An Agent-Oriented Modeling Approach to Complex Biological Pathways

Line Pouchard, Computer Science and Mathematics

Mike Leuze, Computer Science and Mathematics

Richard Ward, Computational Science and Engineering

What is an ontology?

- Domain ontology
  - A domain theory that specifies domain-specific vocabulary of entities, classes, properties, predicates, functions and a set of relationships that necessarily hold among these vocabulary items (Fikes & Farquhar, 1999)
  - An agreed vocabulary of common terms and meanings within a group of people or in some domain for the purpose of knowledge-sharing (Roche)

How does it differ from XML?

- XML provides a syntax for describing taxonomies with some parent/child relationships defined

```
<variable>
        <time units="d">
                <init> 0</init>
                <min>0</min>
                <max>100</max>
        </time>
</variable>
```

- An ontology provides richer relationships between objects (definitions and axioms constraining use)

- An ontology improves interoperability between systems (e.g., transfer annotations between databases)

Ontologies may be expressed using RDF (expression of semantics) & XML (syntax)

- XML (Extensible Markup Language) www.w3.org/XML/
  - Provides syntax
- RDF (Resource Description Framework) www.w3.org/RDF/
  - Provides mechanism for relationships
- Ontology forms innermost tags
  - Describes specifics of the ontology

