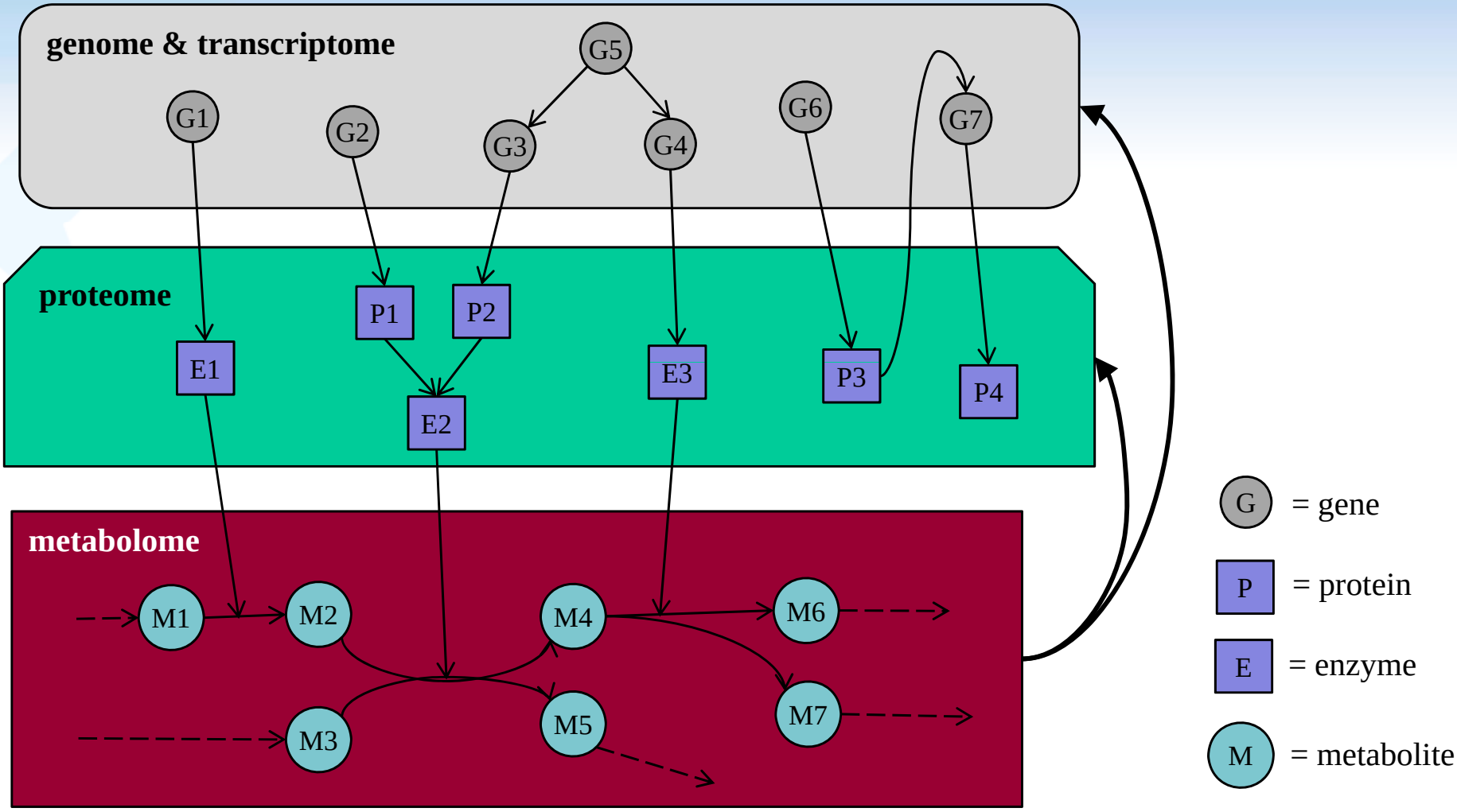
Stoichiometric analysis of metabolic networks

SGN-6156 Computational systems biology II

Antti Larjo
antti.larjo@tut.fi
Department of Signal Processing
Computational Systems Biology group
30.4.2008

