V15 Warmup: Metabolic networks are scale-free ☺

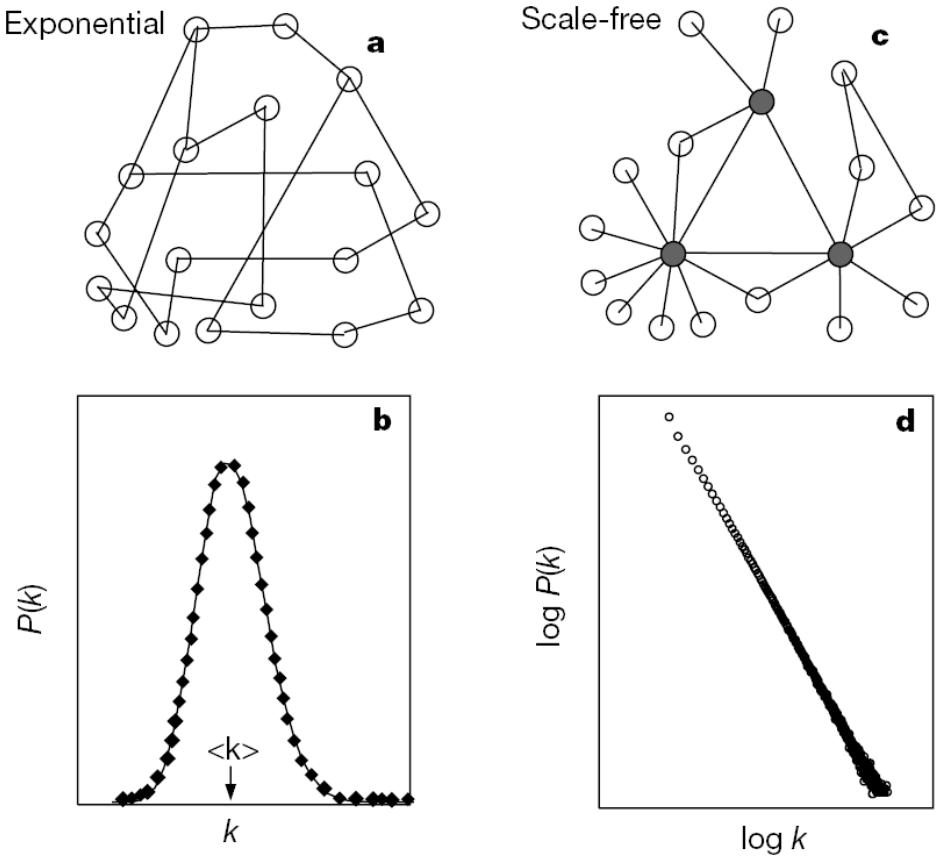
Review of 2 contrasting network topologies.

a, Representative structure of networks generated by the Erdös–Rényi model.

b, For a **random network** the probability, $P(k)$ - that a node has k links - peaks strongly at $k = \langle k \rangle$ and decays exponentially for large k.

c, In the **scale-free network** most nodes have only a few links, but a few nodes, called hubs (dark), have many links.

d, $P(k)$ for a scale-free network has no well-defined peak, and for large k it decays as a power-law, $P(k) \approx k^{-\gamma}$, appearing as a straight line with slope - on a log–log plot.



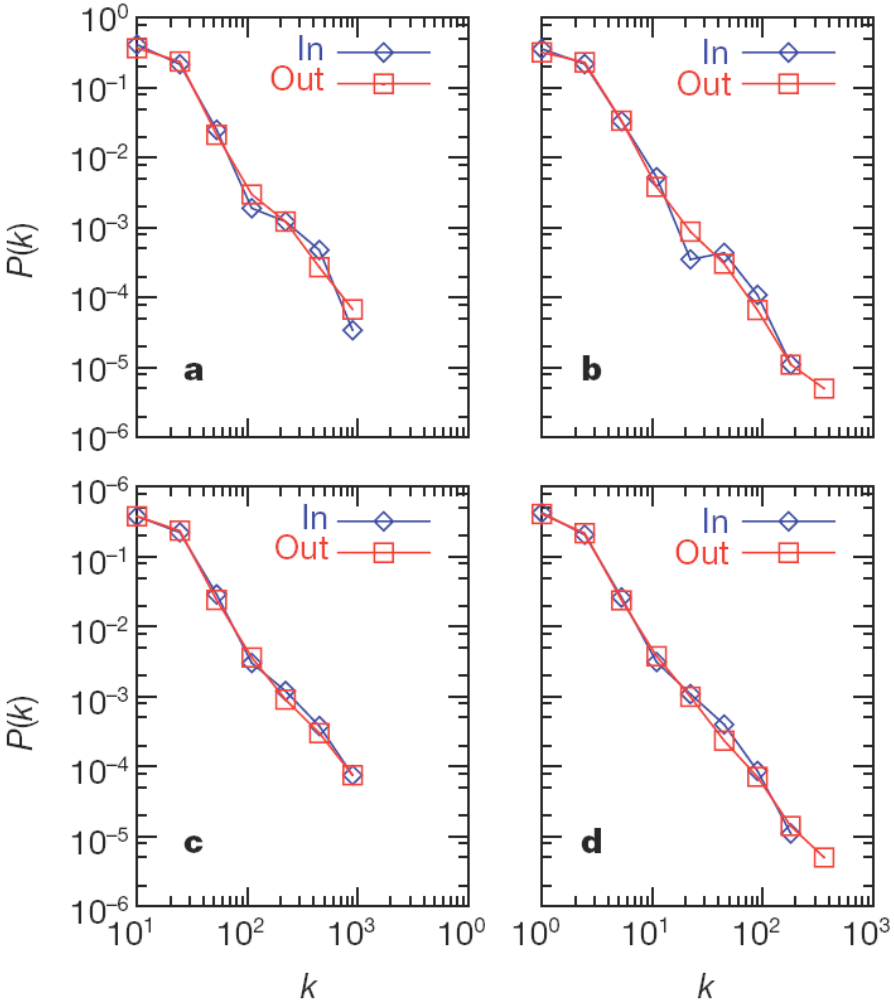
Jeong et al. Nature 407, 651 (2000)

Connectivity distributions *P(k)* for substrates

a, *Archaeoglobus fulgidus* (archae);
b, *E. coli* (bacterium);
c, *Caenorhabditis elegans* (eukaryote)
d, The connectivity distribution averaged over 43 organisms.

x-axis: metabolites participating in *k* reactions
y-axis (*P(k)*): number/frequency of such metabolites

log–log plot, counts separately the incoming (In) and outgoing links (Out) for each substrate.
k_{in} (k_{out}) corresponds to the number of reactions in which a substrate participates as a product (educt).



Jeong et al. Nature 407, 651 (2000)

Properties of metabolic networks

a, The histogram of the biochemical pathway lengths, *l*, in *E. coli*.

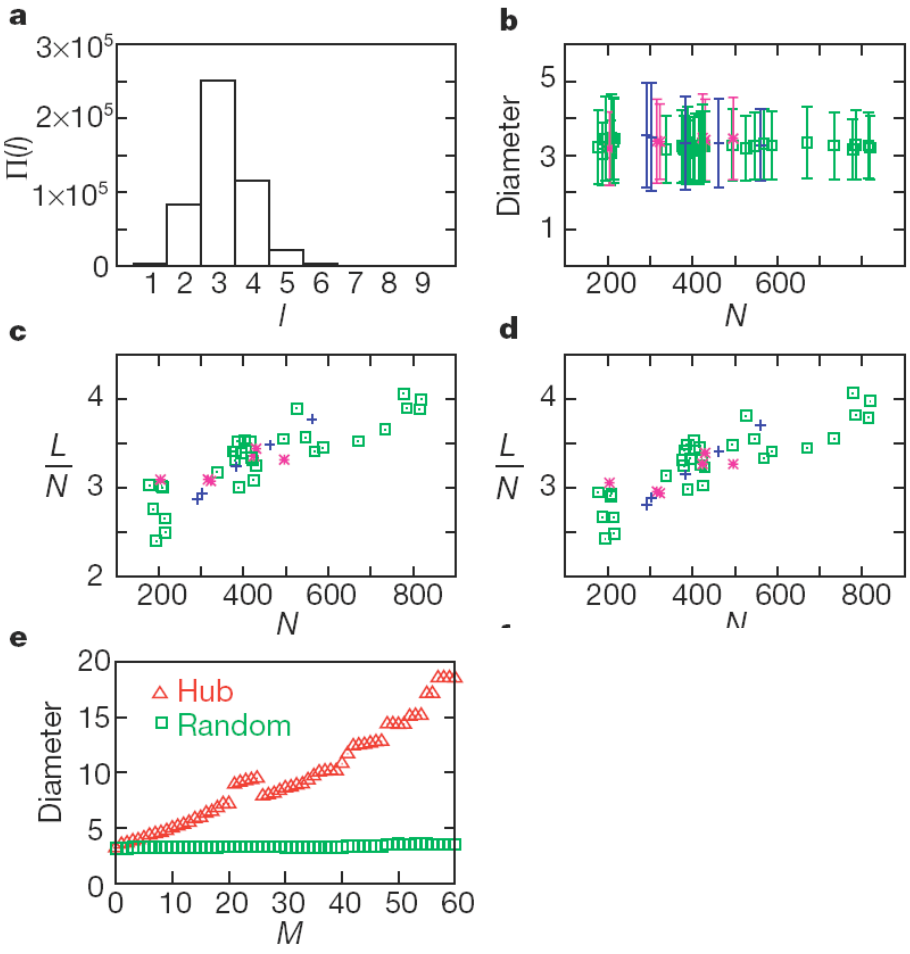
b, The average path length (diameter) for each of the 43 organisms.

N : number of metabolites in each organism

c, **d**, Average number of incoming links (**c**) or outgoing links (**d**) per node for each organism.

e, The effect of substrate removal on the metabolic network diameter of *E. coli*.

In the top curve (red) the most connected substrates are removed first. In the bottom curve (green) nodes are removed randomly. *M* = 60 corresponds to 8% of the total number of substrates in found in *E. coli*.



b–**d**, Archaea (magenta), bacteria (green) and eukaryotes (blue) are shown.

Jeong et al. Nature 407, 651 (2000)

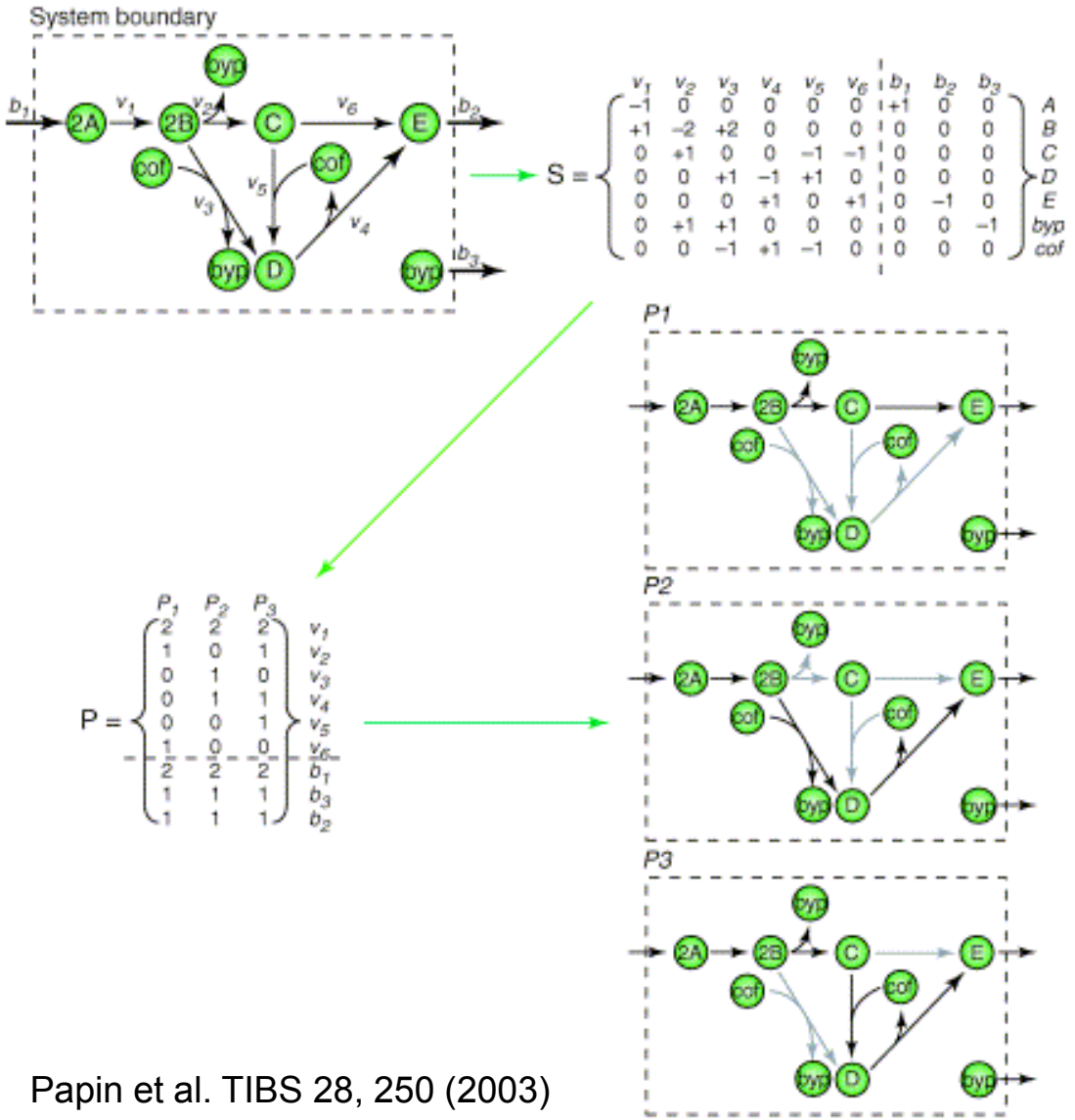
Stoichiometric matrix - Flux Balance Analysis

Stoichiometric matrix S:

m × *n* matrix with stochiometries of the *n* reactions as columns and participations of *m* metabolites as rows.

