Systems Biology

(1) Introduction

David Gilbert

Bioinformatics Research Centre

www.brc.dcs.gla.ac.uk

Department of Computing Science, University of Glasgow

Module outline

- ‘Putting it all together’ - Systems Biology
- Motivation
- Biological background
- Modelling
  - Network Models
  - Data models
- Analysis:
  - Static
  - Dynamic
- Standardisation (sbml & sbw)
- Technologies
- Current approaches
- Systems robustness

Admin

- Term 2; 2007-2008
  - Friday 22nd February, Graham Kerr Seminar Room 10:30 till 11:30,
  - Lab room 101 Davidson Building 11:00 till 12:00 or 12:00 till 1:00
  - LAB TIME STILL TO BE CONFIRMED FOR THIS SESSION

  - Wednesday 27th February, Boyd Orr Room 831, 10:00 till 11:00
  - Lab room 101 Davidson Building, 2:00 till 3:00

  - Thursday 28th February, Graham Kerr Library, 10:00 till 11:00
  - Lab 1 room 101 Davidson Building, 11:00 till 12:00

  - Friday 29th February, Graham Kerr Seminar Room, 10:00 till 11:00
  - Lab room 101 Davidson Building, 11:00 till 12:00

- Module information, resources & reading list: www.brc.dcs.gla.ac.uk/~drg/courses/sysbiomres

- Assessment: 1 Coursework + Exam question

- Summer project - optional

- Course staff
  - Lecturer: Professor David Gilbert
  - Demonstrator: Ms Xu Gu
- Additional things: www.brc.dcs.gla.ac.uk/seminars

Resources

- DRG's handouts
- www.brc.dcs.gla.ac.uk/~drg/bioinformatics/resources.html
- www.ebi.ac.uk/2can
  - Bioinformatics educational resource at the EBI
- International Society for Computational Biology: www.iscb.org
  - very good rates for students, and you get on-line access to the Journal of Bioinformatics.
- Broder S, Venter J C, Whole genomes: the foundation of new biology and medicine, Curr Opin Biotechnol. 2000 Dec;11(6):581-5.
- Kitano H. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. Curr Genet. 2002 Apr;41(1):1-10.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002 Oct 25;298(5594):824-7.
- Yuri Lazebnick. Can a biologist fix a radio? - Or, What I learned while studying Apoptosis. Cancer Cell september 2002 vol 2 179-182.
- Post Genome Informatics Kanehisa. Publisher OUP. Year 2000. Isbn 0198503261. Category background

Lecture outline

- ‘Putting it all together’ - Systems Biology
- Motivation
- Technological drivers
- Some biological background
- Introduction to some (systems biology) databases

Motivation

- The amount and variety of biological data now available, together with techniques developed so far have enabled research in Bioinformatics to move beyond the study of individual biological components (genes, proteins etc) – albeit in a genome-wide context – to attempt to study how individual parts cooperate in their operation.

- Bioinformatics as a scientific activity has now moved closer to the area of Systems Biology which seeks to integrate biological data as an attempt to understand how biological systems function.

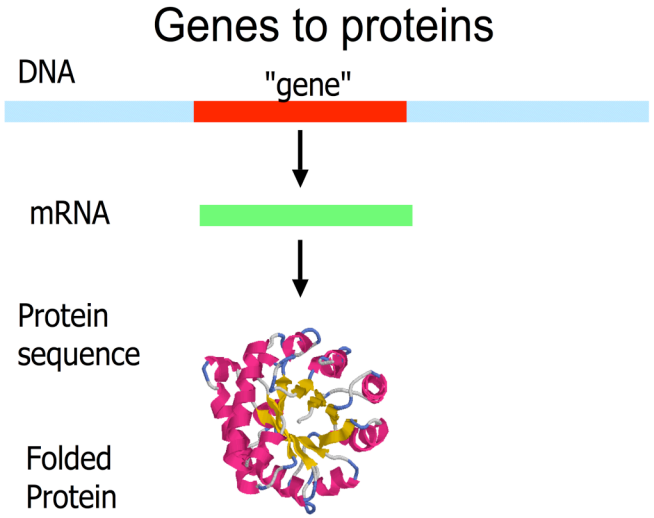
- By studying the relationships and interactions between various parts of a biological system it is hoped that an understandable model of the whole system can be developed.

Central Dogma

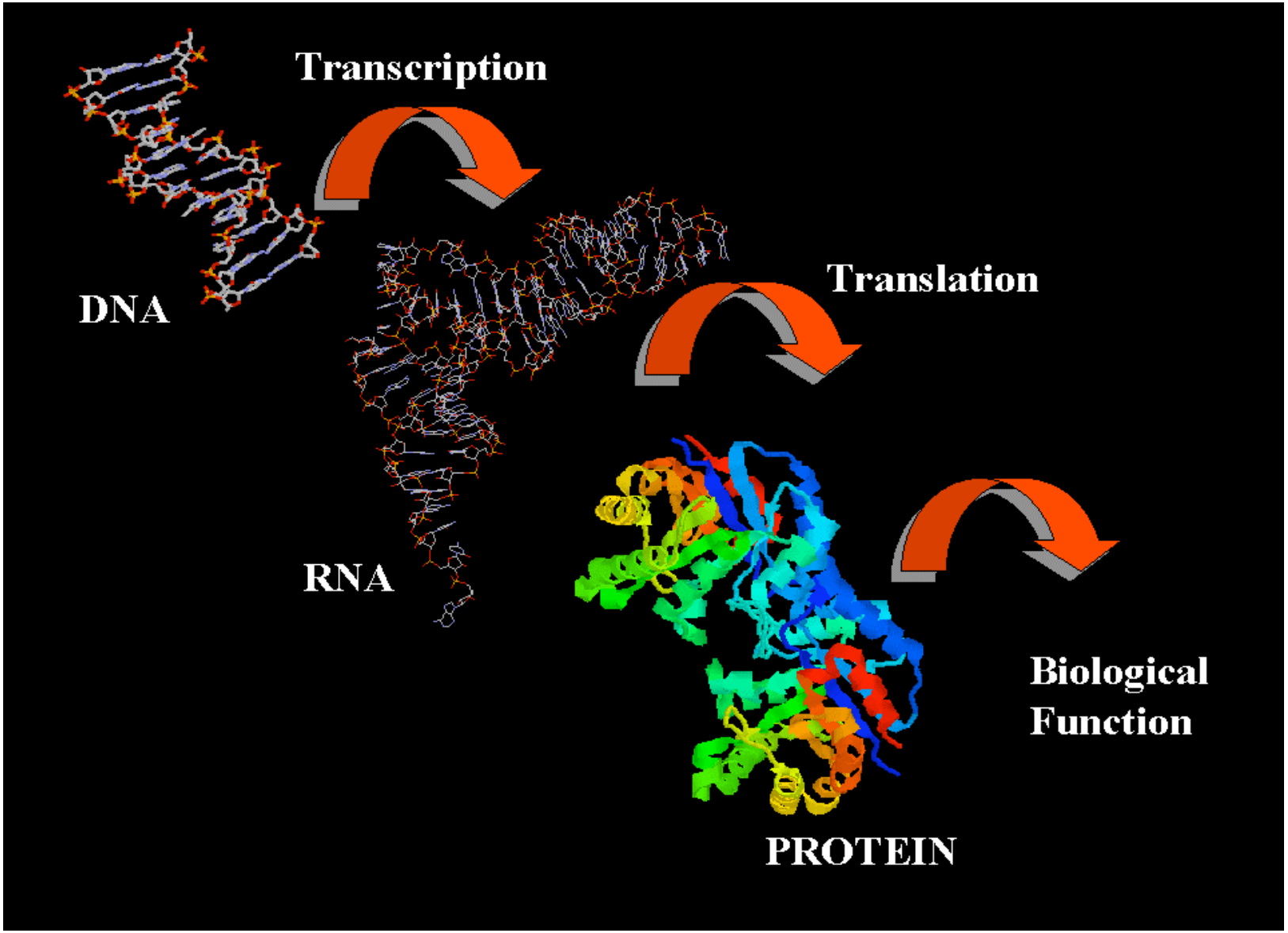
- The central dogma of information flow in biology essentially states that the sequence of amino acids making up a protein and hence its structure (folded state) and thus its function, is determined by transcription from DNA via RNA.

- “This states that once ‘information’ has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein.”
  Francis Crick, On Protein Synthesis, in Symp. Soc. Exp. Biol. XII, 138-167 (1958)

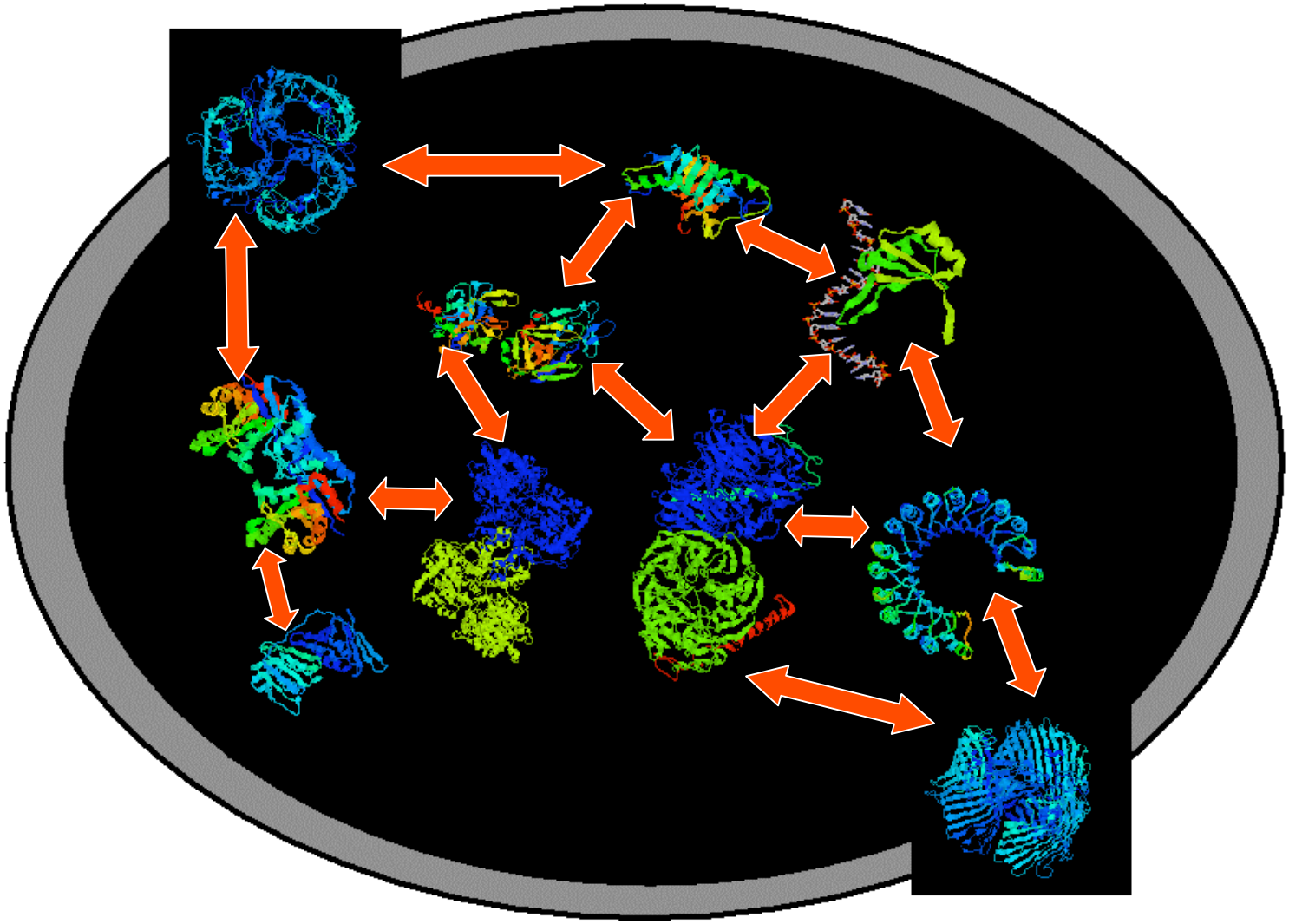
- (Nothing said explicitly about transfer from RNA to DNA)



Behaviour of the gene …



… their interaction



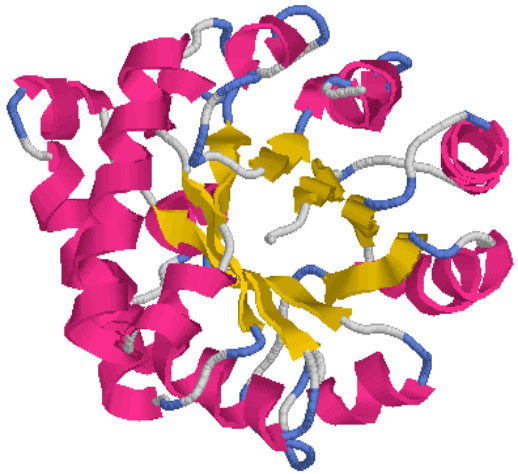
Genes to systems

DNA

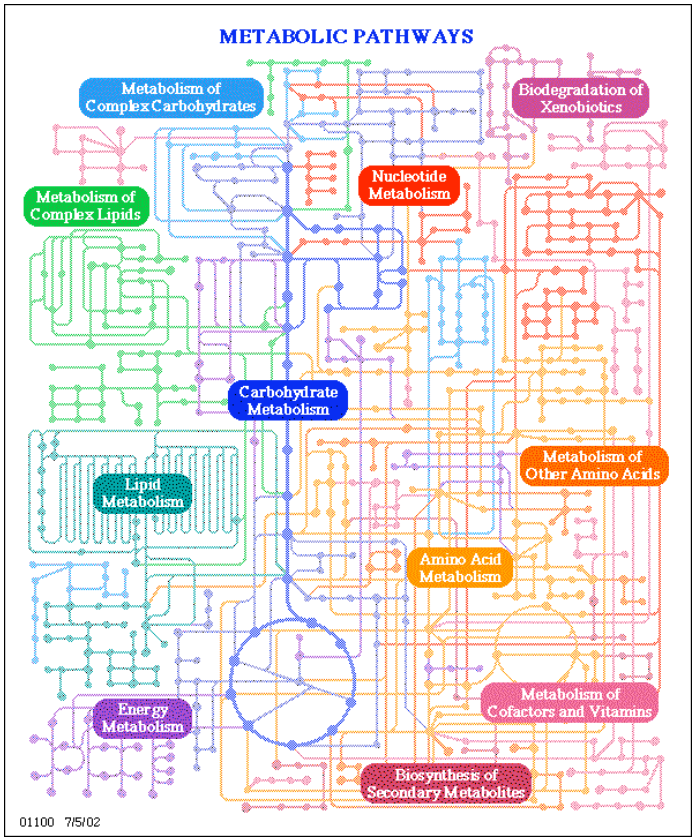
"gene"

mRNA

Protein sequence



Folded Protein

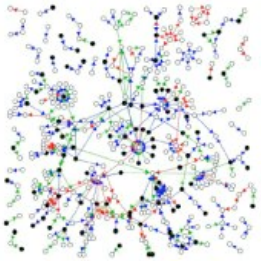


Terminology: Pathways or Networks?

- Pathways implies ‘paths’ - sequences of objects
- Networks - more complex connectivity
- Both are represented by ***graphs***
- Networks: generic; Pathways: specific (?)
  - ‘Signal transduction networks’
  - ‘The ERK signal transduction pathway’

