The Pathway Tools Software and BioCyc Database Collection

Peter D. Karp, Ph.D.

Bioinformatics Research Group

SRI International

pkarp@ai.sri.com

http://www.ai.sri.com/pkarp/talks/

BioCyc.org

EcoCyc.org, MetaCyc.org, HumanCyc.org

Use Cases for Pathway Tools and BioCyc

- Development of organism-specific DBs (model-organism DBs) that span many biological datatypes
- Web publishing of those DBs with a powerful set of query and visualization tools
- Computational inferences of metabolic pathways, pathway hole fillers, operons, transport reactions
- Visual tools for analysis of omics data
- Tools for analysis of biological networks
- Comparative analysis tools
- Metabolic engineering
- BioCyc is a Web portal for genome and pathway information

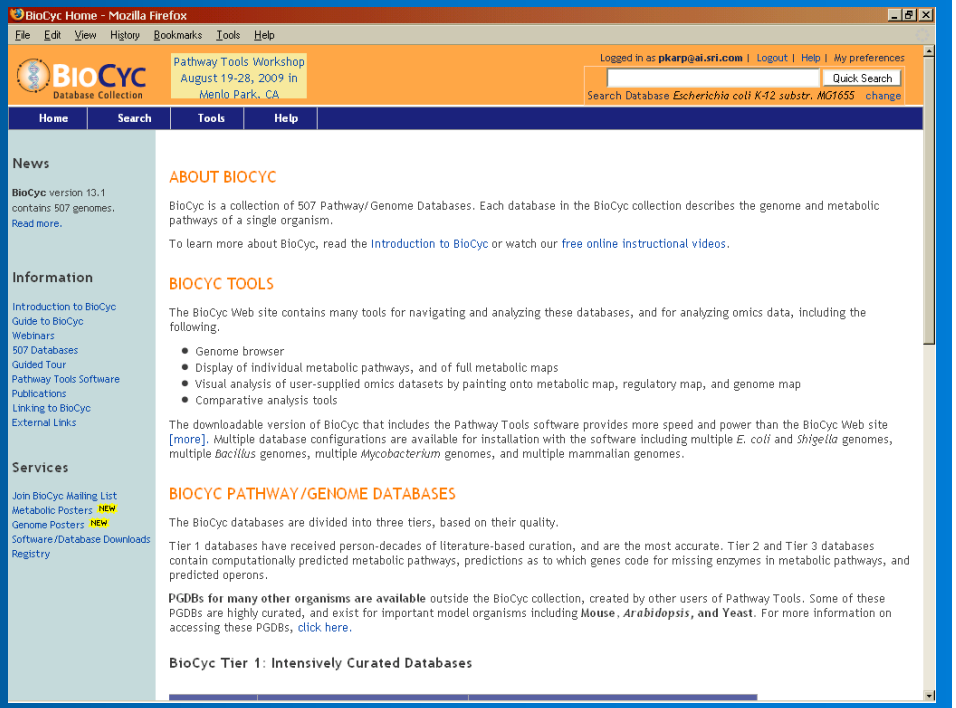
BioCyc Collection of 673 Pathway/Genome Databases

- **Pathway/Genome Database (PGDB) – combines information about**
  - Pathways, reactions, substrates
  - Enzymes, transporters
  - Genes, replicons
  - Transcription factors/sites, promoters, operons

- **Tier 1: Literature-Derived PGDBs**
  - MetaCyc
  - EcoCyc -- *Escherichia coli* K-12

- **Tier 2: Computationally-derived DBs, Some Curation -- 28 PGDBs**
  - HumanCyc
  - Mycobacterium tuberculosis

- **Tier 3: Computationally-derived DBs, No Curation -- 643 DBs**



Pathway Tools Software

- **PathoLogic**
  - Predicts operons, metabolic network, pathway hole fillers, from genome
  - Computational creation of new Pathway/Genome Databases
- **Pathway/Genome Editors**
  - Distributed curation of PGDBs
  - Distributed object database system, interactive editing tools
- **Pathway/Genome Navigator**
  - WWW publishing of PGDBs
  - Querying, visualization of pathways, chromosomes, operons
  - Analysis operations
    - Pathway visualization of gene-expression data
    - Global comparisons of metabolic networks

Briefings in Bioinformatics 11:40-79 2010

Obtaining a PGDB for Organism of Interest

- Find existing curated PGDB
- Find existing PGDB in BioCyc
- Create your own
- Curated pathway DBs now exist for most biomedical model organisms

Pathway Tools Software: PGDBs Created Outside SRI

- **2,100+ licensees: 180 groups applying software to 1,600 organisms**

- ***Saccharomyces cerevisiae*, SGD project, Stanford University**
  - *135 pathways / 565 publications*
- ***Candida albicans*, CGD project, Stanford University**
- **dictyBase, Northwestern University**

- **Mouse, MGD, Jackson Laboratory**
- **Drosophila, FlyBase, Harvard University**
- **Under development:**
  - *C. elegans,* WormBase

- ***Arabidopsis thaliana,* TAIR, Carnegie Institution of Washington**
  - 288 pathways / 2282 publications
- ***PlantCyc*, Carnegie Institution of Washington**
- **Six *Solanaceae* species, Cornell University**
- **GrameneDB, Cold Spring Harbor Laboratory**
- ***Medicago truncatula*, Samuel Roberts Noble Foundation**

MetaCyc: Metabolic Encyclopedia

- Describe a representative sample of every experimentally determined metabolic pathway
- Describe properties of metabolic enzymes

- Literature-based DB with extensive references and commentary

- MetaCyc now assigns more than twice as many reactions to pathways as does KEGG

***Nucleic Acids Research* 2010**

MetaCyc Data -- Version 14.0

Pathways	1,471
Reactions	8,409
Enzymes	6,198
Small Molecules	8,572
Organisms	1,861
Citations	22,459