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<rdf:RDF xmlns:rdf="…">
   xmlns:a=http://www.daml.org/2001/03/daml+oil#>
          <a:Ontology rdf:about="" />
          <a:ObjectProperty rdf:about="...">
             …….
          </a:ObjectProperty>
<rdf:RDF>
```

Ontologies are used in genome and cellular modeling

Examples of Domain Ontologies:

Genome: Gene Ontology (GO) Consortium

www.geneontology.org

GO divided into three levels

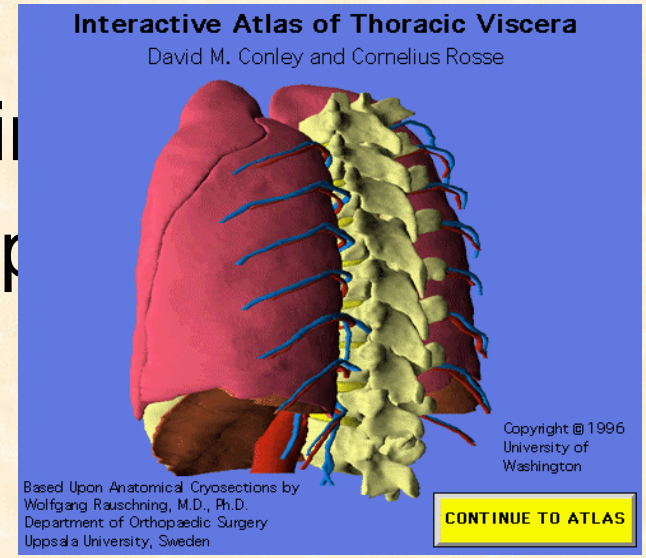
- DNA Metabolism
- Molecular Function
- Cell Processes

Cellular:

- EcoCyc (ecocyc.org)
- eCell (e-cell.org)

Digital Human: create anatomical ontology

- Based on the Digital Anatomist Foundation Model
- NLM Unified Medical Language System (UMLS) **sig.biostr.washington.edu/projects/da/**
- Developed by Rosse and Bri
  - Structural Informatics Group
  - Digital Anatomist Project

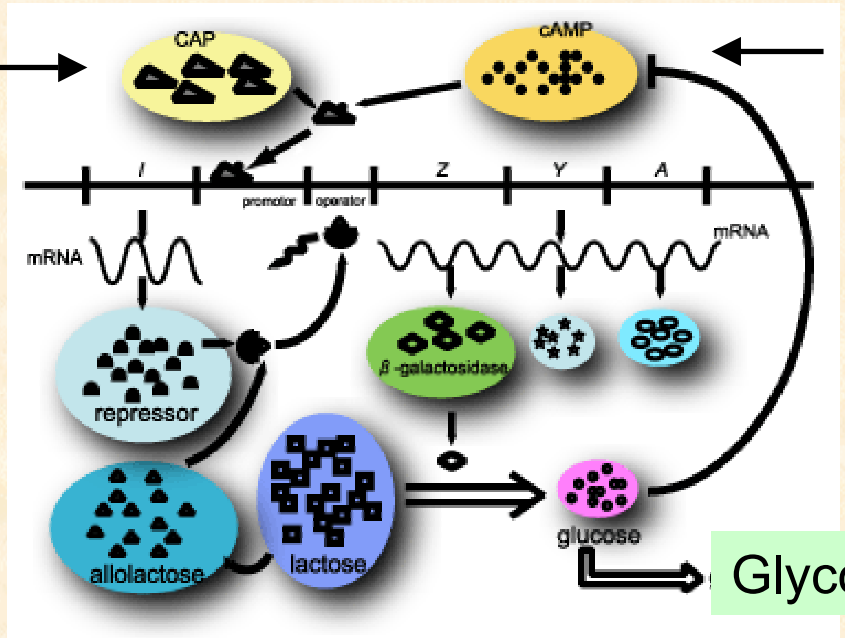


Such an effort will create a universal description of anatomy for use in visualizing Virtual/Digital Human

Regulatory pathways

- Example studied: utilization of lactose in *E. coli.*
