

Validating *Bio*MAS models with mutation

Emanuela Merelli(1,2) and Michal Young(1)

(1)Department of Computer Science University of Oregon

(2)Dipartimento di Matematica e Informatica University of Camerino

Utrecht 25 July 2005 – BioMed05 Workshop



University of Camerino

Outline

- Systems Biology Context
 - MAS for Modelling Biological Systems
 - Model Validation

 - Mutation Testing to Validate BioMAS
 - BioMAs for Carbohydrate Oxidation

 - Conclusion and Future Work
-

Systems Biology

Part of modern biology regarding modelling and simulation of biological processes which aims at system level understanding of biological systems

[Kitano2002]

... new methods and techniques to investigate

- the structure of the systems (components and relations)
- the dynamics of the systems (understand the behaviour in normal and perturbed conditions)
- the functionality of the systems (gain control on mutations, repair malfunctioning cells)

... to be able to design and modify

- new models of the systems for desired properties (organ's cloning)
-

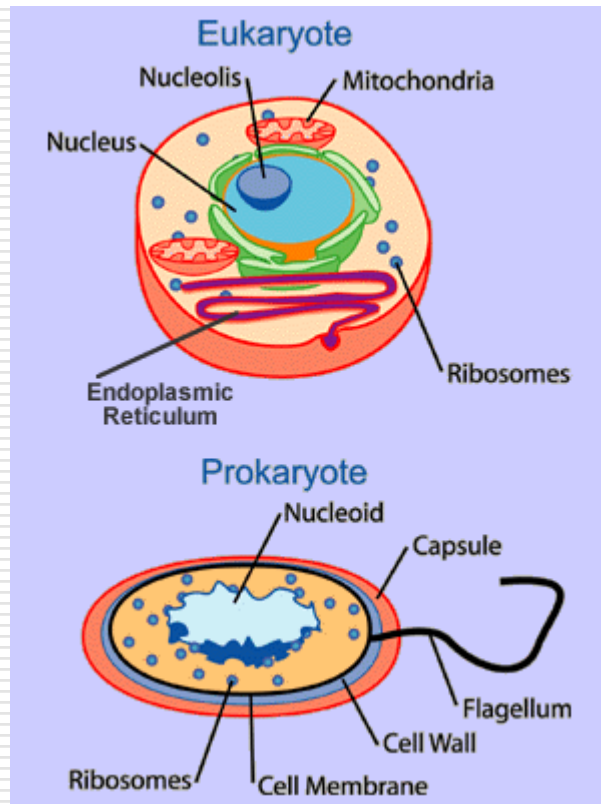
MAS for Modelling Biological Systems

Ongoing experiment: use agent-based paradigm to model biological systems

Autonomous agent is a computer system situated in a dynamic environment ... (Jennings 2000)

Multiagent systems is a collection of interacting autonomous agents ... (Jennings 2001)

