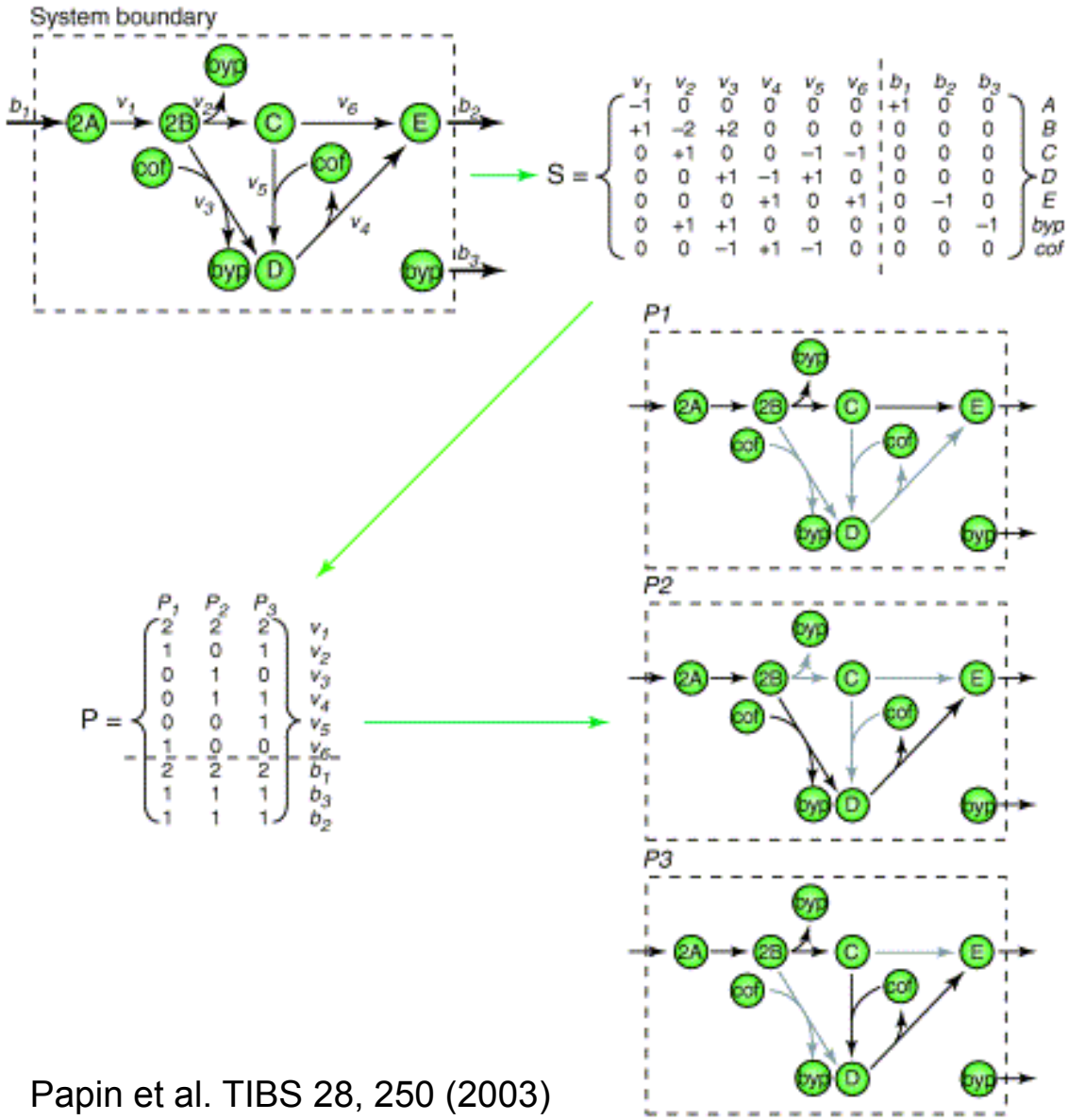
V15 Flux Balance Analysis – Extreme Pathways

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.

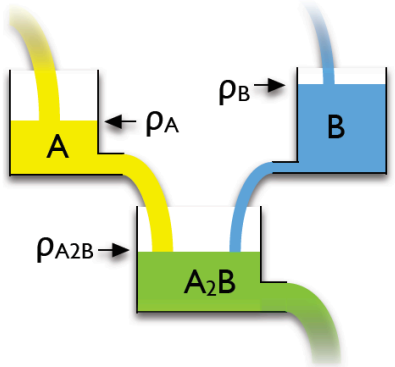
With the matrix, the methods of extreme pathway and elementary mode analyses can be used to generate a unique set of pathways P1, P2, and P3 that allow to express all steady-state fluxes as linear combinations of P1 – P3.



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.



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

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

For each metabolite X_i :

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

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**.

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

Consider

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

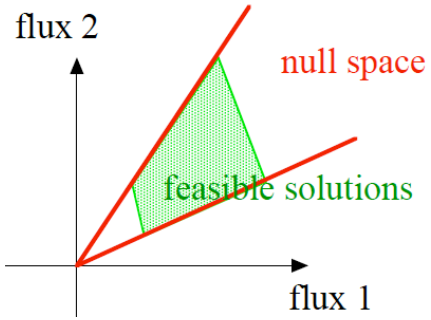
Corresponds to $\quad 2x_2 + x_3 = 0, \quad 3x_1 - x_2 + x_3 = 0 \quad \Leftrightarrow \quad 2x_2 = -x_3, \quad 2x_1 = -x_3$

=> one free parameter: x_3

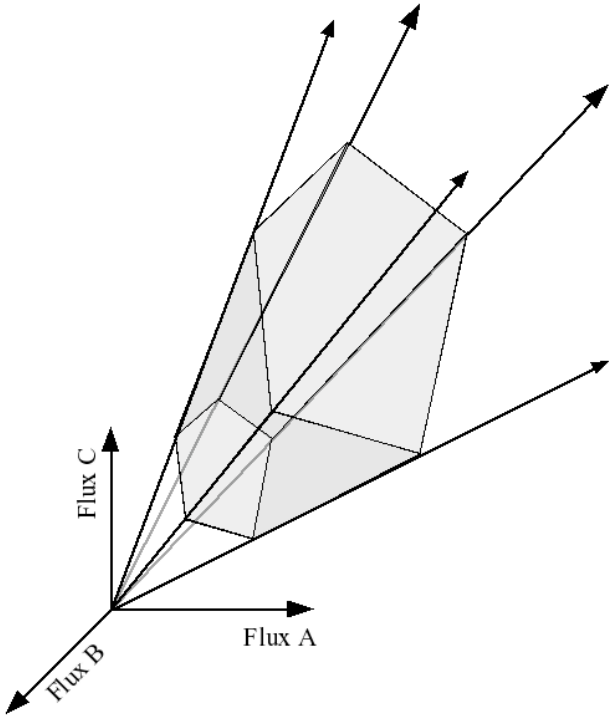
null space: $\vec{x} = \begin{pmatrix} -a \\ -a \\ 2a \end{pmatrix}$

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)

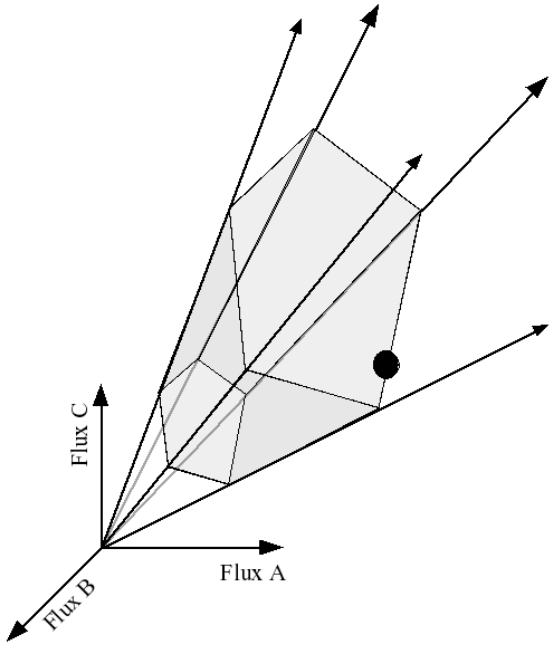
True biological flux

To find the „true“ biological flux in cells (→ e.g. Heinzle, 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, sucA, 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, gatD, gatY, glk, glpK, gntK, gntV, gpsA, lacZ, manA, melA, mtlD, nagA, nagB, nanA, 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, setA, serB, serC, speA, speB, speC, speD, speE, speF, tdcB, tdh, thrA, thrB, thrC, tnaA, trpA, trpB, trpC, trpD, trpE, tyrA, tyrA, tyrB, ygjG, ygjH, alaB* (42), *dapC* (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, gatA, gatB, gatC, glnH, glnP, glnQ, glpF, glpT, gltJ, gltK, gltL, gltP, glt5, 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.

