LOSCHMIDT LABORATORIES

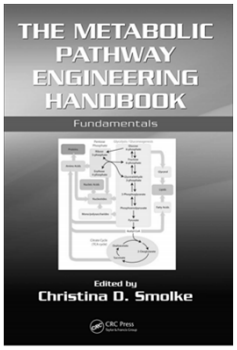
Metabolic Engineering I

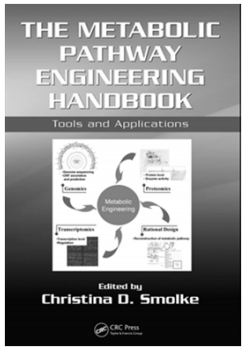
Bi7430 Molecular Biotechnology

Outline

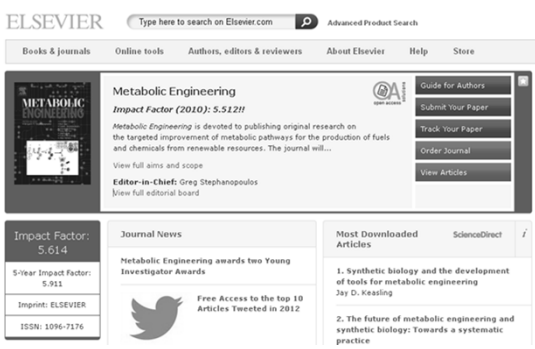
- Recommended sources for study
- Introduction to the metabolic engineering (ME)
- Suitable host organisms for ME
- General workflow of ME project
- Mathematical modelling of metabolic pathways
- Discussion

Sources for study: books

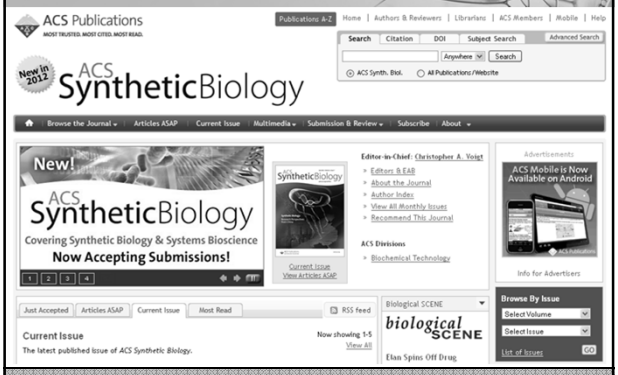




Sources for study: journals



Sources for study: journals



Sources for study: web

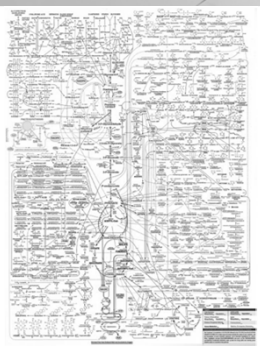
Greg Stephanopoulos (MIT)

Jay D. Keasling (Berkeley)

James C. Liao (UCLA)

Huimin Zhao (University of Illinois)

catabolic pathways: **biodegradation**



anabolic pathways: **biosynthesis**

IUBMB-Nicholson Metabolic Pathways Chart (Sigma-Aldrich): ATP metabolism in mitochondria and chloroplast.

Introduction: a bit of history

"...the improvement of cellular activities by manipulations of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology."

Toward a Science of Metabolic Engineering

JAMES E. BAILEY

Application of recombinant DNA methods to restructure metabolic networks can improve production of metabolite and protein products by altering pathway distributions and rates. Recruitment of heterologous proteins enables extension of existing pathways to obtain new chemical products, alter posttranslational protein processing, and degrade recalcitrant wastes. Although some of the experimental and mathematical tools required for rational metabolic engineering are available, complex cellular responses to genetic perturbations can complicate predictive design.

genetic manipulations is most effective in accomplishing a desired change in cellular function.

