

Modeling Metabolic Pathways- Glycolysis: a few reactions

By

Raquell M. Holmes, Ph.D.

BioQUEST

March, 2005

Computational Biology

- Computational Biology \neq Bioinformatics
 - More than sequences, database searches, and statistics.
- A part of Computational Science
 - Using mathematical modeling, simulation and visualization
 - Complementing theory and experiment

Today's journey

- Bioinformatics
 - From protein sequence
 - Metabolic pathway databases
- Metabolic behaviors
- Computer modeling glycolysis

Bioinformatics

Common starting point

Generic Protein Seq. Record

- Sequence: hexokinase
- E.C. #: 2.7.1.1
- Publication: Stachelek et al 1986
- Organism: Yeast
- Function: main glucose phosphorylating enzyme
- Links: other databases or tools
- Pre-determined Properties:
 - families, folds...
- Swissprot example for enzyme.

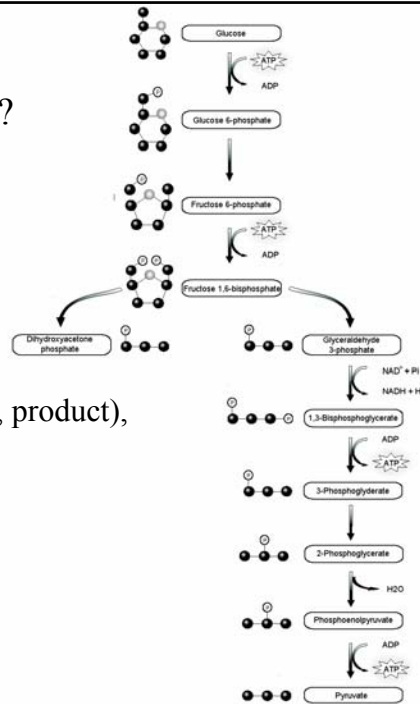
What we learn from protein sequence?

- Similarity searches :
 - Homology with other proteins
 - Conserved domains
 - Active sites
- How do we find the related pathway?
- Are there metabolism databases?

What is a metabolic pathway?

Contains a series of reactions.

Reactions: **Metabolites** (substrate, product),
enzyme, co-factors.



Databases on Molecular Networks

Metabolic Pathways from NAR (5):

- EcoCyc: <http://ecocyc.org>
 - Began with E. coli Genes and Metabolism
 - BioCyc includes additional genomes
- KEGG:
<http://www.genome.ad.jp/kegg/>
 - Encyclopedia of genes and genomes
- Others: WIT2, PathDB, UMBDD

Starting out...

- Metabolism pathway databases
 - Search by name or sequence
 - Compounds, Reactions, Pathway, Genes
 - Associated information
 - Formulas, Names, Synonyms, Links to other databases

Example search results

KEGG: search by compound or enzyme by key word.

Glucose: 82 hits

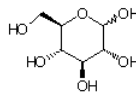
1. cpd:C00029 UDPglucose; UDP-D-glucose; UDP-glucose; Uridine diphosphate glucose
2. cpd:C00031 D-Glucose; Grape sugar; Dextrose
3. cpd:C00092 D-Glucose 6-phosphate; Glucose 6-phosphate; Robison ester
-
79. cpd:C11911 dTDP-D-desosamine; dTDP-3-dimethylamino-3,4,6-trideoxy-D-glucose
80. cpd:C11915 dTDP-3-methyl-4-oxo-2,6-dideoxy-L-glucose
81. cpd:C11922 dTDP-4-oxo-2,6-dideoxy-D-glucose
82. cpd:C11925 dTDP-3-amino-3,6-dideoxy-D-glucose

Grape sugar: FORMULA C₆H₁₂O₆

NAME D-Glucose grape sugar Dextrose

List of reactions involving compound

List of [enzymes](#) acting on compound



C00031

Enzyme	Reaction	Pathway	Compounds
2.7.1.1	R0029	Map0010	C00031

Database Content

Over 11 Data Categories

- Sequence: DNA, RNA, **protein**
 - Structure: genomics, protein, carbohydrate
 - Networks: **metabolic enzymes and pathways** (signaling)
 - Organisms: human/vertebrate, human genes and diseases
 - Expression: microarray and gene expression, proteomics
-
- Nucleic Acid Research
 - Dedicated to review of databases (548-'04, 386-'03)
 - <http://nar.oupjournal.org>
 - Most discussion focus on sequence databases.

Where data comes from...Exp. Why do we care?

- Experiments
 - Sequencing data: Gene or protein identity
 - Enzymatic assays: Biochemical properties
 - Expression studies: Localization, putative cellular function, regulation patterns
 - Protein interactions: complexes and networks

Where data comes from...Comp.

Why do we care?