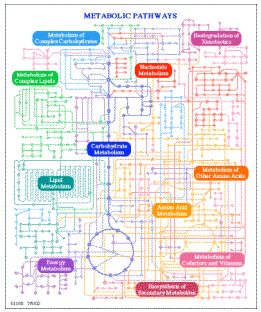
Networks

- Gene regulation



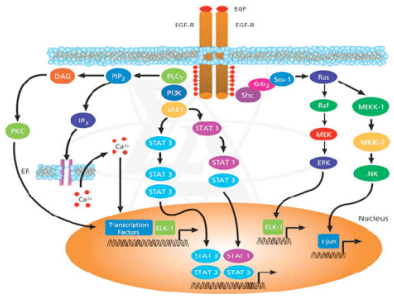
- Protein-protein interaction

- Metabolic

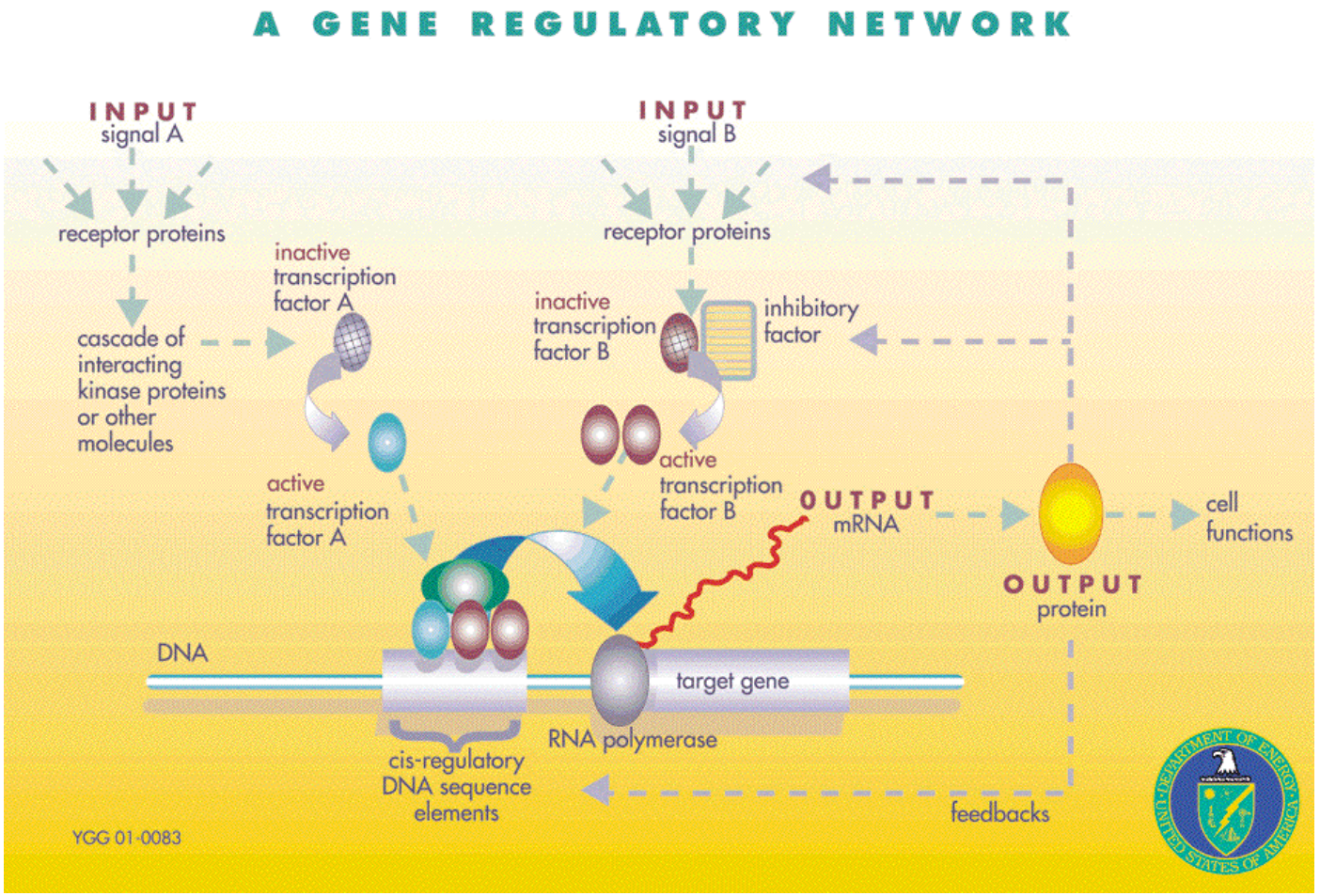


- Developmental

- Signalling

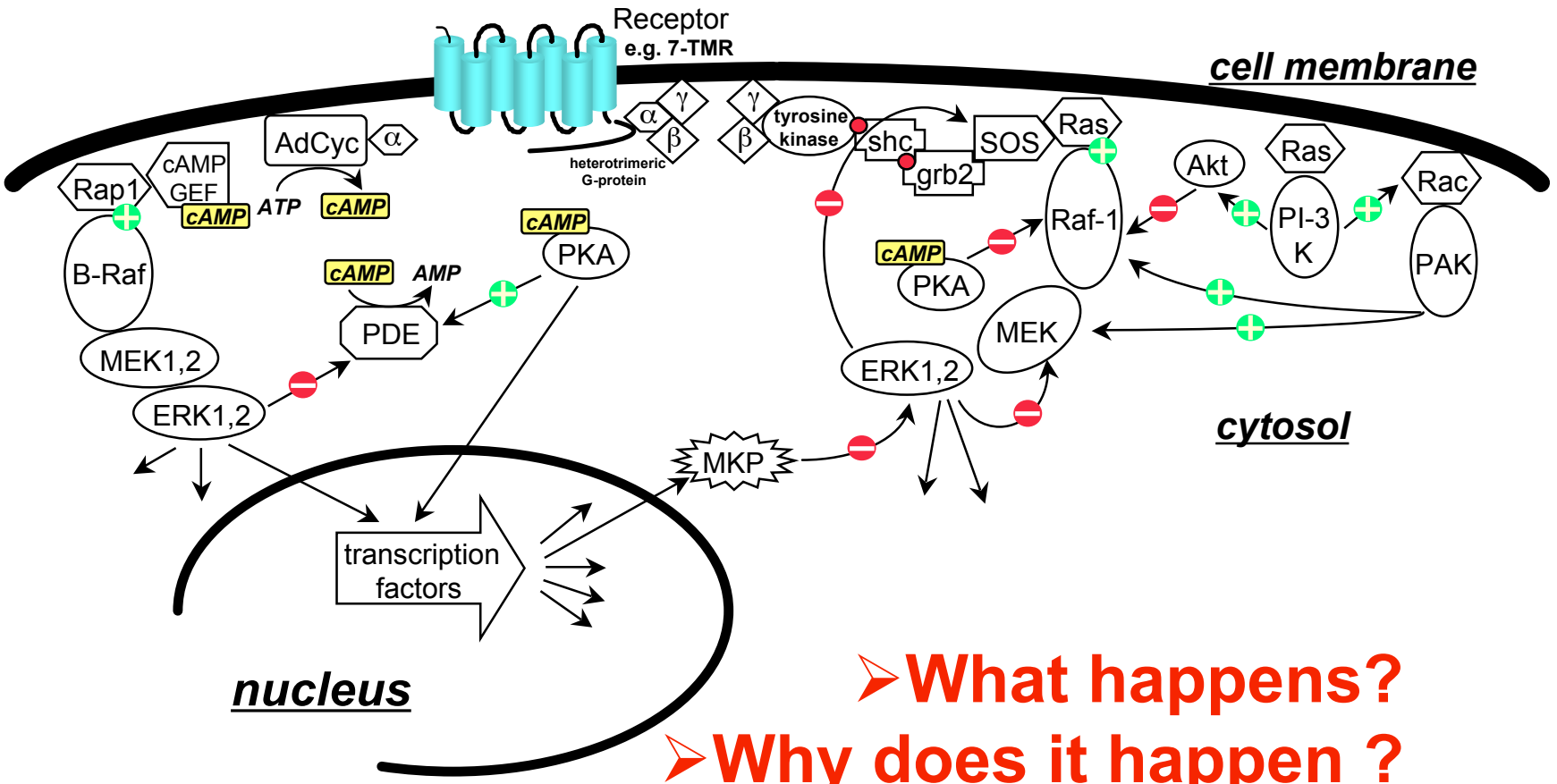


Gene regulation



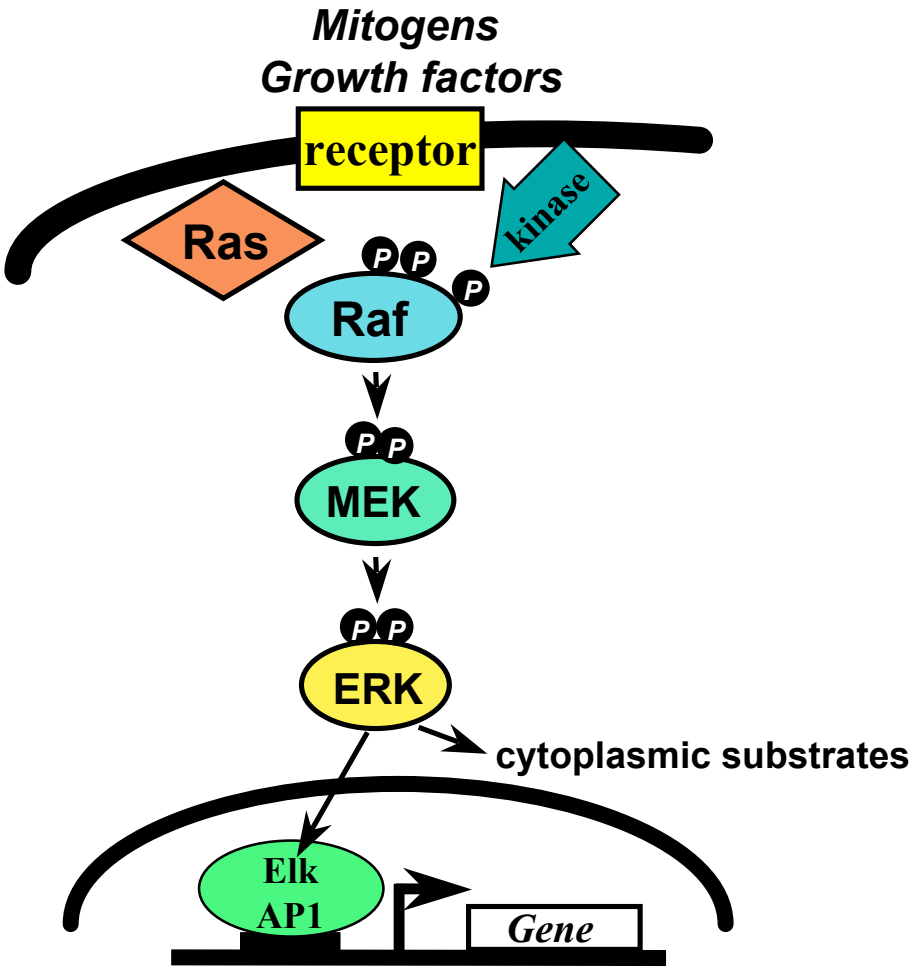
Biochemical networks

We can describe the general topology and single biochemical steps.
However, we do not understand the network function as a whole.

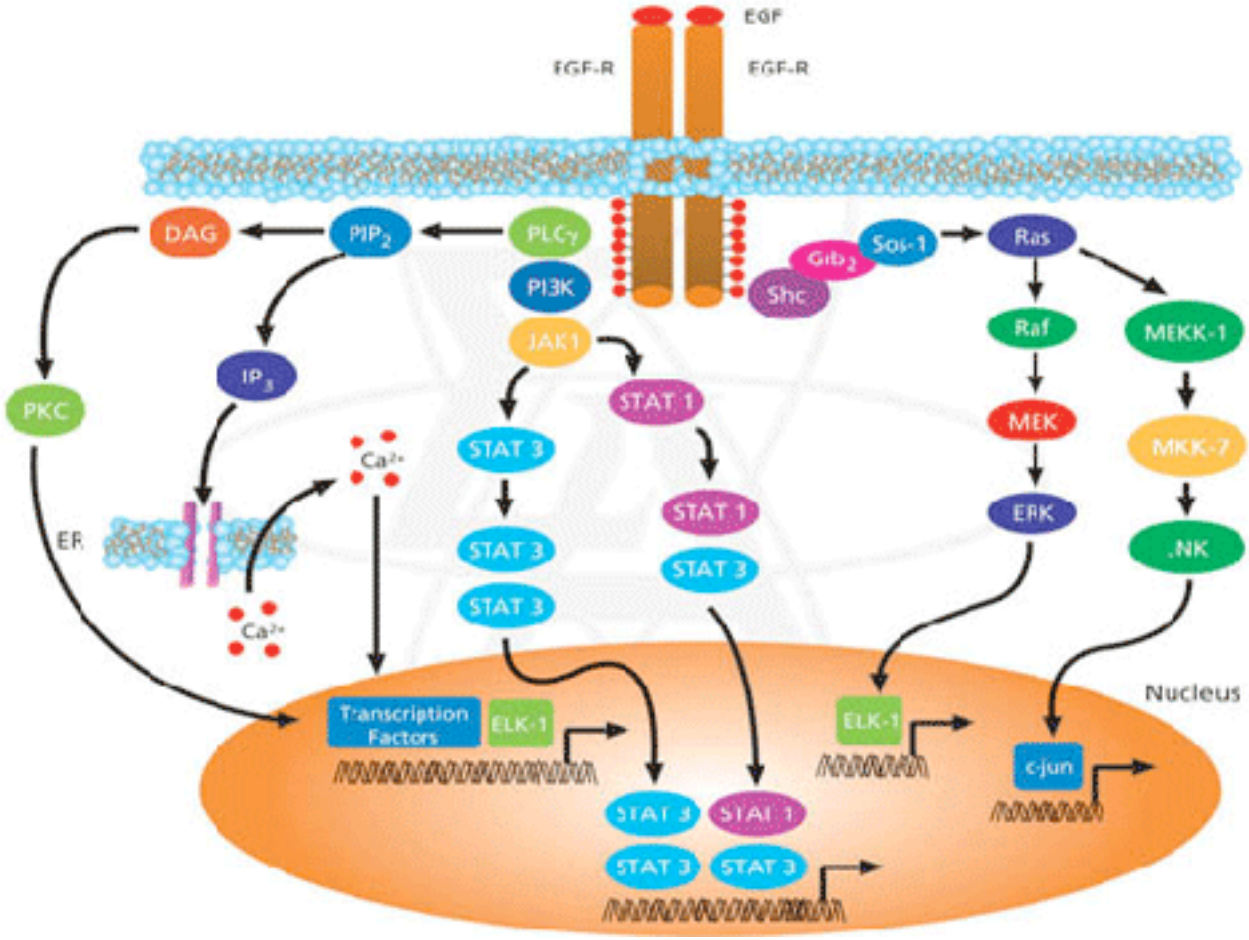


- **What happens?**
- **Why does it happen ?**
- **How is specificity achieved?**

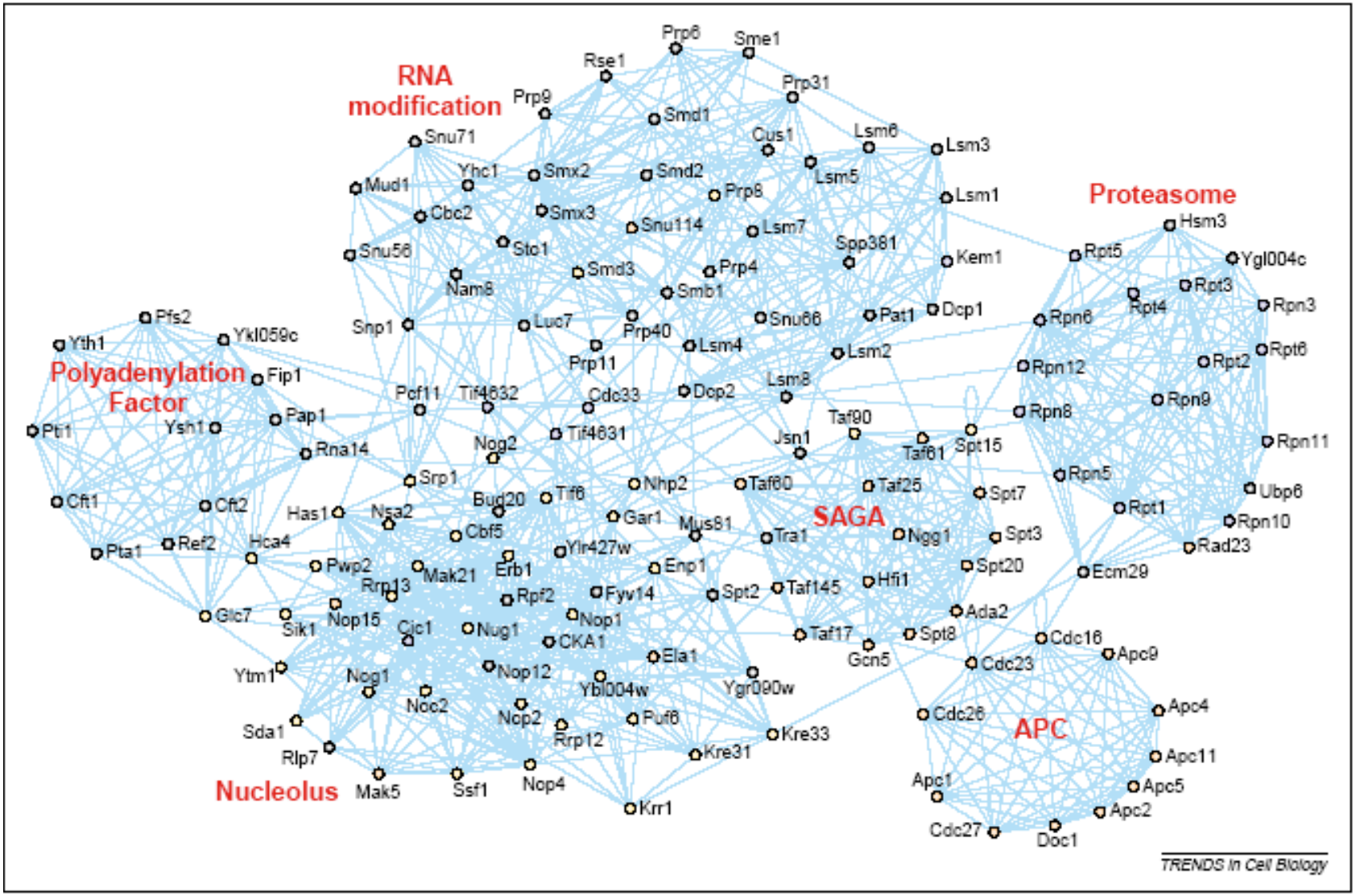
ERK signalling pathway



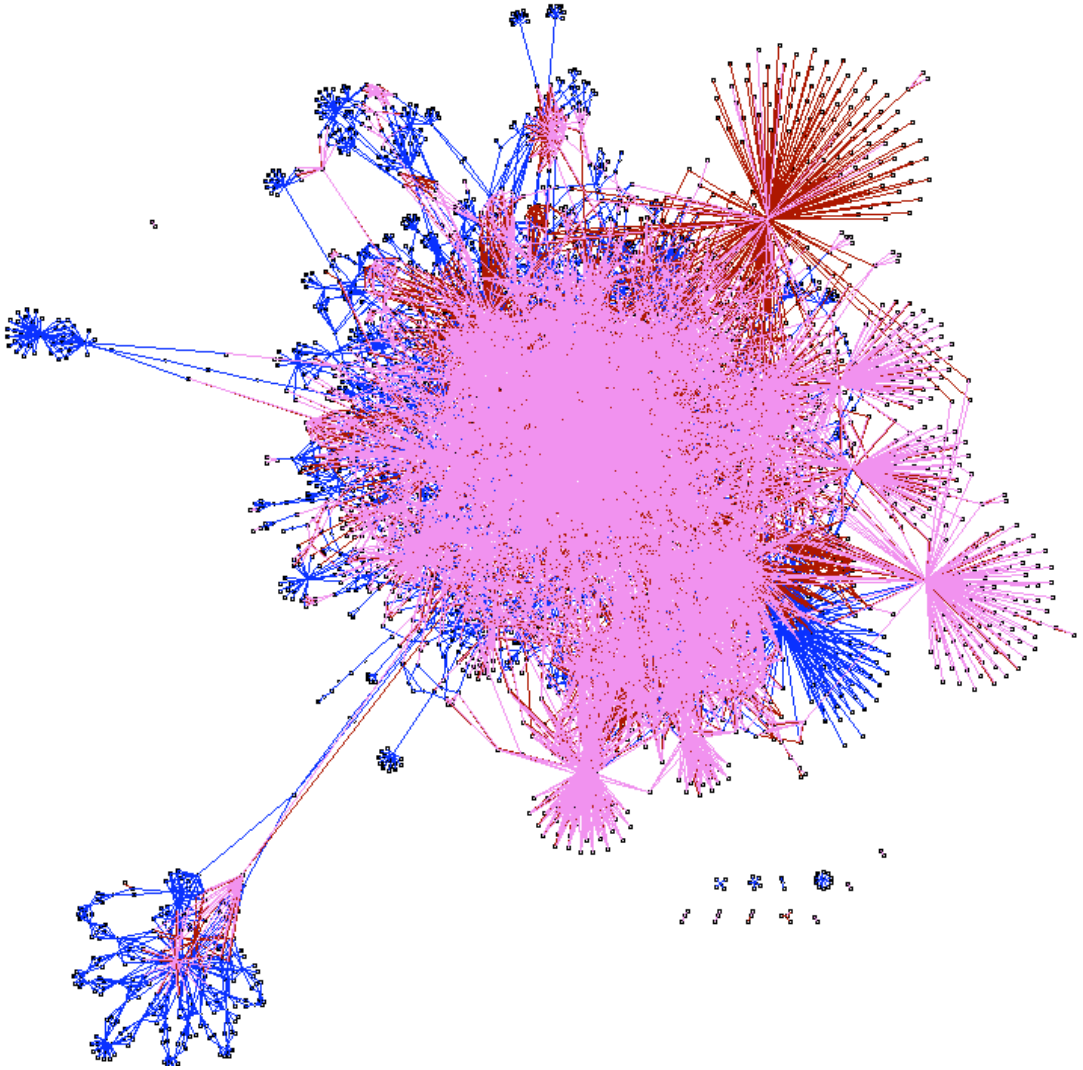
Signal Transduction



Protein-protein interaction in yeast

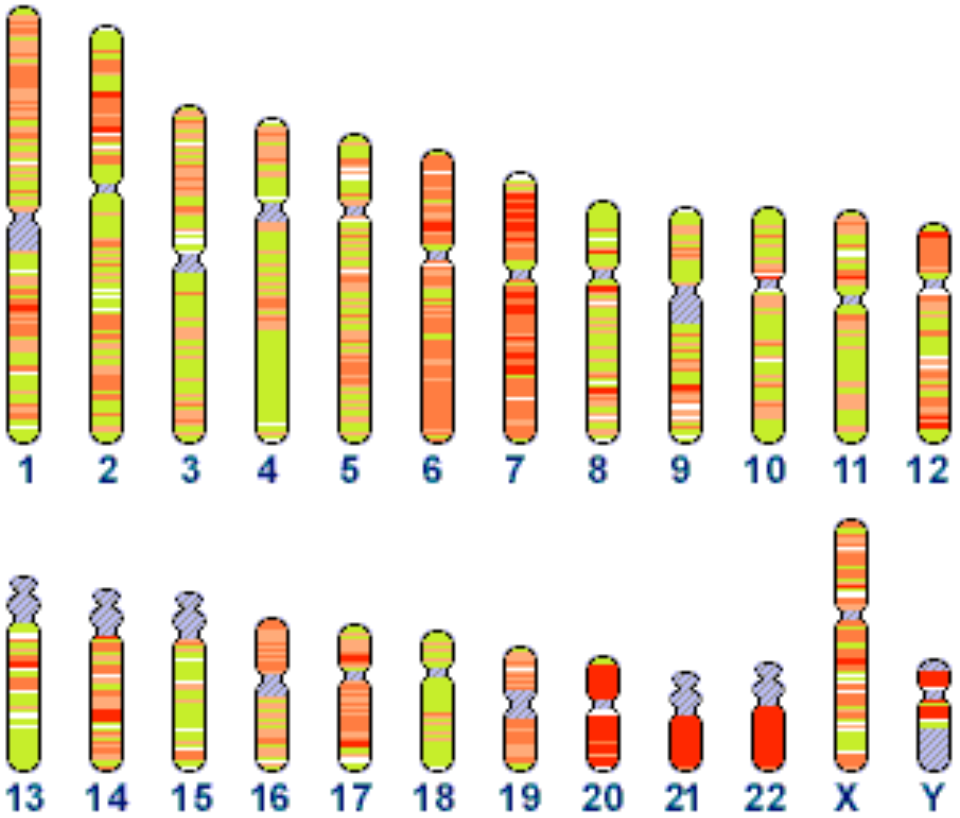


Protein-protein interaction



Developmental pathway

Human Genome

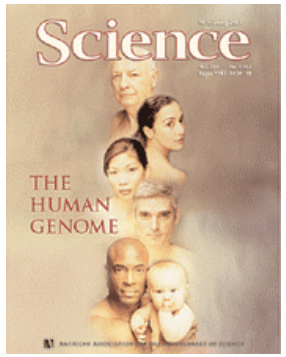


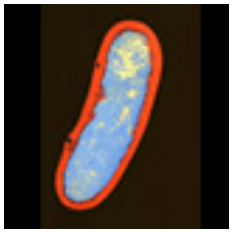
After Human Genome Project (HGP)

