Embedding kinetic pathway models into metabolic networks

Wolfram Liebermeister

Abstract

An algorithm ~~is presented~~ for embedding kinetic metabolic models into larger metabolic networks. As a first step, we build a stoichiometric-kinetic hybrid model with a consistent element mapping and a metabolic state characterised by flux distribution, metabolite concentrations, and thermodynamic forces. In a second step, the reactions in the stoichiometric network are equipped with standard rate laws, and ~~we obtain~~ a complete kinetic model with a thermodynamically feasible steady state by ~~construction~~. The algorithm works as follows. (i) A network model and one or more kinetic models are merged into a common network model in which all element conflicts have been resolved. (ii) To compute a metabolic reference state, a stationary flux distribution is chosen for the entire network, matching the original flux distributions of the kinetic submodels as closely as possible. (iii) The kinetic models are adjusted to this flux distribution, and for the rest of the network, standard rate laws are inserted and adjusted to the flux distribution. The algorithm allows to control the resulting steady state and guarantees a thermodynamically feasible model. Two example cases are presented.

1 Introduction

Large kinetic models would be important to study the effects of external changes, drugs, genetic variants, or genetic modifications on the metabolic state of cells. While small kinetic pathway models exist, large kinetic models are difficult to obtain (because of missing kinetic constants and other data; problems in parameter fitting), and only small regions of the biochemical networks are covered by models. If would be good to complete the other regions, ~~possibly with lower-quality, but still realistic models created automatically~~.

Why is this relevant? Couldn't small kinetic models describe relevant subsystems faithfully? In fact, the biochemical processes in cells are tightly connected. Pathways are embedded in a network and thereby interact with each other. To capture this in models, realistic constraints at their boundaries are necessary [1, 2]. Assuming, for instance, fixed ATP level in a glycolysis model may not be realistic. This becomes relevant when comparing in vitro and in vivo behaviour. If a metabolic pathway ~~were~~ reconstructed in ~~in~~ vitro, it ~~would satisfy~~ simple boundary constraints (e.g., an exact conservation of all chemical elements), whereas the cell, as an open system, may require a different description of its boundaries. This does not only hold for spatial boundaries (fluxes across the cell membrane), but also "boundaries" between parts of the metabolic network. The difference between in vitro and in vivo behaviour poses a problem because often one would like to use in vitro data for in vivo models. Currently, one can only say that this causes errors, which are hard to quantify. A better treatment of boundaries in models could improve data integration and parameter estimation from in-vivo data.

Here we show how existing pathway models ~~could~~ be embedded into larger network, to be modelled automatically (with a lower demand for data and details, lower quality, but maybe enough to improve prediction of dynamic pathway behaviour) The combination of existing models poses a number of issues. The technical issue of matching elements, possibly with different names, and spotting inconsistencies between model elements (e.g., a clash between models with and without cell compartments) has been addressed by the software semanticSBML [3]. So the main remain problem is to match models such that the resulting model is physically consistent. Thus, we need to understand which constraints need to be satisfied, which issues could arise, ~~–~~ and how they could be solved. We need a workflow in which certain conflicts are avoided by construction, while possible remaining conflicts need to be detected and resolved, especially if several pathway models are embedded into a network.

The present paper presents an algorithm, which has been implemented in MATLAB. In general, it can embed several existing kinetic models into a predefined metabolic network; but it can also be used for model combination