Structure of Biological System

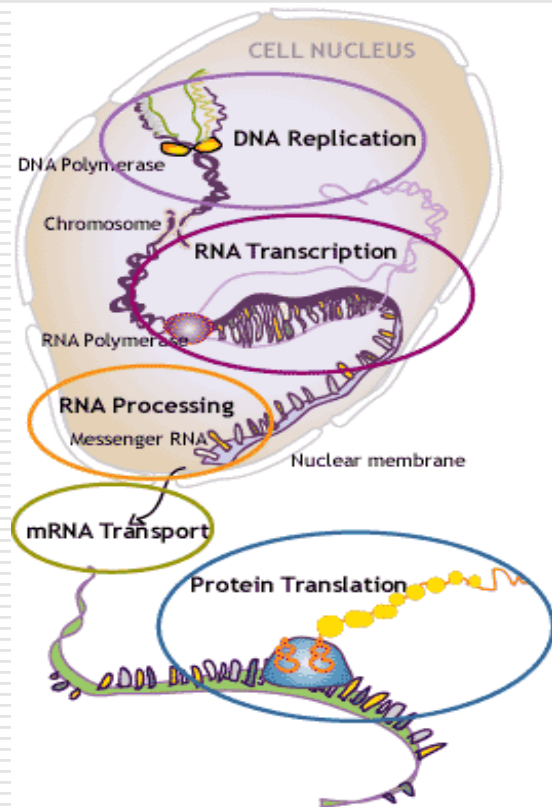


"ACTORS": Components of the Cell

Table 2-1. The Approximate Chemical Composition of a Bacterial Cell

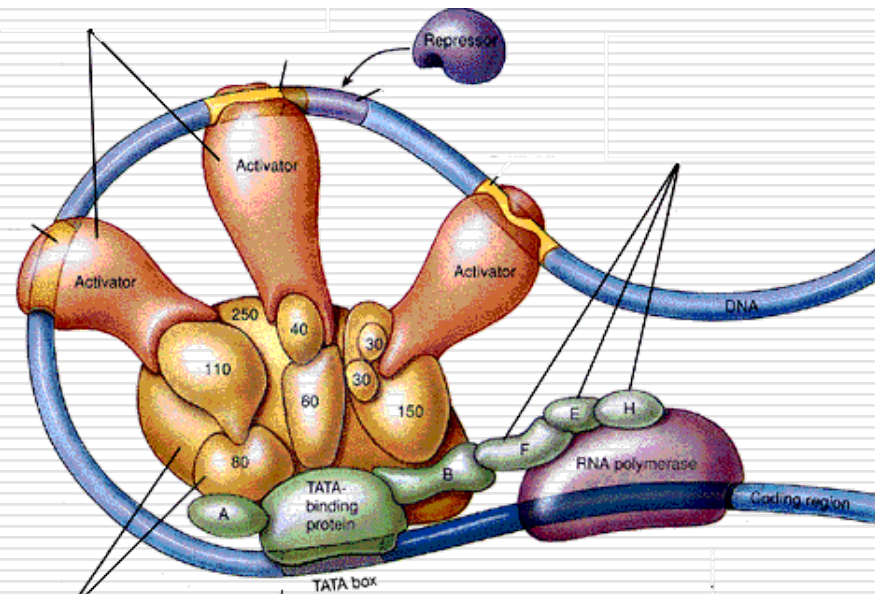
	Percent of Total Cell Weight	Types of Each Molecule
Water	70	1
Inorganic Ions	1	20
Sugars and precursors	1	250
Amino acids and precursors	0.4	100
Nucleotides and precursors	0.4	100
Fatty acids and precursors	1	50
Other small molecules	0.2	~300
Macromolecules (proteins, nucleic acids, and polysaccharides)	26	~3000

Behaviour of Biological System



"ACTIONS": Interactions

The components interact with each other to form new components, more complex structures and to perform the cellular processes and network of interactions (Metabolic Nets, GR Nets, SignalNets, etc.)



Biological System ...

- ❑ ... consists of several components interacting to perform complex functions
- ❑ The complexity of interactions is too far to be fully comprehended by human mind
- ❑ Computer simulation is an essential tool for testing and refining our understanding of structure and behaviour
- ❑ But ...the resulting models can also mislead, either through ordinary software faults (bugs) or through deeper mistake in modelling.

How should we test them?

Model Validation

- To validate a model means to assess how the model we are building responds (is faithful) to the biological one.
 - Did we build the right model

 - *To verify sw properties means to assess the correctness of the sw from the specification to the implementation*
 - *Did we build the system (model) right?*
-

Model Validation in the Classical Approach: Verification Testing

- ❑ The fundamental approach to validating a simulation model is to test it on some scenarios (test scenarios) for which we can distinguish correct behaviours from behaviours that reveal a flaw in design or coding
 - ❑ What test scenarios can shed light on the structural and behavioural fidelity of a model?
-

Model Fidelity

- Model fidelity addresses the question of whether the structure and function of the model accurately reflect the biological system
 - The new challenge comes in choosing the suitable set of scenarios
 - Validation against the intended purpose of the model
-

Mutation in Biological Systems

- A **simulation model** that cosely reflects the modelled system in both structure and function should be amandable to modifications that mimic mutations, and should show similar effects
 - Intentional insertion of faults is a well-known software testing technique, “mutation analysis”
[Hamlet 1977 and DeMillo et al. 1978]
-

Mutation Analysis (MA) in SE

- The seeded faults in conventional software mutation analysis are simple syntactic modifications (for example, changing a comparison " \leq " to a comparison " $=$ ") that bear no relation to biological mutation.
 - The aim of MA is purely verification rather than validation
 - The examination of resulting behaviors stops at distinguishing an incorrect behavior of the "mutant" program from correct behaviors of the unmodified program
-