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

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)

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)

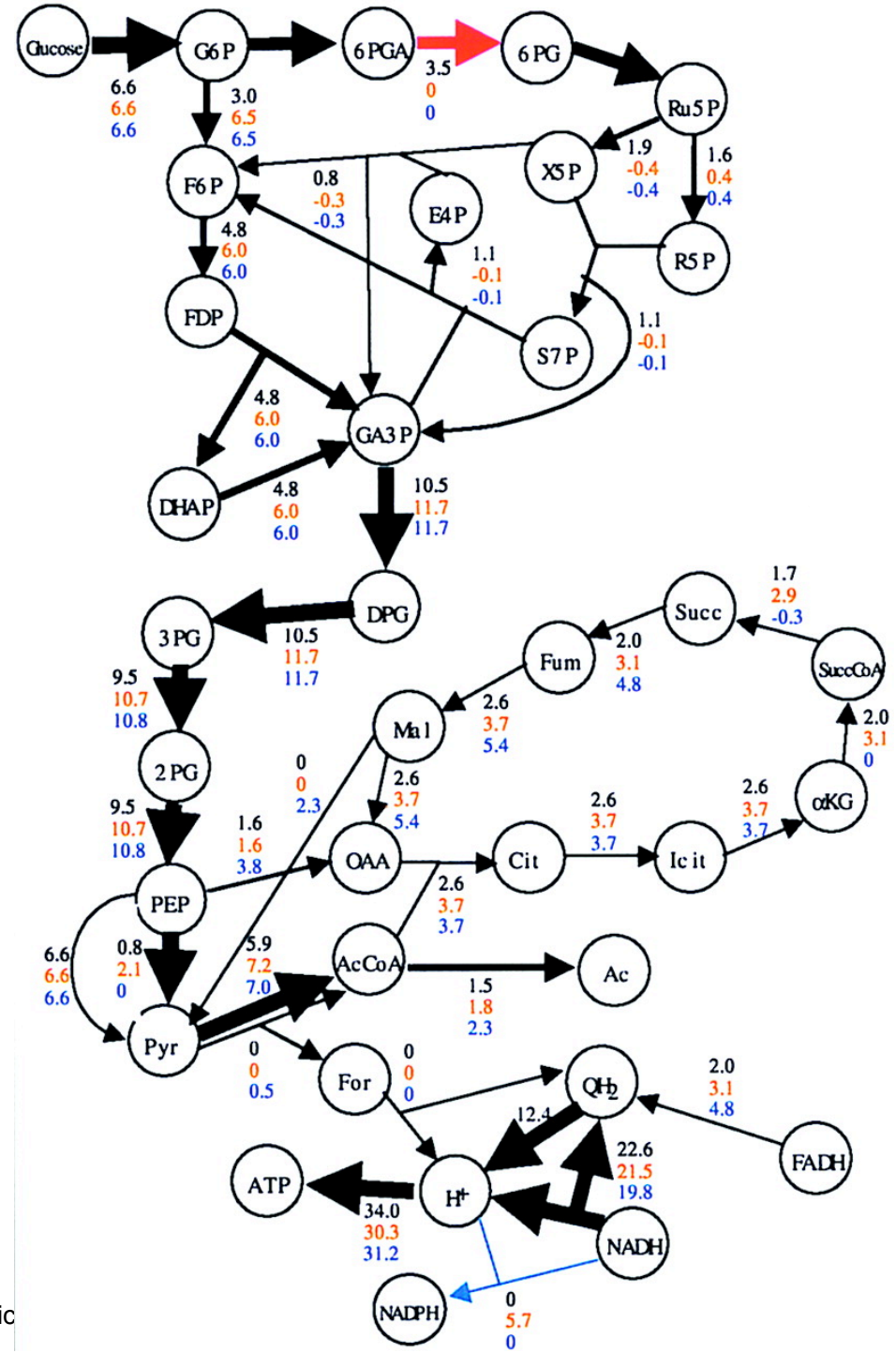
Rerouting of metabolic fluxes

(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.

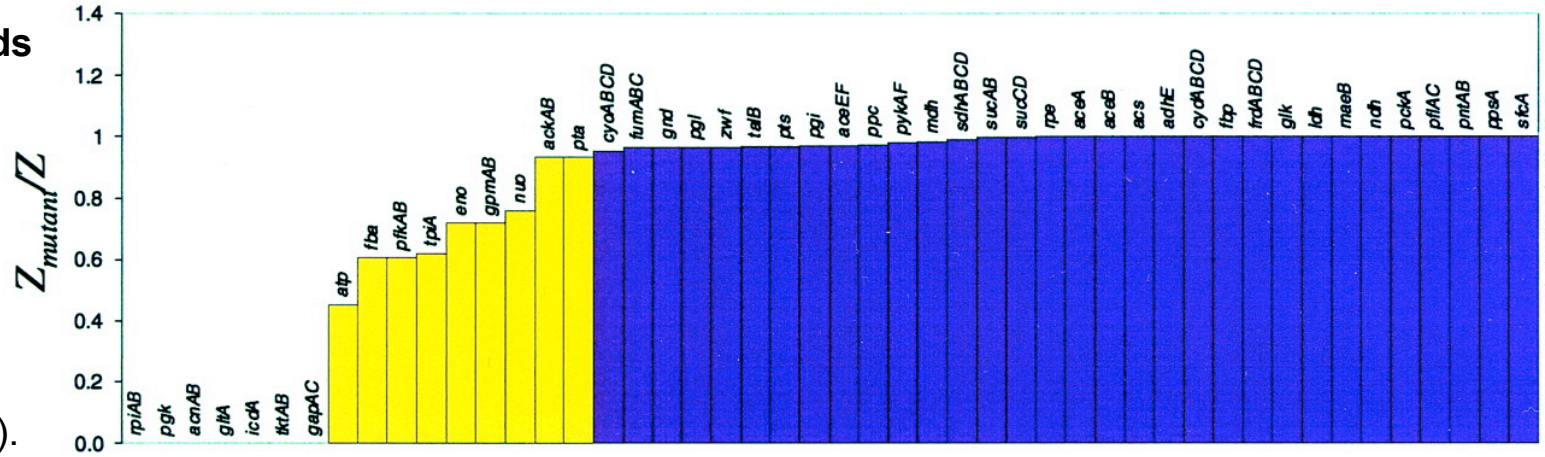
Note how *E.coli in silico* circumvents removal of one critical reaction (red arrow) by increasing the flux through the alternative G6P → P6P reaction.



Edwards & Palsson PNAS 97, 5528 (2000)

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 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.

