Reconstruction and Analysis of Metabolic Networks

Outline

- What is a Reconstruction?
- Data Collection
- Interactions Between Network Components
- Special Considerations
- Applications

Genome-scale Metabolic Model Reconstruction

Genome Annotation

- by homology, location

Biochemical Data

- protein characterized

Physiological Data

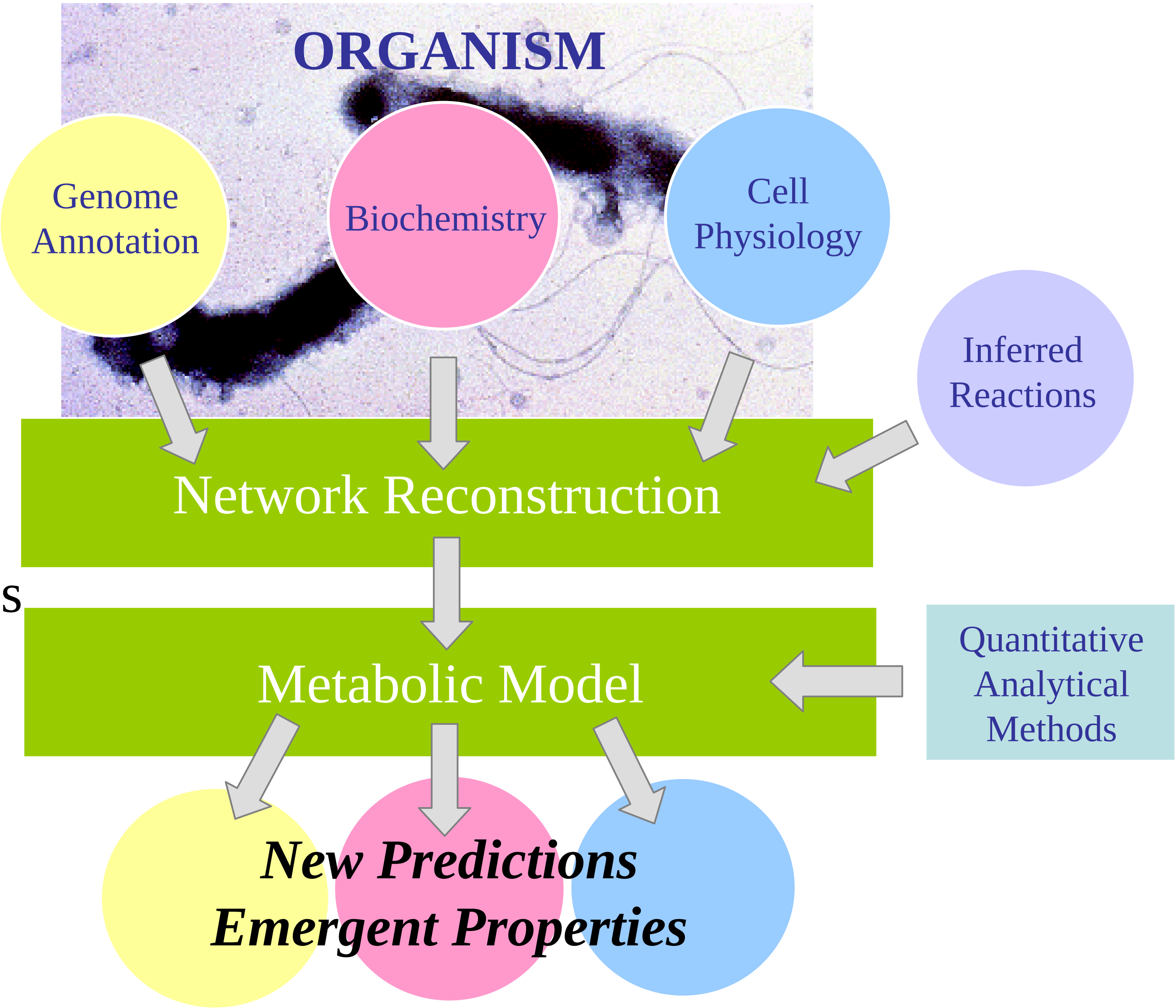
- indirect, pathway known

Inferred Reactions

- indirect, inferred from biomass requirements

Quantitative Analysis

- simulate cell behavior
- drive experimental studies

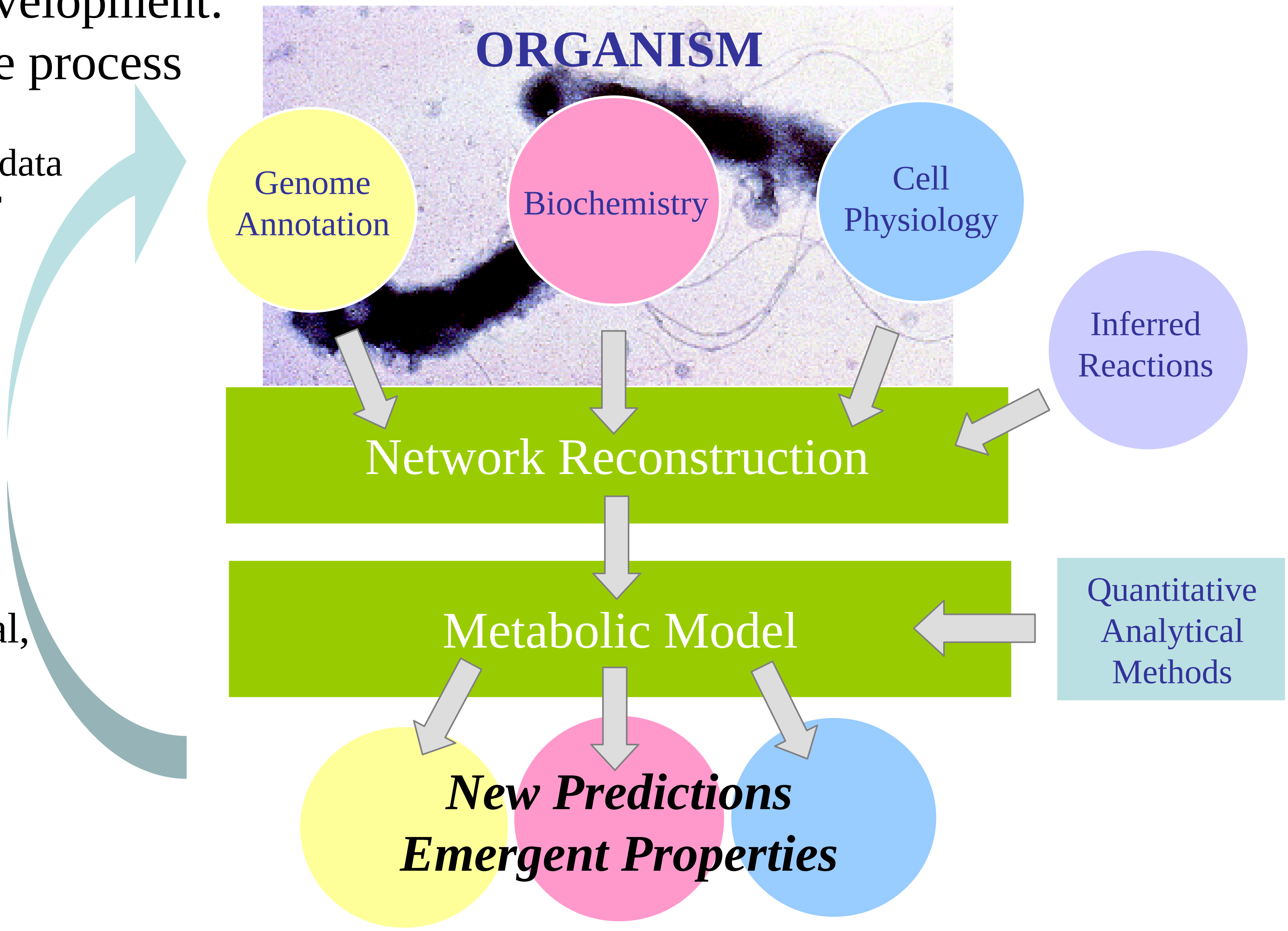


Genome-scale Metabolic Model Reconstruction

Model Development:
an iterative process

- Biochemical data
- Revised ORF assignments

Computational,
Biochemical
Investigation



ORGANISM

Genome Annotation

Biochemistry

Cell Physiology

Inferred Reactions

Network Reconstruction

Metabolic Model

Quantitative Analytical Methods

New Predictions
Emergent Properties

What is in a reconstruction?

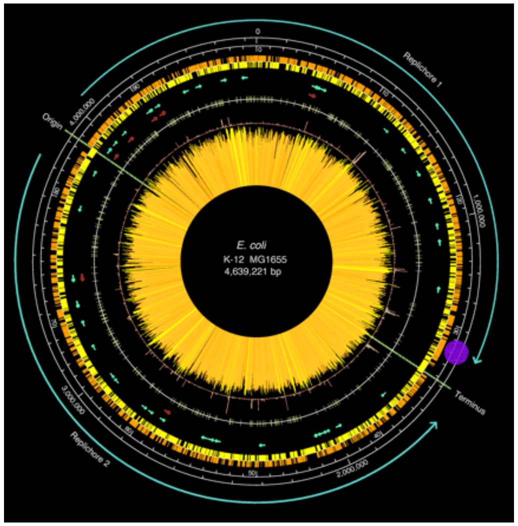
Genome:
- Annotated genes
- Gene location
- Regulatory regions
- Wobble base pairs

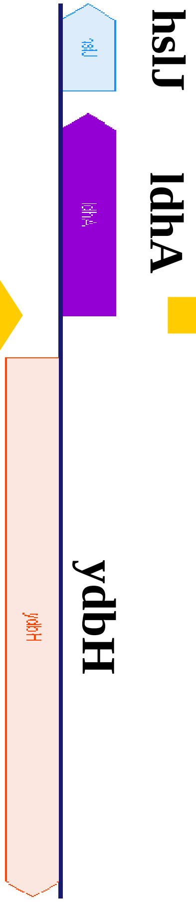
Biochemistry:
- Stereochemistry
- pH and pKa (charge)
- Elemental balance
- Charge balance
- Multiple reactions/enzyme
- Multiple enzymes/reaction

Transcription/translation:
- Gene to transcript to protein to reaction association
- Transcript half-lives
- tRNA abundances
- Ribosomal capacities

Physiology:
- Flux data
- Knock-outs
- Balanced functions
- Overall phenotypic behavior
- Location of gene product compartmentalization

Defining Metabolic Reactions





Lactate Dehydrogenase

1st level: Metabolite Specificity

Primary metabolites: **LAC**, **PYR**

Coenzymes: **NAD**, **NADH**

2nd level: Metabolite Formulas

Neutral Formulas

$C_3H_6O_3$ $C_3H_4O_3$ $C_{21}H_{26}N_7O_{14}P_2$ $C_{21}H_{27}N_7O_{14}P_2$

Charged Formulas

$C_3H_5O_3^{1-}$ $C_3H_3O_3^{1-}$ $C_{21}H_{26}N_7O_{14}P_2^{1-}$ $C_{21}H_{27}N_7O_{14}P_2^{1-}$

3rd level: Stoichiometry

1 LAC + 1 NAD ? 1 PYR + 1 NADH + 1 H

4th level: Thermodynamic Considerations: Directionality

1 LAC + 1 NAD ↔ 1 PYR + 1 NADH + 1 H

5th level: Localization

prokaryotes

[c]: cytoplasm
[e]: extracellular
[p]: periplasm

[n]: nucleus
[g]: golgi aparatus
[v]: vacuole
[l]: lysosome

[m]: mitochondria
[x]: peroxisome
[h]: chloroplast
[r]: endoplasmic reticulum

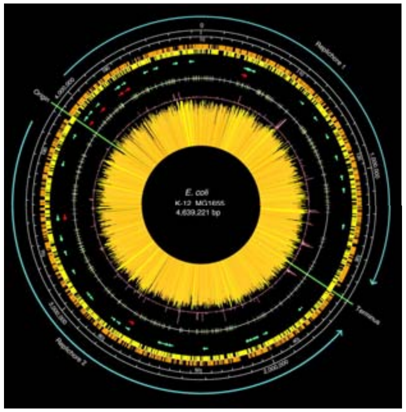
eukaryotes

1 LAC [c] + 1 NAD [c] ↔ 1 PYR [c] + 1 NADH [c] + 1 H [c]

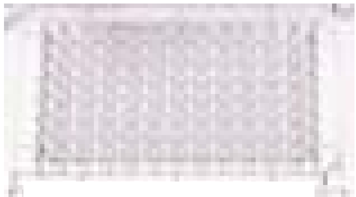
STEPWISE INCORPORATION OF INFORMATION

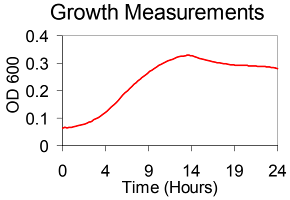
Sources of Information

Genome Sequence & Annotation



Physiological Data





Databases

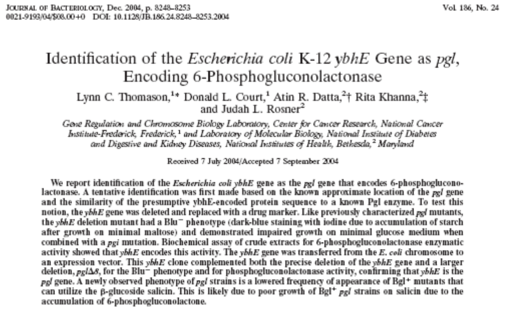
KEGG Kyoto Encyclopedia of Genes and Genomes

EcoCyc

Available Literature

PubMed

National Library of Medicine NLM



JOURNAL OF BACTERIOLOGY, Dec. 2004, p. 8248–8253 Vol. 186, No. 24
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.24.8248–8253.2004

Identification of the *Escherichia coli* K-12 *ybhE* Gene as *pgl*, Encoding 6-Phosphogluconolactonase