The Seven (7) ways the HGP has impacted biology (Hood, 2002)

- Biology is an informational science
- Discovery science enhances global analyses
- A generic parts list provides a toolbox of genetic elements for systems analyses
- High-throughput platforms permit one to carry out global analyses at the DNA, RNA, and protein levels
- Computational, Mathematical, and Statistical tools are essential for handling the explosion of biological information
- Model organisms are Rosetta stones for deciphering biological information
- Comparative genomics is a key to deciphering biological complexity

Each of these seven changes has catalysed the emergence of systems biology

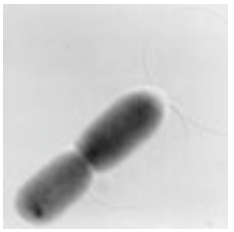




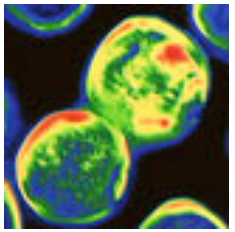
Yersinia pestis

Arabidopsis thaliana

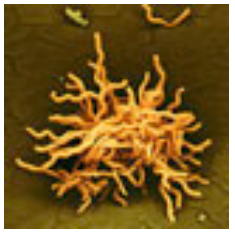
Buchnerasp. APS



Aquifex aeolicus



Archaeoglobus fulgidus

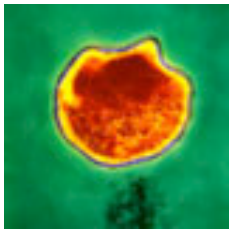


Borrelia burgorferi

Mycobacterium tuberculosis

Caenorhabitis elegans

Campylobacter jejuni

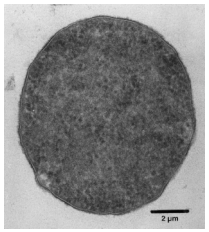


Chlamydia pneumoniae

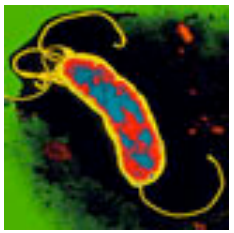
Vibrio cholerae

Drosophila melanogaster

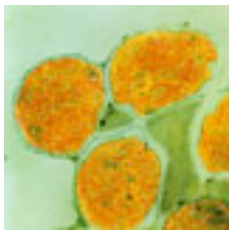
Escherichia coli



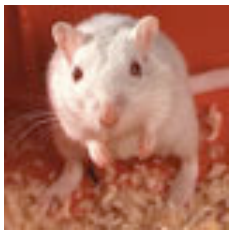
Thermoplasma acidophilum



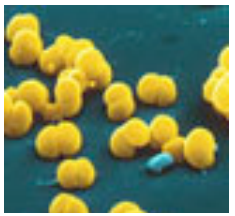
Helicobacter pylori



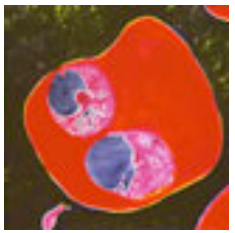
Mycobacterium leprae



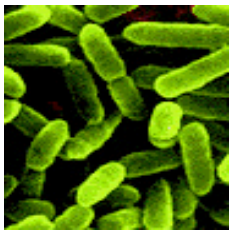
mouse



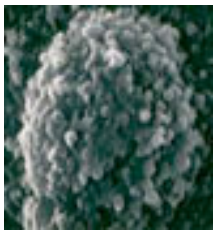
Neisseria meningitidis Z2491



Plasmodium falciparum

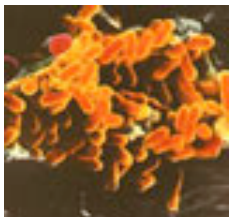


Pseudomonas aeruginosa

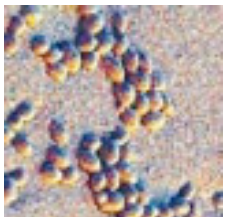


Ureaplasma urealyticum

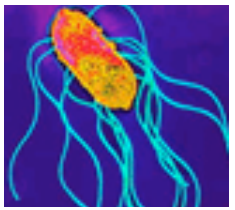
rat



Rickettsia prowazekii

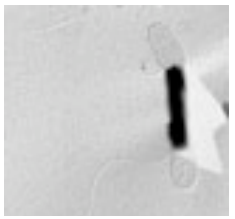


Saccharomyces cerevisiae

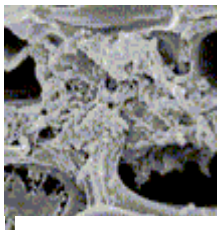


Salmonella enterica

Bacillus subtilis



Thermotoga maritima



Xylella fastidiosa

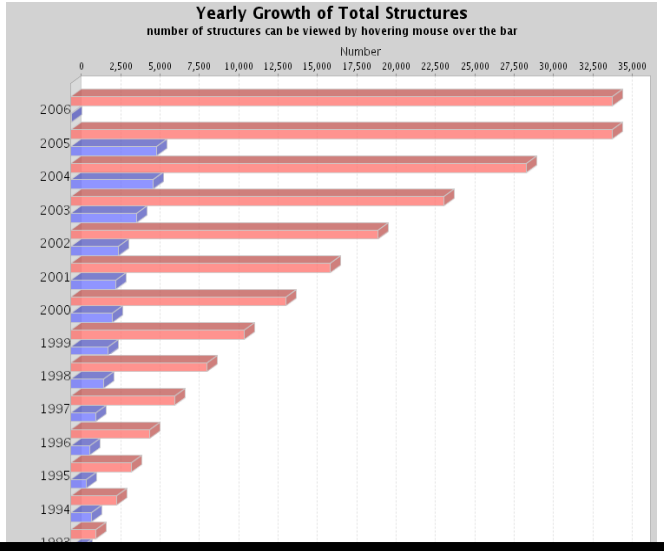
Whole genomes

- Our genomic DNA sequence provides a unique glimpse of the provenance and evolution of our species, the migration of peoples, and the causation of disease.
- Understanding the genome may help resolve previously unanswerable questions, including perhaps which human characteristics are innate or acquired.
- Such an understanding will make it possible to study how genomic DNA sequence varies among populations and among individuals, including the role of such variation in the pathogenesis of important illnesses and responses to pharmaceuticals.
- The study of the genome and the associated proteomics of free-living organisms will eventually make it possible to localize and annotate every human gene, as well as the regulatory elements that control the timing, organ-site specificity, extent of gene expression, protein levels, and post- translational modifications.
- For any given physiological process, we will have a new paradigm for addressing its evolution, development, function, and mechanism.

- Broder S, Venter J C, Whole genomes: the foundation of new biology and medicine, Curr Opin Biotechnol. 2000 Dec;11(6):581-5.

Database Growth

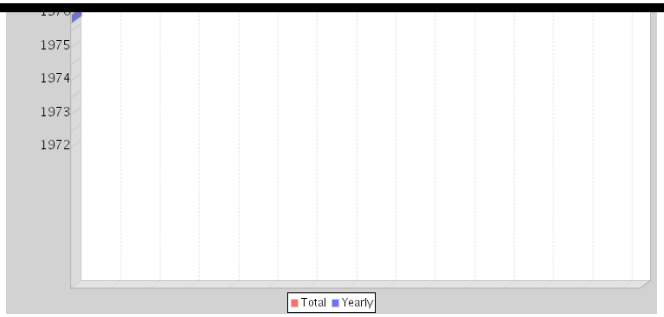
EMBL - sequences



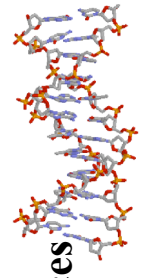
Data deluge is an URBAN MYTH???

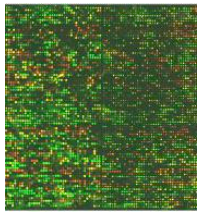
D

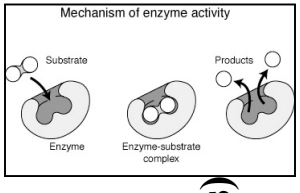
- Nucleotide Seq (GenBank, EMBL, …)
- Biochemical Pathways (KEGG, WIT…)
- Molecular Classifications (SCOP, CATH,…)
- Motif Libraries (PROSITE, Blocks, …)

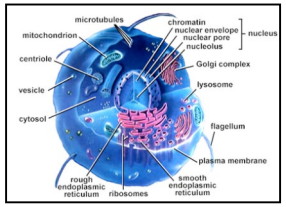


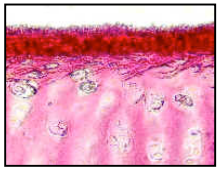
The Complexity of Biological Data

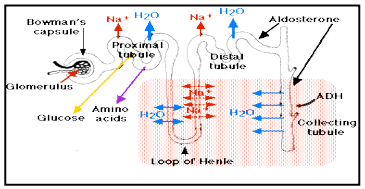












Nucleotide sequences

Nucleotide structures

Gene expressions

Protein Structures

Protein functions

Protein-protein interaction (pathways)

Cell

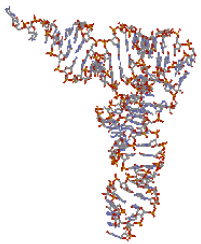
Cell signalling

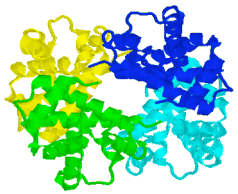
Tissues

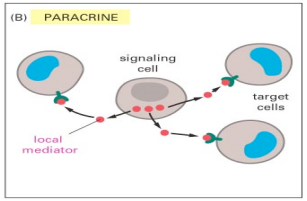
Organs

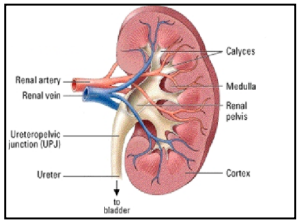
Physiology

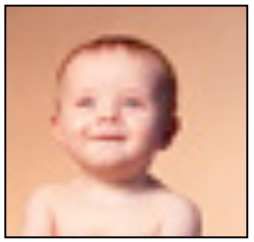
Organisms

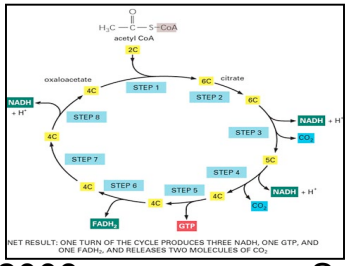




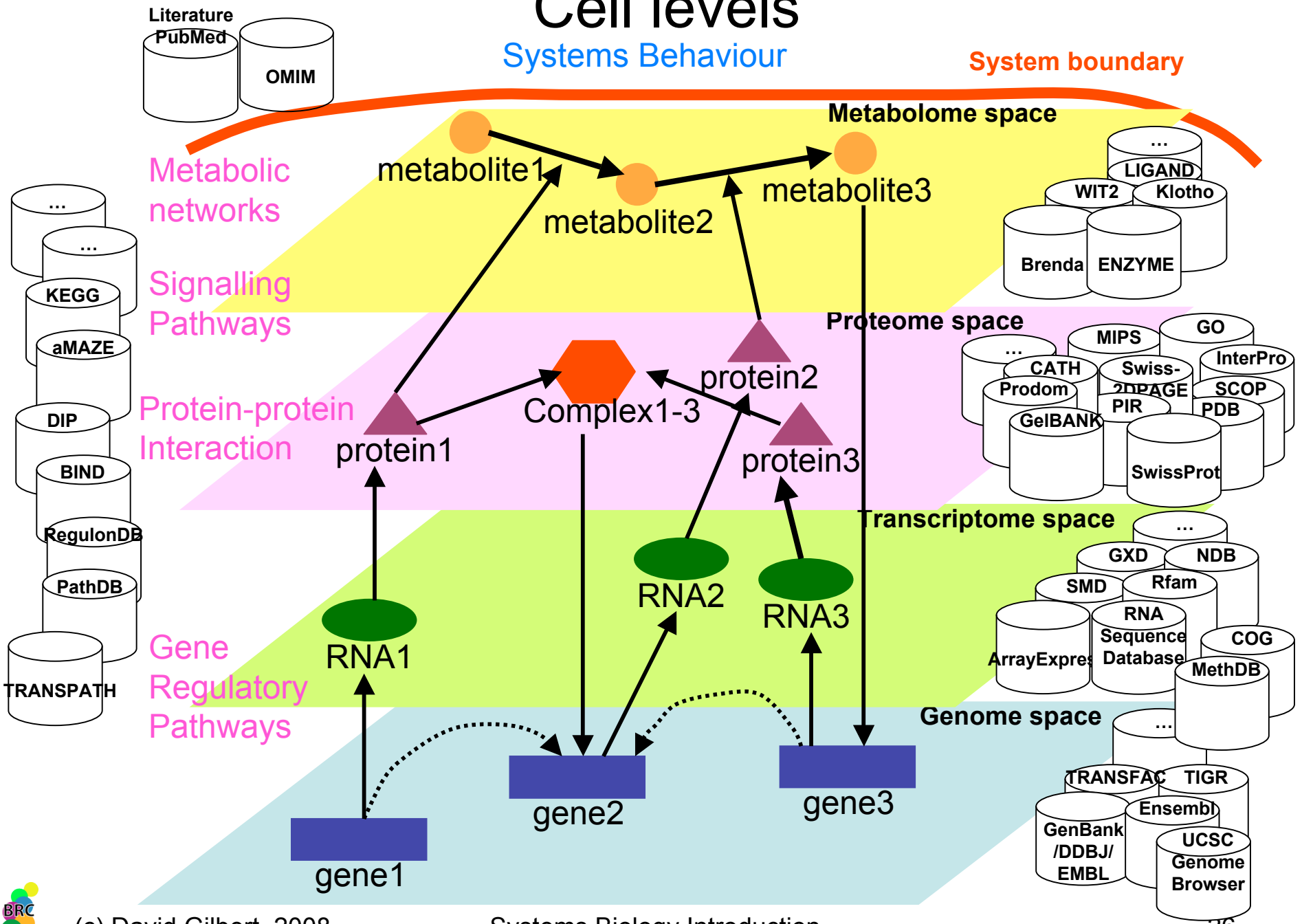








Cell levels


Systems Behaviour
System boundary
Literature
PubMed
OMIM
Metabolic networks
Signalling Pathways
Protein-protein Interaction
Gene Regulatory Pathways
Metabolome space
Proteome space
Transcriptome space
Genome space
metabolite1
metabolite2
metabolite3
protein1
protein2
protein3
Complex1-3
RNA1
RNA2
RNA3
gene1
gene2
gene3
...
LIGAND
WIT2
Klotho
Brenda
ENZYME
KEGG
aMAZE
DIP
BIND
RegulonDB
PathDB
TRANSPATH
MIPS
GO
InterPro
CATH
Swiss-2DPAGE
SCOP
Prodom
PIR
PDB
GelBANK
SwissProt
GXD
NDB
SMD
Rfam
RNA Sequence Database
ArrayExpres
COG
MethDB
TRANSFAC
TIGR
Ensembl
GenBank /DDBJ/ EMBL
UCSC Genome Browser

Rise in system-oriented approaches in genetics and molecular biology

- With the ever-increasing flow of high-throughput gene expression, protein interaction and genome sequence data, researchers gradually approach a system-level understanding of cells and even multi-cellular organisms.
- Systems biology is an emerging field that enables us to achieve in-depth understanding at the system level.
- For this, we need to establish methodologies and techniques that enable us to understand biological systems as systems, which means to understand:
  (1) the structure of the system, such as gene/metabolic/signal transduction networks and physical structures,
  (2) the dynamics of such systems,
  (3) methods to control systems, and
  (4) methods to design and modify systems to generate desired properties.
- However, the meaning of "system-level understanding" is still ambiguous. This paper reviews the current status of the field and outlines future research directions and issues that need to be addressed.

- Kitano H. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. Curr Genet. 2002 Apr;41(1):1-10.

Systems biology – some definitions

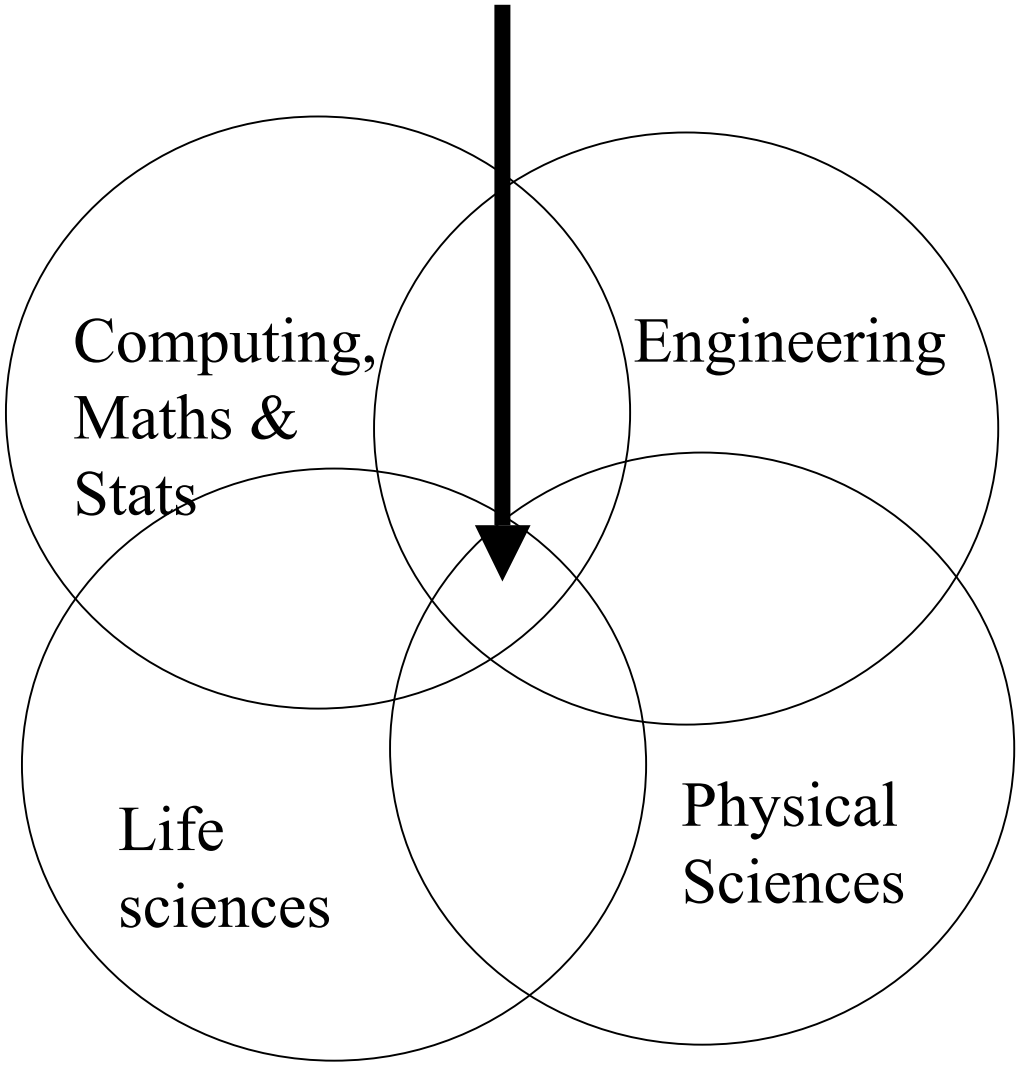
- Systems biology is the study of all the elements in a biological system (all genes, mRNAs, proteins, etc) and their relationships one to another in response to perturbations.
- Systems approaches attempt to study the behaviour of all of the elements in a system and relate these behaviours to the systems or emergent properties

