Validity tests and merging of biochemical models
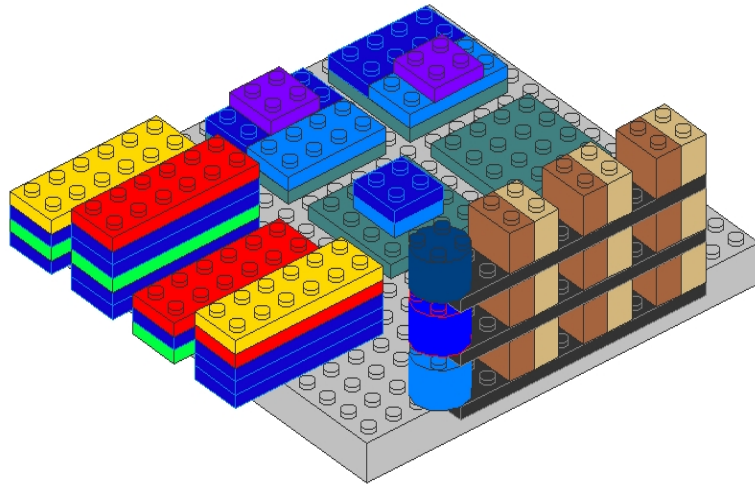
Wolfram Liebermeister, MPI-MG Berlin

Computational Systems Biology
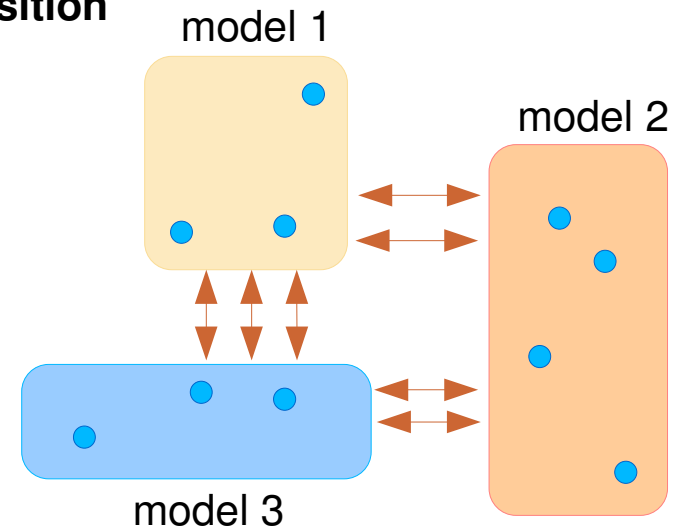
Playing LEGO with biochemical models ?

Model composition

Playing LEGO with biochemical models ?

Model composition

model 1
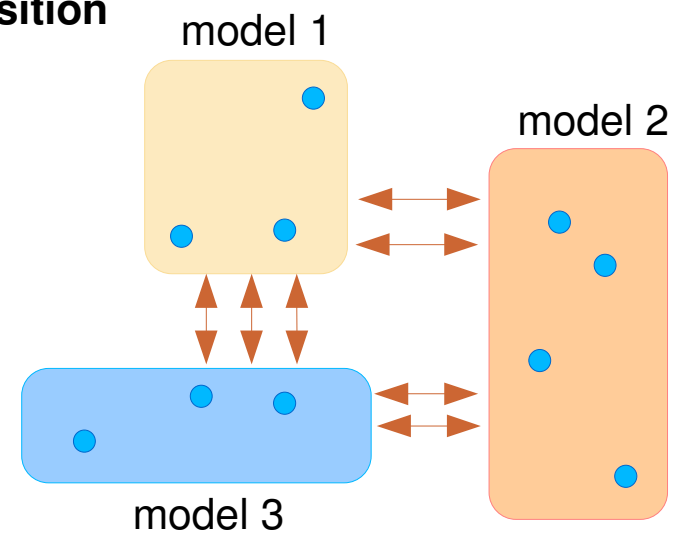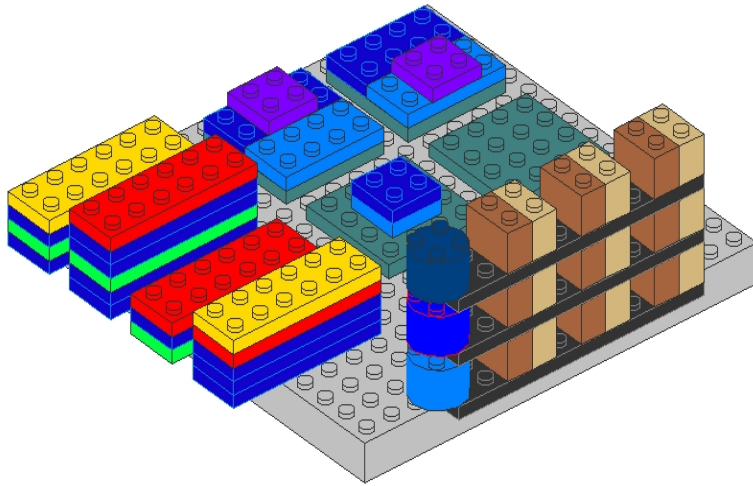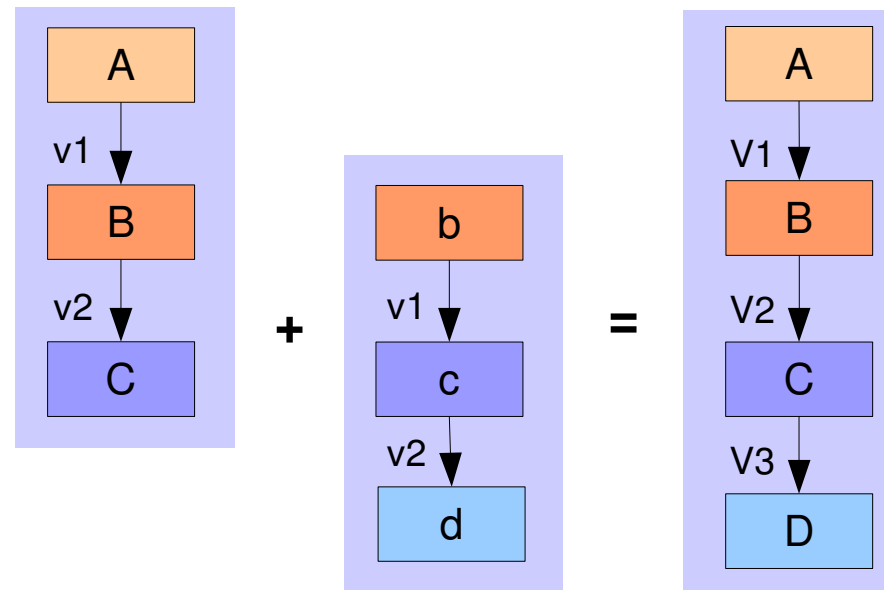
model 2

model 3

Model fusion
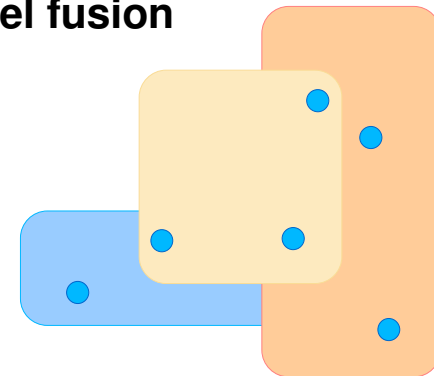
Merging of biochemical models: the basic picture

Questions for this talk

What kinds of models are we talking about ?
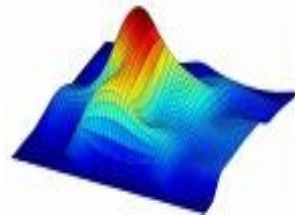
What does "a valid model" mean ?