Recruiting Heterologous Activities for Strain Improvement

Cloning and expression of heterologous genes can serve several useful purposes, including extending an existing pathway to obtain a new product, creating arrays of enzymatic activities that synthesize a novel structure, shifting metabolite flow toward a desired product, and accelerating a rate-determining step. Introduction of a functional heterologous enzyme or transport system into an organism can result in the appearance of new compounds that may subsequently

Bailey , J.E. (1991) *Science*, 252:1668

Introduction: definition of ME

- **ME is the practice of optimizing genetic and regulatory processes within cells to increase the cells' production of a certain substance.[1]**

  (or degradation of certain substance)

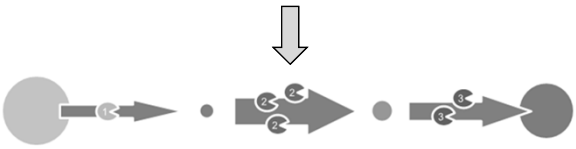
- These processes are series of **biochemical reactions** that allow cells to convert raw materials into molecules necessary for the cell's survival.
- MEs **mathematically model** these reactions, calculate a yield of useful products, and determine the **constraints** for the production of these products.
- MEs use **computational and experimental tools** to overcome these constraints and establish **cost effective process. Maximum yield** of desired substance must be balanced with the natural survival **needs of the cell**.

[1]Yang, Y.T. (1998) *Electronic Journal of Biotechnology*, ISSN 07117-3458

pathway

Boyle, P.M. (2012) *Metabolic Engineering*, 14:223

definition of the problem



solution of the problem

Introduction: characteristics

MULTIDISCIPLINARITY:

- bioinformatics, microbiology, molecular biology, biochemistry, genetics, mathematics... + **common sense (team work!!!)**

COMPLEXITY:

- knowledge of behavior of the entire metabolic pathway(s) in the **context of living organism** (systems biology, all "omics" techniques)

SUSTAINABILITY:

- replacement of fosil fuels by utilization of **renewable resources** in environmentally friendly processes
- **biosynthesis** of value-added chemicals (drugs) and biofuels
- **biodegradation** of toxic chemicals and bioremediation of polluted sites

ME in context of recent world

- **global chemistry market** estimated at **2,292 billion US$**[1]
- **industrial biotechnology** market estimated only **50 billion US$** (does not include pharmaceutical biotechnology and biofuels)
- great **potential for growth** in following decades (up to 20% of global chemistry market could by covered by biotechnological products by 2020)

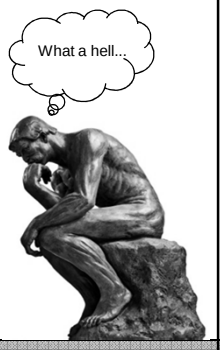
[1]Ghisalba, O. (2010) Industrial Biotransformation. *Encyclopedia of Industrial Biotechnology*.

Suitable host organisms for ME

DILEMMA TO BE SOLVED

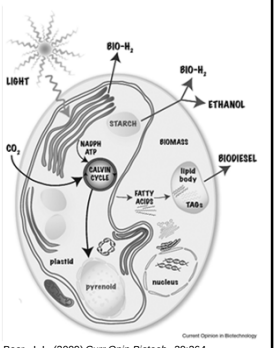
- microorganisms *vs*. multicellular organisms
- prokaryotes *vs*. eukaryotes
- heterotrophic *vs*. autotrophic organisms
- plants *vs*. animals
- *in vivo vs. in vitro*

PROS AND CONS
?



Algae

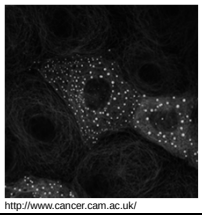
- very promissing for future
- production of H_2, biofuels (lipids) and biodegradable plastics (starch)
- microalgae (*Chlamydomonas reinhardtii*, *Volvox carteri*)
- **(+)** high efficiency in conversion of sun energy, thrive in salt water
- **(-)** lack of genetic tools (random mutagenesis, miRNA), low productivity