Taxonomic Distribution of MetaCyc Pathways – version 13.1

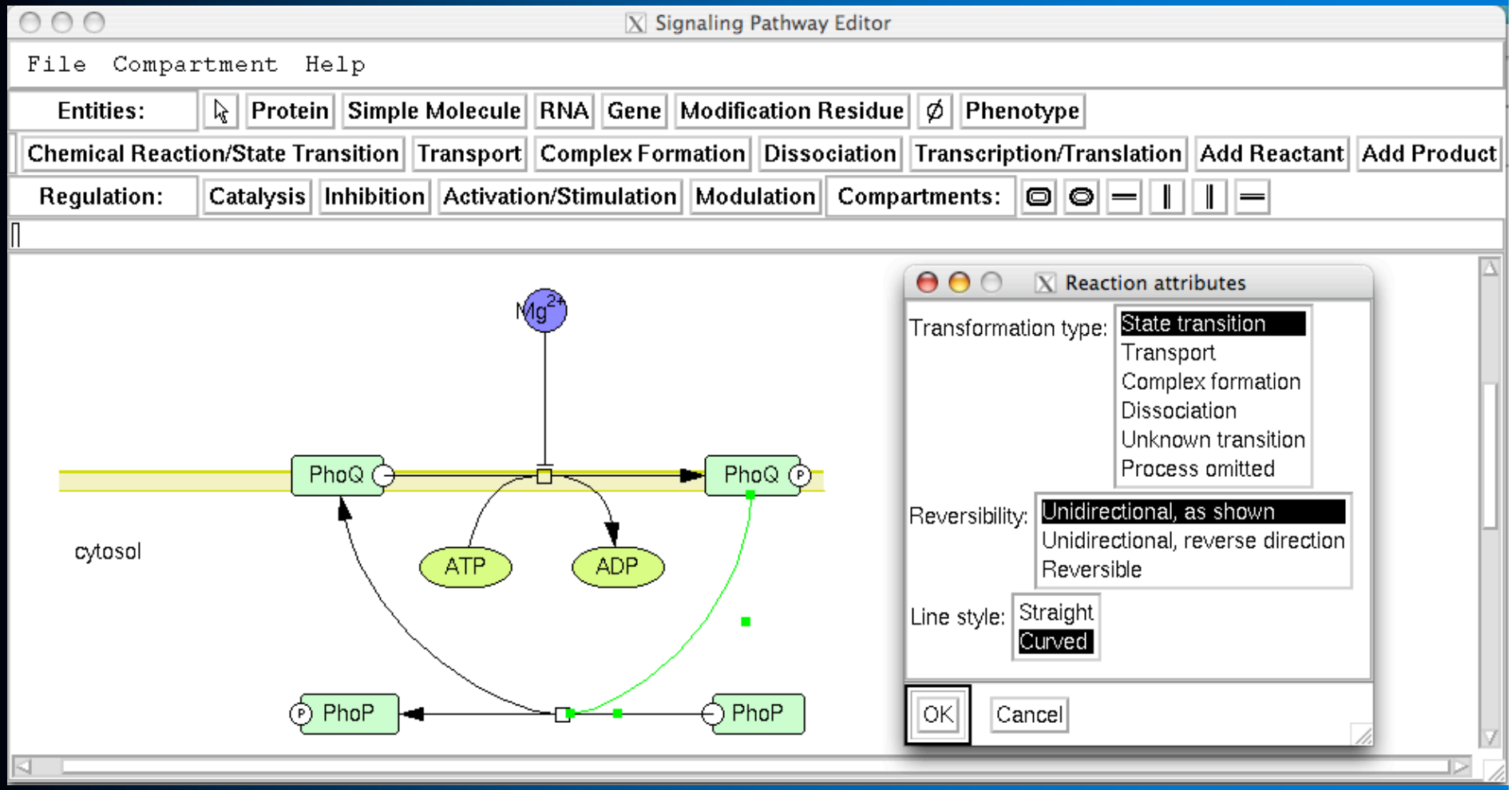
Bacteria	883
Green Plants	607
Fungi	199
Mammals	159
Archaea	112

Pathway Tools Survey Publication

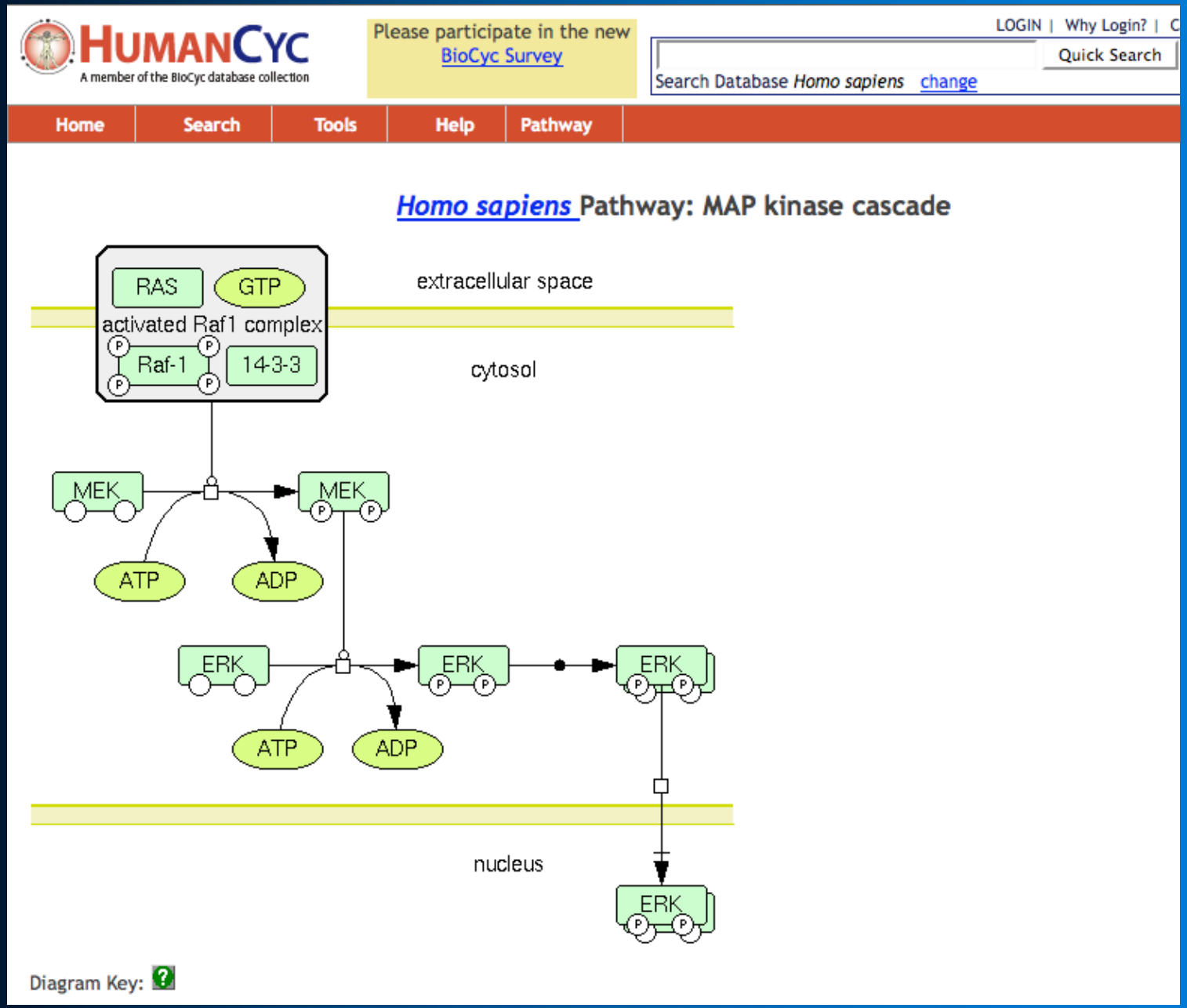
- **Karp et al, *Briefings in Bioinformatics* 2010 11:40-79.**

