

2009-01-21

Jukka Matilainen
jukka.matilainen@medicel.com

Pathways

Network Models in Medicel Integrator

Contents

- **Motivation**
- **Creating** pathways
 - manually (by the Pathway application)
 - programmatically (by workflows)
- **Accessing**, querying and navigating pathways
 - finding instances of motifs
- **Computing**
 - Simulation, pathway analysis, ..

Motivation: Pathways in Integrator

Models

”all models are wrong, but some are useful” – George Box

- Computer-readable models of biological systems for
 - describing the system under study
 - systematic documentation of molecular and cellular functions
 - predicting system behaviour
 - new hypotheses
- Models are needed as a basis for inference

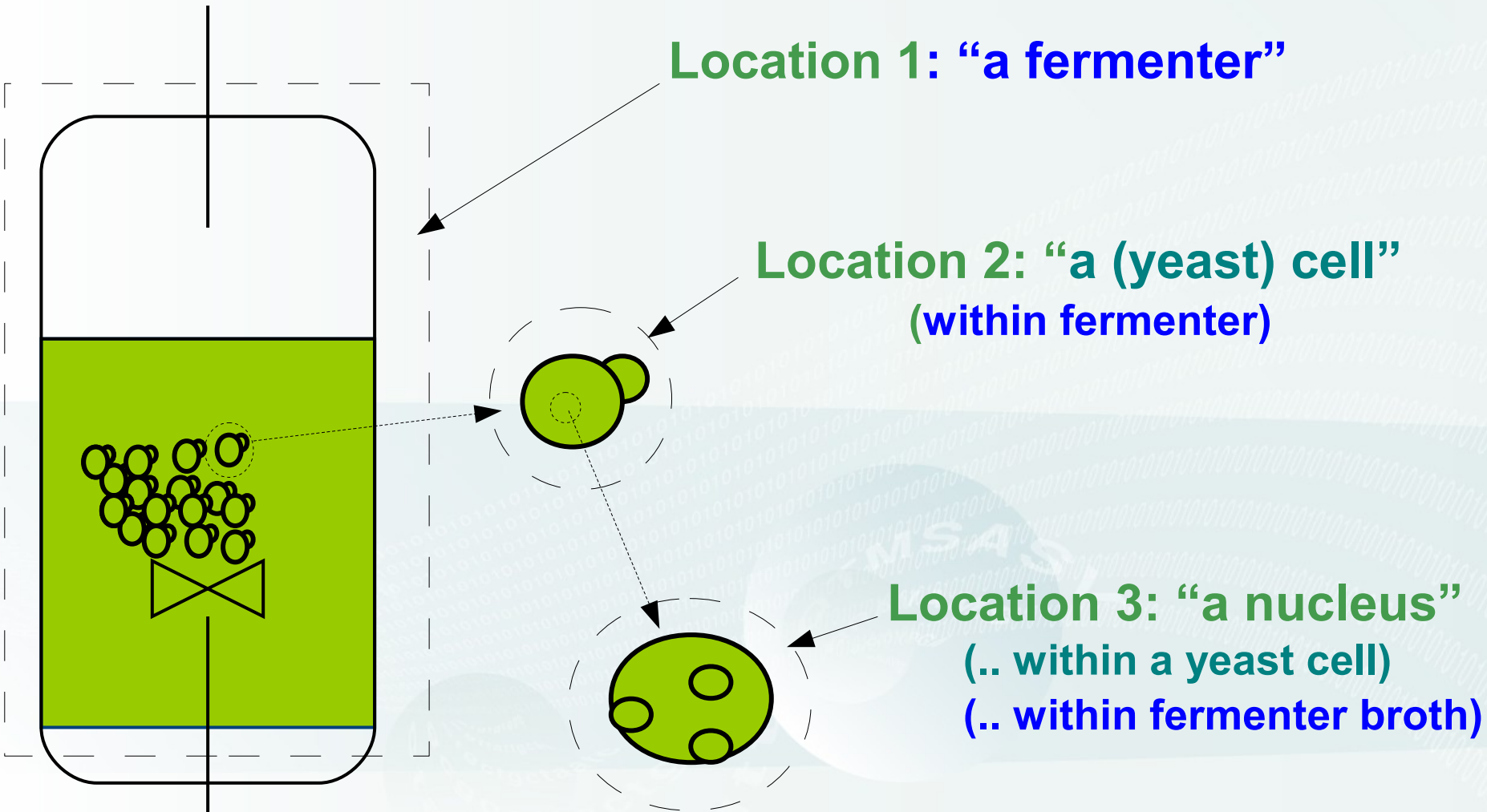
Network Models: Pathways

- Generic conceptual model for representing systems
 - supports different abstraction levels
 - not limited to describing molecular/cellular systems
- Capture general systems theory concepts: system-environment boundary, input, output, process, state, hierarchy...
- The pathway model is the central way to represent biological knowledge in Medicel Integrator

Location

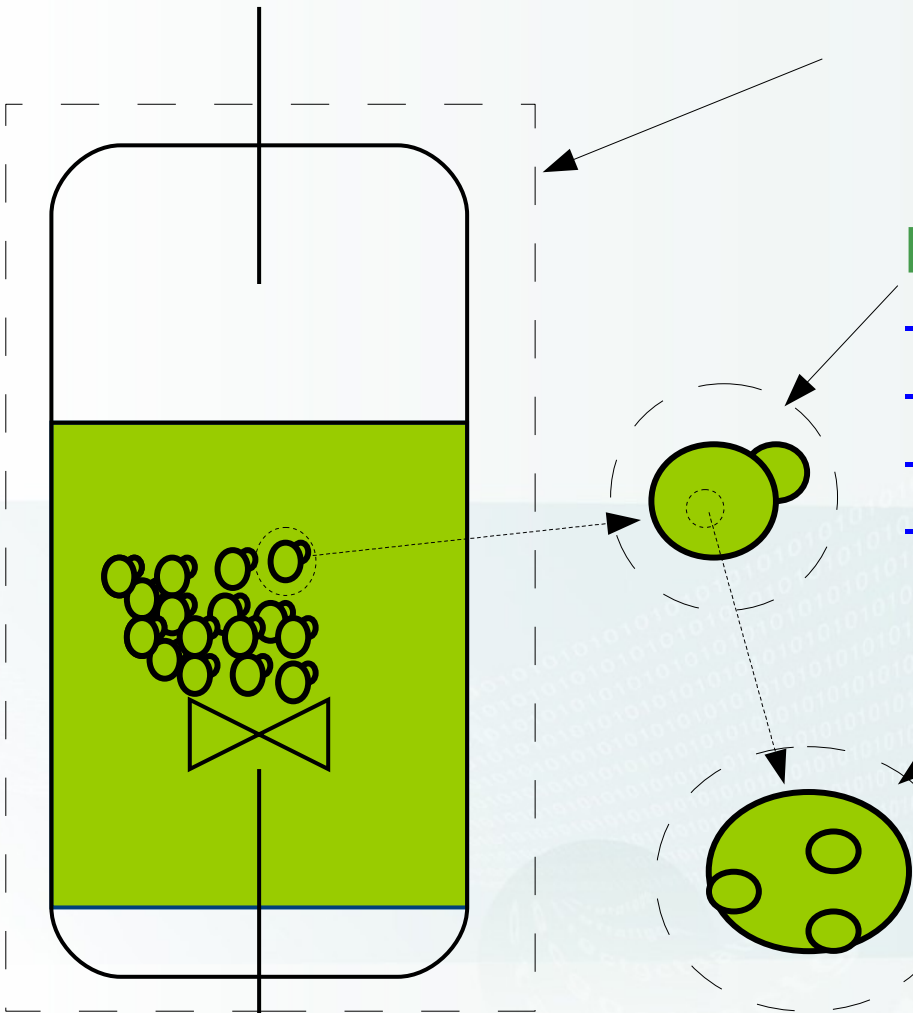
- represents a biological system
- has a description and a definition:
 - is described with name
 - is defined with unique combination of:
 - Environment
 - Population
 - Individual
 - Organism
 - Organ
 - Tissue
 - Cell type
 - Cellular compartment
- a unique *location* required for each distinct context of interest

