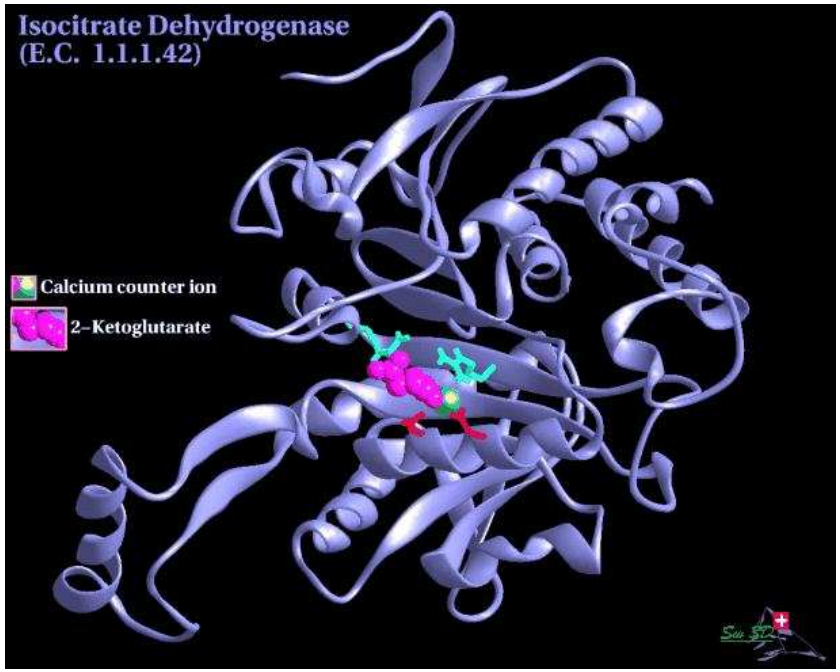
Metabolism and metabolic networks

- Metabolism is the means by which cells acquire energy and building blocks for cellular material
- Metabolism is organized into sequences of biochemical reactions, metabolic pathways
- Pathways are interconnected in many ways, thus their total is a metabolic network, concisting of reactions and compounds (the metabolites).

Reactions and enzymes

- The basic building block of metabolic networks is a (bio)chemical reaction.
- Most reactions that occur within a living cell are catalyzed by enzymes, a class of proteins.
- Enzymes are highly specific, a single enzyme can catalyze only one (or at most a couple) kind of a reaction.
- This enables the cell to control the production of certain metabolites without altering everything else at the same time.



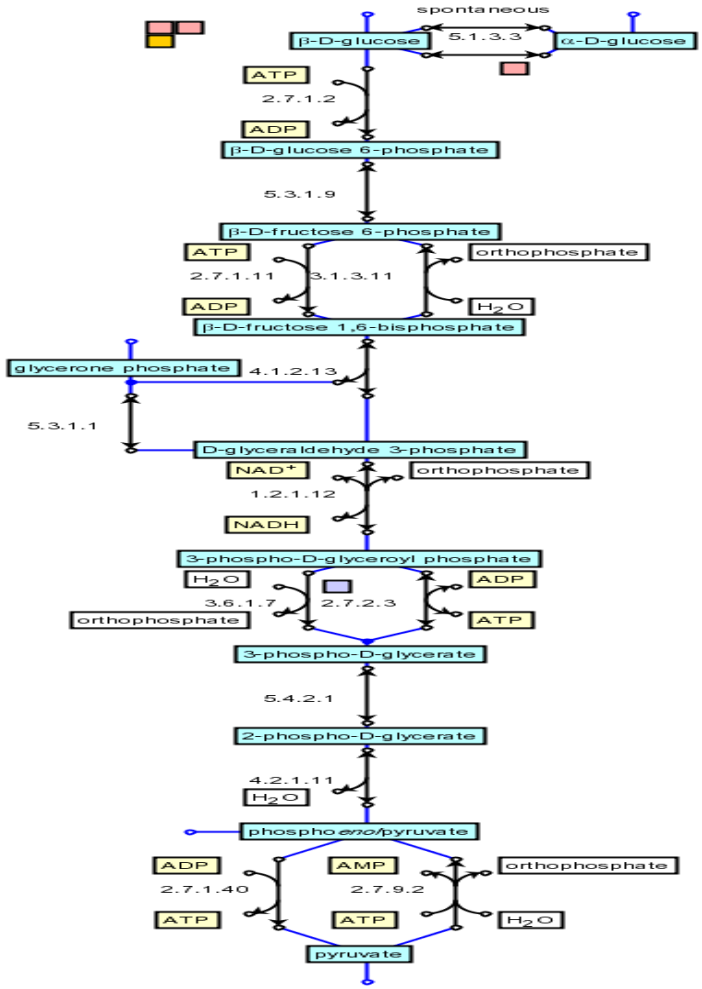
picture from SWISS-3D Database,

http://www.expasy.ch/sw3d/

Metabolic networks

The individual enzymatic reactions are organized into pathways, sequences of reactions. The pathways are interconnected in many ways, which makes the metabolism a directed network.

The network contains both cycles and biconnected components, i.e. alternative routes from one compound to another



(picture:E.Coli glycolysis, EMP database, www.empproject.com/)

Pathway integration

The metabolic pathways are integrated with another in two ways:

- First, some metabolites may take part into two or more reactions. This creates branches into the metabolic network. Forward branches occur when a sinlge metabolite is consumed by two or more reactions, backward branches occur when a metabolite is produced by two or more reactions.
- Second, many reactions involve a co-factor conversions as side-reactions (ATP $\leftrightarrow$ ADP,NAD $\leftrightarrow$ NADH,NADP $\leftrightarrow$ NADPH). The balance of co-factors induces dependencies between otherwise "distant" reactions.

