Validity criteria and merging algorithm for kinetic models

Wolfram Liebermeister, MPI-MG Berlin
Computational Systems Biology

Workshop "Modeling Biological Systems - Syntax, Semantics and Context"
Rostock University, May 22, 2008

Mathematical models in systems biology

- Ordinary differential equations
- Partial differential equations
- Stochastic processes
- Particle-based models
- Constraint-based models
- Optimality-based models

... and many more ...

A model is basically
- a list of mathematical objects ...
- ... linked to biochemical quantities
- and a list of mathematical statements

Statements: facts and rules for calculation
- Predictive deterministic
- Probabilistic
- Properties or constraints
Playing LEGO with biochemical models?

Model composition

Model fusion

Model 1

Model 2

Model 3
Model merging, semantics, and validity

Sensible model merging requires a clear description of:

- biological semantics of model elements
- interpretation of model statements
- validity claims to be satisfied by the model
Part I: Models and model merging

Model merging has to do with the meaning of models
We have an intuition about this meaning
Can we formalise it?
Basic idea: merging of graphical models

Graphical elements are characterised by:

- Name (need not be compatible between models)
- Annotation (what does the element represent?)
  -> used for matching
- Additional information (values, statements, pictures,...)
  -> needs to be matched / selected
Our paradigm: deterministic kinetic models

Graphical model:
- F6P
- ATP
- FBP
- PFK
- FBA

Mathematical statements:
- \( C_{F6P} \) to \( V_{PFK} \) to \( C_{ATP} \)
- \( C_{FBP} \) to \( V_{FBA} \)

SBML code:
```xml
</species>
<species metaid="metaid_0000006"
annotation>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
><rdf:Description rdf:about="#metaid_0000006">
<bqbiol:is>
<rdf:Bag>
<rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/#CHEBI:15422"/>
</rdf:Bag>
</bqbiol:is>
</rdf:Description>
</rdf:RDF>
</annotation>
```

Annotations link model elements to database IDs.

ATP: a metabolite
- Appears in algebraic equation
- Appears in rate equation

FBA: a reaction
### Explicit biochemical models

#### Network scheme

- **F6P** → **ATP**
- **PFK** → **ATP**
- **FBP** → **FBA**

#### Model statements

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Variable</th>
<th>Mathematical assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP conc. [mM]</td>
<td>$c_{\text{ATP}}$</td>
<td>$C_{\text{ATP}} = 0.5$</td>
</tr>
<tr>
<td>F6P conc. [mM]</td>
<td>$c_{\text{F6P}}$</td>
<td>$C_{\text{F6P}} = 0.1$</td>
</tr>
<tr>
<td>FBP conc. [mM]</td>
<td>$c_{\text{FBP}}$</td>
<td>$\frac{dc_{\text{FBP}}}{dt} = v_{\text{PFK}} - v_{\text{FBA}}$; $c_{\text{FBP}}(0) = c_{\text{F6P}}$</td>
</tr>
<tr>
<td>PFK vel. [mM/s]</td>
<td>$v_{\text{PFK}}$</td>
<td>$v_{\text{PFK}} = v_{\text{PFK}}(c_{\text{F6P}}, c_{\text{ATP}})$</td>
</tr>
<tr>
<td>FBA vel. [mM/s]</td>
<td>$v_{\text{FBA}}$</td>
<td>$v_{\text{FBA}} = v_{\text{FBA}}(c_{\text{FBP}})$</td>
</tr>
</tbody>
</table>
Explicit biochemical models

Network scheme

Model statements

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Variable</th>
<th>Mathematical assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP conc. [mM]</td>
<td>( c_{\text{ATP}} )</td>
<td>( \text{ATP} = 0.5 )</td>
</tr>
<tr>
<td>F6P conc. [mM]</td>
<td>( c_{\text{F6P}} )</td>
<td>( \text{F6P} = 0.1 )</td>
</tr>
<tr>
<td>FBP conc. [mM]</td>
<td>( c_{\text{FBP}} )</td>
<td>( \frac{dc_{\text{FBP}}}{dt} = v_{\text{PFK}} - v_{\text{FBA}} ) ( c_{\text{FBP}}(0) = c_{\text{F6P}} )</td>
</tr>
<tr>
<td>PFK vel. [mM/s]</td>
<td>( v_{\text{PFK}} )</td>
<td>( v_{\text{PFK}} = v_{\text{PFK}}(c_{\text{F6P}}, c_{\text{ATP}}) )</td>
</tr>
<tr>
<td>FBA vel. [mM/s]</td>
<td>( v_{\text{FBA}} )</td>
<td>( v_{\text{FBA}} = v_{\text{FBA}}(c_{\text{FBP}}) )</td>
</tr>
</tbody>
</table>