Beer , L.L. (2009) *Curr.Opin.Biotech*., 20:264

Mammalian cells

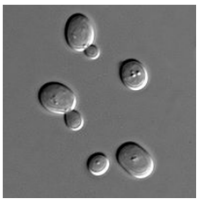
- 60 – 70% of recombinant protein biopharmaceuticals (MAbs - Herceptin)
- approved cell lines: CHO, BHK, HEK-293, NS0, PERC6 (fed-batch)
- ME focused on improvement of product titers (cell density) and quality
- **(+)** posttranslational modifications, human-like system, metabolic diversity
- **(-)** high complexity and sensitivity (apoptosis), slow growth, cumulation of by-products (ammonia, lactate), nutrition requirements



http://www.cancer.cam.ac.uk/

Yeast

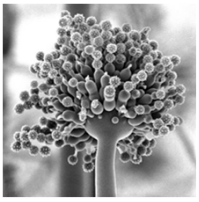
- baker's yeast *Saccharomyces cerevisiae,* the oldest and best known host for biotechnology (bread, wine, beer, ethanol)
- food industry, biofuel production (fermentation products ethanol and glycerol)
- starch and cellulose utilization, production of lactic acids, terpenoids etc.
- **(+)** known genome (13 Mb), eukaryotic microbe (posttranslational modif., cheap cultivation), secretion, number of genetic tools available
- **(-)** recombinant strains not accepted by public (GMO in food), cumulation of by-products



http://en.wikipedia.org/

Fungi

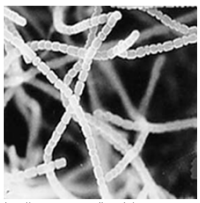
- used for thousands of years in traditional biotechnologies (koji fungi in Japan)
- *Aspergillus niger* (citric acid), *A. oryzae* (α-Amylase), *A. terreus* (statins)
- penicillin, heterologous enzymes, conversion of biomass to commodity chemicals (outlook)
- **(+)** low pH tolerance (potential for production of organic acids), metabolic diversity
- **(-)** lack of genetic tools, formation of by-products, complex protein processing



http://deferre.blogspot.cz/

Streptomyces

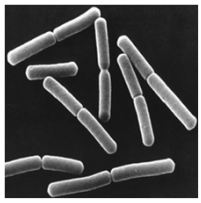
- G+ bacteria with mycelial habit
- producers of antibiotics (50 %), anticancerous agents (bleomycin), enzymes (degradation of cellulose)
- **(+)** non-pathogenic, rich in secondary metabolites, protein excretion, expression of genes with high GC content, cheap cultivation, degradors, decomposers
- **(-)** mycelial growth limits mass transfer, lack of genetic tools (random mutagenesis and screening = not GMO), large genomes with many regulatory proteins



http://streptomyces.nih.go.jp/

Bacillus subtilis

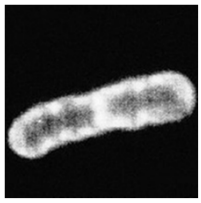
- G+ bacterial model, sporulation
- production of antibiotics, vitamins (H,B1, B2, B5, B6), food enzymes (50% of world market, serine proteases)
- **(+)** non-pathogenic**,** known genome (4.2 Mb), established genetic tools, secretion of enzymes, experience with large-scale fermentations
- **(-)** high complexity of the secretion processes, no endogenous plasmids and instability of recombinant ones – chromosomal expression preffered



http://www.astrobio.net/

Pseudomonas putida

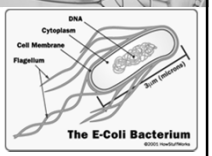
- G-, *P. putida* KT2440, saprophytic soil bacterium
- best characterized *Pseudomonas*, model organism for bioremediation applications and construction of bacterial chassis
- TOL plasmid - natural ability to degrade solvents (toluene, xylene), design of pathways for biodegradation of organic pollutants
- **(+)** non-pathogenic, known genome (6.18 Mb), metabolic versatility
- **(-)** so far less popular bacterial model than *E. coli*