Examples of locations



Location binds related measurement data together

Examples of locations (continued)



Location1:

- environment = fermenter

Location2:

- environment = fermenter
- population = cultivation #102
- organism = *Saccharomyces cerevisiae*
- cell type = yeast cell

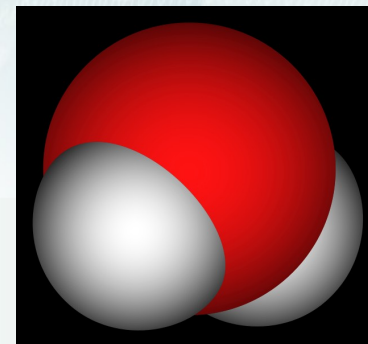
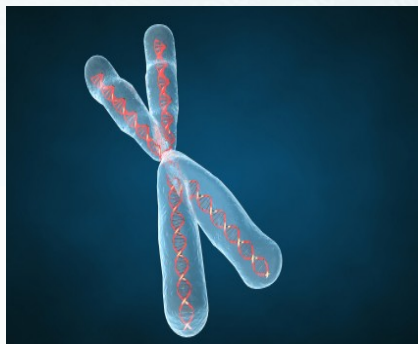
Location3:

- environment = fermenter
- population = cultivation #102
- organism = *Saccharomyces cerevisiae*
- cell type = yeast cell
- cellular compartment = nucleus

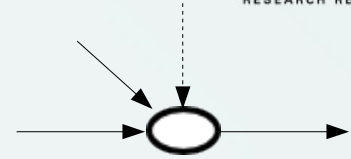
- *Locations* can be related to each other hierarchically

Components

- Various kinds of components
 - *genes, transcripts, proteins, compounds, macromolecular complexes...*
 - but also, at higher levels: *cell types, individuals, organisms, populations, environments*
 - not limited to molecular components
 - the same systems representation can be used for modeling physiology or entire ecosystems



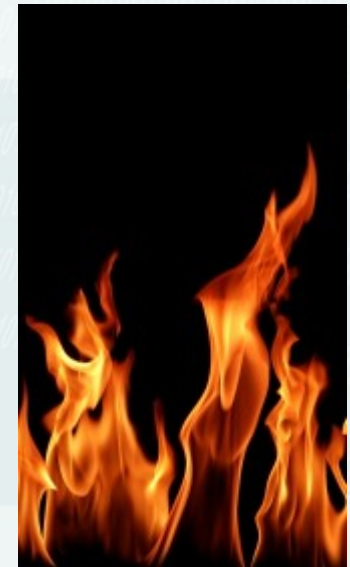
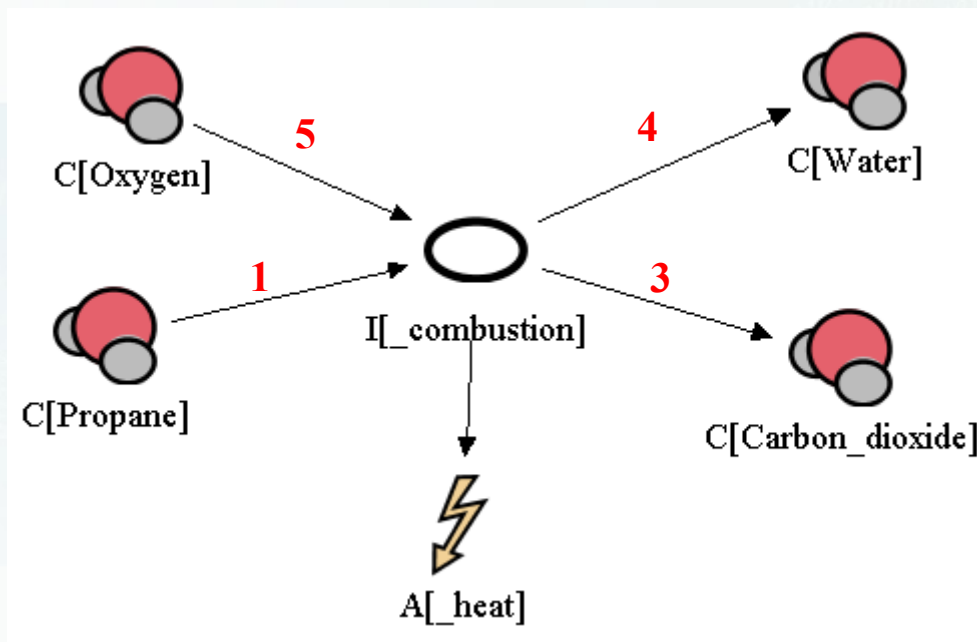
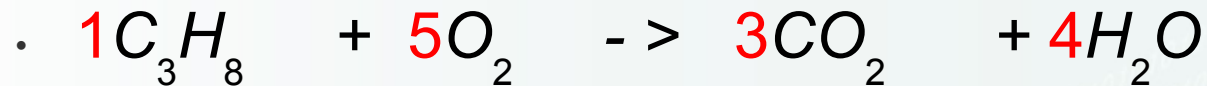
Interaction and connections



- define an event (or a process)
 - typically, a biochemical event, e.g. transcription, translation, chemical reaction, ...
- Components are connected to *interactions* via *connections*
 - different types of *connections*:
 - substrate (is consumed)
 - product (is produced)
 - control (is neither consumed or produced, but affects)
 - outcome (not consumed or produced, but affected)