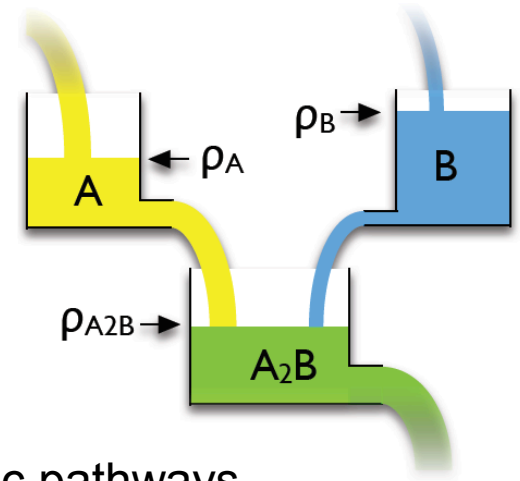
The stochiometric matrix is an important part of the *in silico* model.

With the matrix, the methods of extreme pathway and elementary mode analyses can be used to generate a unique set of pathways P1, P2, and P3 that allow to express all steady-state fluxes as linear combinations of P1 – P3.



Papin et al. TIBS 28, 250 (2003)

Flux balancing



Any chemical reaction requires **mass conservation**.

Therefore one may analyze metabolic systems by requiring mass conservation.

Only required: knowledge about stoichiometry of metabolic pathways.

For each metabolite X_i :

$$dX_i/dt = V_{synthesized} - V_{used} + V_{transported\_in} - V_{transported\_out}$$

Steady state: concentrations are constant
=> flux in = flux out

$$\frac{dA_2B(t)}{dt} = G_{A_2B} - L_{A_2B} = 0$$

Flux balancing

Under **steady-state conditions**, the mass balance constraints in a metabolic network can be represented mathematically by the matrix equation:

$$\mathbf{S} \cdot v = 0$$

where

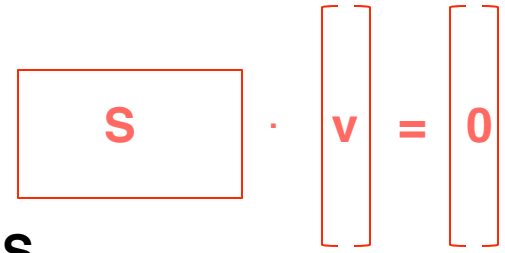
- the matrix **S** is the **stoichiometric matrix** and
- the vector v represents all **fluxes** in the metabolic network, including the internal fluxes, transport fluxes and the growth flux.

Flux balance analysis

Since the number of metabolites is generally smaller than the number of reactions ($m < n$) the flux-balance equation is typically **underdetermined**.

Therefore there are generally multiple feasible flux distributions that satisfy the mass balance constraints.

The set of solutions are confined to the **nullspace** of matrix **S**.



Null space: space of feasible solutions

Consider

$$\begin{pmatrix} 0 & 2 & 1 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

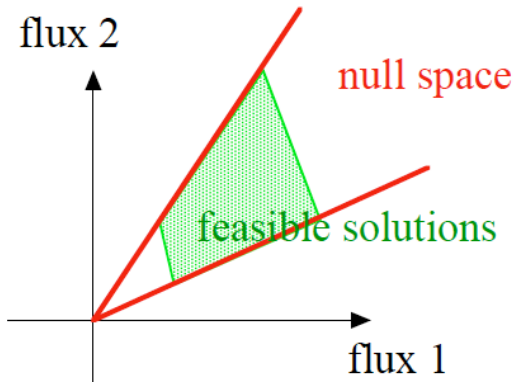
Corresponds to $\quad 2x_2 + x_3 = 0 \quad 3x_1 - x_2 + x_3 = 0 \quad \Leftrightarrow \quad 2x_2 = -x_3 \quad 2x_1 = -x_3$

=> only one free parameter: x_3

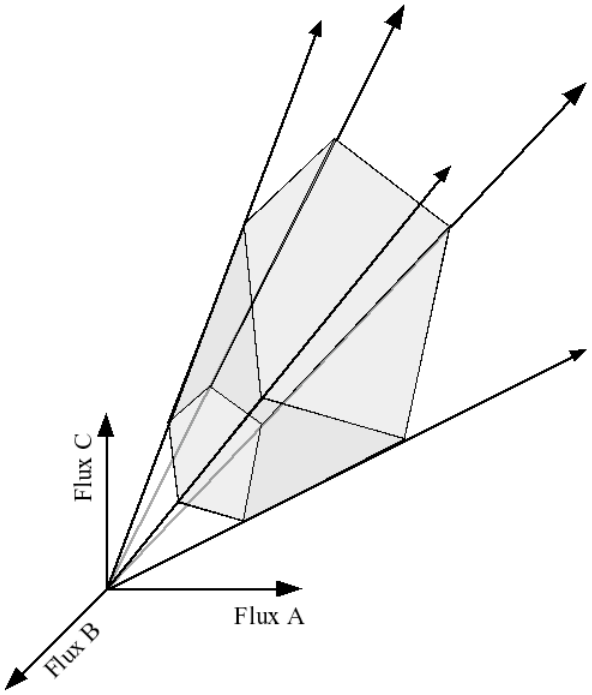
null space: $\vec{x} = \begin{pmatrix} -a \\ -a \\ 2a \end{pmatrix}$

Add inequalities for external fluxes
(here, e.g.: $x_3 \geq 0$)
=> **feasible** solutions for $a \geq 0$

Generally: null space is a cone,
constraints select part of it



Feasible solution set for a metabolic reaction network



The steady-state operation of the metabolic network is restricted to the region within a **pointed cone**, defined as the feasible set.

The feasible set contains all flux vectors that satisfy the physicochemical constrains.

Thus, the feasible set defines the capabilities of the metabolic network. All feasible metabolic flux distributions lie within the feasible set.

Edwards & Palsson PNAS 97, 5528 (2000)

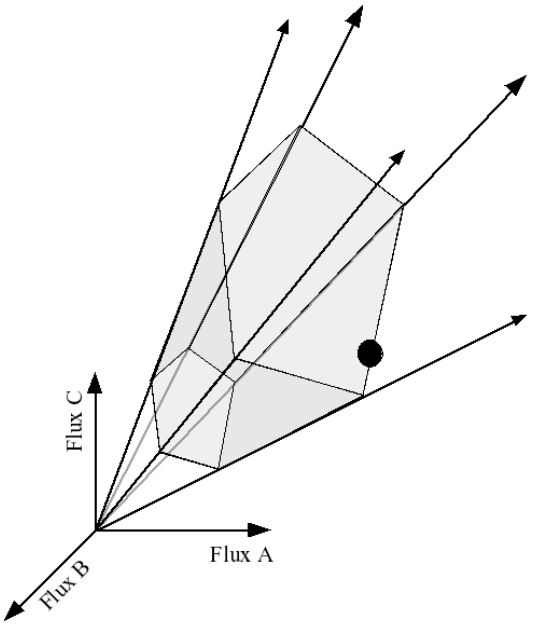
True biological flux

To find the „true“ biological flux in cells ($\rightarrow$ e.g. Heinzle, Wittmann / UdS) one needs additional (experimental) information, or one may impose constraints

$$\alpha_i \leq v_i \leq \beta_i$$

on the magnitude of each individual metabolic flux.

The intersection of the nullspace and the region defined by those linear inequalities defines a region in flux space = the **feasible set of fluxes**.



In the limiting case, where all constraints on the metabolic network are known, such as the enzyme kinetics and gene regulation, the feasible set may be reduced to a single point. This single point must lie within the feasible set.

E.coli in silico

Best studied cellular system: *E. coli*.

In 2000, Edwards & Palsson constructed an *in silico* representation of *E.coli* metabolism.

There were 2 good reasons for this:

(1) genome of *E.coli* MG1655 was already completely sequenced,

(2) Because of long history of *E.coli* research, biochemical literature, genomic information, metabolic databases EcoCyc, KEGG contained biochemical or genetic evidence for every metabolic reaction included in the *in silico* representation. In most cases, there existed both.

Edwards & Palsson

PNAS 97, 5528 (2000)

Genes included in *in silico* model of *E.coli*

Table 1. The genes included in the *E. coli* metabolic genotype (21)

