Applications of structural analysis of metabolic networks.

David Fell

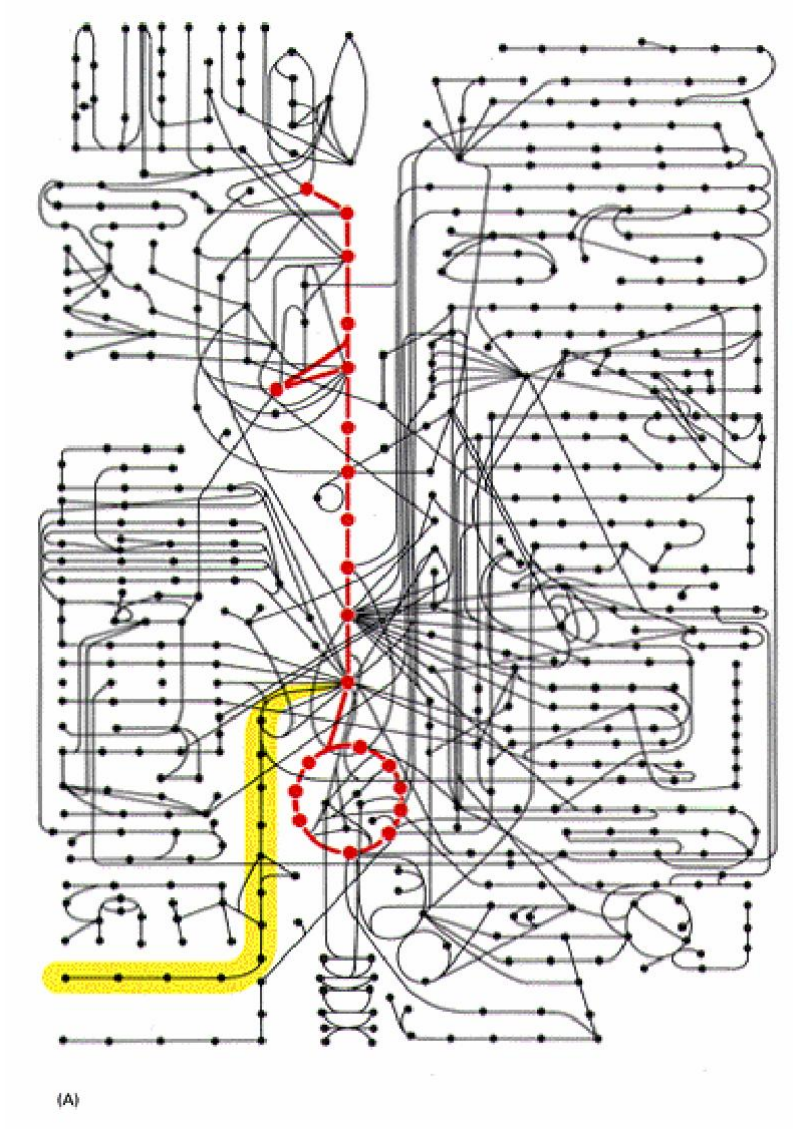
dfell@brookes.ac.uk

OXFORD BROOKES UNIVERSITY

Outline

- **Theoretical approaches to the analysis of metabolic networks**
- Applications of elementary modes analysis:
  - Relating physiology to network structure
  - Identifying productive routes
  - Determining usage of modes
- Substructure in metabolic networks