Thus, to fully understand the behaviour of metabolism one should not blindly restrict one's attention to individual genes or pathways.

Spatial organization

Especially in the eukaryotic cells many of the reactions occur in specific parts of the cell.

- DNA and RNA synthesis occurs in the nucleus
- many fueling and biosynthetic reactions occur in mitochondria
- there are transport mechanisms transferring the molecules from one place to another.
- a given reaction may have different instantiations in the cell, i.e. the same reaction can take place in mitochondria and in cytoplasm.

The "bag of enzymes" model

- However, with present techniques it is difficult to obtain detailed information on the location of the reactions. Therefore, in these lectures we often consider the cell a "bag of enzymes" with no fine structure.
- In this model, a single reaction occurs at most once in the metabolic network. Also, the transport reactions are ignored.
- In the quantitative calculations, we further assume that the cell is "well-mixed" so that the concentration of a molecule is the same in every part of the cell.

Types of reactions

Fueling reactions produce the precursor molecules needed for biosynthesis. In addition they generate energy, in the form of ATP, which is used by biosynthesis, polymerization and assemply reactions.

Biosynthetic reactions produce building blocks used by the polymerization reactions. Biosynthetic reactions are organized into biosynthetic pathways, reation sequences of one to a dozen reactions. All biosynthetic pathways begin with one of 12 precursor molecules.

Polymerization reactions link molecules into long polymeric chains.

Assembly reactions carry out modifications of macromolecules, their transport to prespecified locations in the cell and their association to form cellular structure such as cell wall, membranes, nucleus, etc.

Fuelling reactions

Fuelling reactions serve three purposes

1. Generation of energy, mostly in the form of ATP, to be used in biosynthesis, polymerization and assembly reactions.
2. Production of reducing power, mainly in the form of NADPH, to be used in biosynthesis.
3. Formation of the 12 precursor molecules needed in biosynthesis: glucose-6-phosphate, fructose-6-phosphate, ribose-5-phosphate, erythrose-4-phosphate, dihydroxy acetone phosphate and glyceraldehyde 3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, acetyl CoA,α-ketoglutarate, succinyl CoA, oxaloacetate

Glycolysis

Glycolysis is the sum of all biochemical reactions by which glucose is converted to pyruvate.

The most frequently encountered pathways implementing glycolysis are Embden-Meyerhof-Parnas pathway (EMP), Pentose-phosphate pathway (PP), and Entner-Doudoroff pathway (ED) (only in prokaryotes)

9 of the 12 precursor molecules are produced in EMP and PP pathways.

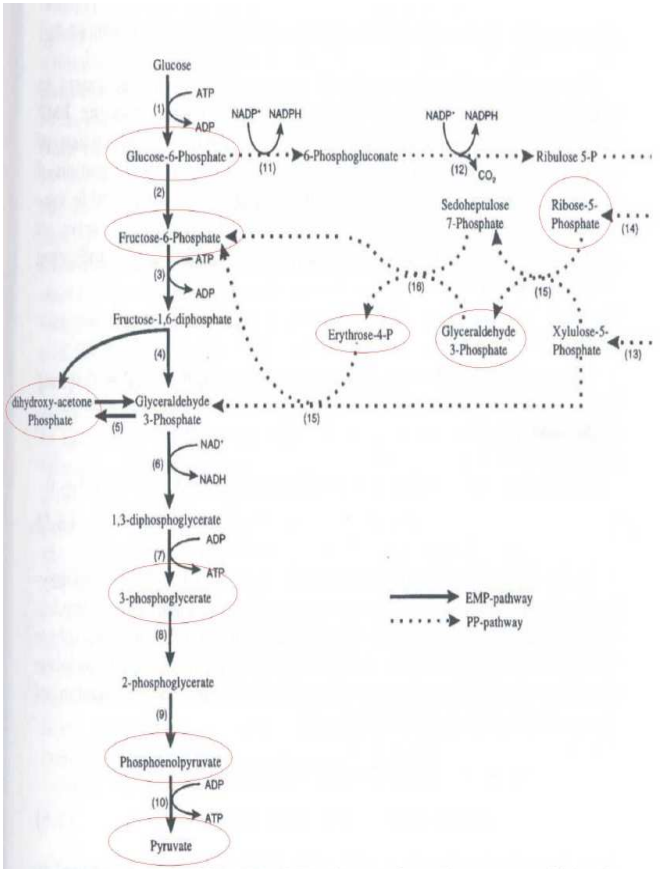


FIGURE 2.6 Overview of the EMP and PP pathways in fungi. The enzymes are (1) hexokinase; (2) phosphohexoisomerase; (3) phosphofructokinase; (4) aldolase; (5) triosephosphate isomerase; (6) 3-phosphoglyceraldehyde dehydrogenase; (7) 3-phosphoglycerate kinase; (8) phosphoglycerate mutase; (9) enolase; (10) pyruvate kinase; (11) glucose-6-phosphate dehydrogenase; (12) 6-phosphogluconate dehydrogenase; (13) ribulosephosphate-3-epimerase; (14) ribosephosphate isomerase; (15) transketolase; (16) and transaldolase.

TCA cycle

In aerobic organisms, most pyruvate enters the TCA cycle, where it is converted into CO_2 and water.

In eukaryotes, TCA cycle recides in mitochondria.

Three of the TCA cycle intermediates, α-ketoglutarate, succinyl-CoA and oxaloacetate, are entry points for biosynthetic pathways.