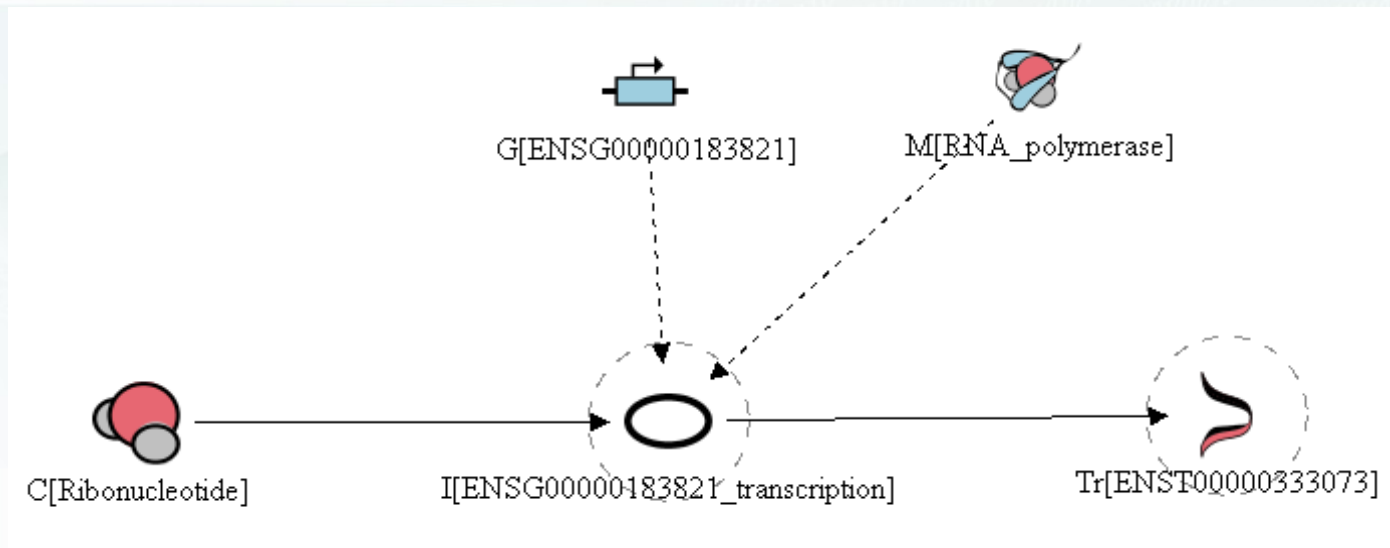
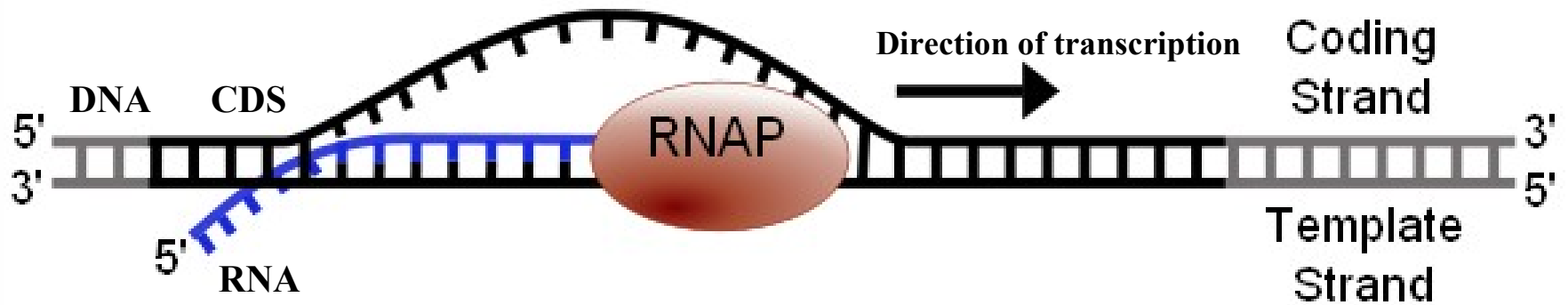
Example: A reaction

- Combustion (burning of propane)