The metabolic network



Pentose phosphate reactions

Gene	Reaction	*E. coli*
Pgi	G6P = F6P	EC4025
Pfk	F6P + ATP – ADP + FP2	EC3916
Fbp	FP2 - F6P + Pi	EC4232
Fba	FP2 = GAP + DHAP	EC2925
TpiA	GAP = DHAP	EC3919
Gap	GAP + Pi + NAD = NADH + BPG	EC1779
Pgk	BPG + ADP = ATP + P3G	EC2926
Gpm	P3G = P2G	EC0755
Eno	P2G = PEP	EC2779
Pyk	PEP + ADP – ATP + Pyr	EC1854
Zwf	G6P + NADP = GO6P + NADPH	EC1852
Pgl	GO6P – P6G	EC0766
Gnd	P6G + NADP = NADPH + CO2 + Ru5P	EC2029
Rpi	Ru5P = R5P	EC2914
Rpe	Ru5P = Xyl5P	EC3386
TktI	R5P + Xyl5P = GAP + Sed7P	EC2935
		EC2465
Tal	GAP + Sed7P = Ery4P + F6P	EC2464
TktII	Xyl5P + Ery4P = F6P + GAP	EC2935
		EC2465
Prs	R5P – R5Pex	EC4383

What pathways make up metabolism?

- Glycolysis

What pathways make up metabolism?

- Glycolysis
- Pentose phosphate pathway

What pathways make up metabolism?

- Glycolysis
- Pentose phosphate pathway
- Electron transport & oxidative phosphorylation

What pathways make up metabolism?

- Glycolysis
- Pentose phosphate pathway
- Electron transport & oxidative phosphorylation
- Fatty acid oxidation

What pathways make up metabolism?

- Glycolysis
- Pentose phosphate pathway
- Electron transport & oxidative phosphorylation
- Fatty acid oxidation
- Photosynthesis

What pathways make up metabolism?

- Glycolysis
- Pentose phosphate pathway
- Electron transport & oxidative phosphorylation
- Fatty acid oxidation
- Photosynthesis
- etc

Metabolic networks

If we can prepare a list of the reactions occurring in the metabolism of an organism, can we decide:

- what nutrients it can utilize and what products it can produce?

Metabolic networks

If we can prepare a list of the reactions occurring in the metabolism of an organism, can we decide:

- what nutrients it can utilize and what products it can produce?
- if there is a route from a particular nutrient to a product?

Metabolic networks

If we can prepare a list of the reactions occurring in the metabolism of an organism, can we decide:

- what nutrients it can utilize and what products it can produce?
- if there is a route from a particular nutrient to a product?
- which route to a product has the highest yield?

Metabolic networks

If we can prepare a list of the reactions occurring in the metabolism of an organism, can we decide:

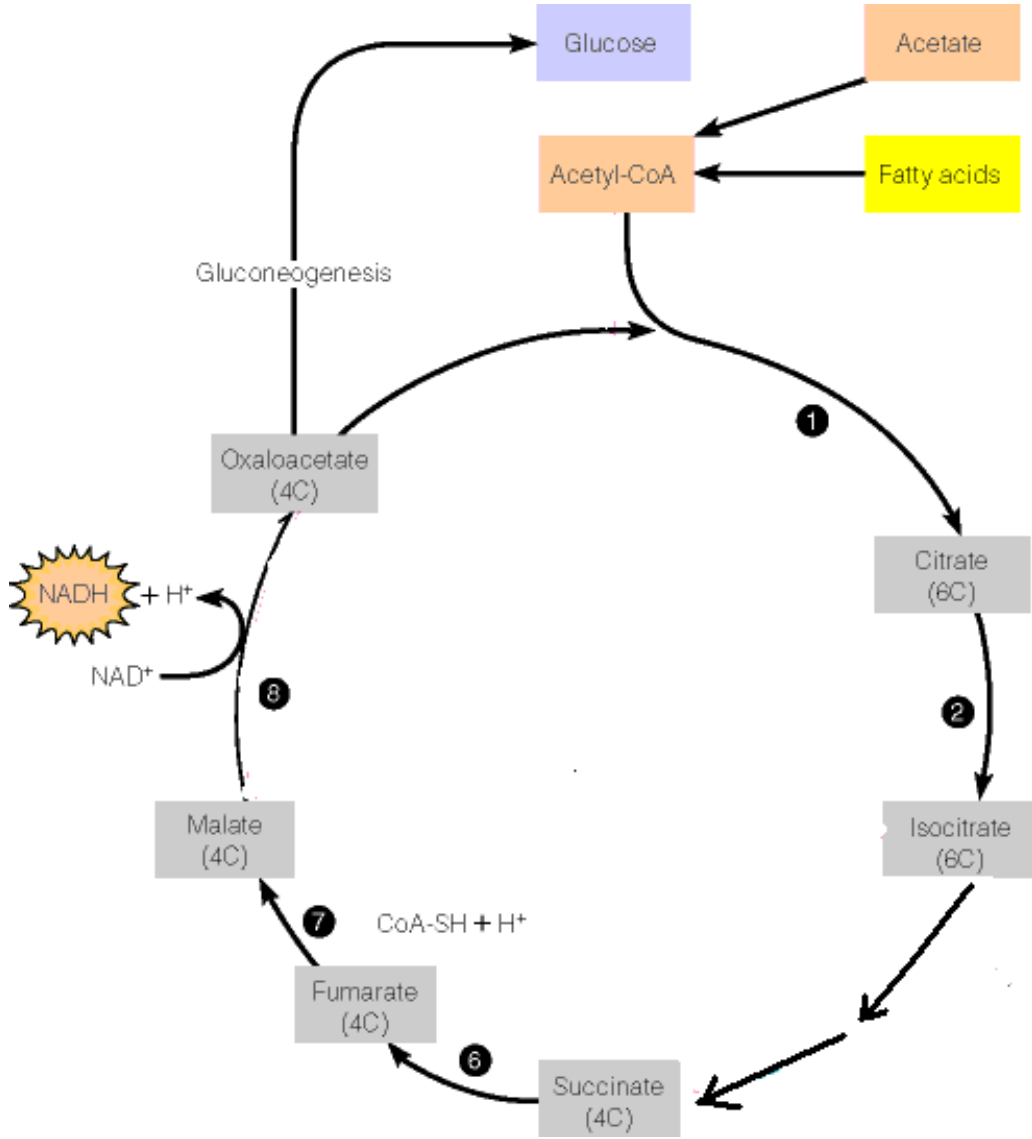
- what nutrients it can utilize and what products it can produce?
- if there is a route from a particular nutrient to a product?
- which route to a product has the highest yield?
- what are the consequences of deleting an enzyme?