A Framework for Systems Biology

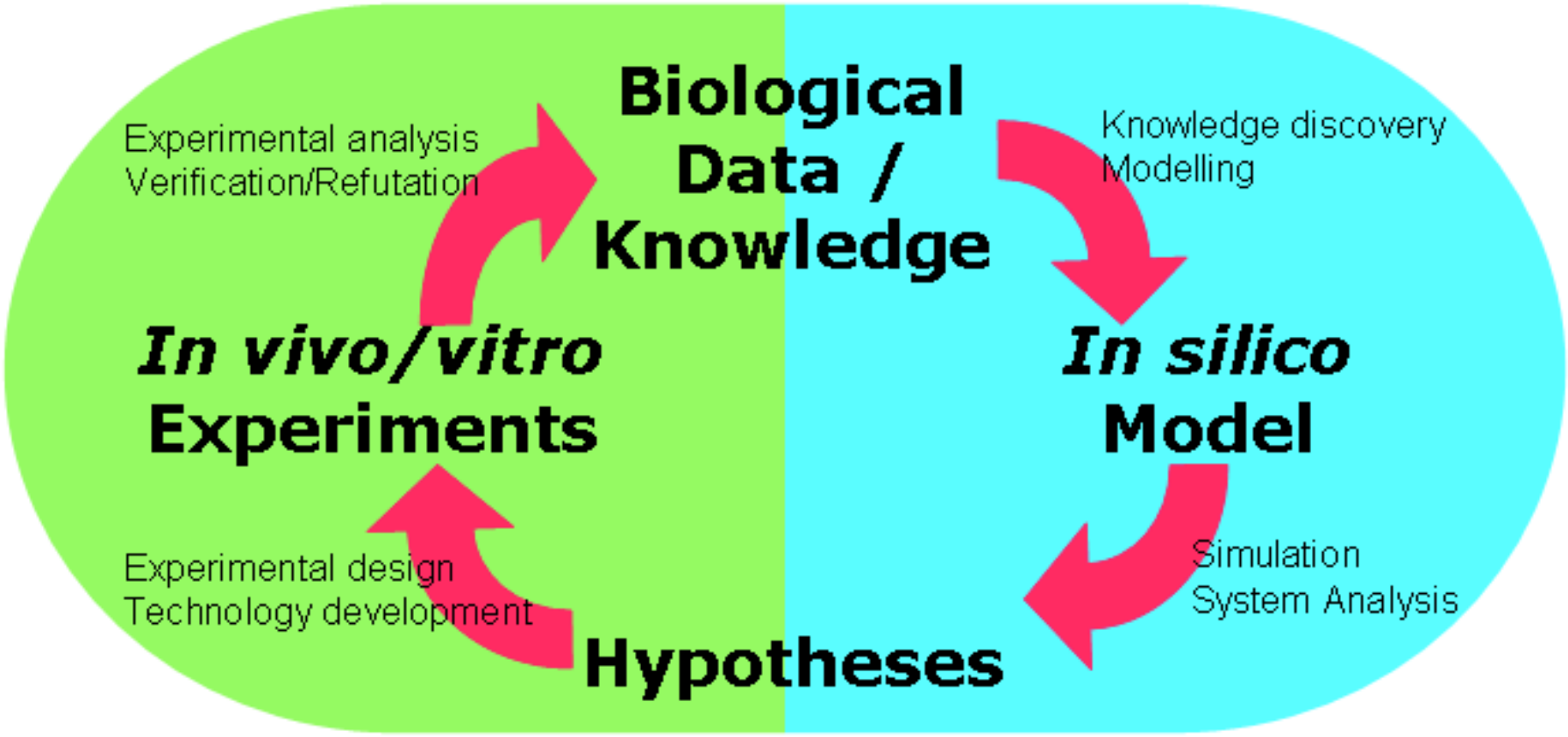
(Ideker, Galitski & Hood, 2001)

- Define all of the components of the system
- Systematically perturb and monitor components of the system
- Reconcile the experimentally observed responses with those predicted by the model
- Design and perform new perturbation experiments to distinguish between multiple or competing model hypotheses

Systems Biology in context - a new discipline?



Bio-Discovery Science



Kitano's SB Challenges

- Methods to identify Network structure (& parameters)
- Methods to quantify the Dynamics of such structure
- Methods to Control Systems
- Methods to Design and modify systems for desired properties

What does it take to carry out Systems Biology?

- Cross-disciplinary team of biologists, computer scientists, chemists, engineers, mathematicians, … who understand each other!
- Integrated teamwork to execute the hypothesis driven, iterative and integrative cycles of systems biology.
- High-throughput facilities for genomics and proteomics technology & the expertise to keep these facilities at the leading-edge of technology development.
- Integration of effort with academia, primarily to encompass intriguing new areas of biology and medicine, & with industry for emerging technologies and support.

Model

- Abstract model: a theoretical construct that represents something, with a set of variables and a set of logical and quantitative relationships between them.
- Such models are constructed to enable reasoning within an idealized logical framework about these processes and are an important component of scientific theories
- The model may make explicit assumptions that are known to be false (or incomplete) in some detail.
- Such assumptions may be justified on the grounds that they simplify the model while, at the same time, allowing the production of acceptably accurate solutions

Simulation

- A simulation is an imitation of some real thing, state of affairs, or process. The act of simulating something generally entails representing certain key characteristics or behaviors of a selected physical or abstract system.

- A computer simulation is an attempt to model a real-life situation on a computer so that it can be studied to see how the system works.
  - By changing variables, predictions may be made about the behaviour of the system.

Analysis & Reasoning

- A model may be used to permit (automated) reasoning about the object / system modelled.

- *Predictive* modelling: the use of a model to predict the behaviour of a system.
  - E.g. predict the effect of drugs on an organism
  - E.g. predict the effect of an inhibitor on a pathway

The Silicon Cell - ultimate goal?

- http://www.bio.vu.nl/hwconf/Silicon/
- The long-term goal of the Silicon Cell (SiC) Consortium is the computation of Life at the cellular level on the basis of the complete genomic, transcriptomic, proteomic, metabolomic and cell-physiomic information that will become available in the forthcoming years.
- 3 major challenges, i.e. networks, space and time; systematic handling of data and results.

- Key objectives

(i) Computational models of catabolism, signal transduction, gene-expression regulation, coupling between supramolecular structures and fluxes, and biochemical cycling.

(ii) Model integration to calculate system properties for two real cells (E. coli and S. cerevisiae).

(iii) Demonstration of the cellular bioinformatics approach: calculating without fitting.

(iv) Methodology for modularisation to accurate mesoscopic descriptions.

(v) Visualisation, systematic data access and a www resource for two real living cells.

- Approach: focus on three different, but interconnected dimensions of cell functioning,

(i) the 'chemical and information dimension': networks of biochemical reactions and their regulation,

(ii) space: gradients and dynamic structures in signal transduction and gene expression (chromatin), and

(iii) biological time: coherent glycolytic and cell-cycle oscillations.

The Silicon Cell - ultimate goal?

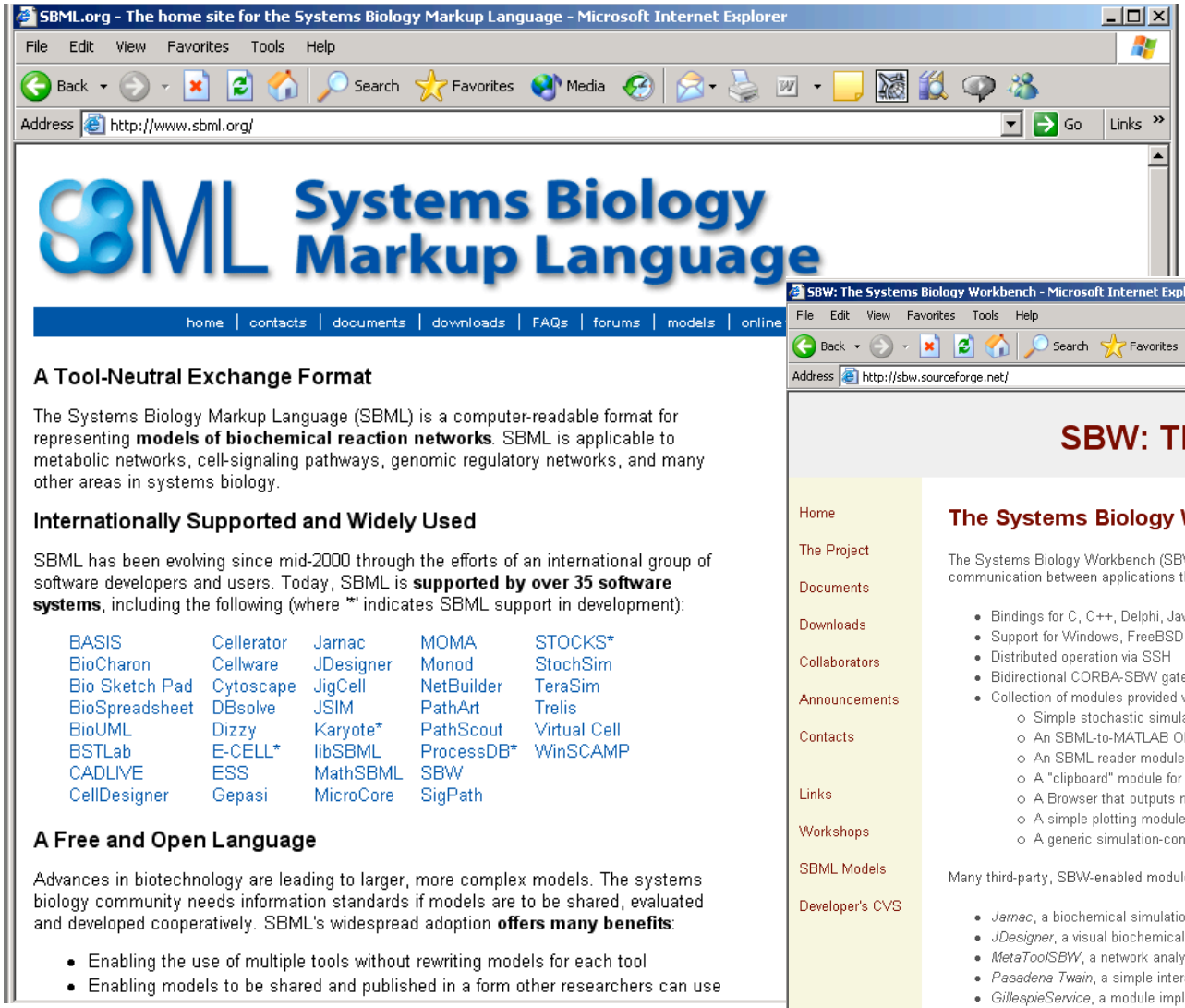
- The specific cases connect to the glucose entry into S. cerevisiae en E. coli, subsequent carbon and energy metabolism, up to their coupling to examples of signal transduction, gene expression regulation and cell-cycle. This work will be coupled to biological experiments.

- Different from most traditional modelling methods, this programme will always start from real experimental data that stem from molecular biology, biochemistry, physics and chemistry. Rather than aiming at an understanding of principles of function (as would be done by theoretical biology or physics) we shall 'merely' compute the implications of the molecular data for system behavior.

- The present program is among the few that integrate all relevant information from various scientific fields (e.g. molecular biology, biochemistry, and physics) into a single model for cell function.

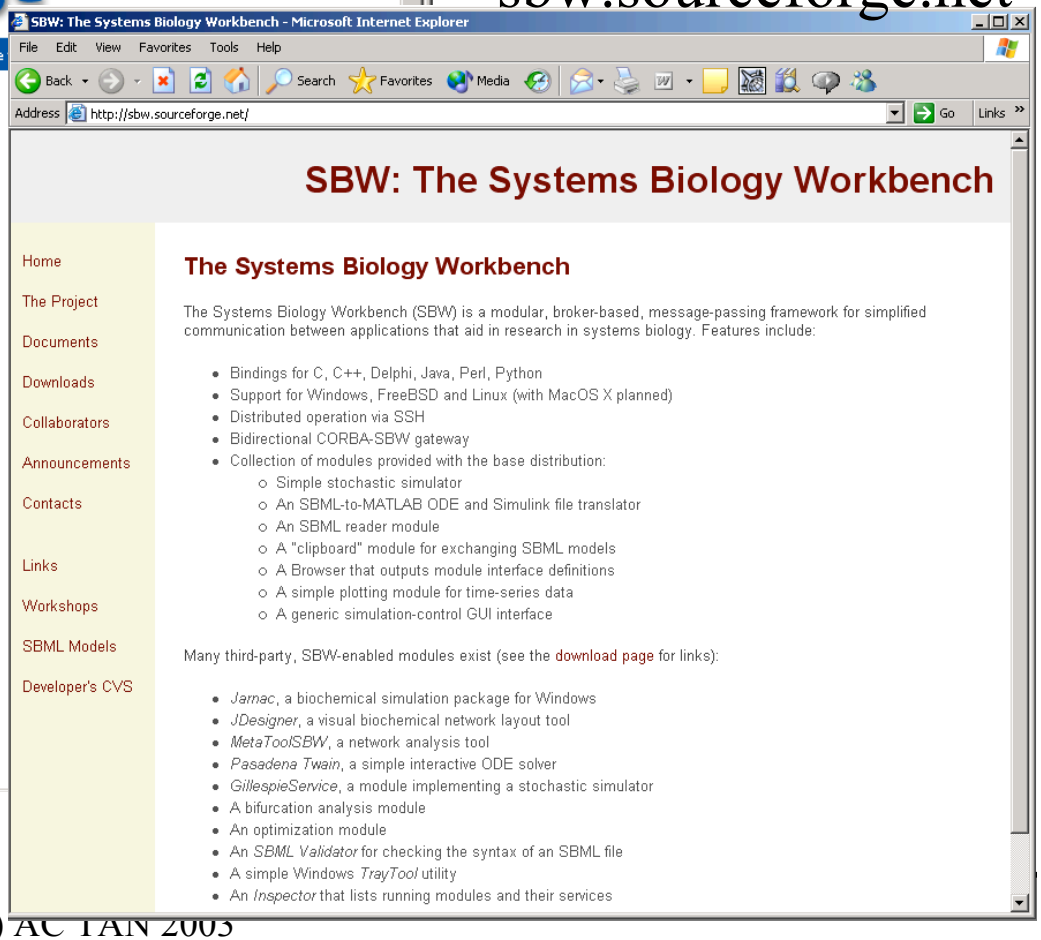
- Until now bioinformatic approaches to the dynamics of cell function have remained 'limited' to categorization of all enzymes, to flux analysis delimitation of metabolic, to metabolic pathway identification, and to the computational biochemistry of isolated metabolic pathways at steady state. For the first time metabolic pathways, their regulation, signal transduction and structure-flux relations will be addressed in a single context, using computational biochemistry, i.e. calculating dynamic concentrations and process rates from molecular data.

SBML and SBW



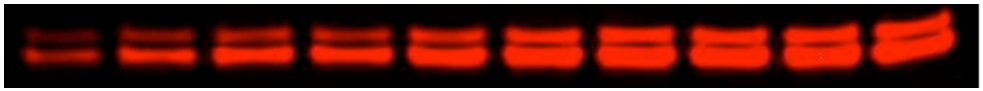
www.sbml.org

sbw.sourceforge.net



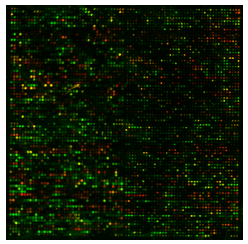
Where do the data come from?

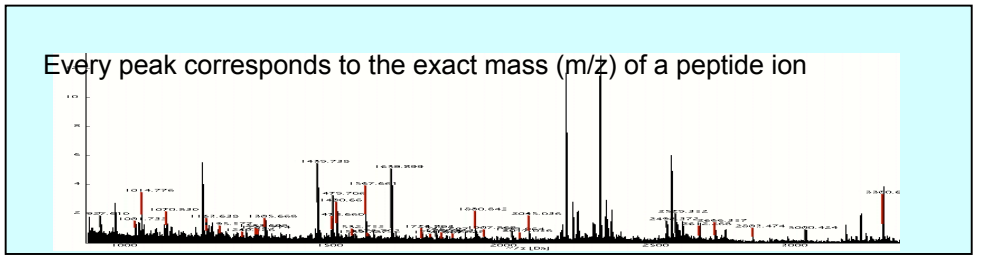
- ‘Traditional’ biochemistry and genetics
  - Reductionist
  - Descriptive
  - “One gene = one career”
  - Published in scientific journals
  - Text mining
- Low throughput: immuno-precipitation,…

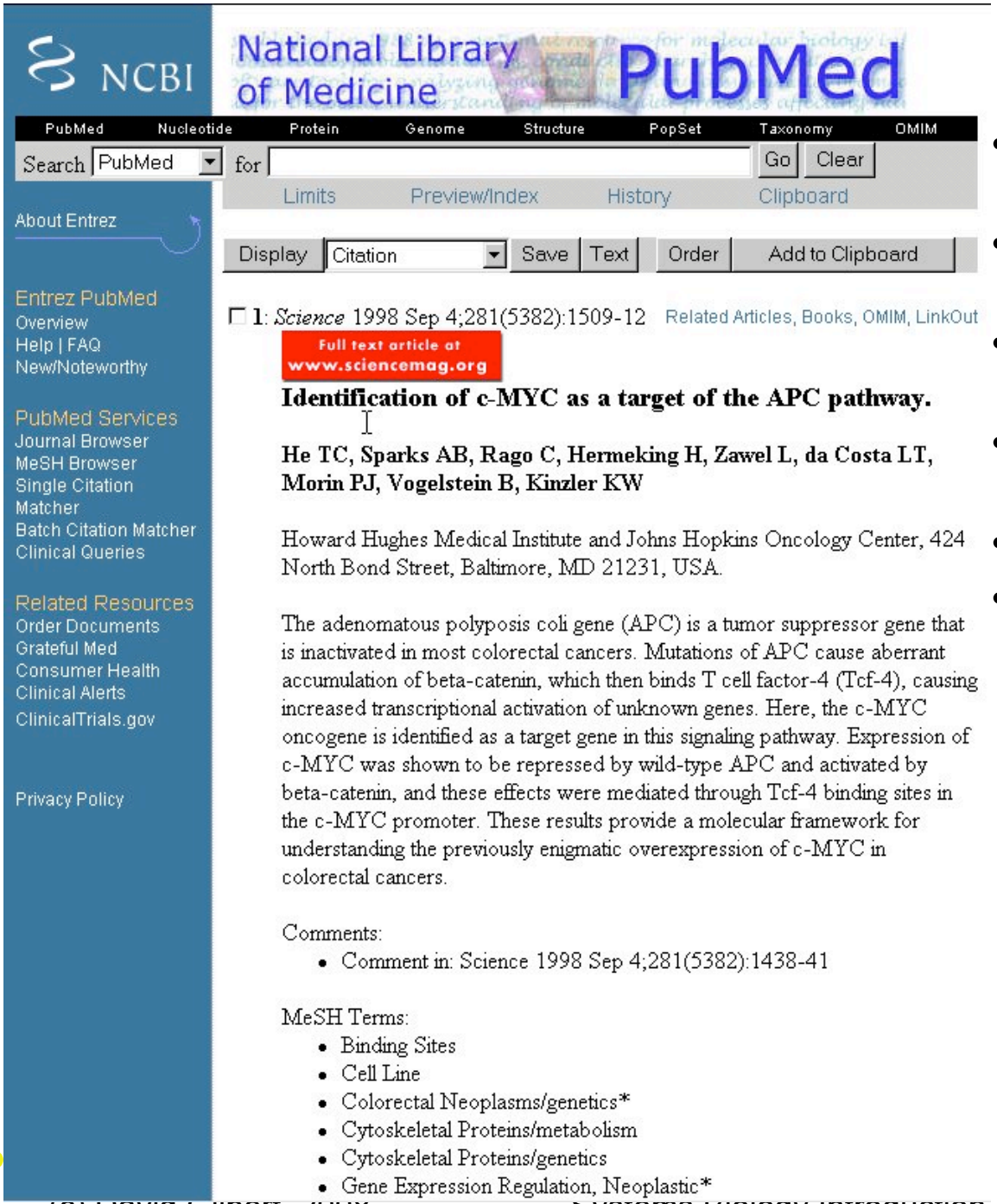


Drivers - technology

- High throughput:
  - Genome sequencing
  - Gene expression array
  - Protein array
  - Mass spectrometry
  - Metabolomics