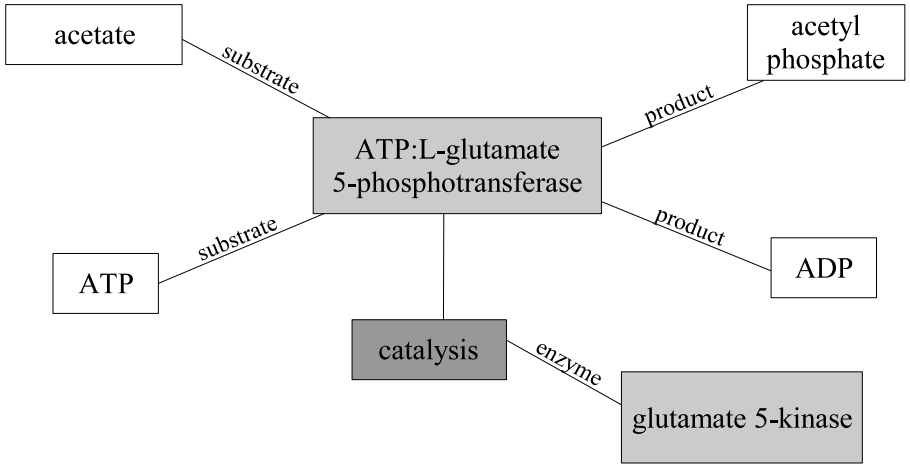
Contents

Networks in cells



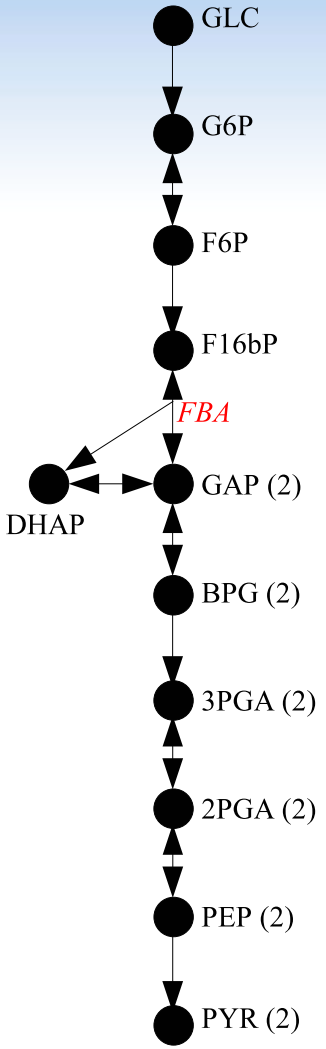
Metabolism

- “Sum of all the chemical reactions that take place in every cell of a living organism, providing energy for the processes of life and synthesizing new cellular material.” (Encyclopædia Britannica)
- ≈ conversion of food to products/biomass/work/warmth/…
- divided into catabolism (destructive) and anabolism (constructive)
- substances are called metabolites
- catalysts are called enzymes (needed by practically every reaction)
- certain parts very similar between organisms, evolved from the same ancient pathway
- structure very constant, but can be changed e.g. by evolution / genetic changes
- only part used at a time
  → metabolic phenotypes



Metabolic networks and pathways

- metabolism forms a network of interconnected metabolites and reactions
- known very well for many organisms
  - reconstructions of the whole-cell (genome-wide) metabolism
  - e.g. *Saccharomyces cerevisiae* 646 metabolites, 1149 reactions [1]
- pathways or networks?
  - metabolic network: set of metabolites connected by reactions, consists of pathways
  - pathway: systems of successive chemical reactions, “set of oriented reactions interacting under given physiological conditions via simple or apparently simple intermediates” [2]
  - pathways sometimes defined by function / topology / …
    → often subjective
  - objective definition later



[1] Duarte, N. C, Herrgård, M. J., and Palsson, B. O., “Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model,” *Genome Research*, 14(7), 1298-1309, 2004.

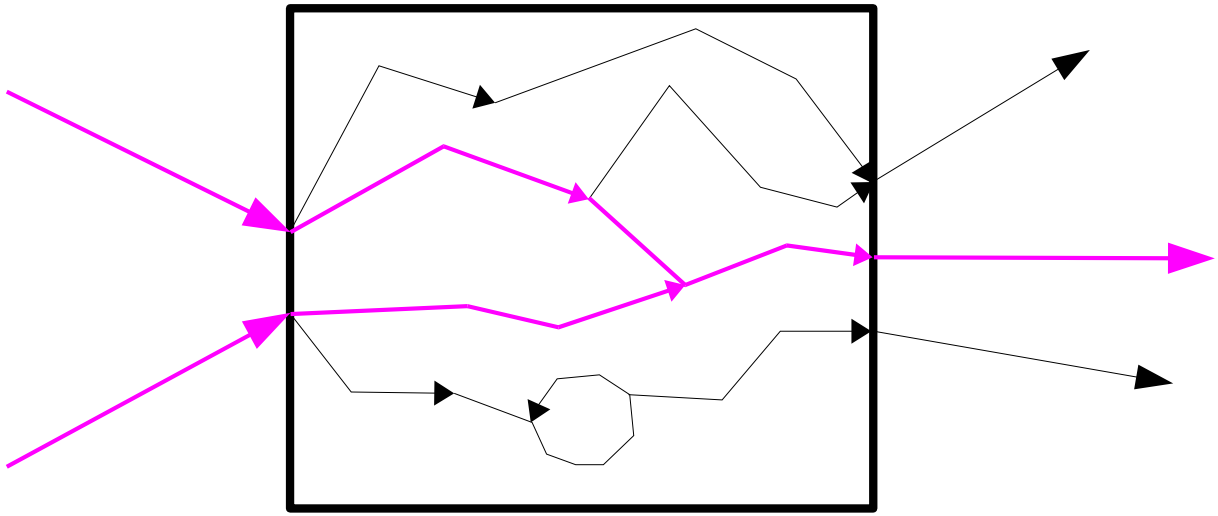
[2] Selkov, E. Jr, Grechkin, Y., Mikhailova, N., and Selkov, E., “MPW: the metabolic pathways database,” *Nucleic Acids Research*, 26(1), 43-45, 1998.