Signaling Pathway Editor

- **Signaling pathways use different visual conventions than metabolic pathways**
- **Look and feel based of our tool based on CellDesigner, SBGN**
- **Manual layout**
  - Can't yet be included in Cellular Overview Diagram



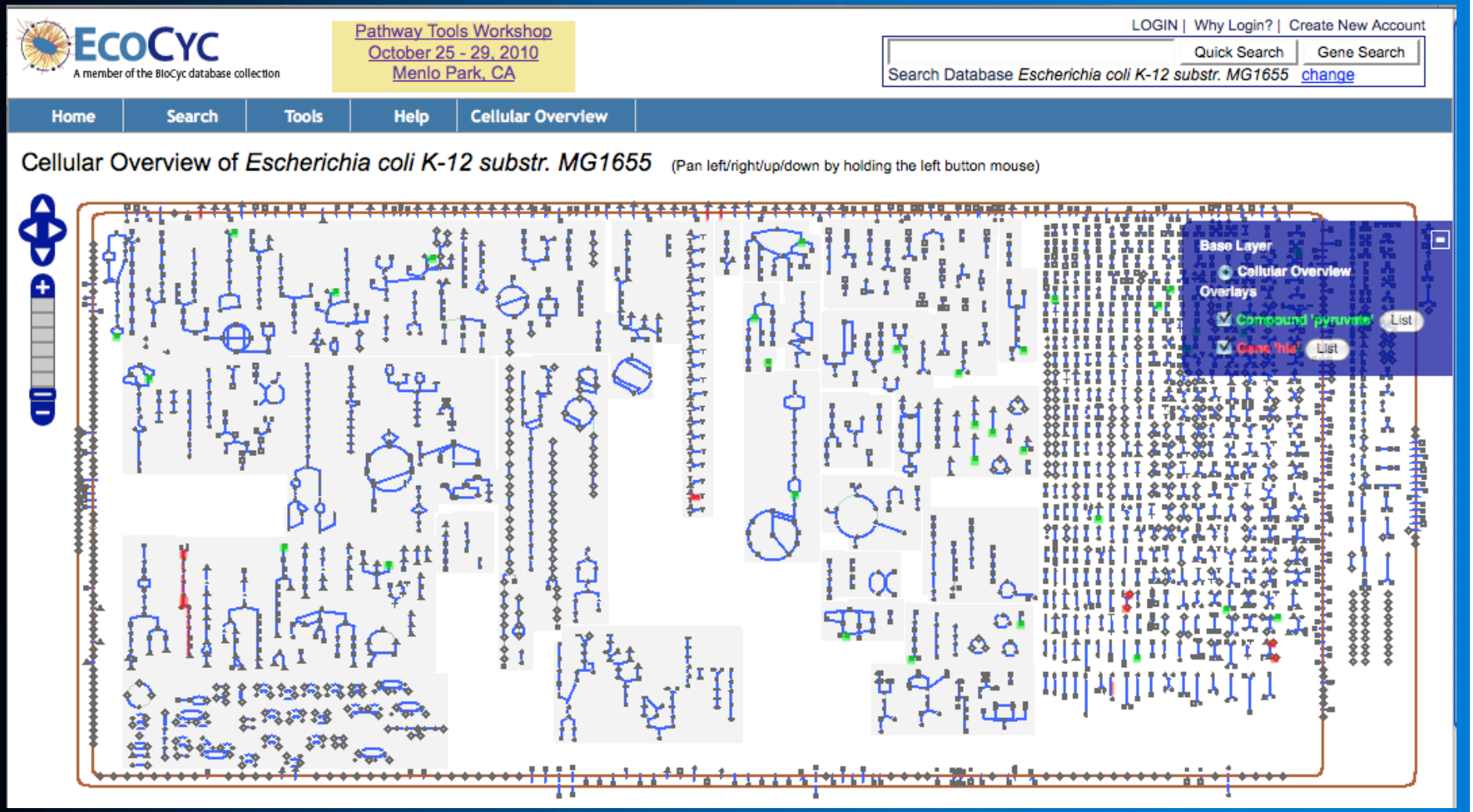
BioCyc Database Collection


HUMANCYC
A member of the BioCyc database collection
Please participate in the new
BioCyc Survey
LOGIN | Why Login? | C
Quick Search
Search Database Homo sapiens change
Home
Search
Tools
Help
Pathway
Homo sapiens Pathway: MAP kinase cascade
RAS
GTP
activated Raf1 complex
Raf-1
14-3-3
extracellular space
cytosol
MEK
MEK
ATP
ADP
ERK
ERK
ERK
ATP
ADP
nucleus
ERK
Diagram Key:

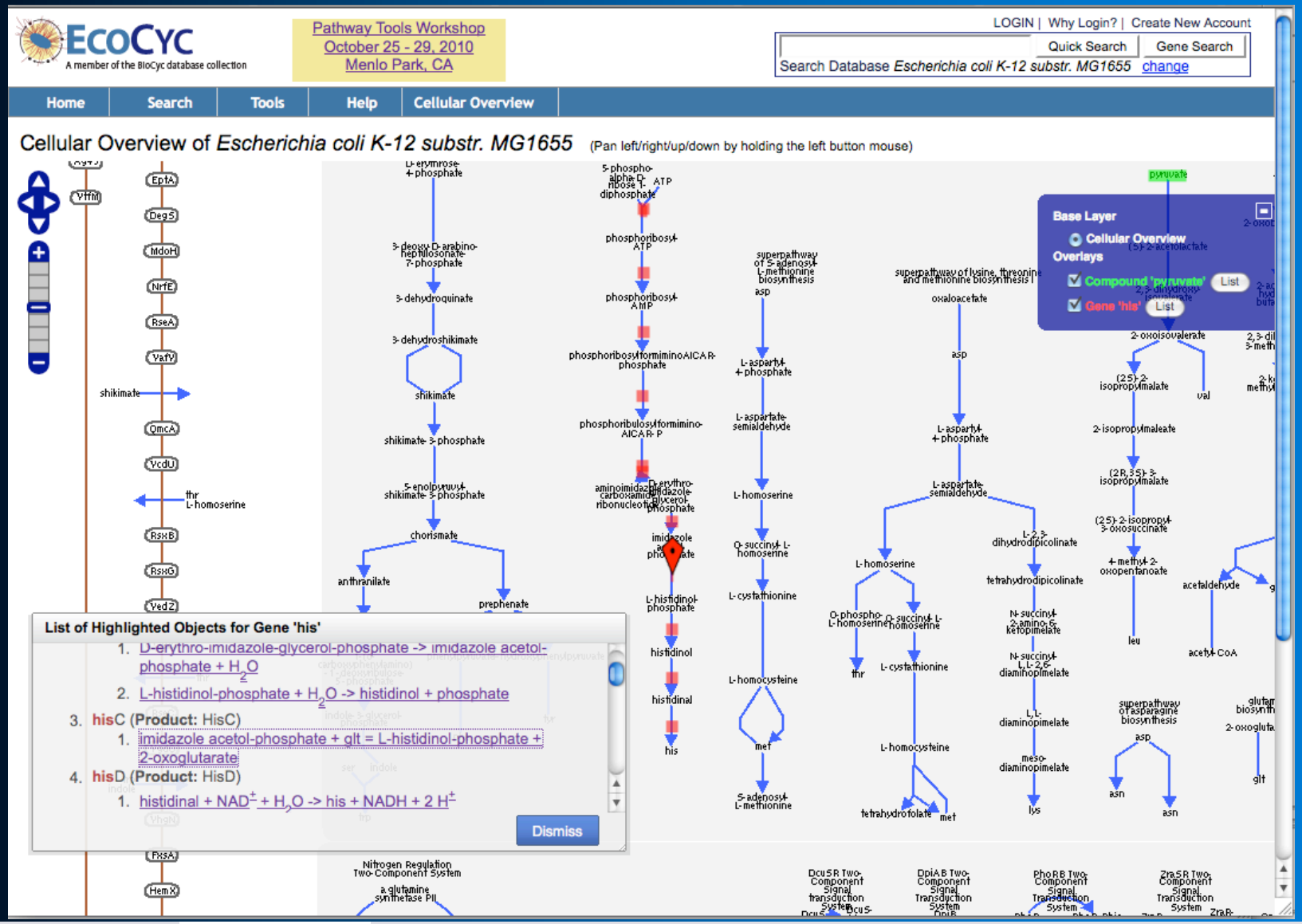
Improved Web Overviews

- **Implemented using OpenLayers**
- **Zoomable, draggable, searchable, paintable**
- **Cellular Overview**
  - Highlight compounds, reactions, enzymes, genes by name, substring, with autocomplete
  - Highlight genes from file
  - Superimpose omics data
- **Regulatory Overview**
  - Draw connections between a gene and its regulators, regulatees
  - Show full diagram or only highlighted genes