Can we ensure that models remain valid during merging ?

Can standards help to prepare reusable models ?

Mathematical models in systems biology

Ordinary diffential equations

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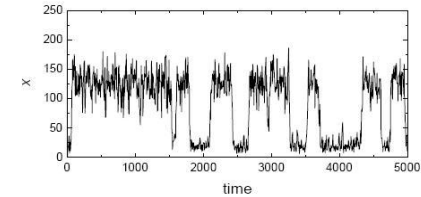
Stochastic processes
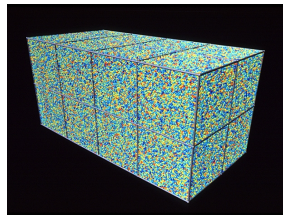
Particle-based models
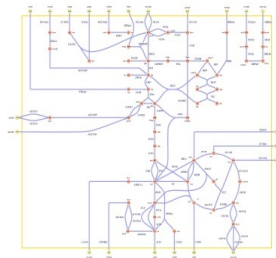
Constraint-based models
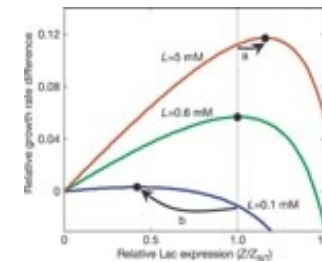
Optimality-based models

A model is basically:

• List of mathematical objects

• Mapping to biochemical objects / quantities

• List of mathematical statements

Statements: facts or rules for computation ?

• Predictive deterministic

• Probabilistic

• Properties or constraints

Correct models ...

Correct models ...

Correct models ...

Ha 656.3

Correct models ...

Hb
Ha

Ha 656.3
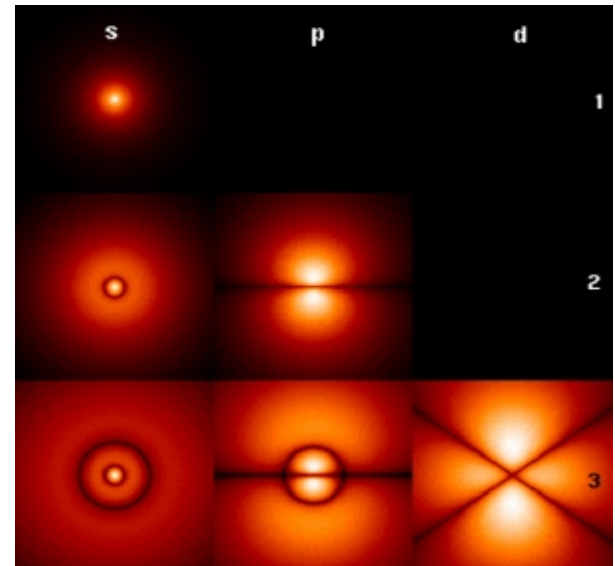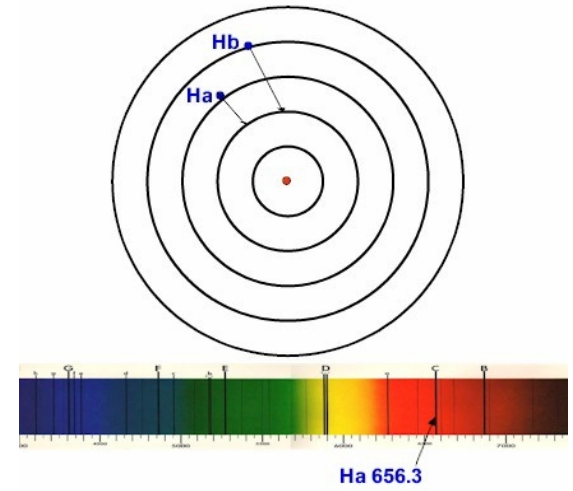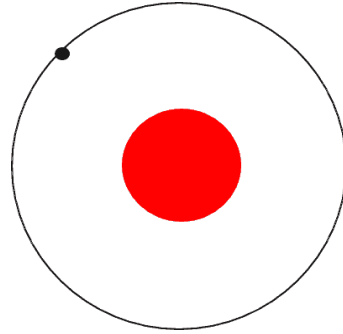
Correct models ...
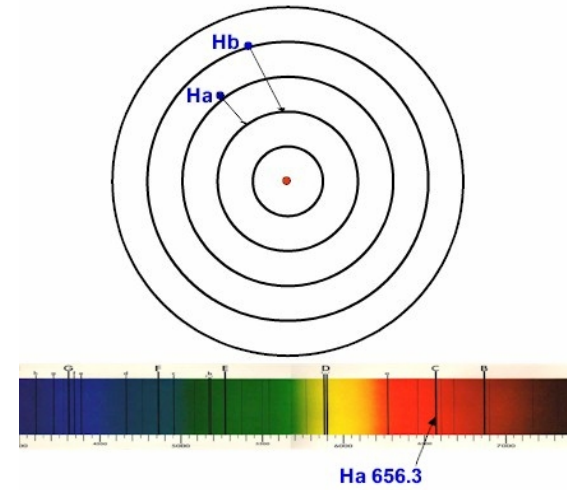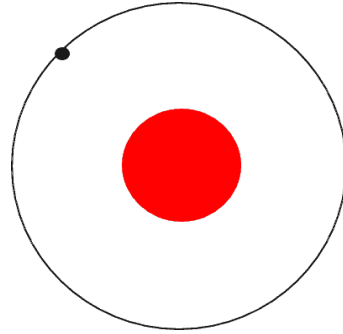
"All models are wrong, but some are useful"

George Box
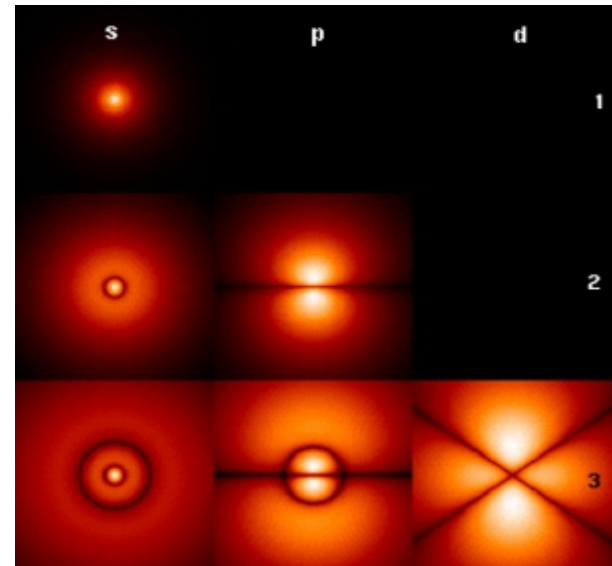
Biochemical models should be ... valid !