Dummy model of metabolism

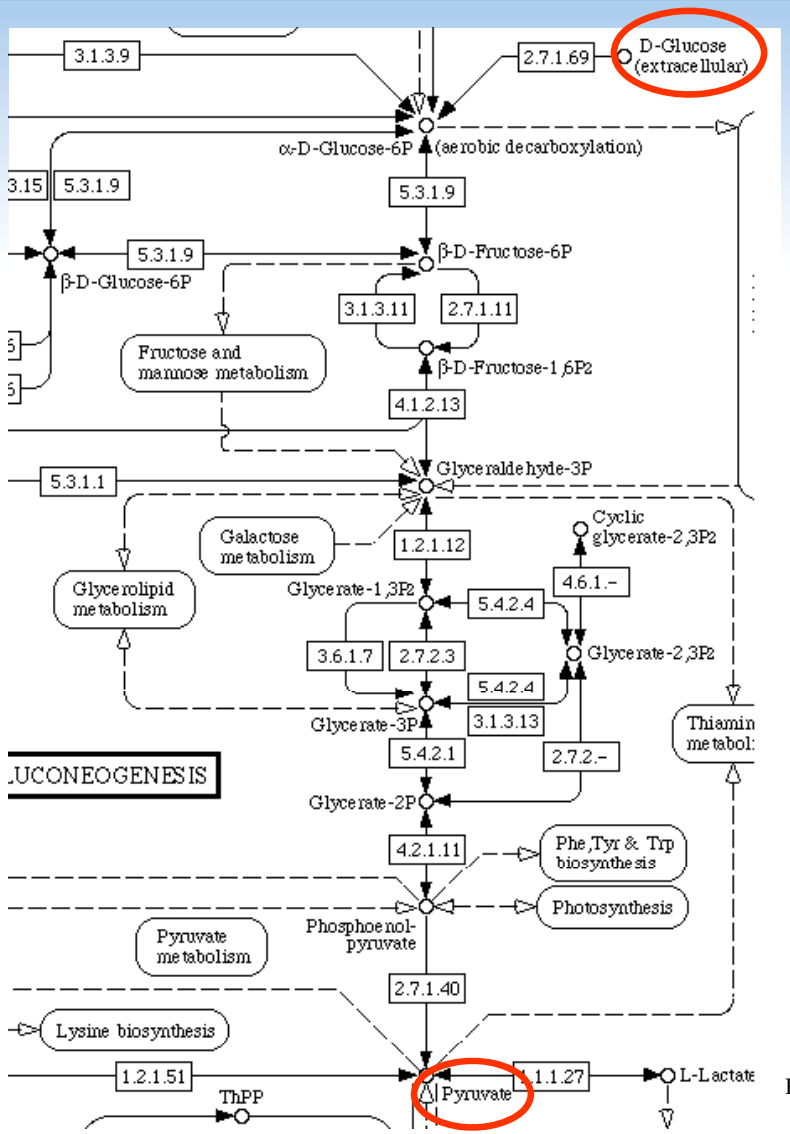
- input: substrates (food, e.g. glucose, oxygen, …)
- output: products (biomass, waste, energy, …)

Inside the black box

- input substances follow conversion routes (pathways) to outputs
- possible routes determined by enzymatic capabilities, reaction directionalities
- used (=active) routes depend on available substrates, enzymes, also other things



Glycolysis



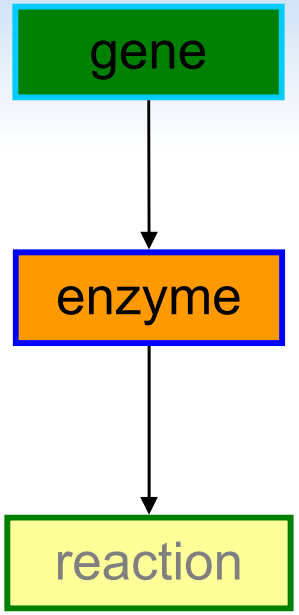
From KEGG (http://www.genome.jp/kegg/)

Reconstruction of metabolic networks

- central dogma of molecular biology: gene → protein
- databases
  - gene annotation
  - biochemical information
  - publications, other databases
- identification of enzyme(=protein) coding genes

→ list of reactions

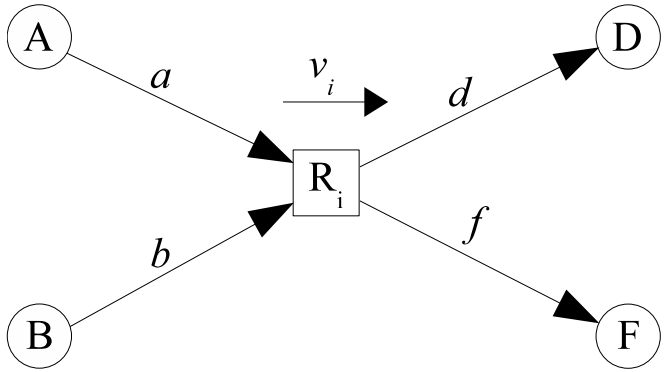
- reconstruction of metabolic network
  - result is a structural (stoichiometric) *in silico* model
- models available from the Internet
  - KEGG (www.genome.jp/kegg)
  - MetaCyc (metacyc.org)



Stoichiometry and fluxes

- "Determination of the proportions (by weight or number of molecules) in which elements or compounds react with one another." (Encyclopædia Britannica)

- stoichiometric coefficients
  - elementary and charge balance
  - constant
  - known for every discovered metabolic reaction
  - definition of direction for reversible reactions



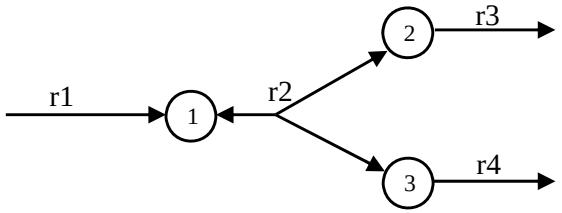
- flux: rate of flow of particles
  - note: not the same as reaction rate (velocity)
  - e.g. the flux from metabolite A to reaction *i* is av_i

Stoichiometric matrix

- systems of several reactions described with stoichiometric matrix
  - rows correspond to metabolites, columns to reactions
  - $2A + B \rightarrow 3C + D$ => $(-2, -1, 3, 1)^T$
- structure of metabolic network defined by stoichiometric matrix S and reversibilities of reactions

