# V15 Flux Balance Analysis – Extreme Pathways

#### Stoichiometric matrix S:

 $m \times n$  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.



# **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:

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

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**.

4

0

=

V

S

#### Null space: space of feasible solutions Consider

$$\left(\begin{array}{ccc} 0 & 2 & 1 \\ 3 & -1 & 1 \end{array}\right) \left(\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right)$$

Corresponds to  $2x_2 + x_3 = 0$  <=>  $2x_2 = -x_3$  $3x_1 - x_2 + x_3 = 0$  <=>  $2x_1 = -x_3$ 

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

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



flux 1

#### 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)

15. Lecture WS 2013/14

## **True biological flux**

To find the "true" biological flux in cells ( $\rightarrow$  e.g. Heinzle, UdS) one needs additional (experimental) information,

or one may impose constraints

$$\alpha_i \le v_i \le \beta_i$$

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) 15. Lecture WS 2013/14

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

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

	Central metabolism (EMP, PPP, TCA cycle, electron transport)	aceA, aceB, aceF, aceF, ackA, acnA, acnB, acs, adhE, agp, appB, appC, atpA, atpB, atpC, atpD, atpE, atpF, atpG, atpA, atpH, atpl, cydA, cydB, cycC, cydD, cyoA, cyoB, cyoC, cyoD, dkl, enc, 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, gnct, gpmA, gpmB, hyaA, hyaB, hyaC, hybA, hybC, hycB, hycE, hycF, hycG, icdA, lctD, ldhA, ipdA, malP, mdh, ndh, nuoA, nuoB, nuoE, nuoF, nuoG, nuoH, nuoI, nu
	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, mtID, nagA, nagB, nanA, pfkB, pai, pam, rbsK, rhaA, rhaB, rhaD, srID, treC, avIA, xvIB
	Amino acid metabolism	adi, aklH, alr, ansA, ansB, argA, argB, argC, argD, argE, argF, argG, argH, argI, araA, aroB, aroC, aroD, aroE, aroF, aroG, aroH, atoK, aroI, asd, asnA, asnB, aspA, aspC, avtA, cadA, carA, carB, cysC, cysD, cysE, cysH, cysI, cysI, cysK, cysM, cysII, dadA, dadX, dapA, dapB, dapD, dapE, dapF, dsdA, gabD, gabT, gadA, gadB, gdfA, glk, glnA, gltB, gltD, glyA, goaG, hisA, hisB, hisC, hisD, hisF, hisG, hisH, hisI, ihA, ihA, ihA, ihA, ihA, ihA, ihA, ihA
	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, purI, 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, epcl, folA, folC, folD, folE, folK, folP, gcvH, gcvF, gcvT, gltX, glyA, gcr, gshA, gshB, hemA, hemB, hemC, hemD, hemE, hemF, hemH, hemK, hemL, hemM, hemX, hemY, ihvC, lig, lpdA, menA, menB, menC, menD, menE, menF, menF, metF, mutT, nadA, nadB, nadC, nadE, ntpA, pabA, pabB, pabC, panB, panC, panD, pdxA, pdxB, pdxH, pdxl, pdxK, pncB, parU, 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), thiN (51), thiN (51), ubiE (52), ubiF (52), arabinose-5-phosphate isomerase (22), phosphop-antetheine adentylyltransforase (50), dephosphoCoA kinase (50), NMN qlycohydrolase (49)
	Lipid metabolism	accA, accB, accD, atcbB, cdh, cdsA, cls, dgkA, fabD, fabH, fadB, gpsA, ispA, ispB, pgpB, pgsA, psd, pssA, pgpA (53)
	Cell wall metabolism	ddW, ddlB, galF, galU, glmS, gimU, httB, kdsA, kdsB, kdtA, lpxA, lpxB, lpxC, lpxD, mtaY, msbB, muxA, murB, murC, murD, murE, murF, murG, murl, rfaC, rfaD, rfaF, rfaG, rfal, rfal, ushA, glmM (54), lpoA (55), rfaE (55), tetraacyldisaccharide 4' kinase (55), 3-deoxy-a-manno-octulosonic-acid 8-phosphete phosphatese (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, gaIP, gatA, gatB, gatC, gInH, ginP, ginQ, gIpF, gIpT, gitL, gitK, gitL, gitP, gItS, gntT, gpt, hisI, hisM, hisP, hisQ, hpt, kdpA, kdpB, kdpC, kgtP, lacY, lamB, linF, linG, linH, linL, linK, linM, hisD, araF, maK, maK, maK, maxF, maxF, maxF, maxF, maxB, maxF, maxB, maxF,
n		riar, yar, maie, mair, maia, maia, maia, maia, mani, mani, mani, mani, mgia, mgia, mgia, mitia, miti, nage, nanī, 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, ptsL
2000)		ptsN, ptsP, purB, putP, rbsA, rbsB, rbsC, rbsD, rhaT, sapA, sapB, sapD, sbp, sdaC, srIA_1, srIA_2, srIB, tdcC, tnaB, treA, treB, trkA, trkG, trkH, tsx, tyrP, ugpA, ugpB, ugpC, ugpE, uraA, xapB, xyIE, xyIE, xyIE, xyIH,
013/14		Bioinformatics III