Valid models satisfy certain predefined quality requirements

An invalid model will either be wrong or won't serve its purpose

1. Syntax

model can be read and processed
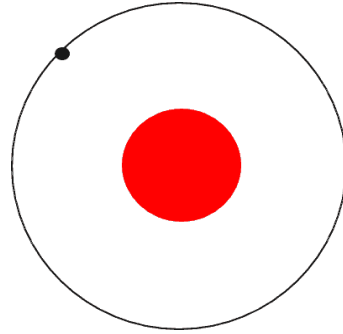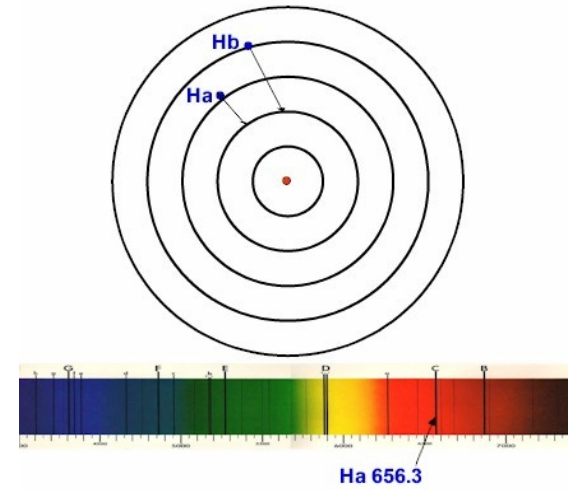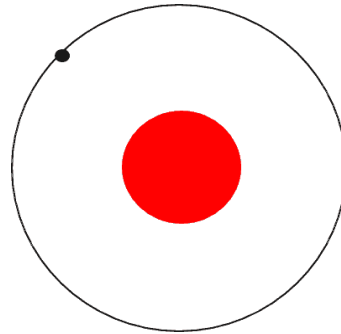
correct and common file format

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 ...

A common modelling paradigm: Deterministic rate equation models

Graphical model

A common modelling paradigm:
Deterministic rate equation models
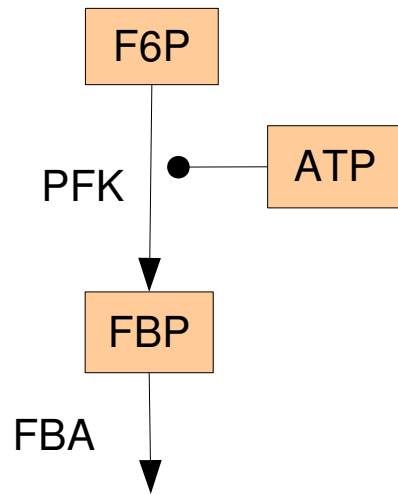
Graphical model

F6P

PFK   ATP

FBP

FBA

Mathematical statements

C_{F6P}

V_{PFK} C_{ATP}

C_{FBP}

V_{FBA}

ATP   a metabolite

FBA   a reaction

Appears in algebraic law

Appears in rate law

A common modelling paradigm: Deterministic rate equation models
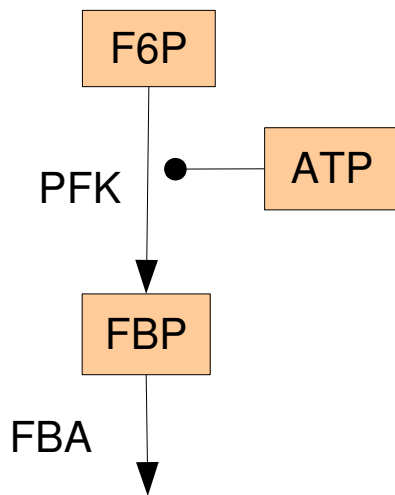
Graphical model

Mathematical statements
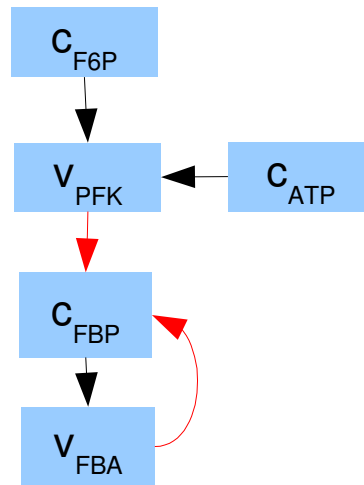
SBML code

```
</species>
 <species metaid="metaid_0000006"
 <annotation>
 <rdf:RDF xmlns:rdf="http://www.w
 <rdf:Description rdf:about="#met
  <bqbiol:is>
   <rdf:Bag>
    <rdf:li rdf:resource="http://
    <rdf:li rdf:resource="http://
   </rdf:Bag>
  </bqbiol:is>
  </rdf:Description>
 </rdf:RDF>
</annotation>
```

ATP — a metabolite

FBA — a reaction

Appears in algebraic law

Appears in rate law

Annotations link model elements to database IDs

Semantic view of rate equation models

Network scheme

F6P

PFK

ATP

FBP

FBA

Model statements

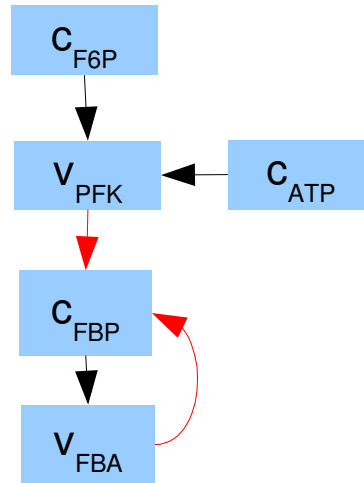|		
ATP conc. [mM]	c_{ATP}	$C_{ATP} = 0.5$
F6P conc. [mM]	c_{F6P}	$C_{F6P} = 0.1$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK} - v_{FBA}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

Quantity   Variable   Mathematical assignment

Semantic view of rate equation models

Network scheme

F6P

PFK ATP

FBP

FBA

Model statements

ATP conc. [mM]	c_{ATP}	$C_{ATP} = 0.5$
F6P conc. [mM]	c_{F6P}	$C_{F6P} = 0.1$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK} - v_{FBA}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

Quantity   Variable   Mathematical assignment

If every single statement **is correct**, then the model **is correct**.

Semantic view of rate equation models

Network scheme

F6P

PFK ●━━━ ATP

FBP

FBA

Model statements

ATP conc. [mM]	c_{ATP}	$C_{ATP} = 0.5$
F6P conc. [mM]	c_{F6P}	$C_{F6P} = 0.1$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK} - v_{FBA}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

Quantity   Variable   Mathematical assignment

If every single statement **is correct**, then the model **is correct**.

If every single statement **is valid**, then the model **is valid**.

Semantic view of rate equation models

Network scheme

F6P

PFK → ATP

FBP

FBA

Model statements

Quantity	Variable	Mathematical assignment
ATP conc. [mM]	c_{ATP}	$C_{ATP} = 0.5$
F6P conc. [mM]	c_{F6P}	$C_{F6P} = 0.1$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK} - v_{FBA}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

Quantity   Variable   Mathematical assignment
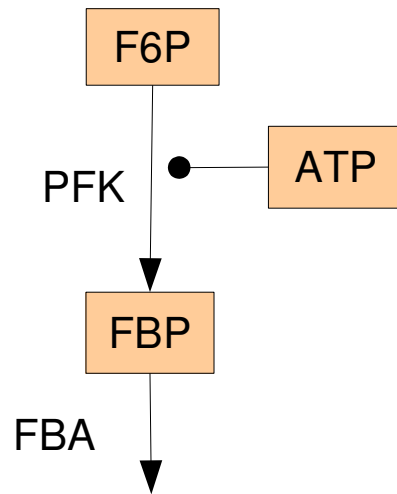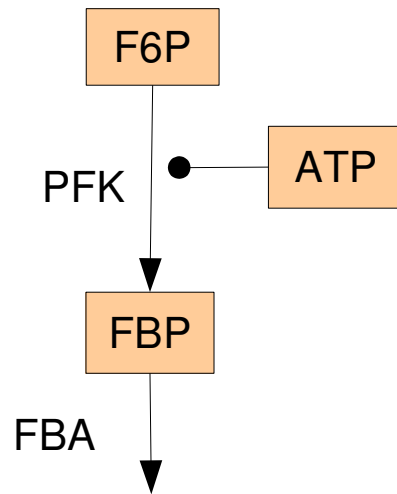
If every single statement **is correct**, then the model **is correct**.