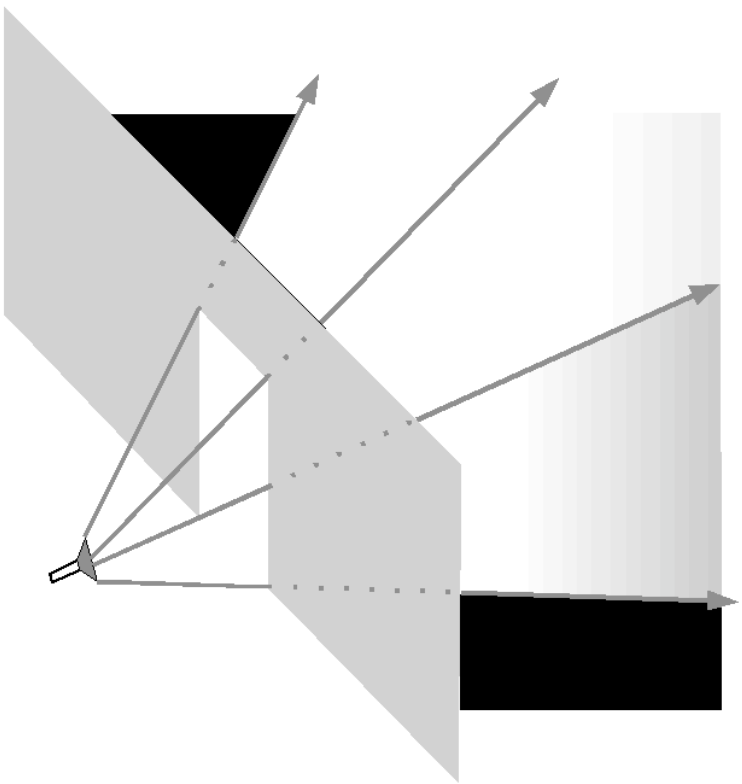
FBA shows that the *E.coli* metabolic network contains relatively **few critical gene products** in central metabolism.
However, the ability to adjust to different environments (growth conditions) may be diminished by gene deletions.

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)

Idea – extreme pathways



A torch is directed at an open door and shines into a dark room ...

What area is lighted ?

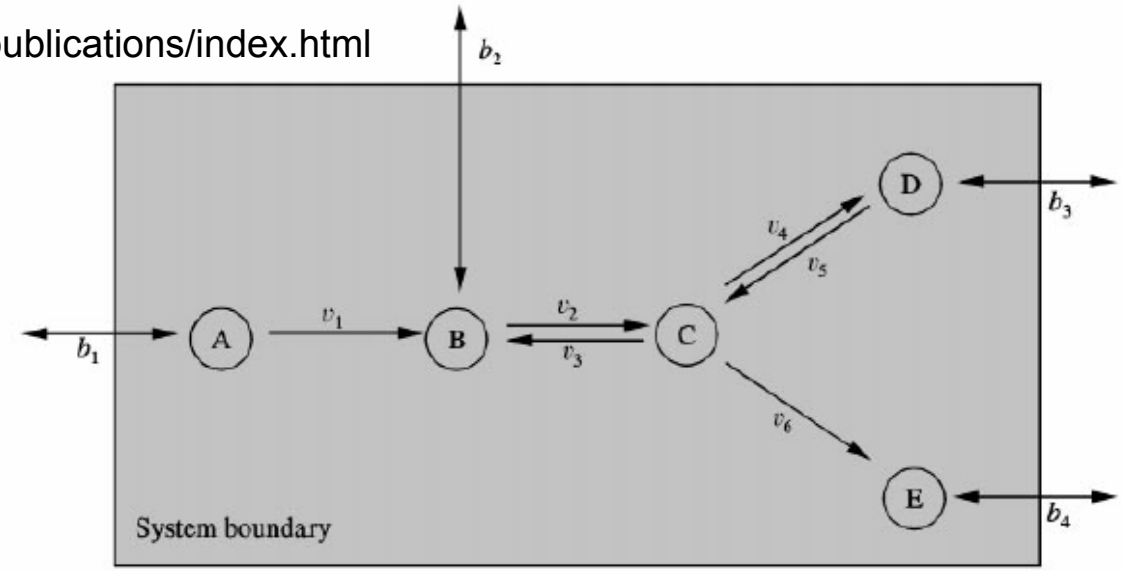
Instead of marking all lighted points individually,
it would be sufficient to characterize the „extreme rays“ that go through the corners of the door.

The lighted area is the area between the extreme rays = linear combinations of the extreme rays.

Extreme Pathways

introduced into metabolic analysis by the lab of Bernard Palsson (Dept. of Bioengineering, UC San Diego). The publications of this lab are available at http://gcrg.ucsd.edu/publications/index.html

The extreme pathway technique is based on the stoichiometric matrix representation of metabolic networks.



All external fluxes are defined as pointing outwards.

Mass balance constraints

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

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

Internal flux constraints

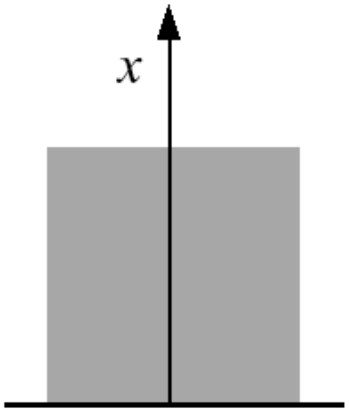
$$v_j \geq 0, \quad j = 1,\ldots,6$$

Exchange flux constraints

$$-\infty \leq b_j \leq +\infty, \quad j = 1,\ldots,4$$

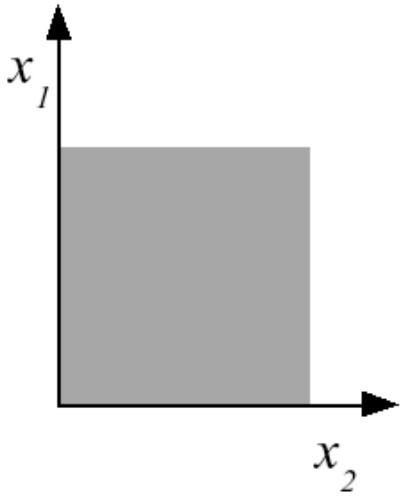
Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

Idea – extreme pathways



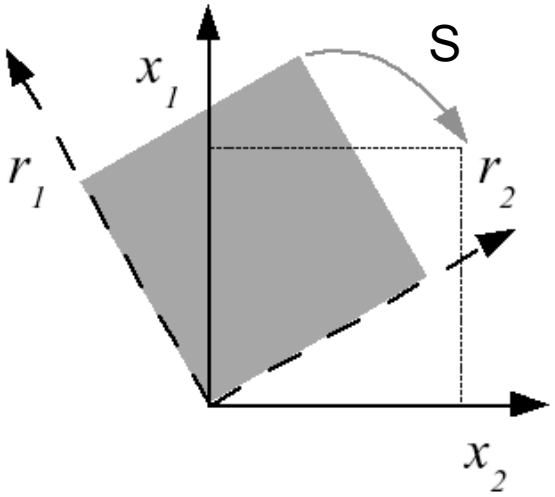
Shaded area:

$x \geq 0$



Shaded area:

$x_1 \geq 0 \wedge x_2 \geq 0$



Either **S** · **x** ≥ **0**

(**S** acts as rotation matrix)

or find optimal vectors

✹ change coordinate system

from x_1, x_2 to r_1, r_2.