If every single statement is true, then the model is true.
Explicit biochemical models: formal semantics and explicit statements

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Variable</th>
<th>Mathematical assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP conc. [mM]</td>
<td>$c_{\text{ATP}}$</td>
<td>$c_{\text{ATP}} = 0.5$</td>
</tr>
<tr>
<td>F6P conc. [mM]</td>
<td>$c_{\text{F6P}}$</td>
<td>$c_{\text{F6P}} = 0.1$</td>
</tr>
</tbody>
</table>
| FBP conc. [mM] | $c_{\text{FBP}}$ | $\frac{dc_{\text{FBP}}}{dt} = v_{\text{PFK}} - v_{\text{FBA}}$
| | | $c_{\text{FBP}}(0) = c_{\text{F6P}}$ |
| PFK vel. [mM/s] | $v_{\text{PFK}}$ | $v_{\text{PFK}} = v_{\text{PFK}}(c_{\text{F6P}}, c_{\text{ATP}})$ |
| FBA vel. [mM/s] | $v_{\text{FBA}}$ | $v_{\text{FBA}} = v_{\text{FBA}}(c_{\text{FBP}})$ |

If every single statement is true, then the model is true.

If every single statement is valid, then the model is valid.
The concatenated model is wrong!!
No experiment is needed to show this.
Requirements for deterministic simulation

Mathematical dependencies

- **C<sub>F6P</sub>**
- **V<sub>PFK</sub>**
- **C<sub>ATP</sub>**
- **C<sub>FBP</sub>**
- **V<sub>FBA</sub>**

- ATP conc. [mM] - \( c_{ATP} \) - \( C_{ATP} = 0.5 \)
- F6P conc. [mM] - \( c_{F6P} \) - \( C_{F6P} = 0.1 \)
- FBP conc. [mM] - \( c_{FBP} \) - \( \frac{dc_{FBP}}{dt} = v_{PFK} - v_{FBA} \)
- PFK vel. [mM/s] - \( v_{PFK} \) - \( v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP}) \)
- FBA vel. [mM/s] - \( v_{FBA} \) - \( v_{FBA} = v_{FBA}(c_{FBP}) \)

Model statements

Simulation requires:
- Explicit statements
- Exactly one statement per variable
- Stepwise evaluation of algebraic equations (no computational cycles)

Other mathematical frameworks have different requirements
Computational cycles should be avoided

Problem: algebraic equations have no solution or cannot be evaluated

→ algebraic loops are forbidden!
Semantics of model elements: biochemical objects, entities, and quantities

**BIOLOGICAL ENTITIES**
- Substances
- Reactions
- Compartments
- States
- Events

**BIOCHEMICAL QUANTITIES**
Measurable quantity (can have a numerical value)

Quantity = (type, unit, entity, place)

“Concentration [mM] of ATP in cytosol”

**Quantity types**

**Amounts**
- Compartment volume
- Substance concentration
- Reaction velocity

**Individual properties**
- Turnover rate
- Rate constant
- Equilibrium constant
- Probability of a state

**Interaction properties**
- $K_m$ value, $K_i$ value
- $K_A$ value, Hill coefficient
Semantic dependence between biochemical quantities

**Quantity = (type, unit, entity, place)**

- “Concentration [mM] of ATP in cytosol”
- “Velocity [mM/s] of PFK in cytosol”
- “Amount [mol] of ATP in cell”

**Semantic dependence between quantities**
“The definition of two quantities implies constraints about their numerical values”

**Postulate about semantic dependence:**
“Either type, or entity, or localisation must be independent”

---

**Quantity types**

- **Amounts**
  - Compartment volume
  - Substance concentration
  - Reaction velocity

- **Individual properties**
  - Turnover rate
  - Rate constant
  - Equilibrium constant
  - Probability of a state

- **Interaction properties**
  - $K_M$ value
  - $K_I$ value
  - $K_A$ value
  - Hill coefficient
Semantic dependence between quantities

Relations for quantity pairs

- independent ➞ no conflict
- identical ➞ *statement conflict*; choose between statements
- interconvertible ➞ *statement conflict*; need to be converted in advance
- semantic overlap ➞ *severe conflict*; models cannot be merged