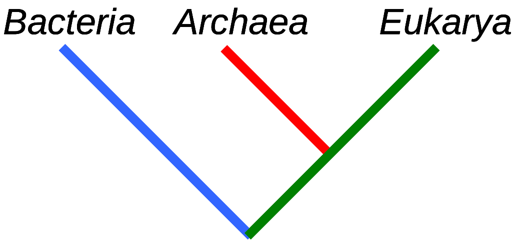
Lynn C. Thomason,[1*] Donald L. Court,[1] Atin R. Datta,[2†] Rita Khanna,[2‡] and Judah L. Rosner[2]

Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, National Cancer Institute-Frederick, Frederick,[1] and Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda,[2] Maryland

Received 7 July 2004/Accepted 7 September 2004

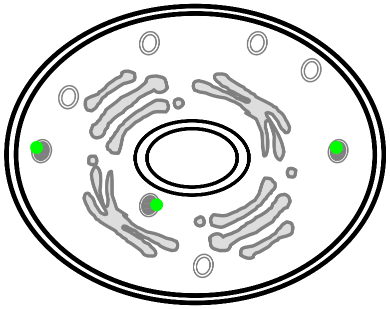
We report identification of the *Escherichia coli ybhE* gene as the *pgl* gene that encodes 6-phosphogluconolactonase. A tentative identification was first made based on the known approximate location of the *pgl* gene and the similarity of the presumptive ybhE-encoded protein sequence to a known Pgl enzyme. To test this notion, the *ybhE* gene was deleted and replaced with a drug marker. Like previously characterized *pgl* mutants, the *ybhE* deletion mutant had a Blu⁻ phenotype (dark-blue staining with iodine due to accumulation of starch after growth on minimal maltose) and demonstrated impaired growth on minimal glucose medium when combined with a *pgi* mutation. Biochemical assay of crude extracts for 6-phosphogluconolactonase enzymatic activity showed that *ybhE* encodes this activity. The *ybhE* gene was transferred from the *E. coli* chromosome to an expression vector. This *ybhE* clone complemented both the precise deletion of the *ybhE* gene and a larger deletion, *pgl*Δ8, for the Blu⁻ phenotype and for phosphogluconolactonase activity, confirming that *ybhE* is the *pgl* gene. A newly observed phenotype of *pgl* strains is a lowered frequency of appearance of Bgl⁺ mutants that can utilize the β-glucoside salicin. This is likely due to poor growth of Bgl⁺ *pgl* strains on salicin due to the accumulation of 6-phosphogluconolactone.

Phylogenetic Data



Localization

Signal sequences
...PLLLLPISGSALP...

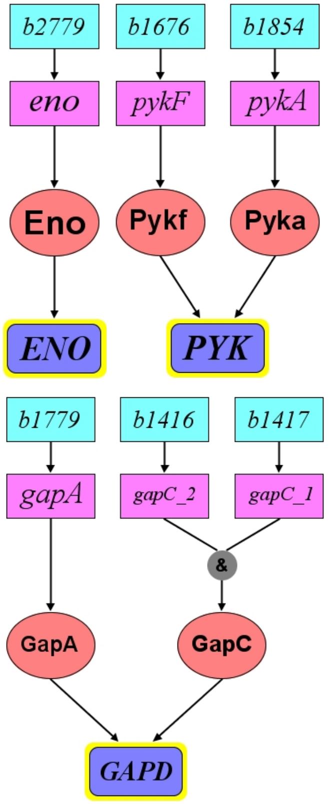


Network Assembly and Representation

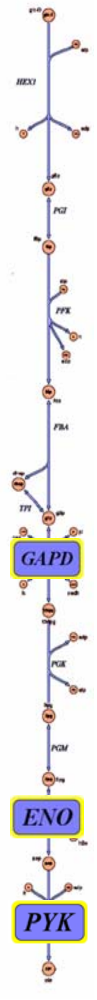
Reconstruction of Glycolytic Pathway

Abbr.	Glycolytic Reactions	Genes
HEX1	[c]glc +atp → g6p + adp	glk
PGI	[c]g6p ↔ f6p	pgi
PFK	[c]atp + f6p → adp + fdp + h	pfkA,pfkB
FBA	[c]fdp ↔ dhap + g3p	fbaA,fbaB
TPI	[c]dhap ↔ g3p	tpiA
GAPD	[c]g3p + nad + pi ↔ 13dpg + h + nadh	gapA,gapC_1,gapC_2
PGK	[c]13dpg + adp ↔ 3pg + atp	pgk
PGM	[c]3pg ↔ 2pg	gpmA,gpmB
ENO	[c]2pg ↔ h2o + pep	eno
PYK	[c]adp + h + pep → atp + pyr	pykA,pykF

	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
atp	-1	0	-1	0	0	0	1	0	0	1
glc	-1	0	0	0	0	0	0	0	0	0
adp	1	0	1	0	0	0	-1	0	0	-1
g6p	1	-1	0	0	0	0	0	0	0	0
h	1	0	1	0	0	1	0	0	0	-1
f6p	0	1	-1	0	0	0	0	0	0	0
fdp	0	0	1	-1	0	0	0	0	0	0
dhap	0	0	0	1	-1	0	0	0	0	0
g3p	0	0	0	1	1	-1	0	0	0	0
nad	0	0	0	0	0	-1	0	0	0	0
pi	0	0	0	0	0	-1	0	0	0	0
13dpg	0	0	0	0	0	1	-1	0	0	0
nadh	0	0	0	0	0	1	0	0	0	0
3pg	0	0	0	0	0	0	1	-1	0	0
2pg	0	0	0	0	0	0	0	1	-1	0
pep	0	0	0	0	0	0	0	0	1	-1
h2o	0	0	0	0	0	0	0	0	1	0
pyr	0	0	0	0	0	0	0	0	0	1

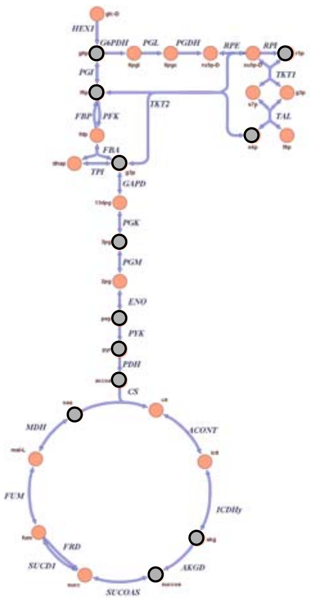


PYK: IF *pykA* OR *pykF*
ENO: IF *eno*
GAPD: IF *gapA* OR (*gapC_1* AND *gapC_2*)

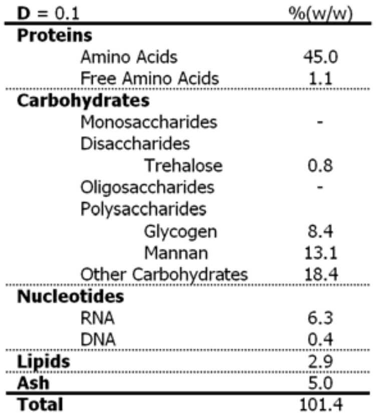


Network Evaluation

- **Precursor Metabolite Formation**

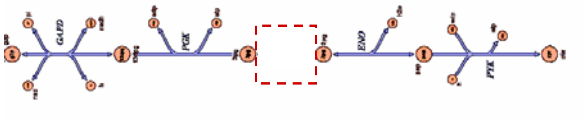


- **Incorporating Biomass Composition**

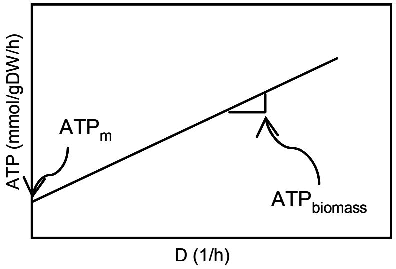


D = 0.1			%(w/w)
Proteins			
	Amino Acids		45.0
	Free Amino Acids		1.1
Carbohydrates			
	Monosaccharides		-
	Disaccharides		
		Trehalose	0.8
	Oligosaccharides		-
	Polysaccharides		
		Glycogen	8.4
		Mannan	13.1
	Other Carbohydrates		18.4
Nucleotides			
	RNA		6.3
	DNA		0.4
Lipids			2.9
Ash			5.0
Total			101.4

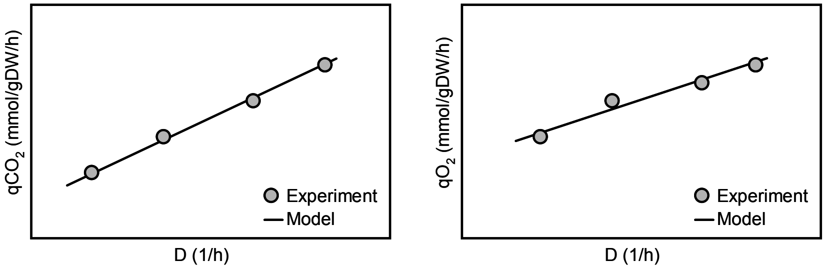
- **Filling Network Gaps**



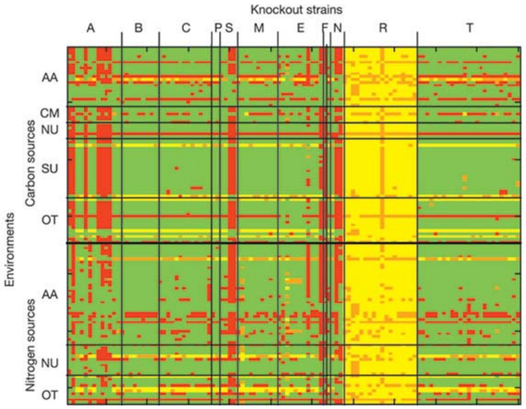
- **ATP Maintenance Calculation**



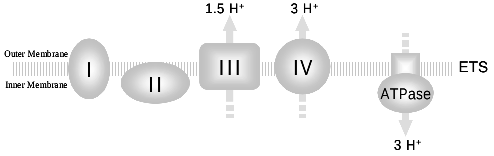
- **Physiological Data Comparison**



- **Knockout Data Comparison**



- **P/O Ratio Calculation**



Data Collection

I. Genome Annotation
II. Biochemistry
III. Physiology

I. Genome Annotation

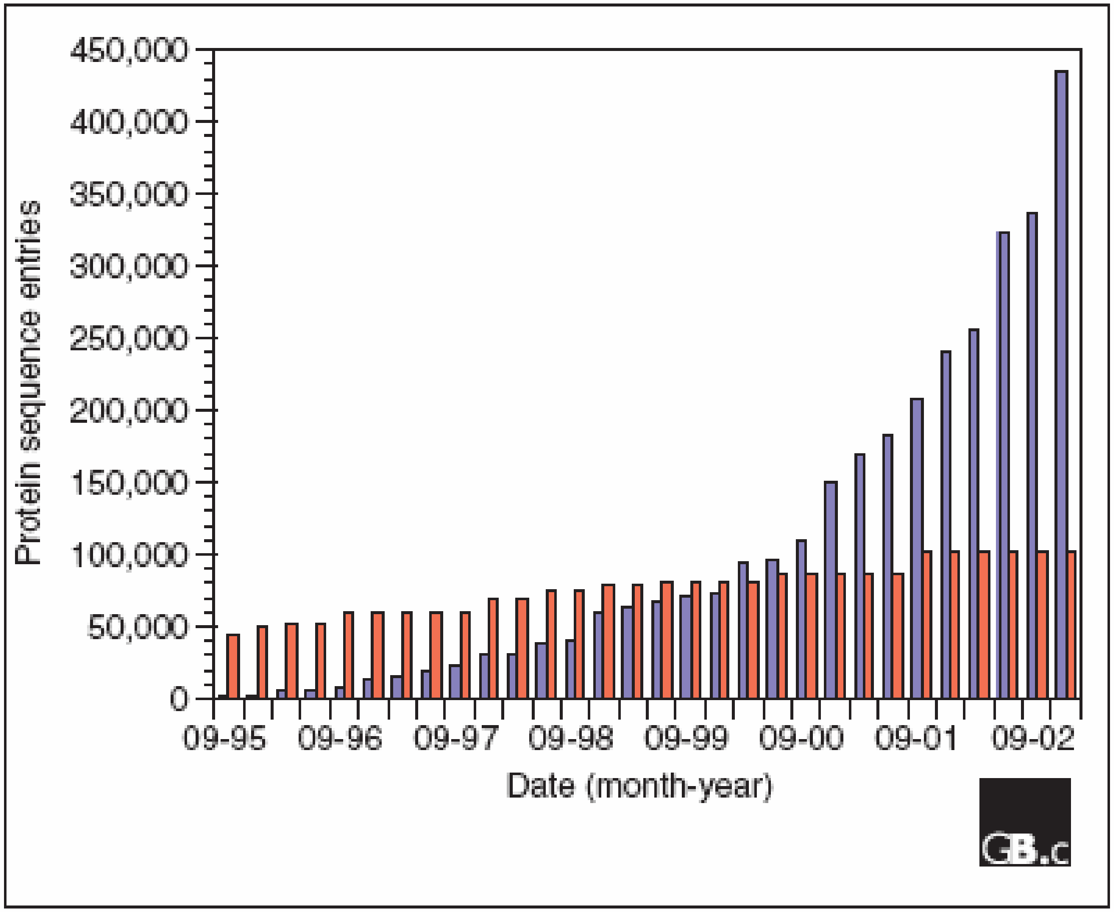


Figure 1
Cumulative number of protein sequence entries (*y*-axis) in completed genomes (CoGenT, in blue) and Swiss-Prot (in red) as a function of time (*x*-axis).