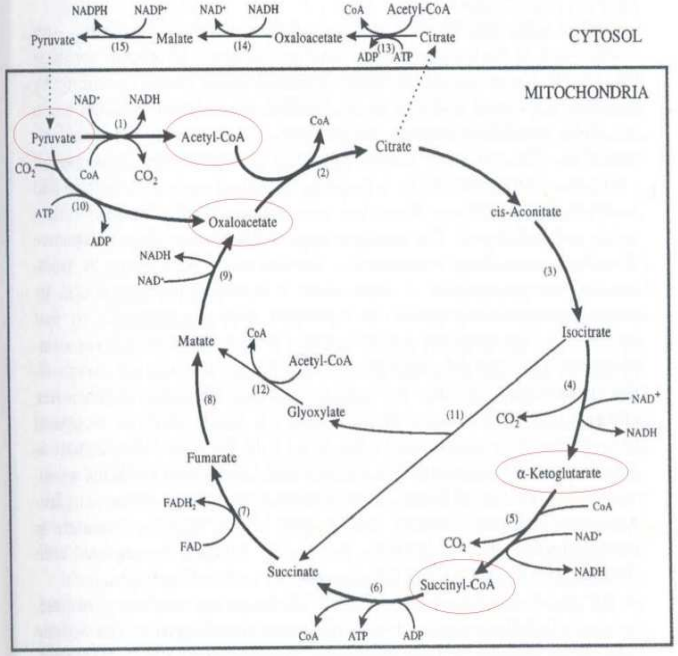


FIGURE 2.10 Overview of the TCA cycle and anaplerotic pathways in fungi. The enzymes are (1) pyruvate dehydrogenase complex; (2) citrate synthase; (3) aconitase; (4) isocitrate dehydrogenase; (5) α-ketoglutarate dehydrogenase; (6) succinate thiokinase; (7) succinate dehydrogenase; (8) fumarase; (9) malate dehydrogenase; (10) pyruvate carboxylase; (11) isocitrate lyase; (12) malate synthase; (13) ATP-citrate lyase; (14) malate dehydrogenase; (15) malic enzyme. In eukaryotic cells, the pyruvate dehydrogenase complex is membrane-bound, such that the conversion of pyruvate to acetyl-CoA is associated with the transport of pyruvate into the mitochondria (there is no carrier for pyruvate in the mitochondrial membrane). Furthermore, the *glyoxylate cycle*, i.e., reactions catalyzed by enzymes (2), (3), (11), (12), and (9), functions not in the mitochondria (as might be suggested from the figure) but in microbodies called *glyoxysomes*, where there is a net synthesis of succinate from acetyl-CoA. Succinate produced in the glyoxylate cycle is then transported to the mitochondria where it enters the TCA cycle. In prokaryotes there is no physical organization of the TCA cycle reactions all occurring in the cytosol, and reactions (13), (14), and (15) consequently are not needed.

Fermentative pathways

In anaerobic organisms and in aerobic organisms when oxygen-supply is limited, pyruvate may be converted into metabolic products like lactic acid or ethanol via fermentative pathways.

When engineering microbes for industrial use, boosting a particular fermentative pathway is often desired.

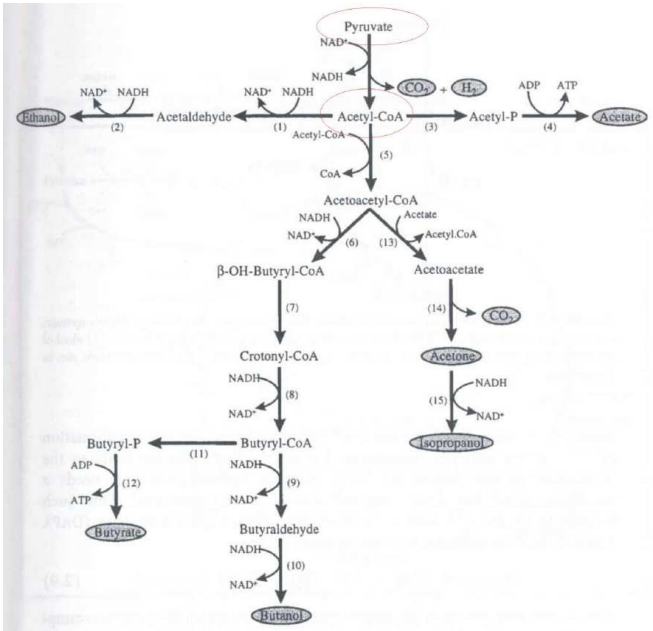
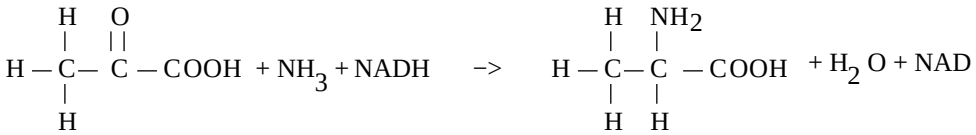


FIGURE 2.8 Mixed fermentation in *C. acetobutylicum*. The enzymes are (1) acetaldehyde dehydrogenase; (2) ethanol dehydrogenase; (3) phosphotransacetylase; (4) acetate kinase; (5) acetyl-CoA-acetyl transferase; (6) L(+)-β-hydroxybutyryl-CoA dehydrogenase; (7) 1,3-hydroxy-acyl-CoA hydrolase; (8) butyryl-CoA dehydrogenase; (9) butyraldehyde dehydrogenase; (10) butanol dehydrogenase; (11) phosphotransbutyrylase; (12) butyrate kinase; (13) CoA transferase; (14) acetoacetate decarboxylase; (15) isopropanol dehydrogenase.