Cellular Overview

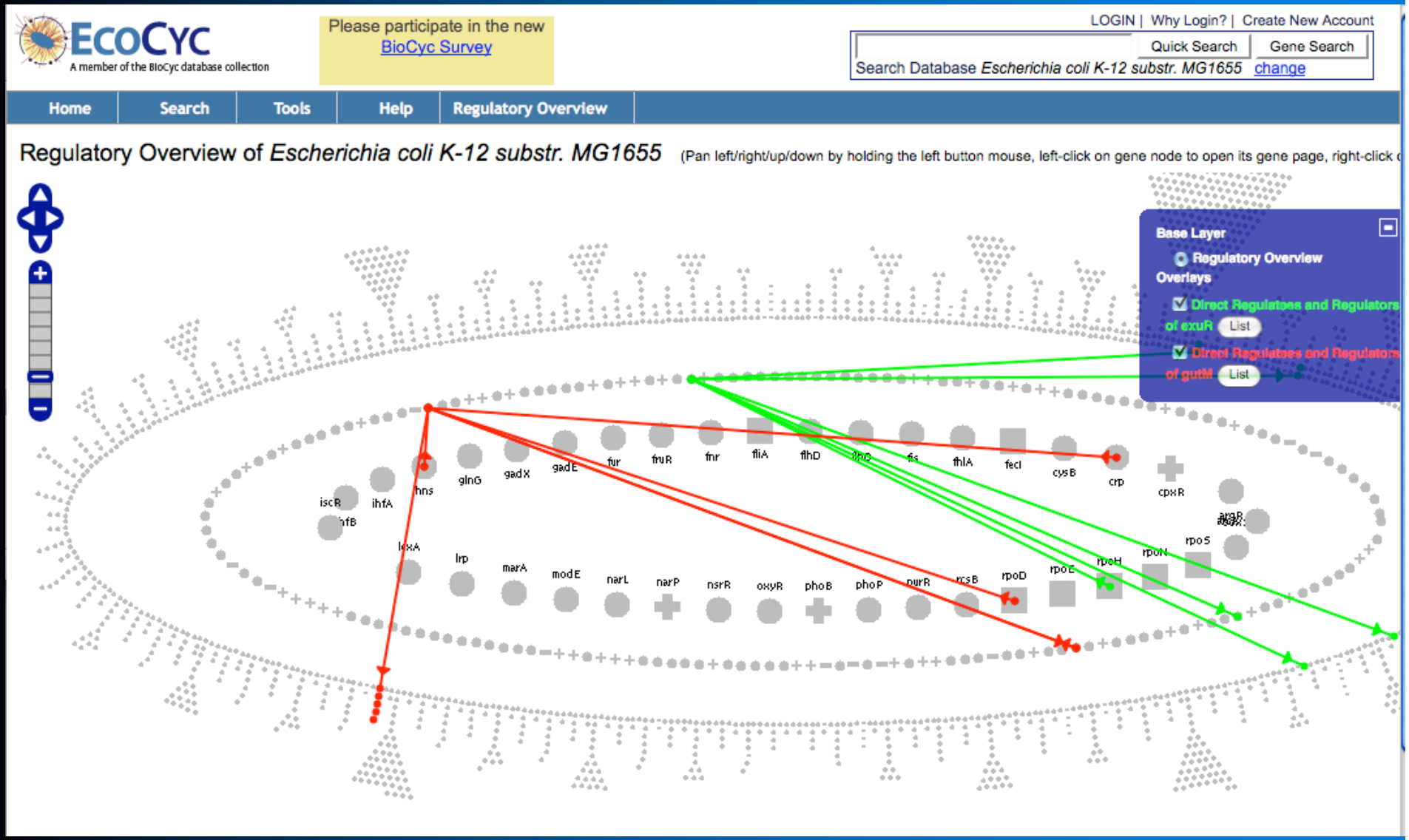


Cellular Overview, zoomed-in view



BioCyc Database Collection

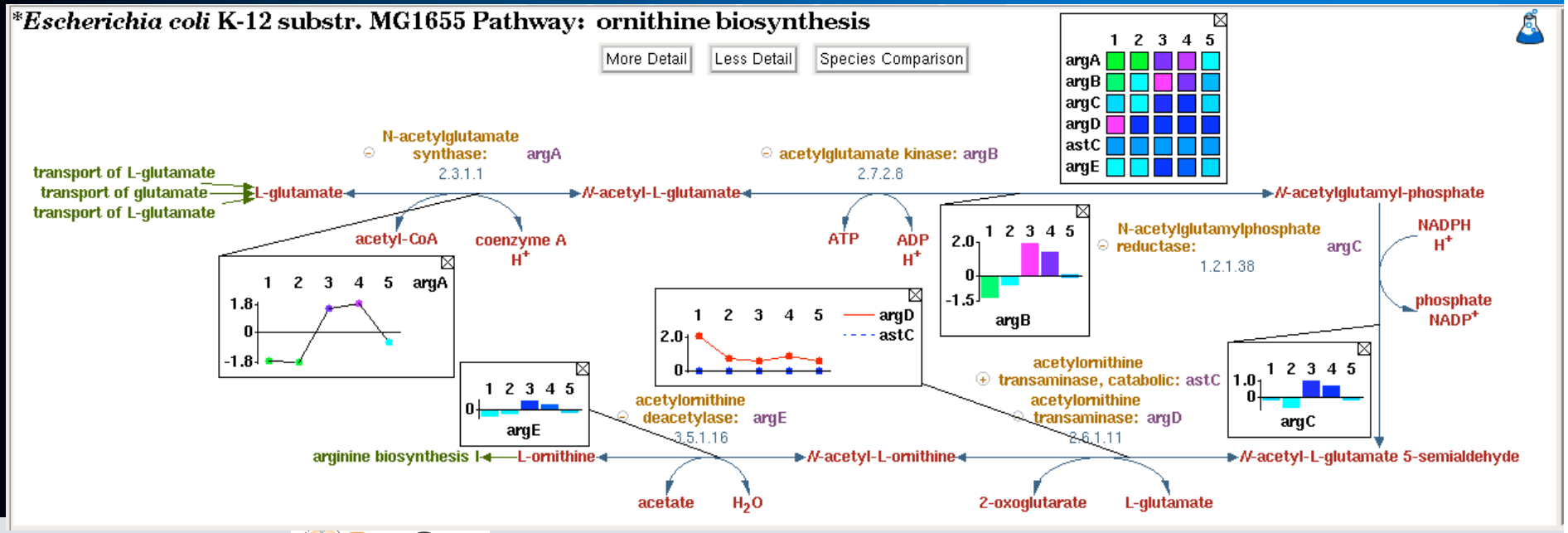
Regulatory Overview



BioCyc Database Collection

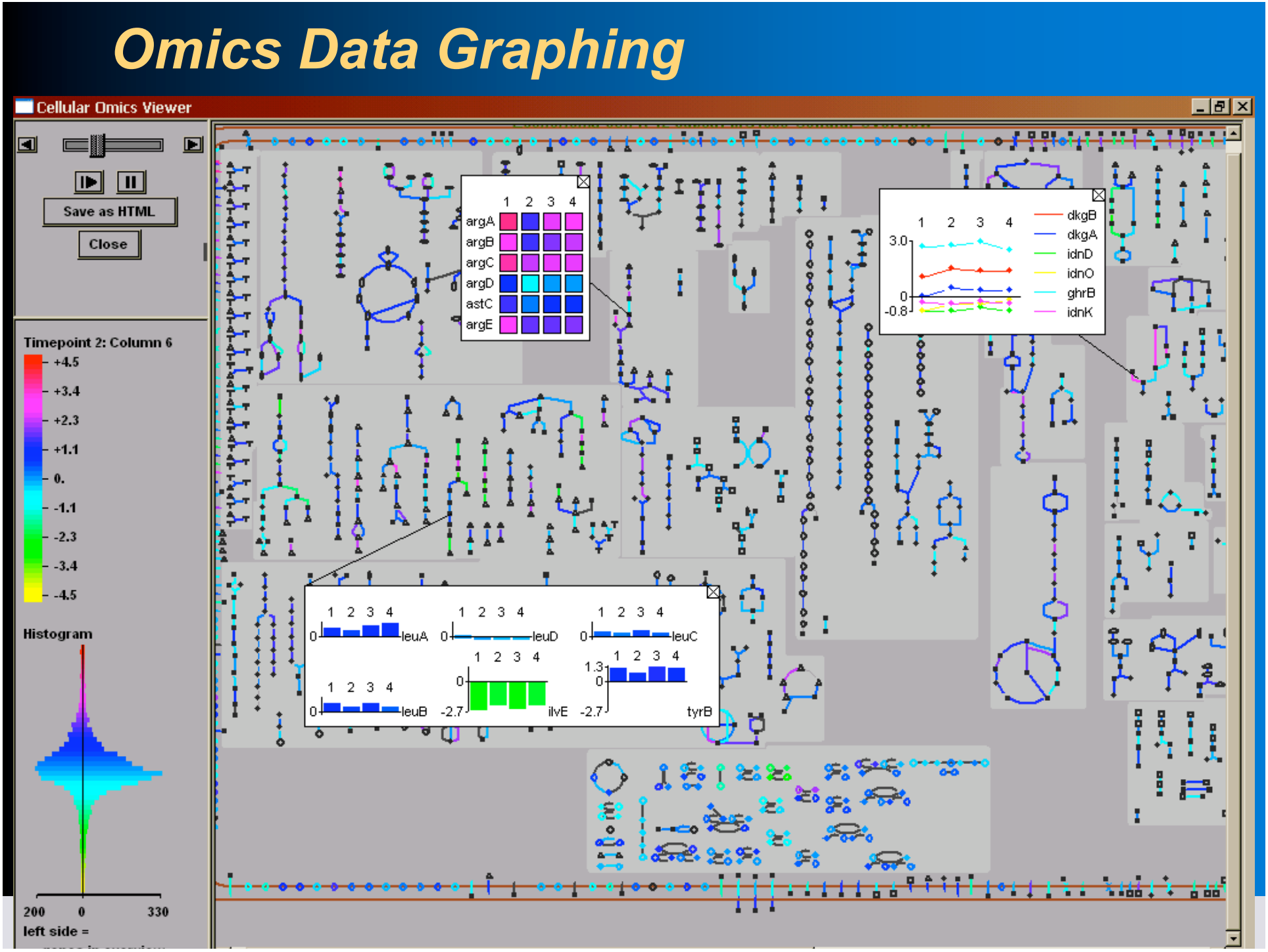
Omics Popups

- Desktop Pathway Tools only