Stoichiometric coefficient is a property of the connection

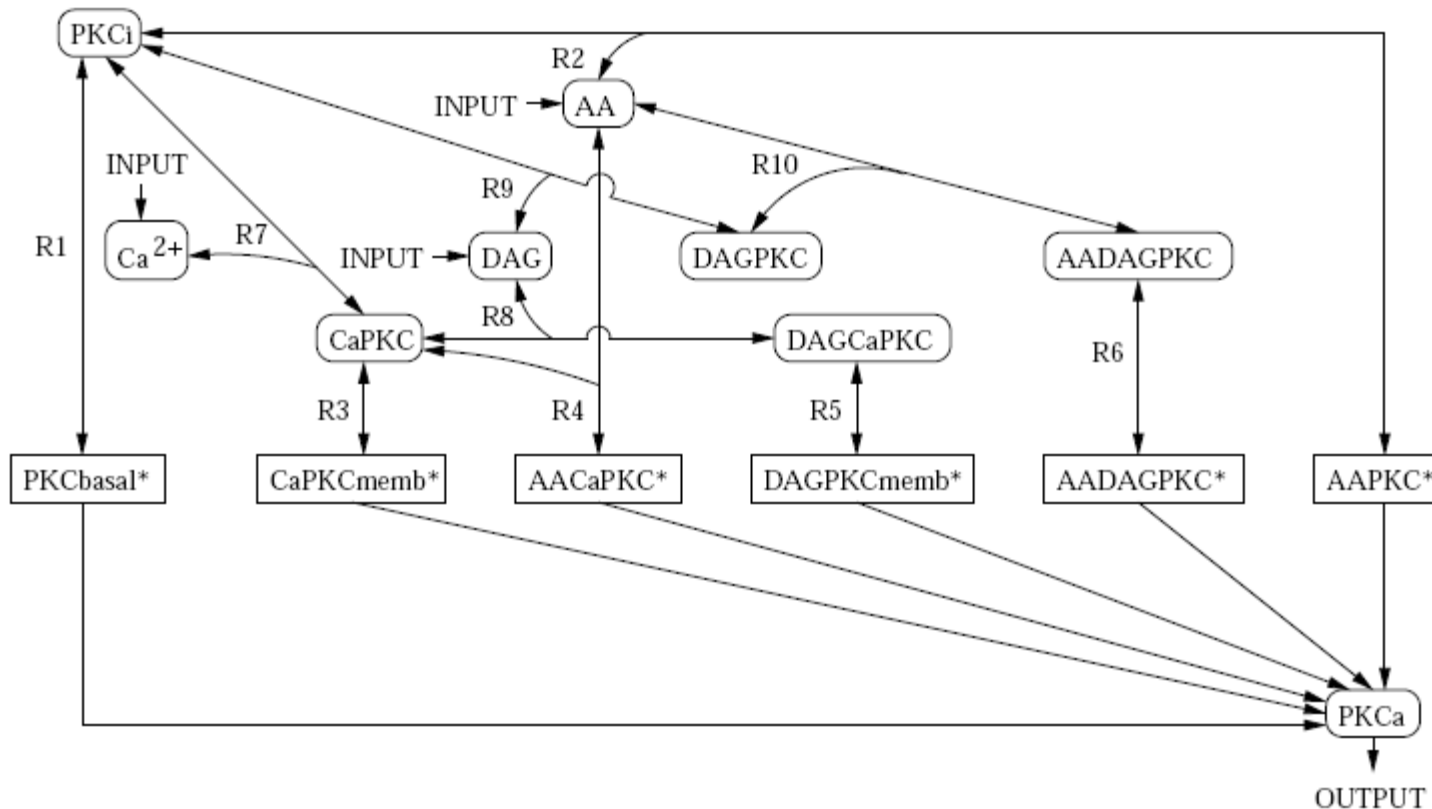
Example: Gene transcription



Pathway

- a network model of a biological system
 - a container for the components and interactions
- there can be multiple pathways for one location
 - for different modeling purposes
 - for example, one pathway of a system might contain kinetic models for ODE (Ordinary Differential Equation) simulation
 - another might contain constraints for flux balance analysis
 - at different abstraction levels
 - alternative models from
 - different origin,
 - different creators,
 - different evidence

Example: A PKC signaling pathway



- Manninen, T., Saarinen, A. & Linne, M-L., Simulation study of differential equation model for protein kinase C signaling, Tampere University of Technology, Institute of Signal Processing, Report, pp. 20 p, (2004)

Hierarchical systems

- a *pathway* can contain other *pathways*
 - typically, the contained *pathways* have different locations associated with them
 - for example, a pathway representing nucleus within a cytoplasm pathway
- in modeling, typically each component has a different pool in each *location*
 - for example, different concentrations in nucleus and cytoplasm for the same molecule species
- interactions can be shared between *pathways*
 - translocation interactions, for moving biomaterial from one location to another

Kinetic laws

- formulas relating the interaction to the state of the system can be attached to an interaction, e.g.
 - equation for the interaction rate, or
 - a set of constraints for the interaction rate, or
 - other information
 - such as a boolean expression telling whether the interaction is feasible given the presence of the products, controls and substrates
- these can be automatically collected from the pathway into a computational model, like
 - an ODE model or a Boolean network for simulation,
 - a constraint set for optimization, etc.

Example: A reaction with kinetics

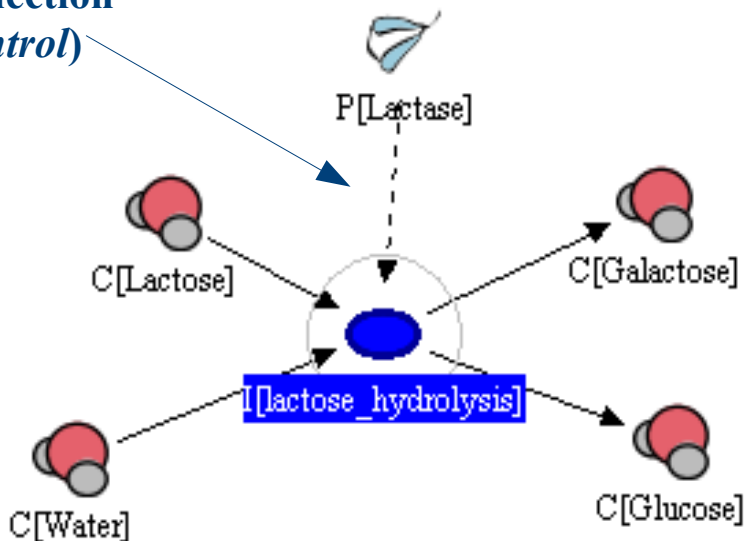
- hydrolysis of lactose into glucose and galactose
lactase



- rate of reaction (Michaelis-Menten kinetics):

$$v = \frac{k \cdot [\text{lactase}] \cdot [\text{lactose}]}{K_m + [\text{lactose}]}$$

connection
(control)



```
Pw[example_of_lactose_hydrolysis]L[N/A]I[lactose_hydrolysis]
V[rate]U[mol/L/s]I[lactose_hydrolysis]=v
v=k1*c1*s2/(k2+s2)
k1=2000
k2=4e-3
s1 = V[concentration]U[mol/L]C[Water]
s2 = V[concentration]U[mol/L]C[Lactose]
c1 = V[concentration]U[mol/L]P[Lactase]
p1 = V[concentration]U[mol/L]C[Glucose]
p2 = V[concentration]U[mol/L]C[Galactose]
```

Creating: Constructing pathways

Sources of pathway data

- External databases (via database population)
- Published models
 - importing models automatically
 - BioPAX (Reactome)
- Results of analysis workflows
 - hypotheses
 - verified results
- Manual entry
 - e.g. “text book knowledge” represented as pathways

Medicel Reference Model

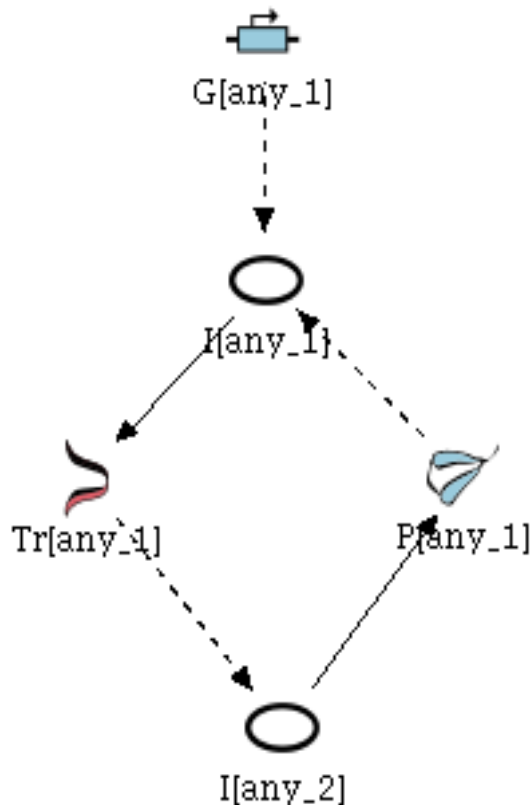
- Medicel reference objects
 - Set of bioentities where information was merged from different sources: genes, transcripts, proteins...
 - identifying the “same” entities – an identity criteria needed for each type of entity
- Medicel reference pathway
 - data in heterogeneous formats interpreted into pathway format and gathered together
 - protein complexes as interactions between proteins (assembly/disassembly)
- Custom reference models can be built in the same way