Mutation Analysis for Biological Models

- The “mutations” are not arbitrary syntactic variations in code

 - Mutations are larger-grain modifications that mimic some known or plausible mutation in the subject system
 - --- which first of all imposes a requirement on the model that it possess a structure admitting of such modifications

 - When the modified model is executed, we expect
 - behaviors that correspond to those of the natural system with a corresponding mutation, or
 - the modification does not correspond to a known natural mutation, we expect a biologically plausible change in behavior
-

MAS Validation

- MAS simulation model is intended to mimic the simulated system in structure and behaviour, and not only in overall output

 - To validate a MAS model means
 - not only to validate predictions (e.g by simulating well-understood scenarios),
 - but also the relation between predictions and the structure and function of the model
-

Mutation Analysis for BioMAS

1. There is no straightforward way to make a software change corresponding to the expression of the biological mutation
 - For example, a mutation might prevent a particular enzyme from being produced. Ideally, in a MAS model, the enzyme itself would be an agent, and preventing production of the enzyme would correspond to suppressing activation of the agent.
 - But, the **enzyme may not exist** as an identifiable structure in the model.
-

Mutation Analysis for BioMAS (2)

2. The behavior of the modified model with biologically-inspired mutation **does not correspond** to the behavior of a biological system with the corresponding change

- This could be **due to a simple software fault** --- a “bug” or it could be due to deeper problems in the design of a model
 - For example, a certain chemical is always in abundant supply during normal functioning. Some behavior of the model may depend on that chemical, and yet it may never be tested. The omission of the test becomes obvious only when a modification introduces a scarcity of that chemical.
-

Validating Carbohydrate Oxidation Simulation

We consider the problem of model validation for a simulation model whose **structure as well as behaviour mimics the Carbohydrate Oxidation**

The Carbohydrate Oxidation is the energy production process performed by two active components of the Cell: Cytoplasm and Mitochondrion

What test scenarios can shed light on the structural and behavioural fidelity of a model?

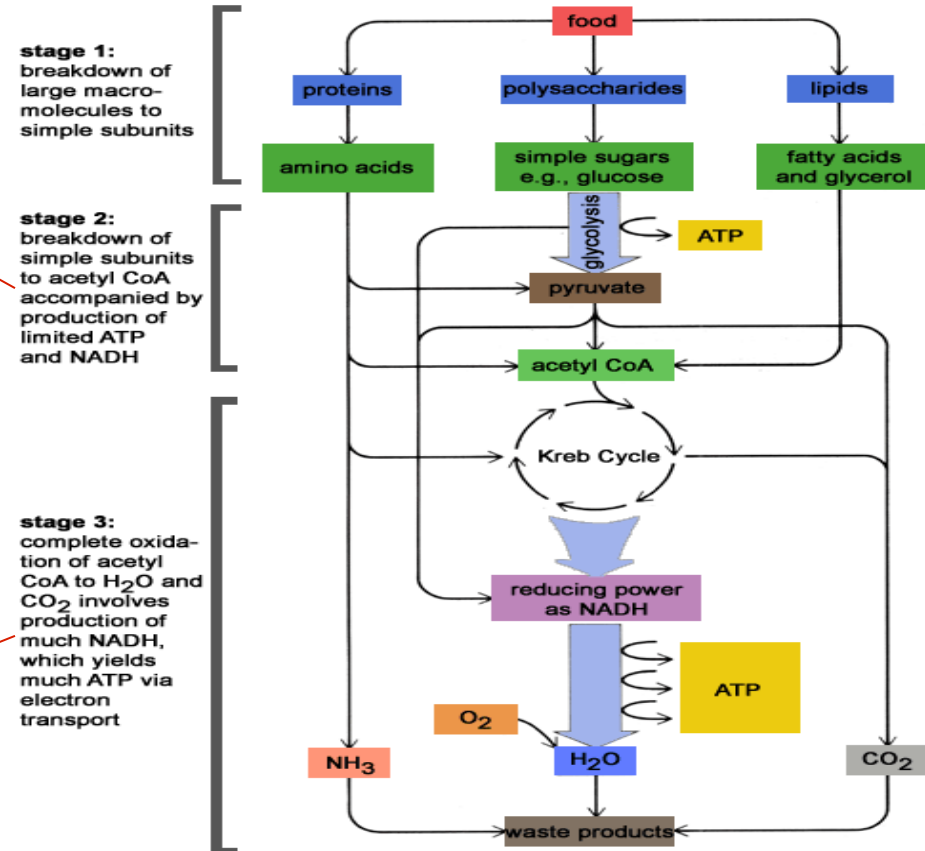
Carbohydrate Oxidation

Cytoplasm (anarobic)