- Can show omics popups for a gene, reaction, pathway
- Use also in Cellular Overview
- Choose from 3 styles: heatmap, bar graph, plot



BioCyc Database Collection

Omics Data Graphing

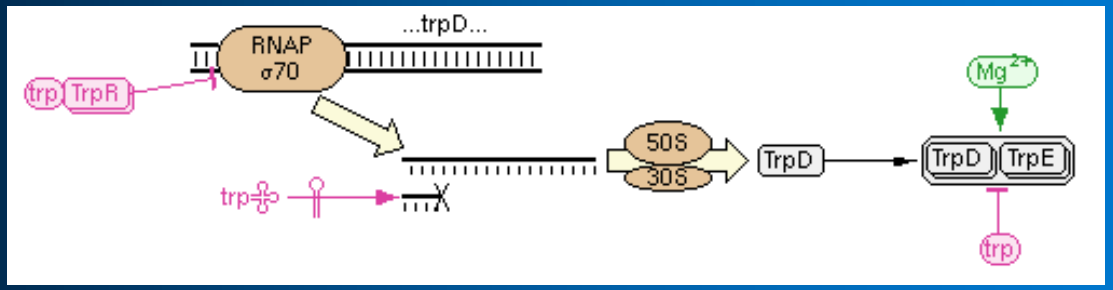


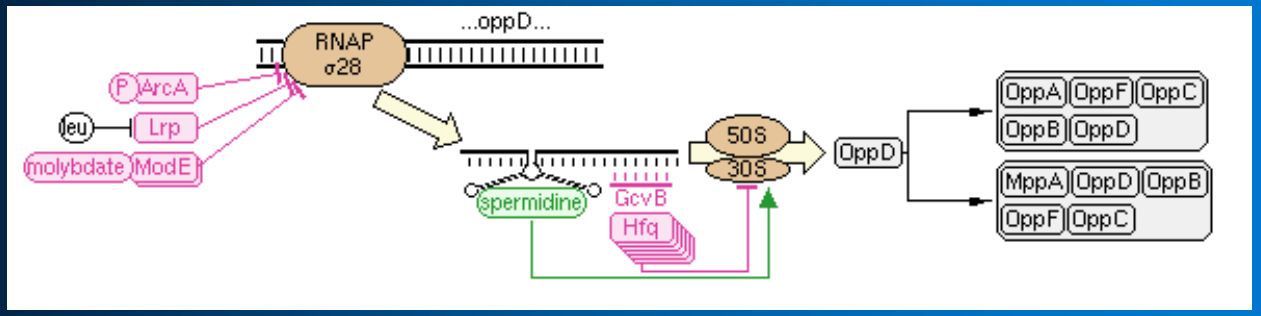
Pathway Tools Captures All Bacterial Regulation Mechanisms