Metabolic networks

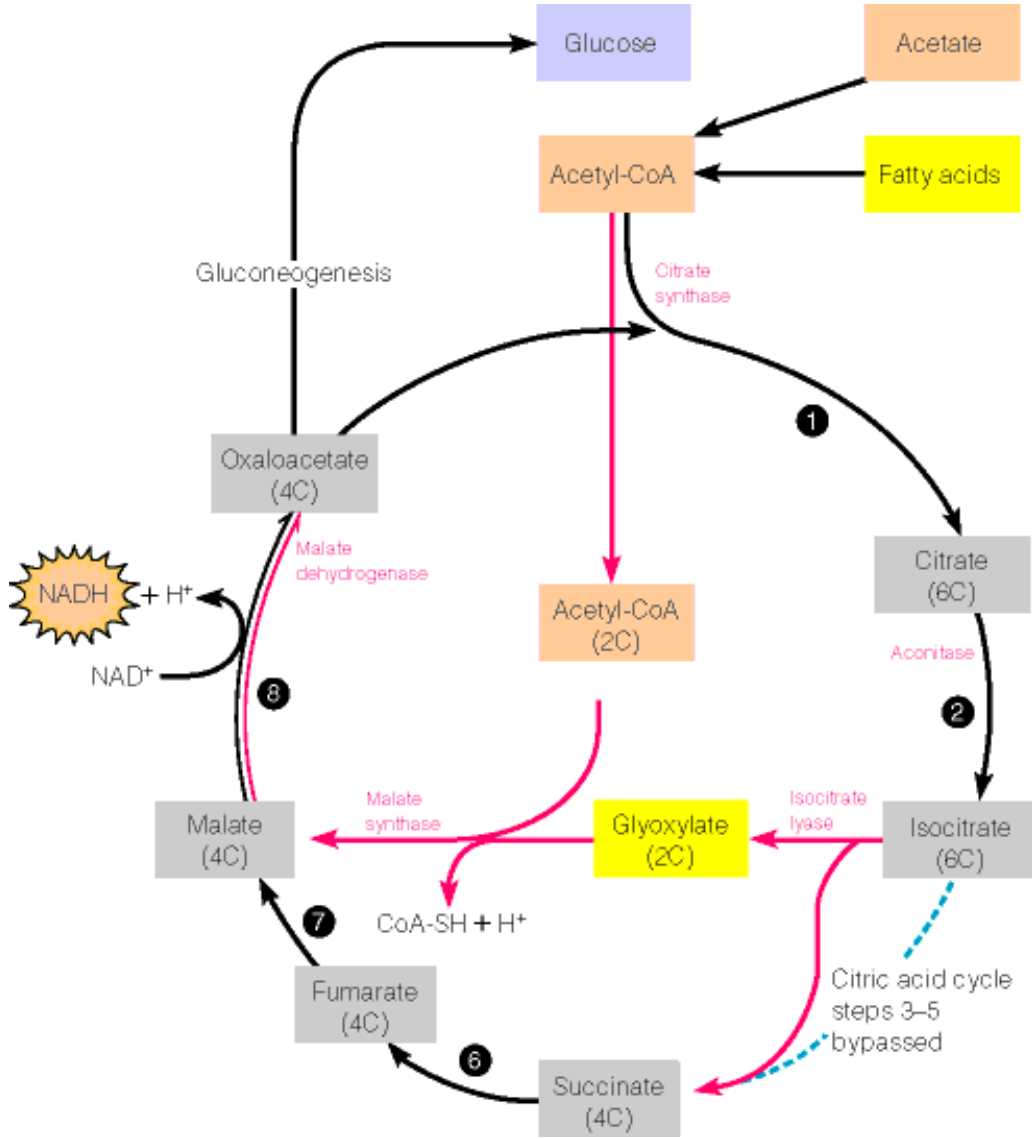
If we can prepare a list of the reactions occurring in the metabolism of an organism, can we decide:

- what nutrients it can utilize and what products it can produce?
- if there is a route from a particular nutrient to a product?
- which route to a product has the highest yield?
- what are the consequences of deleting an enzyme?
- whether genome annotations for an organism generate a connected and self-consistent metabolism?

Do graph algorithms have the answer?

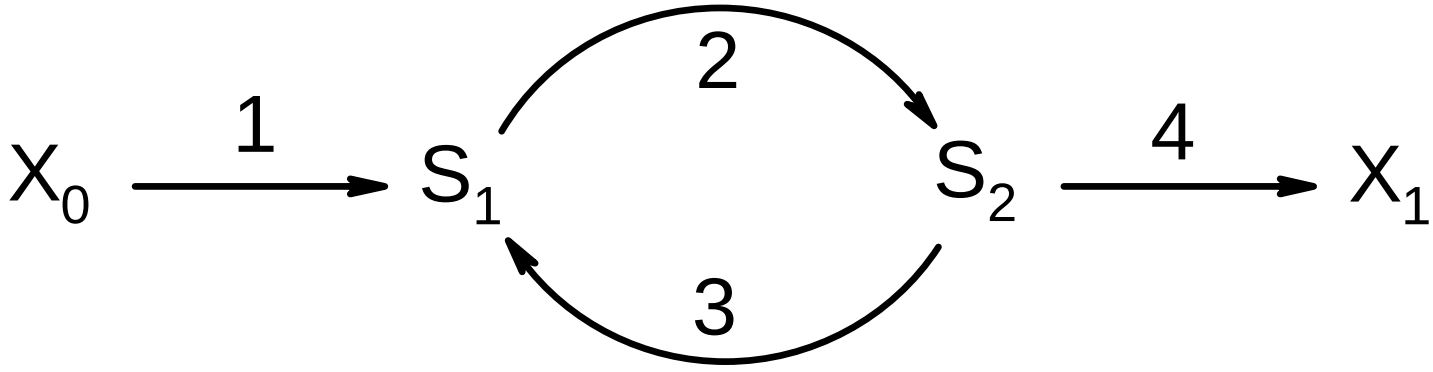


Do graph algorithms have the answer?