Shaded area:

$r_1 \geq 0 \wedge r_2 \geq 0$

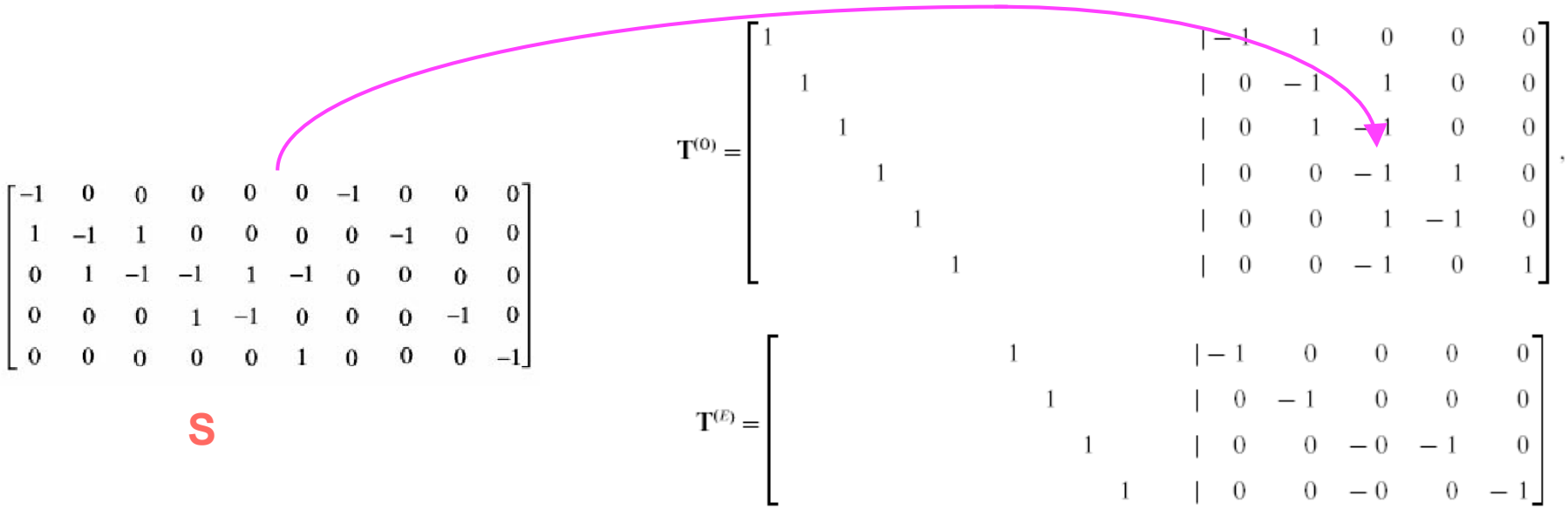
Duality of two matrices S and R.

Edwards & Palsson PNAS 97, 5528 (2000)

Extreme Pathways – algorithm - setup

The algorithm to determine the set of extreme pathways for a reaction network follows the pinciples of algorithms for finding the extremal rays/ generating vectors of convex polyhedral cones.

Combine $n \times n$ identity matrix (**I**) with the transpose of the stoichiometric matrix $\mathbf{S}^T$. **I** serves for bookkeeping.



Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

separate internal and external fluxes

Examine constraints on each of the exchange fluxes as given by

$$\alpha_j \leq b_j \leq \beta_j$$

If the exchange flux is constrained to be positive → do nothing.

If the exchange flux is constrained to be negative → multiply the corresponding row of the initial matrix by -1.

If the exchange flux is unconstrained → move the entire row to a temporary matrix $\mathbf{T}^{(E)}$. This completes the first tableau $\mathbf{T}^{(0)}$.

$\mathbf{T}^{(0)}$ and $\mathbf{T}^{(E)}$ for the example reaction system are shown on the previous slide.

Each element of these matrices will be designated T_{ij}.

Starting with $i = 1$ and $T^{(0)} = T^{(i-1)}$ the next tableau is generated in the following way:

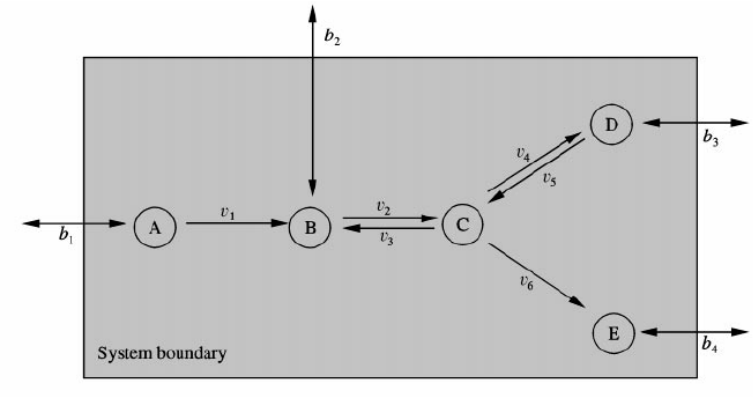
Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

idea of algorithm

(1) Identify all metabolites that do not have an unconstrained exchange flux associated with them.

The total number of such metabolites is denoted by μ.

The example system contains only one such metabolite, namely C (μ = 1).



What is the main idea?

- We want to find balanced extreme pathways that don‘t change the concentrations of metabolites when flux flows through (input fluxes are channelled to products not to accumulation of intermediates).
- The stochiometrix matrix describes the coupling of each reaction to the concentration of metabolites X.
- Now we need to balance combinations of reactions that leave concentrations unchanged. Pathways applied to metabolites should not change their concentrations → the matrix entries need to be brought to 0.

Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

keep pathways that do not change concentrations of internal metabolites

(2) Begin forming the new matrix $\mathbf{T}^{(i)}$ by copying all rows from $\mathbf{T}^{(i-1)}$ which already contain a zero in the column of $\mathbf{S}^T$ that corresponds to the first metabolite identified in step 1, denoted by index *C*. (Here 3rd column of $\mathbf{S}^T$.)

$T^{(0)}$ =

						A	B	**C**	D	E
1						-1	1	0	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1

$T^{(1)}$ =

1						-1	1	0	0	0

+

Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

balance combinations of other pathways

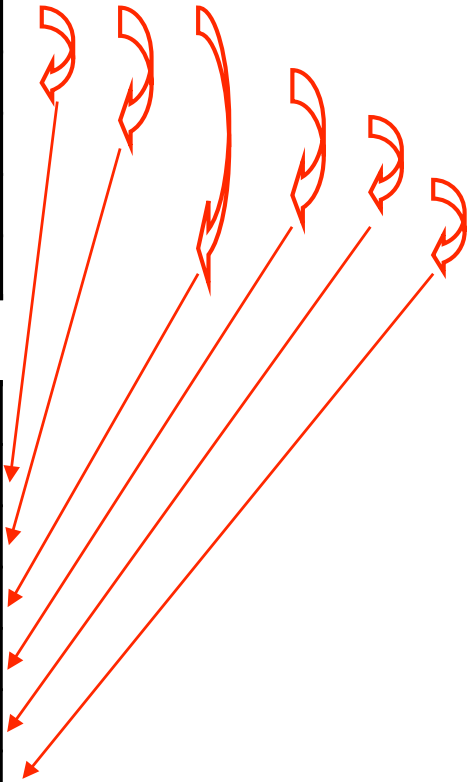
(3) Of the remaining rows in $\mathbf{T}^{(i-1)}$ add together all possible combinations of rows which contain values of the opposite sign in column C, such that the addition produces a zero in this column.

$T^{(0)}$ =

1						-1	1	**0**	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1

$T^{(1)}$ =

1	0	0	0	0	0	-1	1	0	0	0
0	1	1	0	0	0	0	0	**0**	0	0
0	1	0	1	0	0	0	-1	**0**	1	0
0	1	0	0	0	1	0	-1	**0**	0	1
0	0	1	0	1	0	0	1	**0**	-1	0
0	0	0	1	1	0	0	0	**0**	0	0
0	0	0	0	1	1	0	0	**0**	-1	1
1	2	3	4	5	6	7	8	9	10	11



Schilling, et al.
JTB 203, 229

remove “non-orthogonal” pathways

(4) For all rows added to $\mathbf{T}^{(i)}$ in steps 2 and 3 check that no row exists that is a non-negative combination of any other rows in $\mathbf{T}^{(i)}$.

One method for this works is as follows:
let *A(i)* = set of column indices *j* for which the elements of row *i* = 0.