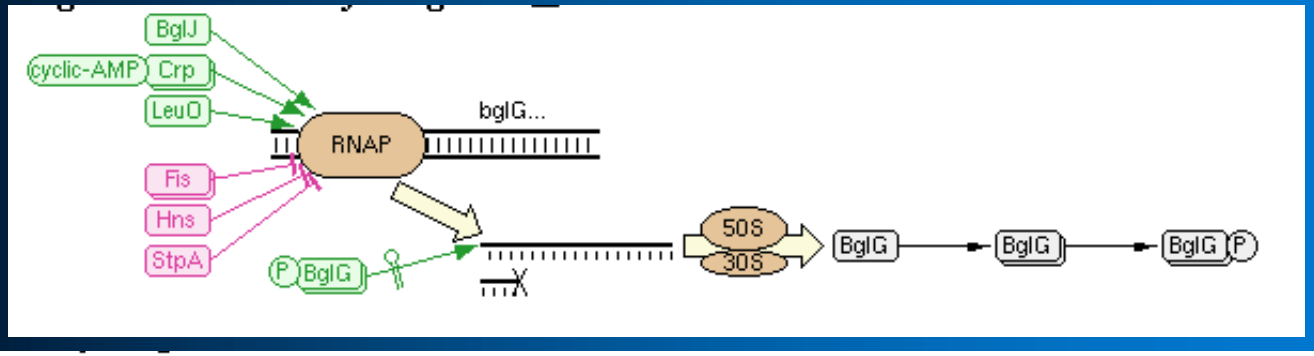
- **Regulation of transcription**
  - By transcription factors
  - By attenuation
- **Regulation of translation**
  - By proteins and small RNAs
- **Regulation of protein activity**
  - By covalent modification (e.g., phosphorylation)
  - By non-covalent modification (e.g., allosteric inhibitors)

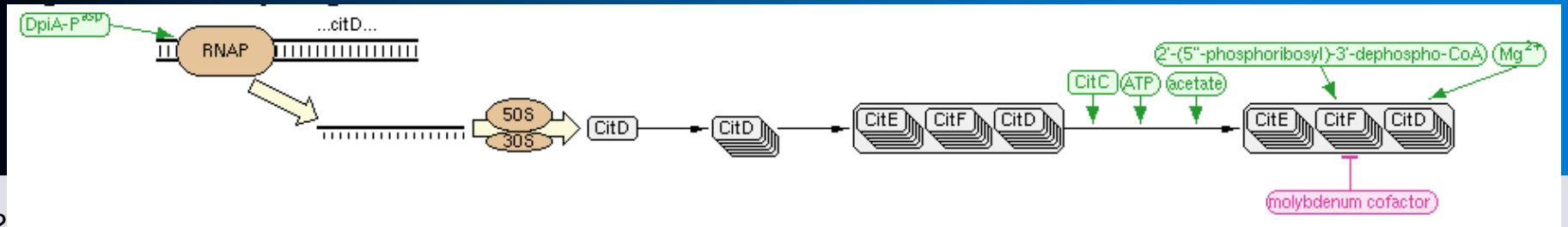
- **Support: Schema, editing tools, display tools**

Regulatory Summary Diagrams









Other Recent Enhancements

- **Phases I and II of upgrade to Pathway Tools Web mode**
  - Phase III still to come
- **Ability to customize pathway displays via Web site**
  - Pathway → Customize

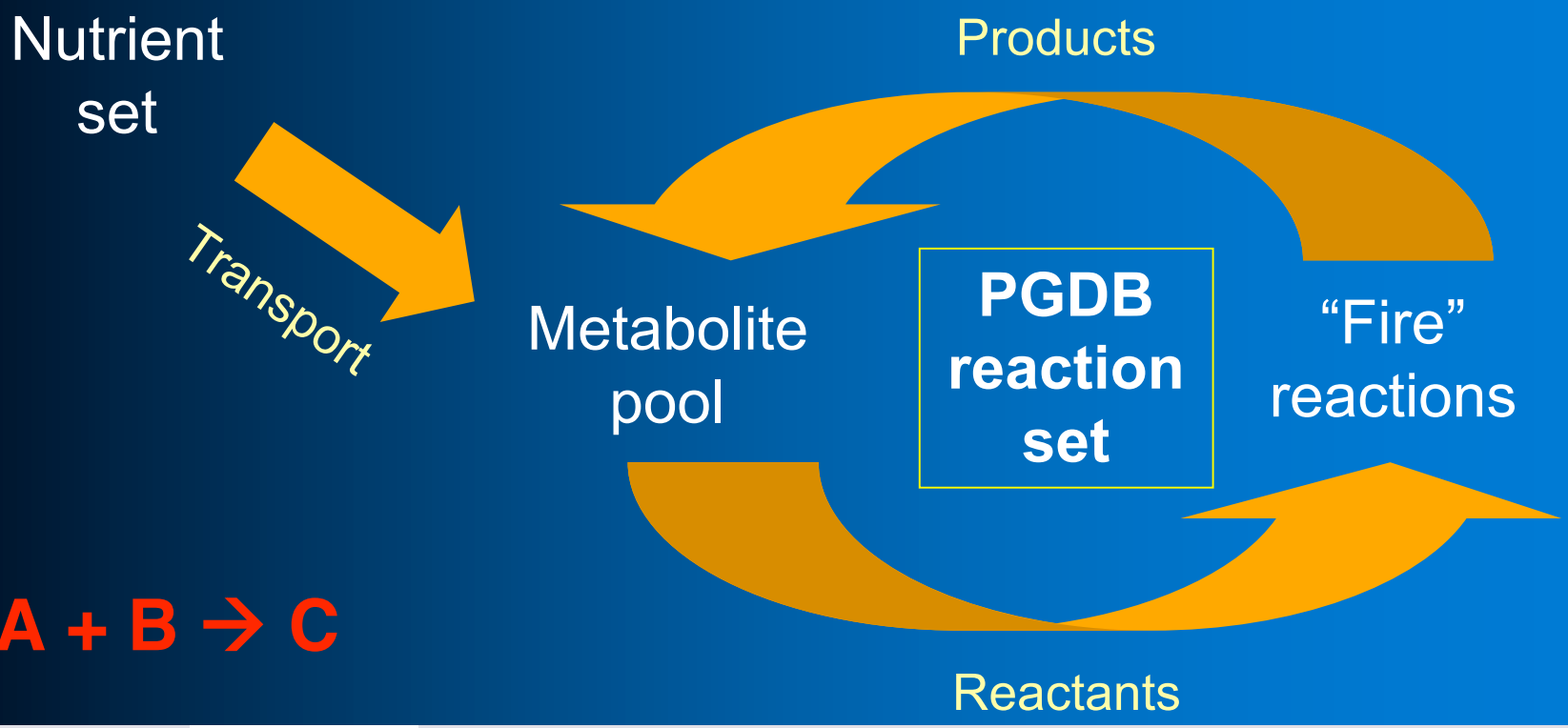
Reachability Analysis of Metabolic Networks