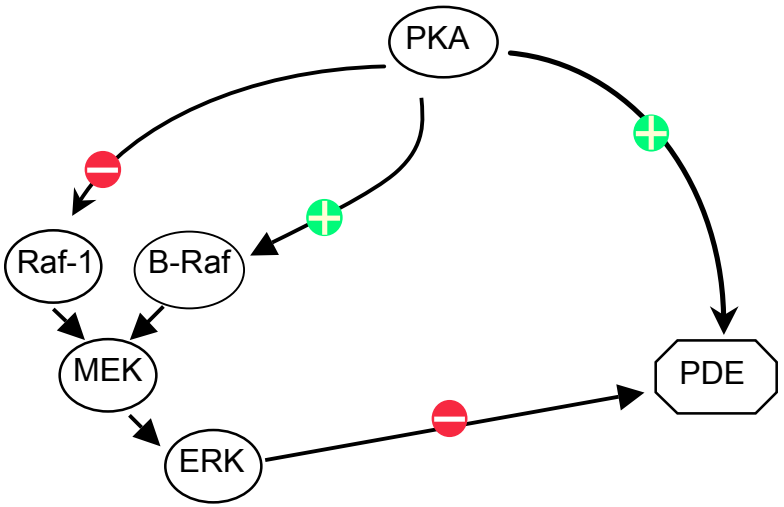


- Maintained by National Library of Medicine
- Free of charge, since 1997
- > 14 million references since 1971
- > 4000 biomedical journals
- > 80% in English
- > 80% have an abstract

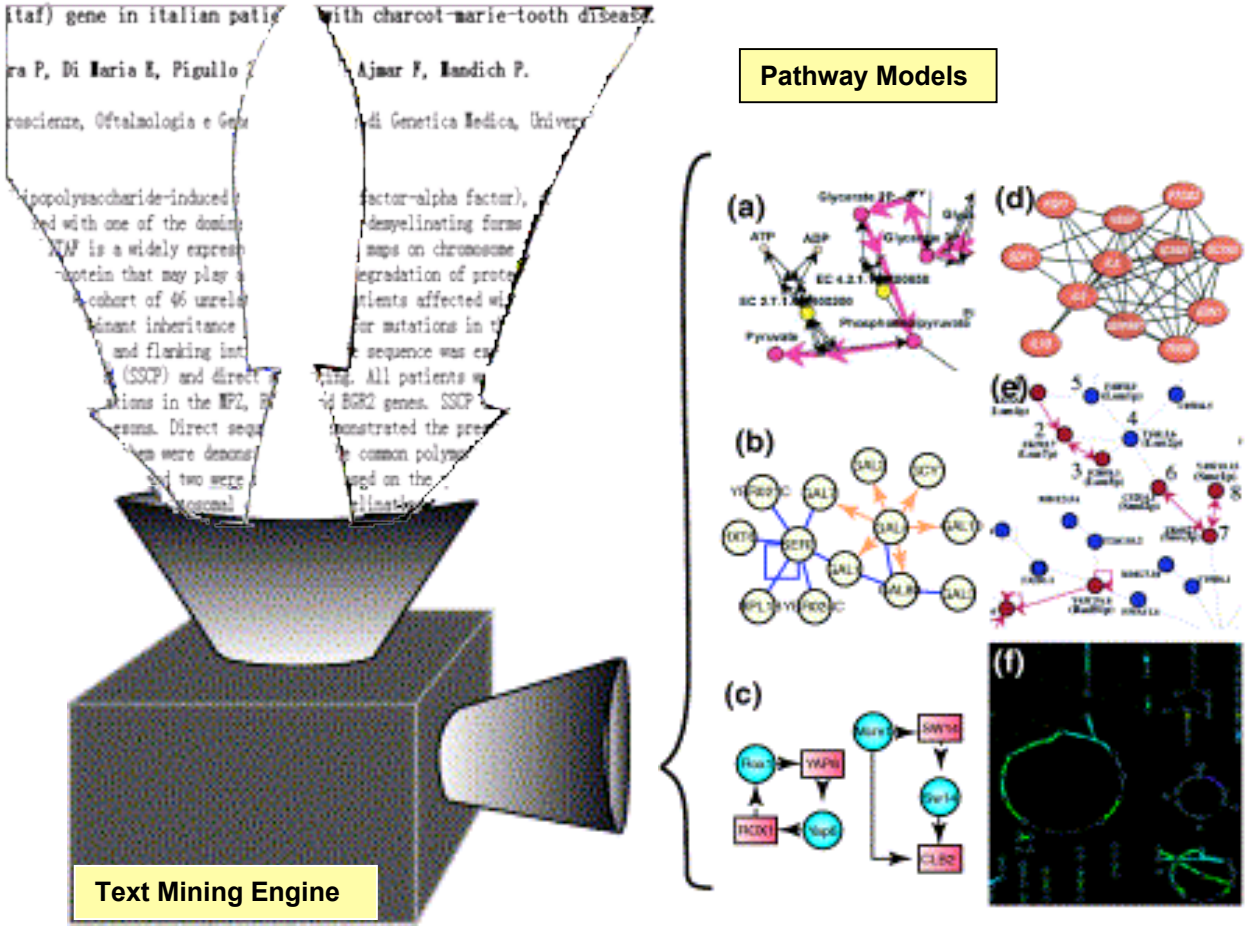
www.ncbi.nlm.nih.gov/entrez

Text Mining: an Example

- …This kinase, which we provisionally denote **MEK** for MAPK/Erk kinase, phosphorylated kinase-inactive **Erk-1** protein primarily on a tyrosine residue and…
- …We further demonstrate in NIH3T3 and Rat 1a cells that **Raf-1** is activated, as measured by its ability to phosphorylate **MEK-1** …
- …immunoblotting and immunoprecipitation experiments demonstrated co-purification of **MEK** activator with **B-Raf**.
- …Signalling by the **Raf-1** kinase can be **blocked** by activation of the cyclic AMP (cAMP)-dependent protein kinase A (**PKA**)…
- …we found that **PKA activates B-Raf** in vitro…
- …**PDE**4 long-form isoenzymes were markedly **inhibited** by **Erk**2 phosphorylation …
- …protein kinase A (**PKA**) **activated** by cAMP can activate **PDE** that hydrolyses …



Biomedical network data mined from scientific texts



GO – Gene Ontology

- In computer science, an **ontology** is the attempt to formulate an exhaustive and rigorous conceptual schema within a given domain, a typically hierarchical data structure containing all the relevant entities and their relationships and rules
- GO - provide controlled vocabularies for the description of the molecular function, biological process and cellular component of gene products.
- These terms are to be used as attributes of gene products by collaborating databases, facilitating uniform queries across them.
- The controlled vocabularies of terms are structured to allow both attribution and querying to be at different levels of granularity.
- http://geneontology.org/

Go ontology

- _all : all (179120)
- _ GO:0008150 : biological_process (113950)
- _ GO:0005575 : cellular_component (105682)
- _ GO:0003674 : molecular_function (113055)
  - _ GO:0003824 : catalytic activity (37486)
    - _ GO:0016491 : oxidoreductase activity (5624)
      - _ GO:0016730 : oxidoreductase activity, acting on iron-sulfur proteins as donors (35)
        - » _ GO:0016731 : oxidoreductase activity, acting on iron-sulfur proteins as donors, NAD or NADP as acceptor (15)
        - » _ GO:0008937 : ferredoxin reductase activity (15)
        - » _ GO:0004324 : ferredoxin-NADP+ reductase activity (7)
  - _ GO:0005215 : transporter activity (9663)
    - _ GO:0005489 : electron transporter activity (1383)
      - _ GO:0008937 : ferredoxin reductase activity (15)
        - » _ GO:0004324 : ferredoxin-NADP+ reductase activity (7)
- _ obsolete_biological_process : obsolete_biological_process (146)
- _ obsolete_cellular_component : obsolete_cellular_component (24)
- _ obsolete_molecular_function : obsolete_molecular_function (1710)

(Bio-)chemical reactions

- Catalyst - substance that increases the rate of a chemical reaction without being consumed in the process.
- Enzyme
  - biological catalyst
  - mainly these are proteins
  - Highly specific for a particular reaction
- Coenzyme - enhances the activity of an enzyme
- substrate (reactant) - consumed in a reaction
- product - produced by a reaction

$$A + B \rightleftharpoons C + D$$

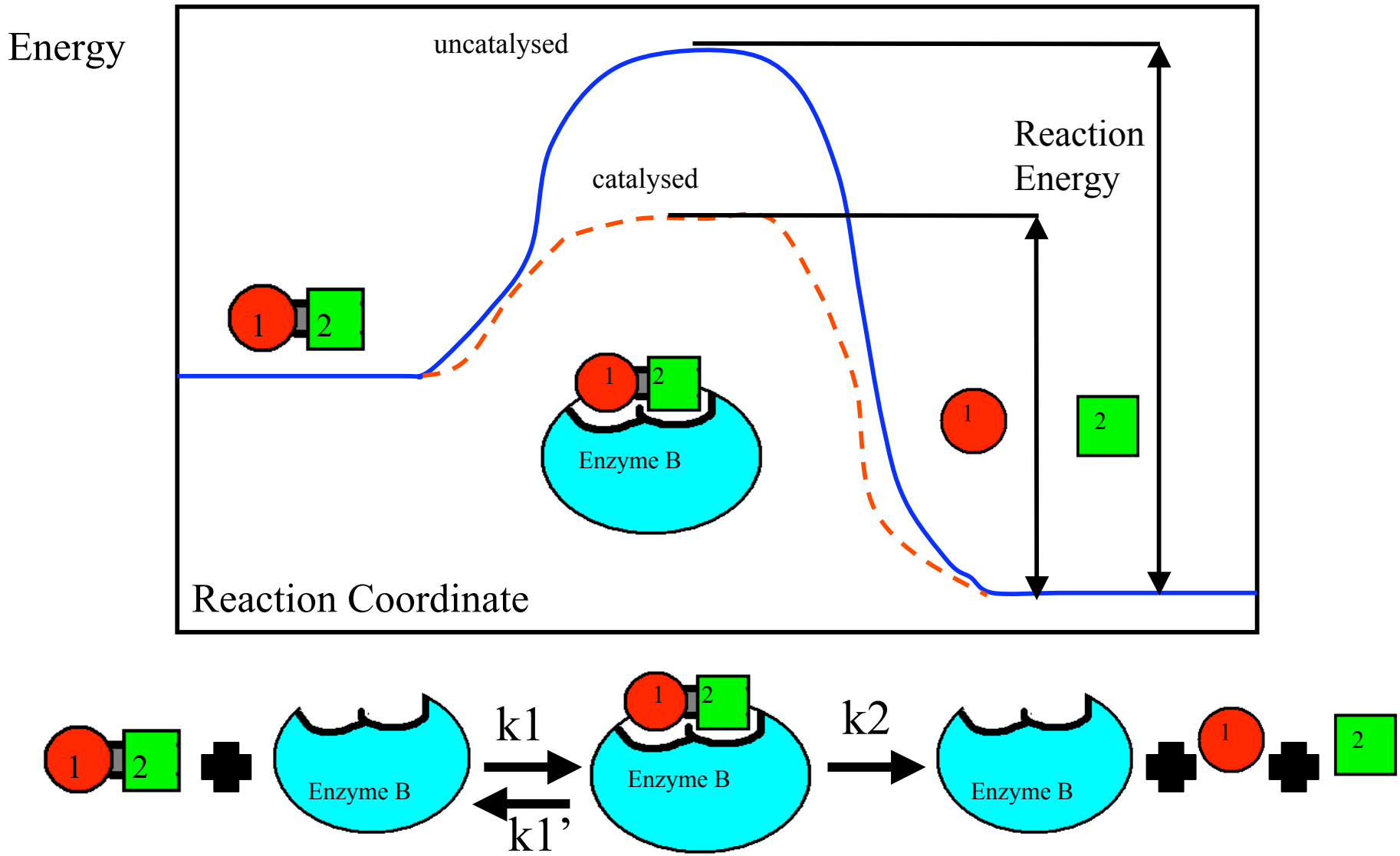
substrates products

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + H^+ + NADH$$

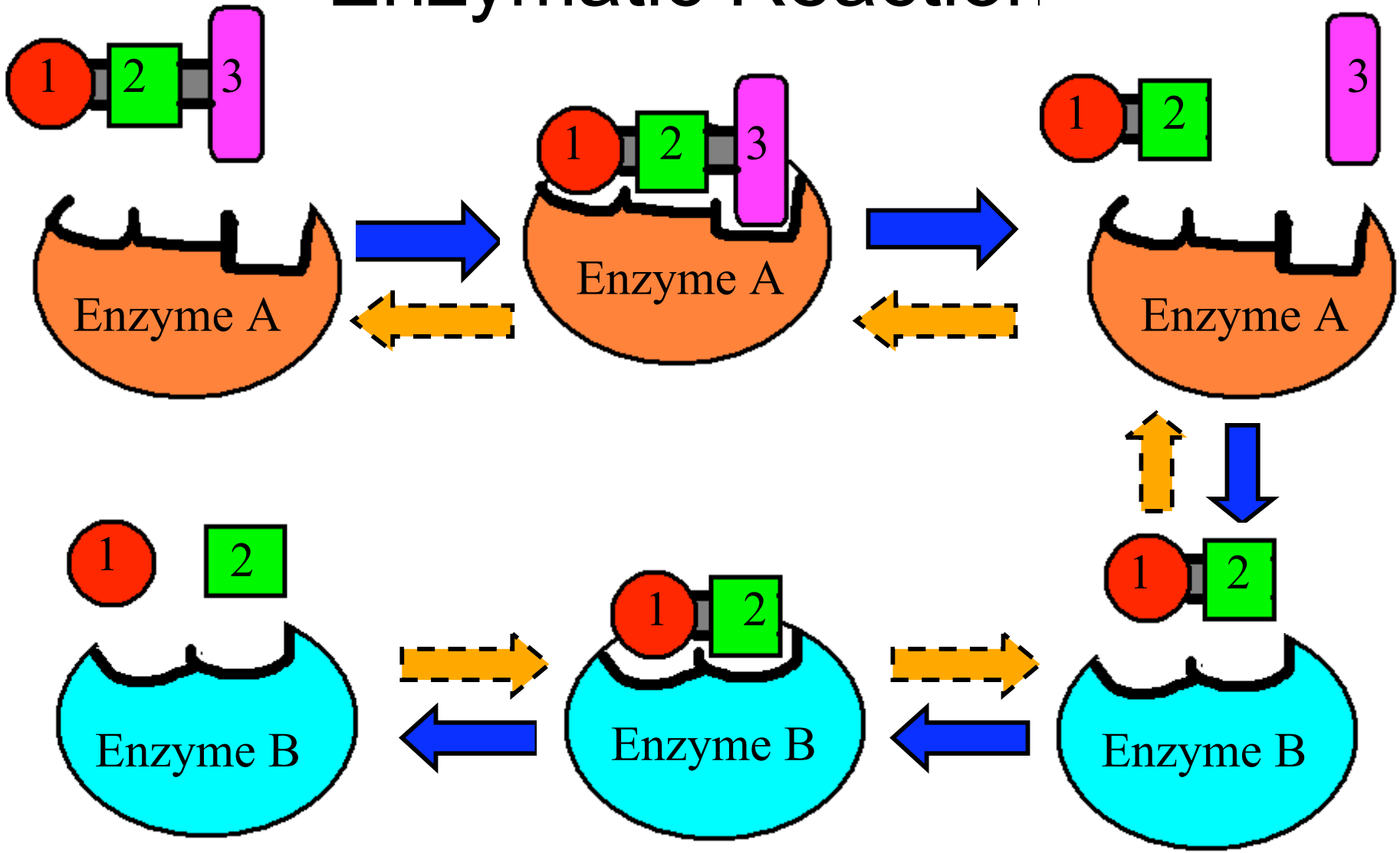
ethanol (alcohol) cofactor acetaldehyde

This reaction is catalysed (accelerated) by alcohol dehydrogenase

Transition State Diagram



Metabolic Pathway = A series of Enzymatic Reactions



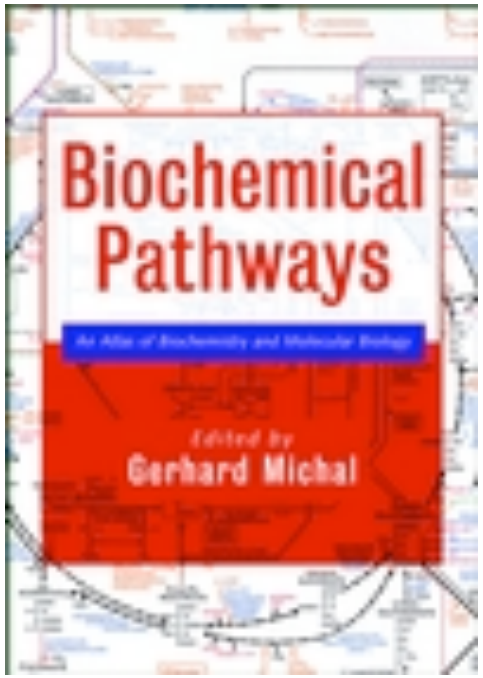
What is a metabolic pathway?

- An ordered sequence of proteins and substrates
- A series of biochemical reactions
- An evolutionary product
- A biological system (living cell)
- A biochemical network/graph

Issues:

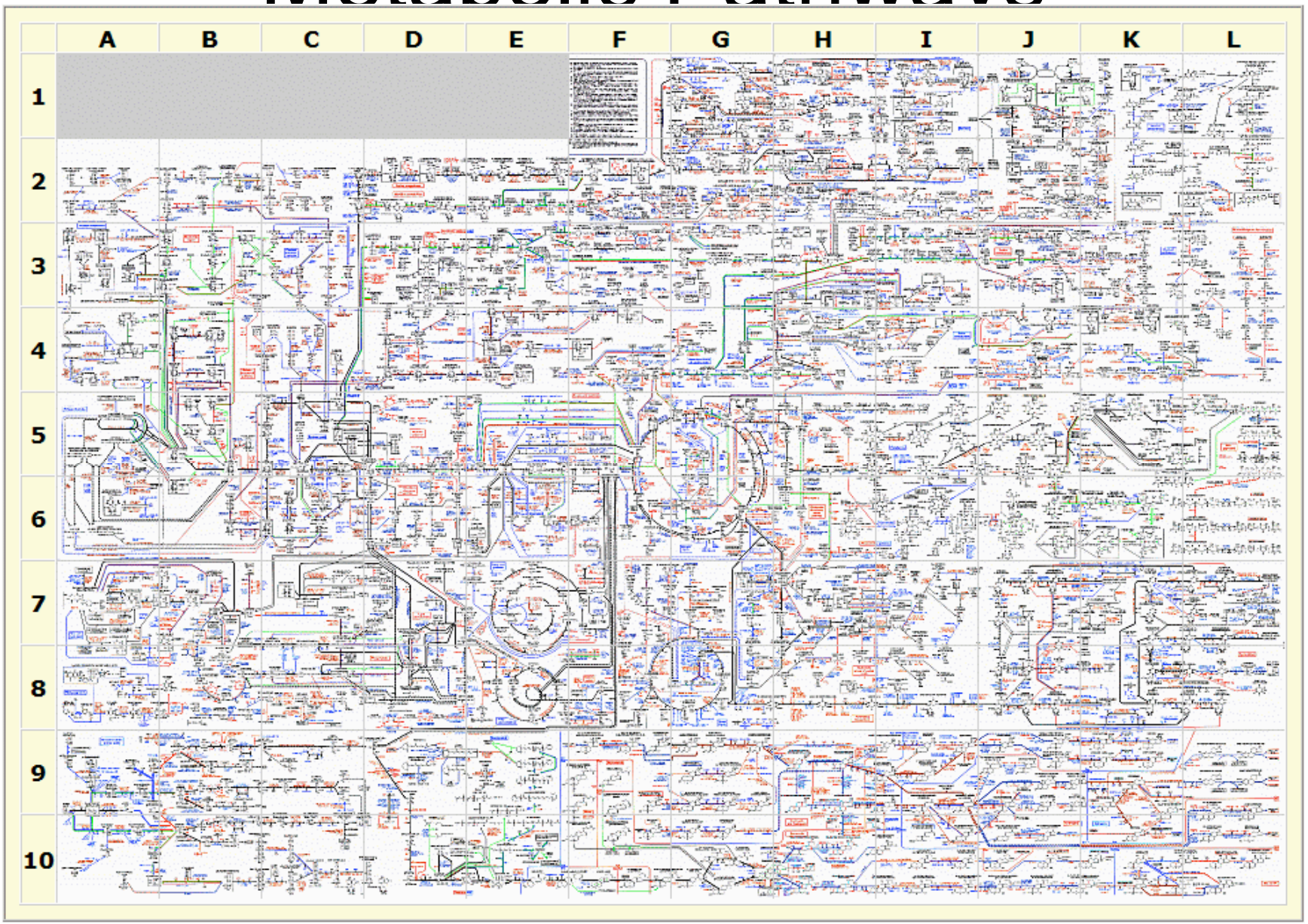
- Organism-specific adaptations
- Which enzyme sets are involved?