- s_{ij} is the stoichiometric coefficient of metabolite *i* in reaction *j*

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$



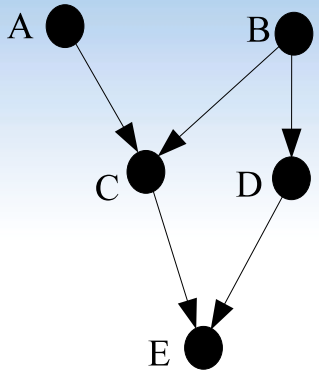
Different metabolic network models

Reaction list:

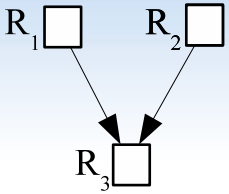
R_1: A + B → C

R_2: B → D

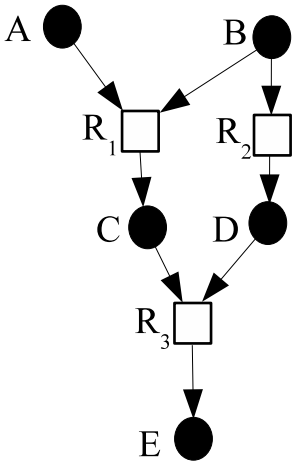
R_3: C + D → E



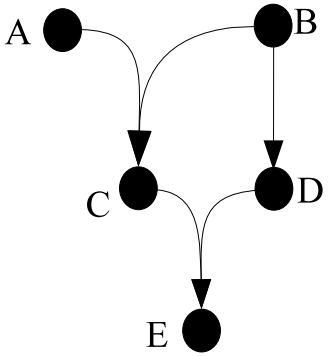
Compound graph



Reaction graph



Bipartite graph

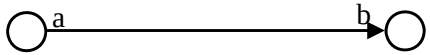


Hypergraph

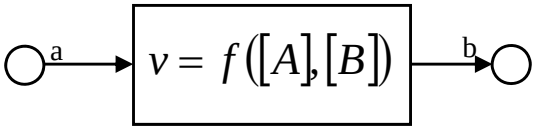
Levels of modeling metabolic networks

- structural
  - only connections between metabolites

- stoichiometric
  - proportions of needed metabolites



- kinetic
  - dynamics of reactions



- regulatory
  - effects of e.g. gene regulation

Modeling the system boundary

- internal metabolite
  - inside the system boundary
  - factors affecting the concentration are included in the system
- external metabolite
  - some factors affecting the concentration not known or excluded from the system
  - concentration assumed constant
  - also called sources or sinks

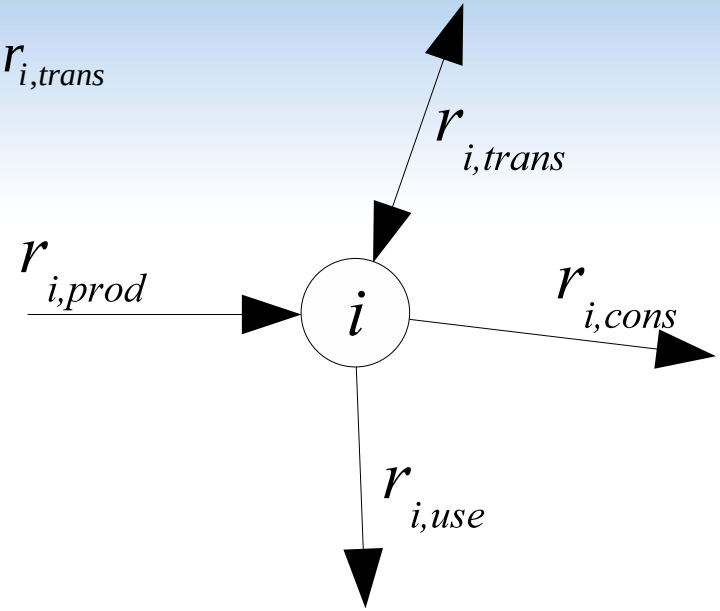
- internal fluxes
  - fluxes whose both sides are inside the system
- exchange flux
  - flux capable of transferring material across the system boundary
  - practically the same as a flux going to an external metabolite

Metabolic pathway analysis

- finding a single flux distribution
  - optimal: Flux balance analysis (FBA)
  - suboptimal: Minimization of metabolic adjustment (MOMA)
- determining all the conversion routes (=pathways)
  - Elementary (flux) modes (EM / EFM)
  - Extreme pathways (EP)
- measuring internal fluxes
  - ^{13}C-labeling

Steady-state

- rate of accumulation[1] $\frac{dc_i}{dt} = r_{i,prod} - r_{i,cons} - r_{i,use} \pm r_{i,trans}$
  - c_i concentration, r_i rates
- dynamic mass balance equation $\mathbf{Sv} = \frac{d\mathbf{c}}{dt} = \mathbf{a}$
  - **S** is the stoichiometric matrix
  - **v** is the reaction rate vector
  - **c** is the concentration vector
  - **a** is the accumulation vector
- **a=0**
  - no accumulation = steady-state (mass balance / flux balance)
  - long time scales
  - large cell populations → average cell state



[1] Schilling, C. H., Edwards, J. S., and Palsson, B. O., “Toward metabolic phenomics: analysis of genomic data using flux balances,” *Biotechnol. Prog.*, 15, 288-295, 1999.