Central metabolism (EMP, PPP, TCA cycle, electron transport)	*aceA, aceB, aceE, aceF, ackA, acnA, acnB, acs, adhE, agp, appB, appC, atpA, atpB, atpC, atpD, atpE, atpF, atpG, atpH, atpI, cydA, cydB, cydC, cydD, cyoA, cyoB, cyoC, cyoD, dld, eno, fba, fbp, fdhF, fdnG, fdnH, fdnI, fdoG, fdoH, fdoI, frdA, frdB, frdC, frdD, fumA, fumB, fumC, galM, gapA, gapC_1, gapC_2, glcB, glgA, glgC, glgP, glk, glpA, glpB, glpC, glpD, gltA, gnd, gpmA, gpmB, hyaA, hyaB, hyaC, hybA, hybC, hycB, hycE, hycF, hycG, icdA, lctD, ldhA, lpdA, malP, mdh, ndh, nuoA, nuoB, nuoE, nuoF, nuoG, nuoH, nuoI, nuoJ, nuoK, nuoL, nuoM, nuoN, pckA, pfkA, pfkB, pflA, pflB, pflC, pflD, pgi, pgk, pntA, pntB, ppc, ppsA, pta, purT, pykA, pykF, rpe, rpiA, rpiB, sdhA, sdhB, sdhC, sdhD, sfcA, suoA, sucB, sucC, sucD, talB, tktA, tktB, tpiA, trxB, zwf, pgl* (30), *maeB* (30)
Alternative carbon source	*adhC, adhE, agaY, agaZ, aldA, aldB, aldH, araA, araB, araD, bglX, cpsG, deoB, fruK, fucA, fucI, fucK, fucO, galE, galK, galT, galU, garD, garY, glk, glpK, gntK, gntV, gpsA, lacZ, manA, melA, mtlD, nagA, nagB, narA, pfkB, pgi, pgm, rbsK, rhaA, rhaB, rhaD, srlD, treC, xylA, xylB*
Amino acid metabolism	*adi, aldH, alr, ansA, ansB, argA, argB, argC, argD, argE, argF, argG, argH, argI, aroA, aroB, aroC, aroD, aroE, aroF, aroG, aroH, aroK, aroL, asd, asnA, asnB, aspA, aspC, avtA, cadA, carA, carB, cysC, cysD, cysE, cysH, cysI, cysJ, cysK, cysM, cysN, dadA, dadX, dapA, dapB, dapD, dapE, dapF, dsdA, gabD, gabT, gadA, gadB, gdhA, glk, glnA, gltB, gltD, glyA, goaG, hisA, hisB, hisC, hisD, hisF, hisG, hisH, hisI, ilvA, ilvB, ilvC, ilvD, ilvE, ilvG_1, ilvG_2, ilvH, ilvI, ilvM, ilvN, kbl, ldcC, leuA, leuB, leuC, leuD, lysA, lysC, metA, metB, metC, metE, metH, metK, metL, pheA, proA, proB, proC, prsA, putA, sdaA, sdaB, serA, serB, serC, speA, speB, speC, speD, speE, speF, tdcB, tdh, thrA, thrB, thrC, tnaA, trpA, trpB, trpC, trpD, trpE, tyrA, tyrA, tyrB, ygjG, ygjH, aleB* (42), *depC* (43), *pat* (44), *prr* (44), *sad* (45), *methylthioadenosine nucleosidase* (46), *5-methylthioribose kinase* (46), *5-methylthioribose-1-phosphate isomerase* (46), *adenosyl homocysteinase* (47), *L-cysteine desulfhydrase* (44), *glutaminase A* (44), *glutaminase B* (44)
Purine & pyrimidine metabolism	*add, adk, amn, apt, cdd, cmk, codA, dcd, deoA, deoD, dgt, dut, gmk, gpt, gsk, guaA, guaB, guaC, hpt, mutT, ndk, nrdA, nrdB, nrdD, nrdE, nrdF, purA, purB, purC, purD, purE, purF, purH, purK, purL, purM, purN, purT, pyrB, pyrC, pyrD, pyrE, pyrF, pyrG, pyrH, pyrI, tdk, thyA, tmk, udk, udp, upp, ushA, xapA, yicP, CMP glycosylase* (48)
Vitamin & cofactor metabolism	*acpS, bioA, bioB, bioD, bioF, coaA, cyoE, cysG, entA, entB, entC, entD, entE, entF, epd, folA, folC, folD, folE, folK, folP, gcvH, gcvP, gcvT, gltX, glyA, gor, gshA, gshB, hemA, hemB, hemC, hemD, hemE, hemF, hemH, hemK, hemL, hemM, hemX, hemY, ilvC, lig, lpdA, menA, menB, menC, menD, menE, menF, menG, metF, mutT, nadA, nadB, nadC, nadE, ntpA, pabA, pabB, pabC, panB, panC, panD, pdxA, pdxB, pdxH, pdxJ, pdxK, pncB, purU, ribA, ribB, ribD, ribE, ribH, serC, thiC, thiE, thiF, thiG, thiH, thrC, ubiA, ubiB, ubiC, ubiG, ubiH, ubiX, yaaC, ygiG, nadD* (49), *nadF* (49), *nadG* (49), *panE* (50), *pncA* (49), *pncC* (49), *thiB* (51), *thiD* (51), *thiK* (51), *thiL* (51), *thiM* (51), *thiN* (51), *ubiE* (52), *ubiF* (52), *arabinose-5-phosphate isomerase* (22), *phosphopantothenate-cysteine ligase* (50), *phosphopantothenate-cysteine decarboxylase* (50), *phospho-pantetheine adenylyltransferase* (50), *dephosphoCoA kinase* (50), *NMN glycohydrolase* (49)
Lipid metabolism	*accA, accB, accD, atoB, cdh, cdsA, cls, dgkA, fabD, fabH, fadB, gpsA, ispA, ispB, pgpB, pgsA, psd, pssA, pgpA* (53)
Cell wall metabolism	*ddlA, ddlB, galF, galU, glmS, glmU, htrB, kdsA, kdsB, kdtA, lpxA, lpxB, lpxC, lpxD, mraY, msbB, murA, murB, murC, murD, murE, murF, murG, murI, rfaC, rfaD, rfaF, rfaG, rfaI, rfaJ, rfaL, ushA, glmM* (54), *lpoA* (55), *rfaE* (55), *tetraacyldisaccharide 4' kinase* (55), *3-deoxy-D-manno-octulosonic-acid 8-phosphate phosphatase* (55)
Transport processes	*araE, araF, araG, araH, argT, aroP, artI, artJ, artM, artP, artQ, brnQ, cadB, chaA, chaB, chaC, cmtA, cmtB, codB, crr, cycA, cysA, cysP, cysT, cysU, cysW, cysZ, dctA, dcuA, dcuB, dppA, dppB, dppC, dppD, dppF, fadL, focA, fruA, fruB, fucP, gabP, galP, garA, garB, garC, glnH, glnP, glnQ, glpF, glpT, gltJ, gltK, gltL, gltP, glts, gntT, gpt, hisJ, hisM, hisP, hisQ, hpt, kdpA, kdpB, kdpC, kgtP, lacY, lamB, livF, livG, livH, livJ, livK, livM, lldP, lysP, malE, malF, malG, malK, malX, manX, manY, manZ, melB, mglA, mglB, mglC, mtlA, mtr, nagE, nanT, nhaA, nhaB, nupC, nupG, oppA, oppB, oppC, oppD, oppF, panF, pheP, pitA, pitB, pnuC, potA, potB, potC, potD, potE, potF, potG, potH, potI, proP, proV, proW, proX, pstA, pstB, pstC, pstS, ptsA, ptsG, ptsI, ptsN, ptsP, purB, putP, rbsA, rbsB, rbsC, rbsD, rhaT, sapA, sapB, sapD, sbp, sdaC, srlA_1, srlA_2, srlB, tdcC, tnaB, treA, treB, trkA, trkG, trkH, tsx, tyrP, ugpA, ugpB, ugpC, ugpE, uraA, xapB, xylE, xylF, xylG, xylH, fruF* (56), *gntS* (57), *metD* (43), *pnuE* (49), *scr* (56)

Edwards & Palsson

PNAS 97, 5528 (2000)

E.coli in silico – Flux balance analysis

Define $\alpha_i = 0$ for irreversible internal fluxes,

$\alpha_i = -\infty$ for reversible internal fluxes (use biochemical literature)

Transport fluxes for PO_4^{2-}, NH_3, CO_2, SO_4^{2-}, K^+, Na^+ were unrestrained.

For other metabolites $0 < v_i < v_i^{max}$ except for those that are able to leave the metabolic network (i.e. acetate, ethanol, lactate, succinate, formate, pyruvate etc.)

Find particular metabolic flux distribution in feasible set by **linear programming**. LP finds a solution that **minimizes** a particular metabolic **objective** –Z (subject to the imposed constraints) where e.g.

$$Z = \sum c_i \cdot v_i = \langle \mathrm{c} \cdot \mathrm{v} \rangle$$

When written in this way, the **flux balance analysis (FBA)** method finds the solution that maximizes the sum of all fluxes = gives maximal biomass.

Edwards & Palsson, PNAS 97, 5528 (2000)

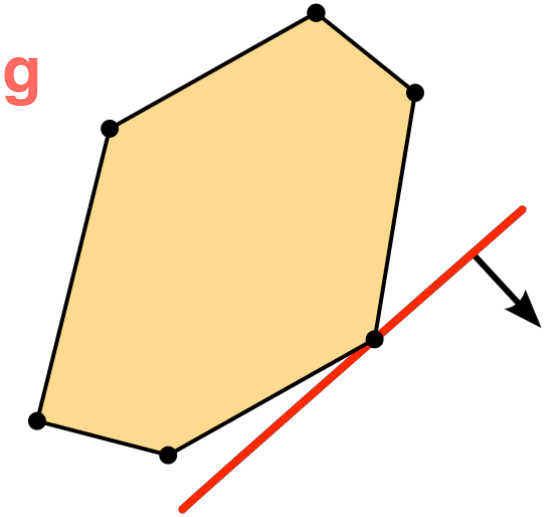
Linear programming

Linear programming is a technique for the **optimization** of a **linear objective function**, subject to linear equality and linear inequality **constraints**.

Its **feasible region** is a convex polytope, which is a set defined as the intersection of finitely many half spaces, each of which is defined by a linear inequality.

Its objective function is a real-valued linear function defined on this polyhedron.

A linear programming algorithm finds a point in the polyhedron where this function has the smallest (or largest) value if such a point exists.



A pictorial representation of a simple linear program with 2 variables and 6 inequalities. The set of feasible solutions is depicted in yellow and forms a polygon, a 2-dimensional polytope.
The linear **cost function** is represented by the red line and the arrow: The arrow indicates the direction in which we are optimizing.

www.wikipedia.org

Linear programming

Linear programs are problems that can be expressed in canonical form as

$$\begin{aligned} \text{maximize} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b} \\ \text{and} \quad & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

where **x** represents the vector of variables (to be determined), **c** and **b** are vectors of (known) coefficients, *A* is a (known) matrix of coefficients, and $(.)^{\mathrm{T}}$ is the matrix transpose.

The expression to be maximized or minimized is called the *objective function* ($c^{\mathrm{T}}x$ in this case).

The inequalities $Ax \leq b$ and $x \geq 0$ are the constraints which specify a convex polytope over which the objective function is to be optimized.

www.wikipedia.org

Integer linear programming

If all of the unknown variables are required to be integers, then the problem is called an integer programming (IP) or integer linear programming (ILP) problem.

In contrast to linear programming, which can be solved efficiently in the worst case, integer programming problems are in many practical situations NP-hard.

The **branch and bound algorithm** is one type of algorithm to solve ILP problems.

www.wikipedia.org

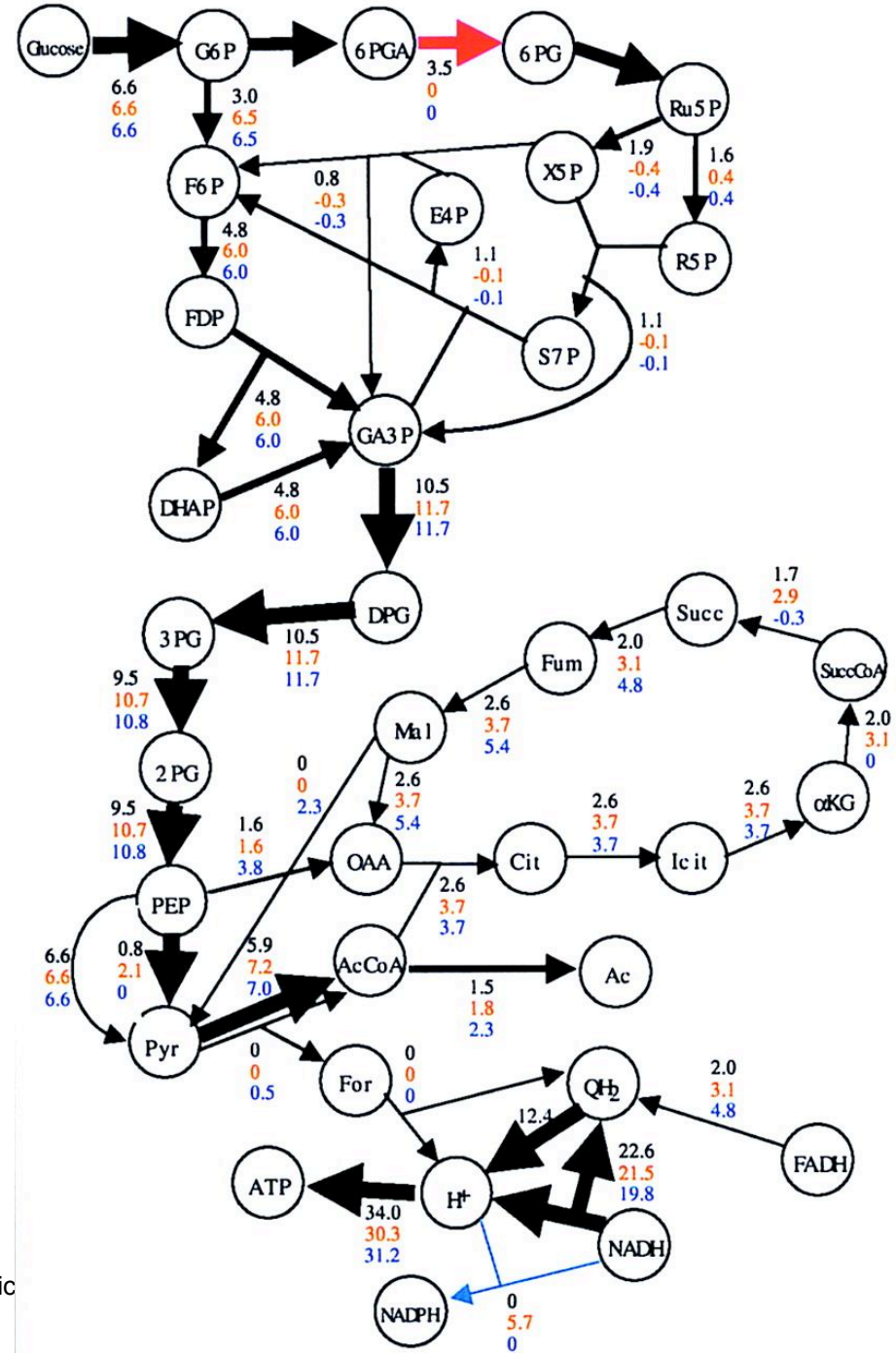
Rerouting of metabolic fluxes

(Black) Flux distribution for the wild-type.