Biosynthetic reactions

Biosynthetic pathways produce the building blocks, i.e. amino acids, nucleotides, lipids, sugars etc. needed for growing larger structures in the cell such as membranes, DNA, RNA etc.

Amino acid biosynthesis, as its simplest, involves acquiring a amino-group (NH2) to the precursor, like it is the case for alanine synthesis from pyruvate.



pyruvate alanine

Biosynthesis of amino acids

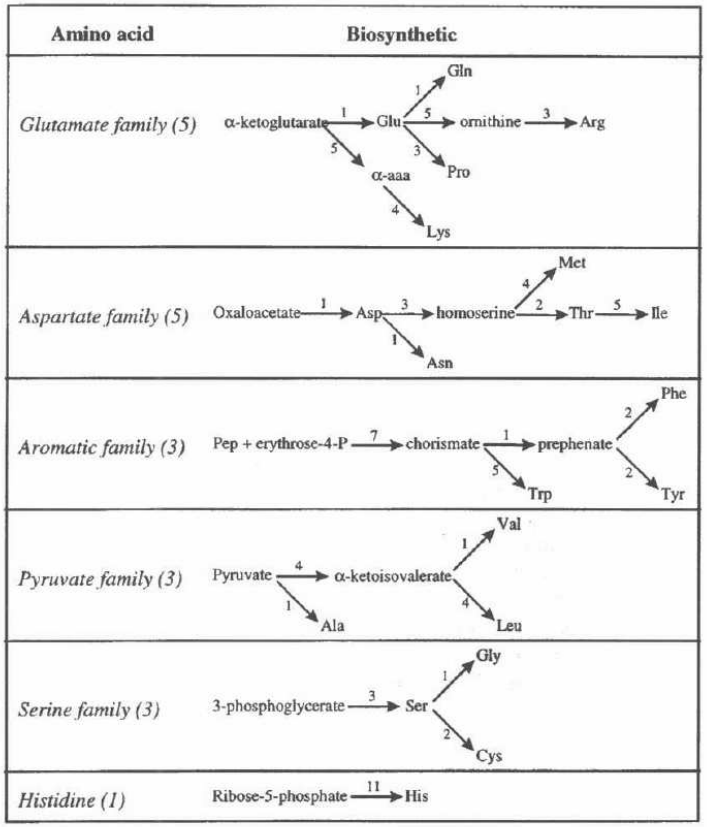


FIGURE 2.12 Overview of amino acid biosynthesis in eukaryotes. The amino acids are classified into five families according to the specific precursor metabolite or amino acid that serves as the starting point for their synthesis. L-Histidine, which has a complex biosynthetic pathway, does not group with any of the other amino acids. The numbers indicate the reaction steps in the pathway. Except for L-lysine, these numbers are the same for bacteria. In bacteria L-lysine is synthesized from aspartate via diaminopimelic acid (an important building block for bacteria cell wall) in a sequence of nine reactions.

How does an enzyme work?

An enzyme works by binding the substrate molecules into the so called active site. In the active site, the substrates end up in such a mutual geometric conformation that the reaction occurs effectively.

The occurence of the reaction causes the enzyme to change its conformation, which releases the products. After that, the enzyme is ready to bind another set of substrates. The enzyme itself stays unchanged in the reaction.

Enzyme activity

The rate of certain enzyme-catalyzed reaction depends on the concentration (amount) of the enzyme and the specific activity of the enzyme (how fast a single enzyme molecule works).

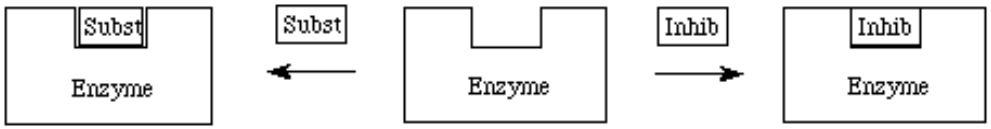
The specific activity of the enzyme depends on

- pH and temperature
- positively on the concentration of the substrates
- negatively on the concentration of the end-product of the pathway (inhibition).

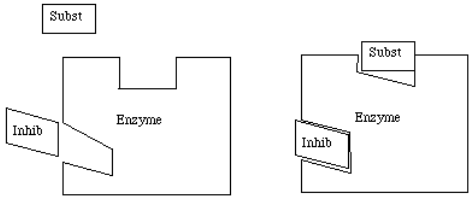
Note that transcription level gene regulation **directly** affects only the concentration of the enzyme.

Inhibition of Enzymes & Metabolic-level regulation

- The activity of enzymes is regulated in the metabolic level by inhibition: certain metabolites bind to the enzyme hampering its ability of catalysing reactions.
- In competitive inhibition, the inhibitor allocates the active site of the enzyme, thus stopping the substrate from entering the active site.



- In non-competitive inhibition, the inhibitor molecule binds to the enzyme outside the active site, causing the active site to change conformation and making the catalysis less efficient.



Metabolic reconstruction problem

From the sequenced genome, the task is to infer the encoded metabolic network.