Metabolic networks in steady-state

- arrange reaction rate vector $\mathbf{v} = \begin{bmatrix} \mathbf{v}_{rev} \\ \mathbf{v}_{irr} \end{bmatrix}$
  - $\mathbf{v}_{rev}$: rates of reversible reactions
  - $\mathbf{v}_{irr}$: rates of irreversible reactions
- arrange the columns of stoichiometric matrix **S** (dimension *mxr*) accordingly

- the set of vectors **v** satisfying the steady-state condition $\mathbf{Sv} = \mathbf{0}$ is given as the null-space $\mathbf{K} = null(\mathbf{S})$
  - linear basis vectors
  - linearly independent, not unique
  - routes not necessarily minimal

- however, $\mathbf{v}_{irr} \geq \mathbf{0}$
  - defines half-spaces in the null-space
  - result is a convex polyhedral cone
  - convex analysis / polyhedral computation needed

Metabolic networks in steady-state (2)

convex polyhedral cone (flux cone) **F**

- convex combination $\forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{F},\ 0 \le \lambda \le 1 : \lambda \mathbf{v}_1 + (1-\lambda)\lambda \mathbf{v}_2 \in \mathbf{F}$
- $\forall \mathbf{v} \in \mathbf{F},\ \alpha \ge 0 : \alpha \mathbf{v} \in \mathbf{F}$

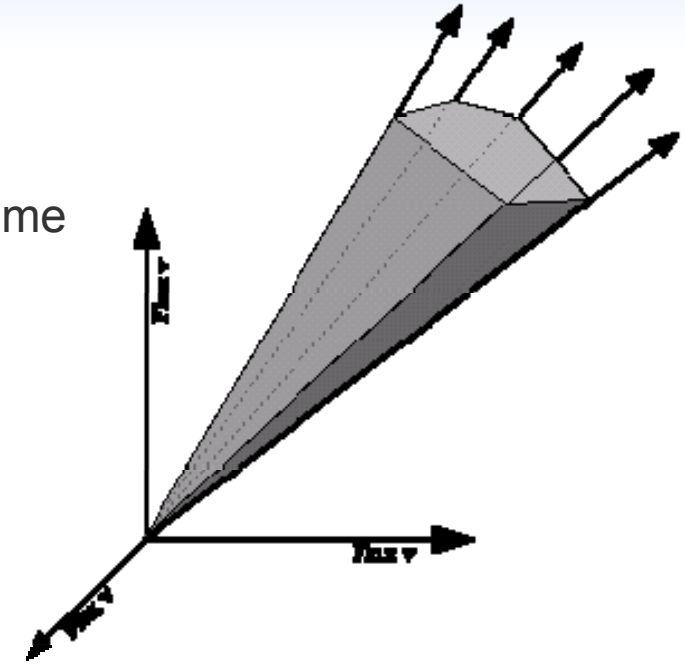
cone **F** defined by

- $\mathbf{Sv} = \mathbf{0}$ and $\mathbf{v}_{irr} \geq \mathbf{0}$ (*H-representation*) or
- combination of generating vectors (also called extreme rays) (*V-representation*)

$$\mathbf{F} = \left\{ \mathbf{v} \in \mathbb{R}^r \middle| \mathbf{v} = \sum_k \lambda_k \mathbf{f}_k + \sum_j \beta_j \mathbf{b}_j,\ \lambda_k, \beta_j \in \mathbb{R},\ \lambda_k \ge 0 \right\}$$

where $\mathbf{f}_k$ are the irreversible generating vectors and $\mathbf{b}_j$ are the reversible generating vectors

- generating vectors unambiguously define the cone
  - not necessarily linearly independent



Flux balance analysis

constraints-based modeling

•thermodynamical constraints (irreversibilities) $\}\alpha_i \leq v_i \leq \beta_i$

•maximal capacities of reactions

- from enzyme kinetics
- from measurements
- → sets a cap to the flux cone

•steady-state (flux balance) $\mathbf{Sv} = \mathbf{0}$

•environment (substrates)

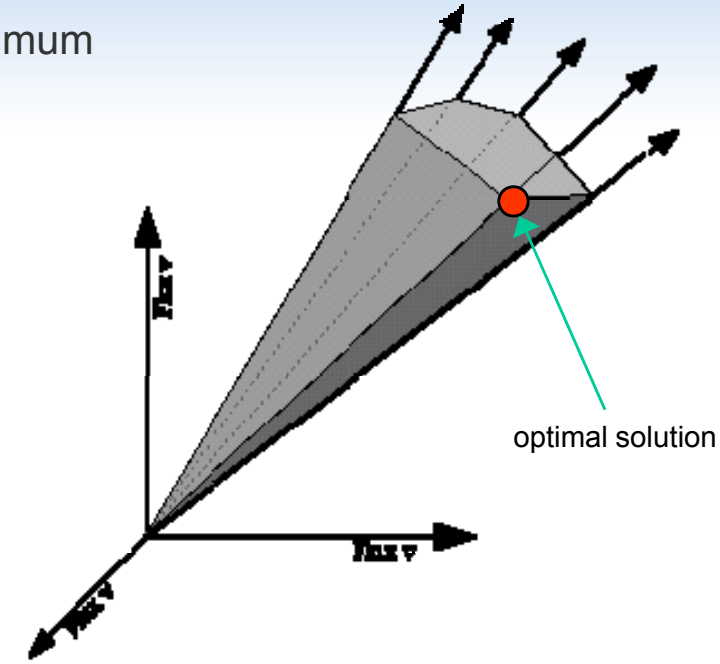
- i.e. definition of external metabolites or exchange fluxes

•the solution space

- defined by the constraints
- contains feasible states of the metabolic network

Flux balance analysis (2)

- objective function $Z = \mathbf{c} \cdot \mathbf{v} = \sum c_i v_i$
  - assumption: evolution drives the organism to the optimum
  - e.g. maximal growth
- maximize objective function
  - use linear programming
  - maximize Z subject to
    $\alpha_i \leq v_i \leq \beta_i$ and $\mathbf{Sv} = \mathbf{0}$
  - → optimal point
- optimal growth rates
- viability of knock-out mutants
- screening for possible drug targets
- consistent with experimental results[1]
- different choices of objective function
  - exploring organism's capabilities, guiding in metabolic engineering



[1] Edwards, J. S., Ibarra, R. U., and Palsson, B. O., "In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data," *Nature Biotechnology*, 19, 125 – 130, 2001.