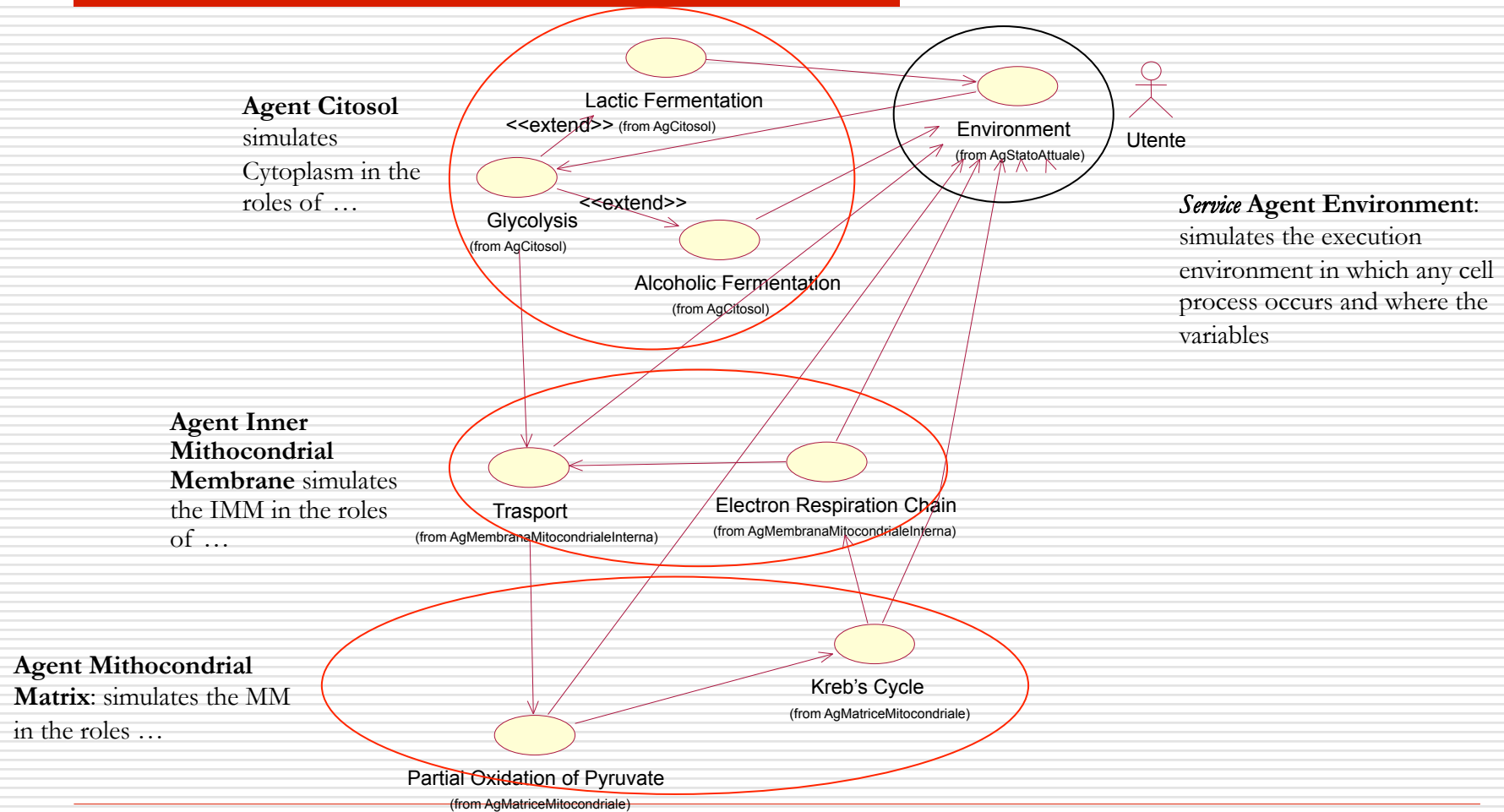
- **Glycolysis** (*convert 1 glucose into 2 pyruvate + ...*)
- Lactic Fermentation (*reduce pyruvate to lactate*)
- Alcoholic Fermentation (*reduce pyruvate to ethanol*)