- 433,238 protein sequences derived from whole genomes (expected to reach 1 million by 2005)
- 101,602 entries in Swiss Prot
- High-quality annotation requires substantial effort

Genome Annotation: how to

- Open Reading Frame (ORF) Identification
  - Start & Stop codons, GLIMMER.
- “Traditional” Annotation Methods
  - Experimental (direct)
  - Sequence homology
  - Generally covers 40-70% of new genomes
- New Annotation Methods
  - Protein-protein interactions
  - Correlated mRNA expression levels
  - Phylogenetic profile clustering
  - Protein fusion
  - Gene neighbors (operon clustering)

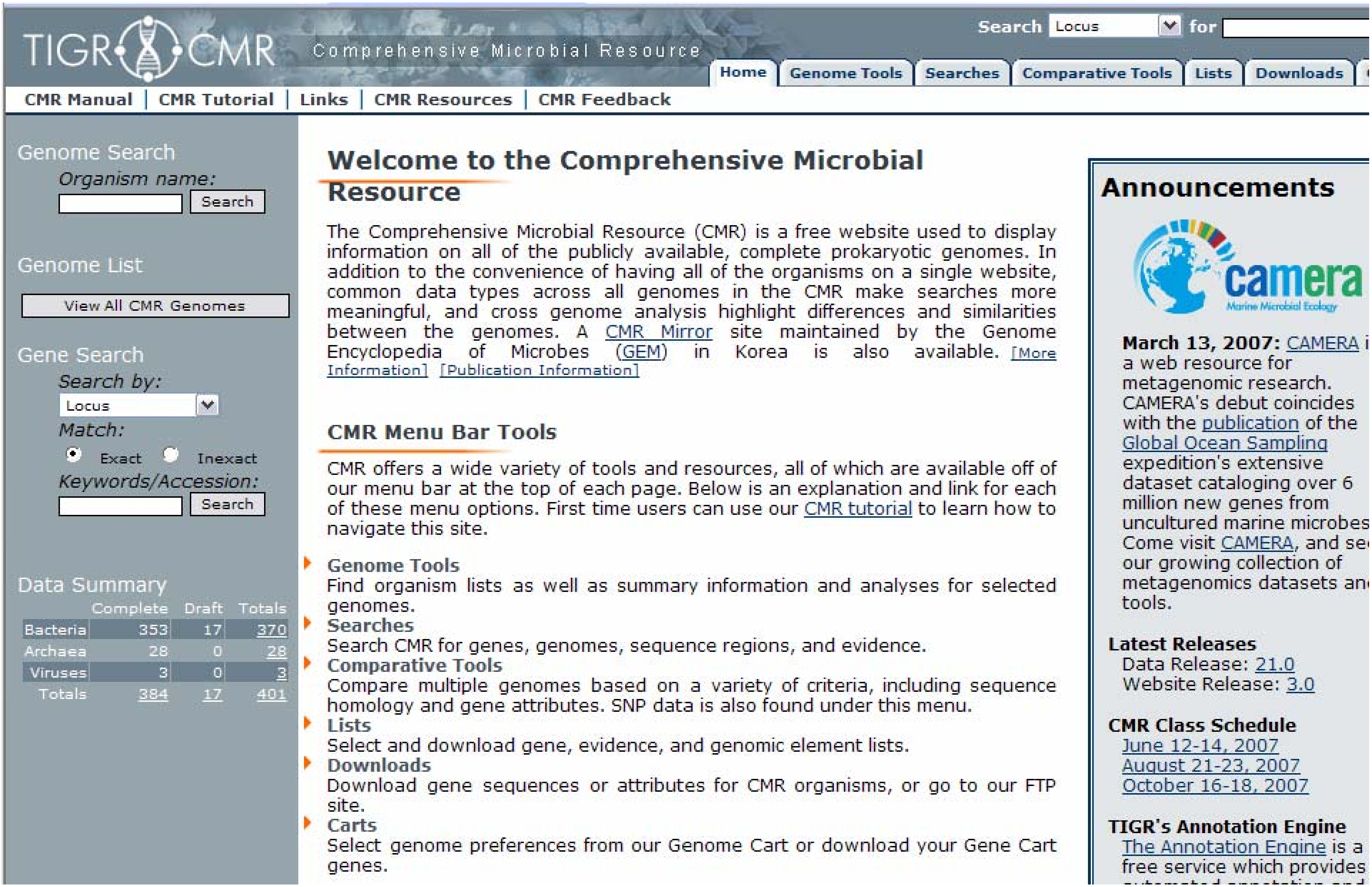
Genome Databases: TIGR

http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi

The Comprehensive Microbial Resource (CMR)

- 353 completed bacterial genomes
- 28 completed archaeal genomes
- Single-genome analysis:
  - Genome overview,
  - list by role category (eg amino acid biosynthesis)
  - analysis methods, searches
- Multi-genome analysis also available

Genome Databases: TIGR



Data Summary

	Complete	Draft	Totals
Bacteria	353	17	370
Archaea	28	0	28
Viruses	3	0	3
Totals	384	17	401

Welcome to the Comprehensive Microbial Resource

The Comprehensive Microbial Resource (CMR) is a free website used to display information on all of the publicly available, complete prokaryotic genomes. In addition to the convenience of having all of the organisms on a single website, common data types across all genomes in the CMR make searches more meaningful, and cross genome analysis highlight differences and similarities between the genomes. A CMR Mirror site maintained by the Genome Encyclopedia of Microbes (GEM) in Korea is also available. [More Information] [Publication Information]

CMR Menu Bar Tools

CMR offers a wide variety of tools and resources, all of which are available off of our menu bar at the top of each page. Below is an explanation and link for each of these menu options. First time users can use our CMR tutorial to learn how to navigate this site.

- **Genome Tools**
  Find organism lists as well as summary information and analyses for selected genomes.
- **Searches**
  Search CMR for genes, genomes, sequence regions, and evidence.
- **Comparative Tools**
  Compare multiple genomes based on a variety of criteria, including sequence homology and gene attributes. SNP data is also found under this menu.
- **Lists**
  Select and download gene, evidence, and genomic element lists.
- **Downloads**
  Download gene sequences or attributes for CMR organisms, or go to our FTP site.
- **Carts**
  Select genome preferences from our Genome Cart or download your Gene Cart genes.

Announcements

March 13, 2007: CAMERA i a web resource for metagenomic research. CAMERA's debut coincides with the publication of the Global Ocean Sampling expedition's extensive dataset cataloging over 6 million new genes from uncultured marine microbes Come visit CAMERA, and se our growing collection of metagenomics datasets an tools.

Latest Releases
Data Release: 21.0
Website Release: 3.0

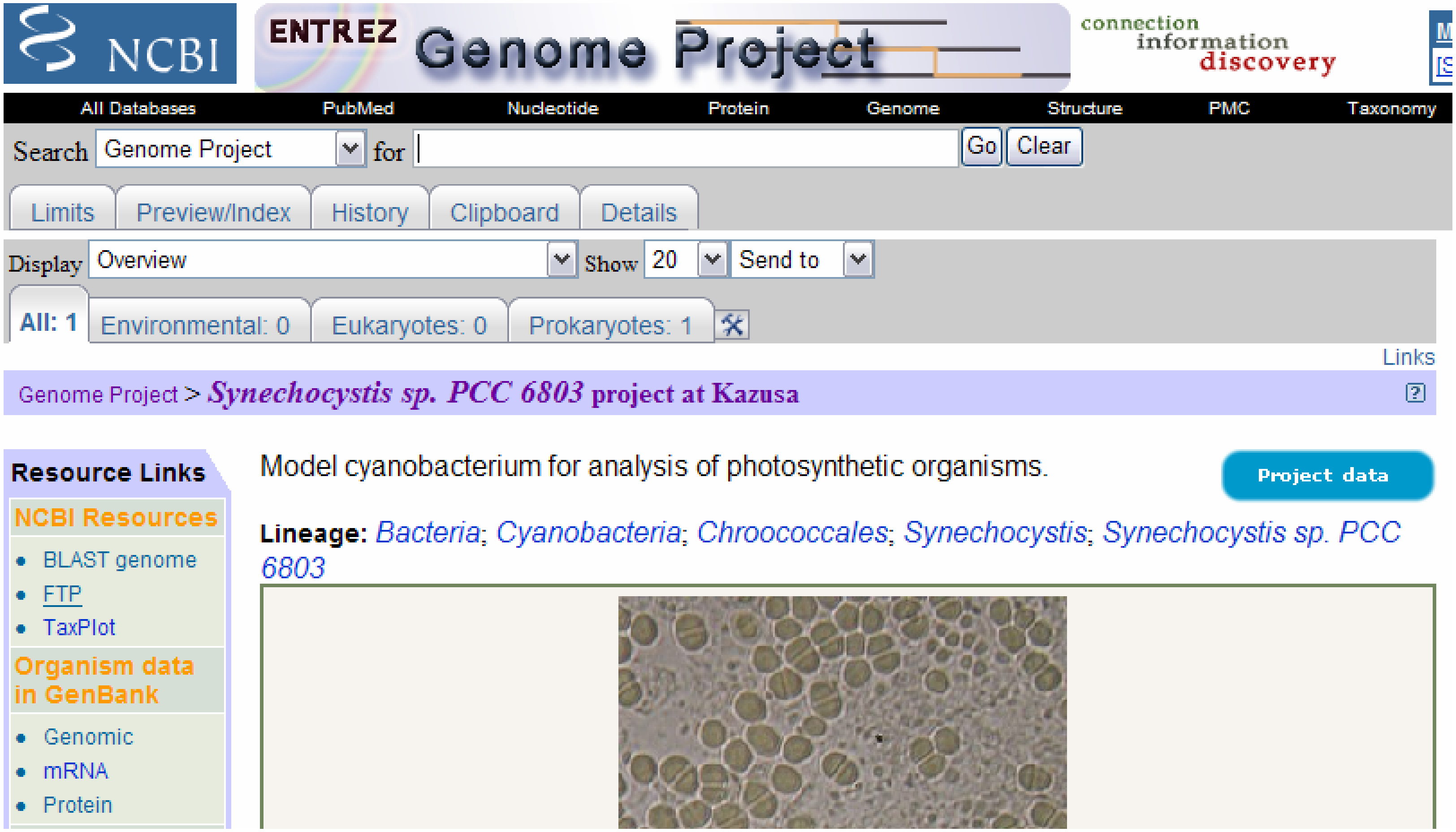
CMR Class Schedule
June 12-14, 2007
August 21-23, 2007
October 16-18, 2007

TIGR's Annotation Engine
The Annotation Engine is a free service which provides

Genome Databases: NCBI

Microbial Genomes Resources

- 595 completed microbial genomes (47 archael)
- FTP sites for Protein Annotations (ptt files)



Search Genome Project for

Display Overview Show 20

Genome Project > *Synechocystis sp. PCC 6803* **project at Kazusa**

Model cyanobacterium for analysis of photosynthetic organisms.

Lineage: *Bacteria; Cyanobacteria; Chroococcales; Synechocystis; Synechocystis sp. PCC 6803*

II. Biochemical Data: Reactions

stoichiometry and reversibility

Gene: glk

Enzyme: Glucokinase

Reaction:

ATP + D-Glucose –>

ADP + D-Glucose 6-phosphate

E.C.: 2.7.1.1

Human ß-cell *glk* http://www.rcsb.org/pdb/

Trust the E.C. Nomenclature!

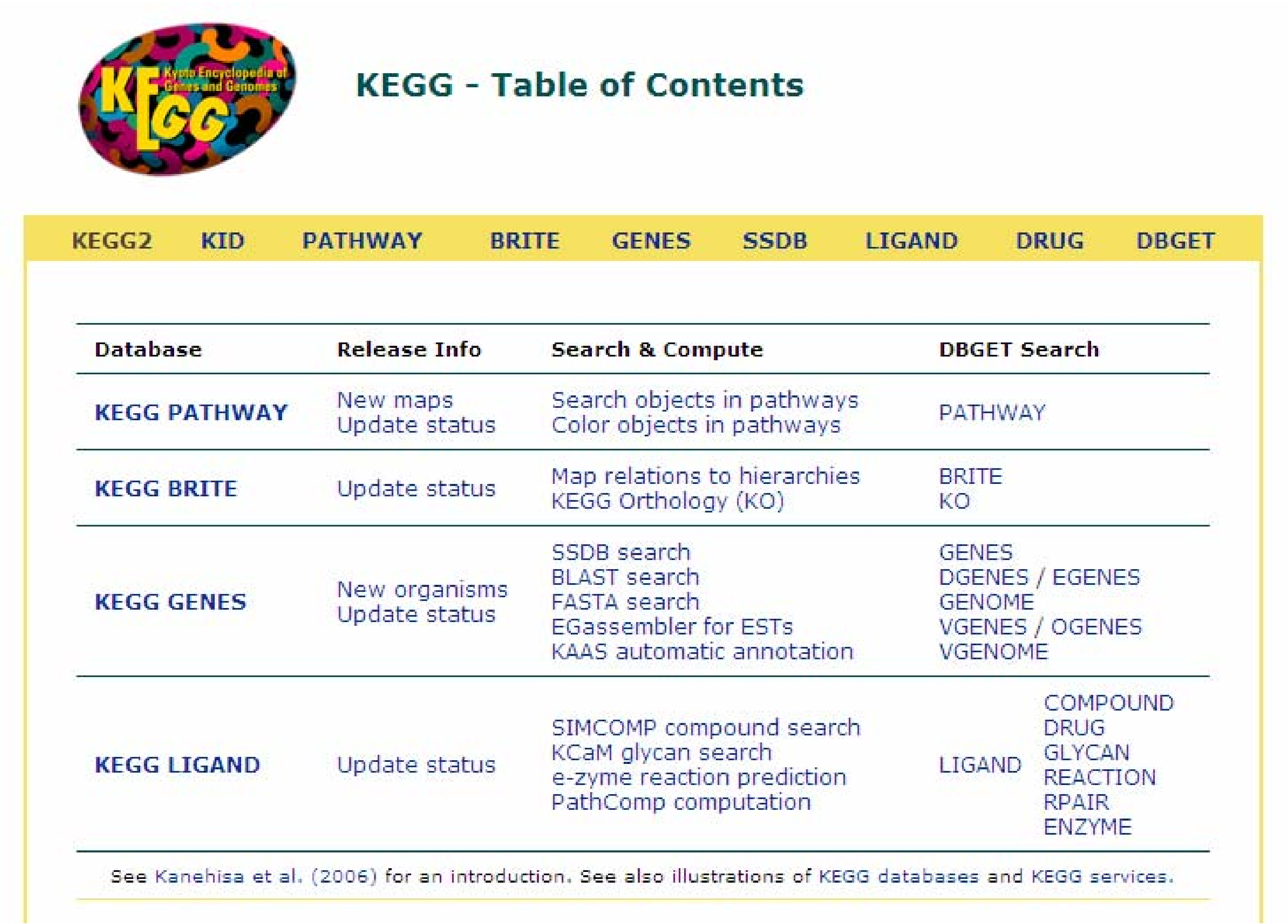
EC 1 Oxidoreductases

EC 1.1	**Acting on the CH-OH group of donors**
EC 1.1.1	With NAD or NADP as acceptor
EC 1.1.2	With a cytochrome as acceptor
EC 1.1.3	With oxygen as acceptor
EC 1.1.4	With a disulfide as acceptor
EC 1.1.5	With a quinone or similar compound as acceptor
EC 1.1.99	With other acceptors
EC 1.2	**Acting on the aldehyde or oxo group of donors**
EC 1.2.1	With NAD or NADP as acceptor
EC 1.2.2	With a cytochrome as acceptor
EC 1.2.3	With oxygen as acceptor
EC 1.2.4	With a disulfide as acceptor
EC 1.2.7	With an iron-sulfur protein acceptor
EC 1.2.99	With other acceptors
EC 1.3	**Acting on the CH-CH group of donors**
EC 1.3.1	With NAD or NADP as acceptor
EC 1.3.2	With a cytochrome as acceptor
EC 1.3.3	With oxygen as acceptor
EC 1.3.5	With a quinone or related compound as acceptor
EC 1.3.7	With an iron-sulfur protein as acceptor