http://www. bacmap.wishartlab.com

Escherichia coli

- G-, (*E. coli* K-12) most important bacterial model organism ("workhorse")
- **(+)** rapid growth on simple synthetic media, fast doubling time (20-30 min), high cell densities, high product content (up to 30% of dry cell mass) well known genome (4.6 Mb), diverse genetic tools, developed cultivation strategies, industrial (and public) acceptance
- **(-)** no glycosilation (and some other pt. modifications), acetate production at high glucose fluxes, limited protein secretion, thorough aeration needed



http://www.telegraph.co.uk/

E. coli..a bit of history

discovery of plasmids

discovery of restriction enzymes and ligases

first recombinant DNA

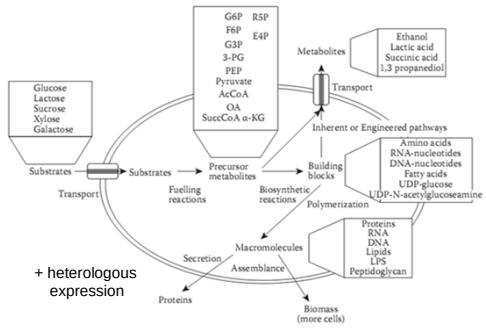
commercial production of insulin (Genentech)

commercial production of AA (Thr, Phe, Trp), **by ME**

commercial production of 1,3-propanediol (Genencor, Du Pont), **by ME**

Smolke, C.D. (2009) *The Metabolic Engineering Handbook: Fundamentals*

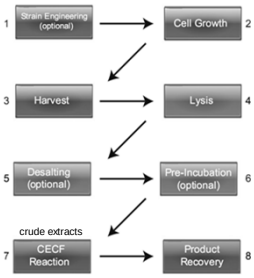
E. coli: biosynthetic potential



Smolke, C.D. (2009) *The Metabolic Engineering Handbook: Fundamentals*

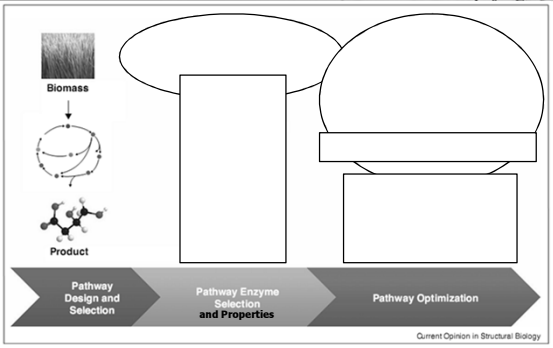
In vitro ME

- cell-free systems, promissing area in ME and synthetic biology
- work with CFE or purified components
- ***in vivo* systems are too complex** (1,3-propanediol production in *E. coli,* 15 years, 575 scientists)
- **(+)** lower complexity, better defined systems, no membranes (faster transport), higher flexibility, no toxicity problems, no GMO
- **(-)** cells needed anyway, lower enzymes' concentrations than in cell, recycling?



Hodgman, C.E. (2011) *Metabolic Engineering*, 14:261

General workflow of ME project



Dhamankar , H. (2011) *Curr.Opin.Struct.Biol.*, 21:1

Pathway design and selection

What is the main goal of the project?

ME can be applied to:

- **improve the yield** and productivity of native products synthesized by organisms
- **extend the range** of substrates or improve the uptake of substrate (including biodegradation)
- **establish production** of products that are new to the host cell

Pathway design and selection

Identification of pathway of interest

Databases of anabolic and catabolic pathways

- biosynthesis: MetaCyc, KEGG
- biodegradation: UM-BBD Pathway Prediction System (183 pathways)

Literature search

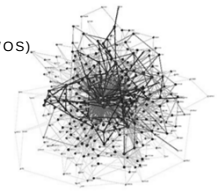
- NCBI Pubmed database, Web of Science (WOS)

