



***The MetaCyc Database and  
The Pathway Tools Software:  
Resources for Metabolic  
Engineering***

**Peter D. Karp, Ph.D.  
Bioinformatics Research Group  
SRI International  
pkarp@ai.sri.com**

**<http://www.ai.sri.com/pkarp/>  
<http://BioCyc.org/>**

***MetaCyc:***

***MetaCyc.org***

## ***Metabolic Encyclopedia***

- **Nonredundant metabolic pathway database**
- **Describe a representative sample of every experimentally determined metabolic pathway**
- **Literature-based DB with extensive references and commentary**
- **Pathways, reactions, enzymes, substrates**
- **460 pathways, 1267 enzymes, 4294 reactions from 160 organisms**
- ***Nucleic Acids Research* 30:59-61 2002.**
- **Jointly developed by SRI and Carnegie Institution**
  - New focus on plant pathways



## *MetaCyc Frequent Organisms*

<i>E. coli</i>	173
<i>Sm. typhimurium</i>	35
<i>Ho. sapiens</i>	31
<i>Sf. sulfataricus</i>	20
<i>B. subtilis</i>	18
<i>Soybean</i>	18
<i>Pseudomonas</i>	17
<i>Hp. influenzae</i>	15
<i>M. capricolum</i>	12
<i>S. cerevisiae</i>	8
<i>P. putida</i>	7
<i>M. pneumoniae</i>	7

## *MetaCyc Enzyme Data*

- **Reaction(s) catalyzed**
- **Alternative substrates**
- **Cofactors / prosthetic groups**
- **Activators and inhibitors**
- **Subunit structure**
- **Molecular weight, pI**
- **Comment, literature citations**
- **Species**

# Terminology

- **Model Organism Database (MOD)** – DB describing genome and other information about an organism

- **Pathway/Genome Database (PGDB)** – MOD that combines information about

- Pathways, reactions, substrates
- Enzymes, transporters
- Genes, replicons
- Transcription factors, promoters, operons, DNA binding sites

- **BioCyc – Collection of 15 PGDBs at BioCyc.org**

- EcoCyc, AgroCyc, YeastCyc

The screenshot shows the BioCyc Knowledge Library homepage. The browser window title is "BioCyc Knowledge Library". The page content includes:

- Home**: [BioCyc Home Page](#)
- Search**: [Database Search](#), [Advanced Database Search](#), [Simple Search Results](#), [Help](#)
- News**: [06/02/02 In Mind News](#), [06/02/02 New BioCyc Website](#), [06/02/02 Blast Search](#), [06/02/02 MINTROPICUS DB](#), [06/02/02 V. cholerae DB](#), [06/02/02 Updated EcoCyc Web](#), [06/02/02 MetaCyc grant](#), [Links](#)
- Services**: [Software Database Downloads](#), [User Support](#), [Subscribe to Mailing List](#)
- Information**: [Introduction to BioCyc](#), [Publications](#), [Links to BioCyc](#), [External Links](#)
- Databases**: [EcoCyc — Escherichia coli](#), [MetaCyc](#), [AgroCyc — Agrobacterium tumefaciens](#), [BsubCyc — Bacillus subtilis](#), [CtraCyc — Chlamydia trachomatis](#), [CauloCyc — Caulobacter crescentus](#), [HspCyc — Helicobacter pylori](#), [HimCyc — Haemophilus influenzae](#), [MtbRvCyc — Mycobacterium tuberculosis](#), [MpnCyc — Mycoplasma pneumoniae](#), [PseudoCyc — Pseudomonas aeruginosa](#), [YeastCyc — Saccharomyces cerevisiae](#), [TriaCyc — Treponema pallidum](#), [VibCyc — Vibrio cholerae](#)
- BioCyc Home Page**: The BioCyc Knowledge Library is a collection of Pathway/Genome Databases. Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism, with the exception of the MetaCyc database, which is a reference source on metabolic pathways from many organisms. [more...](#)
- Literature-derived Pathway/Genome Databases**:
  - EcoCyc — [Escherichia coli](#)
  - MetaCyc — Metabolic pathways and enzymes from 150 species
- Computationally-derived Pathway/Genome Databases**:
  - AgroCyc — [Agrobacterium tumefaciens](#)
  - BsubCyc — [Bacillus subtilis](#)
  - CtraCyc — [Chlamydia trachomatis](#)
  - CauloCyc — [Caulobacter crescentus](#)
  - HspCyc — [Helicobacter pylori](#)
  - HimCyc — [Haemophilus influenzae](#)
  - MtbRvCyc — [Mycobacterium tuberculosis](#)
  - MpnCyc — [Mycoplasma pneumoniae](#)
  - PseudoCyc — [Pseudomonas aeruginosa](#)
  - YeastCyc — [Saccharomyces cerevisiae](#)
  - TriaCyc — [Treponema pallidum](#)
  - VibCyc — [Vibrio cholerae](#)
- Acknowledgments**: The authors of each BioCyc database are listed on the database summary page for each database. The Defense Advanced Research Projects Agency (DARPA) funds development of the MtbRvCyc and the VchoCyc databases under contract N66001-01-C-8011. The Department of Energy funds development of the CauloCyc database under grant DE-FC03-01ER63219. The NIH National Institute for General Medical Sciences funds development of the MetaCyc database under grant 1-R01-GM5466-01. The NIH National Center for Research Resources funds development of the EcoCyc database under grant 1-R01-RR07961-01.