- Not widely available for other types of gene products (T.C. numbers are being developed)
- Kudos to enzymologists
- Make sure to balance elements when writing reaction

KEGG

http://www.genome.jp/kegg/kegg2.html



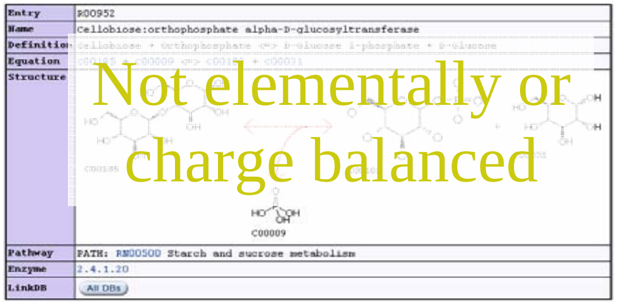
KEGG - Table of Contents

KEGG2 KID PATHWAY BRITE GENES SSDB LIGAND DRUG DBGET

Database	Release Info	Search & Compute	DBGET Search
KEGG PATHWAY	New maps Update status	Search objects in pathways Color objects in pathways	PATHWAY
KEGG BRITE	Update status	Map relations to hierarchies KEGG Orthology (KO)	BRITE KO
KEGG GENES	New organisms Update status	SSDB search BLAST search FASTA search EGassembler for ESTs KAAS automatic annotation	GENES DGENES / EGENES GENOME VGENES / OGENES VGENOME
KEGG LIGAND	Update status	SIMCOMP compound search KCaM glycan search e-zyme reaction prediction PathComp computation	LIGAND COMPOUND DRUG GLYCAN REACTION RPAIR ENZYME

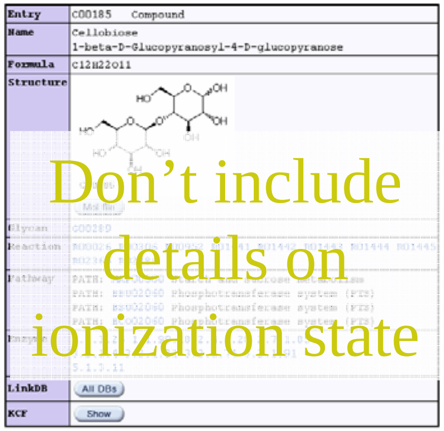
See Kanehisa et al. (2006) for an introduction. See also illustrations of KEGG databases and KEGG services.

How are these reconstructions different than KEGG?



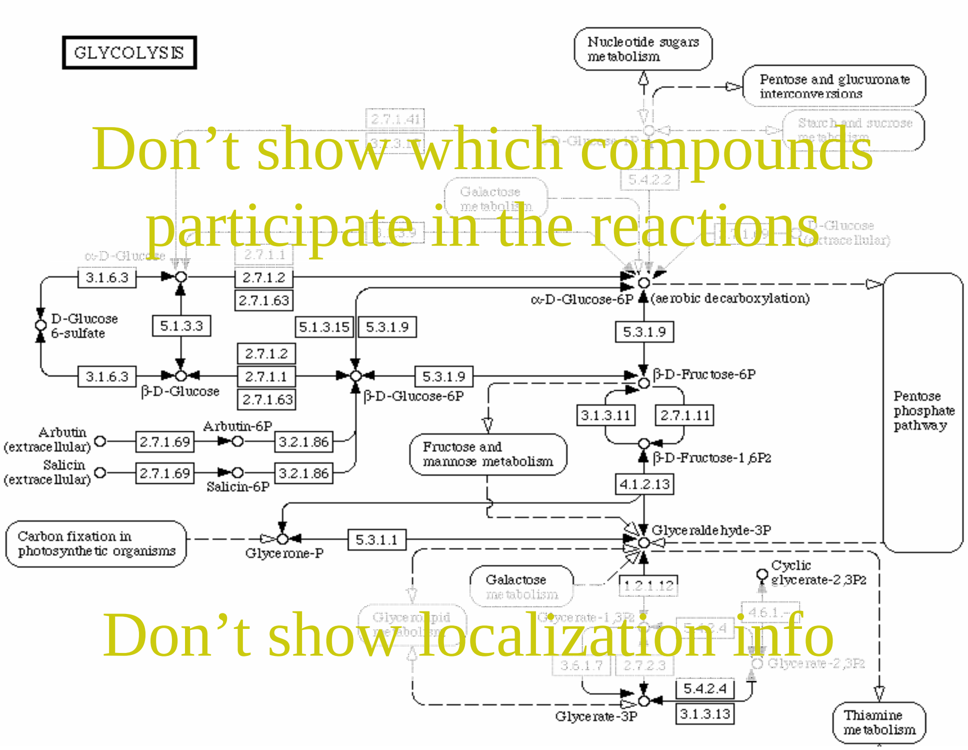
Entry	R00952
Name	Cellobiose:orthophosphate alpha-D-glucosyltransferase
Definition	Cellobiose + Orthophosphate <=> D-Glucose 1-phosphate + D-Glucose
Equation	C00185 + C00009 <=> [illegible] + C00031
Structure	C00185 C00009
Pathway	PATH: RN00500 Starch and sucrose metabolism
Enzyme	2.4.1.20
LinkDB	All DBs

Not elementally or charge balanced



Entry	C00185 Compound
Name	Cellobiose 1-beta-D-Glucopyranosyl-4-D-glucopyranose
Formula	C12H22O11
Structure	
Glycan	G00280
Reaction	
Pathway	
Enzyme	
LinkDB	All DBs
KCF	Show

Don't include details on ionization state



Don't show which compounds participate in the reactions

Don't show localization info

http://www.genome.jp/ligand/

Charge Determination on Metabolite at neutral pH

Identify compound & look up in KEGG
http://www.genome.ad.jp/dbget-bin/www_bfind?compound

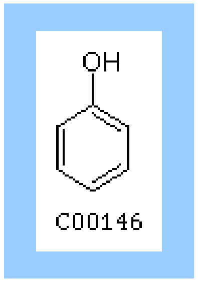
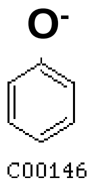
Determine and identify ionizable group

Determine acid and base forms

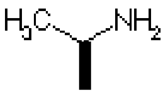
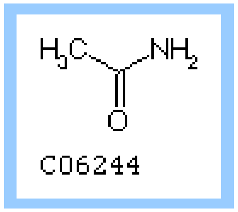
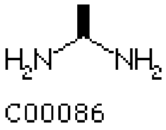
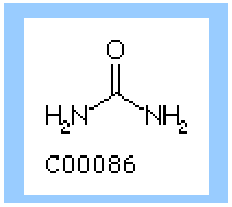
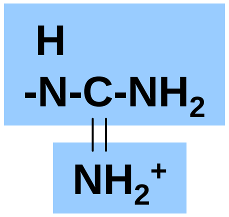
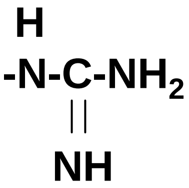
Determine pKa values based on the identifiable group (in the Table)

- If pKa > pH, acid form dominant
- If pKa < pH, base form dominant

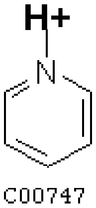
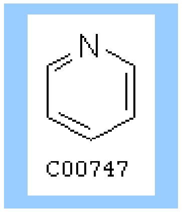
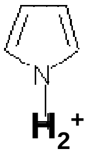
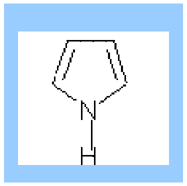
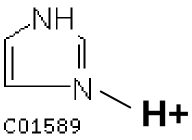
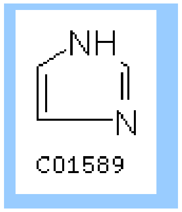
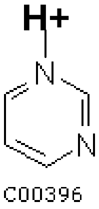
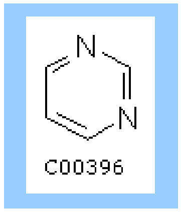
Ionizable Groups 1

GROUP	***ACID***	***BASE + H^+***	***pKa***
Terminal Carboxyl	-COOH	$–COO^-$ + H^+	~4
Primary (Secondary, Tertiary) Amine	$-NH_3^+$	$–NH_2$ + H^+	> 9
Thiol	-SH	$–S^-$ + H^+	~8.5
Phenol		 + H^+	~10

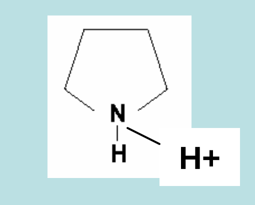
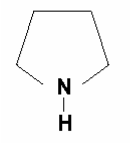
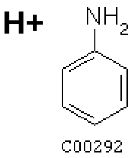
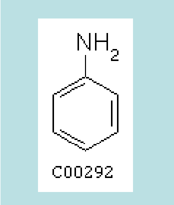
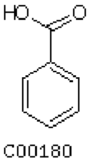
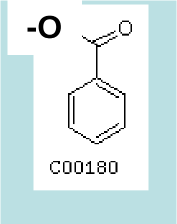
Ionizable Groups 2

GROUP	***ACID***	***BASE + H⁺***	***pKa***
Primary Alcohol	$-CH_2OH$	$–CH_2O^- + H^+$	~15
Acetamide (Amide)	 **OH^+**	 + H^+	~0
Urea (Carbamide)	**OH^+** 	 + H^+	~1
Guanido Group	 $-\overset{H}{N}-\underset{\underset{NH_2^+}{\|}}{C}-NH_2$	 $-\overset{H}{N}-\underset{\underset{NH}{\|}}{C}-NH_2$ + H^+	~12

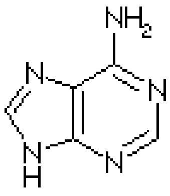
Ionizable Groups 3

GROUP	***ACID***	***BASE + H^+***	***pKa***
Pyridine		 + H^+	~5
Pyrole		 + H^+	~ -1
Imidazole		 + H^+	~7
Pyrimidine		 + H^+	~0

Ionizable Groups 4

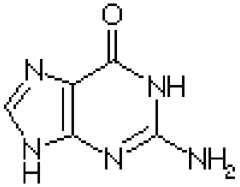
GROUP	***ACID***	***BASE + H$^+$***	***pKa***
Pyrrolidine (Like Sec Amine)		 + H$^+$	~10
Aniline		 + H$^+$	~5
Benzoic Acid		 + H$^+$	~4

Ionizable Groups 5: Purines & Pyrimidines



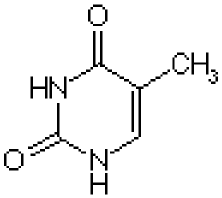
C00147

Adenine



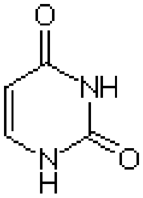
C00242

Guanine



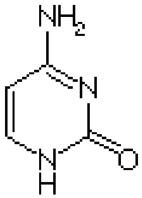
C00178

Thymine



C00106

Uracil

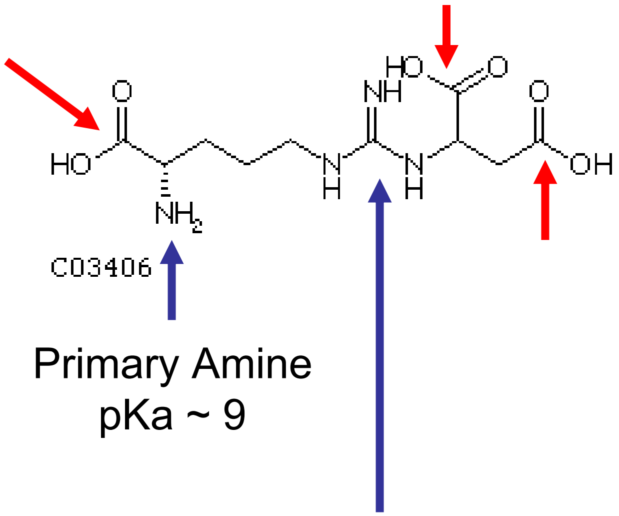


C00380

Cytosine

Net Charge = 0 !