MYSIVKEIIVDPYKRLKWGFIPVKRQVEDLPDDLNS

SIQSHETAENFITTTSEKDQLHFETSSYSEHKDNVN

EYRINEKERSHNKWYSWFKQGTSFKEKKLLIKLDVL

YVSGMKEDLGFQGNDLVHTQVMYTVGNIIFQLPFLI

VGAAYVNSVPHLKAIRFFIGAFEAPSYLAYQYLFGS

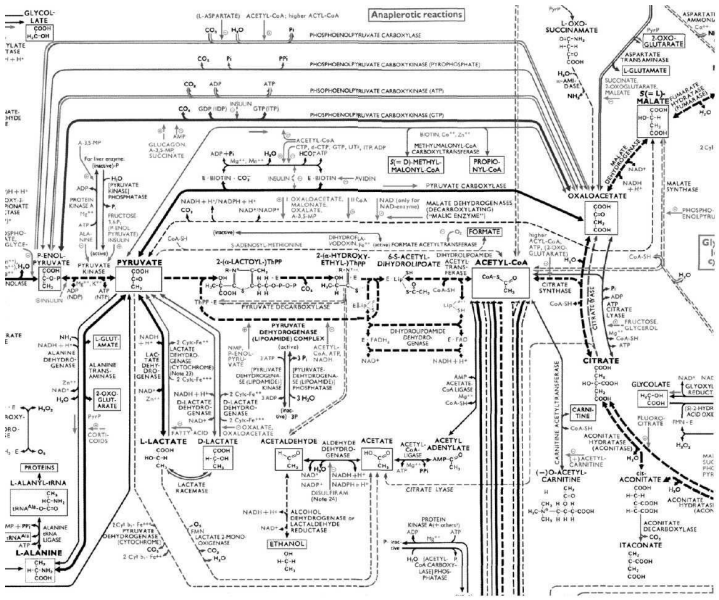
LSAGGIQSAVYSSLNGVNGLEGWRWNFIIDAIVSVV

DDEIRLARKRLKENQTGKSDFETKVFDIKLWKTIFS

GAYLLWLKSLKRYSIPKLNQLSMITPGLGMVYLMLT

IGNSILAAWDVAEGAKWFAFMLQCFGWAMAPVLYST

AQSSTAWISVLVWKTEEAPRYLKGFTFTACSAFCLS



Data sources for Metabolic Reconstruction

The principal kinds of data for reconstruction (roughly in the order of reliability) are:

- Biochemistry: an enzyme has been isolated from an organism, and its function has been demonstrated (experimentally in test tube, or uncovering its 3D structure and simulating its behaviour in a computer).
- Genomics. Functional assigment to open reading frames (ORFs) based on DNA sequence homology. These annotations are often subject to revision and updates.
- Physiology and indirect information. Physiological ability of the cell (e.g. capability to produce certain metabolite) may lead us to "fill in the pathway" so that the resulting network has this ability
- Modeling and simulation studies. The network needs to be able to simulate cell behaviour in silico (e.g. it needs to be able to produce all necessary components of biomass)

Resources in the web

There are numerous online resources that can be used to aid metabolic reconstruction. Rougly, they can be divided into the following categories.

- Databases with annotated genomes and annotation software
- Enzyme databases
- Pathway databases
- Automatic reconstruction tools

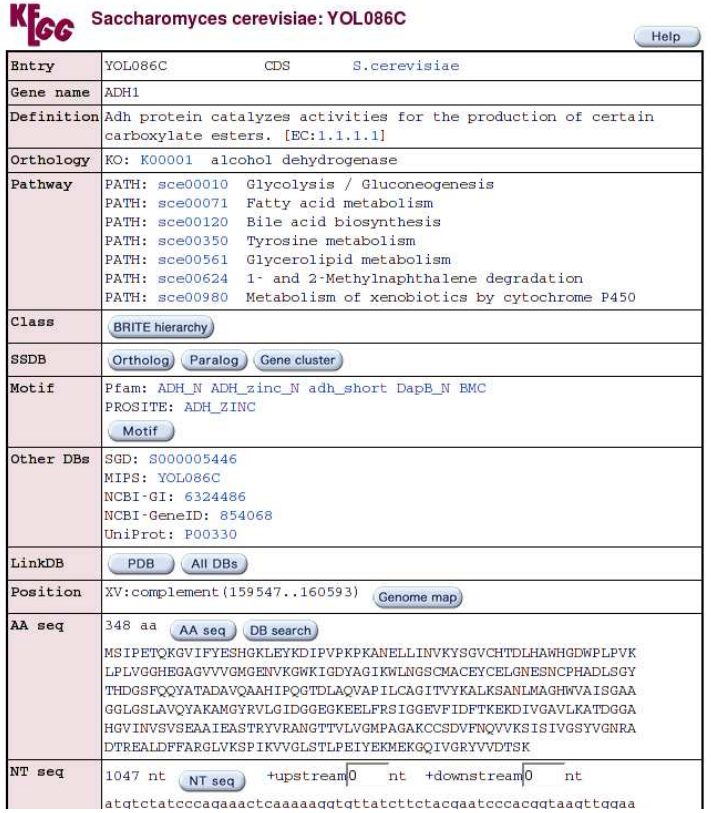
Most services in the web provide some mixture of these tools

KEGG - Kyoto Encyclopedia of Genes and Genomes (http://kegg.com)

- Knowledge base aiming to integrate genetic and higher-level information
- Project initiated in 1995 under the Human Genome Project.
- Genetic information contained in GENES database
- Higher-order functional information in PATHWAY database
- LIGAND databse contains information about chemical compounds, enzyme molecules and enzymatic reactions.
- Downloadable for academic users via ftp://ftp.genome.ad.jp/pub/kegg/.