For the example above
A(1) = {2,3,4,5,6,9,10,11}
A(2) = {1,4,5,6,7,8,9,10,11}
A(3) = {1,3,5,6,7,9,11}
A(4) = {1,3,4,5,7,9,10}
A(5) = {1,2,4,6,7,9,11}
A(6) = {1,2,3,6,7,8,9,10,11}
A(7) = {1,2,3,4,7,8,9}

Then check to determine if there exists another row *(h)* for which *A(i)* is a subset of A*(h)*.

If $A(i) \subseteq A(h),\ i \neq h$
where
$A(i) = \{ j : T_{i,j} = 0,\ 1 \leq j \leq (n+m) \}$
then row *i* must be eliminated from $\mathbf{T}^{(i)}$

Schilling et al.
JTB 203, 229

repeat steps for all internal metabolites

(5) With the formation of $\mathbf{T}^{(i)}$ complete steps 2 – 4 for all of the metabolites that do not have an unconstrained exchange flux operating on the metabolite, incrementing *i* by one up to μ. The final tableau will be $T^{(\mu)}$.

Note that the number of rows in $T^{(\mu)}$ will be equal to *k*, the number of extreme pathways.

Schilling et al.
JTB 203, 229

balance external fluxes

(6) Next we append $\mathbf{T}^{(E)}$ to the bottom of $\mathbf{T}^{(\mu)}$. (In the example here $\mu = 1$.)
This results in the following tableau:

$T^{(1/E)} =$

1										-1	1	0	0	0
	1	1								0	0	0	0	0
	1		1							0	-1	0	1	0
	1				1					0	-1	0	1	0
		1		1						0	1	0	-1	0
			1	1						0	0	0	0	0
				1	1					0	0	0	-1	1
						1				-1	0	0	0	0
							1			0	-1	0	0	0
								1		0	0	0	-1	0
									1	0	0	0	0	-1

Schilling et al.
JTB 203, 229

balance external fluxes

(7) Starting in the *n+1* column (or the first non-zero column on the right side), if $T_{i,(n+1)} \neq 0$ then add the corresponding non-zero row from $\mathbf{T}^{(E)}$ to row *i* so as to produce 0 in the *n+1*-th column.
This is done by simply multiplying the corresponding row in $\mathbf{T}^{(E)}$ by $T_{i,(n+1)}$ and adding this row to row *i* .

Repeat this procedure for each of the rows in the upper portion of the tableau so as to create zeros in the entire upper portion of the *(n+1)* column.

When finished, remove the row in $\mathbf{T}^{(E)}$ corresponding to the exchange flux for the metabolite just balanced.

Schilling et al.
JTB 203, 229

balance external fluxes

(8) Follow the same procedure as in step (7) for each of the columns on the right side of the tableau containing non-zero entries.

(In our example we need to perform step (7) for every column except the middle column of the right side which correponds to metabolite C.)

The final tableau $\mathbf{T}^{(final)}$ will contain the transpose of the matrix **P** containing the extreme pathways in place of the original identity matrix.

Schilling et al.
JTB 203, 229

pathway matrix

$\mathbf{T}^{(final)}$ =

1						-1	1			0	0	0	0	0	0
	1	1								0	0	0	0	0	0
	1		1				-1	1		0	0	0	0	0	0
	1				1		-1		1	0	0	0	0	0	0
		1		1			1	-1		0	0	0	0	0	0
			1	1						0	0	0	0	0	0
				1	1			-1	1	0	0	0	0	0	0

$\mathbf{P}^{T}$ =

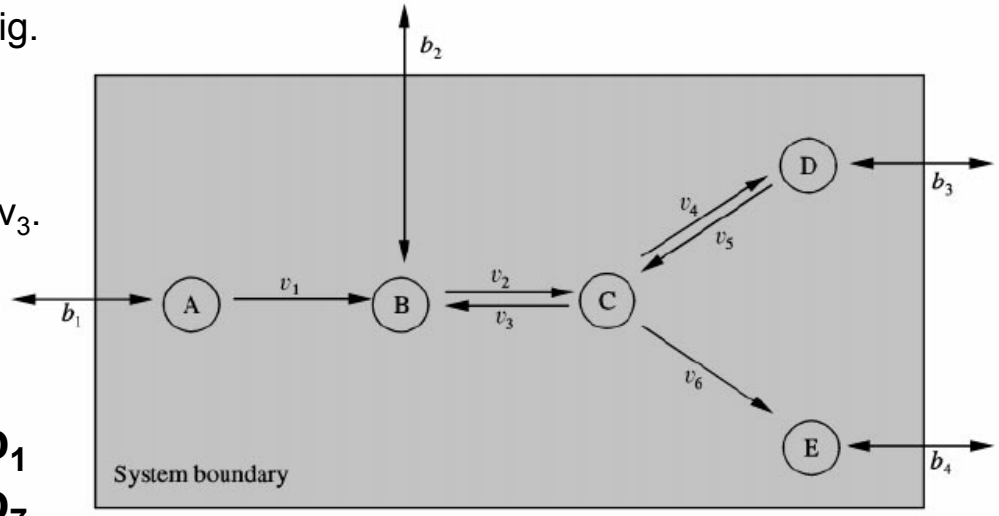
v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4	
1	0	0	0	0	0	-1	1	0	0	$\mathbf{p_1}$
0	1	1	0	0	0	0	0	0	0	$\mathbf{p_7}$
0	1	0	1	0	0	0	-1	1	0	$\mathbf{p_3}$
0	1	0	0	0	1	0	-1	0	1	$\mathbf{p_2}$
0	0	1	0	1	0	0	1	-1	0	$\mathbf{p_4}$
0	0	0	1	1	0	0	0	0	0	$\mathbf{p_6}$
0	0	0	0	1	1	0	0	-1	1	$\mathbf{p_5}$

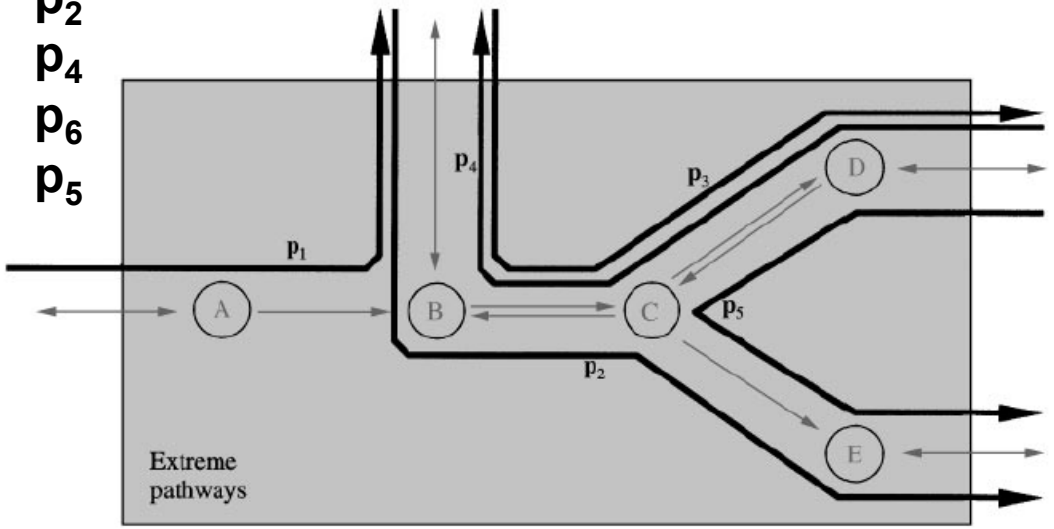
Schilling et al.
JTB 203, 229

Extreme Pathways for model system

2 pathways p_6 and p_7 are not shown in the bottom fig. because all exchange fluxes with the exterior are 0. Such pathways have no net overall effect on the functional capabilities of the network.
They belong to the cycling of reactions v_4/v_5 and v_2/v_3.