Example : Arginosuccinate



Primary Amine
pKa ~ 9

Guanido
pKa ~ 12

Example: Argininosuccinate

Neutral MF: C10H18N4O
pKa: 1.62, 2.70, 4.26, 9.58, >12

Net Charge: -1
Charged MF: C10H17N4O6

Biochemical Data: Curation and Expansion of the Network

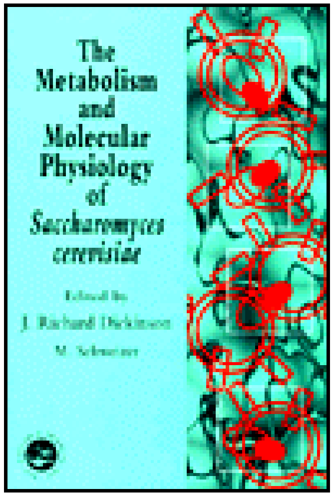
H. pylori Glycolysis according to KEGG:

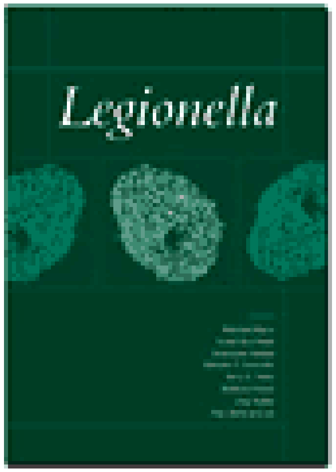
Glucose → **G-6-P** → **F-6-P** 🚫

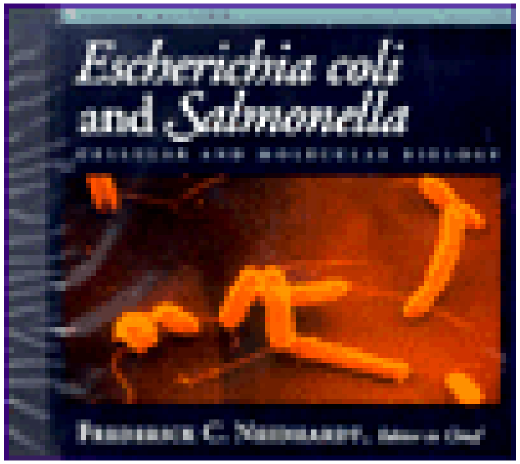
H. pylori Glycolysis according to Hoffman *et al.* (1996):

Glucose → **G-6-P** → **F-6-P** →

Organism-specific Textbooks



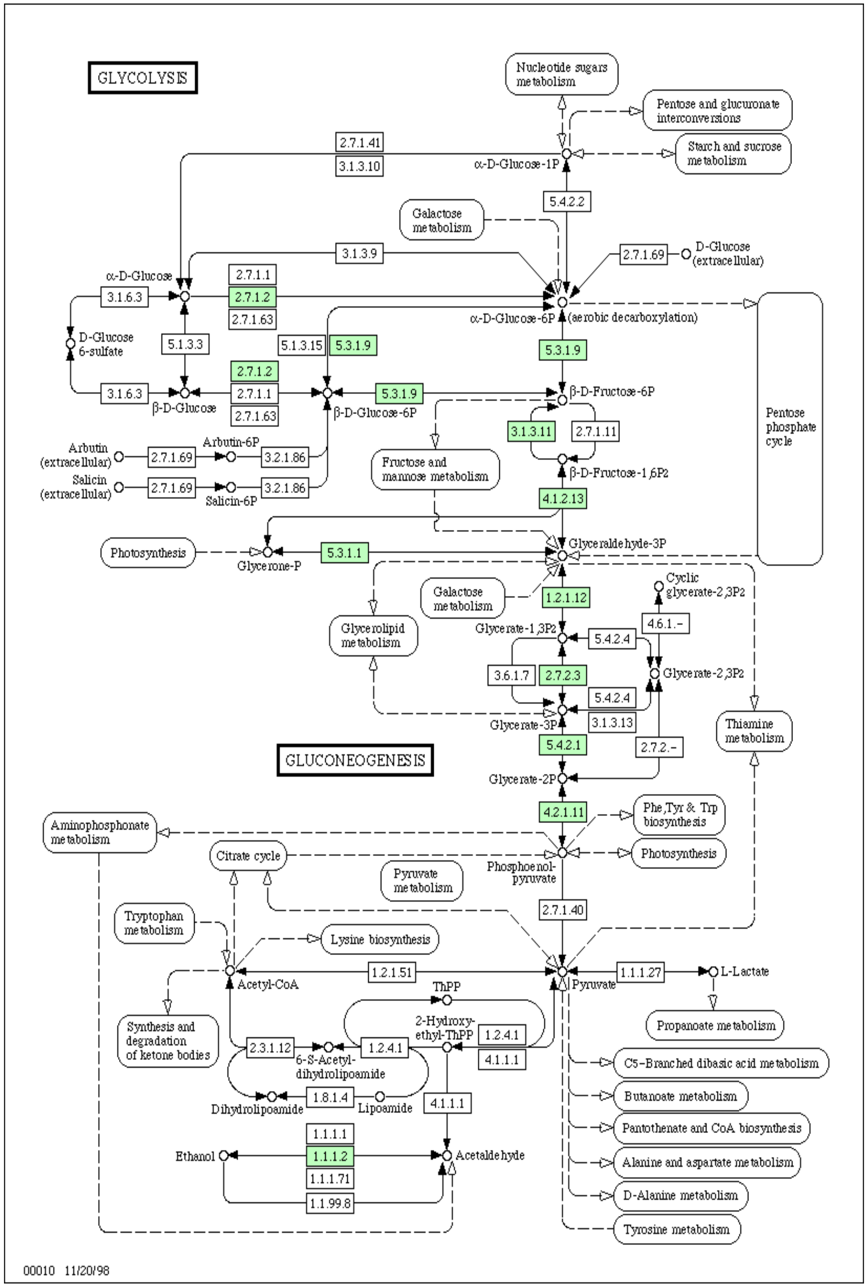




- Great starting point
- Broad view of the organism’s metabolism, biochemistry, physiology, uses, etc.

III. Physiological Information and Inferred Reactions:

Filling in the Gaps based on indirect evidence



Filling in the Gaps – an Example

- Experiments determine which amino acids are taken up by *H. pylori* vs. which can be produced *in vivo*
- Missing steps of amino acid biosynthesis are added if necessary on the basis of this physiological evidence

Amino Acid Requirements

AA	Reynolds	Model
Ala	-	-
Arg	-	-
Asn	+	+
Asp	+	+
Cys	+	+
Gln	+	+
Glu	+	+
Gly	+	+
His	-	-
Ile	-	-
Leu	-	-
Lys	+	+
Met	-	-
Phe	-	-
Pro	+	+
Ser	+	+
Thr	+	+
Trp	+	+
Tyr	+	+
Val	-	-

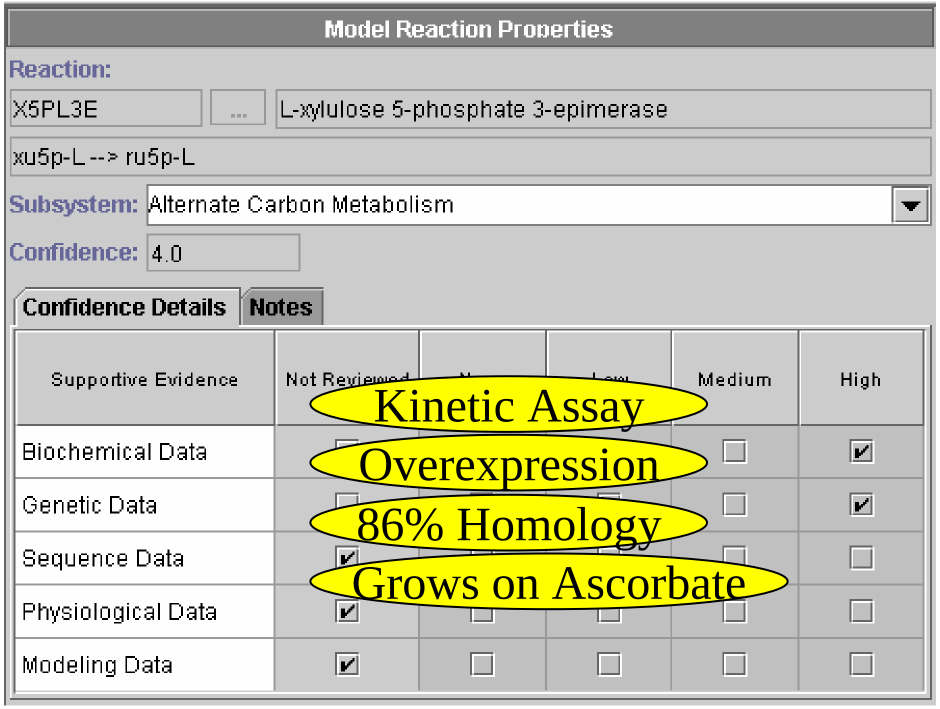
in vivo in silico

Inferred Reactions

- Some reactions are included based on indirect physiological evidence (by inference)
  - Assumption: the cell must be able to produce all biomass components to grow
  - Reactions are added if necessary
  - Generally transporters, etc.
  - Most tentative; should be examined more carefully

Reaction Confidence: Sources of Evidence

- **Biochemical**
  Enzyme has been tested biochemically.
- **Genetic**
  Gene overexpression and purification, gene deletions.
- **Sequence**
  There is significant sequence similarity to another gene with known function.
- **Physiological**
  There is physiological data to support inclusion in the model.
- **Modeling**
  Reaction is included to improve simulation results.



Reaction: X5PL3E L-xylulose 5-phosphate 3-epimerase

xu5p-L --> ru5p-L

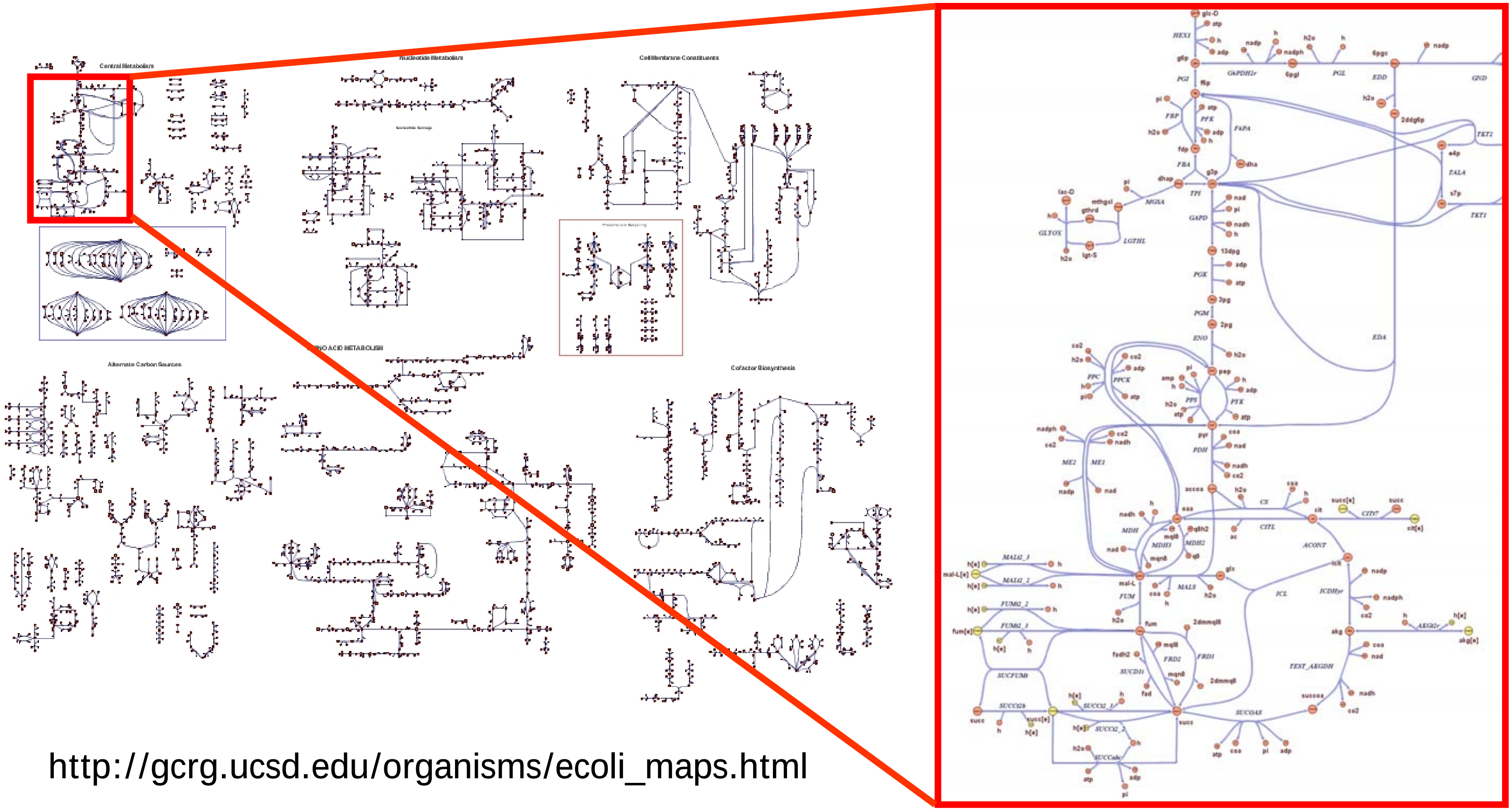
Subsystem: Alternate Carbon Metabolism

Confidence: 4.0

Supportive Evidence	Not Reviewed	[illegible]	Low	Medium	High
Biochemical Data	☐	[illegible]	[illegible]	☐	☑
Genetic Data	☐	[illegible]	☐	☐	☑
Sequence Data	☑	[illegible]	[illegible]	☐	☐
Physiological Data	☑	☐	☐	☐	☐
Modeling Data	☑	☐	☐	☐	☐