Mithocondrion (arobic)

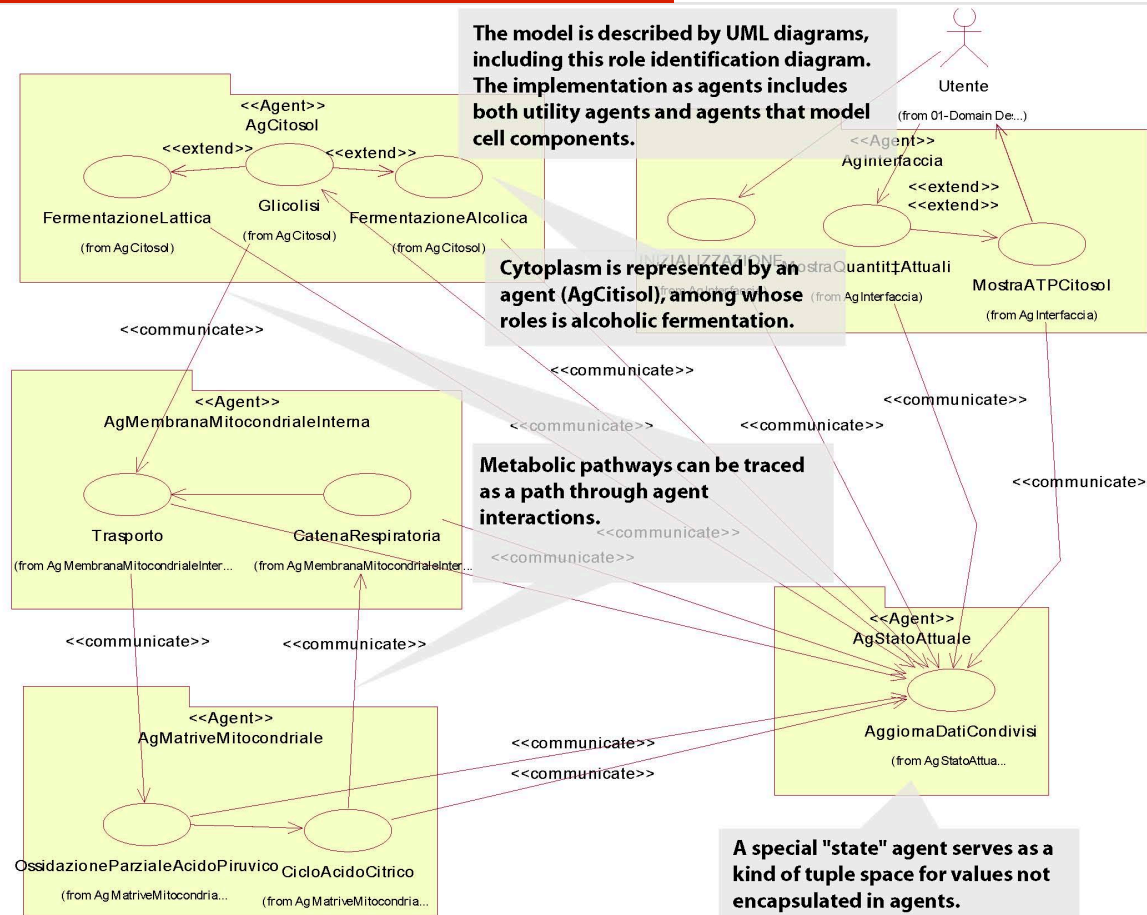
- Mitochondrial Inner Membrane
 - **Transportation** (*pyrutave frpm cytop. to MM*)
 - **Respiration Chain** ()
- Mitochondrial Matrix
 - **Partial oxidation of pyruvate**
 - **Kreb's cycle**



Functional Domains [Corradini et al.05]



Cell Components Stereotypes



Simulation Model [Corradini et al.05]

Control Panel - ©2004 CellSoft D.E.M.