(Red) *zwf*- mutant. Biomass yield is 99% of wild-type result.

(Blue) *zwf- pnt*- mutant. Biomass yield is 92% of wildtype result.

Note how *E.coli in silico* circumvents removal of one critical reaction (red arrow) by increasing the flux through the alternative G6P → P6P reaction.



Edwards & Palsson PNAS 97, 5528 (2000)

E.coli in silico

Examine **changes** in the **metabolic capabilities** caused by hypothetical **gene deletions.**

To simulate a gene deletion, the flux through the corresponding enzymatic reaction was restricted to zero.

Compare optimal value of mutant (Z_{mutant}) to the „wild-type“ objective Z

$$\frac{Z_{mutant}}{Z}$$

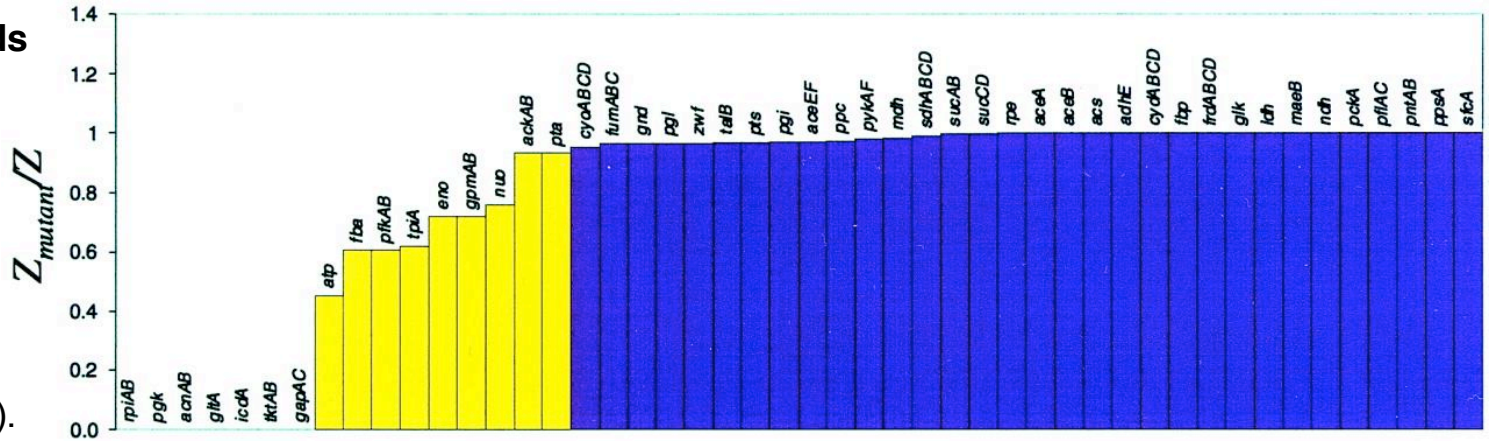
to determine the systemic effect of the gene deletion.

Edwards & Palsson

PNAS 97, 5528 (2000)

Gene deletions in central intermediary metabolism

Maximal biomass yields on glucose for all possible single gene deletions in the central metabolic pathways (gycolysis, pentose phosphate pathway (PPP), TCA, respiration).



The results were generated in a simulated aerobic environment with glucose as the carbon source. The transport fluxes were constrained as follows: glucose = 10 mmol/g-dry weight (DW) per h; oxygen = 15 mmol/g-DW per h.

The maximal yields were calculated by using FBA with the objective of maximizing growth.

Yellow bars: gene deletions that reduced the maximal biomass yield of Z_{mutant} to less than 95% of the *in silico* wild type Z_{wt}.

Edwards & Palsson PNAS 97, 5528 (2000)

Interpretation of gene deletion results

The essential gene products were involved in the 3-carbon stage of glycolysis, 3 reactions of the TCA cycle, and several points within the pentose phosphate pathway (PPP).

The remainder of the central metabolic genes could be removed while *E.coli in silico* maintained the potential to support cellular growth.

This suggests that a large number of the central metabolic genes can be removed without eliminating the capability of the metabolic network to support growth under the conditions considered.

Edwards & Palsson PNAS 97, 5528 (2000)

E.coli in silico – validation

+ and – means growth or no growth.

± means that suppressor mutations have been observed that allow the mutant strain to grow.

4 virtual growth media:
glc: glucose, gl: glycerol, succ: succinate, ac: acetate.

In 68 of 79 cases, the prediction was consistent with exp. predictions.

Red and yellow circles: predicted mutants that eliminate or reduce growth.

Edwards & Palsson

PNAS 97, 5528 (2000)

Table 2. Comparison of the predicted mutant growth characteristics from the gene deletion study to published experimental results with single mutants

Gene	glc	gl	succ	ac
aceA	+/+		+/+	−/−
aceB				−/−
*aceEF**	−/+			
ackA				+/+
acn	−/−			−/−
acs				+/+
cyd	+/+			
cyo	+/+			
eno†	−/+	−/+	−/−	−/−
fbaI	−/+			
fbp	+/+	−/−	−/−	−/−
frd	+/+		+/+	+/+
gap	−/−	−/−	−/−	−/−
glk	+/+			
gltA	−/−			−/−
gnd	+/+			
idh	−/−			−/−
mdh††	+/+	+/+	+/+	
ndh	+/+	+/+		
nuo	+/+	+/+		
pfk†	−/+			
pgi‡	+/+	+/−	+/−	
pgk	−/−	−/−	−/−	−/−
pgl	+/+			
pntAB	+/+	+/+	+/+	
ppc§	±/+	−/+	+/+	
pta				+/+
pts	+/+			
pyk	+/+			
rpi	−/−	−/−	−/−	−/−
sdhABCD	+/+		−/−	−/−
sucAB	+/+		−/+	−/+
tktAB	−/−			
*tpi***	−/+	−/−	−/−	−/−
unc	+/+		±/+	−/−
zwf	+/+	+/+	+/+	

Summary - FBA

FBA analysis constructs the **optimal network utilization** simply using the stoichiometry of metabolic reactions and capacity constraints.

For *E.coli* the *in silico* results are mostly **consistent** with experimental data.

FBA shows that the *E.coli* metabolic network contains relatively **few critical gene products** in central metabolism.
However, the ability to adjust to different environments (growth conditions) may be diminished by gene deletions.

FBA identifies „**the best**“ the cell can do, not how the cell actually behaves under a given set of conditions. Here, survival was equated with growth.

FBA does not directly consider **regulation** or regulatory constraints on the metabolic network. This can be treated separately (see future lecture).

Edwards & Palsson PNAS 97, 5528 (2000)

Metabolic Pathway Analysis (MPA)

Metabolic Pathway Analysis searches for meaningful structural and functional units in metabolic networks.

Today‘s most powerful methods are based on **convex analysis**.

2 such approaches are the **elementary flux modes** [1] and **extreme pathways**[2].

Both sets span the space of feasible steady-state flux distributions by non-decomposable routes, i.e. no subset of reactions involved in an EFM or EP can hold the network balanced using non-trivial fluxes.

Extreme pathways are a subset of elementary modes.
For many systems, both methods coincide.

Klamt et al. Bioinformatics 19, 261 (2003); Trinh et al. Appl. Microbiol Biotechnol. 81, 813-826 (2009)

[1] Schuster & Hilgetag J Biol Syst 2, 165-182 (1994), Pfeiffer et al. Bioinformatics, 15, 251 (1999), Schuster et al. Nature Biotech. 18, 326 (2000)

[2] Schilling et al. J Theor Biol 203, 229-248 (2000)

Applications of Metabolic Pathway Analysis (MPA)

MPA can be used to study e.g.

- metabolic network structure

- functionality of networks (including identification of futile cycles)

- robustness, fragility, flexibility/redundancy of networks

- to identiy all (sub-) optimal pathways with respect to product/biomass yield

- rational strain design

Klamt et al. Bioinformatics 19, 261 (2003) ; Trinh et al. Appl. Microbiol Biotechnol. 81, 813-826 (2009)

Definition of Elementary Flux Modes (EFMs)

A pathway P(**v**) is an **elementary flux mode** if it fulfills conditions C1 – C3.

(C1) **Pseudo steady-state**. **S** · **e** = 0. This ensures that none of the metabolites is consumed or produced in the overall stoichiometry.

(C2) **Feasibility**: rate $e_i \geq 0$ if reaction is irreversible. This demands that only thermodynamically realizable fluxes are contained in **e**.

(C3) **Non-decomposability***:* there is no vector **v** (except the null vector and **e**) fulfilling C1 and C2 and so that P(**v**) is a proper subset of P(**e**).
This is the core characteristics for EFMs and EPs and provides the decomposition of the network into smallest units that are able to hold the network in steady state.

C3 is often called „genetic independence“ because it implies that the enzymes in one EFM or EP are not a subset of the enzymes from another EFM or EP.

Klamt & Stelling Trends Biotech 21, 64 (2003)

Definition of Extreme Pathways (Eps)

The pathway *P(**e**)* is an **extreme pathway** if it fulfills conditions C1 – C3 AND conditions C4 – C5.

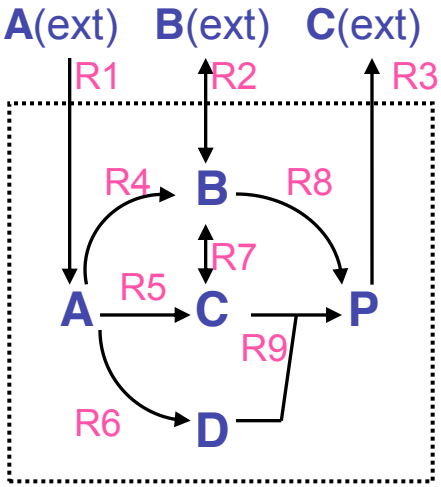
(C4) **Network reconfiguration**: Each reaction must be classified either as exchange flux or as internal reaction.
All **reversible** internal reactions must be **split up** into two separate, irreversible reactions (forward and backward reaction).

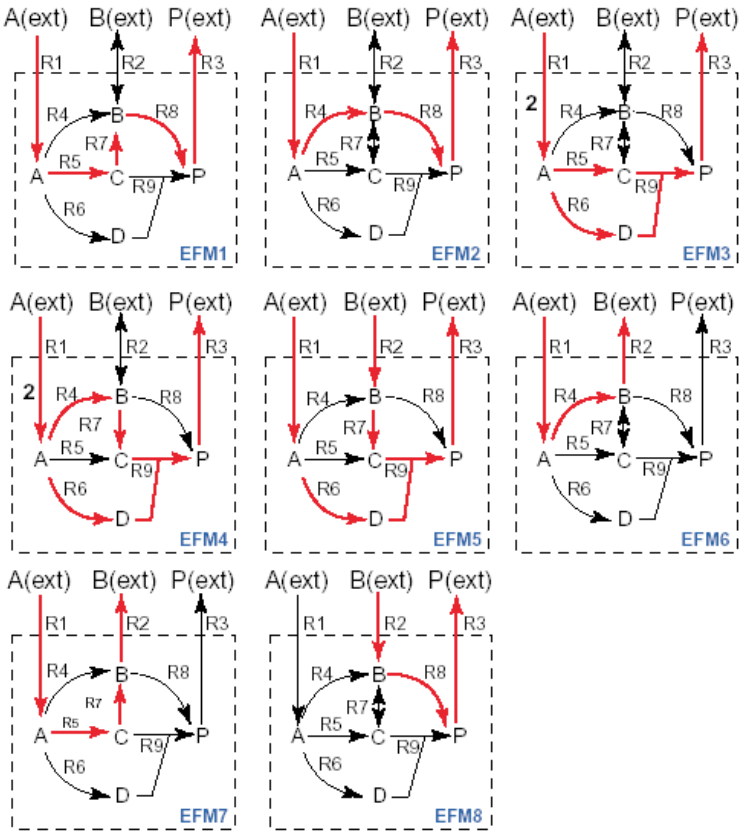
(C5) **Systemic independence**: the set of EPs in a network is the **minimal** set of EFMs that can describe all feasible steady-state flux distributions.

The algorithms for computing EPs and EFMs are quite similar.
We will not cover the algorithmic differences here.