- Annotated sequences
 - Align sequences (full/partial)
 - Homology to other genes, Identity
 - Pattern recognition, property predictions
 - Biochemical properties
 - Motifs, profiles, families
 - Enzyme activity and structure prediction
 - Pathways
 - Homologous function

EcoCyc Encyclopedia of *Escherichia coli* K12 Genes and Metabolism

EcoCyc Home

Quick Search

[Database Search](#)
[Advanced Database Search](#)
[BLAST](#)

Browse
[Pathways](#)
[Genes](#)
[Reactions](#)
[Compounds](#)
[Metabolic Chart](#)
[Expression Viewer](#)

About EcoCyc
[Project Overview](#)
[Guided Tour](#)
[Publications](#)
[Update History](#)
[Advisory Board](#)
[Contributors](#)

Services
[Software/Data Download](#)
[User Support](#)
[Subscribe to Mailing List](#)
[Gene Cross-Reference](#)
[E. coli Links](#)
[Contact Us](#)
[T-shirt](#)

Home page

- Navigation bar
 - List of categories
 - **Compounds**
 - List of results
 - All-Carbohydrates
 - a sugar
 - a sugar phosphate
 - a sugar-1-phosphate
 - Aldonic-Acids
 - Carbohydrate-Derivatives
 - Carbohydrates
 - Disaccharides
 - Oligosaccharides
 - Polysaccharides
 - Fructans
 - Glycogens
 - Large-branched-glucans
 - Long-linear-glucans
 - Pectin
 - Short-glucans
 - Starch

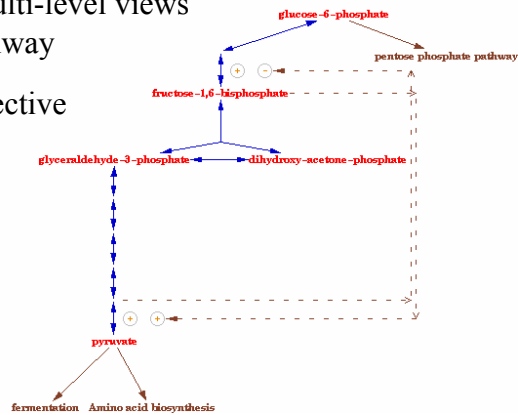
- Keyword search
 - [Results](#) are a list of data
 - Proteins, compounds, reactions, pathways
 - Hunt and peck