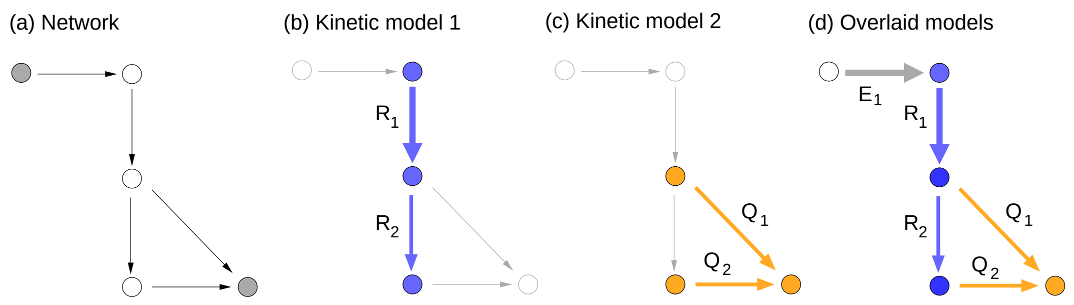

Figure 1: Metabolic network containing predefined kinetic pathway models. (a) Example network (metabolites shown as circles, external metabolites shown in grey). (b) Subnetwork covered by a kinetic model R; a flux distribution (fluxes: blue arrows; rate laws R_1 and R_2) follow~~s~~ from rate laws and concentrations in the kinetic model. (c) Subnetwork covered by kinetic model Q (rate laws Q_1 and Q_2). (d) Each kinetic model, in isolation, shows a set of stationary fluxes. The common stationary flux distribution is chosen to match these fluxes as closely as possible; the flux directions are strictly predefined. The fluxes in the kinetic models are matched to the common flux by adjust~~ed~~ the enzyme levels; ~~i~~n the remaining part of the network (grey), the flux is realised by standard rate laws (E_1).

(without embedding them into a network), ~~and~~ for turning a network into a kinetic model (without embedding existing models). As test~~s~~, ~~we consider~~ several pathway models and ~~simulate and analyse~~ their behaviour in isolation and embedded in the network. The results show how embedding a pathway in a larger system changes metabolic control within the pathway, and how embedding several disconnected pathways creates metabolic control between them.

2 Results

2.1 Model embedding

An algorithm is presented for embedding kinetic pathway models into metabolic networks, allowing ~~us~~ to build large-scale kinetic models of metabolism. The approach is described in the appendix and shown in Figure 1. In brief, it proceeds as follows. We define a metabolic network, parts of which are covered by ~~the~~ kinetic models. Each kinetic model, when run in isolation, defines a stationary flux distribution. ~~Our first aim~~ is to determine a stationary flux distribution for the entire network that approximates these fluxes as closely as possible. It is important that the flux directions are preserved because then, we can realise the [illegible]al fluxes within the kinetic models by ~~simply~~ adjusting ~~the~~ enzyme levels. Next, all remaining fluxes (outside the kinetic models) are realised by standard rate laws.

In the appendix, we present an algorithm that follows this general idea and takes care of a number of consistency issues: (i) the models have to be mapped, and it has to be ensured that their stoichiometric structures ~~actually fit~~; for instance, different levels of granularity in the diferent networks can cause problems; (ii) the kinetic models may overlap; for instance, different models may come with conflicting values of metabolites concentrations; (iii) The resulting model is supposed to be thermodynamically correct (~~and~~ least to the same extent as the initial kinetic models); this requires a proper matching of equilibrium constants; (iv) in the network outside the kinetic models, kinetic constants have to be chosen; this requires that either known kinetic constants are used properly, or that constants are guessed or sampled in sensible ways. The algorithm can handle several kinetic models ~~at a time.~~ I~~f~~ can combine several kinetic models, turn metabolic networks into kinetic scaffold models with a predefined reference states, embed kinetic models into such scaffold models, and even ~~to combine~~ a number of individual kinetic equations into a kinetic pathway model.

