv3} + MgGri23P2[t]\right)}{K23P2Gv9 + Gri23P2f[t] + MgGri23P2[t]}$$

Coherent unit notation

Problems with unit notations are mostly associated with the notation of enzymatic activities and concentrations. It is still common to use units like e. g. "activity per mg freshweight". As pointed out above, reuse of the respective kinetic data makes it necessary to compute the enzyme concentration in the assay. In order to do so, one has to gather all information from the text (if at all possible) about molecular weight, purity etc. This can be quite cumbersome and is probably done multiple times by different people in the community. Instead, it will be much easier if authors do this right away and provide the respective information in the original text.

Conclusions

Computational biochemistry relies more and more on tools that automate and facilitate individual steps in the setting up of models and their computational analysis. In addition to the general requirements of the modelling community, this development adds stronger and different requirements w.r.t. published enzymatic data. Some of these have been discussed above. We hope that enzymatic databases like SABIO-RK [4] and BRENDA [5] will also help by being a useful intermediate layer of information between the primary literature and the modeller being able to curate enzyme kinetic data in such a way that some of the above problems will be resolved.

Acknowledgements

We would like to thank the BMBF and the Klaus Tschira Foundation for funding. We are also grateful to Martin Golebiewski, Ralph Gauges, Stefan Hoops, Renate Kania, Pedro Mendes, Jürgen Pahle, Isabel Rojas, and Andreas Weidemann for fruitful discussions.

References

[1] Gillespie, D.T. (1976) A general method for numerically simulating the stochastic-time evolution of coupled chemical reactions. *J. Comp. Phys.* **22:**403–434.

[2] Holzhütter, H.-G. (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.* **271:**2905–2922.

[3] Olivier, B.G., Snoep, J.L.(2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* **20:**2143–2144.

[4] 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. *Lecture Notes in Bioinformatics* **4075:**94–103.

[5] Schomburg, I., Chang, A., Schomburg, D. (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* **30:**47–49.