GENES database

- Data from 557 genomes, majority completely sequenced
- $\approx$ 2 000 000 entries
- For each gene
  - Identification
  - Classification according to KEGG/PATHWAYS
  - Known sequence motifs
  - Chromosomal position
  - Amino acid and nucleotide sequences
  - Links to other databases (Genbank, SWISS-PROT)



KEGG Saccharomyces cerevisiae: YOL086C

Help

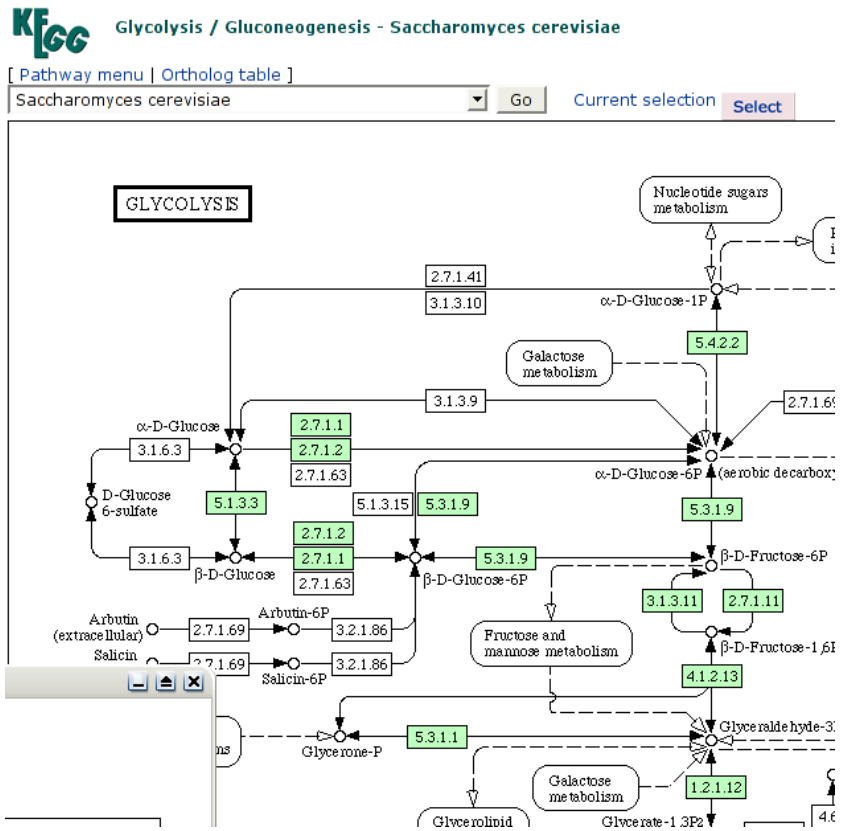
Entry	YOL086C CDS S.cerevisiae
Gene name	ADH1
Definition	Adh protein catalyzes activities for the production of certain carboxylate esters. [EC:1.1.1.1]
Orthology	KO: K00001 alcohol dehydrogenase
Pathway	PATH: sce00010 Glycolysis / Gluconeogenesis PATH: sce00071 Fatty acid metabolism PATH: sce00120 Bile acid biosynthesis PATH: sce00350 Tyrosine metabolism PATH: sce00561 Glycerolipid metabolism PATH: sce00624 1- and 2-Methylnaphthalene degradation PATH: sce00980 Metabolism of xenobiotics by cytochrome P450
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: ADH_N ADH_zinc_N adh_short DapB_N BMC PROSITE: ADH_ZINC Motif
Other DBs	SGD: S000005446 MIPS: YOL086C NCBI-GI: 6324486 NCBI-GeneID: 854068 UniProt: P00330
LinkDB	PDB All DBs
Position	XV:complement(159547..160593) Genome map
AA seq	348 aa AA seq DB search MSIPETQKGVIFYESHGKLEYKDIPVPKPKANELLINVKYSGVCHTDLHAWHGDWPLPVK LPLVGGHEGAGVVVGMGENVKGWKIGDYAGIKWLNGSCMACEYCELGNESNCPHADLSGY THDGSFQQYATADAVQAAHIPQGTDLAQVAPILCAGITVYKALKSANLMAGHWVAISGAA GGLGSLAVQYAKAMGYRVLGIDGGEGKEELFRSIGGEVFIDFTKEKDIVGAVLKATDGGA HGVINVSVSEAAIEASTRYVRANGTTVLVGMPAGAKCCSDVFNQVVKSISIVGSYVGNRA DTREALDFFARGLVKSPIKVVGLSTLPEIYEKMEKGQIVGRYVVDTSK
NT seq	1047 nt NT seq +upstream 0 nt +downstream 0 nt atgtctatcccagaaactcaaaaaggtgttatcttctacgaatcccacggtaagttcgaa

LIGAND database

- http://www.genome.ad.jp/dbget/ligand.html
- A database of enzymatic reactions
- $\approx$ 4673 enzymes, 14000 compounds and 6870 reactions
- Supports similairty searches between compounds, and reaction prediction between compunds
- Pathway computation capability, i.e. queries returning all possible pathways between two compounds.