Visualizing Data

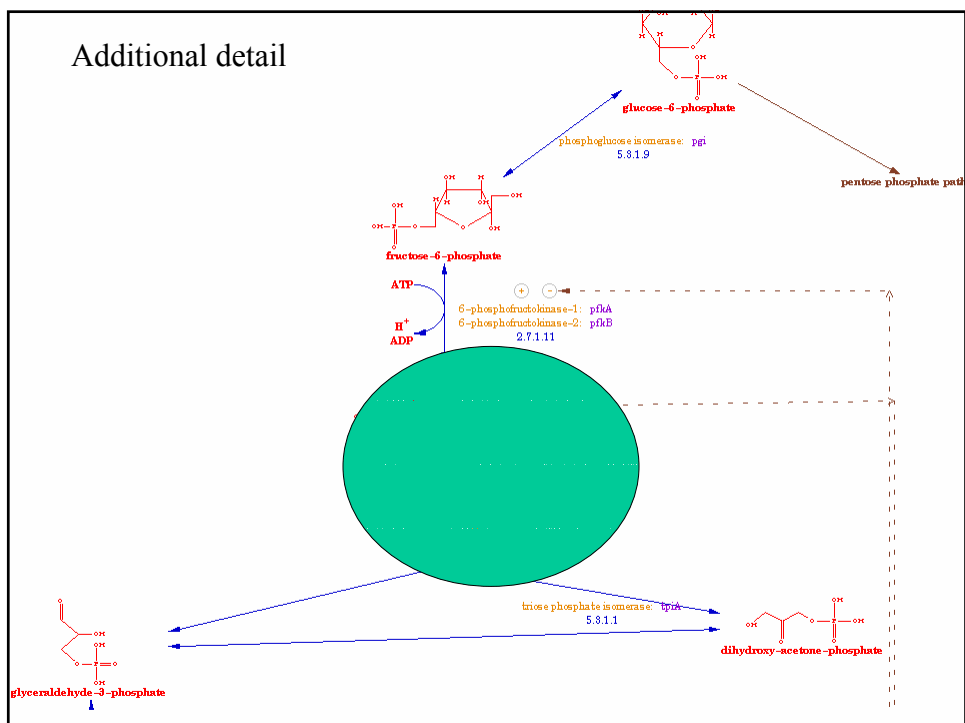
Visualization: Data levels

EcoCyc draws multi-level views
Based on the pathway

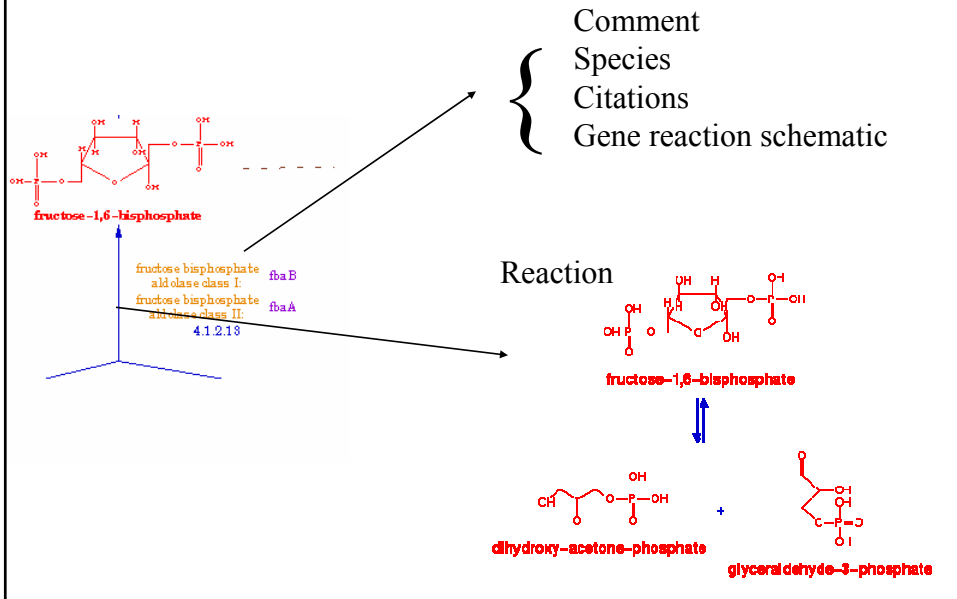
Metabolite perspective



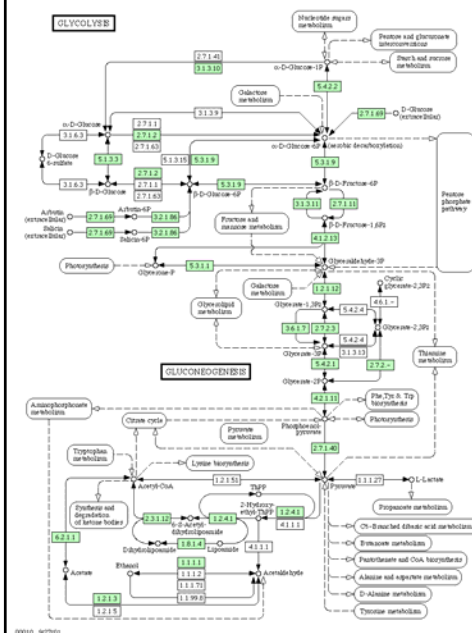
Additional detail



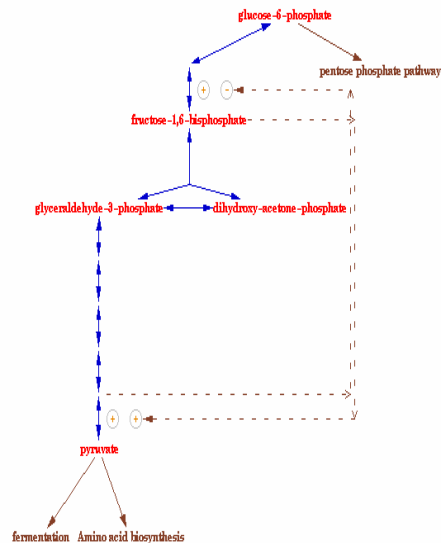
E. coli K-12 Reaction: 4.1.2.13



KEGG: E.coli



EcoCyc: E.Coli



Summary I

- Protein ->enzyme->pathway database
- Pathway Database Content:
 - Species specific (BioCyc), general (KEGG)
 - Known and proposed enzymes, co-factors, metabolites.
 - Searches return similar results (compounds, reactions...) with different appearance.

Different species

<http://biocyc.org/server.html>

BioCyc Query Page

Curator positions available at
Carnegie Institution and SRI

This form provides several different mechanisms for querying Pathway/Genome Databases.

Select a dataset:

Query

To retrieve objects by name
name as a substring will be

Choose from a

Browse Ontology

Each dataset contains class
classification system to bro

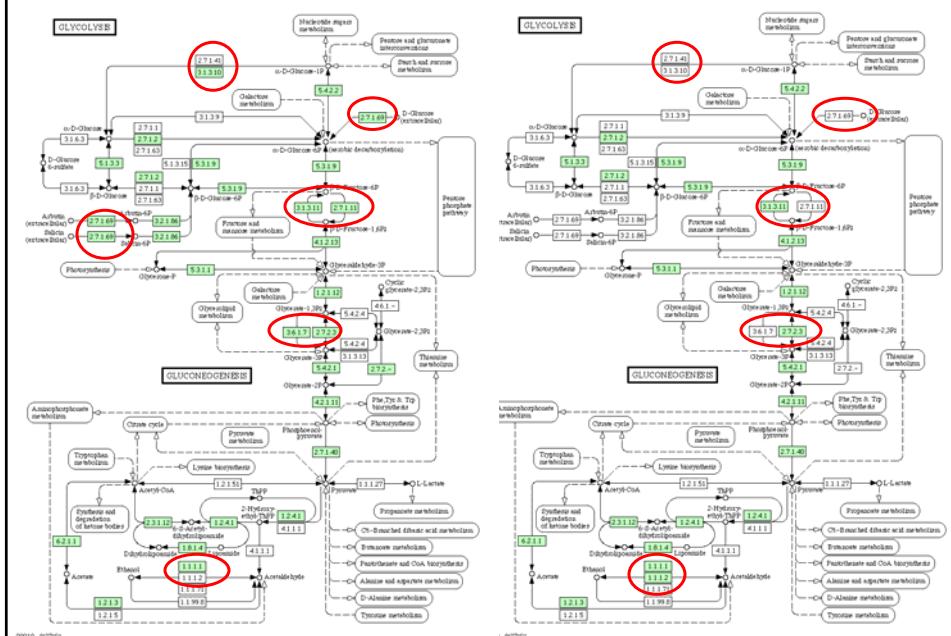
A. tumefaciens C58
B. subtilis
Cb. crescentus
C. trachomatis D-UW-3-CX
E. coli K-12
E. coli O157:H7
Hm. influenzae
Hb. pylori 26695
H. sapiens
MetaCyc
M. tb. CDC1551
M. tb. H37Rv
Mp. pneumoniae
P. falciparum 3D7
S. flexneri 2457T
T. pallidum Nichols
V. cholerae N16961

ect you wish to retrieve, then enter the name of the object
er multiple names or EC numbers, separating them with

thways, for reactions (the enzyme nomenclature system),

- Links to summary information about the selected organism:

Pseudomona



See Exercise handout for ideas

Populating metabolic databases

Where do the pathways come from?