Pathway maps

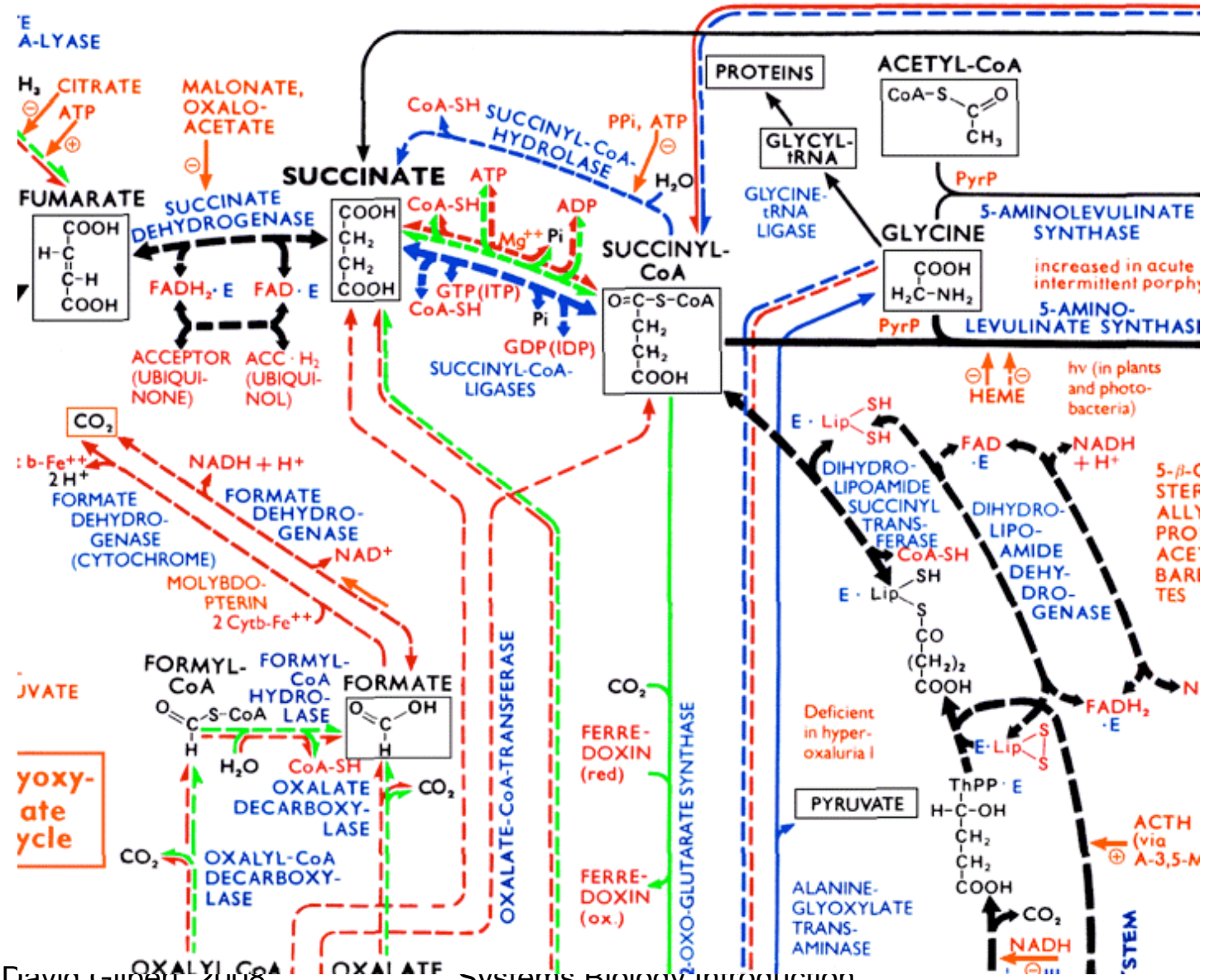


Gerhard Michal

Metabolic Pathways



→ general biochemical pathways, → animals,
→ higher plants, → unicellular organisms



Fundamental questions

- How does a cell extract energy (and reducing power) form its environment?
- How does a cell synthesise the building blocks of its macromolecules and then the macromolecules themselves?

⇨METABOLISM

- A highly integrated network of chemical reactions

Overwhelming?

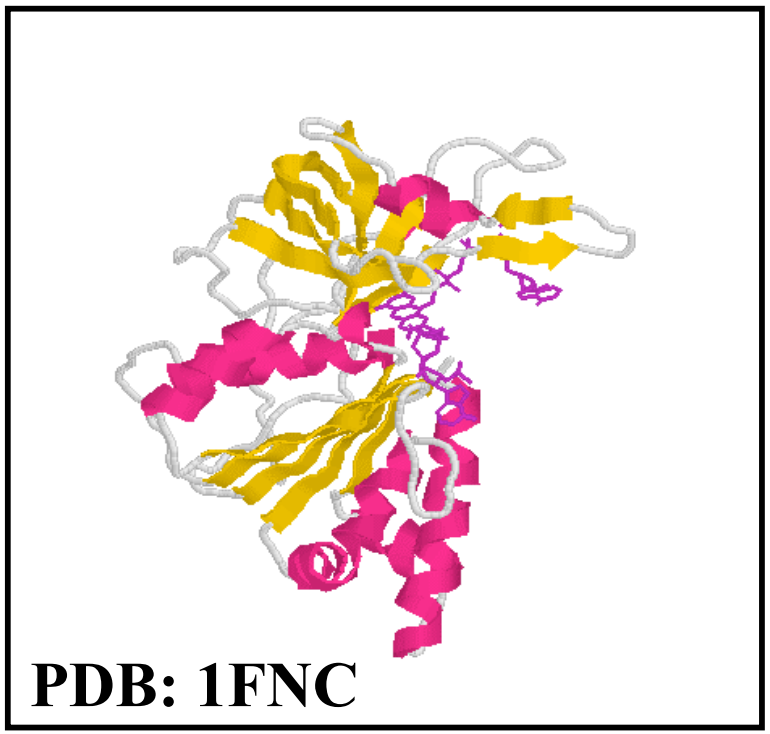
- More than 1000 chemical reactions take place even in a simple organism like *Escherichia coli*
- But:
- Metabolism contains many common motifs
  - Use of a universal energy currency (ATP)
  - Use of a limited number of activated intermediates
  - ⇨ Around 100 molecules play central roles
- Only a small number of different *kinds* of reactions
- Mechanism of these reactions is usually simple
- Pathways are regulated in a common way

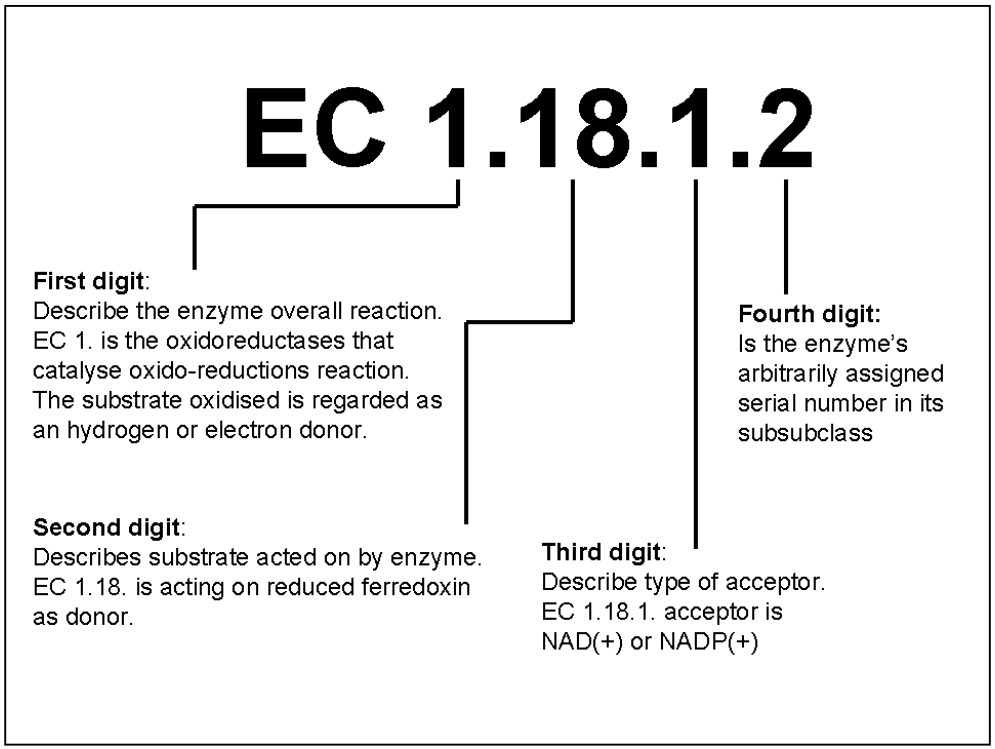
EC Classification (EC)

- Classified according to Enzyme Nomenclature

www.chem.qmw.ac.uk/iubmb/enzyme

- INTERNATIONAL UNION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY (IUBMB)
- Six major biochemical reactions
- Denoted in four figures (EC X.X.X.X) according to the reaction





EC classes

- EC 1 Oxidoreductases
- EC 2 Transferases
- EC 3 Hydrolases
- EC 4 Lyases
- EC 5 Isomerases
- EC 6 Ligases

EC 1. Oxidoreductases

- EC 1.1 Acting on the CH-OH group of donors
- EC 1.2 Acting on the aldehyde or oxo group of donors
- EC 1.3 Acting on the CH-CH group of donors
- EC 1.4 Acting on the CH-NH2 group of donors
- EC 1.5 Acting on the CH-NH group of donors
- EC 1.6 Acting on NADH or NADPH
- EC 1.7 Acting on other nitrogenous compounds as donors
- EC 1.8 Acting on a sulfur group of donors
- EC 1.9 Acting on a heme group of donors
- EC 1.10 Acting on diphenols and related substances as donors
- EC 1.11 Acting on a peroxide as acceptor
- EC 1.12 Acting on hydrogen as donor
- EC 1.13 Acting on single donors with incorporation of molecular oxygen (oxygenases)
- EC 1.14 Acting on paired donors, with incorporation or reduction of molecular oxygen
- EC 1.15 Acting on superoxide radicals as acceptor
- EC 1.16 Oxidising metal ions
- EC 1.17 Acting on CH or CH2 groups
- EC 1.18 Acting on iron-sulfur proteins as donors
- EC 1.19 Acting on reduced flavodoxin as donor
- EC 1.20 Acting on phosphorus or arsenic in donors
- EC 1.21 Acting on X-H and Y-H to form an X-Y bond
- EC 1.97 Other oxidoreductases

EC 1.18.

- EC 1.18.1 With NAD+ or NADP+ as acceptor
- EC 1.18.2 With dinitrogen as acceptor (transferred to EC 1.18.6)
- EC 1.18.3 With H+ as acceptor
- EC 1.18.6 With dinitrogen as acceptor
- EC 1.18.96 With other, known, acceptors
- EC 1.18.99 With H+ as acceptor

EC 1.18.1.

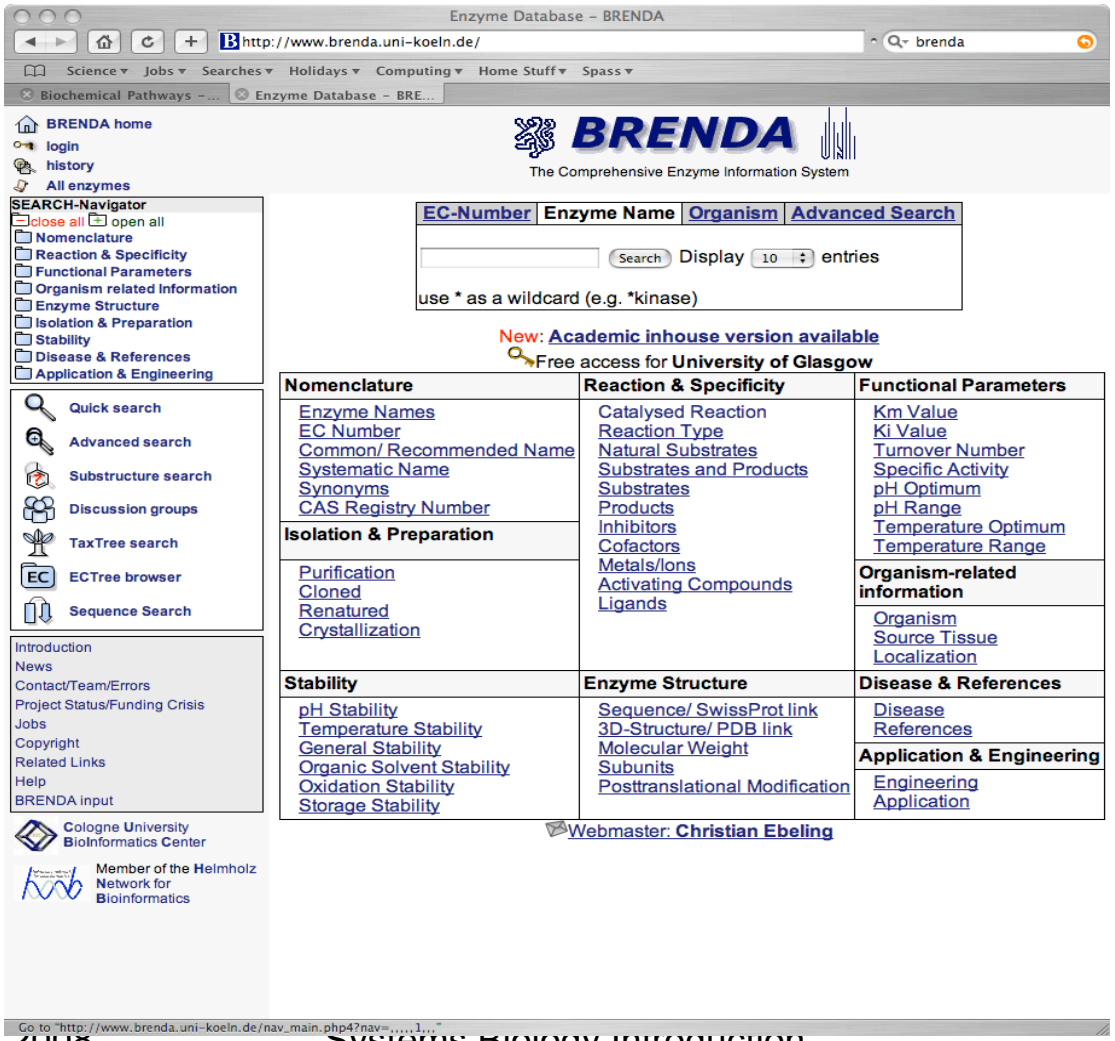
- EC 1.18.1.1 rubredoxin—NAD+ reductase
  EC 1.18.1.2 ferredoxin—NADP+ reductase
  EC 1.18.1.3 ferredoxin—NAD+ reductase
  EC 1.18.1.4 rubredoxin—NAD(P)+ reductase

EC 1.18.1.2

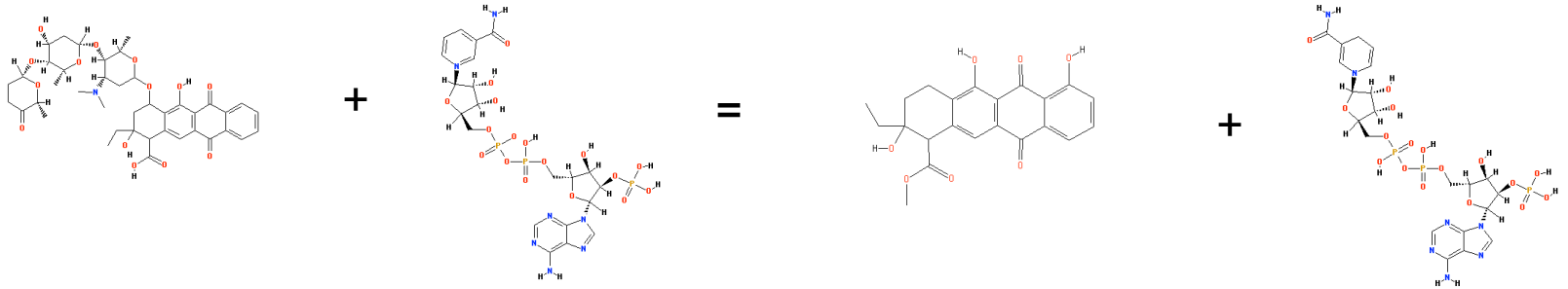
- **IUBMB Enzyme Nomenclature**
- **EC 1.18.1.2**
- **Common name:** ferredoxin—NADP+ reductase **Reaction:** reduced ferredoxin + NADP+ = oxidized ferredoxin + NADPH + H+
- For diagram click here.
- **Other name(s):** adrenodoxin reductase; ferredoxin:NADP+ oxidoreductase; ferredoxin-nicotinamide adenine dinucleotide phosphate reductase; ferredoxin-NADP reductase; TPNH-ferredoxin reductase; ferredoxin-NADP oxidoreductase; NADP:ferredoxin oxidoreductase; ferredoxin-TPN reductase; reduced nicotinamide adenine dinucleotide phosphate-adrenodoxin reductase; ferredoxin-NADP-oxidoreductase; NADPH:ferredoxin oxidoreductase; ferredoxin-nicotinamide-adenine dinucleotide phosphate (oxidized) reductase; ferredoxin—NADP reductase
- **Systematic name:** ferredoxin:NADP+ oxidoreductase
- **Comments:** A flavoprotein. Formerly EC 1.6.7.1 and EC 1.6.99.4. Can also reduce flavodoxin.
- **Links to other databases**: BRENDA, EXPASY, KEGG, ERGO, PDB, CAS registry number: 9029-33-8
- **References:**
- 1. Omura, T., Sanders, E., Estabrook, R.W., Cooper, D.Y. and Rosenthal, O. Isolation from adrenal cortex of a nonheme iron protein and a flavoprotein functional as a reduced triphosphopyridine nucleotide-cytochrome P-450 reductase. *Arch. Biochem. Biophys.* 117 (1966) 660-673.
- 2. Shin, M., Tagawa, K. and Arnon, D.I. Crystallization of ferredoxin-TPN reductase and its role in the photosynthetic apparatus of chloroplasts. *Biochem. Z.* 338 (1963) 84-96.