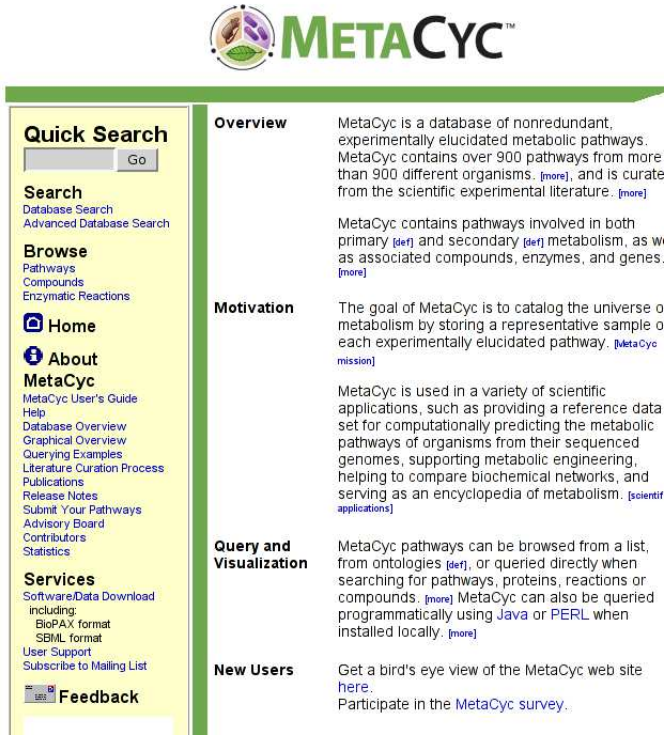
PATHWAY database

- PATHWAY database contains maps of metabolic pathways of many organimsm.
- The enzymes and compounds are clickable in the map and lead to the LIGAND and GENES database entries.
- Kegg PATHWAY maps are frequently used by biologists in their presentations



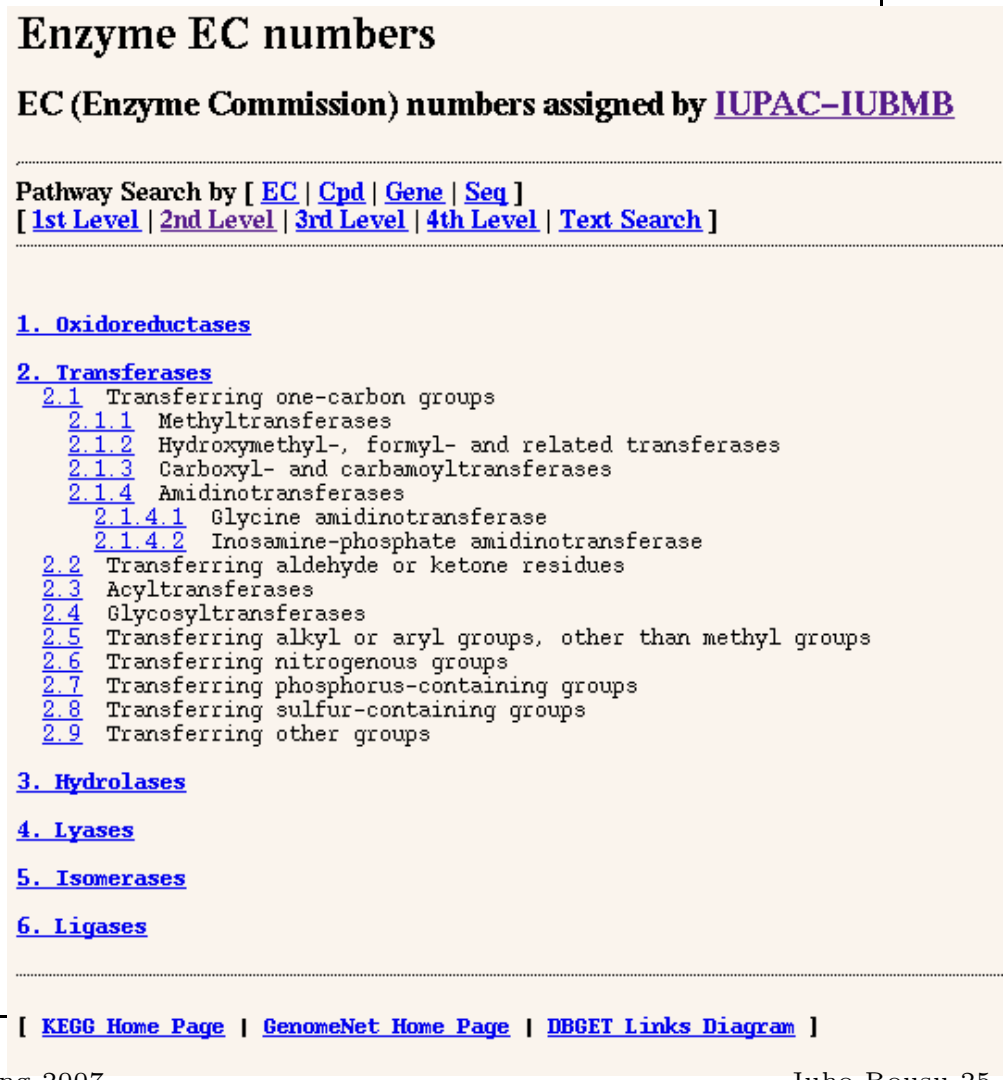
BioCyc (http://www.biocyc.org/)

- A competitor of KEGG, BioCyc is a collection of over 200 pathway/genome databases, mostly containing whole databases dedicated to certain organisms.
- One organism specific database, EcoCyc, is a highly detailed bioinformatics database on the genome and metabolic reconstruction of Escherichia Coli, including thorough descriptions of the various signaling pathways.
- MetaCyc, an encyclopedia of metabolic pathways, contains a wealth of information on metabolic reactions derived from over 600 different organisms.