If every single statement **is valid**, then the model **is valid**.

Concatenation of models can lead to conflicts
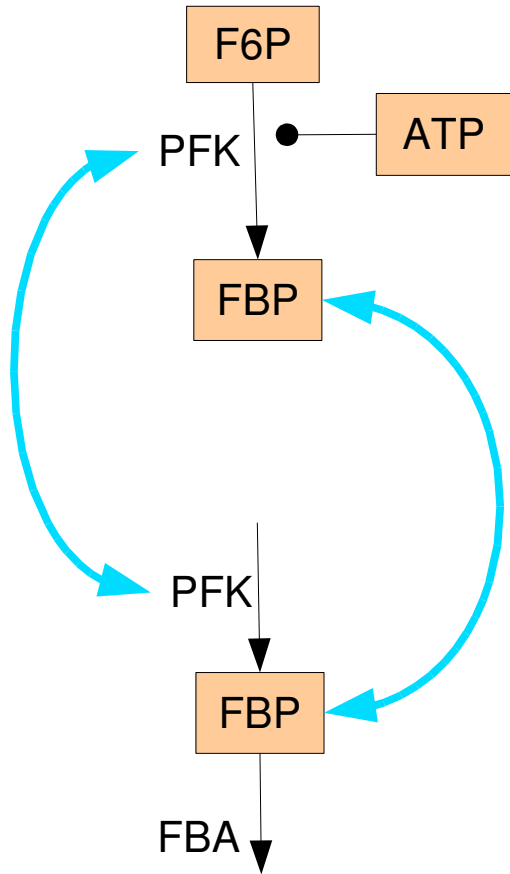
Model 1

F6P conc. [mM]	c_{F6P}	$C_{F6P} = 0.1$
ATP conc. [mM]	c_{ATP}	$C_{ATP} = 0.5$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$

Model 2

PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = 0.5$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK-} v_{FBA}$ $c_{FBP}(0) = 0.1$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

No experiment is needed to see that the concatenated model is wrong !!

Does the model allow for deterministic simulation?

Mathematical dependencies

c_{F6P}

v_{PFK} ← c_{ATP}

c_{FBP}

v_{FBA}

← Appears in algebraic law

← Appears in rate law

Model statements

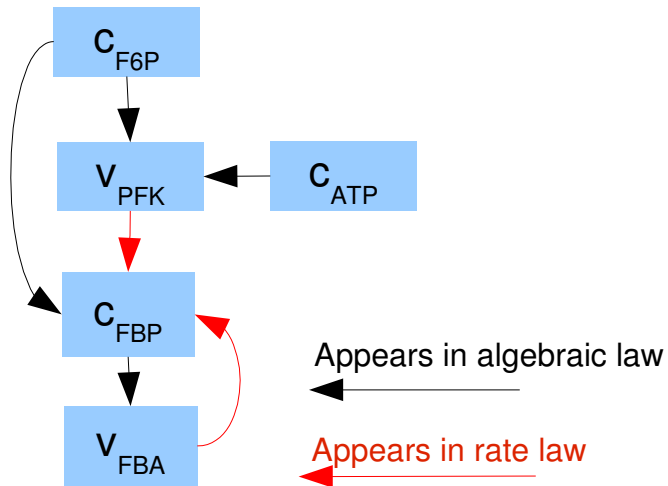|		
ATP conc. [mM]	c_{ATP}	$c_{ATP} = 0.5$
F6P conc. [mM]	c_{F6P}	$c_{F6P} = 0.1$
FBP conc. [mM]	c_{FBP}	$dc_{FBP}/dt = v_{PFK} - v_{FBA}$ $c_{FBP}(0) = c_{F6P}$
PFK vel.[mM/s]	v_{PFK}	$v_{PFK} = v_{PFK}(c_{F6P}, c_{ATP})$
FBA vel. [mM/s]	v_{FBA}	$v_{FBA} = v_{FBA}(c_{FBP})$

Simulation requires:

- Each variable must come with a statement
- Only one statement per variable
- Stepwise evaluation of algebraic assignments (no algebraic loops!)

Other mathematical frameworks have different requirements

Requirements for merging of rate equation models

1. Syntax

model can be read and processed

2. Computation

model can be used for predictive simulations

3. Semantic correctness

model statements agree with the model semantics

4. Empirical correctness

model agrees with physical and biochemical facts

5. Relevance

model performs well and suits its purpose

Requirements for merging of rate equation models

1. Syntax

model can be read and processed

2.

m

3

m

4.

m

5

m

Minimal validity criteria for merged models

- Correct syntax

- Consistent use of variable names (= element IDs)

- Model has to allow for deterministic simulation

- Same statements as in the input models (if possible)

- But: no conflicting statements

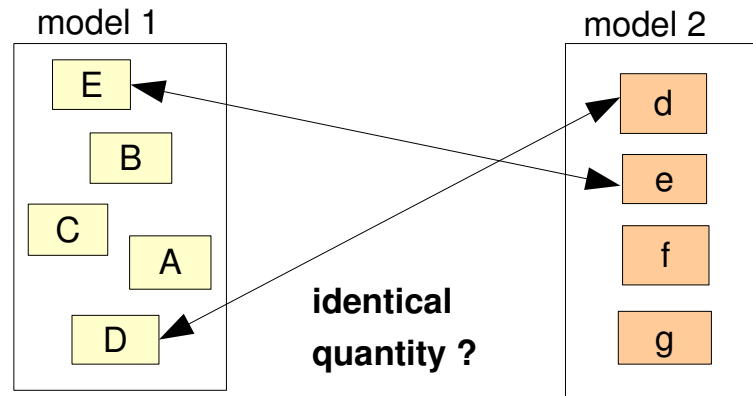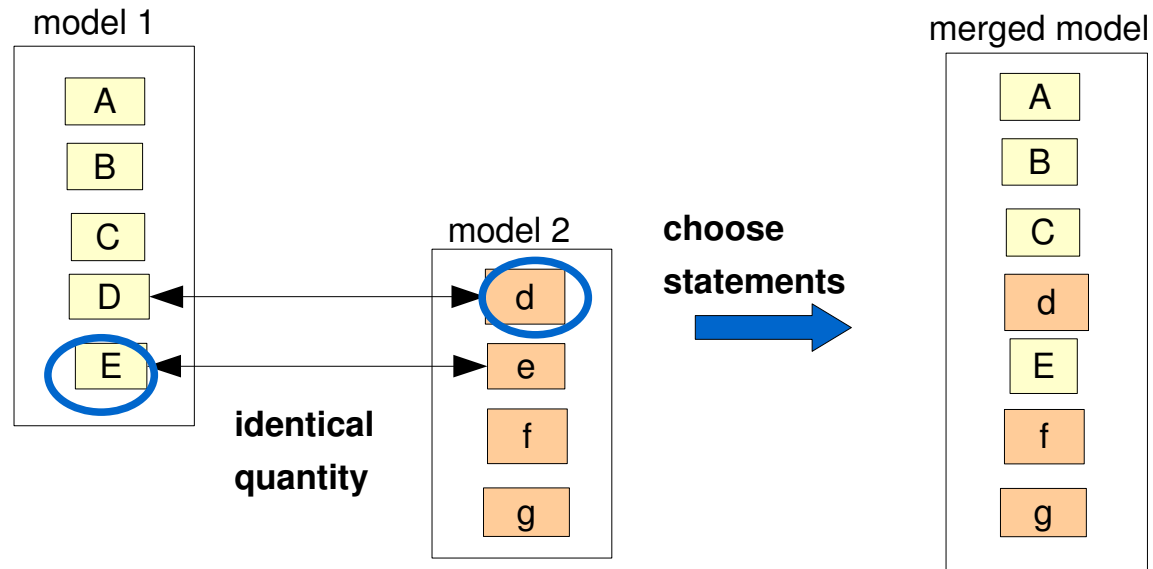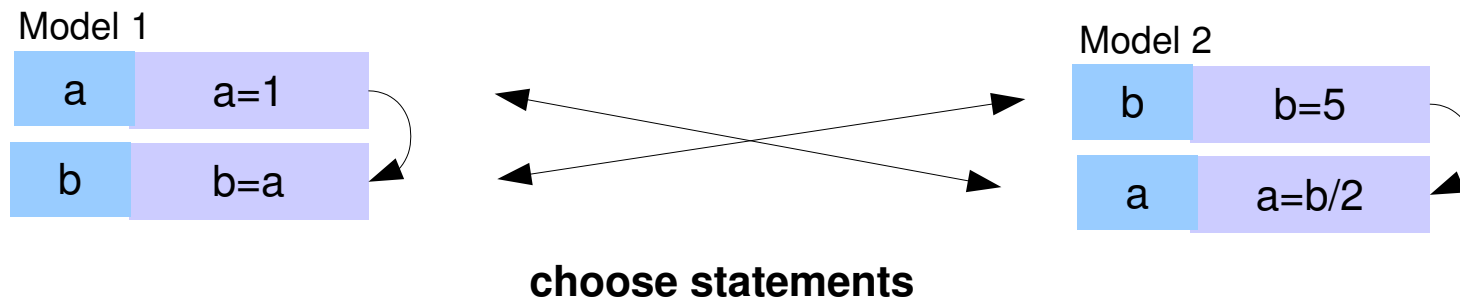**Precondition: input models must be correct!**