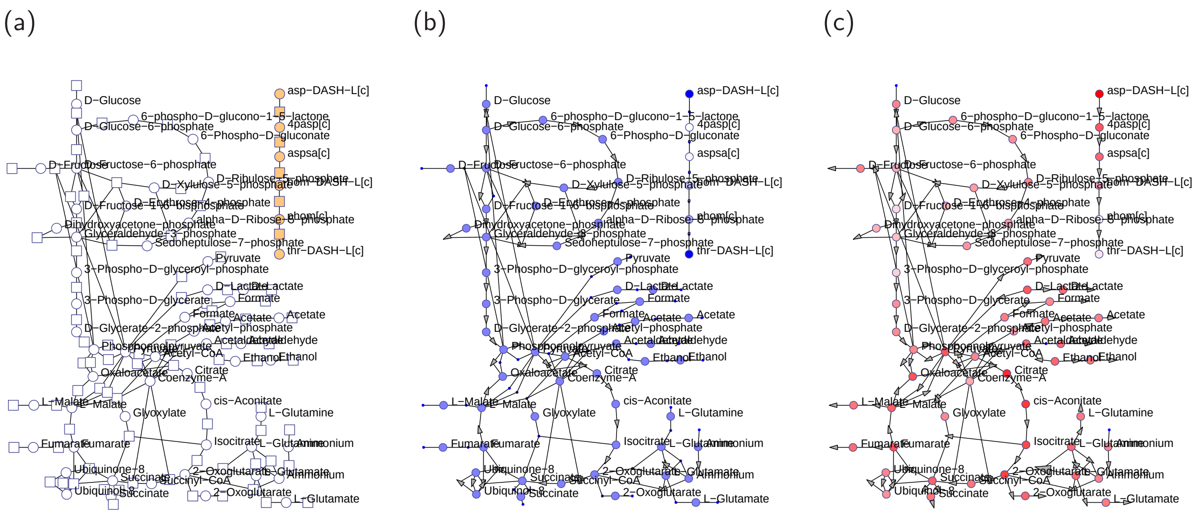

Figure 2: Threonine pathway, coupled to a model of central metabolism. (a) Network model (circles and squares denote metabolites and reactions); a part of the network is described by a kinetic model (brown elements, Chassagnole model [4] in SBML format obtained from BioModels Database [5]). The central metabolism model is an example model provided by the cobra toolbox [6]. (b) Stationary reference state. Fluxes (arrows) are the predetermined stationary fluxes; concentrations (shades of blue) are predefined within the initial kinetic model, other concentrations have been chosen by the algorithm. (c) Chemical potentials (shades of red) in the reference state.

2.2 Threonine pathway coupled to network model for ATP and NADH regeneration

The threonine synthesis pathway in *E. coli* consumes ATP and NADH. Chasagnole's kinetic model [4] displays a steady flux if the cofactors, as well as the substrate aspartate and the product threonine, are kept constant. ~~Of course,~~ this requires supply and degradation reactions, which are not part of the model ~~and~~ which exactly balance the consumption or production by the pathway. If the pathway was induced in a real cell, this would cause the ATP level to go down; however, a higher ATP demand might probably induce the ATP-producing reactions, which would restore the ATP level, partially and with a certain time lag. This ATP dynamics would feed back on the dynamics of the threonine pathway.

To model this dynamic interplay between ATP consumption and production, we coupled the original threonine pathway model to a model of central metabolism (see Figure 2). Unlike a simple ATP-generating reaction (which also could have been used to restore the ATP balance), this model captures time delays, and possibly the adaption of other pathways or growth to an increased ATP demand in the threonine pathway. In this example model, aspartate and threonine (the substrate and product of the threonine pathway) are still treated as fixed. In the model, this allows ~~one~~ to apply controlled perturbations to them and to study the response of the synthesis flux. Alternatively, they could also be coupled to a larger network model.

2.3 Yeast glycolysis and biomass production

In the previous example, a pathway was augmented by a network for cofactor supply. Now, we try the opposite: we consider a kinetic model of central metabolism and connect it to a second kinetic model, describing macromolecule (biomass) synthesis, which consumes ATP and other intermediates. Both models are connected by an "external" network providing "glue" reactions to connect the two kinetic models (see Figure 3).