# *Pathway Tools Software*



- **PathoLogic**

- Prediction of metabolic network 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

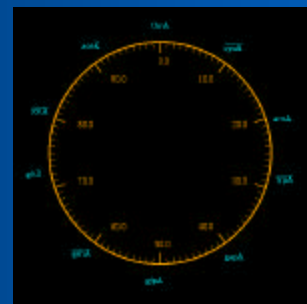
- **Bioinformatics 18:S225 2002**

# Pathway Tools Algorithms

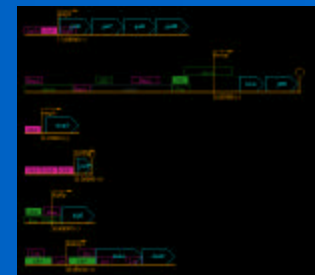
- Visualization and editing tools for following datatypes
- Full Metabolic Map
  - Paint gene expression data on metabolic network; compare metabolic networks
- Pathways
  - Pathway prediction
- Reactions
  - Balance checker
- Compounds
  - Chemical substructure comparison
- Enzymes, Transporters, Transcription Factors
- Genes
- Chromosomes
- Operons
  - Operon prediction; visualize genetic network



Gene	Product	Enzyme	Reaction
...	...	...	...
...	...	...	...
...	...	...	...



Gene	Product	Enzyme	Reaction
...	...	...	...
...	...	...	...
...	...	...	...



# BioCyc Collection of Pathway/Genome DBs



- Literature-based Datasets:

- *MetaCyc*

- *Escherichia coli (EcoCyc)*

## Computationally Derived Datasets:

- *Agrobacterium tumefaciens*

- *Caulobacter crescentus*

- *Chlamydia trachomatis*

- *Bacillus subtilis*

- *Helicobacter pylori*

- *Haemophilus influenzae*

- *Mycobacterium tuberculosis RvH37*

- *Mycobacterium tuberculosis CDC1551*

- *Mycoplasma pneumonia*

- *Pseudomonas aeruginosa*

- *Saccharomyces cerevisiae*

- *Treponema pallidum*

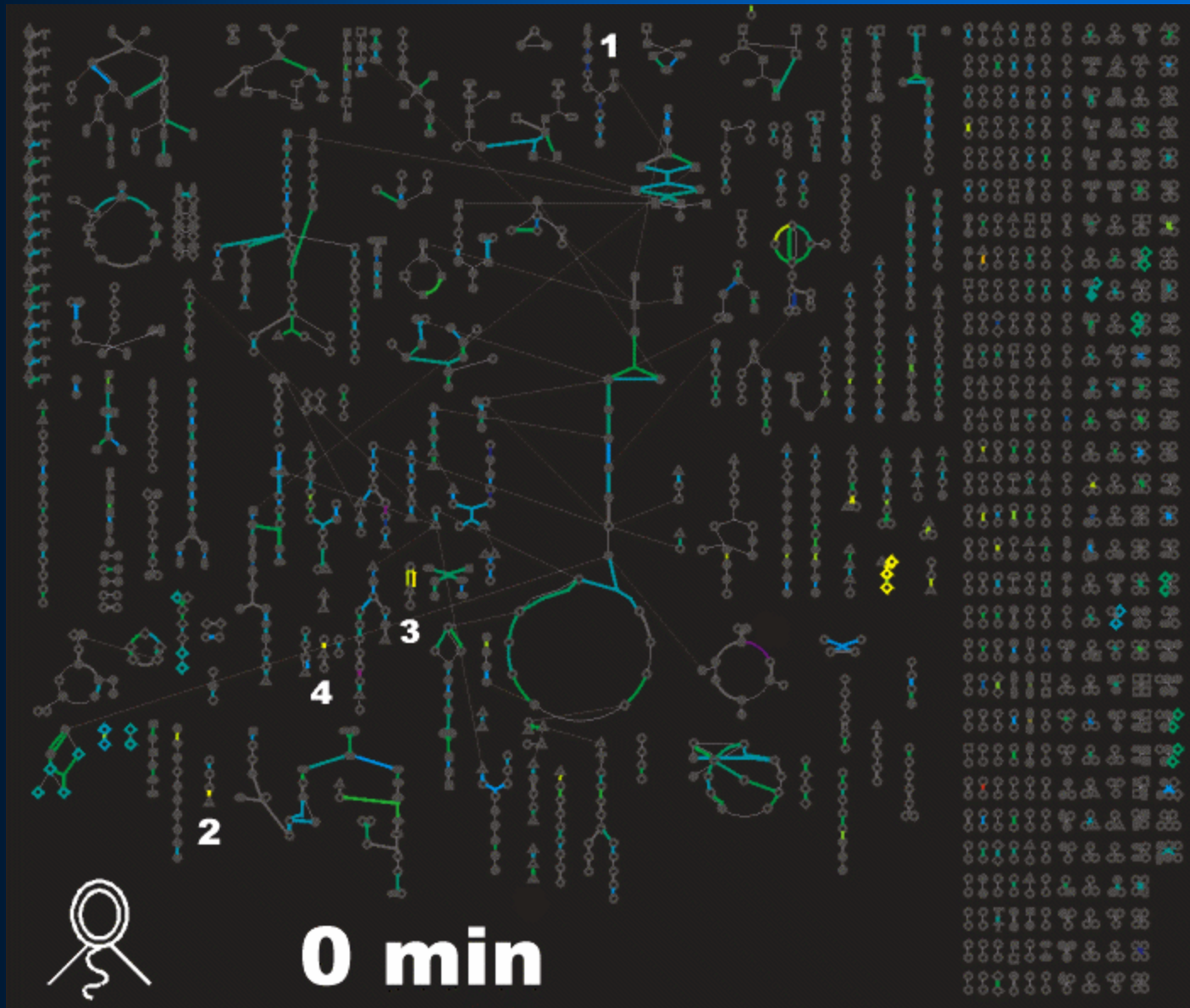
- *Vibrio cholerae*

- *Yellow = Open Database*

<http://BioCyc.org/>



# *C. crescentus* Cell Cycle Gene Expression





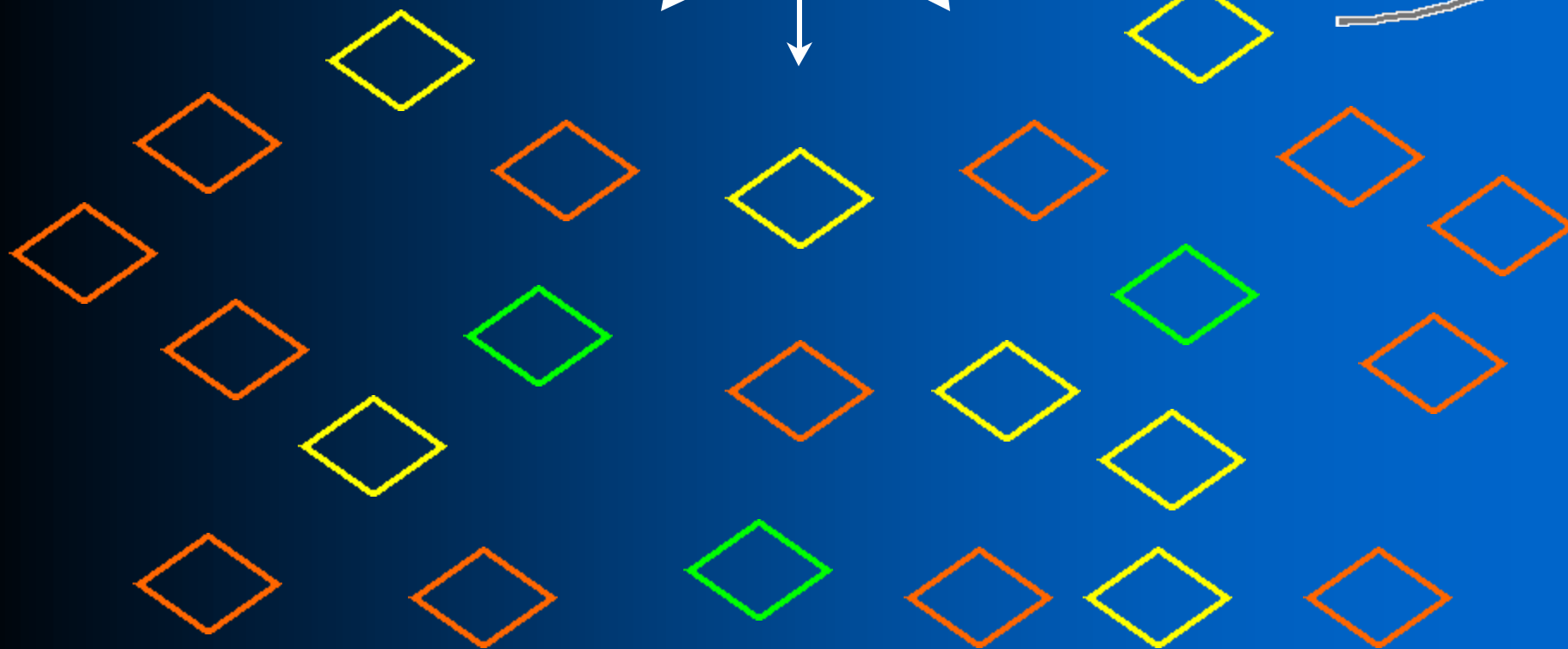
## ***Pathway/Genome DBs Created by External Users***

SRI International  
Bioinformatics

- ***Plasmodium falciparum, Stanford University***
  - [plasmocyc.stanford.edu](http://plasmocyc.stanford.edu)
- ***Arabidopsis thaliana and Synechocystis, Carnegie Institution of Washington***
  - [Arabidopsis.org:1555](http://Arabidopsis.org:1555)
- **Other PGDBs in progress by 10 other users**

# *Family of Pathway/Genome Databases*

SRI International  
Bioinformatics



# **Acknowledgements**

## ● **SRI**

- Suzanne Paley, Pedro Romero, John Pick, Cindy Krieger, Martha Arnaud

## ● **EcoCyc Project**

- Julio Collado-Vides, Ian Paulsen, Monica Riley, Milton Saier

## ● **MetaCyc Project**

- Sue Rhee, Lukas Mueller, Peifen Zhang, Chris Somerville

## ● **Stanford**

- Gary Schoolnik, Harley McAdams, Lucy Shapiro, Russ Altman, Iwei Yeh

## ● **Funding sources:**

- NIH National Center for Research Resources
- NIH National Institute of General Medical Sciences
- NIH National Human Genome Research Institute
- Department of Energy Microbial Cell Project
- DARPA BioSpice, UPC

**BioCyc.org**