Representation

Consider a simple pathway, e.g.:



$$\begin{matrix} S_1 \\ S_2 \end{matrix} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

r1: $X_0 \rightarrow S_1 \sim$
r2: $S_1 \rightarrow S_2 \sim$
r3: $S_2 \rightarrow S_1 \sim$
r4: $S_2 \rightarrow X_1 \sim$

Separating structure and kinetics

The rate at which the substrate concentrations are changing is given by $\mathbf{N}.\mathbf{v}$, where $\mathbf{N}$ is the stoichiometry matrix, and $\mathbf{v}$ are the enzyme kinetic functions. So for our substrate cycle pathway:

$$\begin{bmatrix} \frac{dS_1}{dt} \\ \frac{dS_2}{dt} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} . \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

where each v_i is the rate function for enzyme i, depending on the metabolites, V_{m}, K_{m} etc.

Steady state solutions

Any metabolic pathway at steady state satisfies the relationship $\mathbf{N}.\mathbf{v} = \mathbf{0}$, where $\mathbf{N}$ is the stoichiometry matrix, exemplified by the substrate cycle pathway:

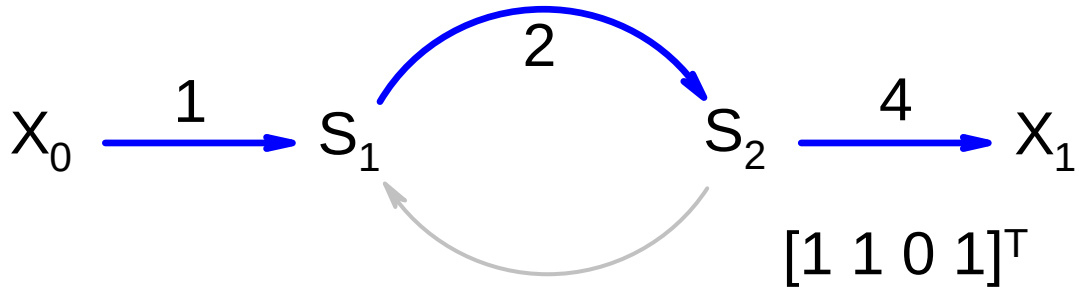
$$\begin{array}{c} S_1 \\ S_2 \end{array} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} . \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

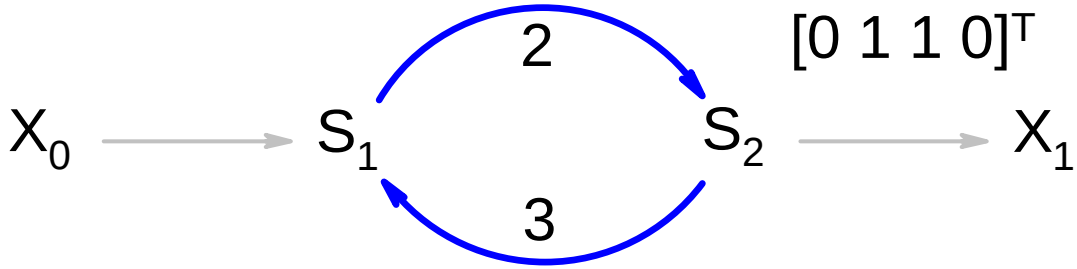
Steady state solutions – 2

Any observed set of velocities at steady state will be a linear combination of a set of vectors $\mathbf{K}$ referred to as the null space of the stoichiometry matrix. In this case,

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Null space vectors as pathways