Databases: KEGG, WIT, **BioCyc**

Resources:

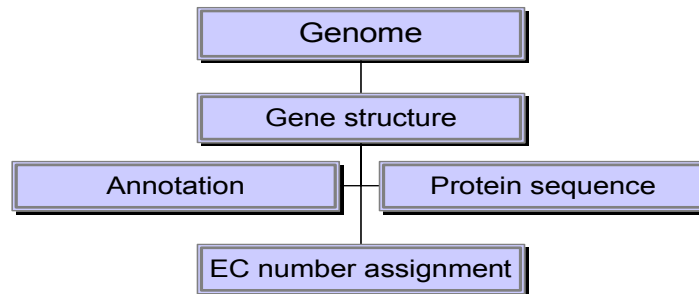
- Experimental data based on literature

- Genomic data from other databases

Determination of metabolic pathway

- Comparison to known pathways.

Linking Genome to Reactions: EC



[EC #](#) provides information on catalyzed reaction and synonyms.

Search for EC# in EcoCyc database.

Assign reactions and possible pathway association.

How correct is the pathway assignment?

PathLogic/EcoCyc scoring new species pathway

X=# reactions for pathway 4

Y=# reactions found 2

Z=# found in other pathways 1

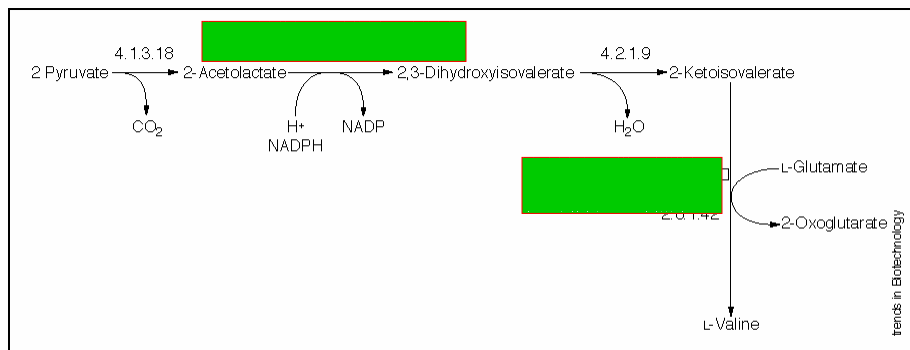


Figure 2

The predicted *Helicobacter pylori* pathway for valine biosynthesis, to which PathoLogic assigned a score of 4,2,1.

Karp et al., 1999. TIBT

Probably, possibly, not

- Probability score depends on X:Y ratio
- 4,2,1 has a 4:2 ratio which equals 0.5
- Probable $X:Y \geq 0.5$
- Possible $0 < X:Y < 0.5$
- Not $Y=0$

Pathway Evidence Glyph



[Homo sapien](#) glycolysis pathway

[E.Coli](#) pathway

Key to edge colors:

- green: reactions in which the enzyme is present
- black: reactions for which the enzyme is not identified in this organism
- orange: reactions in which the enzyme is unique to this pathway
- magenta: reactions that are spontaneous, or edges that do not represent reactions at all (e.g. in polymerization pathways)

Summary II

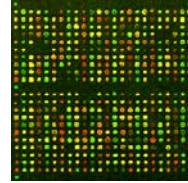
- Genomes used to create database of reactions
- EC# link gene product to enzyme in reactions
- Pathways in the database vary in degrees of probability.

From Data to Dynamics

- Static Data
- What is the behavior of the pathway?
 - Expression data
 - Dynamics

Expression data

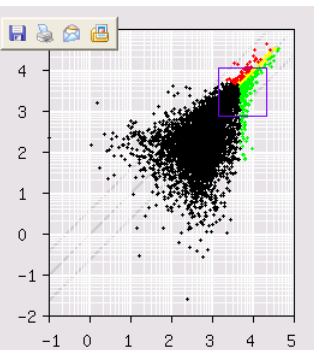
- KEGG:
 - retrievable expression data sets
- EcoCyc:
 - input expression data to view in relation to metabolic data.
- Expression data is one way of viewing the behavior of a system.