- Well studied experimentally.
- Differential equation model exists.
- Connected to the metabolic pathway to be studied.

Catabolic activator protein



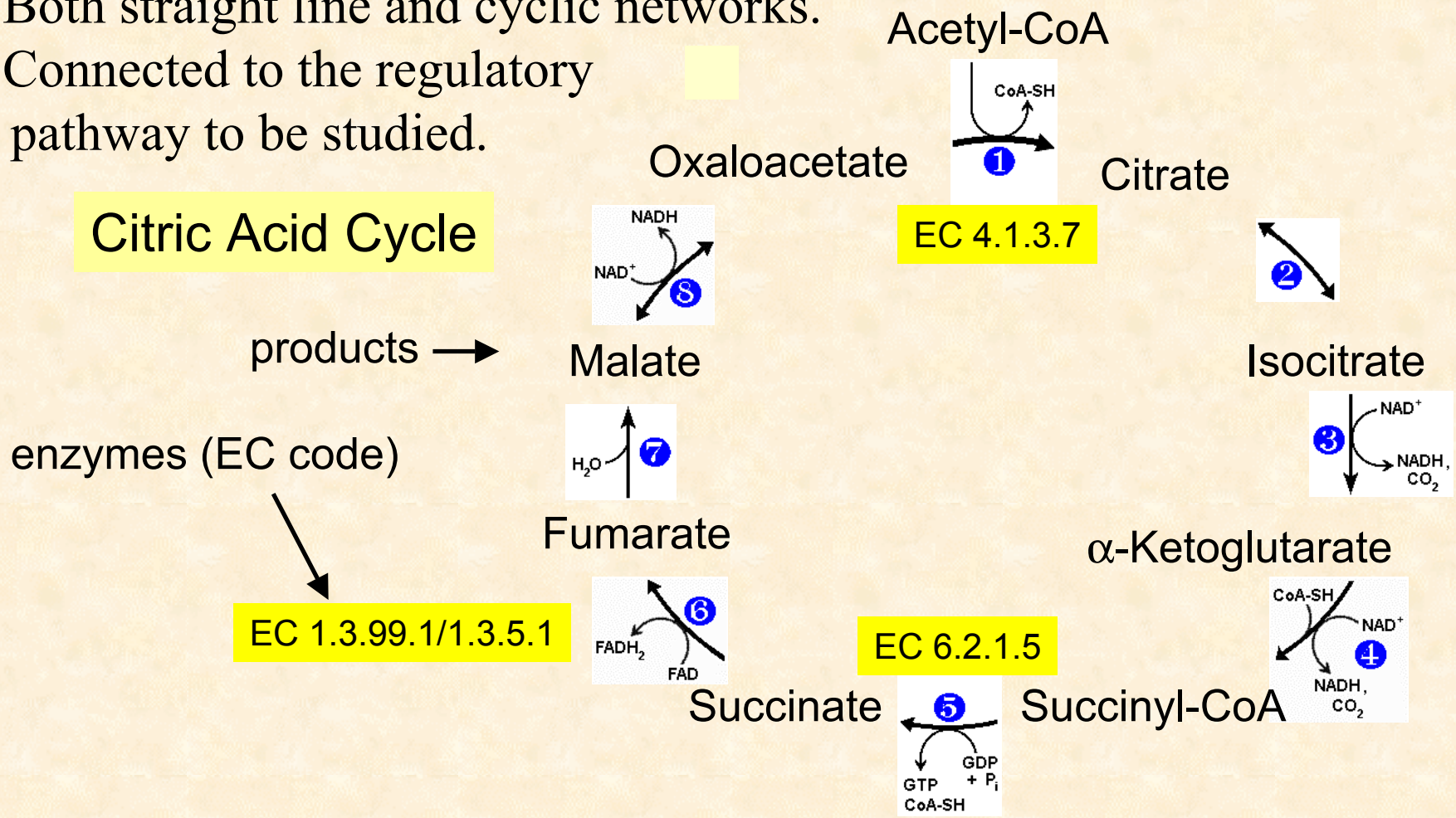
Cyclic AMP plays crucial role

Glycolysis pathway

Gene regulation of the *lac* operon

Metabolic Pathways

- Examples considered: glycolytic pathway and the citric acid cycle.
- Both straight line and cyclic networks.
- Connected to the regulatory pathway to be studied.



Modeling of a cell is a major challenge

- Complex pathways are inter-dependent.
- Understanding the genome and its function is not sufficient to understand the organism.
- Networks of pathways are required for modeling the organism.
- Emerging need and a growing body of research for an infrastructure integrating some complex biological systems.
- In silico experiments integrate large number of components through simulation.
- Some directions exist and efforts are underway but no known system uses agents
  - Interest in large scale, coupled models for predictions of behavior in cell networks.