Klamt & Stelling Trends Biotech 21, 64 (2003)

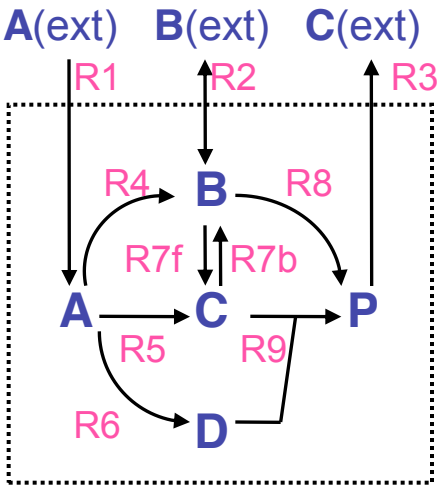
Comparison of EFMs and EPs

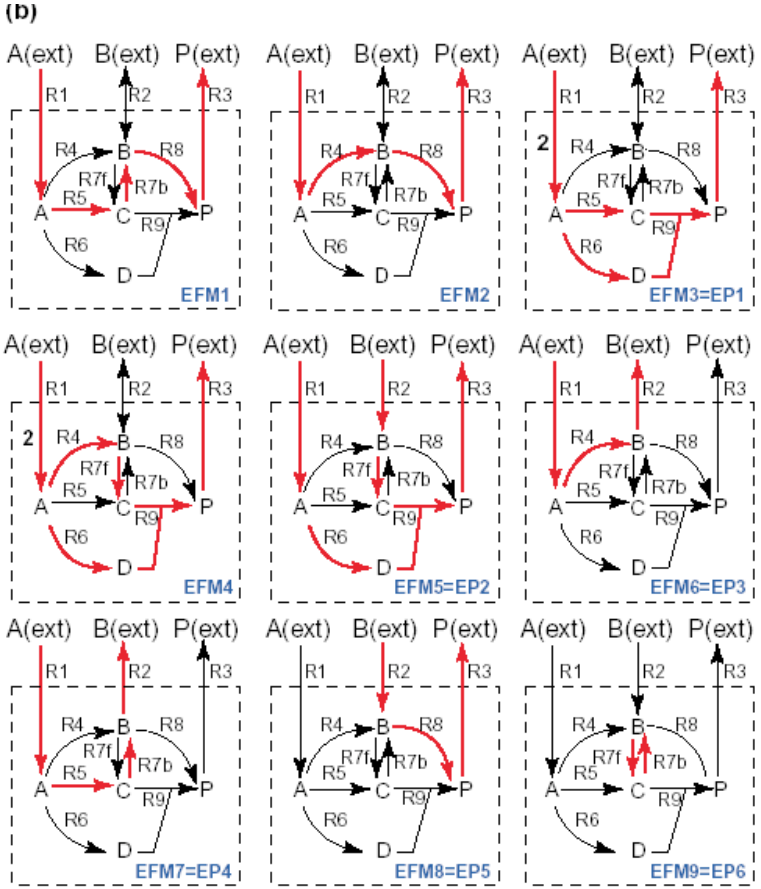




Klamt & Stelling Trends Biotech 21, 64 (2003)

Reconfigured Network: split up R7





3 EFMs are not systemically independent:
EFM1 = EP4 + EP5
EFM2 = EP3 + EP5
EFM4 = EP2 + EP3

Klamt & Stelling Trends Biotech 21, 64 (2003)

Property 1 of EFMs

The only difference in the set of EFMs emerging upon reconfiguration consists in the **two-cycles** that result from splitting up reversible reactions.
However, two-cycles are not considered as meaningful pathways.

Valid for any network: Property 1
Reconfiguring a network by splitting up reversible reactions leads to the same set of meaningful EFMs.

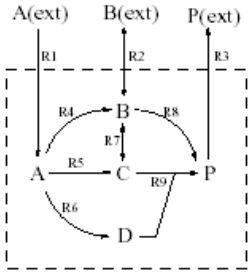
Klamt & Stelling Trends Biotech 21, 64 (2003)

EFMs vs. EPs

What is the consequence when all exchange fluxes (and hence all reactions in the network) are made irreversible?

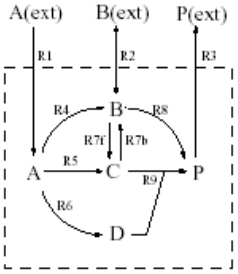
Table 1. Configurations of the example network (upper part N1 and N3; lower part N2 and N4), with corresponding elementary flux modes (EFM) and extreme pathways (EP) (see also Fig. 1)

N1 (R2 and R7 reversible) N3 (as N1 but R2 irreversible)



N1		N3								
EFMs		EFMs	R1	R2	R3	R4	R5	R6	R7	R8	R9
EFM1		×	1	0	1	0	1	0	−1	1	0
EFM2		×	1	0	1	1	0	0	0	1	0
EFM3		×	2	0	1	0	1	1	0	0	1
EFM4		×	2	0	1	1	0	1	1	0	1
EFM5		×	1	1	1	0	0	1	1	0	1
EFM6			1	−1	0	1	0	0	0	0	0
EFM7			1	−1	0	0	1	0	−1	0	0
EFM8		×	0	1	1	0	0	0	0	1	0

N2 (R2 reversible, R7 split up) N4 (as N2 but R2 irreversible)



N2		N4										
EFMs	EPs	EFMs	EPs	R1	R2	R3	R4	R5	R6	R7f	R8	R9	R7b
EFM1		×	EP1′	1	0	1	0	1	0	0	1	0	1
EFM2		×	EP2′	1	0	1	1	0	0	0	1	0	0
EFM3	EP1	×	EP3′	2	0	1	0	1	1	0	0	1	0
EFM4		×	EP4′	2	0	1	1	0	1	1	0	1	0
EFM5	EP2	×	EP5′	1	1	1	0	0	1	1	0	1	0
EFM6	EP3			1	−1	0	1	0	1	0	0	0	0
EFM7	EP4			1	−1	0	0	1	0	0	0	0	1
EFM8	EP5	×	EP6′	0	1	1	0	0	0	0	1	0	0
EFM9	EP6	×	EP7	0	0	0	0	0	0	1	0	0	1

Then EFMs and EPs always co-incide!

Klamt & Stelling Trends Biotech 21, 64 (2003)

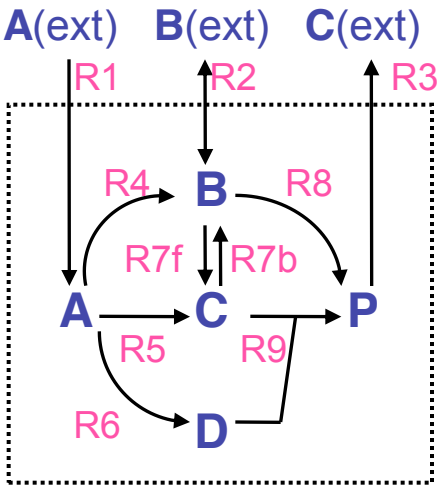
Property 2 of EFMs

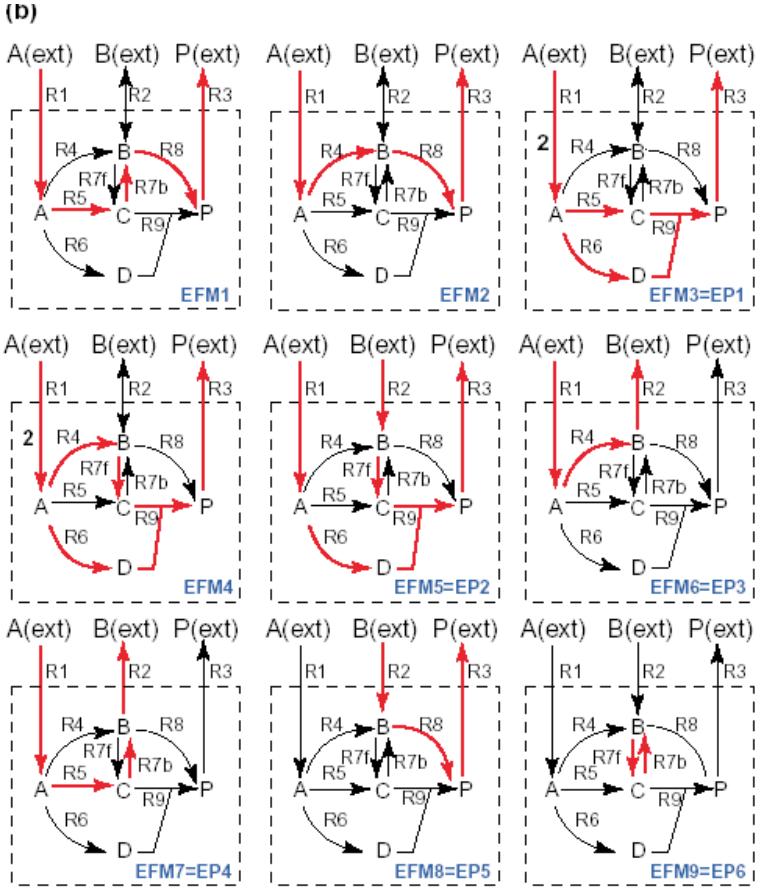
Property 2

If all exchange reactions in a network are irreversible then the sets of meaningful EFMs (both in the original and in the reconfigured network) and EPs coincide.

Klamt & Stelling Trends Biotech 21, 64 (2003)

Reconfigured Network



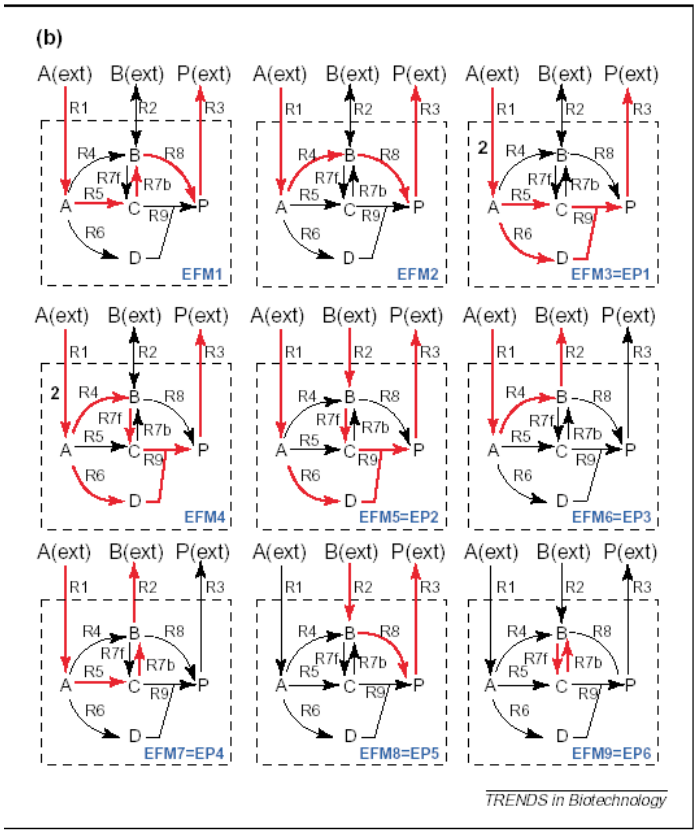


3 EFMs are not systemically independent:
EFM1 = EP4 + EP5
EFM2 = EP3 + EP5
EFM4 = EP2 + EP3

Klamt & Stelling Trends Biotech 21, 64 (2003)

Operational modes

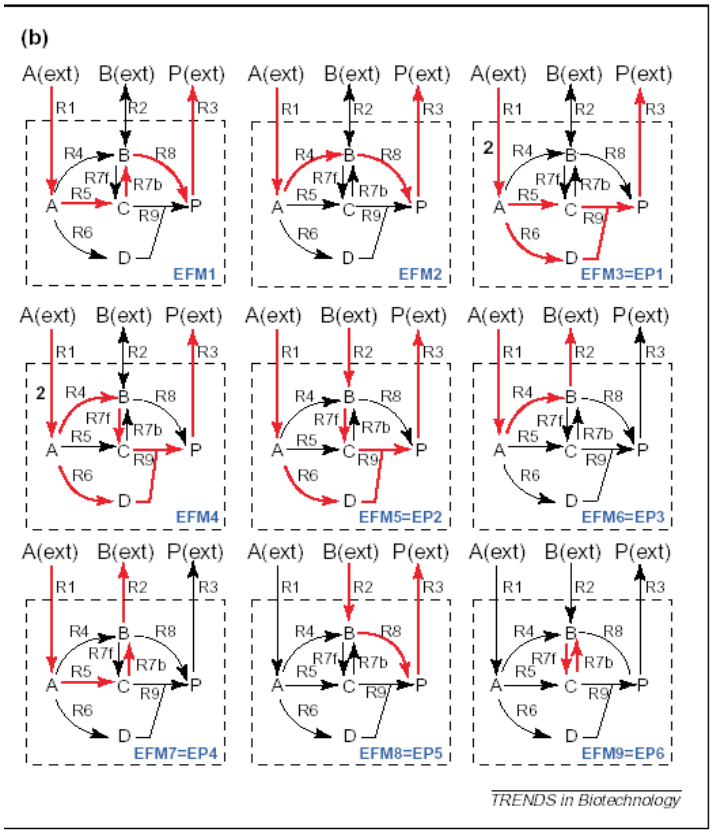
Problem	EFM (network N1)	EP (network N2)
Recognition of operational modes: routes for converting exclusively A to P.	4 genetically independent routes (EFM1-EFM4)	Set of EPs does not contain all genetically independent routes, only EP1. No EP leads directly from A to P via B.