Carboidrato

Nome: Quantità: Lunghezza Catena Polisaccaride:

Ossigeno: Tipo Fermentazione:

START SIMULATION >>

ATP: 0 **ADP: 0**

AcetilCoA:	0	CoA:	0	FADrid:	0	NADox:	0
Acido Lattico:	0	Etanolo:	0	H2O:	0	Pi:	0
CO2:	0	FADox:	0	NADrid:	0	Piruvato:	0

Simulation Model for Mutation

- Our validation approach imposes a new requirement on the model
 - we should, in addition, obtain reasonable results **when the model is altered** in ways that correspond to known or plausible mutations

 - Mutations with known effects are most useful, because their effects **provide a standard** to compare simulation results
 - carbohydrate oxidation can take place within two different environmental conditions, in presence of oxygen (aerobic) or in its absence (anaerobic)
 - the former takes place in the mitochondria,
 - the latter in the cytoplasm

 - ... as consequence of malfunction due to DNA mutation in the mitochondrion, the aerobic pathway is blocked and metabolism is forced to change behavior with respect to new condition
-

Consequence of Mutation in a Biological Carbohydrate Oxidation

- **A mitochondrion DNA mutation** can lead to a lack or disappearance of an enzyme involved in a metabolic pathways
 - It can provide different ATP or in the worst case the block the process
 - An aerobic microorganism can become anaerobic

 - **Example**
a mutation of the gene that produces one of the enzyme involved in the partial oxidation of pyruvate (i.e. it allows the passage of pyruvate from the cytoplasm to the mitochondrion, through the mitochondrial membrane, by allowing the aerobic respiration)
 - If the mutation provokes the disappearance of an involved enzyme the microorganism is forced to adapt to a new context by using the anaerobic pathway.

 - In this case, the pyruvate is transformed in lactate by producing NADox, allowing the cell to maintain glycolysis and to produce ATP
-

Mutation Analysis for Corradini et al. BioMAS Model Validation

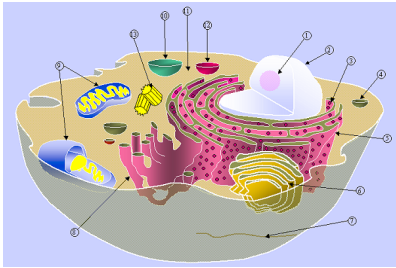
- **The Cell-MAS model was not designed with such modifications in mind**, and (not too surprisingly), we found that the correspondence between the software system and components in the biological system is imperfect
 - the enzyme is not explicitly represented, but we can easily simulate the effect of its absence by a simple modification to the agent representing the membrane
 - aerobic and anaerobic pathways are selected artificially as a (user-settable) model parameter instead of arising from chemical conditions, in particular the anaerobic pathway could be activated by failure of the aerobic pathway

 - we identified a **characteristic of the model that should be refined** to improve the correspondence between software model and natural system
 - Specification diagrams describe **aerobic and anaerobic pathways to proceed in parallel** when both are present; this should have allowed anaerobic pathways to engage when the aerobic pathway was blocked by mutation.
 - However, the corresponding modification to the implementation did not work because dependence on that aspect of the model was spread among other implementation components (agents).

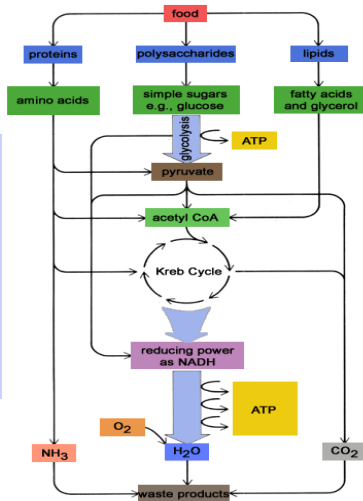
 - this deviation from agent-based design criteria and can be considered **an implementation fault**, but it became evident only in validation
-