BRENDA

- http://www.brenda.uni-koeln.de/
- Contains information on enzyme function



Extract of entry of ferredoxin-NADP+ reductase (EC-Number 1.18.1.2)



aclacinomycin A + NADP+ = 7-deoxyaklavinone + NADPH

Spinacia oleraceaunder anaerobic conditions

See also more info
http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=1.18.1.2

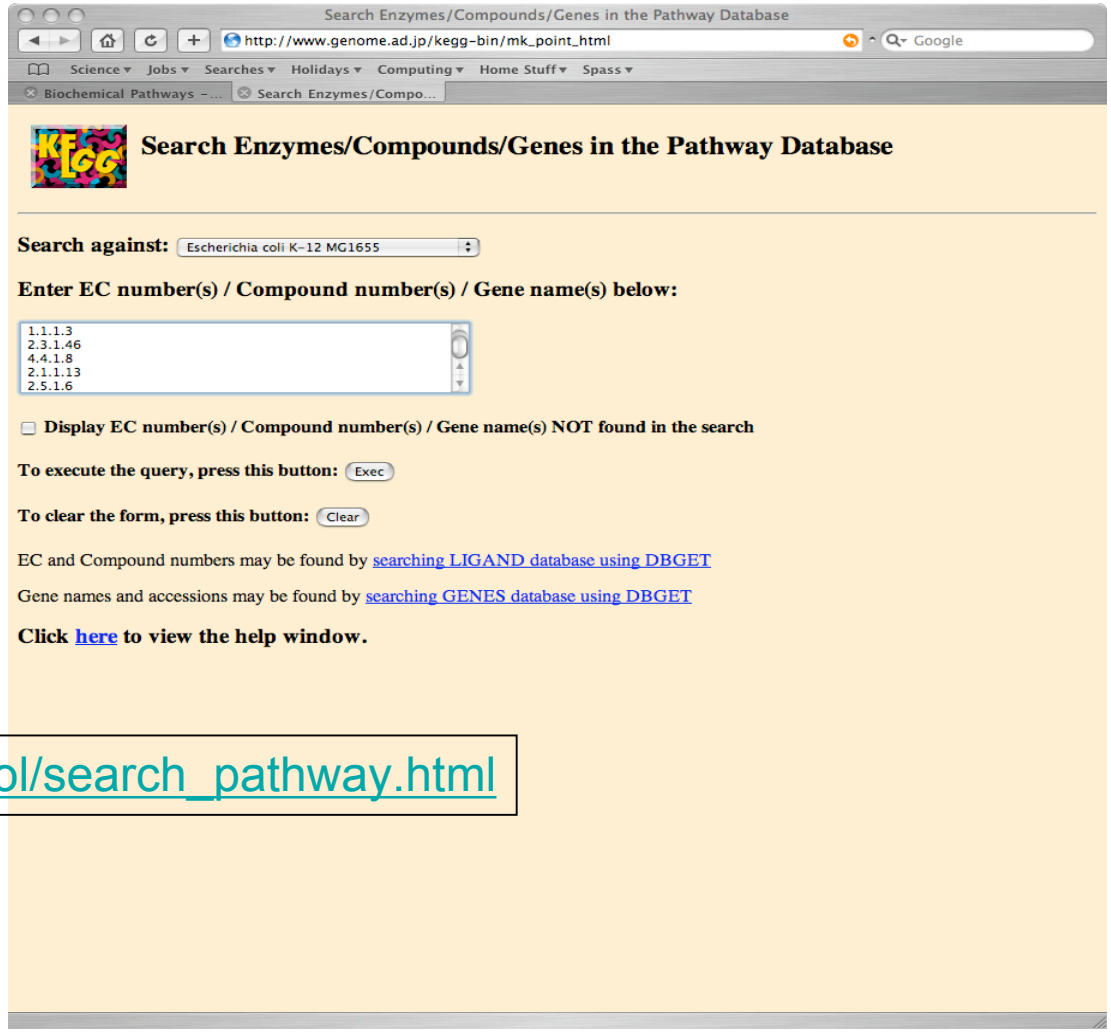
239 species…

http://www.genome.jp/kegg/
http://www.genome.jp/about_genomenet/service.html

Search in KEGG pathways

Find a pathway in E.coli enzymatic reactions

- 2.7.2.4
- 1.2.1.11
- 1.1.1.3
- 2.3.1.46
- 4.4.1.8
- 2.1.1.13
- 2.5.1.6



http://www.genome.jp/kegg/tool/search_pathway.html

Pathway Search Result

map00260 Glycine, serine and threonine metabolism

EC 1.1.1.3	Homoserine dehydrogenase
EC 1.2.1.11	Aspartate-semialdehyde dehydrogenase
EC 2.7.2.4	Aspartate kinase; Aspartokinase

map00271 Methionine metabolism

EC 2.1.1.13	5-Methyltetrahydrofolate--homocysteine S-methyltransferase; Methionine synthase; Tetrahydropteroylglutamate methyltramsferase
EC 2.3.1.46	Homoserine O-succinyltransferase; Homoserine O-transsuccinylase
EC 2.5.1.6	Methionine adenosyltransferase
EC 4.4.1.8	Cystathionine beta-lyase; beta-Cystathionase; Cystine lyase

map00272 Cysteine metabolism

EC 4.4.1.8	Cystathionine beta-lyase; beta-Cystathionase; Cystine lyase

map00300 Lysine biosynthesis

EC 1.1.1.3	Homoserine dehydrogenase
EC 1.2.1.11	Aspartate-semialdehyde dehydrogenase
EC 2.7.2.4	Aspartate kinase; Aspartokinase

map00450 Selenoamino acid metabolism

EC 2.5.1.6	Methionine adenosyltransferase
EC 4.4.1.8	Cystathionine beta-lyase; beta-Cystathionase; Cystine lyase

map00670 One carbon pool by folate

EC 2.1.1.13	5-Methyltetrahydrofolate--homocysteine S-methyltransferase; Methionine synthase; Tetrahydropteroylglutamate methyltramsferase

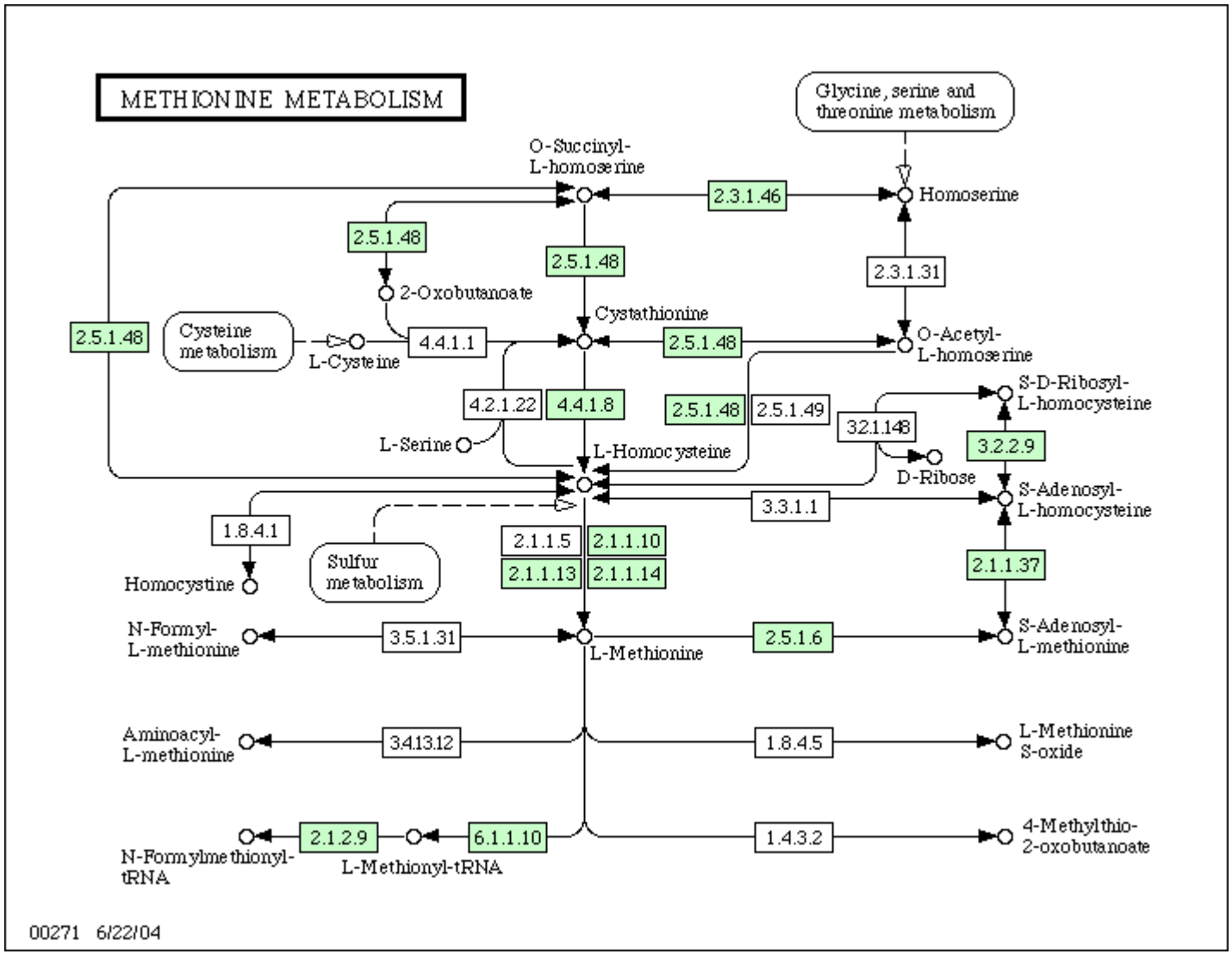
map00910 Nitrogen metabolism

EC 4.4.1.8	Cystathionine beta-lyase; beta-Cystathionase; Cystine lyase

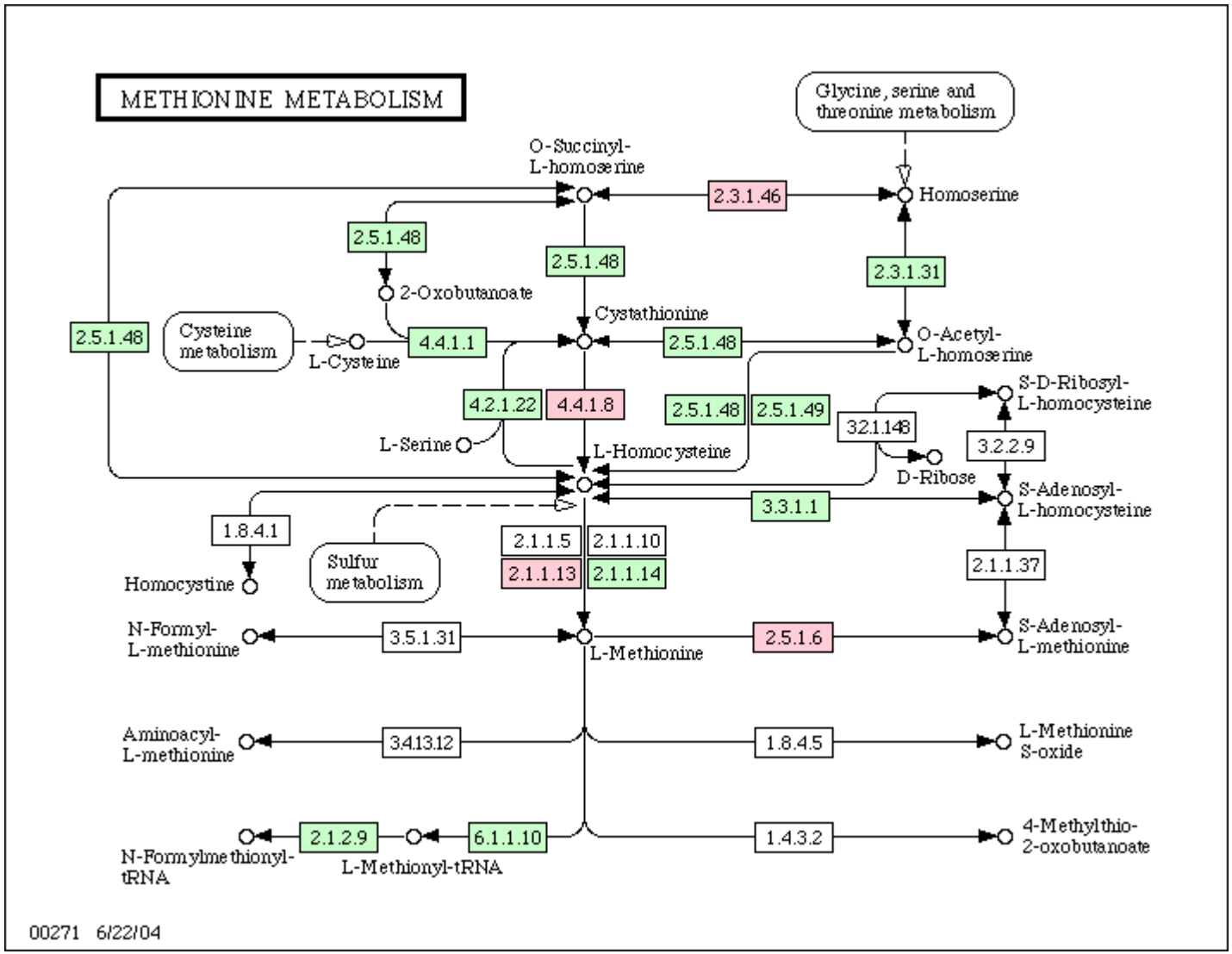
map00920 Sulfur metabolism

EC 2.3.1.46	Homoserine O-succinyltransferase; Homoserine O-transsuccinylase
EC 4.4.1.8	Cystathionine beta-lyase; beta-Cystathionase; Cystine lyase

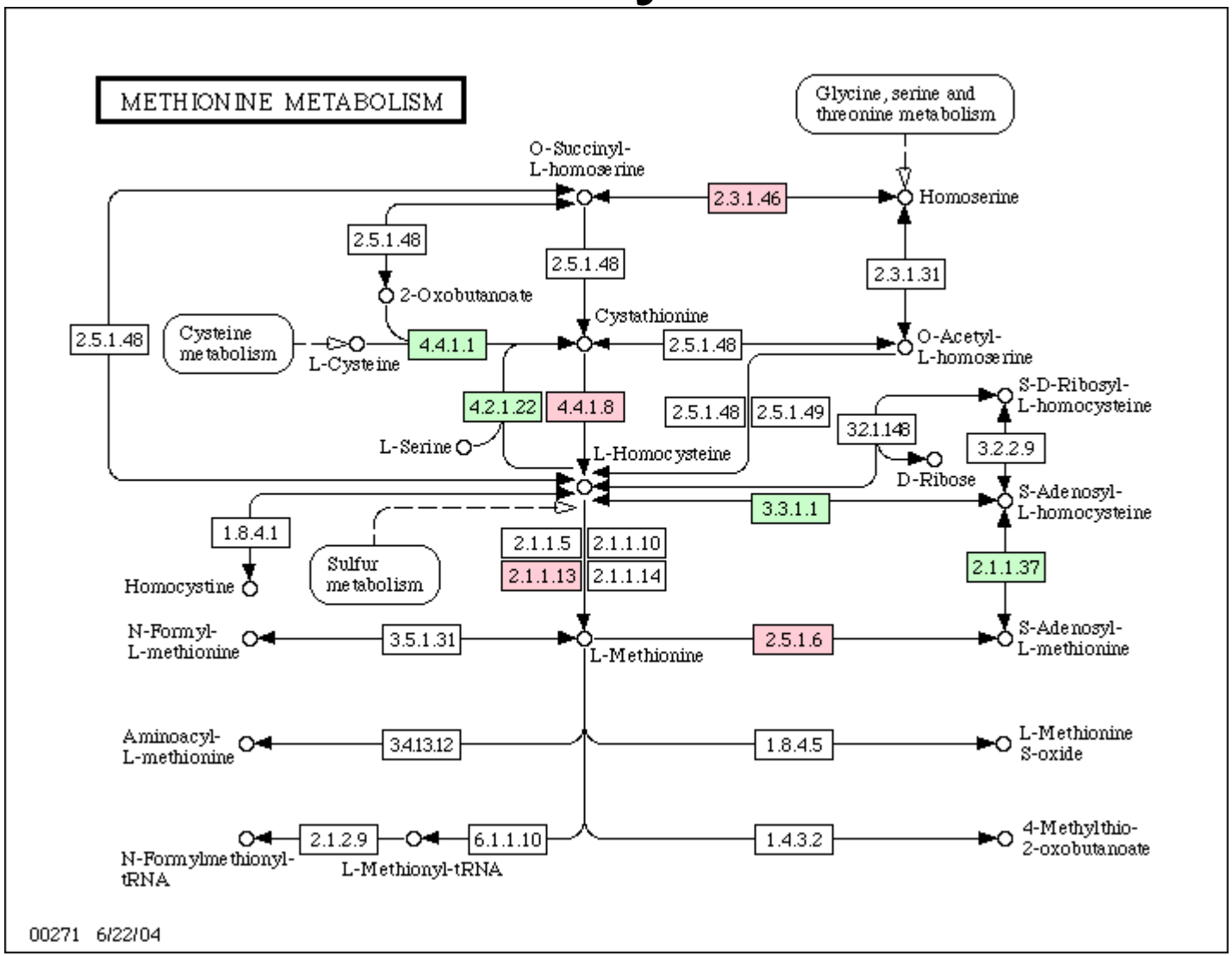
Escherichia coli K-12 MG1655



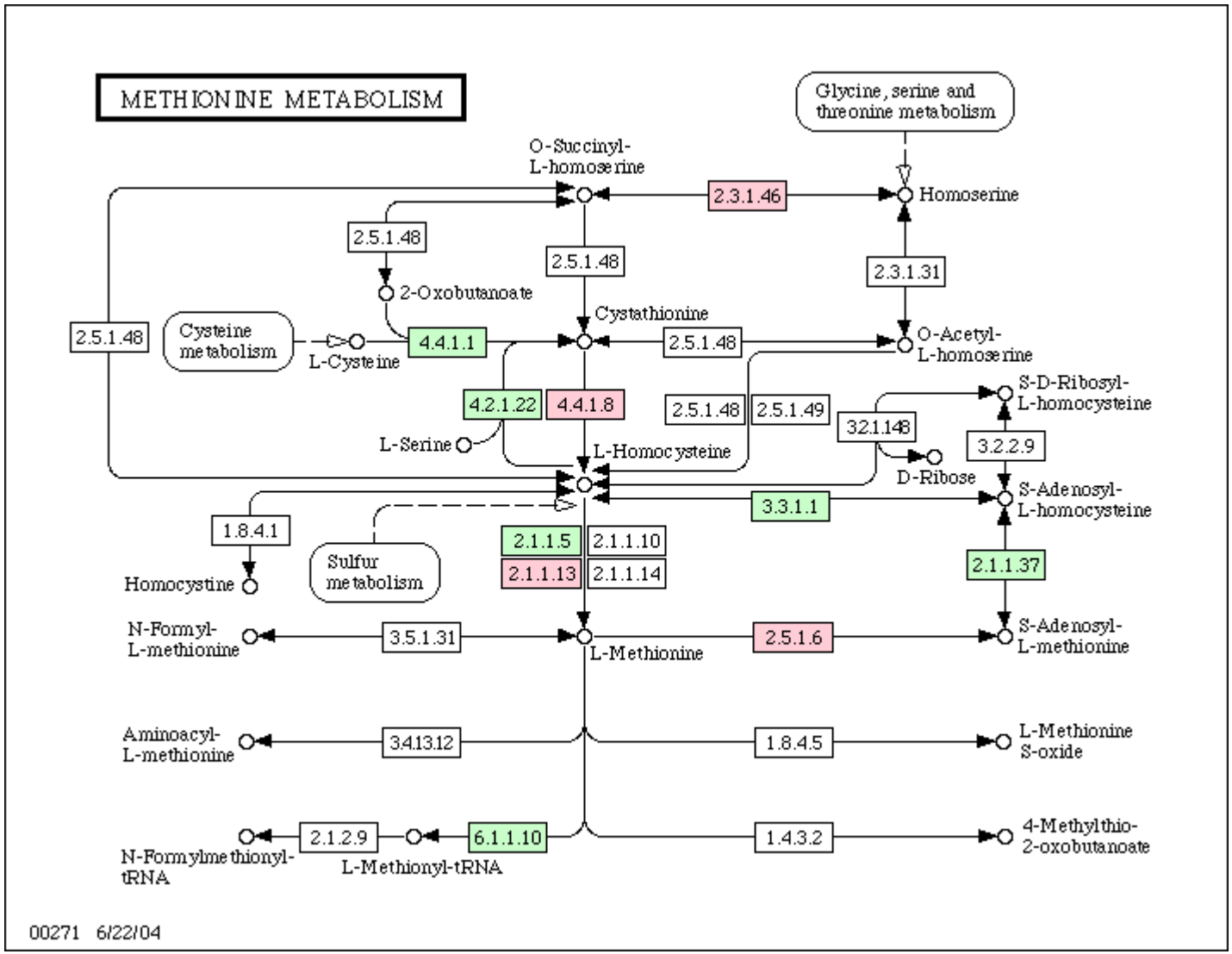
Yeast


METHIONINE METABOLISM
Glycine, serine and threonine metabolism
O-Succinyl-L-homoserine
2.3.1.46
Homoserine
2.5.1.48
2.5.1.48
2.3.1.31
2-Oxobutanoate
Cystathionine
2.5.1.48
Cysteine metabolism
4.4.1.1
L-Cysteine
2.5.1.48
O-Acetyl-L-homoserine
S-D-Ribosyl-L-homocysteine
4.2.1.22
4.4.1.8
2.5.1.48
2.5.1.49
3.2.1.148
3.2.2.9
L-Serine
L-Homocysteine
D-Ribose
S-Adenosyl-L-homocysteine
3.3.1.1
1.8.4.1
2.1.1.5
2.1.1.10
2.1.1.37
Sulfur metabolism
2.1.1.13
2.1.1.14
Homocystine
N-Formyl-L-methionine
3.5.1.31
L-Methionine
2.5.1.6
S-Adenosyl-L-methionine
Aminoacyl-L-methionine
3.4.13.12
1.8.4.5
L-Methionine S-oxide
N-Formylmethionyl-tRNA
2.1.2.9
6.1.1.10
L-Methionyl-tRNA
1.4.3.2
4-Methylthio-2-oxobutanoate
00271 6/22/04