Edwards & Palsso

PNAS 97, 5528 (20

15. Lecture WS 2013/14

#### DIDITIONNALICS III

#### 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<sub>4</sub><sup>2-</sup>, NH<sub>3</sub>, CO<sub>2</sub>, SO<sub>4</sub><sup>2-</sup>, K<sup>+</sup>, Na<sup>+</sup> 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.

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

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) 15. Lecture WS 2013/14

#### **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  $\rightarrow$  P6P reaction.

Edwards & Palsson PNAS 97, 5528 (2000)

15. Lecture WS 2013/14



## Gene deletions in central intermediary metabolism



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) Bioinformatics III

# 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.

# 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) 15. Lecture WS 2013/14 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				-/-
nave	aceEF*	-/+			
	ackA				+/+
	acn	-/-			-/-
•	acs				+/+
	cyd	+/+			
	cyo	+/+			
	eno†	-/+	-/+	-/-	-/-
	fbal	-/+			
	fbp	+/+	-/-	-/-	-/-
	frd	+/+		+/+	+/+
	gap	-/-	-/-	-/-	-/-
	glk	+/+			
	gitA	-/-			-/-
	gnd	+/+			
	idh	-/-			-/-
	mdh <sup>++</sup>	+/+	+/+	+/+	
	ndh	+/+	+/+		
•	nuo –	+/+	+/+		
	pfk <sup>†</sup>	-/+			
	pgi‡	+/+	+/-	+/-	
	pgk	-/-	-/-	-/-	-/-
	pgl	+/+			
	pntAB	+/+	+/+	+/+	
	ppc⁵	$\pm / +$	-/+	+/+	
	pta				+/+
	pts	+/+			
wth	pyk	+/+			
vv li i.	🛑 rpi	-/-	-/-	-/-	-/-
	sdhABCD	+/+		-/-	-/-
	sucAB	+/+		-/+	-/+
	tktAB	-/-			
	tpi**	-/+	-/-	-/-	-/-
	unc	+/+		±/+	-/-
Diginformation !!!	zwf	+/+	+/+	+/+	
DIDITION TRACES III					

#### **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

defined as pointing outwards.											$v_1$		
	Mass	bala	nce co	onstra	ints						<i>v</i> <sub>2</sub>		Internal flux constraints
	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	0	0	0	0	0	-1	0	0	0	03 04		$v_j \ge 0,  j = 1, \dots, 6$
	0	1	-1	-1	1	-1	0	0	0	0	<i>U</i> 5	_ 0	
	0	0	0	1	-1	0	0	0	-1	0	$b_1$	0	Exchange flux constraints
	0	0	0	0	0	1	0	U	0	-1]	b.	[0]	
Schilling, Letscher, Palsson,					(6						h		$-\infty \leq b_j \leq +\infty,  j=1,,4$
J. theor. Biol. 203, 229 (2000) 15. Lecture WS 2013/14					(5	• <b>v</b> = 0	0)				$b_4$		

#### Idea – extreme pathways



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 **S**<sup>T</sup>. I serves for bookkeeping.