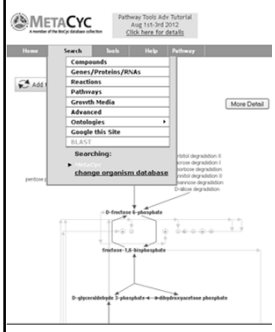
Visualization of metabolic networks

- Cytoscape, GraphViz, Systrip

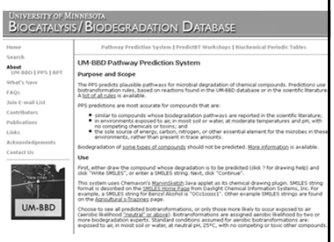
Visualization of reaction networks

- CellDesigner, GLAMM





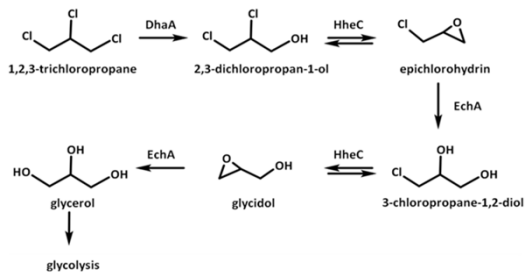
http://metacyc.org/



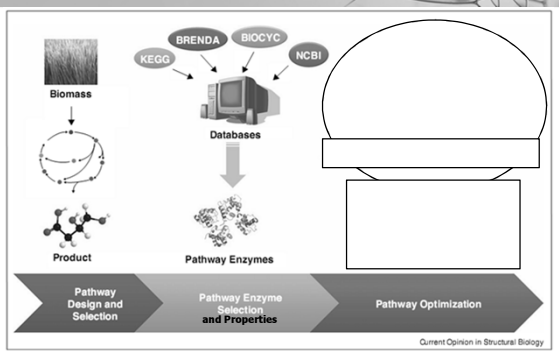
http://umbbd.ethz.ch/predict/

Our model pathway

Synthetic pathway for biodegradation of 1,2,3-trichloropropane (TCP)



General workflow of ME project



Dhamankar , H. (2011) *Curr.Opin.Struct.Biol.*, 21:1

Pathway enzyme properties

Enzyme databases:

- BRENDA, BioCyc, KEGG

Enzyme properties:

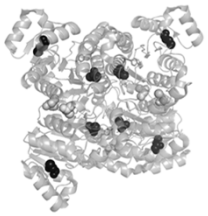
- BRENDA, ExPASy server

Searching for sequence homology:

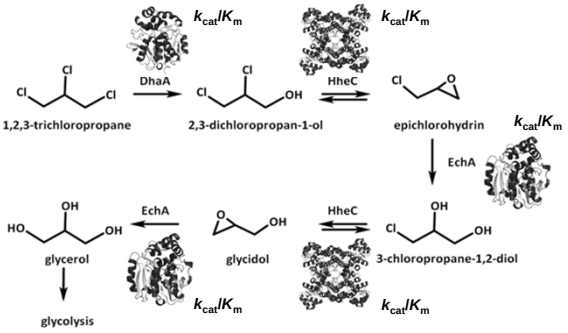
- BLAST (NCBI) – basic local alignment

Enzyme structure and its visualization

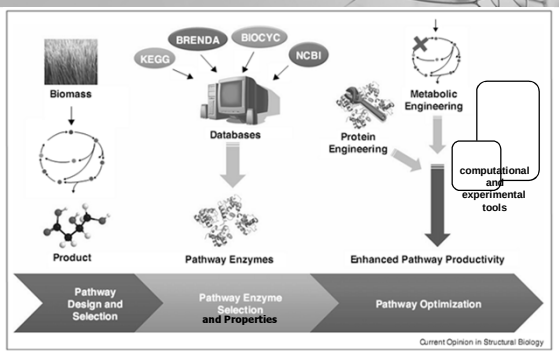
- Protein Data Bank (PDB), PyMOL



+ physical/chemical properties of metabolites



General workflow of ME project



Dhamankar , H. (2011) *Curr.Opin.Struct.Biol.*, 21:1

Pathway optimization

Original "historic" approach:

- combination of **random mutagenesis** with exhaustive **screening** of candidates with improved production of desired compound
- chemical (e.g. ethidium bromide) and physical (UV irradiation) mutagens, "mutator strains" (*Epicurian coli* XL1-Red)
- **(+)** success achieved in production of AA, antibiotics, vitamins.
- **(-)** toxic agents, resulting variants with more undefined mutations, demanding screening

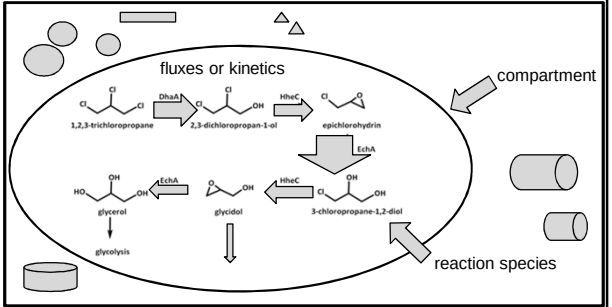
Pathway optimization

Recent strategies try to be more rational and focused.

Understanding of cellular metabolism requires the knowledge of:

1) The **topology** of the metabolic network (**where** - **compartments**)
2) The **concentrations** of metabolites (**what and how much** - **species**)
3) The **flow** of metabolites through pathways (**how fast** – **fluxes or kinetics of enzymes**)
4) Overall **context** of the living cell (**"-omics" techniques**)

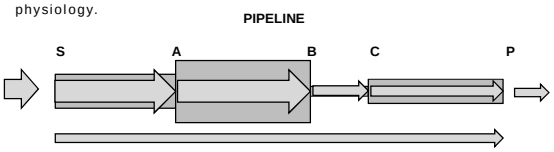
Study of 1, 2, 3 and 4 requires different theoretical and experimental methods and tools.



whole-cell environment (studied by "-omics" techniques)

Pathway flux

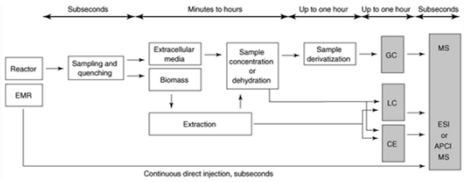
- **FLUX is the reaction rate connecting two metabolites (A→B) or rate at which material is processed through the whole pathway,** with unit [mol.dm^{-3}.s^{-1}] .[1]
- Fluxes are calculated in **steady-state, DO NOT describe dynamics** of enzymatic reactions (**kinetic parameters** of enzymes needed)
- Along with intracellular **metabolite concentrations**, fluxes define the minimal information needed for describing metabolism and cell physiology.



[1]Stephanopoulos , G. (1999) *Metabolic Engineering*, 1:1

Metabolomics

- Metabolomics completes the information from flux analysis with information about **concentrations of metabolites**.
- **METABOLOME** is a quantitative set of all molecules present in the cell at certain time.
- Experimental analytical techniques: **GC-MS, LC-MS, CE-MS, NMR**.



Meyer, A. (2007) *Current Opinion in Microbiology* 10:246

Mathematical models of MPs

- Mathematical models of metabolic pathways play a **central role in metabolic engineering**.
- Models help to **identify reactions which need to be modified** to improve the performance of the pathway.
- Once such reaction(s) is (are) identified, **experimental techniques** are applied in order to **target corresponding gene(s) or regulatory mechanisms**.
- **There are three types of modelling of metabolism**

1) Flux Balance Analysis (chemostat)

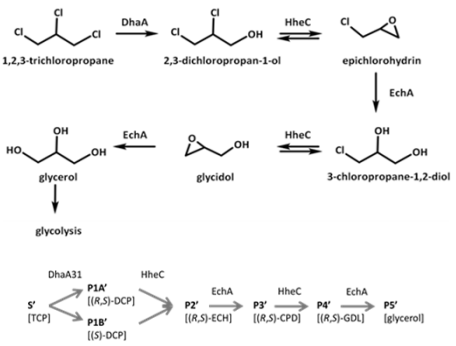
2) Metabolic Flux Analysis (chemostat)

3) Kinetic modelling (real conditions)