Fly


METHIONINE METABOLISM
Glycine, serine and threonine metabolism
O-Succinyl-L-homoserine
2.3.1.46
Homoserine
2.5.1.48
2.5.1.48
2.3.1.31
2-Oxobutanoate
Cystathionine
2.5.1.48
Cysteine metabolism
4.4.1.1
L-Cysteine
2.5.1.48
O-Acetyl-L-homoserine
S-D-Ribosyl-L-homocysteine
4.2.1.22
4.4.1.8
2.5.1.48
2.5.1.49
3.2.1.148
3.2.2.9
L-Serine
L-Homocysteine
D-Ribose
S-Adenosyl-L-homocysteine
3.3.1.1
1.8.4.1
2.1.1.5
2.1.1.10
2.1.1.37
Sulfur metabolism
2.1.1.13
2.1.1.14
Homocystine
N-Formyl-L-methionine
3.5.1.31
L-Methionine
2.5.1.6
S-Adenosyl-L-methionine
Aminoacyl-L-methionine
3.4.13.12
1.8.4.5
L-Methionine S-oxide
N-Formylmethionyl-tRNA
2.1.2.9
6.1.1.10
L-Methionyl-tRNA
1.4.3.2
4-Methylthio-2-oxobutanoate
00271 6/22/04

Human


METHIONINE METABOLISM
Glycine, serine and threonine metabolism
O-Succinyl-L-homoserine
2.3.1.46
Homoserine
2.5.1.48
2.5.1.48
2.3.1.31
2-Oxobutanoate
Cystathionine
2.5.1.48
Cysteine metabolism
4.4.1.1
L-Cysteine
2.5.1.48
O-Acetyl-L-homoserine
S-D-Ribosyl-L-homocysteine
4.2.1.22
4.4.1.8
2.5.1.48
2.5.1.49
3.2.1.148
3.2.2.9
L-Serine
L-Homocysteine
D-Ribose
S-Adenosyl-L-homocysteine
3.3.1.1
1.8.4.1
2.1.1.5
2.1.1.10
2.1.1.37
Sulfur metabolism
2.1.1.13
2.1.1.14
Homocystine
N-Formyl-L-methionine
3.5.1.31
L-Methionine
2.5.1.6
S-Adenosyl-L-methionine
Aminoacyl-L-methionine
3.4.13.12
1.8.4.5
L-Methionine S-oxide
N-Formylmethionyl-tRNA
2.1.2.9
6.1.1.10
L-Methionyl-tRNA
1.4.3.2
4-Methylthio-2-oxobutanoate
00271 6/22/04

KEGG searches

- At http://www.genome.jp/kegg/tool/search_pathway.html
- Try to search with 1.18.1.2
- Also try at http://www.genome.jp/kegg/ligand.html

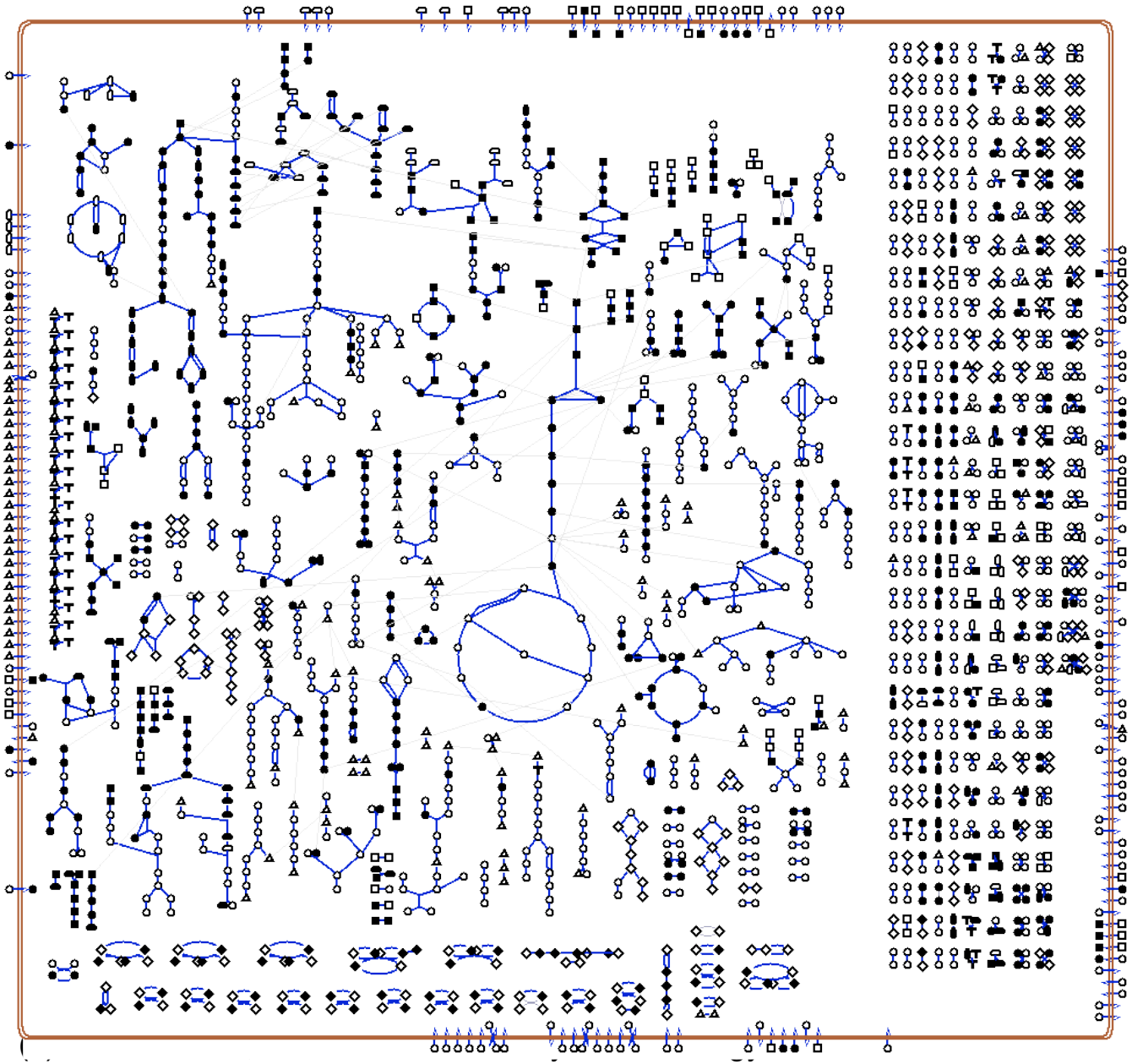
BioCyc Database Collection

- Collection of Pathway/Genome Databases.
- **Literature-derived Pathway/Genome Databases**
- EcoCyc -- *Escherichia coli* K12
- MetaCyc -- Metabolic pathways and enzymes from 150 species

- **Computationally-derived Pathway/Genome Databases**
- AgroCyc -- *Agrobacterium tumefaciens*
- AnthraCyc -- *Bacillus anthracis*
- BsubCyc -- *Bacillus subtilis*
- CtraCyc -- *Chlamydia trachomatis*
- CauloCyc -- *Caulobacter crescentus*
- EcoO157Cyc -- *Escherichia coli* O157:H7
- FrantCyc -- *Francisella tularensis*
- HinCyc -- *Haemophilus influenzae*
- HpyCyc -- *Helicobacter pylori*
- HumanCyc -- *Homo sapiens*
- MtbCdcCyc -- *Mycobacterium tuberculosis* CDC1551
- MtbRvCyc -- *Mycobacterium tuberculosis* H37Rv
- MpneuCyc -- *Mycoplasma pneumoniae*
- PlasmoCyc -- *Plasmodium falciparum*
- ShigellaCyc -- *Shigella flexneri*
- TpalCyc -- *Treponema pallidum*
- VchoCyc -- *Vibrio cholerae*

http://biocyc.org/

EcoCyc - overview of E.coli metabolic map



http://ecocyc.org/

MetaCyc

- Database of nonredundant, experimentally elucidated metabolic pathways
- From more than 240 different organisms.
- Curated from the scientific experimental literature.
- Predominantly qualitative information rather than quantitative data

http://metacyc.org/

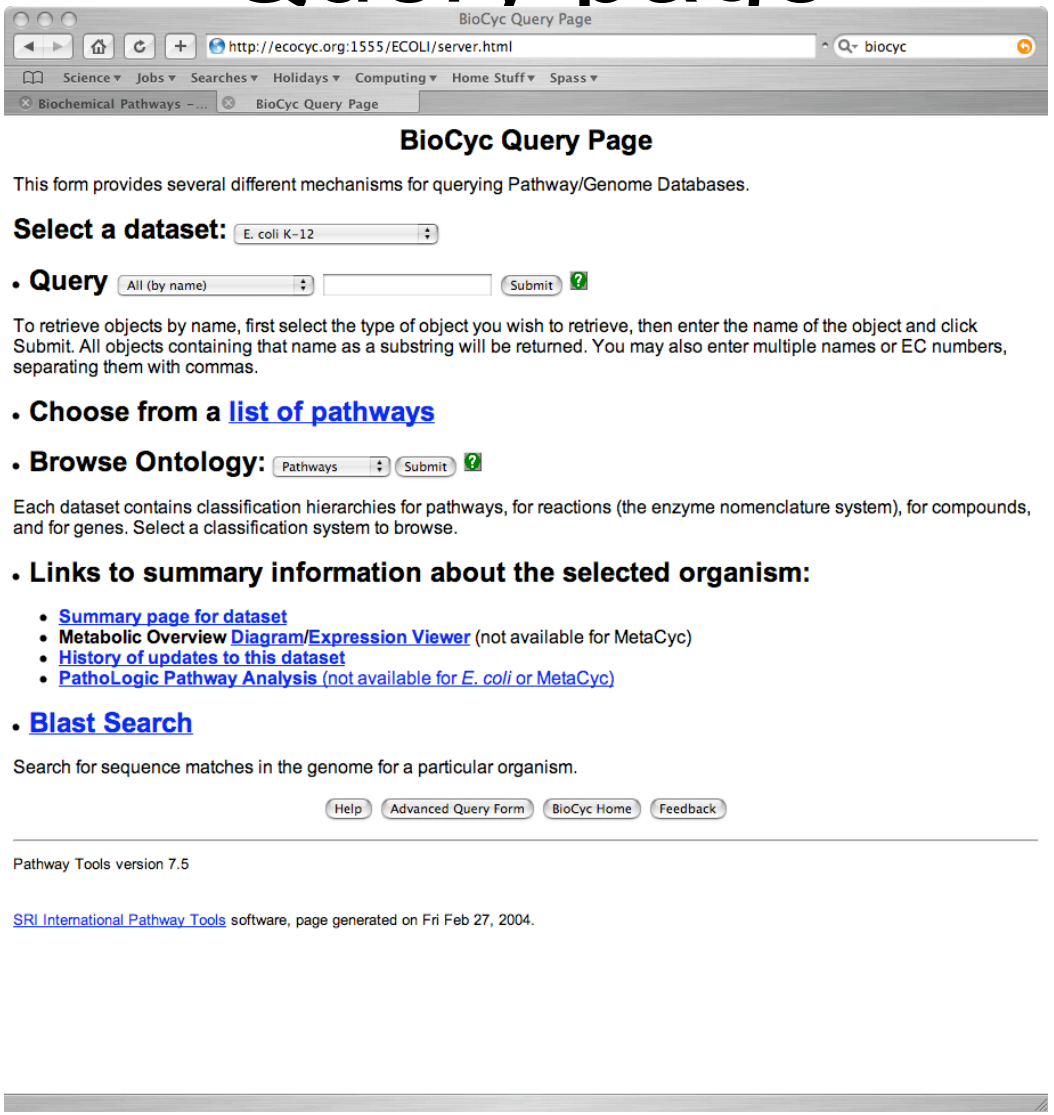
MetaCyc

Query and Visualization for pathways, proteins, reactions and compounds,

- **Text-based search**, when trying to find information without knowing exactly how an object is named.
- **Browse using ontologies**, when one wants to search by proceeding from general categories to specific instances (In computer science, an **ontology** is the attempt to formulate an exhaustive and rigorous conceptual schema within a given domain, a typically hierarchical data structure containing all the relevant entities and their relationships and rules)
- **Direct queries**, when an identifier is known
- (desktop program):
- Compare the overall metabolic maps of different organisms
- Compare specific pathways between two organisms
- Compare the genomic maps of two organisms

http://metacyc.org/META/server.html

Query page


BioCyc Query Page
http://ecocyc.org:1555/ECOLI/server.html
biocyc
Science Jobs Searches Holidays Computing Home Stuff Spass
Biochemical Pathways -...
BioCyc Query Page
BioCyc Query Page
This form provides several different mechanisms for querying Pathway/Genome Databases.
Select a dataset: E. coli K-12
Query All (by name) Submit
To retrieve objects by name, first select the type of object you wish to retrieve, then enter the name of the object and click Submit. All objects containing that name as a substring will be returned. You may also enter multiple names or EC numbers, separating them with commas.
Choose from a list of pathways
Browse Ontology: Pathways Submit
Each dataset contains classification hierarchies for pathways, for reactions (the enzyme nomenclature system), for compounds, and for genes. Select a classification system to browse.
Links to summary information about the selected organism:
Summary page for dataset
Metabolic Overview Diagram/Expression Viewer (not available for MetaCyc)
History of updates to this dataset
PathoLogic Pathway Analysis (not available for E. coli or MetaCyc)
Blast Search
Search for sequence matches in the genome for a particular organism.
Help Advanced Query Form BioCyc Home Feedback
Pathway Tools version 7.5
SRI International Pathway Tools software, page generated on Fri Feb 27, 2004.

- Find information related to 6-phosphofructokinase but you have forgotten its precise name. All you remember is that the enzyme is a kinase involving fructose. **S**earch MetaCyc for all objects (proteins, reactions, genes,…) that contain the words "kinase" and "fructose"…

- **Proteins**

- 1-phosphofructokinase *(**fructose**-1-phosphate **kinase**)*
- 1-phosphofructokinase monomer *(**fructose**-1-phosphate **kinase** monomer)*
- 6-phosphofructokinase-1 *(**fructose**-6-p-1-**kinase**)*
- 6-phosphofructokinase-2 *(**fructose**-6-p-1-**kinase**)*
- **fructose**lysine 6-**kinase**
- **Reactions**

- ATP + fructose-1-phosphate = ADP + fructose-1,6-bisphosphate *(**Fructose** 1-phosphate **kinase**)*
- D-fructose-6-phosphate + pyrophosphate = phosphate + fructose-1,6-bisphosphate *(Diphosphate-dependent 6-phospho**fructose**-1-**kinase**)*

Biochemical Pathway databases - URLs

Database	URL
PRL	http://www.cbio.mskcc.org/prl/index.php Pathway Resource List…
aMAZE	www.amaze.ulb.ac.be (WorkBench for the representation, management, annotation and analysis of information on networks of cellular processes: genetic regulation, biochemical pathways, signal transductions.)
KEGG	www.genome.ad.jp/kegg
BRENDA	www.brenda.uni-koeln.de
PathDB	www.ncgr.org/pathdb Install on local machine
BioCyc	www.biocyc.org
CSNbd	geo.nihs.go.jp/csndb (defunct?) Cell Signalling
BioBase	http://www.gene-regulation.com (TRANSFAC, TRANSPATH,…)
RegulonDB	www.cifn.unam.mx/Computational_Genomics/regulondb
DPinteract	arep.med.harvard.edu/dpinteract (DNA-protein interactions)
EBI	www.ebi.ac.uk/services

EBI On-line text databases

MEDLINE
The premier literature database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. Can be searched using SRS.

MIM
The Online Mendelian Inheritance in Man (OMIM) database is a catalog of human genes and genetic disorders. Can be searched using SRS.

OLDMEDLINE
Contains citations to articles from international biomedical journals covering the fields of medicine, preclinical sciences and allied health sciences from 1953 through 1965. Can be searched using SRS.

Patent Abstracts
This is a set of biotechnology-related abstracts of patent applications derived from data products of the European Patent Office (EPO). Can be searched using SRS.

Taxonomy
The taxonomy database of the International Sequence Database Collaboration contains the names of all organisms that are represented in the sequence databases. Can be searched using SRS.

EBI ‘Protein’ databases

GO
The EBI's Gene Ontology consortium pages. GO is an international consortium of scientists with the editorial office based at the EBI.

GOA
Provides assignments of gene products to the Gene Ontology (GO) resource.

IntAct
IntAct is a protein interaction database and analysis system. It provides a query interface and modules to analyse interaction data.

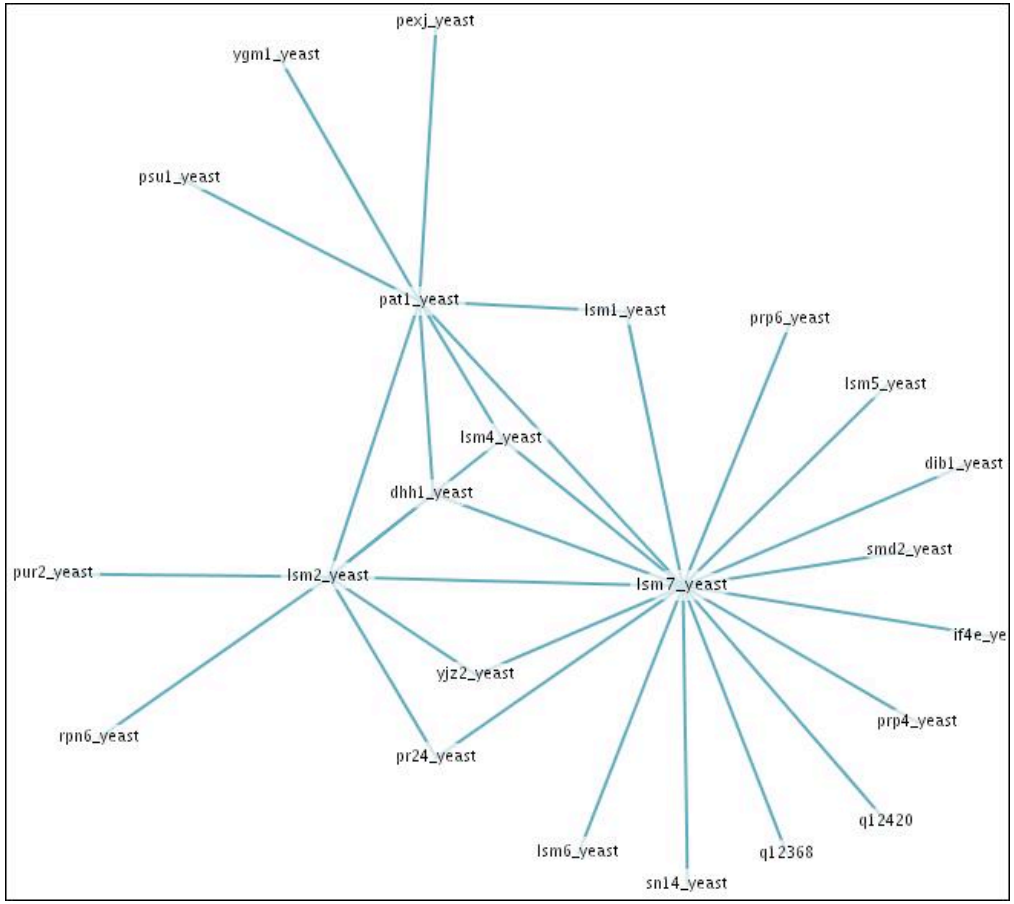
IntEnz
The Integrated relational Enzyme database (IntEnz) will contain enzyme data approved by the Nomenclature Committee. The goal is to create a single relational enzyme database.

Proteome Analysis
Statistical and comparative analysis of the predicted proteomes of fully sequenced organisms.

Reactome
A curated database of biological processes in humans. Reactome will not only be useful to general biologists as an online textbook of biology, but also to bioinformaticians for making new discoveries about biological pathways.

IntAct

- http://www.ebi.ac.uk/intact/index.jsp



Lecture summary

- ‘Putting it all together’ - Systems Biology
- Motivation
- Technological drivers
- Some biological background
- Introduction to some (systems biology) databases