$[1 \;\; 1 \;\; 0 \;\; 1]^T$ and $[0 \;\; 1 \;\; 1 \;\; 0]^T$

Steady state solutions – 3

Any feasible set of velocities at steady state is a linear combination of these null space vectors, e.g.:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a \\ a+b \\ b \\ a \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Problems with the null space

Shortcomings of the set of basis vectors as metabolic routes:

- Is not a unique solution
- May not respect thermodynamic direction
- Not necessarily ’simple’
- Can mislead about the impact of enzyme deletion

The linear programming approach

The substrate cycle pathway can also be expressed in terms of the full stoichiometry matrix $\mathbf{N}$:

$$\begin{array}{c} X_0 \\ S_1 \\ S_2 \\ X_1 \end{array} \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} -x \\ 0 \\ 0 \\ \geq 0 \end{bmatrix}$$

Linear programming gives a single route satisfying a particular set of constraints and an optimization function. It does not tell of other similar routes.

Elementary modes

An elementary mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions working in the thermodynamically favoured direction, and enzymes weighted by the *relative* flux they carry.

Elementary modes

An elementary mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions working in the thermodynamically favoured direction, and enzymes weighted by the *relative* flux they carry.

'Steady state' implies that there is only net production or consumption of external metabolites. Production and consumption of all internal metabolites is balanced.

Elementary modes

An elementary mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions working in the thermodynamically favoured direction, and enzymes weighted by the *relative* flux they carry.

'Steady state' implies that there is only net production or consumption of external metabolites. Production and consumption of all internal metabolites is balanced.

'Minimal' means that deleting any enzyme in the set would prevent a steady state. By definition, an elementary mode is not decomposable into component elementary modes.

Elementary modes

An elementary mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions working in the thermodynamically favoured direction, and enzymes weighted by the *relative* flux they carry.

'Steady state' implies that there is only net production or consumption of external metabolites. Production and consumption of all internal metabolites is balanced.

'Minimal' means that deleting any enzyme in the set would prevent a steady state. By definition, an elementary mode is not decomposable into component elementary modes.

Hence the set of elementary modes of a reaction network is unique.

Outline

- Theoretical approaches to the analysis of metabolic networks

- **Applications of elementary modes analysis:**
  - **Relating physiology to network structure**
  - Identifying productive routes
  - Determining usage of modes

- Substructure in metabolic networks

Crassulacean acid metabolism

This is a specialized form of photosynthesis adopted by some plants that live in arid areas, and that collect CO_2 by night and close their stomata by day to save water.

- Considering phase 3 of CAM metabolism - daylight metabolism of stored malate with no net CO_2 uptake.
- CAM plants can be divided into 2 groups according to whether malic enzyme or PEP carboxykinase is the main enzyme of malate utilization.
- Christopher & Holtum (1996) proposed a further division into producers of chloroplastic starch and producers of extrachloroplastic sucrose or hexose, giving 4 categories.

Elementary modes analysis of CAM

- There are 6 elementary modes, not 4; 3 for malic enzyme, 3 for PEPCK plants.
- Production of hexose and starch together is a distinct route from production of either starch or hexose, differing in transport between chloroplast and cytosol.
- *Aloe vera* is an example, given by the original authors, of a PEPCK plant forming both starch and hexose.
- The missing class: a malic enzyme plant forming both starch and hexose - *Mesembryanthemum crystallinum*. (Thanks to Dr Annie Borland, Newcastle, for finding this for me.)