Submitted by:
Hirota MORI

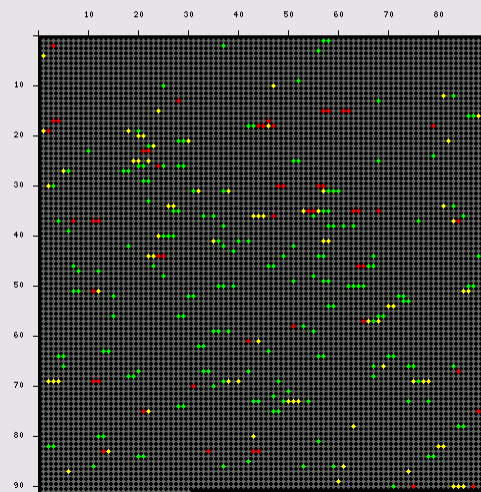
Organism: E.coli

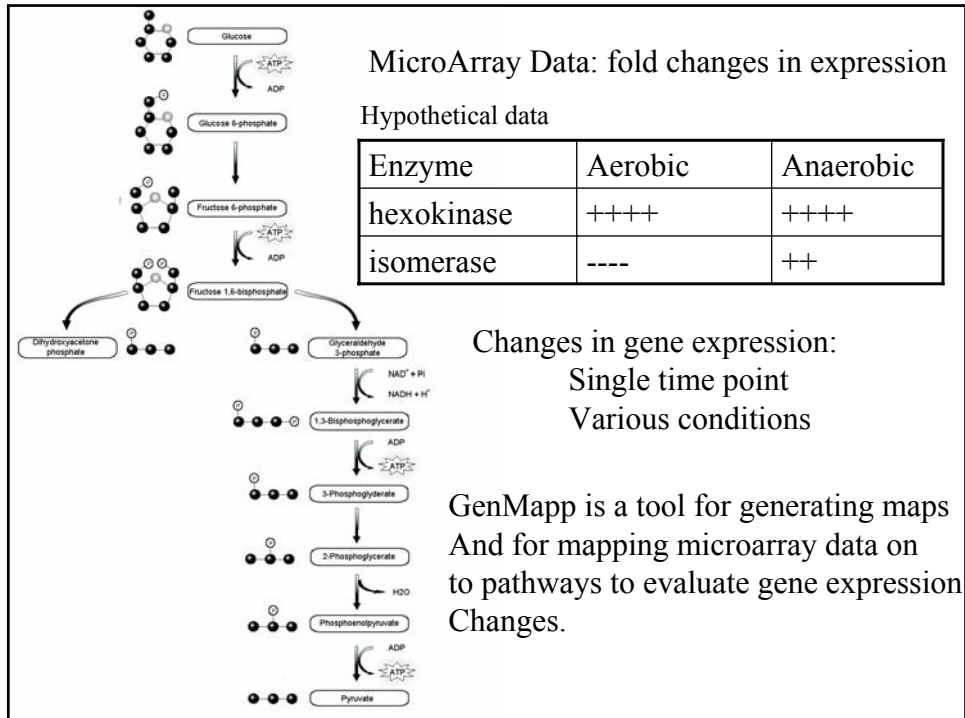
[Raw data](#)



KEGG/Expression - Array image whole view
Entry: ex0000287 Organism: E.coli_J Intensity threshold: 5000 Ratio threshold: 1.2
GENOME PATHWAY List
Red: ☒ Green: ☒ Yellow: ☒ Black: ☐

Search[KEGG ID: Search]





Dynamics: Answering different questions

- If glucose concentration is 20 mM outside of the cell,
 - How quickly is glucose converted to glucose 6-phosphate?
 - How does the concentration of glucose 6 phosphate change over time?

Yeast Glycolysis

suspended

Liquid, flasks, temp, ph



Experiment:
Sample media or characterize
cell content over time

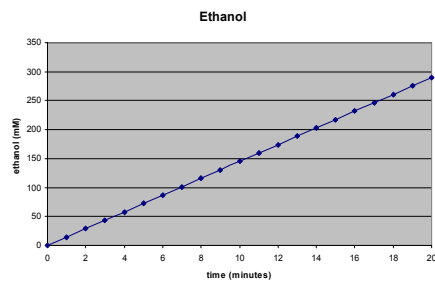
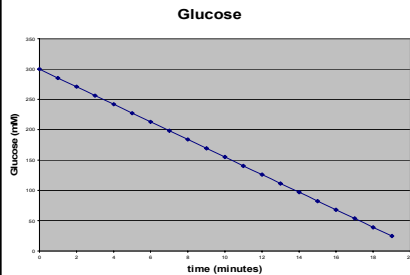
Repeat under different conditions

immobile

Agar, temp, ph



Results may look like



Computer Modeling

- Method for analyzing what we know about a biological process
- Used to describe mechanisms behind changes
- Determines what can be seen or predicted

Walking through a Computational Model

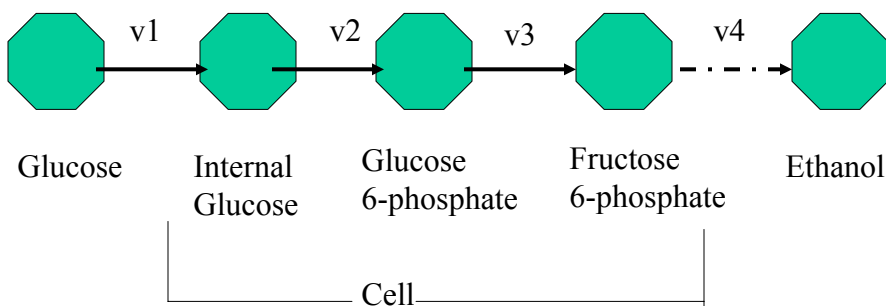
- Concept Map
- Factors and relationships between factors
- Describe relationships mathematically
- Solve equations: using computer tools
- View and interpret results

Designing a dynamic experiment

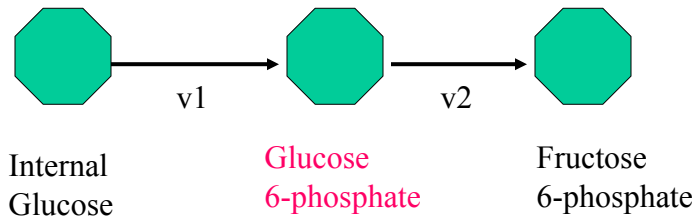
- What **components** are involved?
 - Glucose, glucose 6 phosphate, fructose 6 phosphate...
- What **chemical reactions** involved?
 - Transport, chemical conversions...