	v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4
p_1	1	0	0	0	0	0	-1	1	0	0
p_7	0	1	1	0	0	0	0	0	0	0
p_3	0	1	0	1	0	0	0	-1	1	0
p_2	0	1	0	0	0	1	0	-1	0	1
p_4	0	0	1	0	1	0	0	1	-1	0
p_6	0	0	0	1	1	0	0	0	0	0
p_5	0	0	0	0	1	1	0	0	-1	1



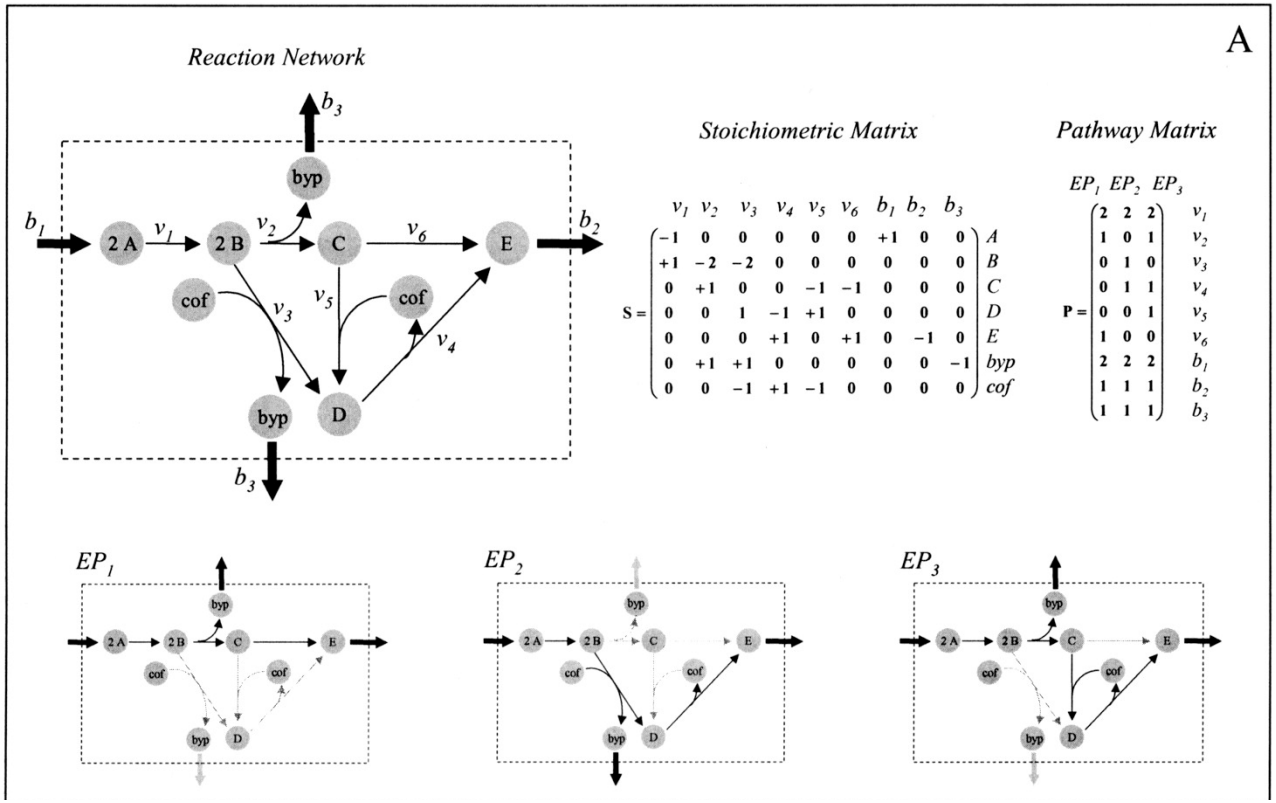


Schilling et al.
JTB 203, 229

How reactions appear in pathway matrix

In the matrix **P** of extreme pathways, each column is an EP and each row corresponds to a reaction in the network.

The numerical value of the *i,j*-th element corresponds to the relative flux level through the *i*-th reaction in the *j*-th EP.



Papin, Price, Palsson,
Genome Res. 12, 1889 (2002)

Properties of pathway matrix

After normalizing **P** to a matrix with entries 0 or 1,
the symmetric Pathway Length Matrix $\mathbf{P}_{LM}$ can be calculated:

$$\mathbf{P}_{LM} = \mathbf{P}^T \cdot \mathbf{P}$$

where the values along the diagonal correspond to the length of the EPs.

Pathway Length

$$\mathbf{P} = \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \tilde{\mathbf{P}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \tilde{\mathbf{P}}^T \bullet \tilde{\mathbf{P}} = \begin{array}{c} \begin{matrix} EP_1 & EP_2 & EP_3 \end{matrix} \\ \begin{pmatrix} 6 & 4 & \textcircled{5} \\ & 6 & 5 \\ & & 7 \end{pmatrix} \end{array} \begin{matrix} EP_1 \\ EP_2 \\ EP_3 \end{matrix}$$

Comments:

1) The lengths of EP_1, EP_2,and EP_3 are 6, 6, and 7, respectively, the highlighted diagonal elements of the final matrix.

2) EP_2 and EP_3 have a shared length of 5 (indicated by the circle). As seen in the schematics above, they share reactions v_1, v_4, b_1, b_2, and b_3.

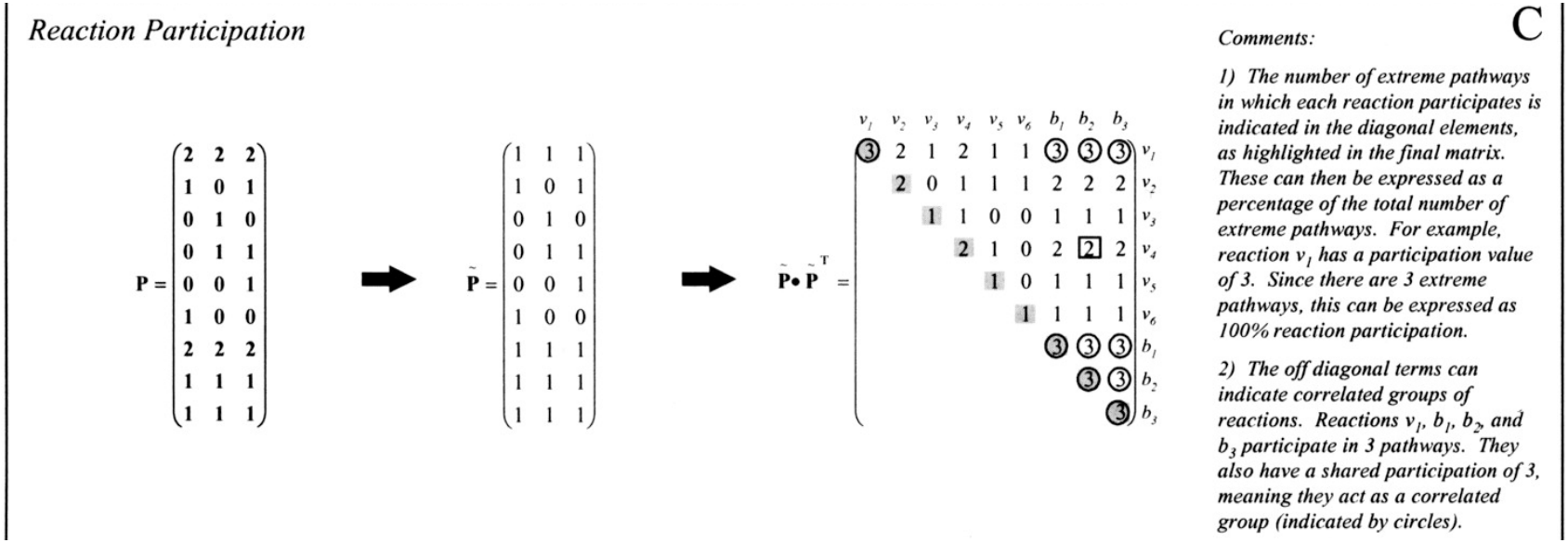
The off-diagonal terms of $\mathbf{P}_{LM}$ are the number of reactions that a pair of extreme pathways have in common.