Accessing: Querying pathways

Pathway analysis

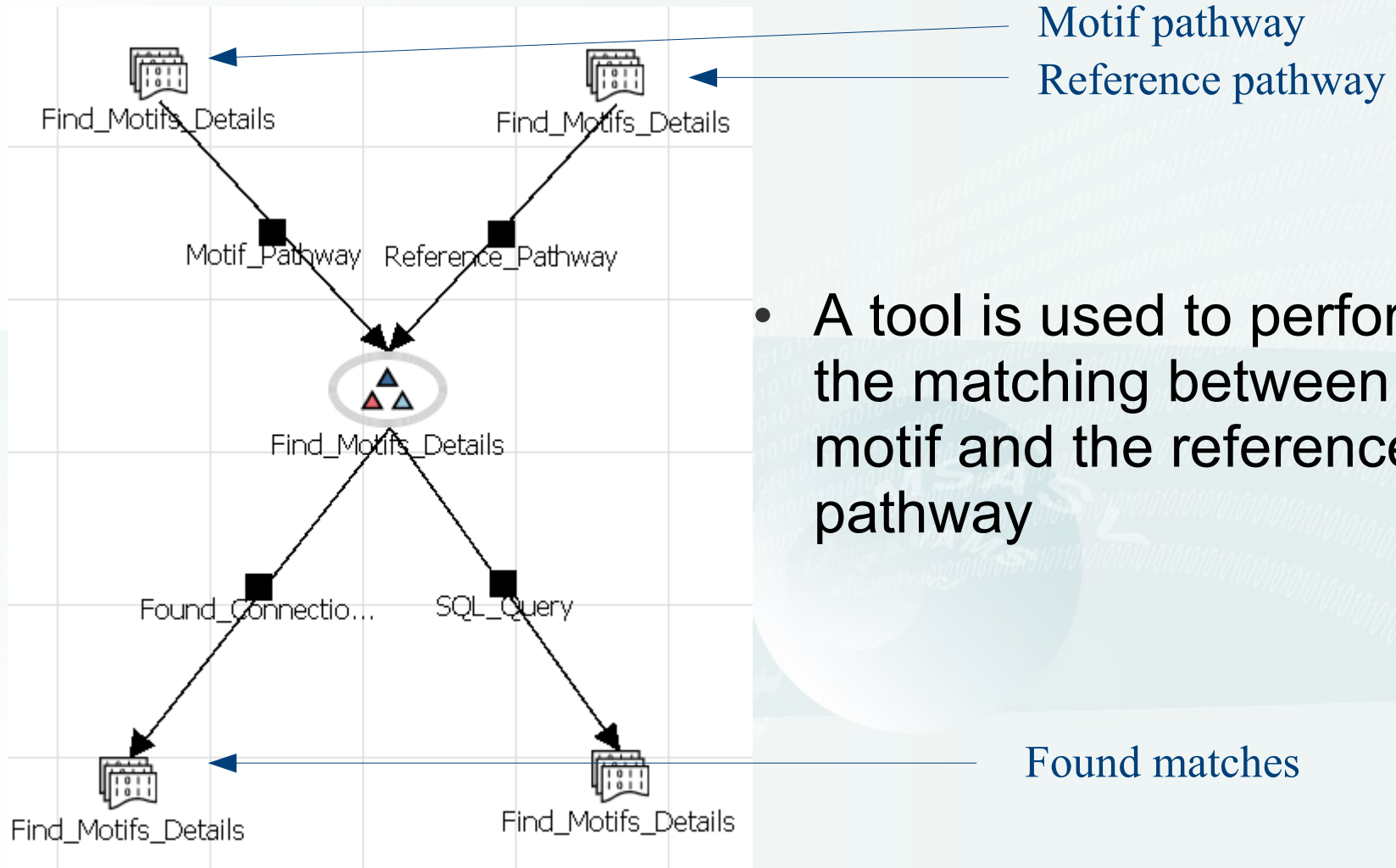
- Custom pathways can be created by the user
 - The process can be automated by creating a workflow which reads in the source pathways, performs the necessary modifications and transformations and writes the result pathway to the database
- Network motifs
 - Identifying recurring patterns in pathway topology
 - Finding instances of patterns from pathways
 - Creating pathways by creating instances of motifs

Motif example: autoregulation



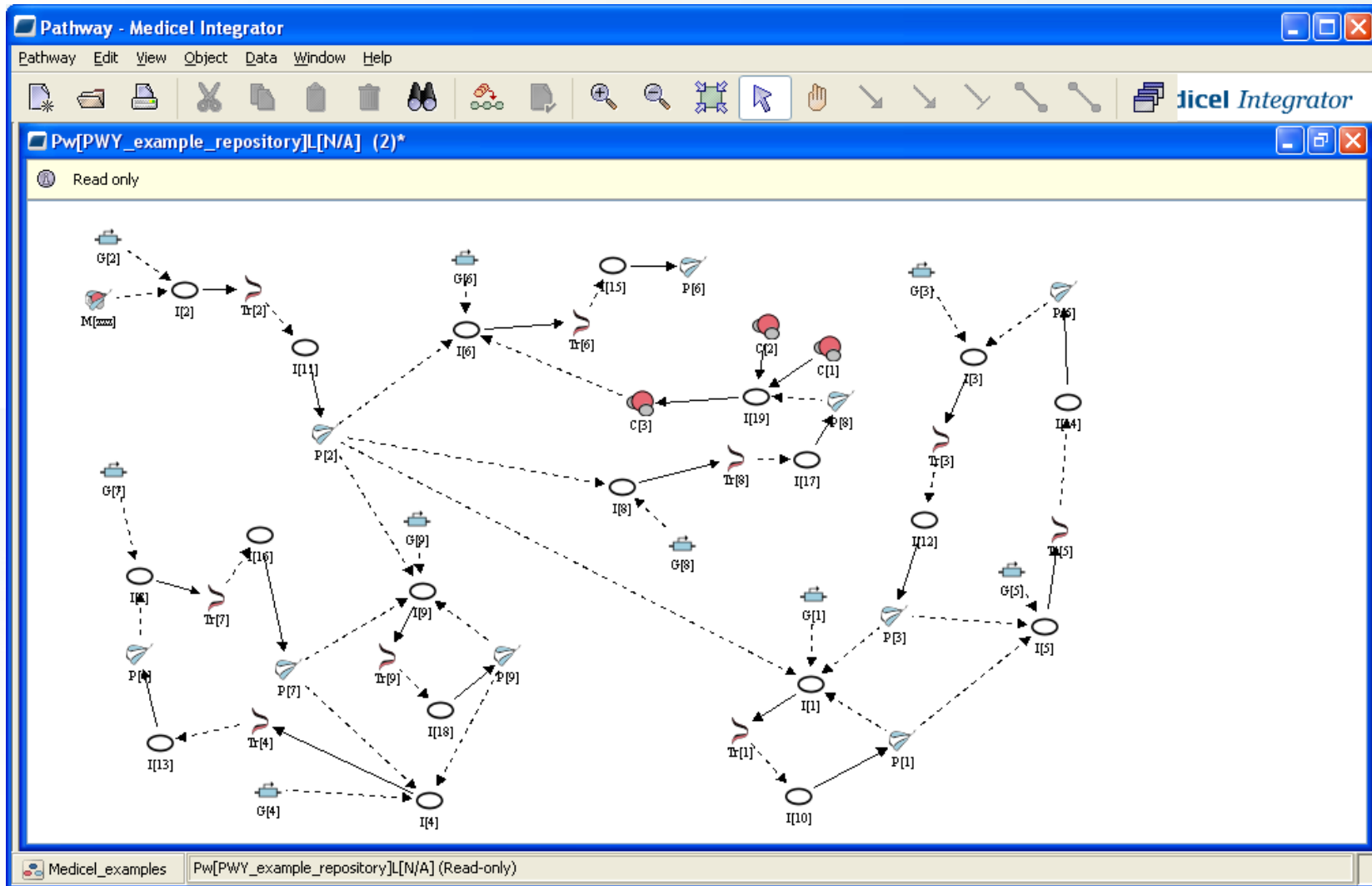
- Placeholder components (named with *any**) are used as placeholders for biological components and interactions
- This motif can be used to match autoregulation from a transcriptional regulatory network