SBMLmerge, first version

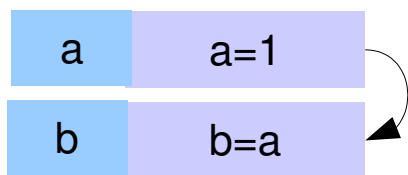## I. Compare the quantities
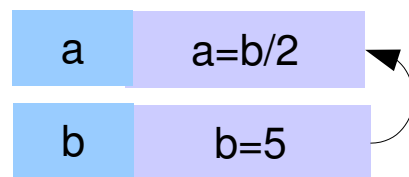

II. Choose between statements

Problem: avoid algebraic loops

Model 1

a	a=1
b	b=a

choose statements

Model 2

b	b=5
a	a=b/2

Possible choices:

Merged model
= model 1

a	a=1
b	b=a

Merged model
= model 2

a	a=b/2
b	b=5

Merged model
(allowed)

a	a=1
b	b=5
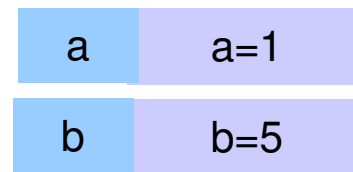
Merged model
(forbidden!!)
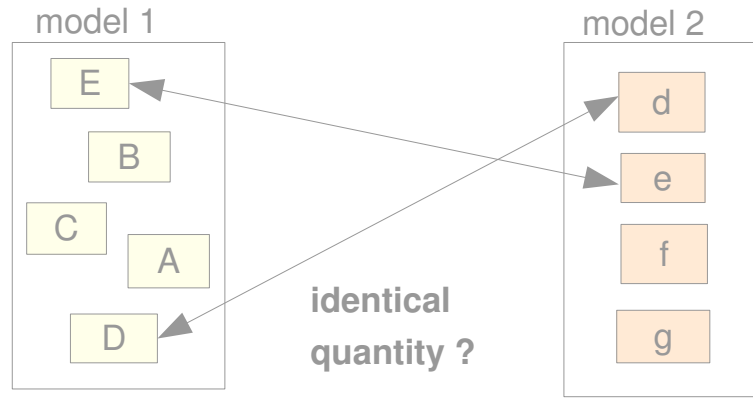
a	a=b/2
b	b=a
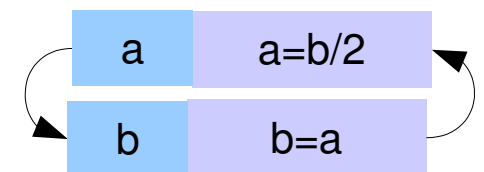
Problem: algebraic equations have no solution or cannot be evaluated
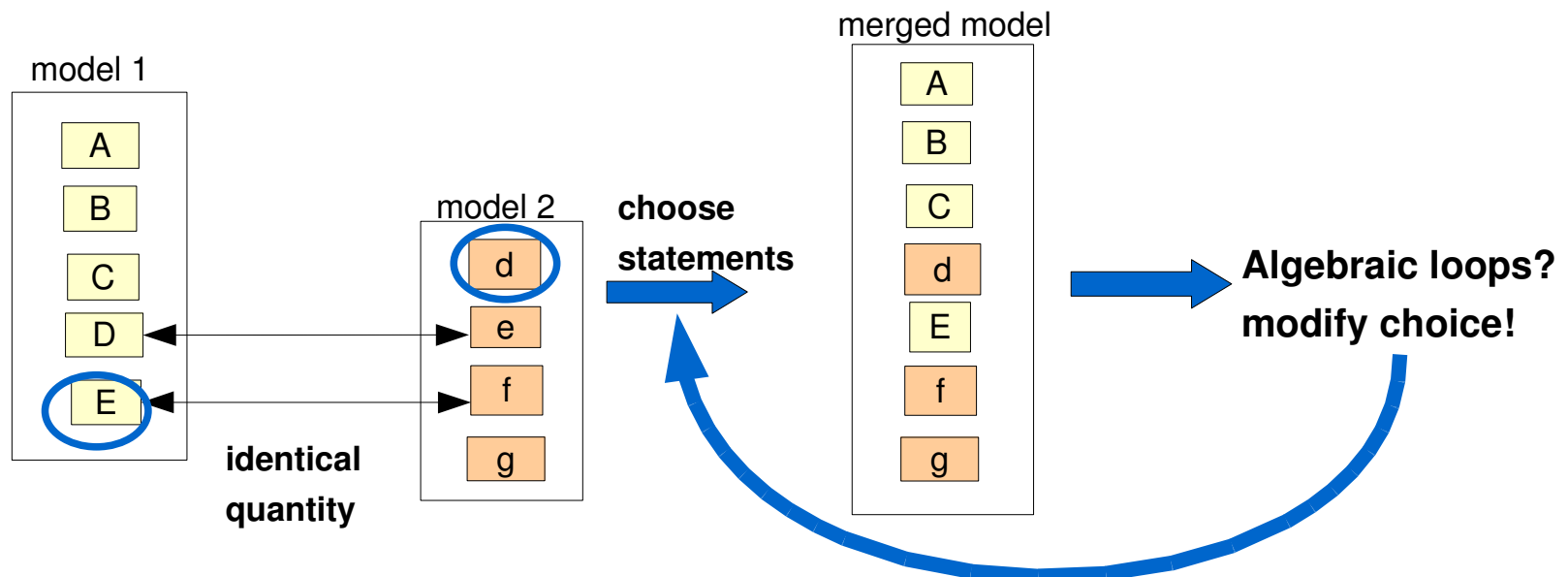
➡ **algebraic loops are forbidden!**
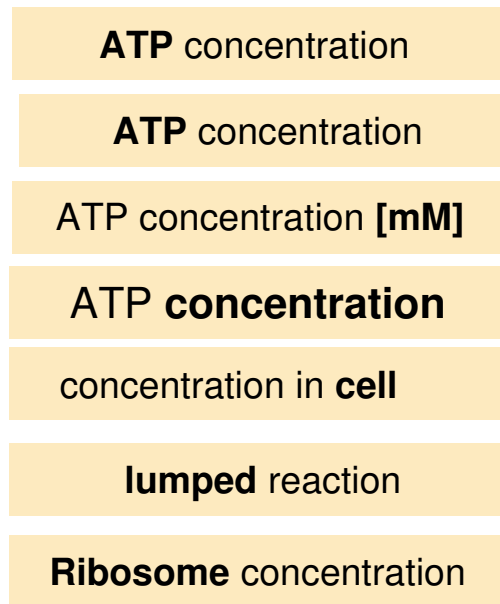
SBMLmerge, second version
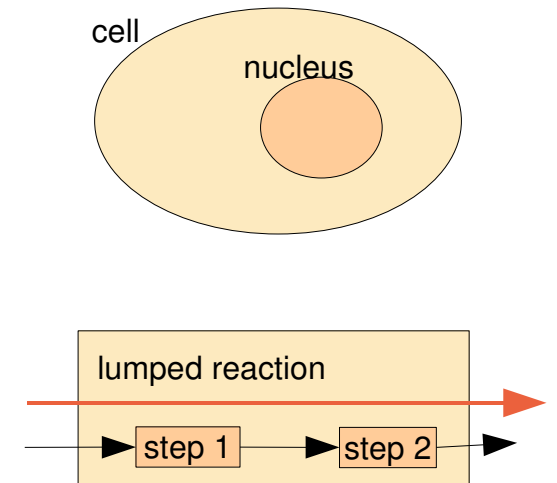
I. Compare the quantities