Glycolysis: Concept Map

Often drawings, schematics or chemical reactions



Examples of relationships



[Glucose 6-phosphate] is determined by increase from Glucose conversion and decrease by conversion to Fructose 6-phosphate

Amount of glucose 6 phosphate = amount produced - amount converted

Designing a dynamic experiment

- Describing relationship **mathematically**

Relationship in terms of rates of change

The **rate of change** of Glucose-6-phosphate (S_2) is the **rate** of Glucose conversion (v_1) minus the **rate** of conversion (v_2) to Fructose-6-phosphate.

$$\frac{dS_2}{dt} = v_1 - v_2$$

Designing a dynamic experiment

- Describing relationship mathematically
- What **rate laws** are known to describe the enzymatic reaction?
 - Types of rate laws/kinetic models
 - Constant, mass action, michaelis menten...

Simplify

- Glucose transport (v_1)
- Facilitated diffusion

$$v = V^+ \frac{\frac{Glc_{out} - Glc_{in}}{K_{Glc}}}{1 + \frac{Glc_{out}}{K_{Glc}} + \frac{Glc_{in}}{K_{Glc}} + K_i \frac{Glc_{out} Glc_{in}}{K_{Glc}}}$$

Rate Equations

Substrates

- Glucose: S_1
- Glucose-6-phosphate: S_2

• Mass action kinetics are used here to describe the enzymatic reactions.

$$v_1 = k_1 S_1$$

$$v_2 = k_2 S_2$$

Rate constants

- Enzyme1: k_1
- Enzyme2: k_2

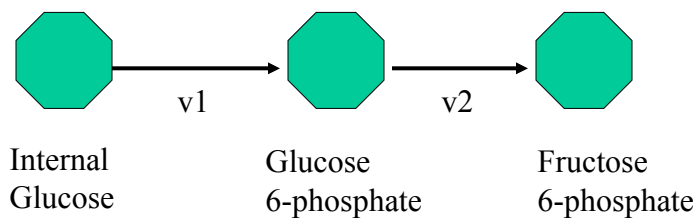
• This is a simplification of the enzyme kinetics for this example.

Initial conditions

- Concentrations of components
 - External glucose (i.e. 20mM)
- Enzymatic rates
 - Rate constant k (i.e. 50mM/min)
 - Michaelis-Menten constants, Hill Coefficients

Values taken from Richard et. al., yeast

The model



Ordinary differential equation

$$\frac{dS_2}{dt} = v_1 - v_2$$

Rate equations

$$v_1 = k_1 S_1$$

$$v_2 = k_2 S_2$$

Initial conditions*

$$S_1 = 20\text{mM}$$

$$k_1 = 55\text{mM/min}$$

$$k_2 = 9.8\text{mM/min}$$

*Taken from....

Walking through a Computational Model

- Concept Map
 - Factors and relationships between factors
 - Describe relationships mathematically
-
- Solve equations: using computer tools
 - View and interpret results

Some Available Tools

General

- | | |
|-------------|---|
| 1. Stella | 1. Concept mapping and system dynamics. |
| – Install | • Glycolysis examples available at |
| – Mac or PC | • http://eot.bu.edu/ccb/GettingStarted/Modeling/biologicalProcess.htm |
| 2. Excel | |
| – Install | |
| – Mac or PC | 2. Discrete events, algebraic equations |

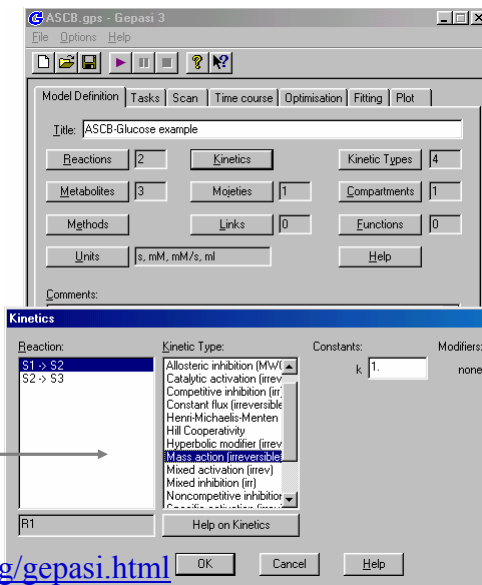
Customized

- | | |
|-----------------|---|
| 3. GEPASI | 3. Biochemical kinetics and kinetic analyses. |
| – Install | |
| – Mac or PC | |
| 4. Virtual Cell | 4. Icon mapping, dynamics and space |
| – Browser: Java | |
| – Mac or PC | |