Motif example: matching

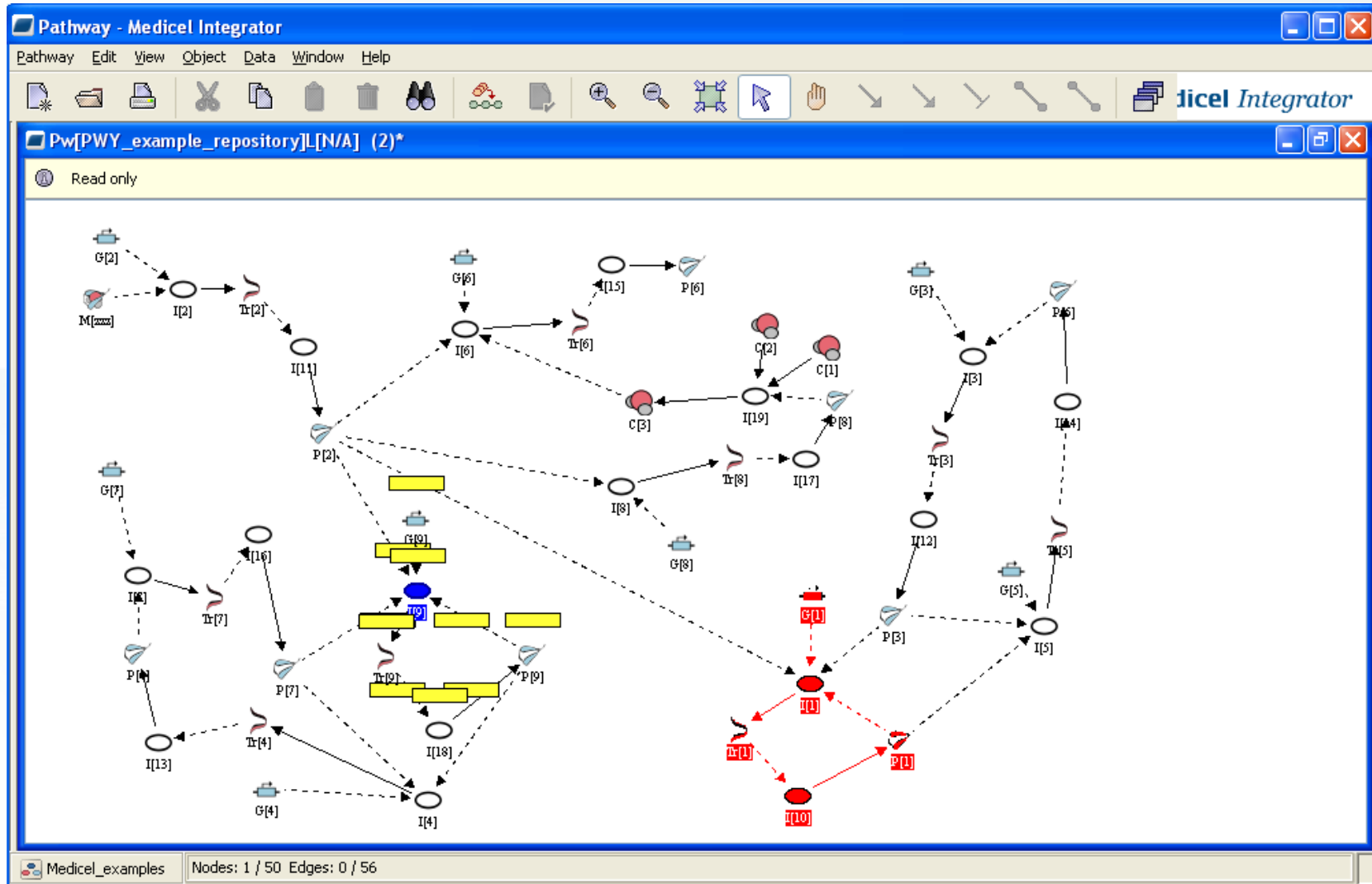


- A tool is used to perform the matching between the motif and the reference pathway