Model 1

- ATP concentration
- ATP concentration
- ATP concentration [mM]
- ATP *concentration*
- concentration in *cell*
- lumped reaction
- Ribosome concentration

Model 2

- ADP concentration
- ATP concentration
- ATP concentration [M]
- ATP *amount*
- concentration in *nucleus*
- individual reaction steps
- total RNA concentration

Model 1 and Model 2 are related through various quantities and their concentrations in the cell and nucleus.
A merging algorithm for explicit semantic models

I. Compare the quantities

Semantic overlap?? merging is impossible

II. Choose between statements

Algebraic loops? modify choice!
A merging algorithm for explicit semantic models

I. Compare the quantities

II. Choose between statements

Algebraic loops? modify choice!
Interpretation of statements
Part II: Claims, conflicts, and validity of models

We have an intuition about the purpose of a model

Can we formalise it?
What is a hydrogen atom?
What is a hydrogen atom?
What is a hydrogen atom?
What is a hydrogen atom?
“All models are essentially wrong, but some are useful”

George Box

- What is the purpose of a model?
- What criteria does it have to satisfy to suit its purpose?
Purpose and claims of biochemical models

Some possible claims

- "All model elements are described unambiguously"
- "The model can be solved numerically; it has a unique solution."
- "The model has a steady state"
- "The model is thermodynamically feasible"
- "The model parameters are biologically plausible"
- "All parameter values are supported by experimental data"

What are the claims good for?

- Claims can imply that the model is suited for certain tasks (simulation, prediction, ...)
- Validity means that all claims are fulfilled
- Conflicts / problems mean that a claim is not fulfilled

Practical questions concerning claims

- Which claims are relevant for a model / for a purpose?
- How can different claims be checked? What information is necessary?
- Can we expect / do we require that claims are still satisfied after merging?
- How can we ensure this in a merging algorithm?
## Classification of claims

**Valid models** satisfy certain predefined quality requirements  
An invalid model will either be wrong or won't serve its purpose

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Valid Statements</th>
</tr>
</thead>
</table>
| **1. Syntax**                | model can be read and processed  
correct and common file format  |
| **2. Computation**           | model can be used for predictive simulations  
statements are unique and complete, can be evaluated sequentially, ... |
| **3. Semantic correctness**  | model statements agree with the model semantics  
valid statements, no semantic dependencies, ... |
| **4. Empirical correctness** | model agrees with physical and biochemical facts  
realistic numerical values,  
correct thermodynamics,  
correct reaction balances, ... |
| **5. Relevance**             | model performs well and suits its purpose  
agreement with data, plausible assumptions, no irrelevant parts,  
model becomes a paradigm ... |
Local and global claims

Claims related to single statements:

- "All model elements are annotated"
- "Atom numbers are conserved in all reactions"

Claims related to several statements or the entire model:

- Model does not contain computational loops
A formal description of claims in SBML?

Suggestions (to be discussed):

Mathematical scope:
• Specify dynamic behaviour ("...shows a Hopf bifurcation") TEDDY?
• Specify solvability ("...has a single deterministic solution")

Physical scope
• "Thermodynamically feasible parameters"
• "Atom numbers are conserved"

Biological scope:
• Specify organism ("... supposed to hold for E. coli"; NCBI ID)
• "... contains only enzymes found in this organism"
• "... contains all relevant processes of glycolysis"
• Specify experimental /natural conditions
  ("... only for aerobic growth conditions"; ontology needed)
• Specify data to be explained (link to annotated data set)
Links and literature

SemanticSBML our tool for model annotation, checking, merging)

Home page:  http://sysbio.molgen.mpg.de/semanticsbml/
Web interface:  http://sysbio.molgen.mpg.de/semweb/start
Sourceforge:  http://sourceforge.net/projects/semanticsbml/

Paper on SBMLmerge (previous version of SemanticSBML)
Schulz M., Uhlendorf J., Klipp E., Liebermeister W. (2006), SBMLmerge, a system for combining biochemical network models, Genome Informatics Series 17 (1).

Paper on model merging and validity (topics of this talk)

Can be found at:  www.molgen.mpg.de/~lieberme/pages/publications.html
Acknowledgements

MPI Berlin
Falko Krause
Jannis Uhlendorf
Marvin Schulz
Edda Klipp

Universität Heidelberg
Ursula Kummer
Sven Sahle
Ralph Gauges

EML Heidelberg
Isabel Rojas
Martin Golebiewski

MPI Magdeburg
Martin Ginkel
Michael Rempel

Funding by European integrated project
BaSysBio