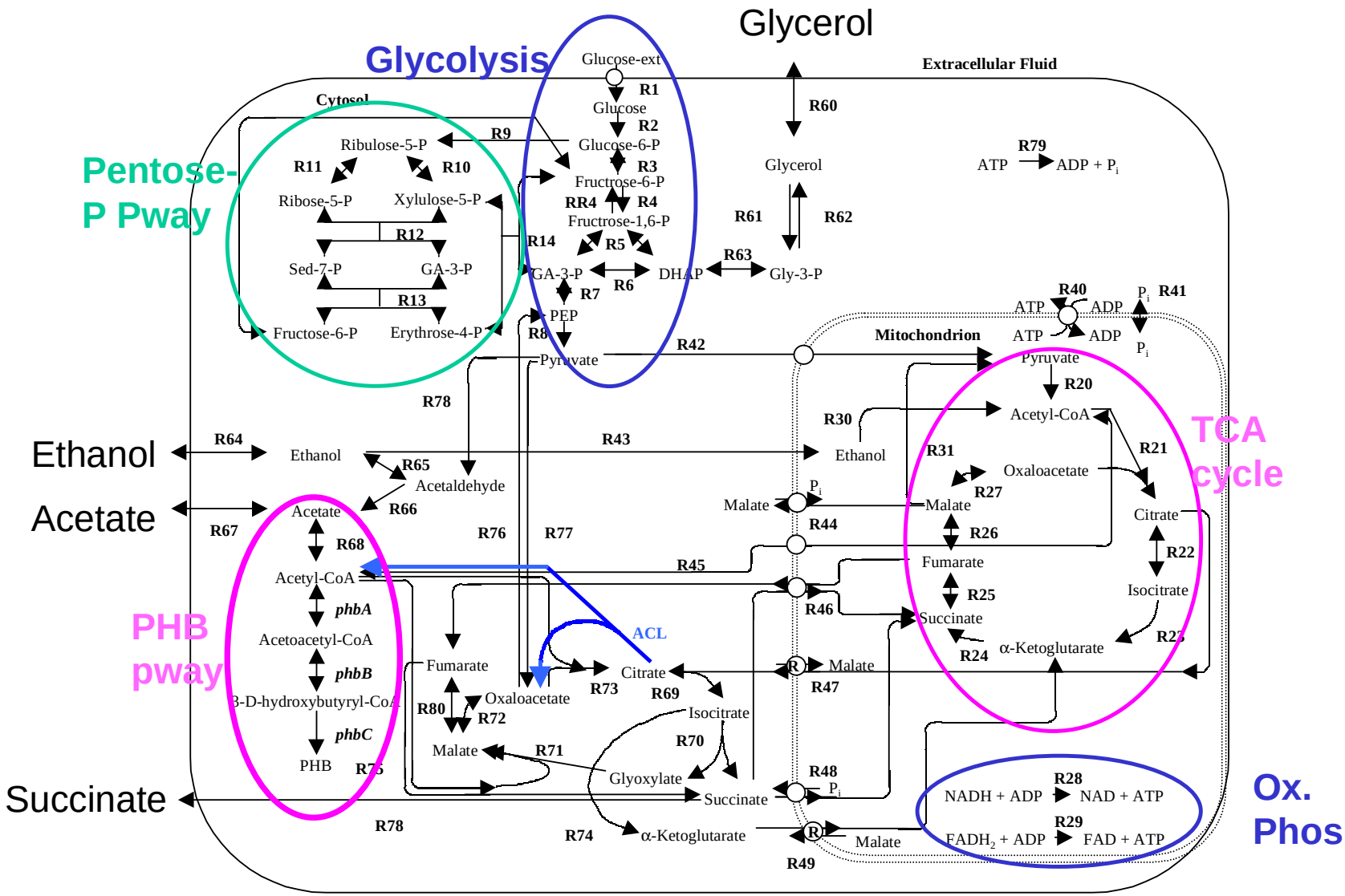
Outline

- Theoretical approaches to the analysis of metabolic networks

- **Applications of elementary modes analysis:**
  - Relating physiology to network structure
  - **Identifying productive routes**
  - Determining usage of modes

- Substructure in metabolic networks

Polyhydroxybutyrate synthesis in yeast



Optimal yields of PHB synthesis

Wild-type yeast + PHB pathway

1. 2 Acetate + EtOH $\rightarrow$ PHB + 2 CO_2 0.67
2. 65 Ac. + 31 EtOH $\rightarrow$ 30 PHB + 72 CO_2 0.63

(Number following each mode is the fractional carbon conversion.)