Example biomass reaction

(1.1348) 13BDglcn + (0.4588) ala-L + (0.046) amp + (0.1607) arg-L + (0.1017) asn-L + (0.2975) asp-L + (59.276) atp + (0.0447) cmp + (0.0066) cys-L + (0.0036) damp + (0.0024) dcmp + (0.0024) dgmp + (0.0036) dtmp + (0.0007) ergst + (0.1054) gln-L + (0.3018) glu-L + (0.2904) gly + (0.5185) glycogen + (0.046) gmp + (59.276) h2o + (0.0663) his-L + (0.1927) ile-L + (0.2964) leu-L + (0.2862) lys-L + (0.8079) mannan + (0.0507) met-L + (0.000006) pa_SC + (0.00006) pc_SC + (0.000045) pe_SC + (0.1339) phe-L + (0.1647) pro-L + (0.000017) ps_SC + (0.000053) ptd1ino_SC + (0.1854) ser-L + (0.02) so4 + (0.1914) thr-L + (0.0234) tre + (0.000066) triglyc_SC + (0.0284) trp-L + (0.102) tyr-L + (0.0599) ump + (0.2646) val-L + (0.0015) zymst

→ (59.276) adp + (58.7162) h + (59.305) phosphate

From N. Duarte, M. Herrgård, and B. Palsson, "Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model," Genome Research, 14(7):1298-309, 2004.

FBA: Phenotype phase plane analysis

- FBA gives only particular solutions
- phenotype phase plane analysis
  - select two fluxes and calculate FBA as the function of these (by changing the values of α_i and β_i)
- dependency of optimal solution on some flux constraints

Minimization of Metabolic Adjustment (MOMA)

- optimal growth may be a good assumption for wild-type but not for knock-out mutants
  - not enough time & evolutionary pressure in lab
- alternative approximation: steady-state flux distribution responds minimally to perturbation
- denote by Φ^j the feasible space of mutant *j* and by $\mathbf{v}^{WT}$ the wild-type optimal solution (FBA)
- find vector $\mathbf{x} \in \Phi^j$ minimizing the Euclidean distance $D(\mathbf{v}^{WT}, \mathbf{x}) = \sqrt{\sum_{i=1}^{N} (v_i^{WT} - x_i)^2}$
- can be written as a standard quadratic programming (QP) problem $f(\mathbf{x}) = \mathbf{L}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$
- shows much higher correlation with measurement data than FBA[1]

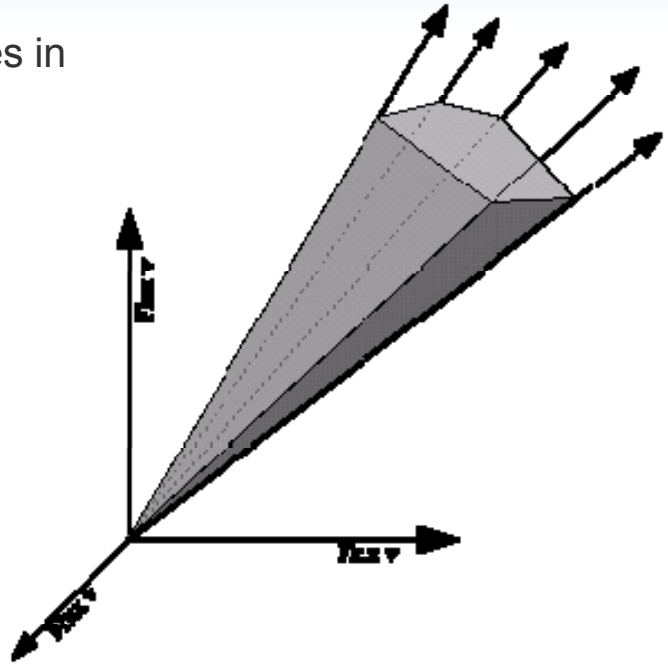
[1] D. Segrè , D. Vitkup, and G. M. Church, “Analysis of optimality in natural and perturbed metabolic networks,” *PNAS*, 99(23), 15112 – 15117, 2002.

Elementary modes

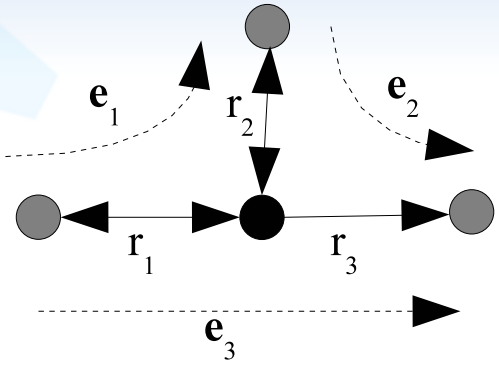
- **v** is an elementary (flux) mode if it fulfills
  - steady-state: **Nv = 0**
  - feasibility: $\mathbf{v}_{irr} \geq \mathbf{0}$
  - non-decomposability: setting any of the nonzero rates in **v** to zero will make the whole mode zero
- unique up to scaling
- not necessarily linearly independent
- all feasible states given as non-negative linear combinations of EMs