Biological model

Eukariote Cell

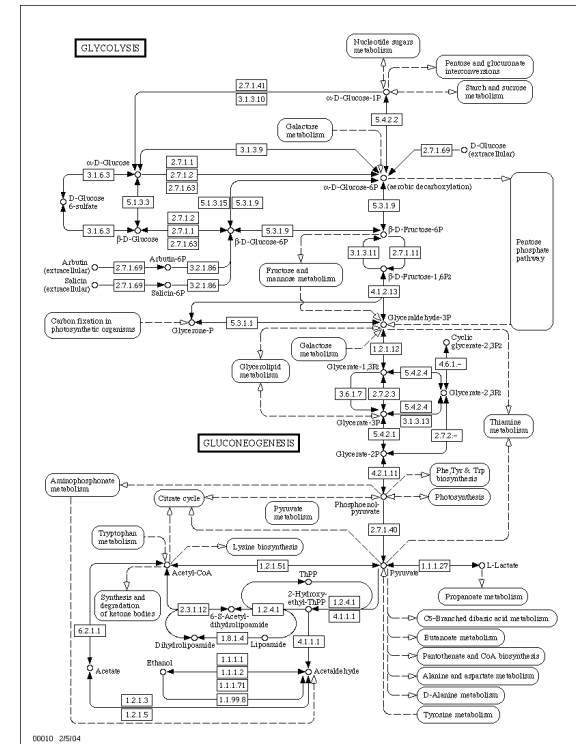


Cell Metabolism



Semi-formal description of the biological model

Glycolysis semi-formal description



System Modelling

Formal description of the biological model

$$(a,A) \rightarrow \text{SKIP}; (b,B) \rightarrow \text{SKIP}; (c,C) \rightarrow \text{SKIP}; (d,D) \rightarrow \text{SKIP}; \prod_{i=1}^3 P_i;$$

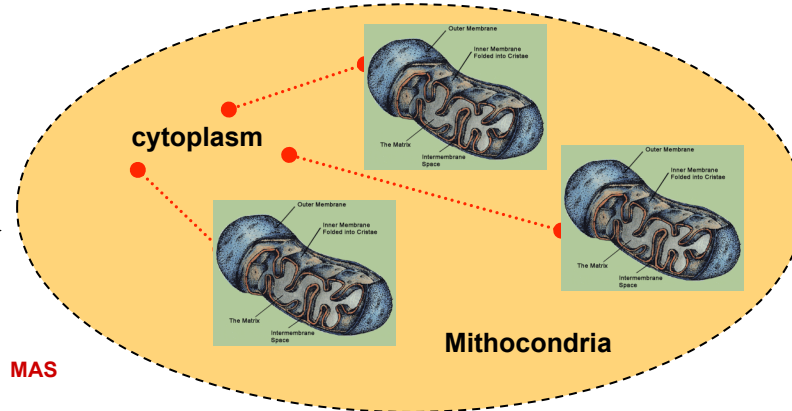
$$(n,N) \rightarrow \text{SKIP}; \prod_{j=1}^4 Q_j; (x,X) \rightarrow \text{SKIP}$$

where $\prod_{i=1}^3 P_i = (f,F) \rightarrow \text{SKIP}; (g,G) \rightarrow \text{SKIP}; (h,H) \rightarrow \text{SKIP}$
 and $\prod_{j=1}^4 Q_j = (p,P) \rightarrow \text{SKIP}; (q,Q) \rightarrow \text{SKIP}; (r,R) \rightarrow \text{SKIP}; (s,S) \rightarrow \text{SKIP}.$

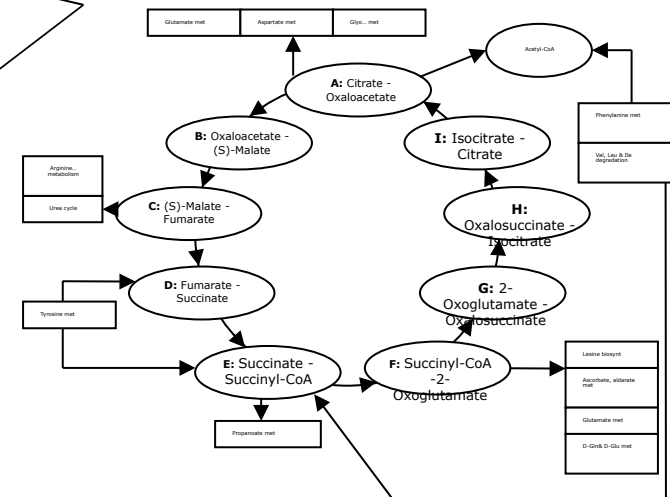
Model Validation by model checking

MAS Model Validation

Validation by testing scenarios



System Simulation



Conclusions

- Agent-based model to understand and precisely describe the behaviour of biological processes
 - Agent-based simulation supports scalable approach
 - Agent-based model for biological system simulation can potentially establish fine-grained structural and behavioural correspondence
 - Agent-based model supports the validation based on “mutations” (i.e. changes to the model correspond to changes in the natural system)
-



Thank you!