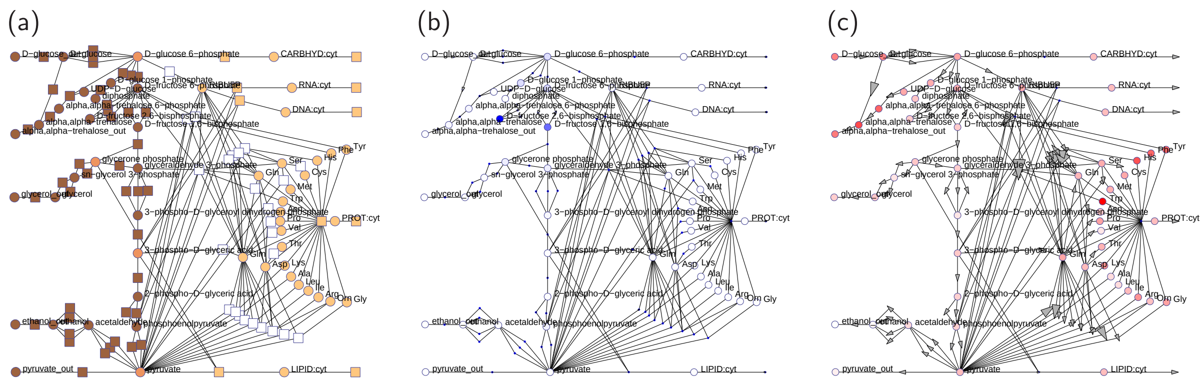

Figure 3: Yeast glycolysis coupled to biomass-producing reactions. (a) Network model (circles and squares denote metabolites and reactions); glycolysis is described by a kinetic model (brown elements). Biomass production is described by another kinetic model (orange elements). Both models had been built automatically based on a network structure, flux analysis, and standard rate laws. (b) Stationary reference state. Fluxes (arrows) are the predetermined stationary fluxes; concentrations (shades of blue) are predefined within the initial kinetic model, other concentrations have been chosen by the algorithm. (c) Chemical potentials (shades of red) in the reference state.

3 Discussion

Challenges Model embedding, as presented here, has to address a couple of challenges: (i) Element matching, especially with lumped elements. (ii) Compartments. (iii) Conflicts between model statements and quantities. Some of these challenges have been discussed in the context of SemSBML. Kinetic models have to be prepared to show sensible outflows: whenever there will be an outflow towards the surrounding network, the metabolites has to be set external already; there has to be a mismatch in the balance in the kinetic model; and the metabolite has to be marked to become internal after the combination.

References

[1] W. Liebermeister, U. Baur, and E. Klipp. Biochemical network models simplified by balanced truncation. *FEBS Journal*, 272(16):4034 – 4043, 2005.

[2] W. Liebermeister. Validity and combination of biochemical models. In *Proceedings of 3rd International ESCEC Workshop on Experimental Standard Conditions on Enzyme Characterizations*, 2008.

[3] F. Krause, J. Uhlendorf, T. Lubitz, E. Klipp, and W. Liebermeister. Annotation and merging of SBML models with semanticSBML. *Bioinformatics*, 26(3):421–422, 2010.

[4] C. Chassagnole, B. Raïs, E. Quentin, D. A. Fell, and J. Mazat. An integrated study of threonine-pathway enzyme kinetics in Escherichia coli. *Biochem J*, 356:415–423, 2001.

[5] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M Schilstra, B. Shapiro, J.L. Snoep, and M. Hucka. Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34:Database Issue:D689–91, 2006.

[6] J. Schellenberger, R. Que, R.M.T. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, D.R. Hyduke, and B.Ø. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nature Protocols*, 6:1290–1307, 2011.

Acknowledgements

This work is funded by the German Research Foundation (LI 1676/2-1).

A Algorithm for model embedding

A.1 Preconditions

- **Kinetic models.** A number of kinetic models, each defined by stoichiometric matrix, rate laws with all parameters, stationary concentrations and fluxes, list of external metabolites, and unique identifiers for metabolites and reactions.
- **Metabolic states of kinetic models.** Each submodel comes with a (stationary or non-stationary) state. For each flux, penalty weights for deviations must be defined.
- **Stoichiometric network model.** A stoichiometric model defined by a stoichiometric matrix, list of external metabolites, and unique identifiers for metabolites and reactions.
- **Criteria for network fluxes and concentrations.** FBA-type constraints or goal functions, for instance, a biomass production flux, fluxes to be approximated, or similar kinds of about concentrations, equilibrium constants etc.
- **Element mapping** A mapping of all elements in the kinetic models to elements in the stoichiometric model. All metabolites and reactions in the kinetic models must be mapped to metabolites and reactions in the stoichiometric model. If elements do not exist in the stoichiometric model, they will be created by the algorithm.
- **Synonymous elements.** Metabolites (or reactions) in different kinetic models that map to the same metabolite (reaction) are called synonymous. Reactions can appear in several models, but each belongs primarily to one model, which defines its rate law. Metabolites can appear in several models, but each metabolite belongs primarily to one model, which defines its concentration.
- **Resolving conflicts** Information about decisions taken by the user, in particular: (i) if several reactions in the kinetic models map to the same stoichiometric reaction: which of them will be used in the combined model? (ii) For metabolites with different statements about being internal or external, or different concentrations, how information should be chosen or combined.