Properties of pathway matrix

One can also compute a reaction participation matrix $\mathbf{P}_{PM}$ from $\mathbf{P}$:

$$\mathbf{P}_{PM} = \mathbf{P} \cdot \mathbf{P}^{T}$$

where the diagonal correspond to the number of pathways in which the given reaction participates.



EP Analysis of *H. pylori* and *H. influenza*

Amino acid synthesis in *Heliobacter pylori* vs. *Heliobacter influenza* studied by EP analysis.

Table 4. Summary of the Statistical Analyses of Extreme Pathway Lengths

H. pylori Target product	Number of EPs	Pathway length average	maximum	minimum	coefficient of variation
Asparagine	340	44	54	28	15%
Aspartic Acid	491	43	52	24	14%
Cysteine	1022	59	71	45	10%
Glutamine	315	41	53	23	18%
Glutamic Acid	493	41	53	25	17%
Glycine	377	51	60	38	10%
Lysine	611	54	66	39	12%
Proline	867	43	56	15	16%
Serine	355	45	54	33	12%
Threonine	469	48	60	31	14%
Tryptophan	1958	64	73	51	6%
Tyrosine	1008	58	68	44	7%
Equimolar Amino Acids	6032	106	112	99	2%
E. coli Ratio Amino Acids	5553	106	112	99	2%
H. influenzae Target product	Number of EPs	Pathway length average	maximum	minimum	coefficient of variation
Alanine	1739	36	49	18	10%
Asparagine	445	39	52	29	13%
Aspartic Acid	690	35	49	27	14%
Glutamine	690	37	46	28	11%
Glycine	456	39	48	35	7%
Histidine	1507	65	74	61	3%
Isoleucine	1480	47	61	37	9%
Leucine	3884	42	55	31	10%
Lysine	1168	47	61	37	9%
Methionine	1343	48	63	40	8%
Phenylalanine	1758	51	64	43	7%
Proline	2624	38	51	25	11%
Serine	690	37	50	30	10%
Threonine	1318	42	55	32	10%
Tryptophan	3540	58	69	49	6%
Tyrosine	1758	51	64	43	7%
Valine	1739	39	52	23	9%

The coefficient of variation is the standard deviation normalized to the average (expressed as a percent). Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass. EPs, extreme pathways.

Table 1. Number of Reactions Involved in the Production of the Indicated Target Product

H. pylori Target product	Essential reactions	Utilized reactions
Tryptophan	32	105
Tyrosine	28	101
Cysteine	25	102
Glycine	22	97
Lysine	22	102
Serine	16	91
Threonine	14	96
Asparagine	13	91
Aspartic Acid	12	91
Proline	10	91
Glutamic Acid	7	91
Glutamine	6	91
Equimolar Amino Acids	85	140
E. coli Ratio Amino Acids	85	140
H. influenzae Target product	Essential reactions	Utilized reactions
Histidine	51	112
Tryptophan	41	108
Phenylalanine	36	108
Tyrosine	36	108
Methionine	34	106
Isoleucine	31	108
Lysine	31	108
Glycine	29	82
Threonine	26	103
Asparagine	25	98
Serine	25	97
Leucine	23	105
Aspartic Acid	22	97
Glutamine	21	102
Proline	18	103
Valine	17	102
Alanine	12	99

See Fig. 3 for the indicated network inputs and outputs. Essential reactions refers to the number of reactions that were used in every extreme pathway (region I in Fig. 4). Utilized reactions refers to the number of reactions that were used at least once in the set of extreme pathways for the production of the associated product (region II in Fig. 4). The individual amino acids are sorted in descending order according to the number of essential reactions. Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Extreme Pathway Analysis

Calculation of EPs for increasingly large networks is computationally intensive and results in the generation of large data sets.

Even for integrated genome-scale models for microbes under simple conditions, EP analysis can generate thousands or even millions of vectors!

It turned out that the number of reactions that participate in EPs that produce a particular product is usually poorly correlated to the product yield and the molecular complexity of the product.

Possible way out?

Matrix diagonalisation – eigenvectors: only possible for quadratic $n \times n$ matrices with rank n.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Quasi-diagonalisation of pathway matrix by SVD

Suppose M is an $m \times n$ matrix with real or complex entries.

Then there exists a factorization of the form

$M = U \Sigma V^*$ where

U : $m \times m$ unitary matrix, ($U^*U = UU^* = I$)

Σ : is an $m \times n$ matrix with nonnegative numbers on the diagonal and zeros off the diagonal,

V^* : the transpose of V, is an $n \times n$ unitary matrix of real or complex numbers.

Such a factorization is called a **singular-value decomposition** of M.

U describes the rows of M with respect to the base vectors associated with the singular values.

V describes the columns of M with respect to the base vectors associated with the singular values. Σ contains the singular values.

One commonly insists that the values $\Sigma_{i,i}$ be ordered in non-increasing fashion. The, the diagonal matrix Σ is uniquely determined by M (but not U and V).

Single Value Decomposition of EP matrices

For a given EP matrix $\mathbf{P} \in \Re^{n \times p}$, SVD decomposes **P** into 3 matrices

$$\mathbf{P} = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n \times p} \mathbf{V}^T$$

where $\mathbf{U} \in \Re^{n \times n}$: orthonormal matrix of the left singular vectors,

$\mathbf{V} \in \Re^{p \times p}$: an analogous orthonormal matrix of the right singular vectors,

$\mathbf{\Sigma} \in \Re^{r \times r}$:a diagonal matrix containing the singular values $\sigma_{i=1..r}$ arranged in descending order where *r* is the rank of **P**.

The first *r* columns of **U** and **V**, referred to as the left and right singular vectors, or modes, are unique and form the orthonormal basis for the column space and row space of **P**.

The singular values are the square roots of the eigenvalues of $\mathbf{P}^T\mathbf{P}$. The magnitude of the singular values in Σ indicate the relative contribution of the singular vectors in **U** and **V** in reconstructing **P**.