Motif example: reference pathway



Motif example: visualizing matches



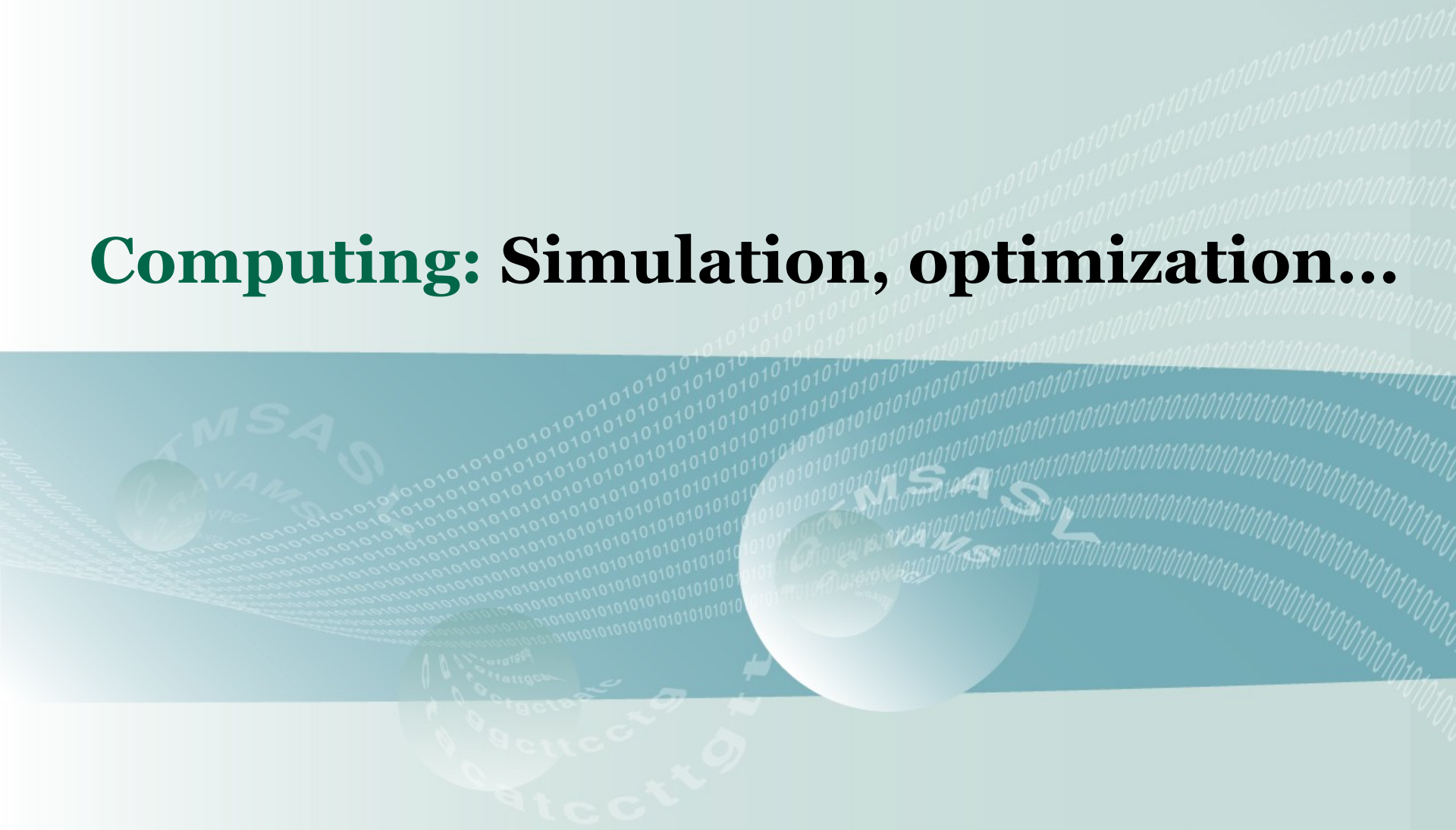
Querying pathways by motif

- What compound is used as a substrate when a specific product xyz is produced?
- What compound is produced when a specific substrate xyz is consumed?
- What protein is an enzyme when a specific product xyz is produced?
- What genes are regulated by specific protein?

Constructing pathways by motif

- Also can use networks motifs as building blocks to create pathways
 - Use a pathway motif as a template; fill in the variables.

Computing: Simulation, optimization...



Utilizing system data

- Uses of integrated systems descriptions
 - capturing and documenting knowledge
 - both systems data brought in from external sources
 - and results of own analysis
 - modeling the systems quantitatively
 - dynamic models (ODEs, BNs...)
 - FBA, etc.
 - plug in your own modeling software
 - analysis of topology
 - discovering topological patterns
 - finding instances of patterns
 - “converting” names (protein->transcript etc.)

Modeling: ODE models

- Each interaction can optionally have a specified kinetic law which specifies the kinetics of the reaction
 - typically, this is the interaction rate, i.e. how fast the interaction takes place
 - typically, the interaction rate is a function of concentrations of substrates, controls, and products
 - expressed in DT_RULES notation

Modeling: ODE models

The screenshot shows a software window titled "0: Pw[test_pkc1/pathway-2]L[N/A]I[test_pkc/R2]". The main text area contains the following code:

```

V[rate]U[mol/s]I[test_pkc/R2]=v
v=V[volume]*(k1*s1*c1 - k2*p1)
k1=120
k2=0.1
s1=V[concentration]U[mol/L]M[test_pkc1/PKCi]
c1=V[concentration]U[mol/L]M[test_pkc1/AA]
p1=V[concentration]U[mol/L]M[test_pkc1/AAPKC_]
    
```

Below the code is a "Feedback" section with the message "Law saved." and three buttons: "Check", "Save", and "Update Topology".

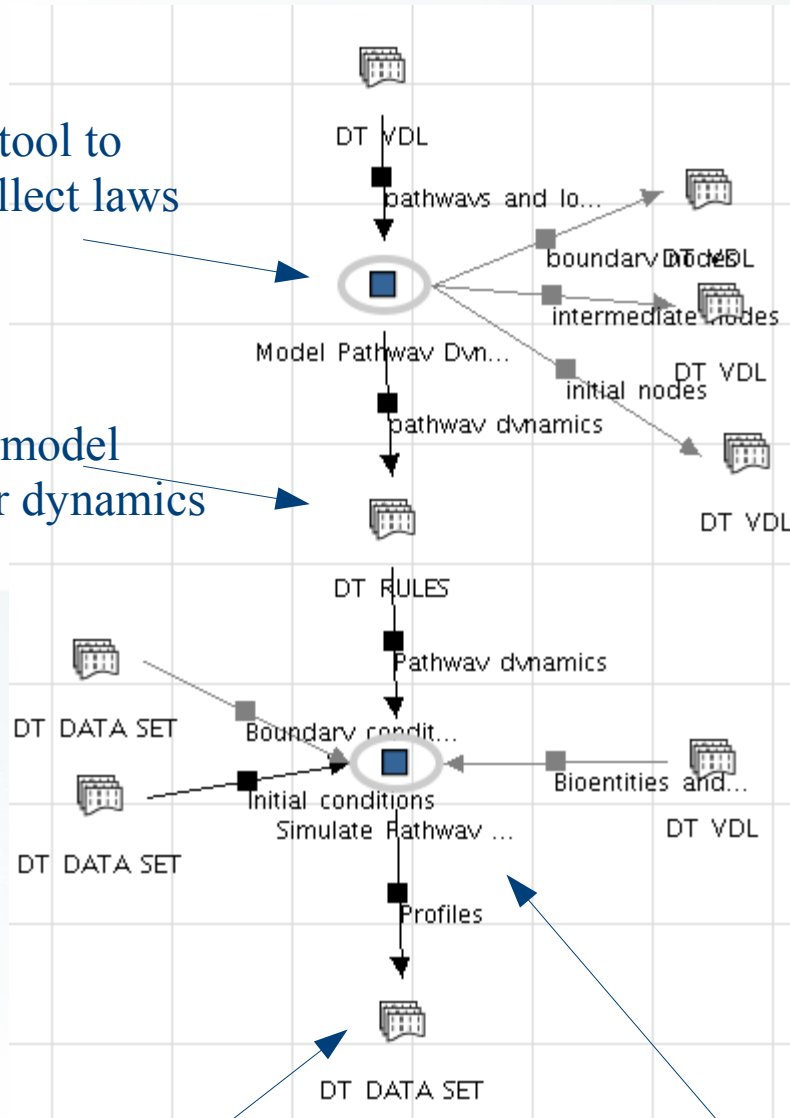
To the right of the window is a diagram of the PKC model. It shows a central node "I[test_pkc/R2]" (circled in blue) with arrows pointing to it from several other nodes: "M[test_pkc1/CaPKCmemb_]", "M[test_pkc1/DAGPKCmemb_]", "M[test_pkc1/AADAGPKC_]", "M[test_pkc1/AAPKC_]", and "M[test_pkc1/PKCbasal_]". Below "I[test_pkc/R2]" is another node "I[test_pkc1/sum]", which has an arrow pointing to "M[test_pkc1/PKCa]".