II. Choose between statements

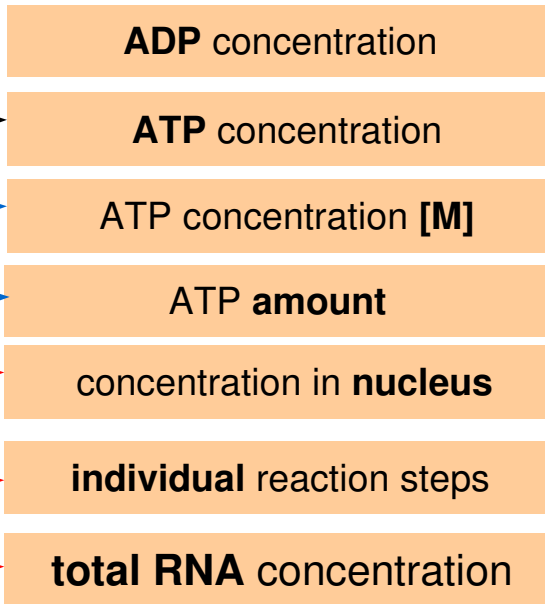# Another problem: avoid "semantic overlap"

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 |

cell

nucleus

lumped reaction
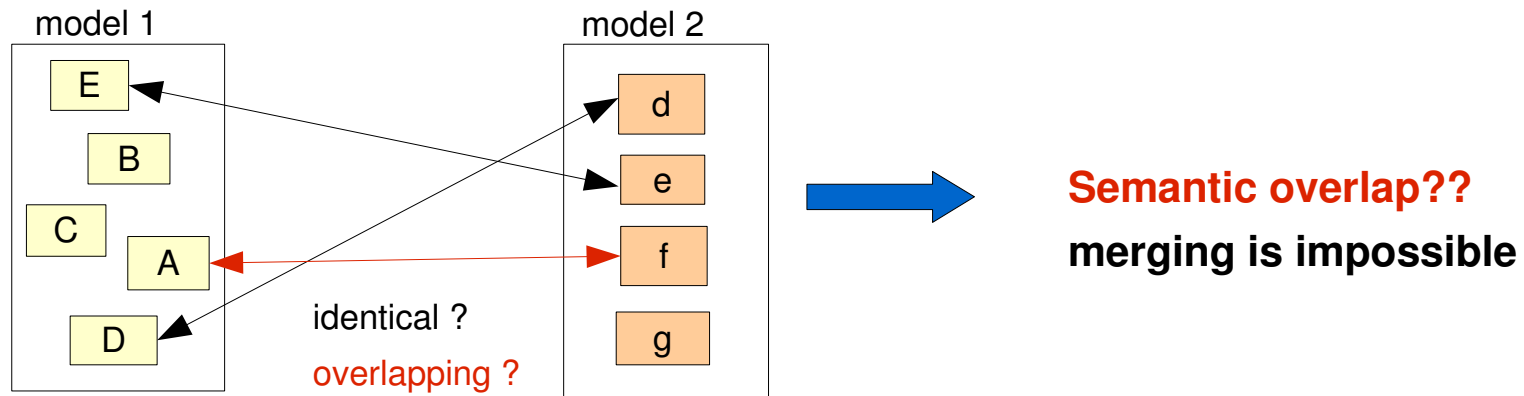
step 1   step 2

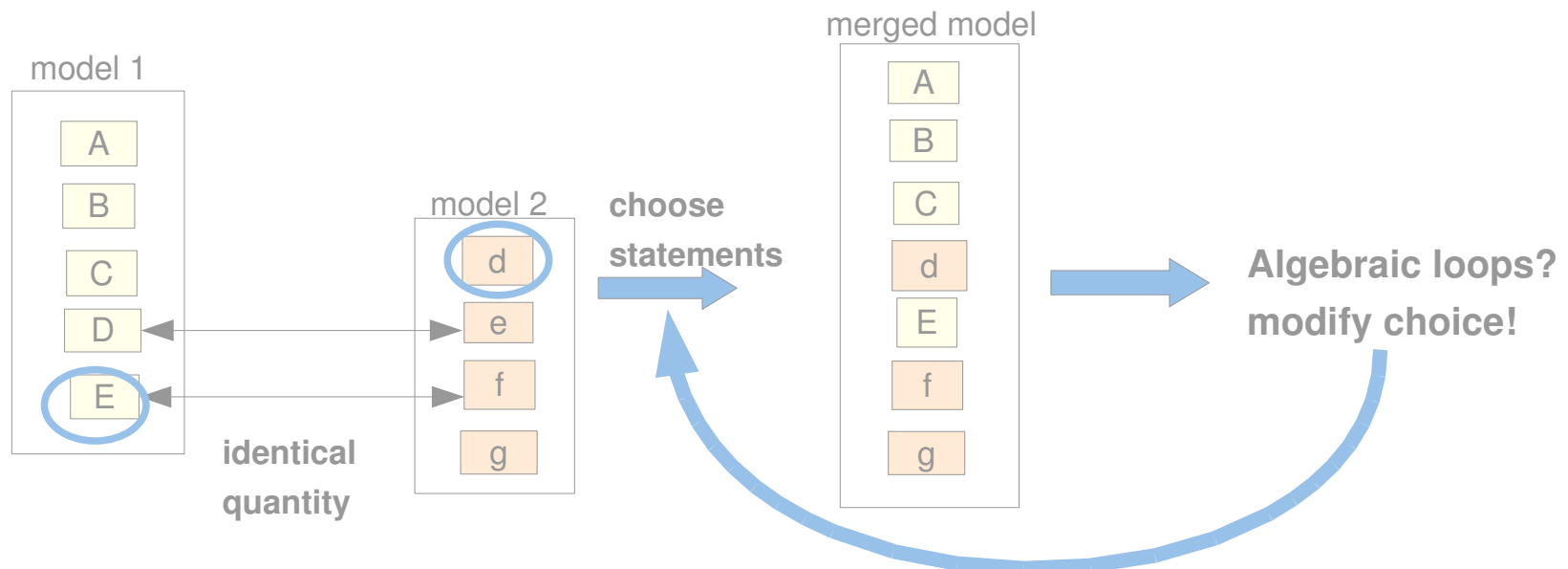## Relations between element pairs:

- 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

SBMLmerge, final version

I. Compare the quantities

Semantic overlap??
merging is impossible

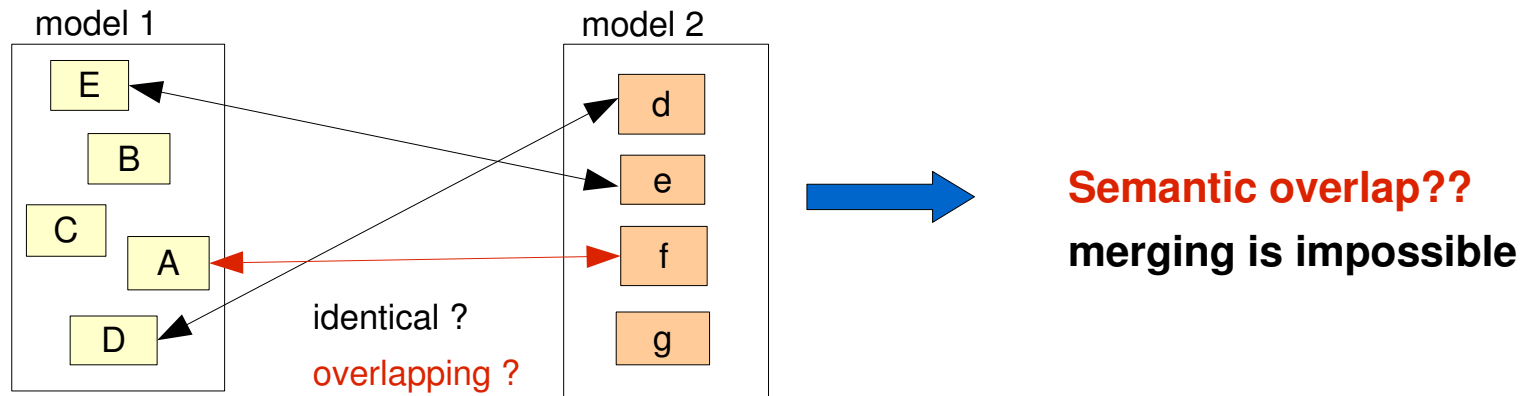## II. Choose between statements

SBMLmerge, final version

I. Compare the quantities

Semantic overlap??

merging is impossible

II. Choose between statements

choose statements
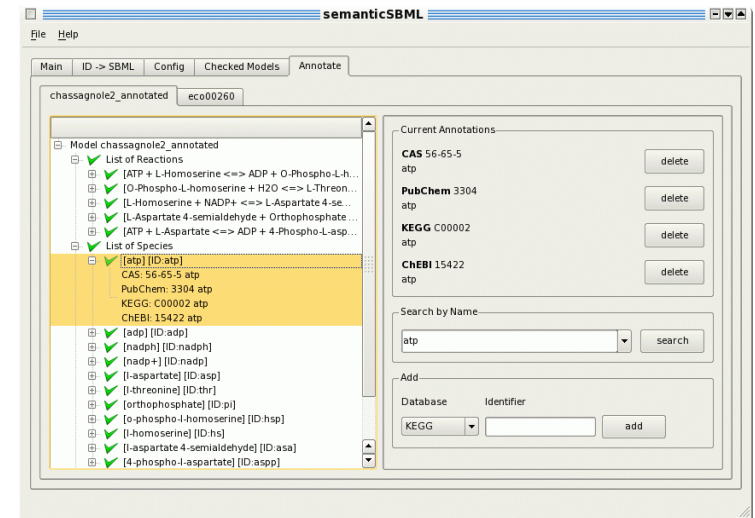
Algebraic loops?

modify choice!
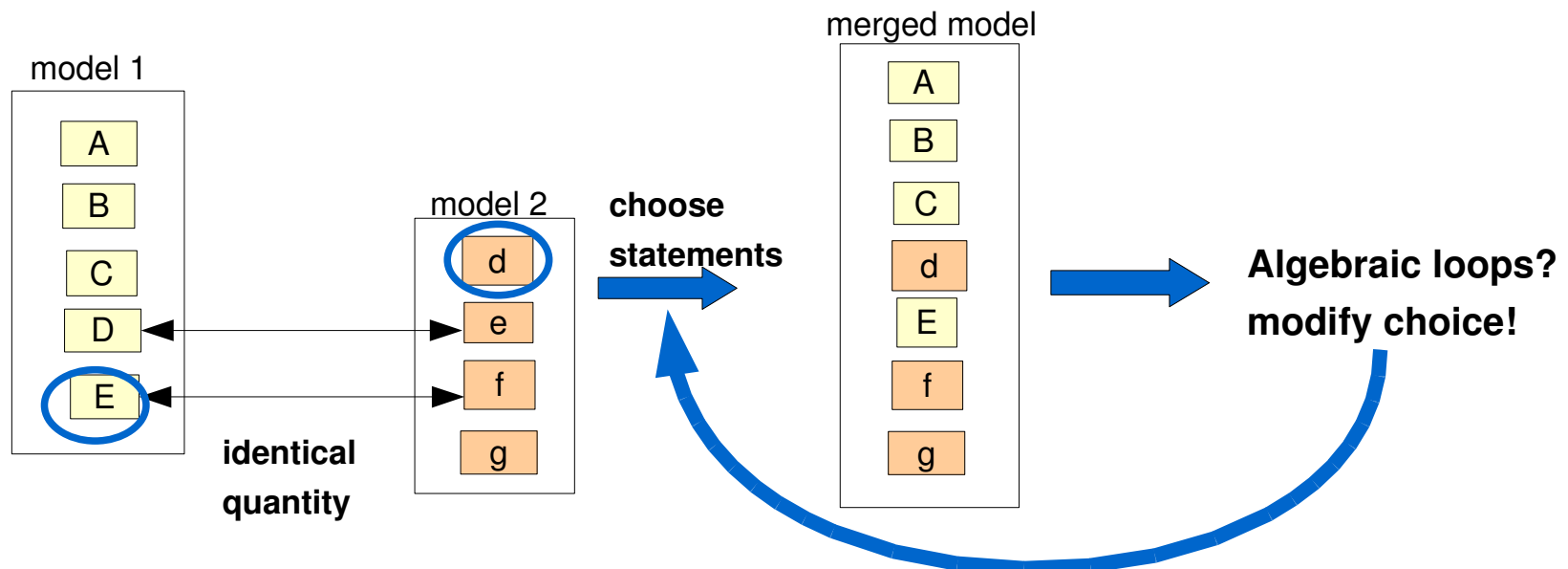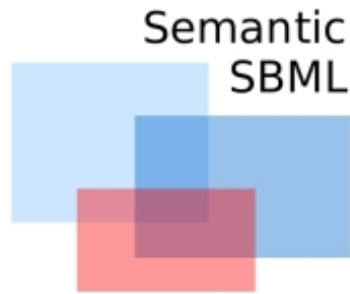
A tool for annotation, checking, and merging of SBML models

semanticSBML

- GUI + command line tool
- written in python, based on libSBML
- requires python2.4, libSBML 2.3.4, QT4, graphviz