$$\mathbf{v} = \sum_j \alpha_j \mathbf{v}_j, \quad \alpha_j \geq 0$$

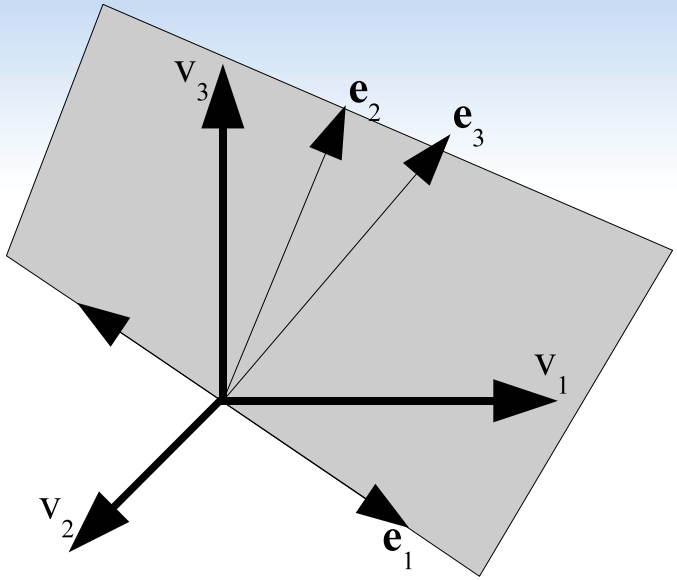
- problem: combinatorial explosion
  - computation difficult (impossible) for big networks
  - analysis of results cumbersome



Example of elementary modes

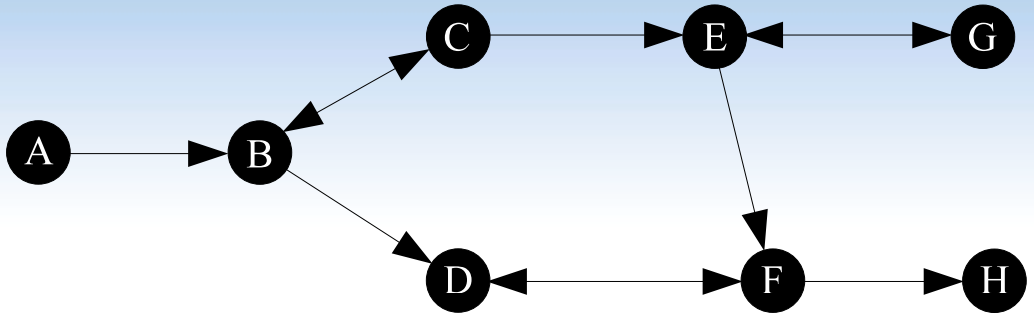


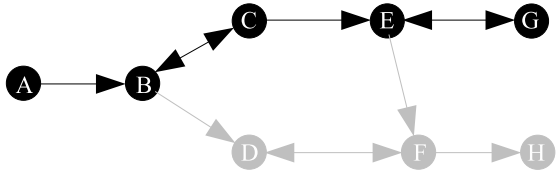
a)



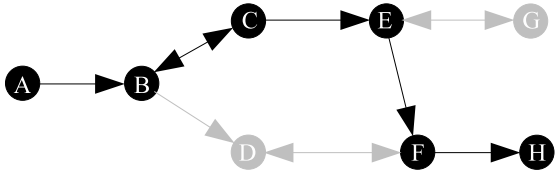
b)

Example of elementary modes (2)