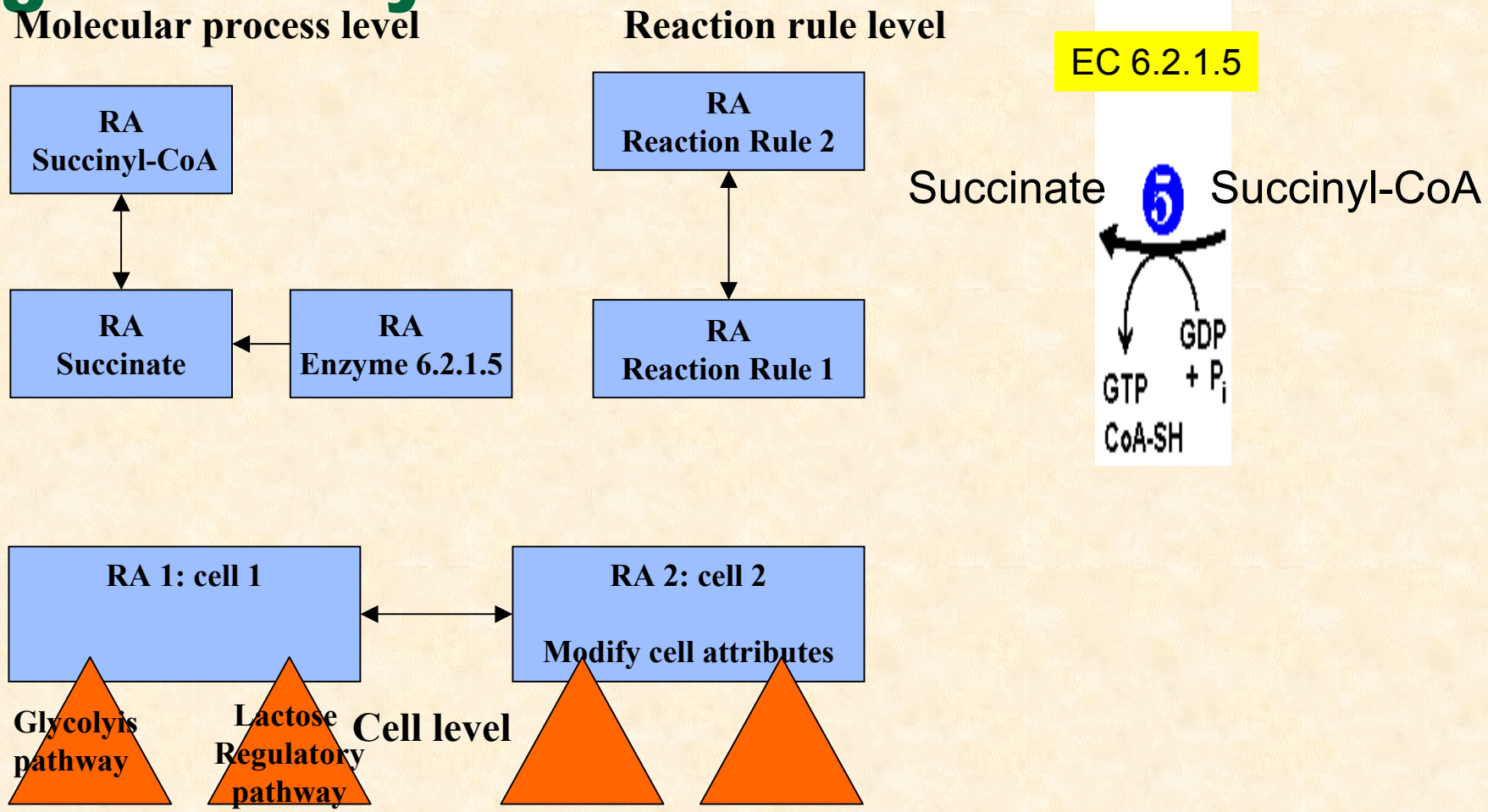
Existing Cell Modeling Approaches

- The E-cell approach is limited in the number of components active in a pathway.
- The Virtual Cell models spatial structures in the cell.
- Kinetic models require
  - quantitative information that is not always known
  - incorporate very little knowledge.
- Cell modeling overextends the resources of a single machine.

An agent-oriented approach can handle the complexity of cell modeling in a robust manner

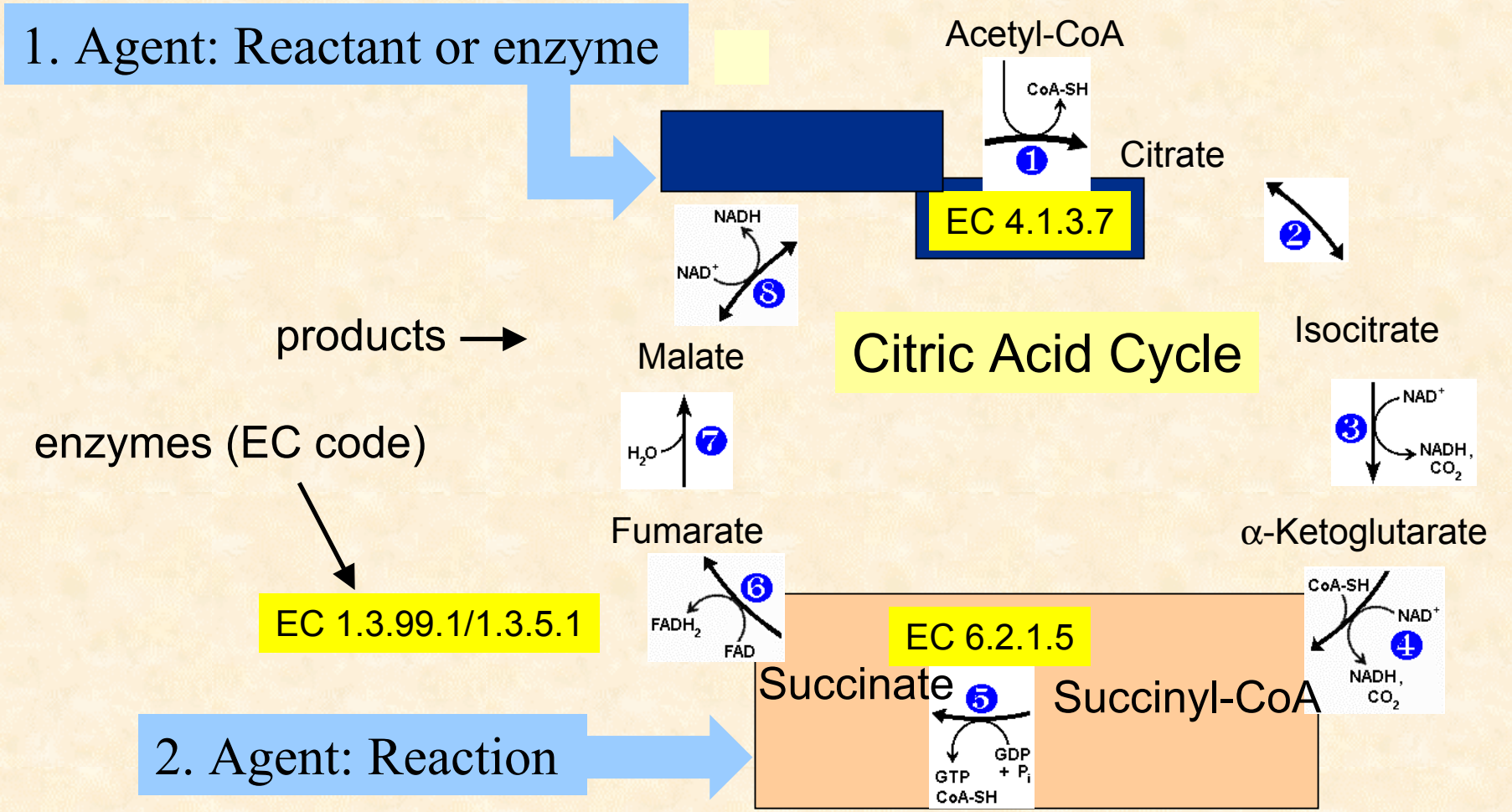
- Concentrates on a well-known pathway for proof-of-principle
  - the model is adjusted if computational results do not compare to experimental ones.
- Extends the framework to another pathway.
- Allows annotations from other sources to be brought into the model.
- Investigates issues for semi-genome scale analysis.
- Investigates the challenges for this approach when applied to new pathways.
- Agents also bring flexibility, a distributed environment, and qualitative modeling

Agents can model different levels of granularity

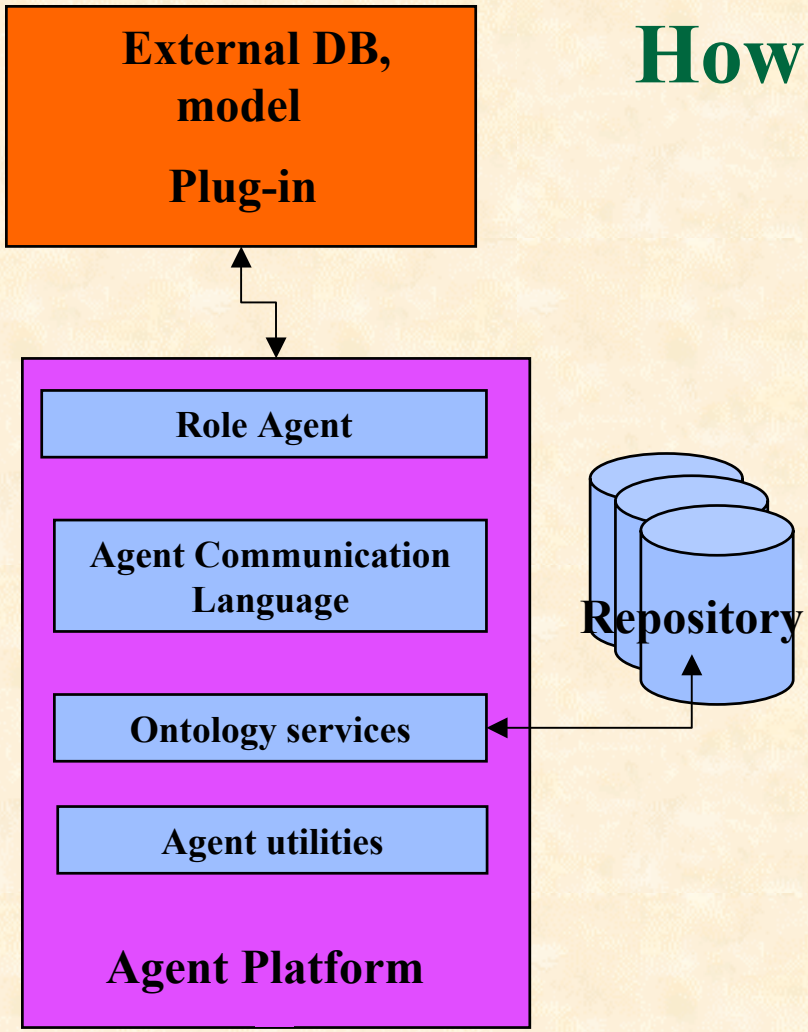