GEPASI

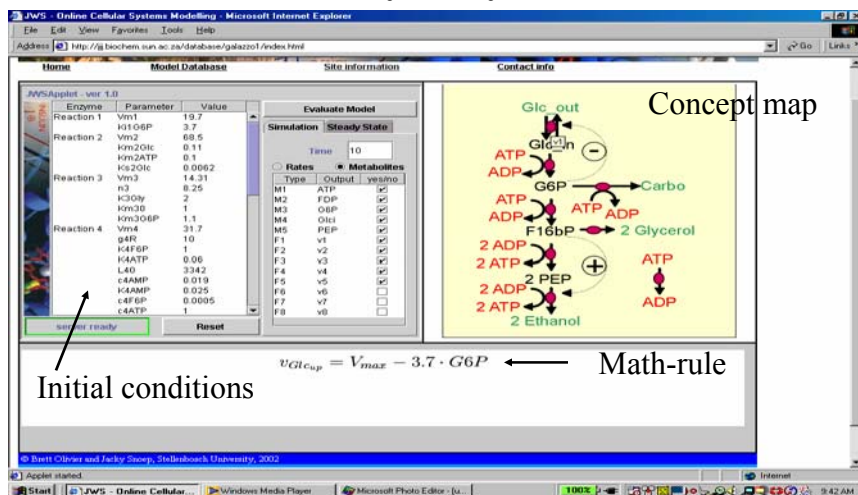
Chemical Equations

Math



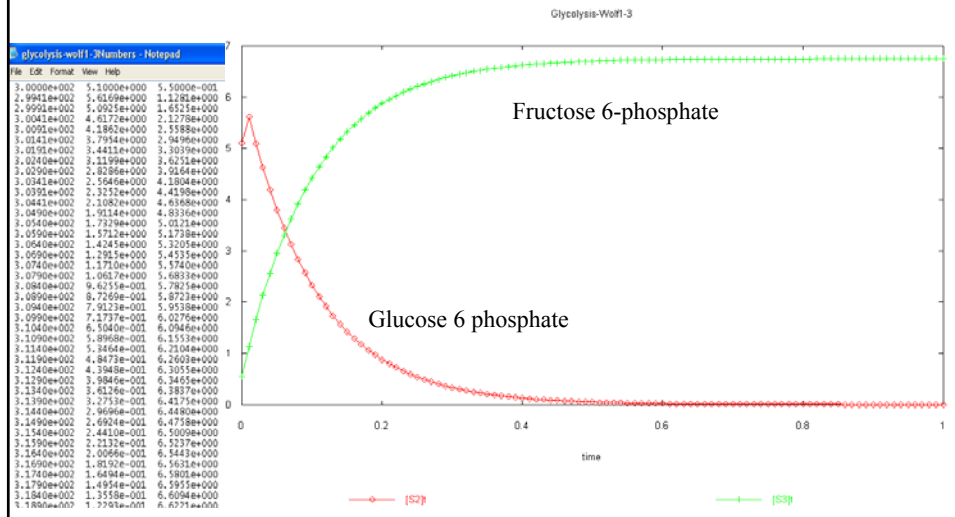
<http://www.gepasi.org/gepasi.html>

Online Glycolysis models



<http://jij.biochem.sun.ac.za/database/>

Results



Conclusions...

- Model based discoveries in glycolysis:
 - Oscillations in concentrations of some but not all metabolites.
 - Control of process distributed throughout pathway
 - Development of theoretical models
- Method integrates knowledge of pathway factors to examine pathway behaviors.

Modeling

- Requires formalizing assumptions
 - Rate equations
 - Inclusion or exclusion from model
- Worst case scenario
 - See what you believe
- Best case scenario
 - See something unexplainable
 - Create new laboratory experiments



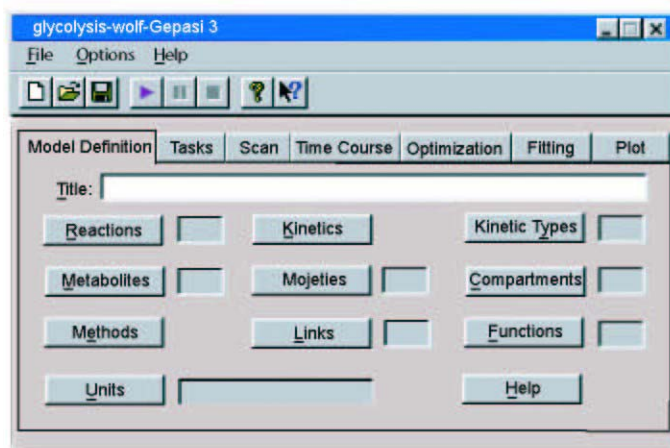
Hands On Section

See Handouts

Modeling Glycolysis

- What part of glycolysis do you want to model?
 - Draw out a concept map
 - Write out the biochemical reactions
 - $\text{Glucose} + \text{ATP} \rightarrow \text{Glucose 6-phosphate} + \text{ADP}$
 - Create a legend of symbols to molecule names
 - Try to use standard biochemical notation, but this is not required.

Gepasi Interface



Write your reactions in Gepasi

- Gepasi uses standard biochemical notation for entering Reactions.
 - Table of reactions provided are a reference for a simplified glycolytic model.
- Enter Kinetics for reactions
- Enter Metabolite concentrations
- Select Tasks to be performed

Reactions

- Influx of Glucose OR \rightarrow Glucose
- Glucose + ATP \rightarrow glucose-6-phosphate + ADP*
- Glucose 6-phosphate \leftrightarrow fructose 6-phosphate
- Fructose 6-Phosphate + ATP \rightarrow Fructose 1,6bisphosphate + ADP
- **Symbolic Notation**
 - \rightarrow S1
 - $S1 + A3 \rightarrow S2 + A2$
 - $S2 \leftrightarrow S3$
 - $S3 + A3 \rightarrow S4 + A2$