Gene to Reaction Connections

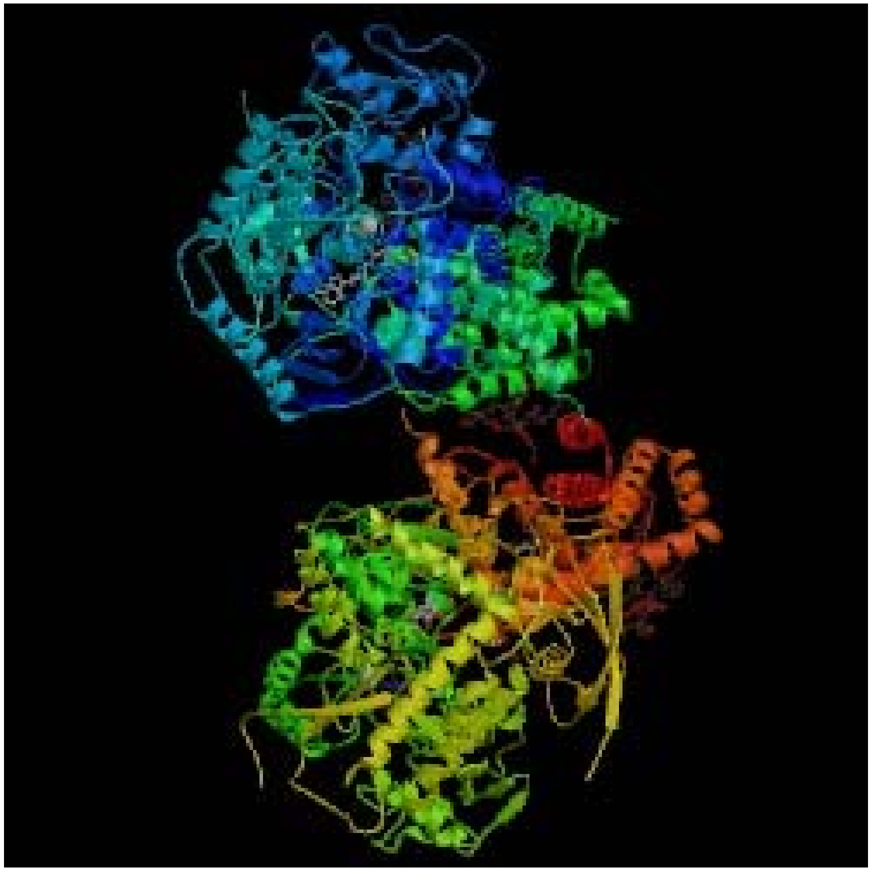
Escherichia coli Metabolism



http://gcrg.ucsd.edu/organisms/ecoli_maps.html

From Genes to Reactions

Not all genes have a one-to-one relationship with their corresponding enzymes or reactions