N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

Finding optimal routes

Problem	EFM (network N1)	EP (network N2)
Finding all the optimal routes: optimal pathways for synthesizing P during growth on A alone.	EFM1 and EFM2 are optimal because they yield one mole P per mole substrate A (i.e. R3/R1 = 1), whereas EFM3 and EFM4 are only sub-optimal (R3/R1 = 0.5).	One would only find the suboptimal EP1, not the optimal routes EFM1 and EFM2.

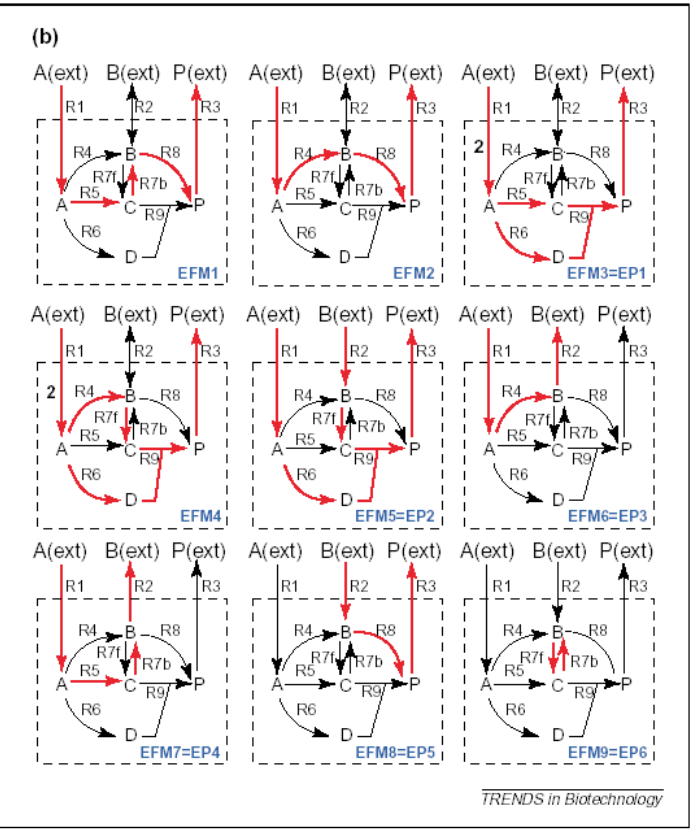


l and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

Network flexibility (structural robustness, redundancy)

Problem

Analysis of network flexibility: relative robustness of exclusive growth on A or B.



N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

EFM (network N1)

4 pathways convert A to P (EFM1-EFM4), whereas for B only one route (EFM8) exists.

When one of the internal reactions (R4-R9) fails, 2 pathways will always „survive“ for production of P from A.
By contrast, removing reaction R8 already stops the production of P from B alone.

EP (network N2)

Only 1 EP exists for producing P by substrate A alone (EP1), and 1 EP for synthesizing P by (only) substrate B (EP5).

This suggests that both substrates possess the same redundancy of pathways, but as shown by EFM analysis, growth on substrate A is much more flexible than on B.

Klamt & Stelling Trends Biotech 21, 64 (2003)

Relative importance of single reactions

Problem

Relative importance of single reactions: relative importance of reaction R8.

EFM (network N1)

R8 is essential for producing P by substrate B (EFM8), whereas for A there is no structurally „favored“ reaction (R4-R9 all occur twice in EFM1-EFM4).

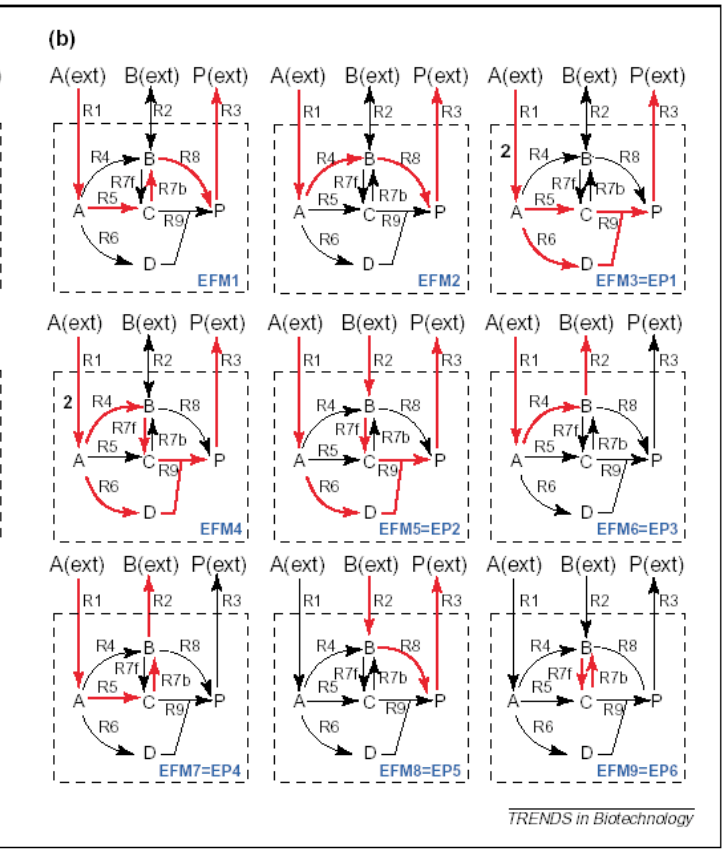
However, considering the optimal modes EFM1, EFM2, one recognizes the importance of R8 also for growth on A.

EP (network N2)

Consider again biosynthesis of P from substrate A (EP1 only).

Because R8 is not involved in EP1 one might think that this reaction is not important for synthesizing P from A.

However, without this reaction, it is impossible to obtain optimal yields (1 P per A; EFM1 and EFM2).



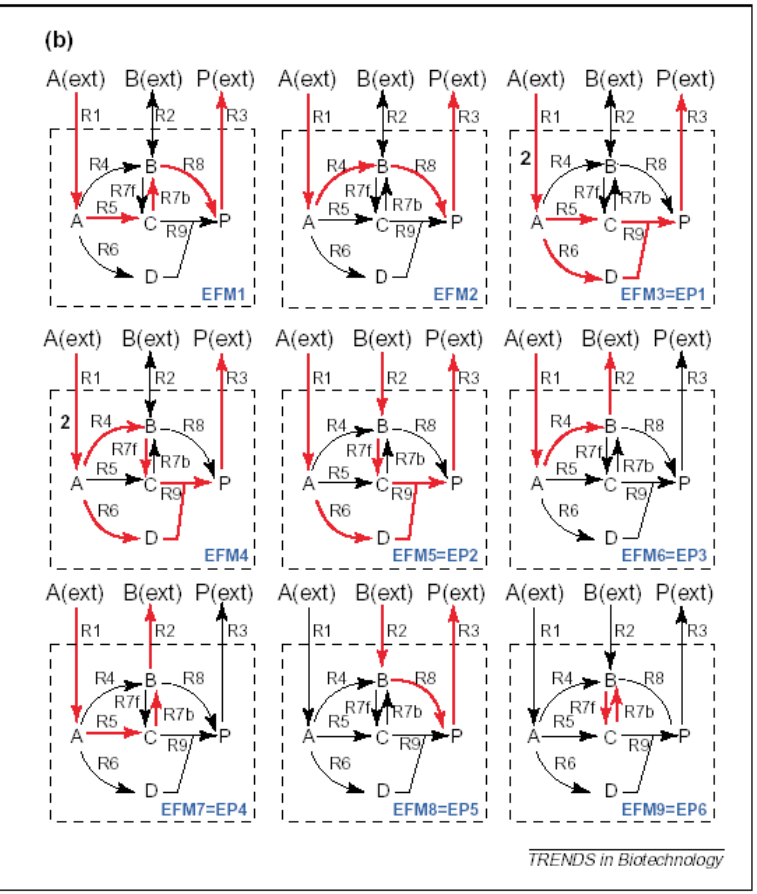
k N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

Klamt & Stelling Trends Biotech 21, 64 (2003)

Enzyme subsets and excluding reaction pairs

Problem

Enzyme subsets and excluding reaction pairs: suggest regulatory structures or rules.



k N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

EFM (network N1)

R6 and R9 are an **enzyme subset**.

By contrast, R6 and R9 never occur together with R8 in an EFM.

Thus (R6,R8) and (R8,R9) are excluding reaction pairs.
(In an arbitrary composable steady-state flux distribution they might occur together.)

EP (network N2)

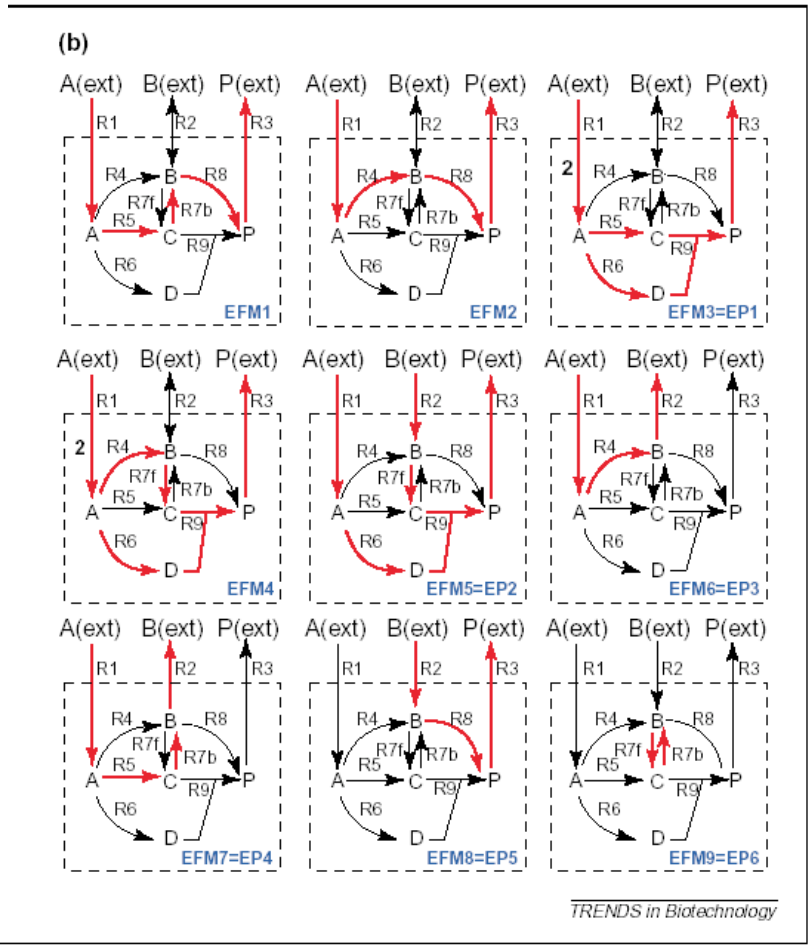
The EPs pretend R4 and R8 to be an excluding reaction pair – but they are not (EFM2).

The enzyme subsets would be correctly identified in this case. However, one can construct simple examples where the EPs would also pretend wrong enzyme subsets (not shown).

Klamt & Stelling Trends Biotech 21, 64 (2003)

Pathway length

Problem	EFM (network N1)	EP (network N2)
Pathway length: shortest/longest pathway for production of P from A.	The shortest pathway from A to P needs 2 internal reactions (EFM2), the longest 4 (EFM4).	Both the shortest (EFM2) and the longest (EFM4) pathway from A to P are not contained in the set of EPs.