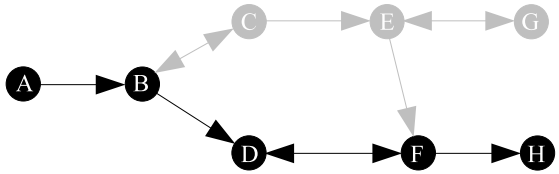




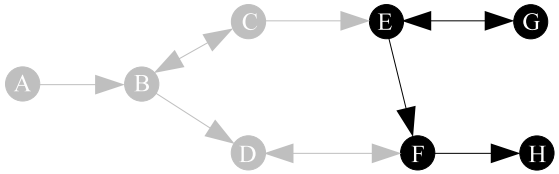
EM1



EM2



EM3



EM4

metabolites A, G, and H are external

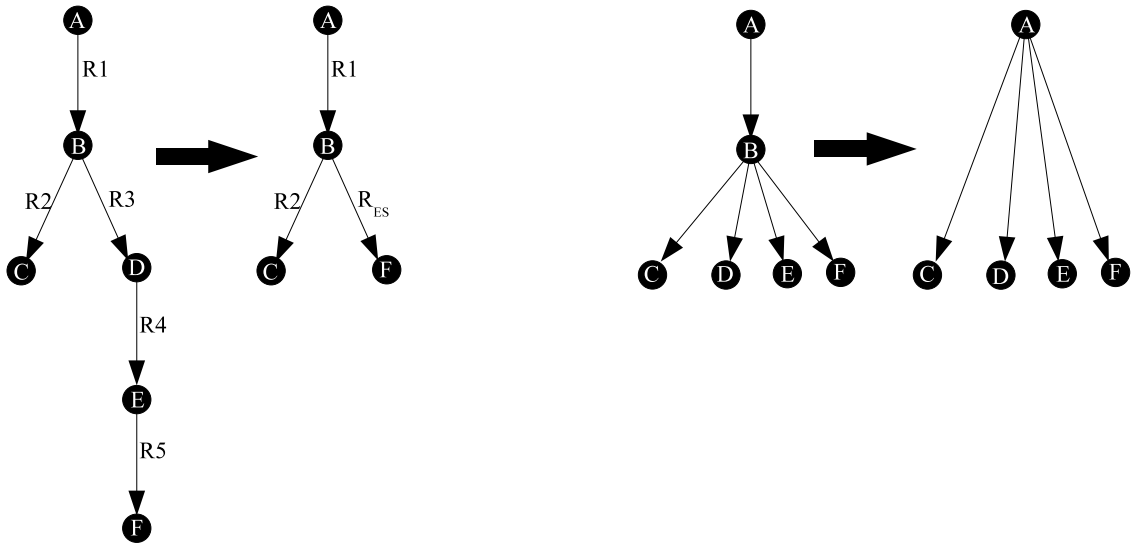
Extreme pathways

- minimal possible conversion routes
  - steady-state S**v=0**
  - network reconfiguration: split reversible internal fluxes into two irreversible fluxes
  - thermodynamic constraints $v_i \geq 0$
  - non-decomposability (minimality)
  - systemic independence
- unique set
- all other possible routes given as their linear combinations ≈ basis
- EPs are the minimal set of EMs needed to span the feasible steady-state
  - (proper) subset of elementary modes
- combinatorial explosion
  - whole-cell analysis practically impossible

Redundancy removal / network compression

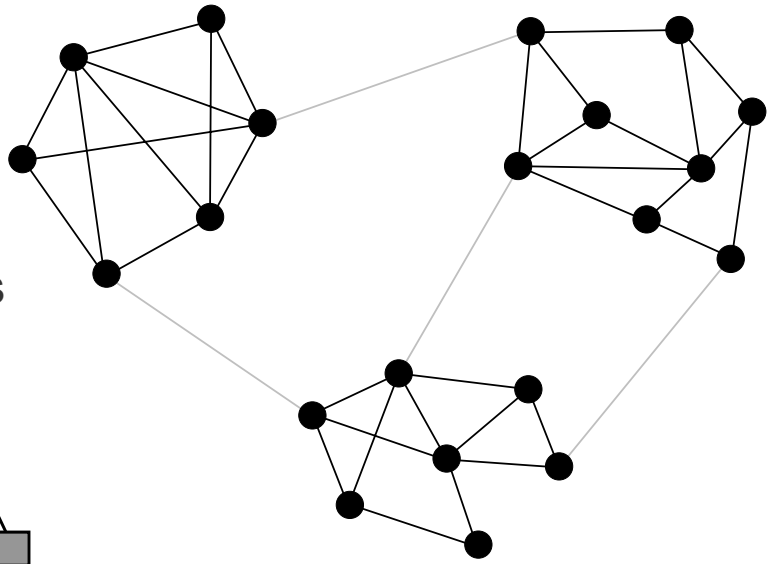
some methods can be used to alleviate the computational problems

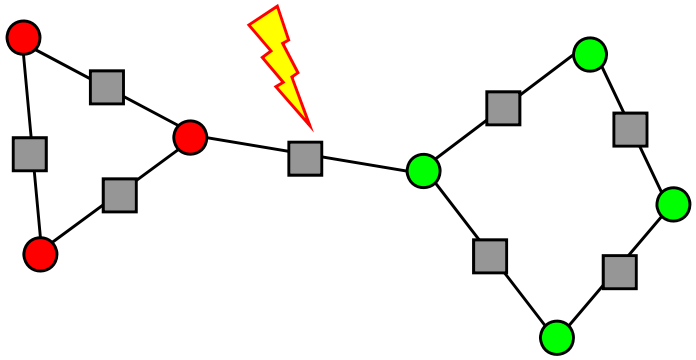
- enzyme subsets
  - reactions that necessarily operate together in steady-state
  - can be found from null-space matrix as the rows whose values are proportional to each other
- uniquely produced / consumed metabolites



Decomposition of metabolic networks

- allows a sort of divide-and-conquer approach
- reduces computational burden
- metabolic networks have modular structure
  - molecules → modules → networks → cells
- Girvan-Newman method
  - shortest paths for all pairs of nodes
  - edges between modules
- compute EMs for subnetworks
- combine EMs to yield whole-network EMs
- parallelization easy





EM / EP: Applications

- objective (mathematical) definition of pathways
  - however, dependent on classification of metabolites to external
- constraints, basis
  - every feasible steady-state given as linear combination of EMs / EPs
- optimal growth rates & maximal yields
- testing network models
  - model must be able to produce certain products from given substrates
- knock-out mutant viability
  - removing a reaction removes all EMs / EPs that contain this reaction
  - if all vital EMs removed, the organism dies
- identifying possible drug targets
  - finding the smallest set of reactions whose removal blocks a certain “disease metabolism” (minimal cut sets)

EM / EP: Applications (2)

- enhancement points for metabolic engineering
  - identification of bottlenecks in production
- correlated reaction sets
  - hypothesis: these could be under the same regulatory control
- robustness of networks
  - e.g. how many alternative routes there are between any two metabolites
- can also be applied to e.g. genetic networks

^{13}C-labeling

- problems with the above methods: parallel pathways sometimes indistinguishable, balancing of energy metabolites very tricky
- feeding of ^{13}C-labeled substrate (e.g. glucose)
  - in steady-state
  - until isotopical steady-state
- isotopomer of a metabolite with *n* carbon atoms: one of the 2^n different labeling states
  - measurement (NMR, MS)
  - put the isotopomer fractions to the labeling state vector $\mathbf{x}$
- isotopomer labeling balance equation $f\left(\mathbf{v}, \mathbf{x}^{inp}, \mathbf{x}\right) = \mathbf{0}$
  - $\mathbf{x}^{inp}$ contains the isotopomer fractions of input metabolites
- more details in [1]

[1] Wiechert, W., "^{13}C Metabolic flux analysis," *Metabolic Engineering*, 3, 195 – 206, 2001.