Many genes, one reaction: *frdABCD*

Four subunits combine to form fumarate reductase enzyme, catalyzing

FUM + $FADH_2$ → SUCC + FAD

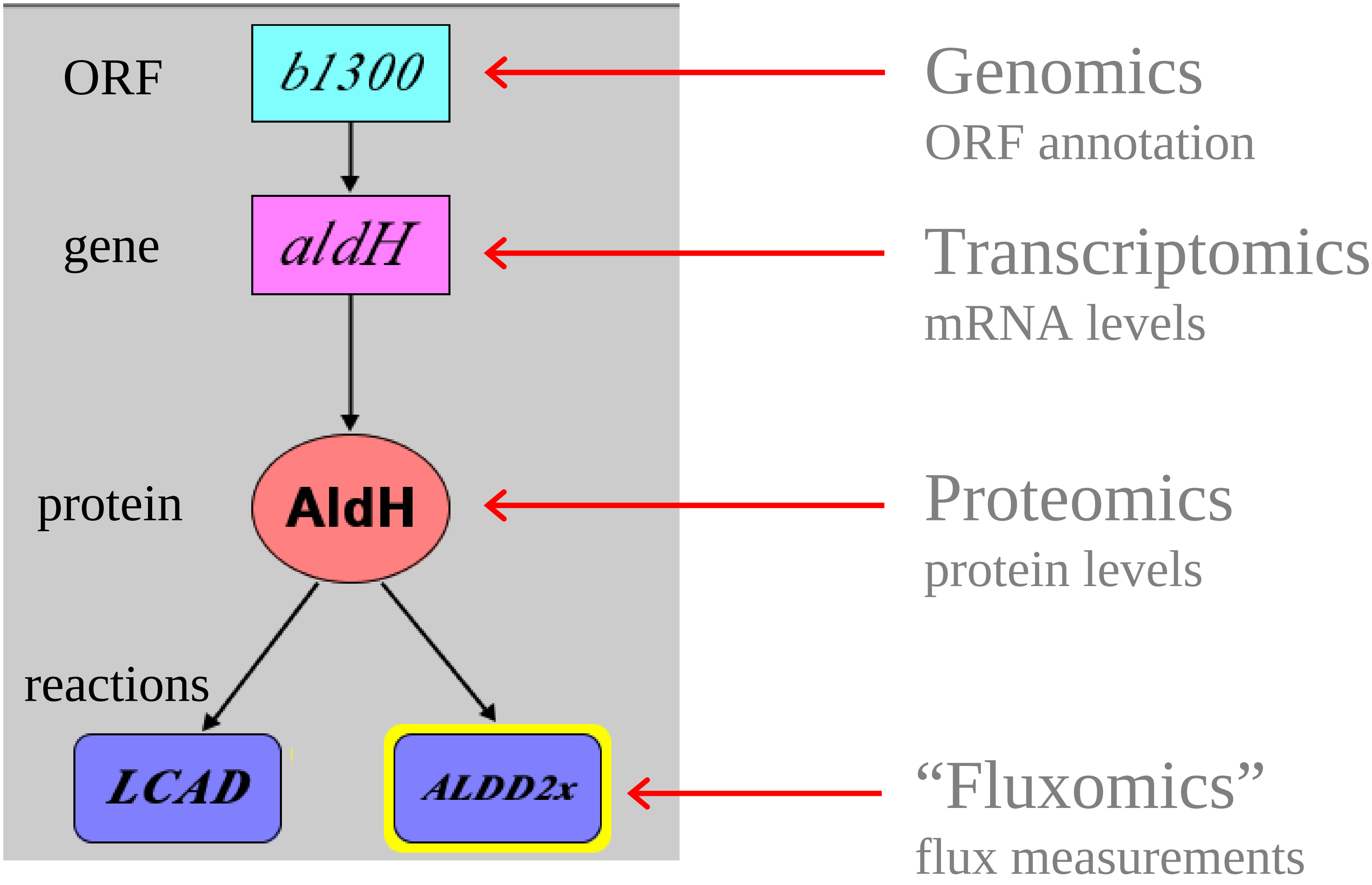
One gene, many reactions: *tktA*

One gene encodes transketolase I enzyme, catalyzing

R5P + X5P → T3P1 + S7P

E4P + X5P → T3P1 + F6P

Integrating “-omics” Data



ORF
b1300
gene
aldH
protein
AldH
reactions
LCAD
ALDD2x

Genomics
ORF annotation

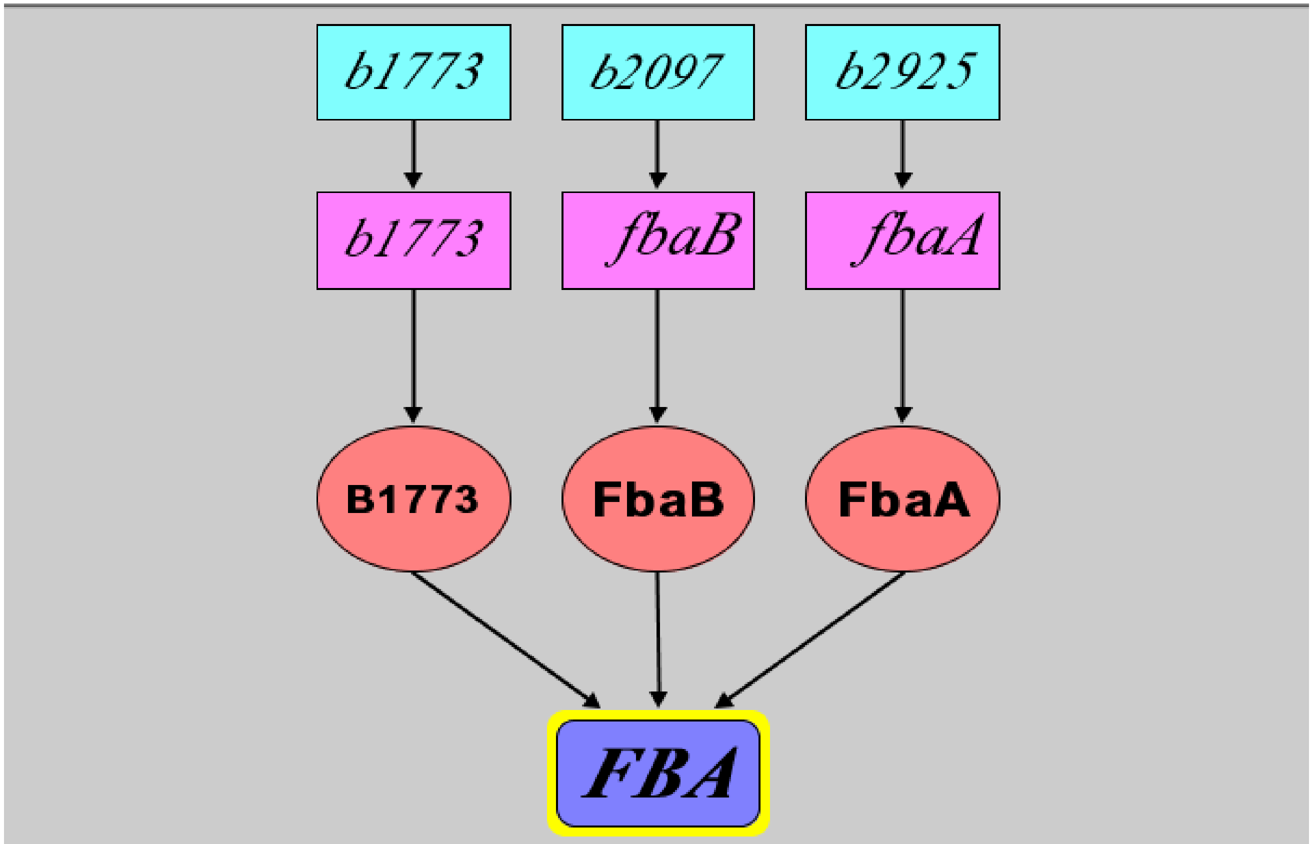
Transcriptomics
mRNA levels

Proteomics
protein levels

“Fluxomics”
flux measurements

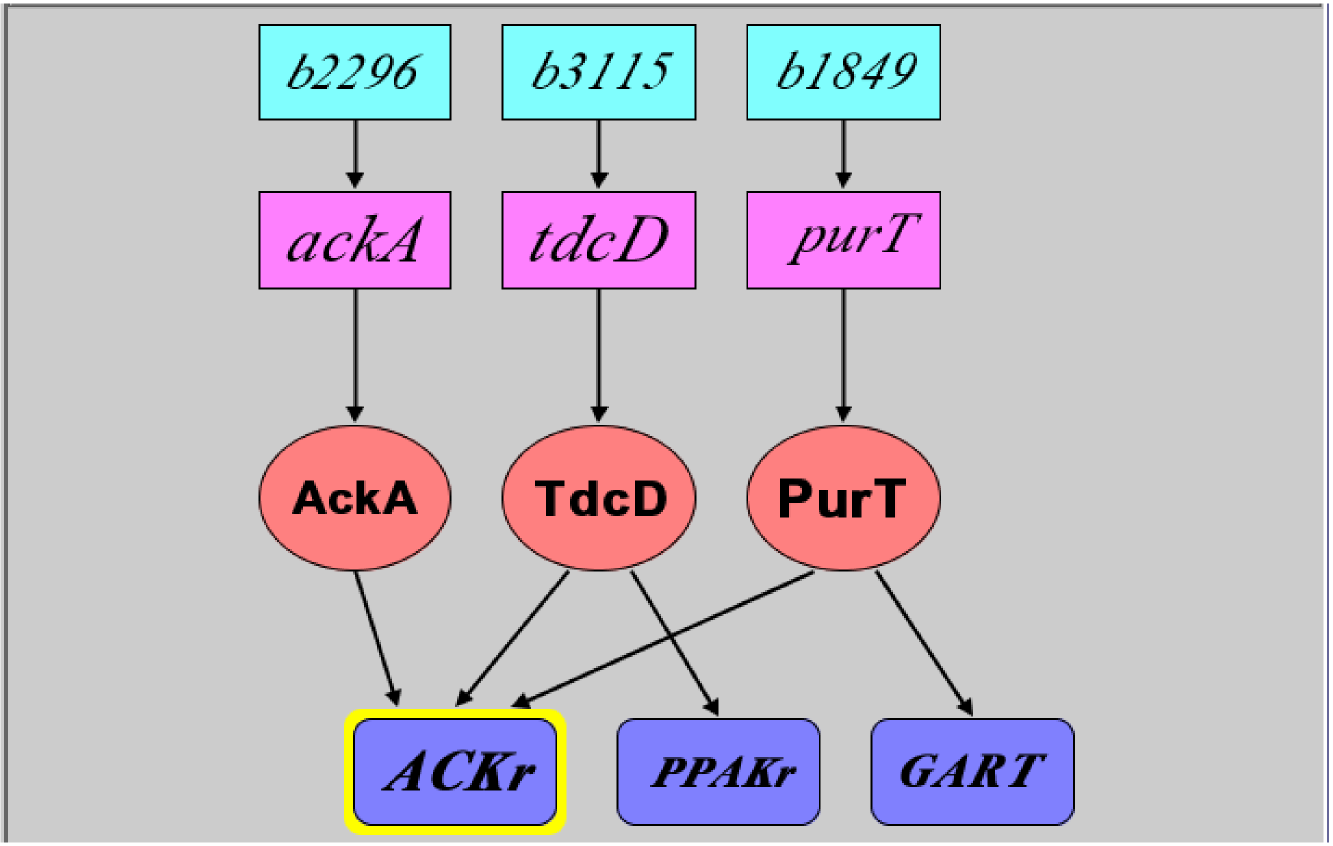
Example of Isozymes

fructose-1,6-bisphosphate aldolase

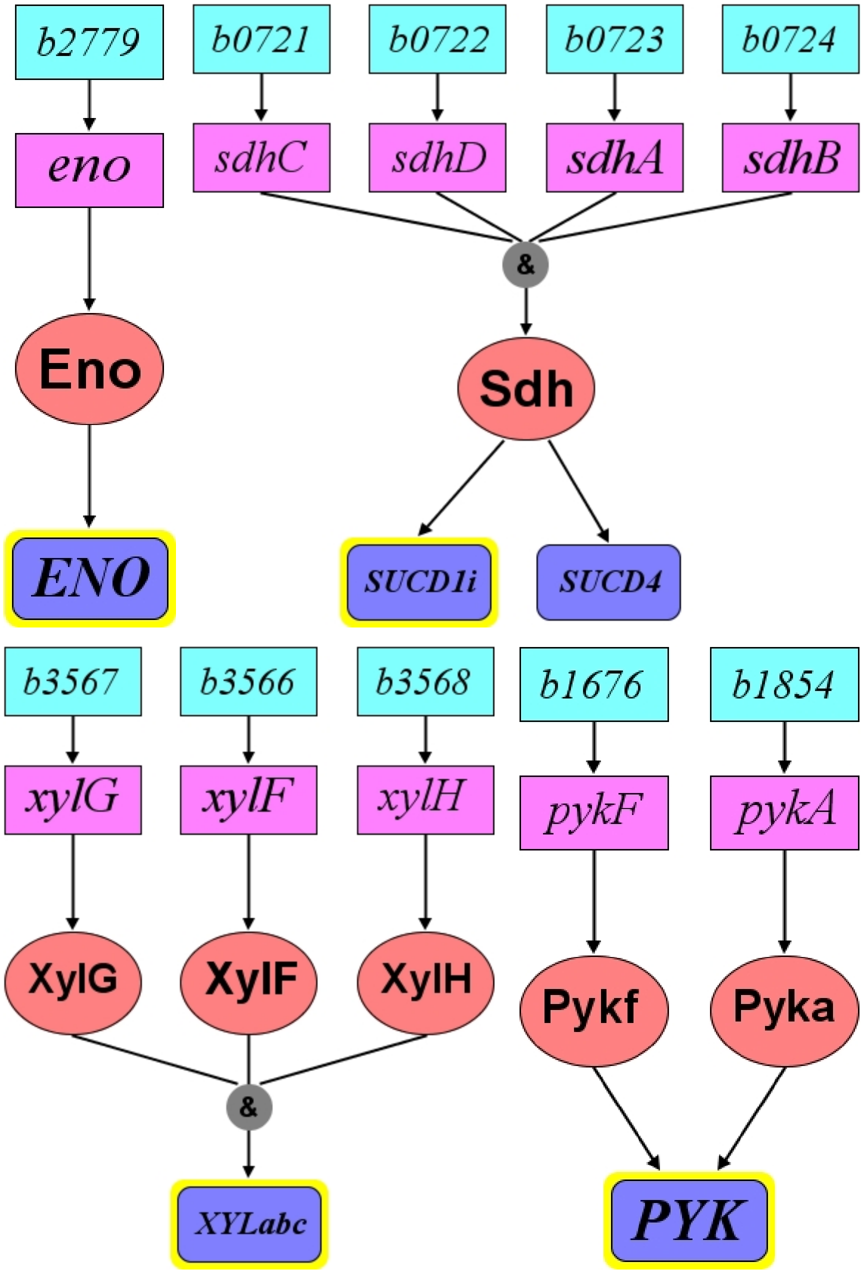


A More Complex Example

Pyruvate Metabolism



http://gcrg.ucsd.edu/organisms/ecoli_GPR.html



Enolase

1 gene
1 reaction

(342/904 Genes
are one-to-one)

Succinate Dehydrogenase

4 subunits
2 reactions

(135/904 Genes
interact as Subunits)

Xylose ABC Transporter

Protein complex
(3 proteins)
1 reaction

(153/904 Genes
interact as PCs)

Pyruvate Kinase

2 isozymes
1 reaction

(149/931 Reactions
have multiple isozymes)

Special Considerations

Biomass Composition

- Indicates demands of the system (more detail in modeling section of class)
- Precursors may also be used for smaller networks
- Approximation of Biomass composition for less-characterized organisms (*H. pylori*, *H. influenzae*)

Metabolite	Demand (mmol)
ATP	**41.3**
NAD$^+$	**3.5**
NADPH	**18.2**
G6P	**0.2**
F6P	**0.1**
R5P	**0.9**
E4P	**0.4**
GA3P	**0.1**
3PG	**1.5**
PEP	**0.5**
PYR	**2.8**
ACCOA	**3.7**
OXA	**1.8**
AKG	**1.1**
SUCCOA	**(trace)**

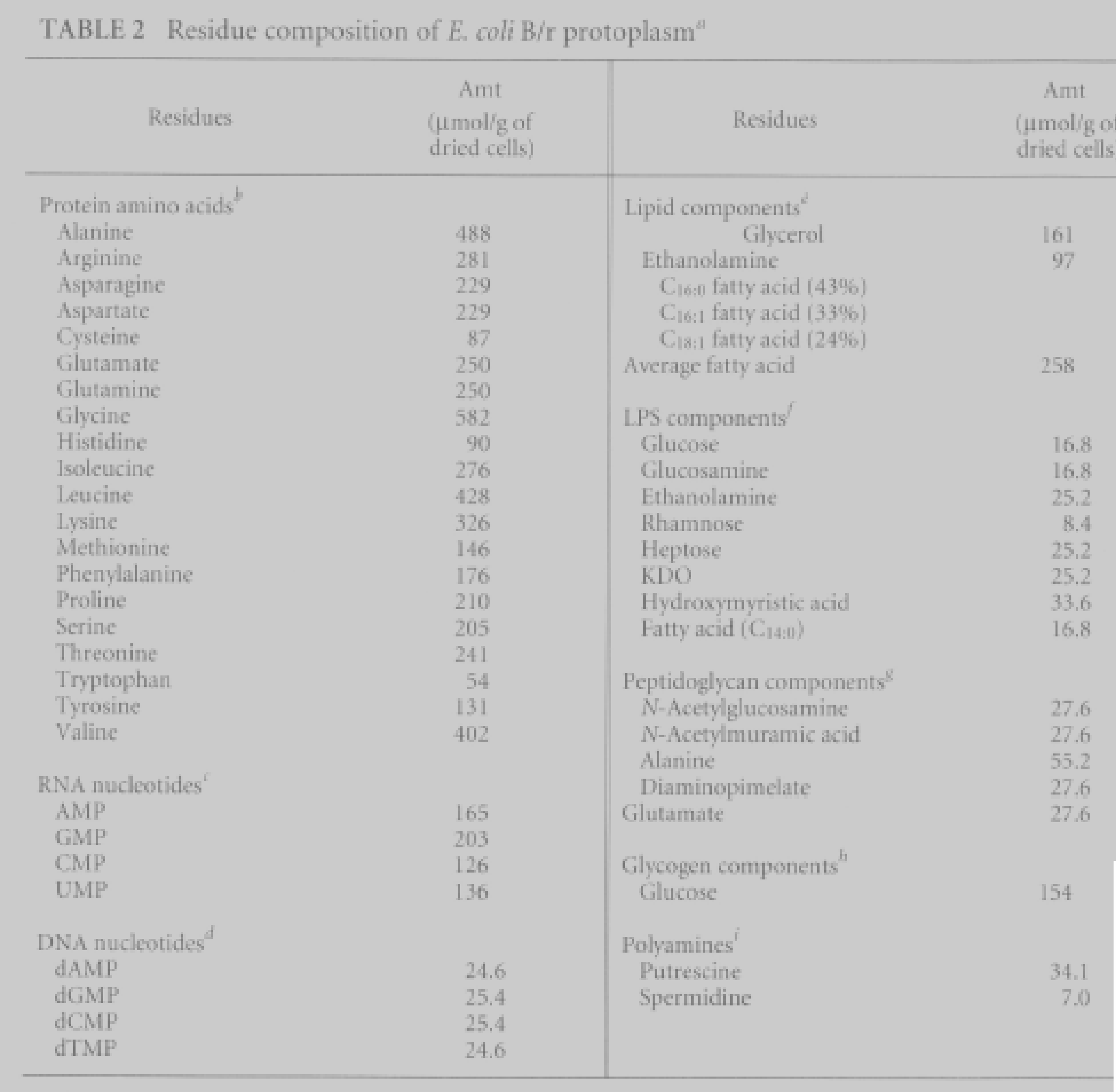
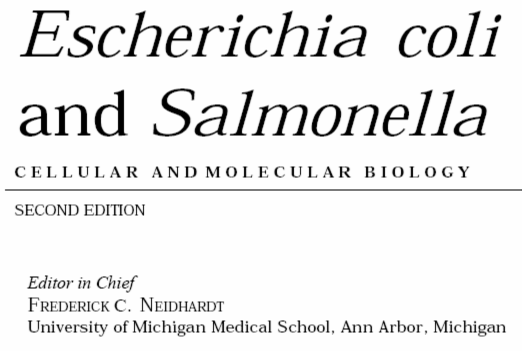


TABLE 2 Residue composition of *E. coli* B/r protoplasm[a]

Residues	Amt (µmol/g of dried cells)	Residues	Amt (µmol/g of dried cells)
Protein amino acids[b]		Lipid components[e]	
Alanine	488	Glycerol	161
Arginine	281	Ethanolamine	97
Asparagine	229	$C_{16:0}$ fatty acid (43%)	
Aspartate	229	$C_{16:1}$ fatty acid (33%)	
Cysteine	87	$C_{18:1}$ fatty acid (24%)	
Glutamate	250	Average fatty acid	258
Glutamine	250		
Glycine	582	LPS components[f]	
Histidine	90	Glucose	16.8
Isoleucine	276	Glucosamine	16.8
Leucine	428	Ethanolamine	25.2
Lysine	326	Rhamnose	8.4
Methionine	146	Heptose	25.2
Phenylalanine	176	KDO	25.2
Proline	210	Hydroxymyristic acid	33.6
Serine	205	Fatty acid ($C_{14:0}$)	16.8
Threonine	241		
Tryptophan	54	Peptidoglycan components[g]	
Tyrosine	131	*N*-Acetylglucosamine	27.6
Valine	402	*N*-Acetylmuramic acid	27.6
		Alanine	55.2
RNA nucleotides[c]		Diaminopimelate	27.6
AMP	165	Glutamate	27.6
GMP	203		
CMP	126	Glycogen components[h]	
UMP	136	Glucose	154
DNA nucleotides[d]		Polyamines[i]	
dAMP	24.6	Putrescine	34.1
dGMP	25.4	Spermidine	7.0
dCMP	25.4		
dTMP	24.6		



Escherichia coli and *Salmonella*

CELLULAR AND MOLECULAR BIOLOGY

SECOND EDITION

Editor in Chief
FREDERICK C. NEIDHARDT
University of Michigan Medical School, Ann Arbor, Michigan

Charge Considerations

- An underappreciated aspect of building reaction networks—electrical charge should be conserved in all reactions
- Phosphofructokinase (from KEGG):

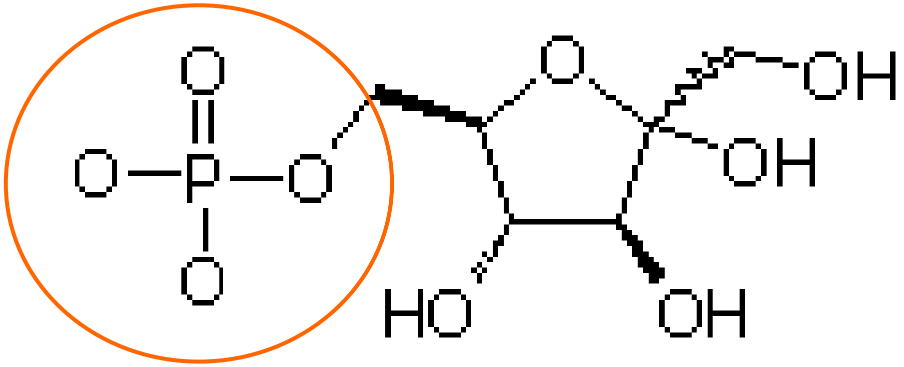
ATP + F6P => ADP + FDP $+ H^+$

charges -4 + -2 -3 + -4 + 1

= -6 = -7 = -6

Finding Compound Charges

- Consult diagram and look at each chemical group independently
- Determine if H's are attached or dissociated at cellular pH (attached if pKa<pH)—can find pKa values
- F6P:



Both H's dissociate at cellular pH, leaving a charge of -2

Balancing Reactions: An Example

pyruvate decarboxylase

Solution # 1:

$$\mathrm{pyr} \rightarrow \mathrm{acald} + \mathrm{co}_2$$

C3H3O3	C2H4O	CO2	ELEMENTALLY IMBALANCED (H and O)
(-1)	(0)	(0)	CHARGE IMBALANCED

Solution # 2:

$$\mathrm{h}_2\mathrm{o} + \mathrm{pyr} \rightarrow \mathrm{acald} + \mathrm{co}_2 + ½\ \mathrm{o}_2 + \mathrm{h}^+$$