  http://sysbio.molgen.mpg.de/semanticsbml/

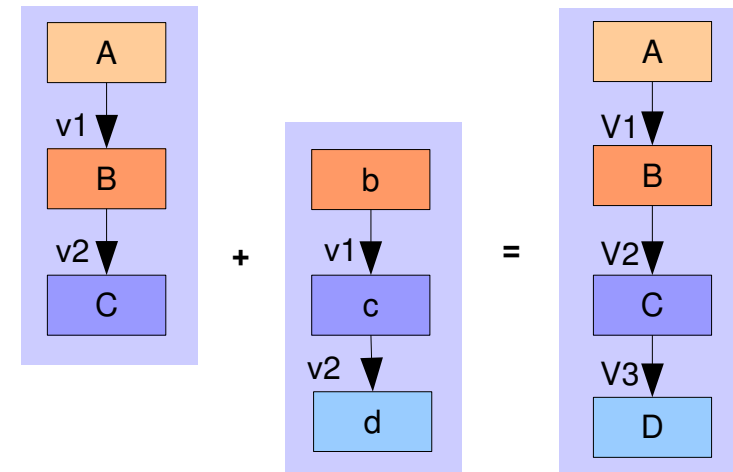| **SBMLannotate** | Check, add, and modify annotations ("MIRIAM" style annotation tags) |

| **SBMLcheck** | Apply model checks based on annotations (e.g., balances of atom numbers) |

| **SBMLmerge** | Merge two SBML models • detect conflicts between them • resolve conflicts or produce a warning |

Summary: model merging

Model merging: difficulties and solutions

Incompatible names   compare elements by annotations (SBML: "MIRIAM" style) , not by names

Comparing the elements   find duplicates and conflicting elements; use database for comparisons

Conflicting elements   must not appear together in a model -> abandon merging

Contradicting statements for duplicates   user chooses between conflicting statements

Algebraic loops   do not allow algebraic rules; constrain the choice between statements

Guidelines for preparing reusable models

Make your model accessible (see MIRIAM proposal)

- publish all information that is necessary to reproduce the fitting, simulations, etc

- provide the model in a standard (preferably free) format (SBML is perfect)

- submit the model to a repository like **JWS online** or **biomodels.net**

Explain the meaning of model elements

- use IDs or unambiguous terms to describe the substances and reactions

- put the annotations into the model code (following the MIRIAM proposal)

Construct the model such that its parts will still work after merging

- globally fitted parameters (in top-down modelling) may loose their meaning after merging

- lumped reactions, metabolites should be carefully described

Concerning the experiments:

support your local standardisation effort (EU projects??) & STRENDA!!

Acknowledgements

MPI Berlin

Edda Klipp

Falko Krause

Jannis Uhlendorf

Marvin Schulz

Anselm Helbig

Universität Heidelberg

Ursula Kummer

Sven Sahle

Ralph Gauges

EML Heidelberg

Isabel Rojas

Martin Golebiewski
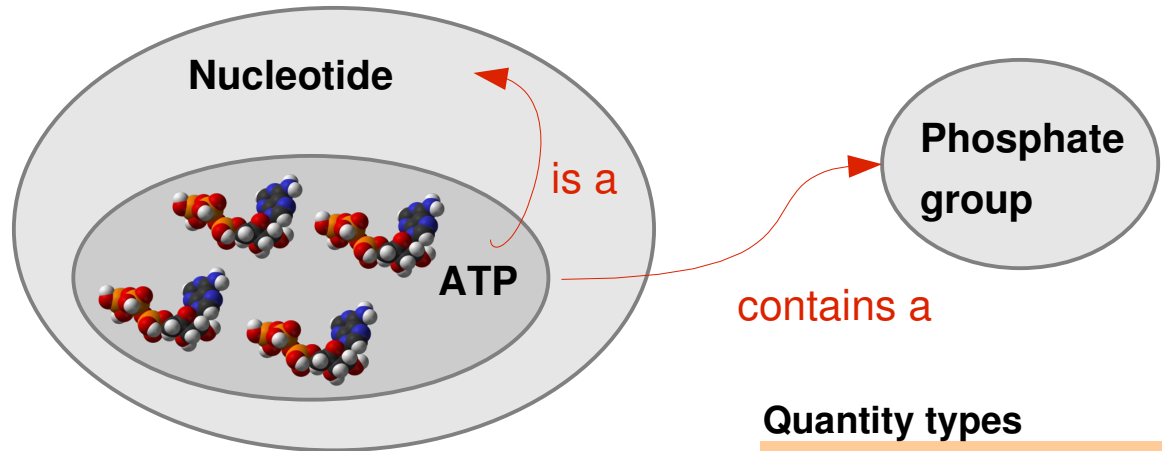
MPI Magdeburg

Martin Ginkel

Michael Rempel

BaSysBio

Thank you
for your attention !!

Biochemical objects, entities, and quantities

BIOLOGICAL ENTITIES

Substances
Reactions
Compartments
States
Events

Nucleotide

ATP

is a

contains a

Phosphate group

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 entities

Semantic dependence between (different) biochemical entities, alternative criteria:

- "Each instance of A is also an instance of B or vice versa"
  (ATP is a nucleotide)

- "The presence of an instance of A implies the presence of an instance of B"
  (Ribosomes in cell -> RNA in cell)

- For compartments: "Compartments A and B are spatially overlapping"
  (nucleus is within the cell)
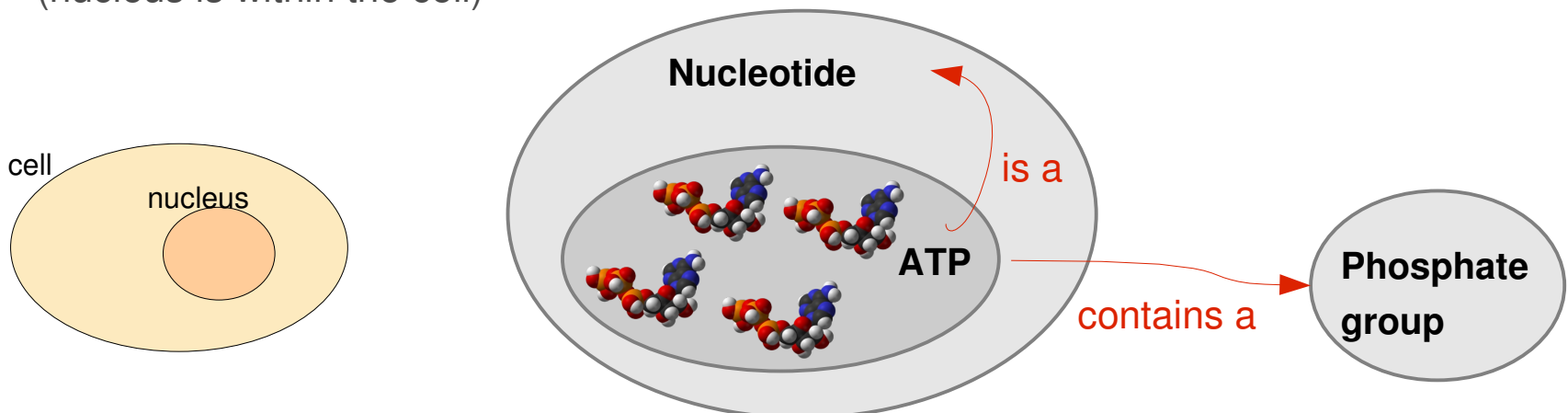
Semantic dependence between biochemical quantities

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
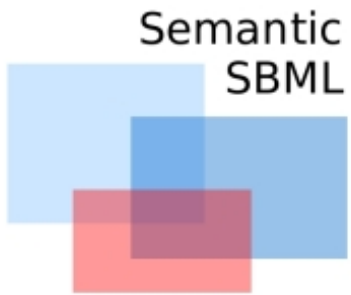
Handling annotations in SBMLannotate