-> S1

$S1 + A3 \rightarrow S2 + A2$

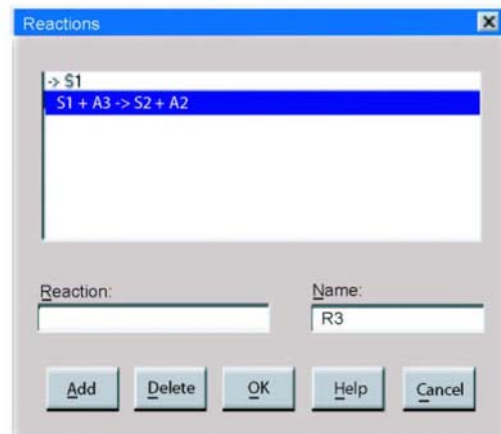
$S2 \leftrightarrow S3$

$S3 + A3 \rightarrow S4 + A2$

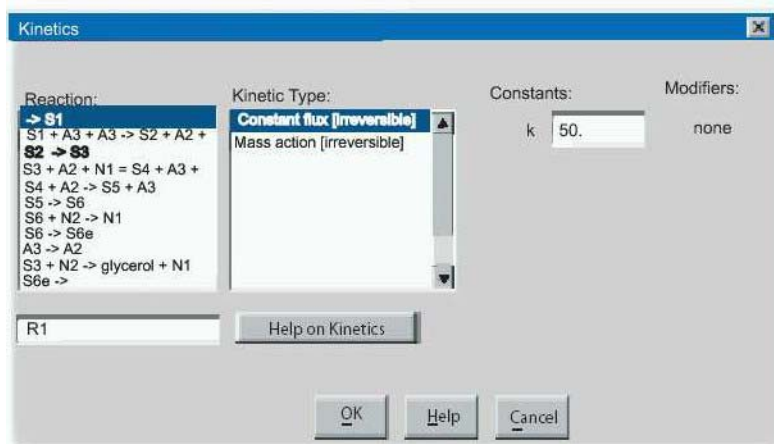
One modification to writing
Reversible reactions

$S2 \leftrightarrow S3$

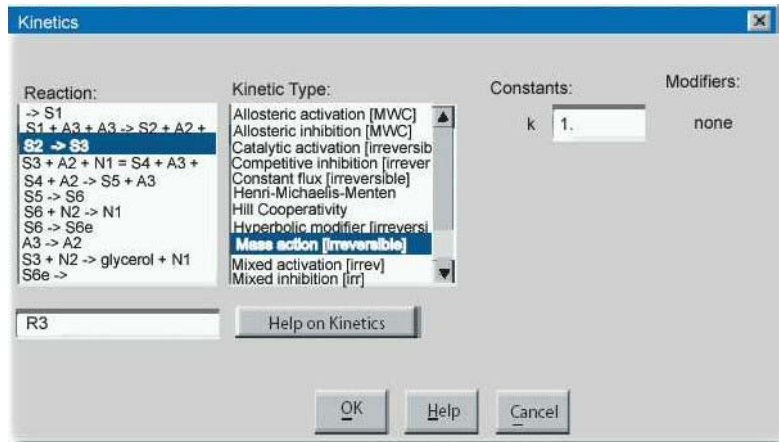
$S2 = S3$



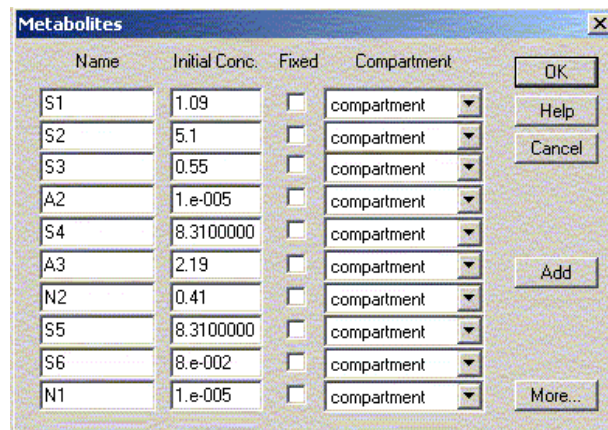
Selecting kinetics



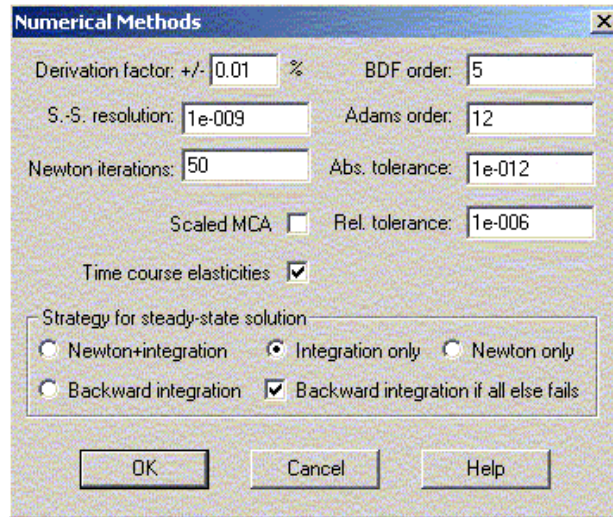
Selecting Kinetics contd



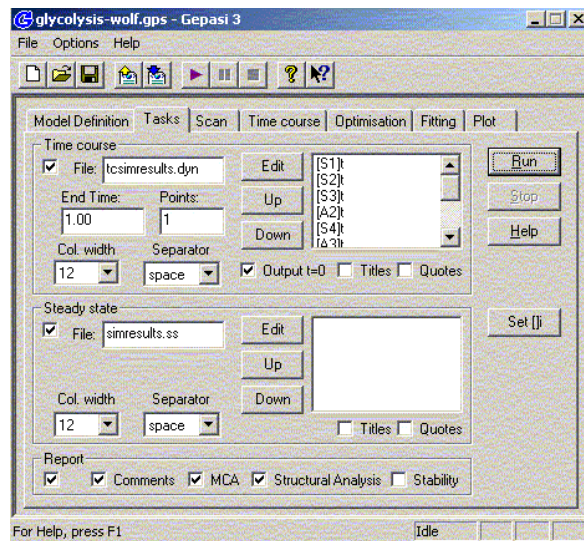
Metabolites



Methods



Tasks



Plot