Mapping a Metabolic Pathway to an Agent Architecture

Depending on the granularity chosen:
either 1) reactant or 2) entire reaction can be mapped to a single agent



How are ontologies used?



- Collected in a repository for re-use: EcoCyc and MetaCyc ontology services for E-coli
  - –ontologies are becoming current in genomics due to the very large number of inter-dependent concepts present in a biological system
- GOBO
  - –Ontologies built on RDF schemas
  - –No pathway ontologies yet
  - –Some pathways in KEGG, WIT and MEtaCyc

Importance of a declarative ontological framework for handling the complexity

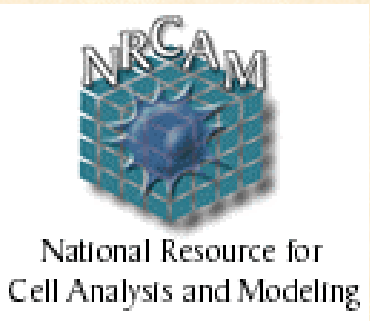
- A declarative framework
  - constructs a structured, precisely defined data model for metabolic pathways and provides a database schema
- A taxonomy with definitions and relationships between terms
  - can be dynamically updated as new pathways, compounds and reactions are being discovered
  - can incorporate and organize large amounts of information
- Ontology objects allow incorporation of annotations into the model
- May need to be extended to incorporate our concepts (e.g. compounds, chemical reactions)
  - Possible re-use of existing terms and relationships for

Evaluation of the approach

Agent-based approach will be validated against

differential equations (kinetic) approach using:

- **E-cell**
- **Virtual Cell (National Resource for Cell Analysis and Modeling)**



Goals:

- Validation of the agent-based approach
- Elimination of any noise propagation resulting from analysis of large biological pathway
- Understand the scalability issues.

Future Directions

- ***Constructing new pathways:*** **long-term goal**
- ***The focus in our approach*****: major tasks**
  - **Building a prototype for part of glycolysis**
  - **Designing the agents, designing the ontology**
  - **Investigating how our approach can be scaled to model an entire pathway**
  - **Investigating issues related to extension of the model by incorporating another type of pathway**
- ***Need for a committed domain expert in the team.***

Government Disclaimer

- **The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.**