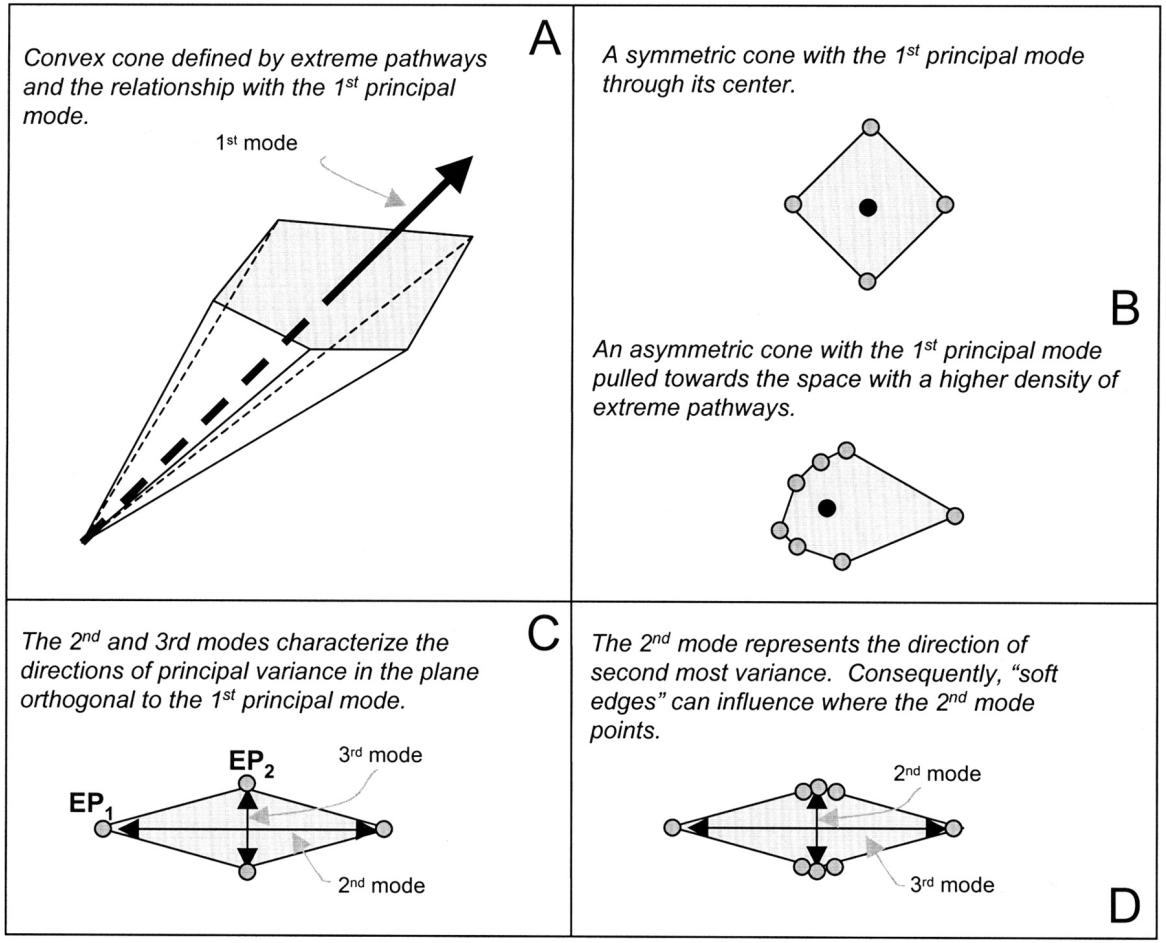
E.g. the second singular value contributes less to the construction of **P** than the first singular value etc.

Price *et al.* Biophys J 84, 794 (2003)

Single Value Decomposition of EP: Interpretation

The first mode (as the other modes) corresponds to a valid biochemical pathway through the network.

The first mode will point into the portions of the cone with highest density of EPs.



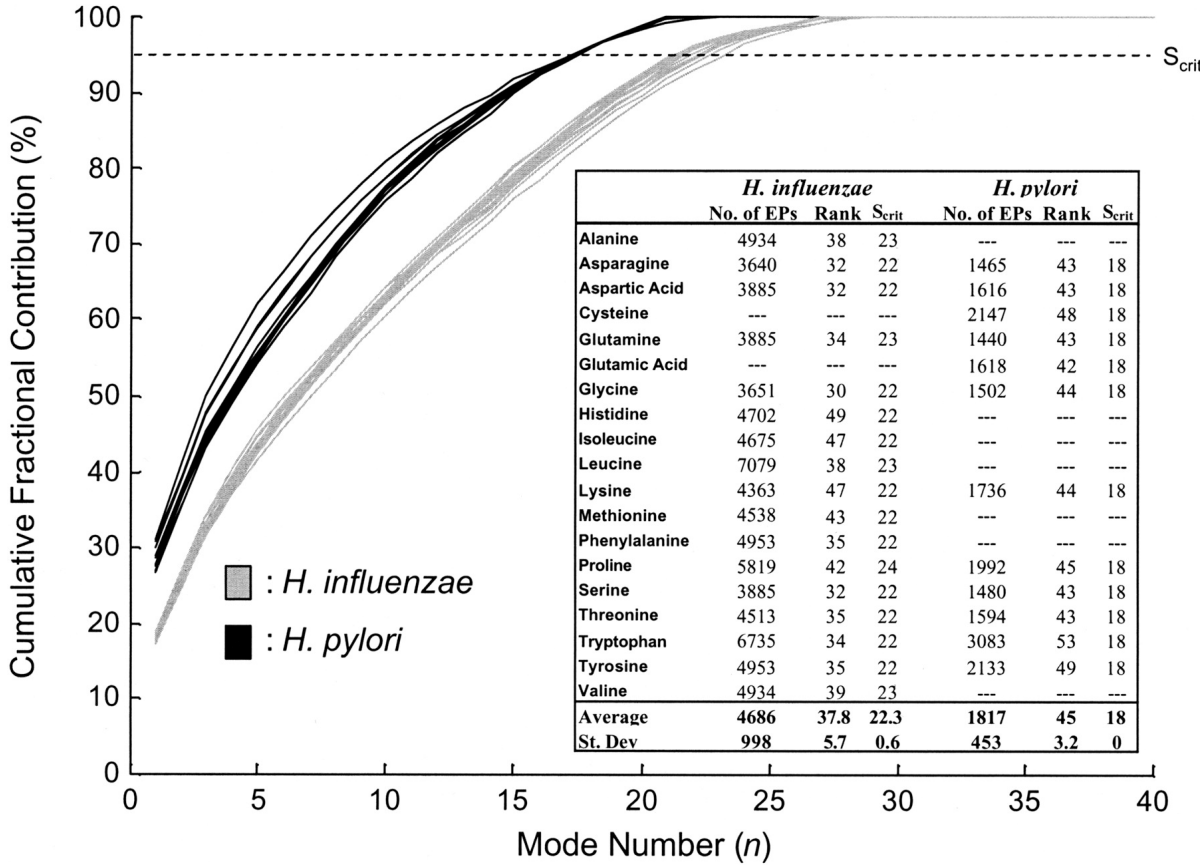
Price *et al.* Biophys J 84, 794 (2003)

SVD applied for *Heliobacter* systems

Cumulative fractional contributions for the SVD of the EP matrices of *H. influenza* and *H. pylori*.

This plot represents the contribution of the first *n* modes to the overall description of the system.

Ca. 20 modes allow describing most of the metabolic activity in the Network.



	H. influenzae			*H. pylori*		
	No. of EPs	Rank	S_{crit}	No. of EPs	Rank	S_{crit}
Alanine	4934	38	23	---	---	---
Asparagine	3640	32	22	1465	43	18
Aspartic Acid	3885	32	22	1616	43	18
Cysteine	---	---	---	2147	48	18
Glutamine	3885	34	23	1440	43	18
Glutamic Acid	---	---	---	1618	42	18
Glycine	3651	30	22	1502	44	18
Histidine	4702	49	22	---	---	---
Isoleucine	4675	47	22	---	---	---
Leucine	7079	38	23	---	---	---
Lysine	4363	47	22	1736	44	18
Methionine	4538	43	22	---	---	---
Phenylalanine	4953	35	22	---	---	---
Proline	5819	42	24	1992	45	18
Serine	3885	32	22	1480	43	18
Threonine	4513	35	22	1594	43	18
Tryptophan	6735	34	22	3083	53	18
Tyrosine	4953	35	22	2133	49	18
Valine	4934	39	23	---	---	---
Average	**4686**	**37.8**	**22.3**	**1817**	**45**	**18**
St. Dev	**998**	**5.7**	**0.6**	**453**	**3.2**	**0**

Cumulative fractional contribution : sum of the first n fractional singular values. This value represents the contribution of the first n modes to the overall description of the system. The rank of the respective extreme pathway matrix is shown for nonessential amino acids. S_{crit}: number of singular values that account for 95% of the variance in the matrices. Entries with "- - -" correspond to essential amino acids.

Price *et al.* Biophys J 84, 794 (2003)

Summary – Extreme Pathways

Extreme Pathway Analysis is a standard technique for analysis of metabolic networks.

Number of EPs can become extremely large – hard to interpret.

EP is an excellent basis for studying systematic effects of reaction cut sets.

SVD could facilitate analysis of EPs. Has not been widely used sofar.

It will be very important to consider the interplay of metabolic and regulatory networks.