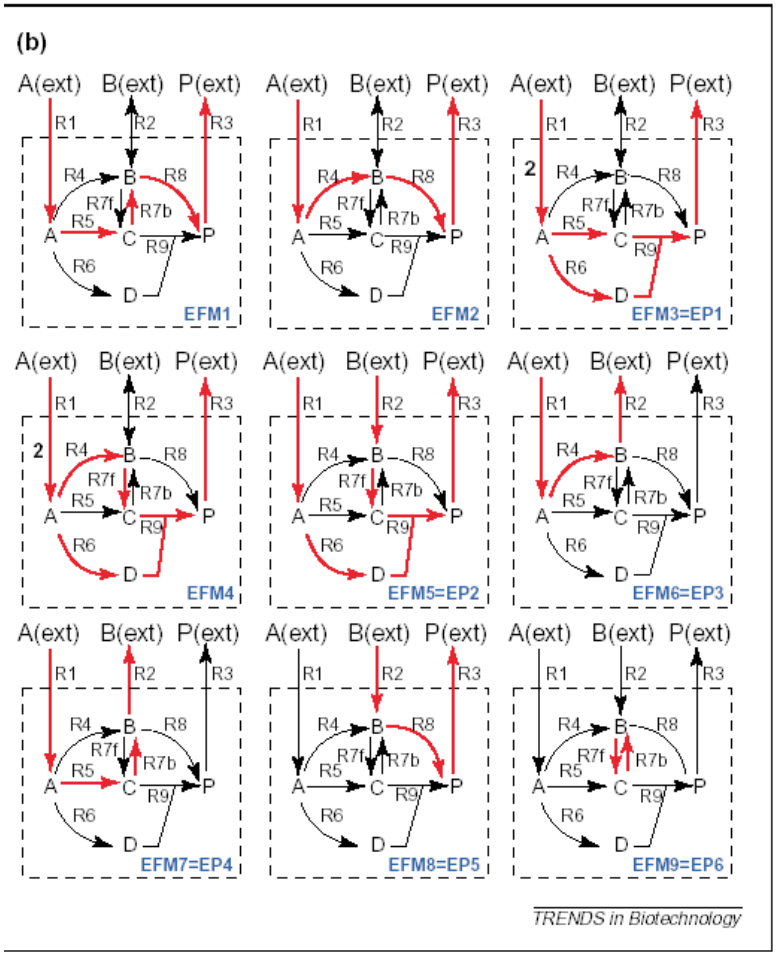


Klamt & Stelling Trends Biotech 21, 64 (2003)

Removing a reaction and mutation studies

Problem

Removing a reaction and mutation studies: effect of deleting R7.



and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

EFM (network N1)

All EFMs not involving the specific reactions build up the complete set of EFMs in the new (smaller) sub-network.

If R7 is deleted, EFMs 2,3,6,8 „survive“. Hence the mutant is viable.

EP (network N2)

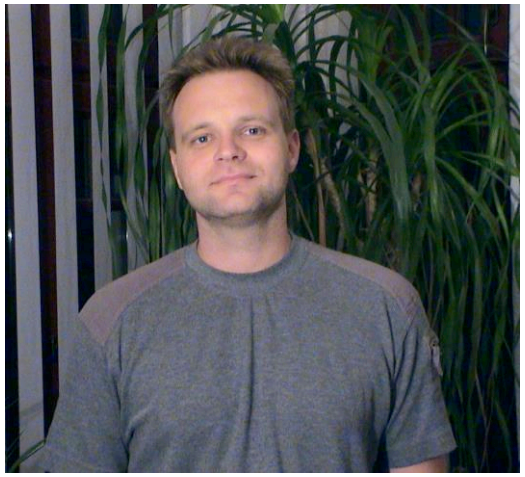
Analyzing a subnetwork implies that the EPs must be newly computed.

E.g. when deleting R2, EFM2 would become an EP.

For this reason, mutation studies cannot be performed easily.

Klamt & Stelling Trends Biotech 21, 64 (2003)

Software: FluxAnalyzer, based on Matlab



Steffen Klamt.

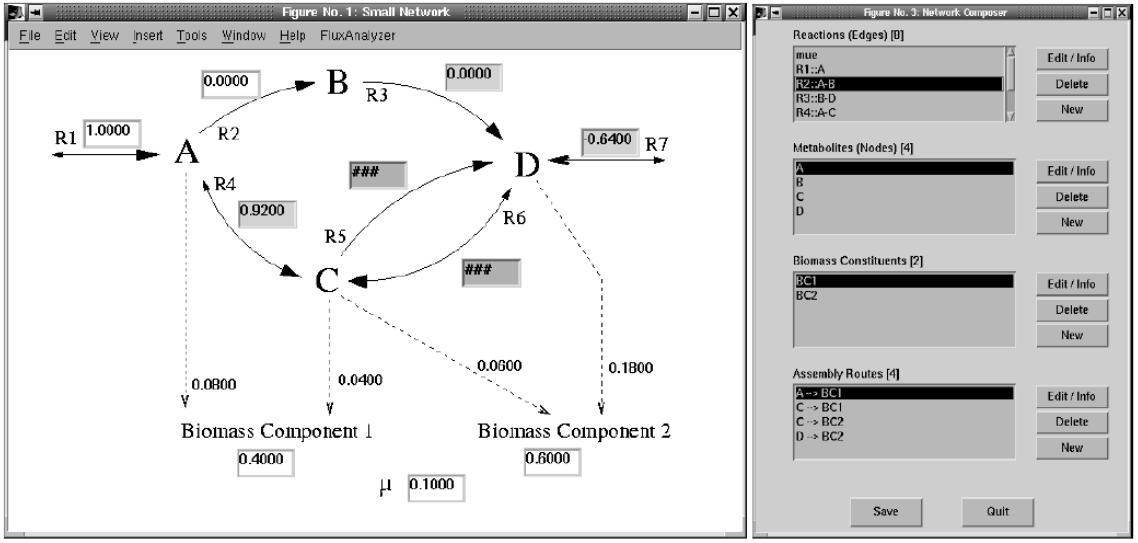


Fig. 1. The network project of 'SMALLNET' constructed by the FluxAnalyzer. Left: interactive flux map displaying a flux scenario (unknown rates are denoted by '###'). Right: network composer.

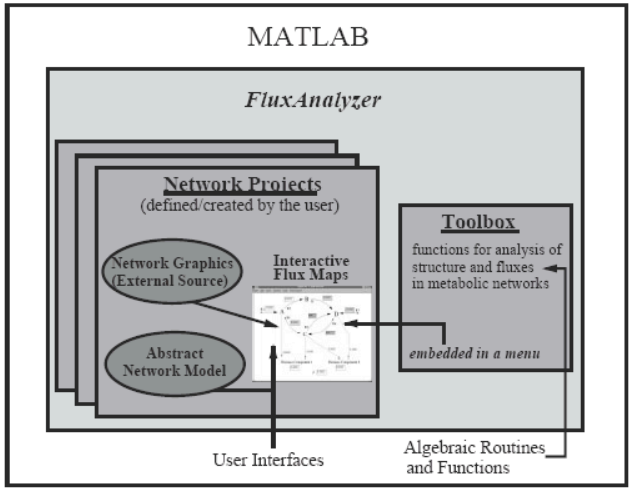


Fig. 2. Structural setup of the FluxAnalyzer.

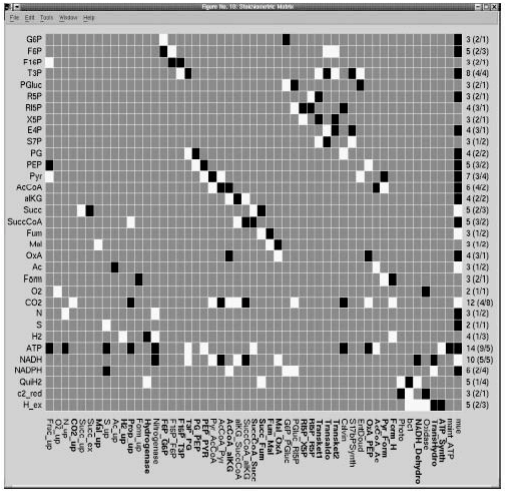


Fig. 3. Concise graphical representation of the stoichiometric matrix (here: catabolic part of the network studied in Klamt *et al.*, 2002)

`FluxAnalyzer` has both EPs and EFMs implemented.

Allows convenient studies of metabolicsystems.

Klamt et al.
Bioinformatics 19, 261 (2003)

Strain optimization based on EFM-analysis

Metabolic Engineering 12 (2010) 112–122

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid

Pornkamol Unrean, Cong T. Trinh, Friedrich Srienc*

Department of Chemical Engineering and Materials Science, and BioTechnology Institute, University of Minnesota, 240 Gortner Laboratory, 1479 Gortner Ave, St. Paul, MN 55108, USA

Carotenoids (e.g. DPL and DPA) are light-harvesting pigments, UV-protecting compounds, regulators of membrane fluidity, and antioxidants.

They are used as nutrient supplements, pharmaceuticals, and food colorants.

Aim: increase carotenoid synthesis in *E.coli*

Unrean et al. Metabol Eng 12, 112-122 (2010)

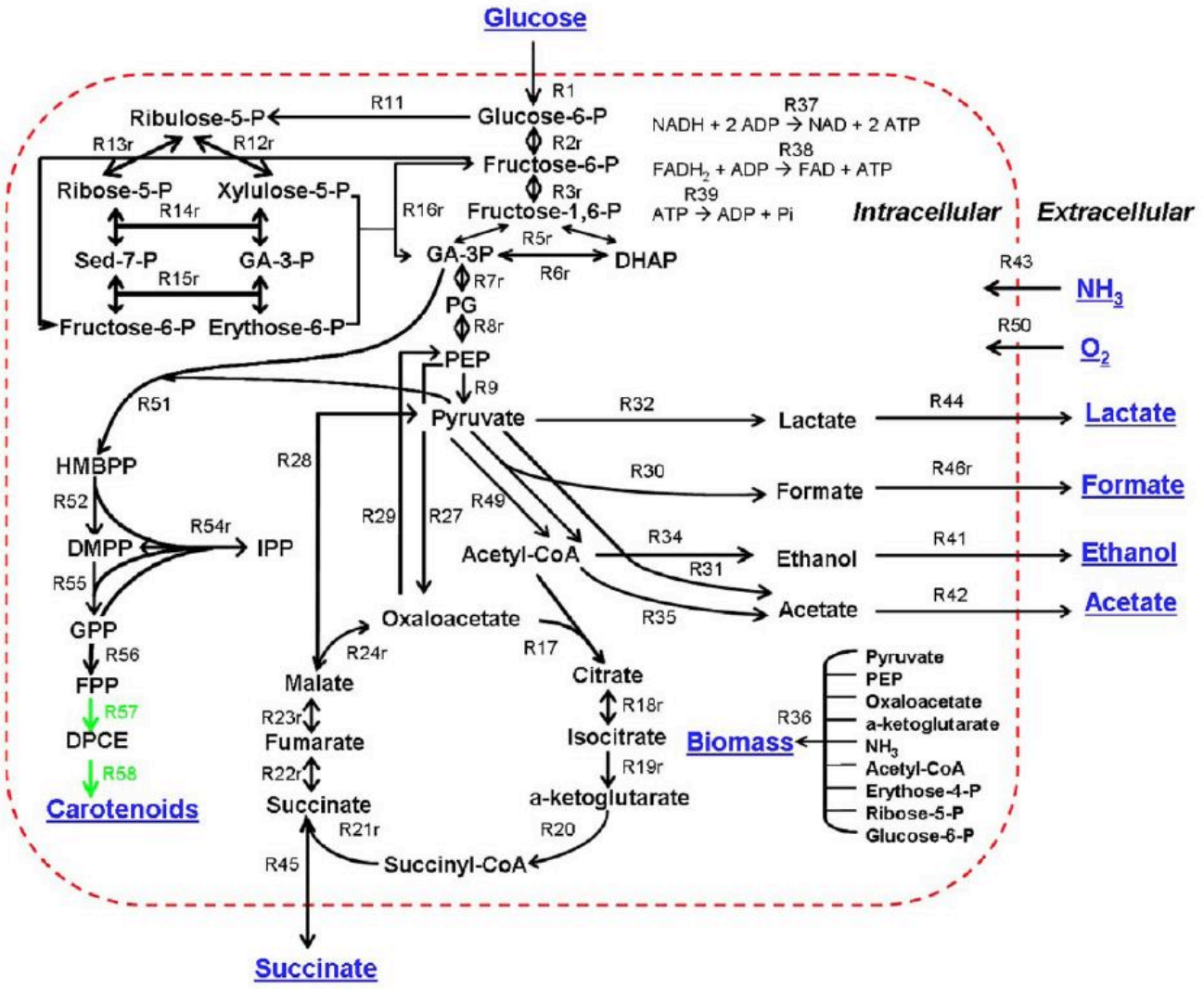
Metabolic network of recombinant *E.coli*