Kinetic modelling

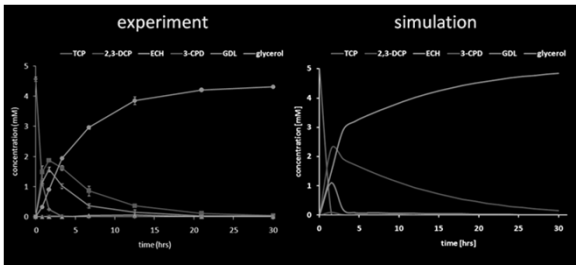
- Kinetic models employ kinetic parameters of pathway enzymes to **describe dynamic behavior** of metabolic pathway.
- **Parameters**: concentrations of enzymes, K_m, k_{cat} or V_{max}, K_i, K_d, (BRENDA database).
- **Kinetic equations:** e.g. Michaelis-Menten mechanism
- **Application for simulation** of the pathway reaction courses **and prediction** of pathway behaviour
- Computing tools: E-Cell, COPASI, CellDesigner, Scientist, Matlab
- **(+)** the most realistic, dynamic description of the system.
- **(-)** missing kinetic data for majority of enzymes, kinetic parameters often measured *in vitro* in optimal (not physiological) conditions

Kinetic modelling: example



Kinetic modelling: example

Modelling of TCP pathway.



Summary

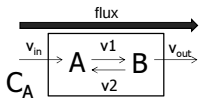
KEY POINTs of ME project

- **selection of suitable host organism**
- **thorough selection/design of the pathway**
- **maximal knowledge of basic building blocks (enzymes and metabolites) of target MP**
- **mathematical modelling and metabolomics play a key role in ME for definition of problematic reactions**
- **genetic tools are used for solution of the problem (Lecture 2)**

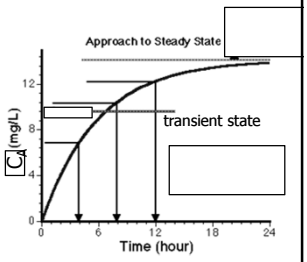
Discussion

Appendices

Pathway flux



- under continuous feed of A, the fluxes become stationary after some time
- MEs are interested in **steady-state fluxes** (disadvantage)
- in steady-state d[A]/dt = 0



http://www.boomer.org/ (modified)

Mathematical models: SBML

SBML (Systems Biology Markup Language)

- Standard (common) language used for sharing mathematical models describing metabolic networks (+ other biological networks, systems and processes).
- **Structure of SMBL model**
  - list of **compartments** (*in vivo* vs. *in vitro*)
  - list of **species** involved in reaction
  - list of **reactions**: for each reaction
    - list of **reactants**
    - list of **products**
    - **kinetic law** (list of parameters)

Mathematical models: SBML

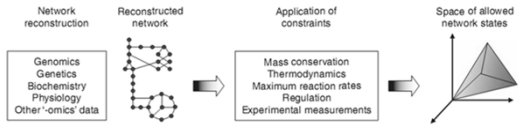
```
<?xml version="1.0" encoding="UTF-8" ?>
<sbml xmlns="http://www.sbml.org/sbml/level2/version4" xmlns:celldesigner="http://www.sbml.org/2001/ns/celldesigner" level="2" version="4">
- <model metaid="in_vitro" id="in_vitro">
  - <annotation>
    - <celldesigner:extension>
        <celldesigner:modelVersion>4.0</celldesigner:modelVersion>
        <celldesigner:modelDisplay sizeX="600" sizeY="400" />
        <celldesigner:listOfCompartmentAliases />
        <celldesigner:listOfComplexSpeciesAliases />
      - <celldesigner:listOfSpeciesAliases>
        - <celldesigner:speciesAlias id="sa1" species="s1">
            <celldesigner:activity>inactive</celldesigner:activity>
            <celldesigner:bounds x="2.0" y="119.5" w="70.0" h="25.0" />
            <celldesigner:font size="12" />
            <celldesigner:view state="usual" />
          - <celldesigner:usualView>
              <celldesigner:innerPosition x="0.0" y="0.0" />
              <celldesigner:boxSize width="70.0" height="25.0" />
              <celldesigner:singleLine width="1.0" />
              <celldesigner:paint color="ffccff66" scheme="Color" />
            </celldesigner:usualView>
          - <celldesigner:briefView>
              <celldesigner:innerPosition x="0.0" y="0.0" />
              <celldesigner:boxSize width="80.0" height="60.0" />
              <celldesigner:singleLine width="1.0" />
              <celldesigner:paint color="3fff0000" scheme="Color" />
            </celldesigner:briefView>
            <celldesigner:info state="empty" angle="-1.5707963267948966" />
          </celldesigner:speciesAlias>
        - <celldesigner:speciesAlias id="sa2" species="s2">
            <celldesigner:activity>inactive</celldesigner:activity>
            <celldesigner:bounds x="131.0" y="121.5" w="70.0" h="25.0" />
            <celldesigner:font size="12" />
            <celldesigner:view state="usual" />
          - <celldesigner:usualView>
```