A.2 Algorithm

Phase 1: Mapping between kinetic models and scaffold network

- **Check element mapping** Check if all metabolites and reactions from the kinetic models exist in the stoichiometric model. If some are missing, the user must check if some of them are identical between the kinetic models, add these missing metabolites and reactions to the stoichiometric model, and start again.
- **Unique naming** Create direct mappings (kinetic models to scaffold model) with numerical indices (not identifiers). If necessary, rename the metabolites and reactions in the stoichiometric model according to the names in the kinetic models.
- **Synonymous elements** Check whether reactions or metabolites in different kinetic models are synonymous. If this is the case, issue a warning (in particular, indicate that altered concentrations will alter the fluxes in kinetic models).
- **Conflicting model statements** Check for possible conflicts about metabolites being external/internal; in cases of conflict, let the user decide.

Phase 2: Choosing the flux distribution, concentrations, and equilibrium constants In the algorithm, fluxes and concentrations are constructed for the complete network. They must be thermodynamically feasible. Fluxes and concentrations in the original models should not be changed or, if this is impossible, the changes should be as small as possible.

- **Stationary fluxes.** Compute a stationary flux distribution in the large model, keeping the fluxes in the kinetic parts fixed. For instance, run a minimal sum of fluxes. If there is no solution, determine a stationary flux distribution in the large model that resembles the fluxes in the kinetic model (in some least-square sense, possibly with cost weights for different fluxes); and update the kinetic models such that they realise these fluxes (next point).
- **Concentrations in kinetic models.** Choose concentrations in kinetic models. The concentrations from the kinetic models should be preserved. If kinetic models overlap, the higher-priority models determine metabolite concentrations in lower-priority models. We recalculate the fluxes in the kinetic models, and if any of them change their directions, the algorithm should stop and warn the user.
- **Equilibrium constants in kinetic models.** Numerically compute equilibrium concentrations in the kinetic models (set everything internal and integrate); if some of them are exactly zero, replace them by small positive values.
- **Chemical potentials in kinetic models** Compute chemical potentials (from concentrations and equilibrium concentrations) of the metabolites in the kinetic parts.
- **Concentrations and equilibrium constants in surrounding network.** In choosing the metabolite levels in the surrounding network, the concentrations from the kinetic models must be preserved. All other concentrations must be determined to yield feasible thermodynamic forced and to satisfy certain other constraints (approximately given values, upper and lower bounds etc). Equilibrium constants or formation energies may be given or also be determined.
- **Chemical potentials in surrounding network.** Find chemical potentials such that there is agreement between fluxes and thermodynamic forces. If this is possible, use thermodynamic parameter balancing to obtain realistic values. If it is not possible, stop and ask the user to change the assumptions about the flux distribution; or continue and accept that the model will not be thermodynamically balanced.

Phase 3: Choosing or adapting the rate laws

- **Rate laws in surrounding network.** ~~in~~ Run first-order elasticity sampling for the reactions of the embedding network. This yields rate laws and kinetic constants for these reactions.
- **Adjust enzyme levels in embedded models.** If the original flux distributions from the kinetic models could not be realised in the network, the rate laws in the kinetic models must be updated to yield the new fluxes. Since the flux directions remained unchanged in the updating (otherwise the algorithm would have stopped), we can adjust the models by simply adjusting the Vmax values, or, in particular, the enzyme levels. The original enzyme levels are multiplied by a scaling factor V^{flux}/V^{kin}, i.e., updated flux divided by the original kinetic network flux (possibly, after having updated the reactant concentrations).
- **Build model** Construct the complete model (accoding to user's directives about how to choose among redundant reactions in different kinetic models) and export it as a matlab network structure and as an SBML file.

Phase 4 (optional): Replacing the surrounding network by reduced models

A.3 Implementation

The algorithm has been implemented in MATLAB.