58 metabolic reactions
22 reversible
36 irreversible

57 metabolites

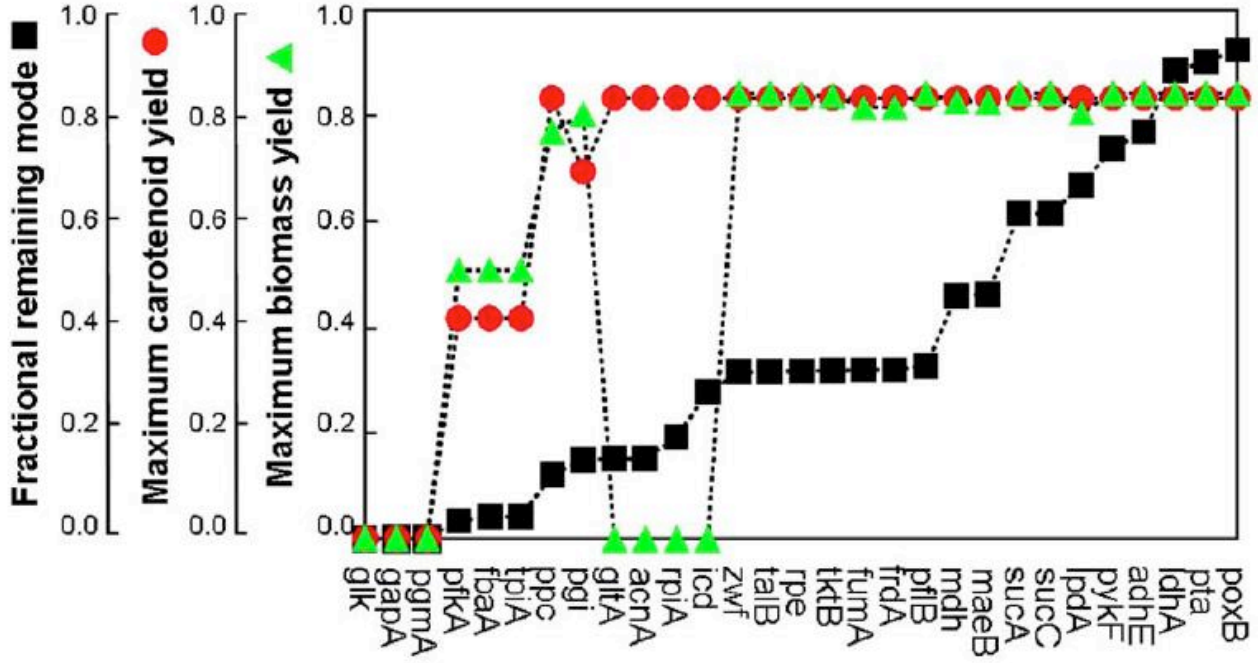
29532 EFMs

In 5923 EFMs, the production of biomass and DPA are coupled.



Unrean et al. Metabol Eng 12, 112-122 (2010)

Effect of single gene deletions



Results of virtual gene knockout calculations (counting number of EFMs and computing their yield from reaction stochiometries).

Select target genes where knockouts still maintain a maximum possible yield of carotenoid production, a reasonable yield of biomass while the largest number of EFMs is eliminated.

Unrean et al. Metabol Eng 12, 112-122 (2010)

Effect of single gene deletions

Strain	Total modes	Aerobic modes	Anaerobic modes	Predicted CRT yield[a]
Wild-type	29,532	24,155	5377	0.0-426
Δ*ldhA*	15,662	13,405	2257	0.0-426
Δ*ldhA*Δ*frdA*	8573	7810	763	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*	7541	6861	680	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*Δ*pta*	6171	5600	571	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*Δ*pta*Δ*adhE*	4099	4099	0	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*Δ*pta*Δ*adhE*Δ*pykF*	2573	2573	0	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*Δ*pta*Δ*adhE*Δ*pykF*Δ*zwf*	375	375	0	0.0-426
Δ*ldhA*Δ*frdA*Δ*poxB*Δ*pta*Δ*adhE*Δ*pykF*Δ*zwf*Δ*maeB*	5	5	0	0.4-426

[a] Yield is in mg-diapolycopendioic acid/g–glucose.

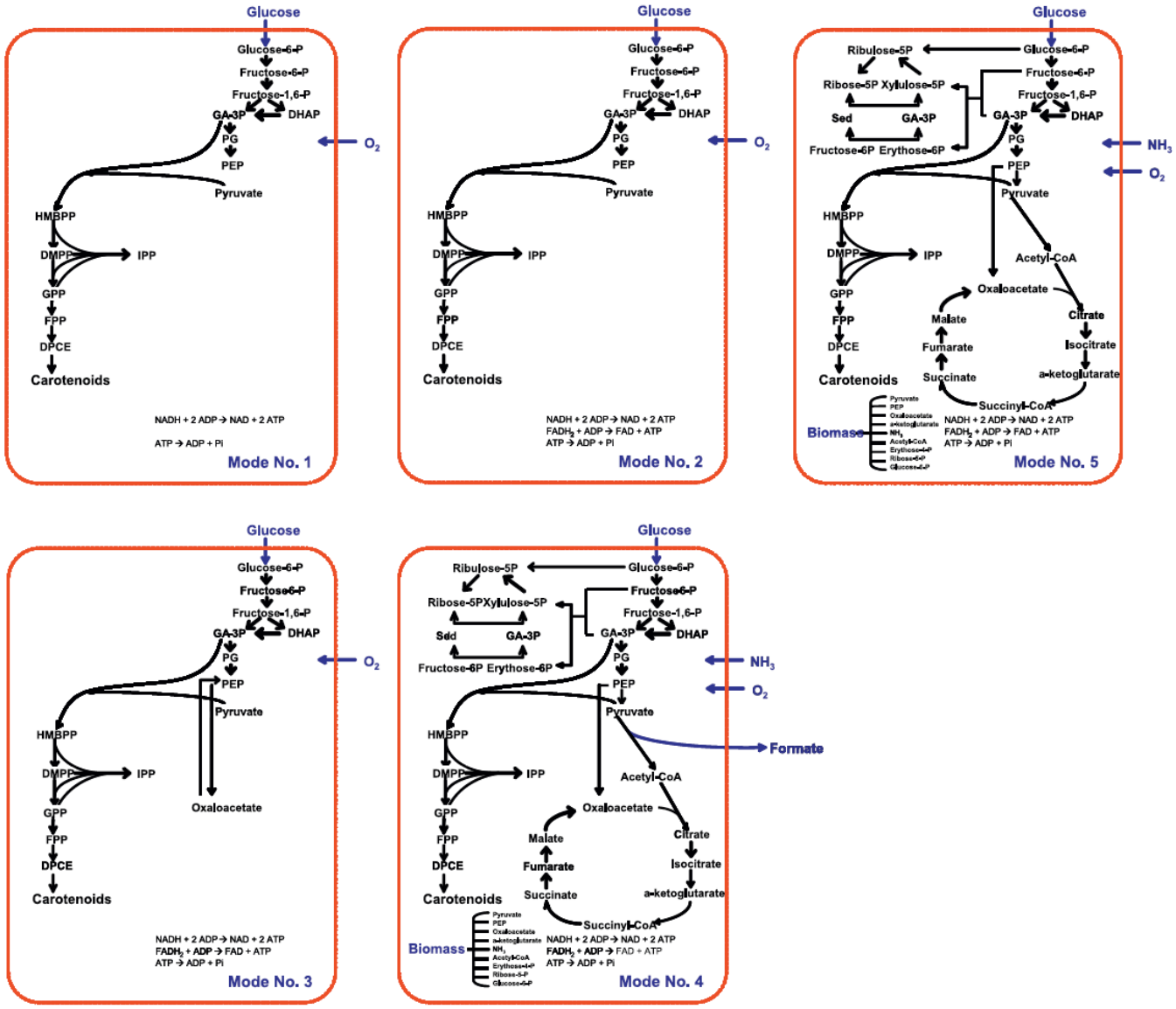
Deleted Reaction	Corresponding gene	Enzyme	Pathway
R9	*pykF*	Pyruvate kinase	Glycolysis
R11	*zwf*	Glucose-6-phosphate-1-dehydrogenase	Pentose phosphate
R22	*frdA*	Fumarate reductase	Fermentation
R28	*maeB*	Malate dehydrogenase	Anapleurotic
R31	*poxB*	Pyruvate oxidase	Fermentation
R32	*ldhA*	Lactate dehydrogenase	Fermentation
R34	*adhE*	Alcohol dehydrogenase	Fermentation
R35	*pta*	Phosphate acetyltransferase	Fermentation

Optimal: 8 gene knockouts lead to predicted over-production of DPL and DPA.

Only 5 EFMs remain.

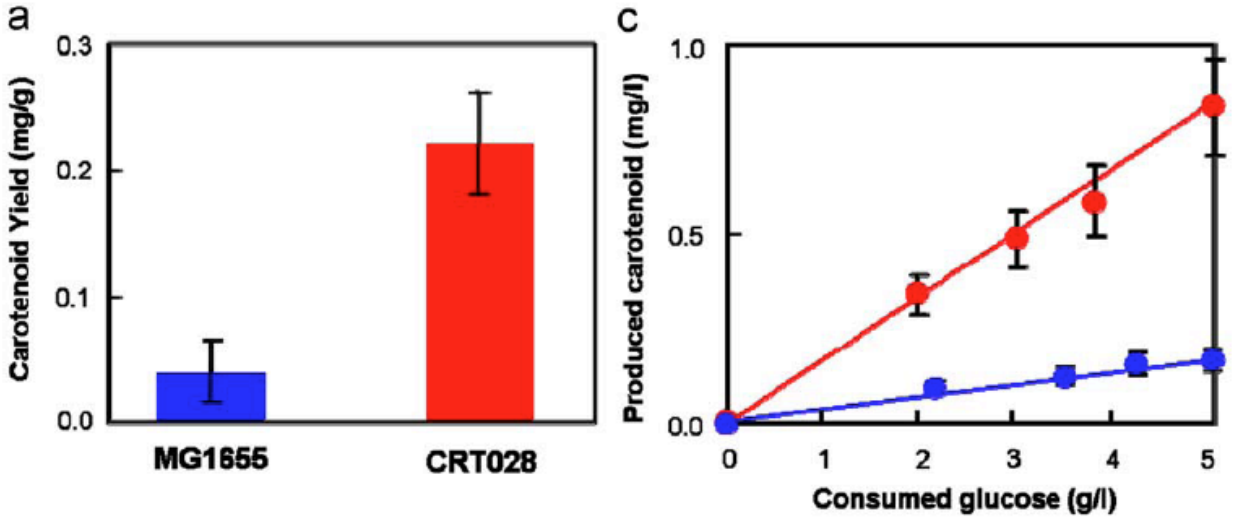
Unrean et al. Metabol Eng 12, 112-122 (2010)

Remaining EFMs



Unrean et al. Metabol Eng 12, 112-122 (2010)

Experimental verification: increased carotenoid yield



Mutant grows slower, but CRT production is increased 4 times.

	MG1655/ pACMNOx	CRT028/ pACMNOx
Growth rate (/h)	0.17 ± 0.02	0.13 ± 0.01
Carotenoid production (mg/l)	0.19 ± 0.02	0.83 ± 0.20
Carotenoid yield (mg carotenoid/g glucose)	0.04 ± 0.00	0.17 ± 0.04
Specific production (mg carotenoid/g cell dry weight-h)	0.01 ± 0.00	0.10 ± 0.02

Unrean et al. Metabol Eng 12, 112-122 (2010)

Complexity of finding and enumerating EFMs

Theorem: Given a stochiometric matrix *S*, an elementary mode can be found in polynomial time.

Theorem: In case all reactions in a metabolic network are reversible, the elementary modes can be enumerated in polynomial time.

The enumeration task becomes dramatically more difficult if the reactions are irreversible.
In this case, the modes of the network form a cone, and the elementary modes are the rays of the cone.

Theorem: Given a flux cone and two coordinates *i* and *j*, deciding if there exists and extreme ray of the cone that has both r_i and r_j in its support is NP-complete.

Theorem: Given a matrix *S* and a number *k*, deciding the existence of an elementary mode with at most *k* reactions in its support is NP-complete.

The question whether all elementary modes of a general network can be enumerated in polynomial time is an open question.

Acuna et al. BioSystems 99, 210-214 (2010); BioSystems 95, 51-60 (2009)

Summary EFMs

EFMs are a robust method that offers great opportunities for studying functional and structural properties in metabolic networks.

The **decomposition** of a particular flux distribution (e.g. determined by experiment) as a linear combination of EFMs is **not unique**.

Klamt & Stelling suggest that the term „elementary flux modes“ should be used whenever the sets of EFMs and EPs are identical.
In cases where they don‘t, EPs are a subset of EFMs.

It remains to be understood more thoroughly how much valuable information about the pathway structure is lost by using EPs.

Ongoing Challenges:
- study really large metabolic systems by subdividing them into sub-systems
- combine metabolic model with model of cellular regulation.

Klamt & Stelling Trends Biotech 21, 64 (2003)