FBA and MFA

- Used for calculation of intracellular **fluxes**. Employ information about intracelullar **concentrations of metabolites** (**^{13}C** labelling)**.**
- Results in proposals for **deletion** (knock-outs) or **overexpression** of certain gene(s) coding for pathway enzymes.
- **(+)** can deal with **genome-scale metabolic networks**, does not require the knowledge of enzyme kinetics
- **(-)** fluxes represent only **static approximation** of dynamic and complex reality in living cell at certain conditions set up by experimenter. Reliable *in vivo* metabolic models are still rare.

Flux Balance Analysis

- FBA is completely **theoretical concept.**[1]
- It is a direct application of linear programming to biological systems that uses the **stoichiometric coefficients** for each reaction in the system as the set of **constraints** for the optimization. It simply requires that the total flux of any compound in the system is 0.
- **computational tools**: CellDesigner, MetNetMaker, COBRA toolbox (requires MATLAB), models in SBML format

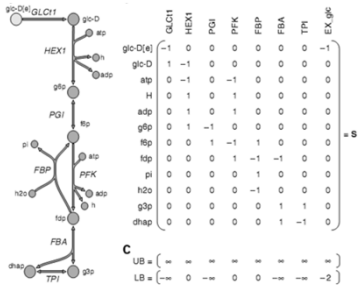


Becker, S.A. (2007) *Nature protocols* doi:10.1038/nprot.2007.99

[1] see Wikipedia for detailed description

Flux Balance Analysis

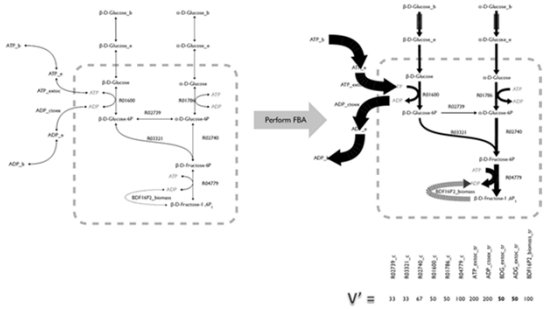
An example of stoichiometric matrix for a network representing the top of glycolysis prepared for FBA.



Becker, S.A. (2007) *Nature protocols* doi:10.1038/nprot.2007.99

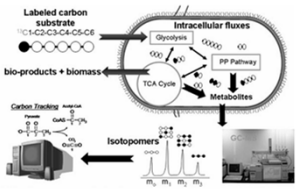
Flux Balance Analysis

The results of FBA can be represented identically as a vector of fluxes, or by weighting the lines representing the reactions according to the flux they carry.



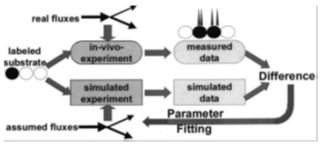
http://en.wikipedia.org/

Metabolic Flux Analysis



^{13}C MFA workflow

Tang, Y.J. (2009) *Mass Spectrometry Reviews* 28:362



Wiechert, W. (2001) *Metabolic Engineering* 3:195