- **Given:**
  - A PGDB for an organism
  - A set of initial metabolites
- **Infer:**
  - What set of products **can be** synthesized by the small-molecule metabolism of the organism
- **Motivations:**
  - Quality control for PGDBs
    - Verify that a known growth medium yields known essential compounds
  - Experiment with other growth media
  - Experiment with reaction knock-outs
- **Limitations**
  - Cannot properly handle compounds required for their own synthesis
  - Nutrients needed for reachability may be a superset of those required for growth

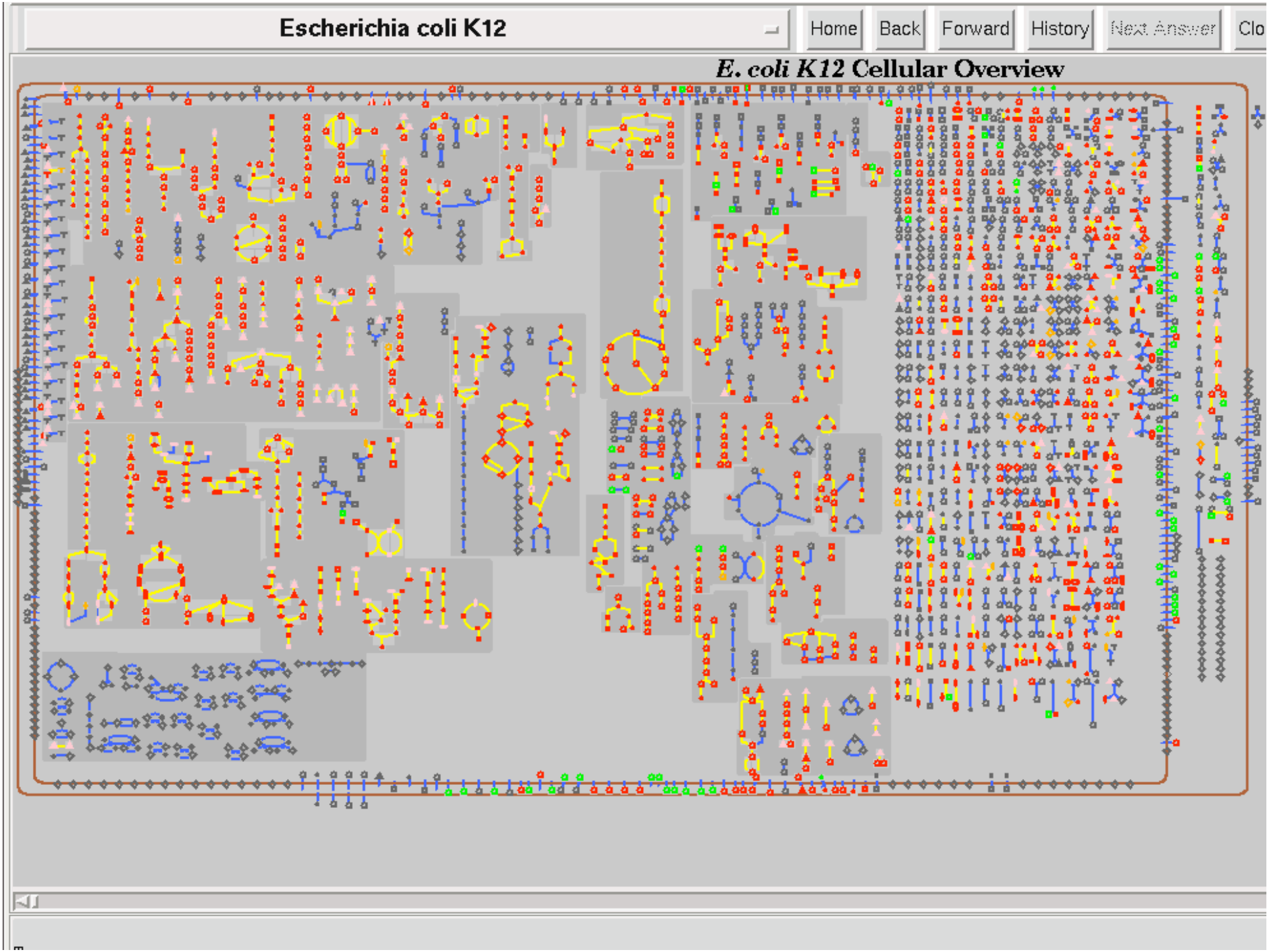
Romero and Karp, ***Pacific Symposium on Biocomputing,* 2001**

Algorithm: Forward Propagation Through Production System

- **Each reaction becomes a production rule**
- **Each of the 21 metabolites in the nutrient set becomes an axiom**



A + B → C


Escherichia coli K12
Home
Back
Forward
History
Next Answer
Clo
E. coli K12 Cellular Overview

Coming Soon

- **BioCyc / EcoCyc / HumanCyc will support Web services for data retrieval**
- **iPhone app for BioCyc / EcoCyc / HumanCyc and other PGDBs**

Acknowledgements

- SRI
  - Suzanne Paley, Ron Caspi, Ingrid Keseler, Carol Fulcher, Markus Krummenacker, Alex Shearer, Tomer Altman, Joe Dale, Fred Gilham, Pallavi Kaipa
- EcoCyc Collaborators
  - Julio Collado-Vides, Robert Gunsalus, Ian Paulsen
- MetaCyc Collaborators
  - Sue Rhee, Peifen Zhang, Kate Dreher
  - Lukas Mueller, Anuradha Pujar
- Funding sources:
  - NIH National Institute of General Medical Sciences
  - NIH National Center for Research Resources

BioCyc.org

Learn more from BioCyc webinars: biocyc.org/webinar.shtml