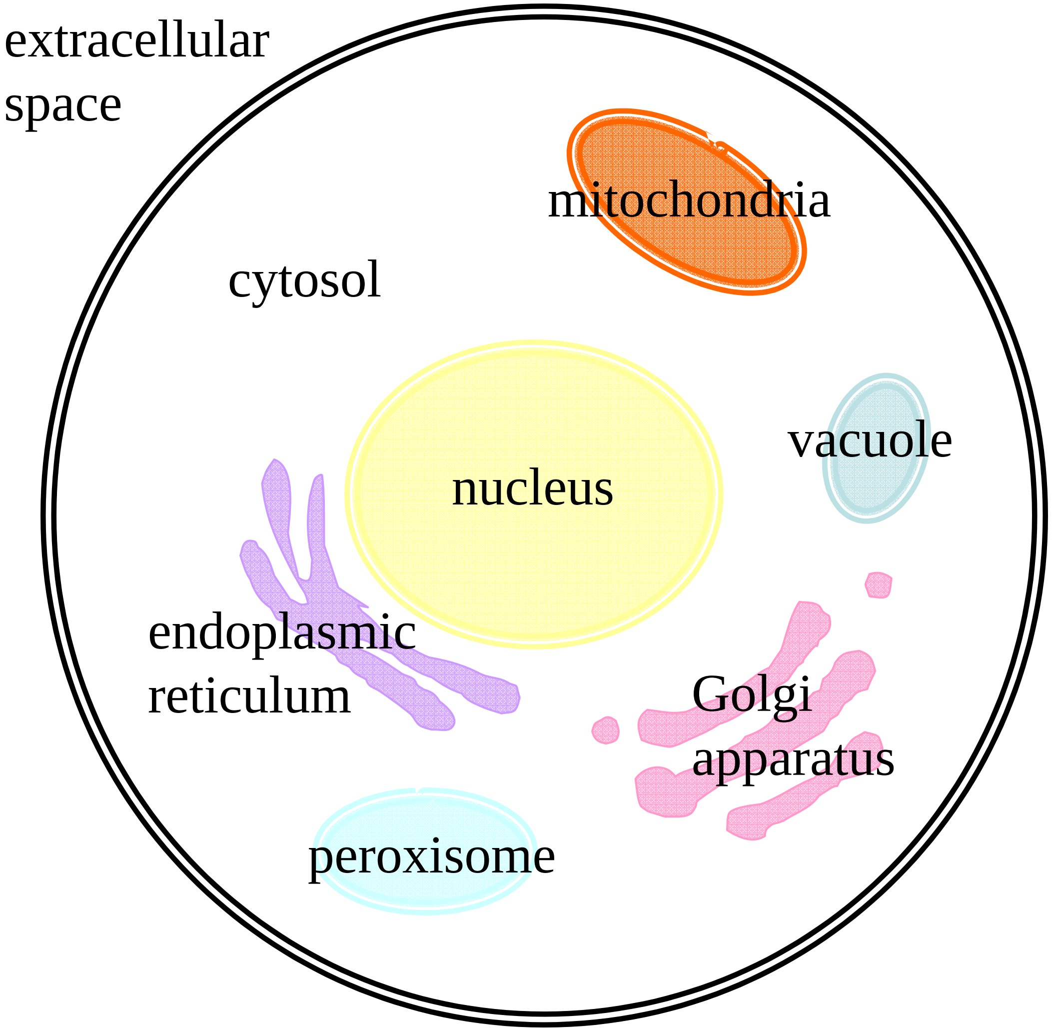
H2O	C3H3O3	C2H4O	CO2	O2	H	ELEMENTALLY BALANCED
(0)	(-1)	(0)	(0)	(0)	(+1)	

Solution # 3:

$$\mathrm{h}^+ + \mathrm{pyr} \rightarrow \mathrm{acald} + \mathrm{co}_2$$

H	C3H3O3	C2H4O	CO2	ELEMENTALLY BALANCED
(+1)	(-1)	(0)	(0)	

Compartmentalization



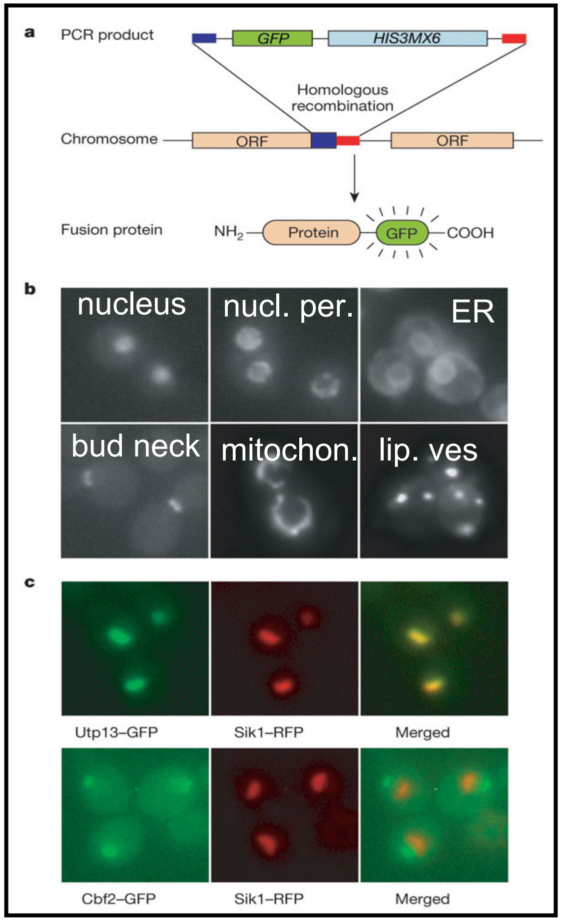
There are 8 compartments included in our yeast model

May need to infer transport reactions between compartments

H+, ATP, NADH, NADPH must be balanced within each compartment

Protein Localization

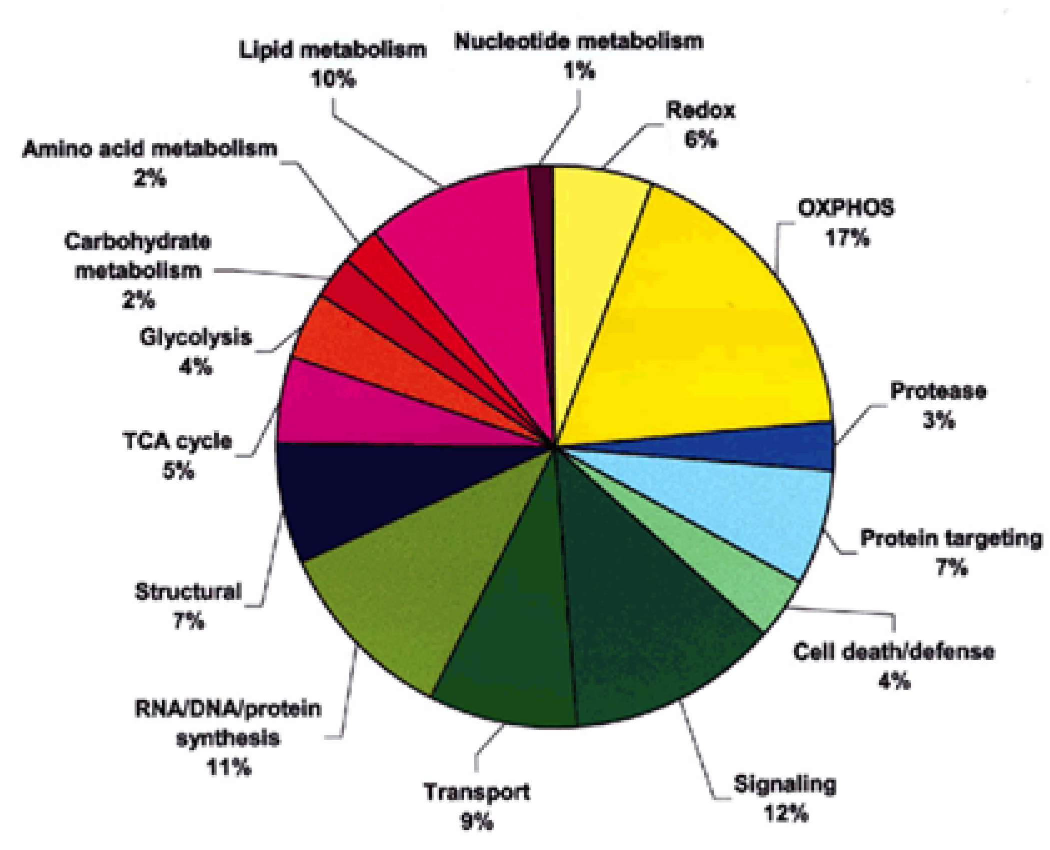
- Compartmentalization is a key part of network reconstruction
- Both a component (static localization) and state (dynamic localization) data type
- Techniques typically based on GFP-tagging of proteins or isolation of specific organelles
- Potential problems:
  - Effect of the GFP tag on localization
  - Usually human assistance is required in image analysis
  - Condition dependence: e.g. mitochondrial localization agrees only in 30% of cases in yeast between three different data sets



Huh et al. Nature 425:686 (2003)

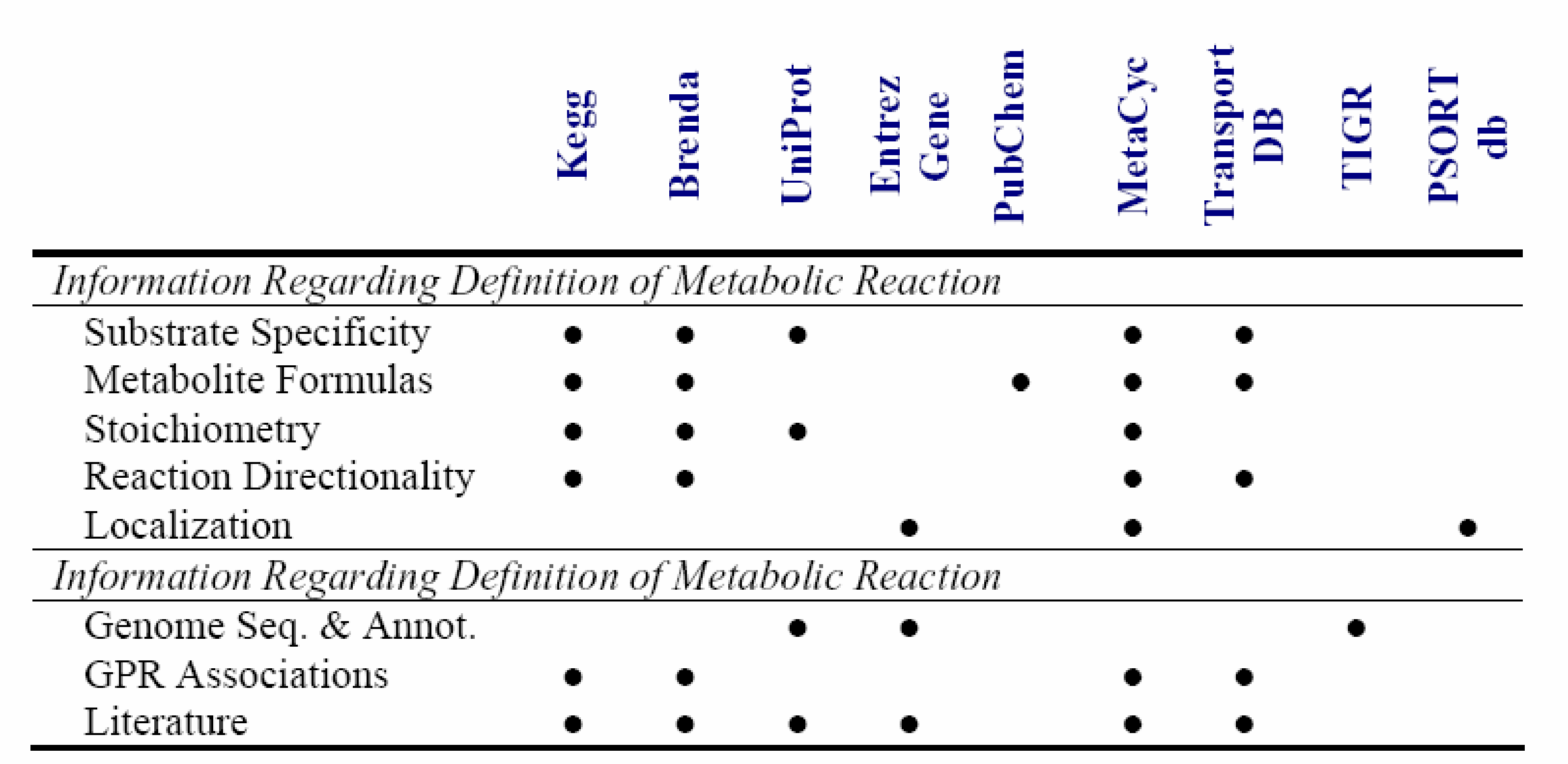
Mitochondrial Isolation: Protein Identification

- 657 distinct proteins
- 498 (81%) functionally classified into 15 cellular processes.
- 153 unique enzymatic activities
  - Glycolysis, TCA cycle, oxidative phosphorylation, urea cycle, fatty acid oxidation, lipid and heme biosynthesis.



Taylor S. *et al*, Nature Biotech (21), 2003

Metabolic Databases



	Kegg	Brenda	UniProt	Entrez Gene	PubChem	MetaCyc	Transport DB	TIGR	PSORT db
Information Regarding Definition of Metabolic Reaction									
Substrate Specificity	●	●	●			●	●		
Metabolite Formulas	●	●			●	●	●		
Stoichiometry	●	●	●			●			
Reaction Directionality	●	●				●	●		
Localization				●		●			●
Information Regarding Definition of Metabolic Reaction									
Genome Seq. & Annot.			●	●				●	
GPR Associations	●	●				●	●		
Literature	●	●	●	●		●	●		

Genome-Scale Reconstructions

	Genes	SKI	N_G	N_m	N_R	Status	Ref
BACTERIA							
Bacillus subtilis	4,225	4.8	614	637	754	C, E	93
Escherichia coli	4,405	55.1	904	625	931	C, E	39
			720	438	627	C, E	90
			961	NA	1,107	C	53
Francisella tularensis	1,804		350[a]	NA	429	C	68
Geobacter sulfurreducens	3,530		588	541	523	C, E	d
Haemophilus influenzae	1,775	8.9	296	343	488	C, E	94
			400	451	461	C, E	95
Helicobacter pylori	1,632	13	341	485	476	C, E	61
			291	340	388	C, E	96
			301[a,c]	442[c]	533[c]	C	63
Lactococcus lactis	2,310		358	422	621	C,E	97
Mannheimia succiniproducens	2,463		335	352	373	C, E	98
Pseudomonas putida	5,441		523	NA	581	C, E	e
Pseudomonas aeruginosa	5,640	5.7	516	NA	647	C, E	e
			718	623	800	C	67
Staphylococcus aureus	2,702	16	619	571	641	C, E	4
Streptomyces coelicolor	8,042	0.13	700	500	700	C, E	36
ARCHAEA							
Methanococcus jannaschii	1,821	0.3	436[a]	510	609	C	64
Methanosacrcina barkerii	5,072		692	558	619	C, E	f
EUKARYA							
Arabidopsis thaliana	28,848		1,418	NA	894	C	66
Homo sapiens	28,783	48.5	2,709[a]	661	1,093	C	65
Mus musculus	28,287	15.6	1,156[b]	872	1,220	C, E	92
Plasmodium falciparum	5,342		737[a]	525	697	C	3
Saccharomyces cerevisiae	6,183	10.6	750	646	1,149	C, E	45
			708	584	1,175	C, E	91