- this is the PKC model seen earlier

Model & Simulate using Workflow

A tool to collect laws

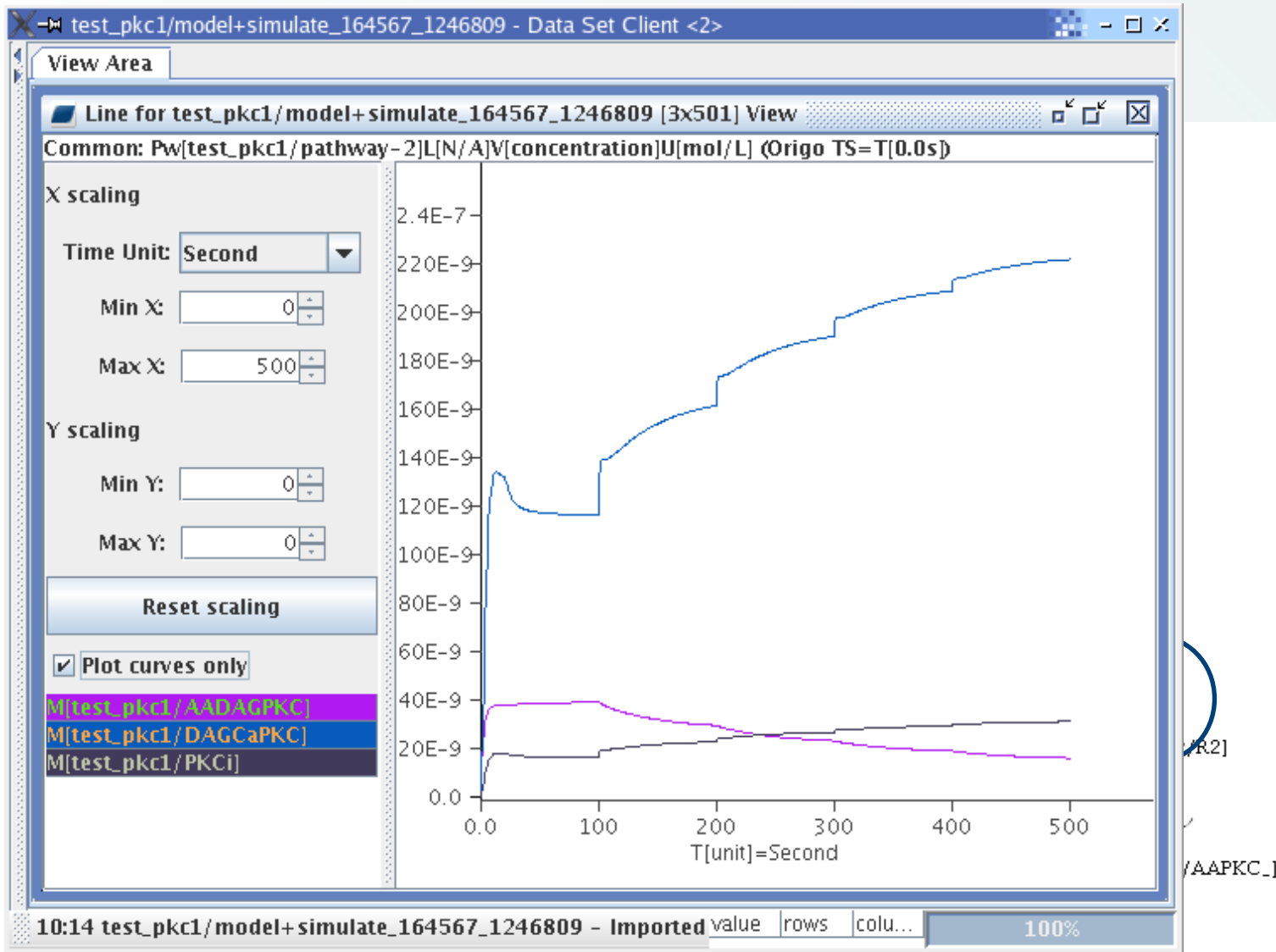
A model for dynamics



Simulation results

Simulation

- A tool can be used to collect the kinetic laws from the pathway along with the stoichiometry
 - produces a model for the dynamics as a set of ordinary differential equations (ODE)
- Another tool can be then used to simulate the the dynamic behavior, given initial and boundary conditions
 - produces concentration profiles as a function of time



- **Simulation results** in collaboration with Antti Ylipää, Tampere Univ. of Technology

I[test_pkc1/sum]



M[test_pkc1/PKCa]

Modeling: FBA

- Flux Balance Analysis
- Instead of giving the formula for the interaction rates, the kinetic laws be used to express constraints for the interaction rate
 - These can be collected from the pathway and a constraint model can be produced for the whole pathway
 - Linear optimization can be used on that model to find the material fluxes when the system is in steady state, optimizing a given target function
 - such as the production of biomass

Modeling: Boolean Network simulation

- instead of modeling the interaction rates and concentrations as in ODE modeling, we choose the state variables to have only on/off (1/0) values.
 - these values can be given suitable interpretations, e.g.
 - 1 = the interaction is possible/compound can be present
 - 0 = the interaction is not possible/compound cannot be present
- custom tools and workflows can be used to
 - collect the model from a pathway
 - simulate the state transitions from a start state, reaching an attractor cycle

Summary

Summary

- Pathways used to systematically document models of biological systems in a computer readable format
- Pathway concepts in Medicel Integrator:
 - location, interaction, connection, pathway, klaws
- Pathway data comes from
 - external databases, published models, from literature by manual data entry, analysis pathways
- Pathway data can be utilized for
 - ODE modeling, Boolean modeling, FBA, functional analysis/comparison, structure prediction
 - sharing the understanding of known biological relationships

Thank you!