Taxonomy of enzyme function: EC classification

- The Enzyme Commission (EC) classification scheme divides enzymes classes based on their function.
- The scheme has four levels, the three first level specifying the general kind of the reaction (oxidation, hydrolysis, which kind of bonds are acted on, which co-factors are used and so on.The fourth level contains individual enzymes.
- The EC scheme is the current standard for denoting enzyme function



Enzyme EC numbers

EC (Enzyme Commission) numbers assigned by IUPAC–IUBMB

Pathway Search by [EC | Cpd | Gene | Seq]
[1st Level | 2nd Level | 3rd Level | 4th Level | Text Search]

1. Oxidoreductases

2. Transferases
- 2.1 Transferring one-carbon groups
  - 2.1.1 Methyltransferases
  - 2.1.2 Hydroxymethyl-, formyl- and related transferases
  - 2.1.3 Carboxyl- and carbamoyltransferases
  - 2.1.4 Amidinotransferases
    - 2.1.4.1 Glycine amidinotransferase
    - 2.1.4.2 Inosamine-phosphate amidinotransferase
- 2.2 Transferring aldehyde or ketone residues
- 2.3 Acyltransferases
- 2.4 Glycosyltransferases
- 2.5 Transferring alkyl or aryl groups, other than methyl groups
- 2.6 Transferring nitrogenous groups
- 2.7 Transferring phosphorus-containing groups
- 2.8 Transferring sulfur-containing groups
- 2.9 Transferring other groups

3. Hydrolases

4. Lyases

5. Isomerases

6. Ligases

[KEGG Home Page | GenomeNet Home Page | DBGET Links Diagram]

Metabolic reconstruction workflow

- Start from a sequenced genome of an organism
- Obtain annotations for ORFs via sequence homology and pick those with annotated enzymatic reaction (EC class)
- Pick reactions that have multiple polypeptides associated and decide if they correspond to protein complexes or isozymes. (If available protein-protein interaction data could be used here)
- Fill in gaps in the metabolism: metabolites that cannot be produced by the reactions although they are empirically observed. Here sources other than sequence homology data are useful (phylogenetic profiling, metabolite concentrations, literature)

Constructing whole-genome metabolic reconstructions is a non-trivial exercise: each such reconstruction is typically worth a publication.

Genome annotation

Since few organism have extensive biochemical information available, reconstruction relies heavily on an annotated genome sequence.

Traditional techniques for annotation include

- Experimental methods: gene cloning or knockout and observation of changes in the phenotype
- Sequence homology: comparing the sequence to genes with known function in other organisms

Genome annotation

More recent techniques include:

- Protein-protein interaction data: if two enzymes are known to form a complex, it is likely that they together catalyze the same or adjacent reactions in the metabolic network
- Correlated mRNA expression: an enzyme that has similar expression profile (over a set of conditions) might have a similar function
- Phylogenetic profiling: based on the assumption that proteins that function together in a pathway or structural complex are likely to evolve in a correlated fashion. Functionally linked proteins tend to same similar occurrence profiles accross species.

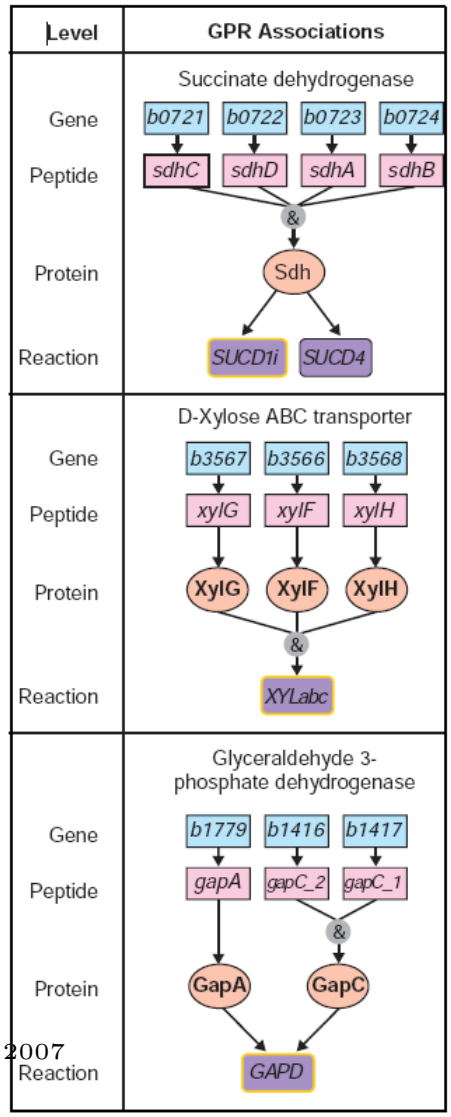
Finding similar sequences

- Alignment: Use the BLAST or FASTA family of methods to align ORFs with the sequences of known enzymes function contained in enzyme databases such as IntEnz (www.ebi.ac.uk/intenz) or Uni-Prot (www.expasy.ch). Function can be reliably assigned for sequences that are evolutionarily close but it is not reliable for distant homologs.
- Conserved motifs: find groups of conserved amino acids, 'motifs' that are stored in a database such as PROSITE (www.expasy.ch/prosite/). The idea is to define certain conserved amino acid patterns that are related to function, e.g. they are residues close to the active site. These methods are more sensitive for function determination than alignment techniques.

Gene-protein-reaction interactions

- Peptides from several genes may be used to encode single protein which may catalyze several reactions (top picture)
- Several proteins may form a complex to catalyze a single reaction (middle picture)
- Different genes may encode isozymes (proteins with identical function) that catalyze the same reaction (bottom picture)

(picture taken from Reed, Vo, Schilling and Palsson. Genome Biology 4, 2003)



Pathway Tools (http://bioinformatics.ai.sri.com/ptools/)

- One of the few software packages that assists in the construction of pathway/genome databases such as EcoCyc.
- PathoLogic tool takes an annotated genome for an organism and infers probable metabolic pathways to produce a new pathway/genome database.
- This can be followed by application of the Pathway Hole Filler, which predicts likely genes to fill "holes" (missing steps) in predicted pathways.
- In addition there are Navigation and editing tools by which the user can visualize, analyze, access and update the database.
- The rationale: Pathway Tools give a rapid first blueprint of the metabolic network that can be iteratively refined.