#### separate internal and external fluxes

Examine constraints on each of the exchange fluxes as given by

 $\alpha_j \le \mathbf{b}_j \le \beta_j$ 

If the exchange flux is constrained to be positive  $\rightarrow$  do nothing.

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

If the exchange flux is unconstrained  $\rightarrow$  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)

15. Lecture WS 2013/14

# 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  $\mu$ .

The example system contains only one such metabolite, namely C ( $\mu$  = 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  $\rightarrow$  the matrix entries Schilling, Letscher, Palsson, need to be brought to 0.

J. theor. Biol. 203, 229 (2000)

15. Lecture WS 2013/14

# 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  $\mathbf{S}^{T}$ .)





Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000) 15. Lecture WS 2013/14 Bioinformatics III

#### balance combinations of other pathways

(3) Of the remaining rows in  $T^{(i-1)}$  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.



#### remove "non-orthogonal" pathways

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

One method for this works 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 **T**<sup>(i)</sup>

Schilling et al. JTB 203, 229 15. Lecture WS 2013/14

#### repeat steps for all internal metabolites

(5) With the formation of  $T^{(i)}$  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  $\mu$ . 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 15. Lecture WS 2013/14

#### **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:



#### **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  $T^{(E)}$  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

15. Lecture WS 2013/14

#### 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  $\mathbf{P}$  containing the extreme pathways in place of the original identity matrix.

Schilling et al. JTB 203, 229 15. Lecture WS 2013/14

#### pathway matrix



**p**<sub>1</sub>

**p**<sub>7</sub>

 $\mathbf{p}_3$ 

**p**<sub>2</sub>

**p**<sub>4</sub>

p<sub>6</sub>

**p**<sub>5</sub>

 $v_1$   $v_2$   $v_3$   $v_4$   $v_5$   $v_6$   $b_1$   $b_2$   $b_3$   $b_4$ 

**P**<sup>⊤</sup> =

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

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  $v_2/v_3$ .

$$v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad b_1 \quad b_2 \quad b_3 \quad b_4$$

1	0	0	0	0	0	-1	1	0	0
0	1	1	0	0	0	0	0	0	0
0	1	0	1	0	0	0	-1	1	0
0	1	0	0	0	1	0	-1	0	1
0	0	1	0	1	0	0	1	-1	0
0	0	0	1	1	0	0	0	0	0
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  $\mathbf{P}$  to a matrix with entries 0 or 1,

the symmetric Pathway Length Matrix  $\mathbf{P}_{\rm LM}$  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.



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 **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.

fficient of variation
15%
14%
10%
18%
17%
10%
12%
16%
12%
14%
6%
7%
2%
2%
fficient of variation
10%
13%
14%
11%
7%
3.96
3/0
9%
9% 10%
9% 10% 9%
9% 10% 9% 8%
9% 10% 9% 8% 7%
9% 10% 9% 8% 7% 11%
9% 10% 9% 8% 7% 11% 10%
9% 10% 9% 8% 7% 11% 10% 10%
9% 10% 9% 7% 11% 10% 6%
9% 10% 9% 7% 11% 10% 6% 7%

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

<i>H. pylori</i> 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		
11 · · ·	1.0			

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 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)

15. Lecture WS 2013/14

# **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)

 $\Sigma$ : 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.  $\Sigma$  contains the singular values.

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

15. Lecture WS 2013/14

### **Single Value Decomposition of EP matrices**

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

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

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

- $\mathbf{V} \in \mathfrak{R}^{p \times p}$  : an analogous orthonormal matrix of the right singular vectors,
- $\Sigma \in \Re^{r \times r}$  :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  $\mathbf{P}^{\mathsf{T}}\mathbf{P}$ .

The magnitudes of the singular values in  $\Sigma$  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.

Price *et al.* Biophys J 84, 794 (2003) 15. Lecture WS 2013/14



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.

#### **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.