Optimal yields of PHB synthesis

Wild-type yeast + PHB pathway

1. 2 Acetate + EtOH $\rightarrow$ PHB + 2 CO_2 0.67
2. 65 Ac. + 31 EtOH $\rightarrow$ 30 PHB + 72 CO_2 0.63

Wild-type yeast + ATP–citrate lyase + PHB pathway

3. 12 EtOH $\rightarrow$ 5 PHB + 4 CO_2 0.83
4. 77 EtOH + 31 Glycerol $\rightarrow$
   48 PHB + 4 Ac. + 47 CO_2 0.78

(Number following each mode is the fractional carbon conversion.)

Outline

- Theoretical approaches to the analysis of metabolic networks

- **Applications of elementary modes analysis:**
  - Relating physiology to network structure
  - Identifying productive routes
  - **Determining usage of modes**

- Substructure in metabolic networks

Relative flux in elementary modes

The steady state requirement is:

$$\mathbf{N}.\mathbf{v} = \mathbf{0}$$

But each elementary mode vector $\mathbf{e_i}$ in a matrix of elementary mode vectors $\mathbf{E}$ is also a steady state solution, i.e.:

$$\mathbf{N}.\mathbf{E} = \mathbf{0}$$

Relative flux in elementary modes

The steady state requirement is:

$$\mathbf{N}.\mathbf{v} = \mathbf{0}$$

But each elementary mode vector $\mathbf{e_i}$ in a matrix of elementary mode vectors $\mathbf{E}$ is also a steady state solution, i.e.:

$$\mathbf{N}.\mathbf{E} = \mathbf{0}$$

So the actual velocities, $\mathbf{v}$ are a weighted combination of the elementary modes:

$$\mathbf{v} = \mathbf{E}.\mathbf{w}$$

where $\mathbf{w}$ is a vector of weighting factors.

Solutions for the weighting vector $\mathbf{w}$

Although there is not a unique solution for $\mathbf{w}$ because $\mathbf{E}$ is generally non–invertible and the system under–determined:

$$\hat{\mathbf{w}} = \mathbf{E}^{\#}\mathbf{v}$$

where $\mathbf{E}^{\#}$ is the generalised Penrose inverse of $\mathbf{E}$. Even if we have not measured all the fluxes in $\mathbf{v}$, we can partition observed $\mathbf{v}_o$ and non–observed $\mathbf{v}_x$ fluxes:

$$\mathbf{v} = \left[\mathbf{v}_o, \mathbf{v}_x\right]^T, \text{ and } \mathbf{E} = \begin{bmatrix} \mathbf{E}_o \\ \mathbf{E}_x \end{bmatrix}.$$

Then:

$$\hat{\mathbf{w}} = \mathbf{E}_o^{\#}\mathbf{v}_o$$

Properties of the solution for $\hat{\mathbf{w}}$

Why select $\hat{\mathbf{w}}$ from the many feasible values of $\mathbf{w}$?

- It is the minimum norm solution. i.e. it minimizes

$$\sqrt{\sum \hat{w}_i^2}$$

Properties of the solution for $\hat{\mathbf{w}}$

Why select $\hat{\mathbf{w}}$ from the many feasible values of $\mathbf{w}$?

- It is the minimum norm solution. i.e. it minimizes

$$\sqrt{\sum \hat{w}_i^2}$$

- This ensures that zero flux is assigned to cycles with no overall stoichiometry.

Properties of the solution for $\hat{\mathbf{w}}$

Why select $\hat{\mathbf{w}}$ from the many feasible values of $\mathbf{w}$?

- It is the minimum norm solution. i.e. it minimizes

$$\sqrt{\sum \hat{w}_i^2}$$

- This ensures that zero flux is assigned to cycles with no overall stoichiometry.
- It assigns flux as evenly as possible over available modes (the ‘democratic’ option).

Properties of the solution for $\hat{\mathbf{w}}$

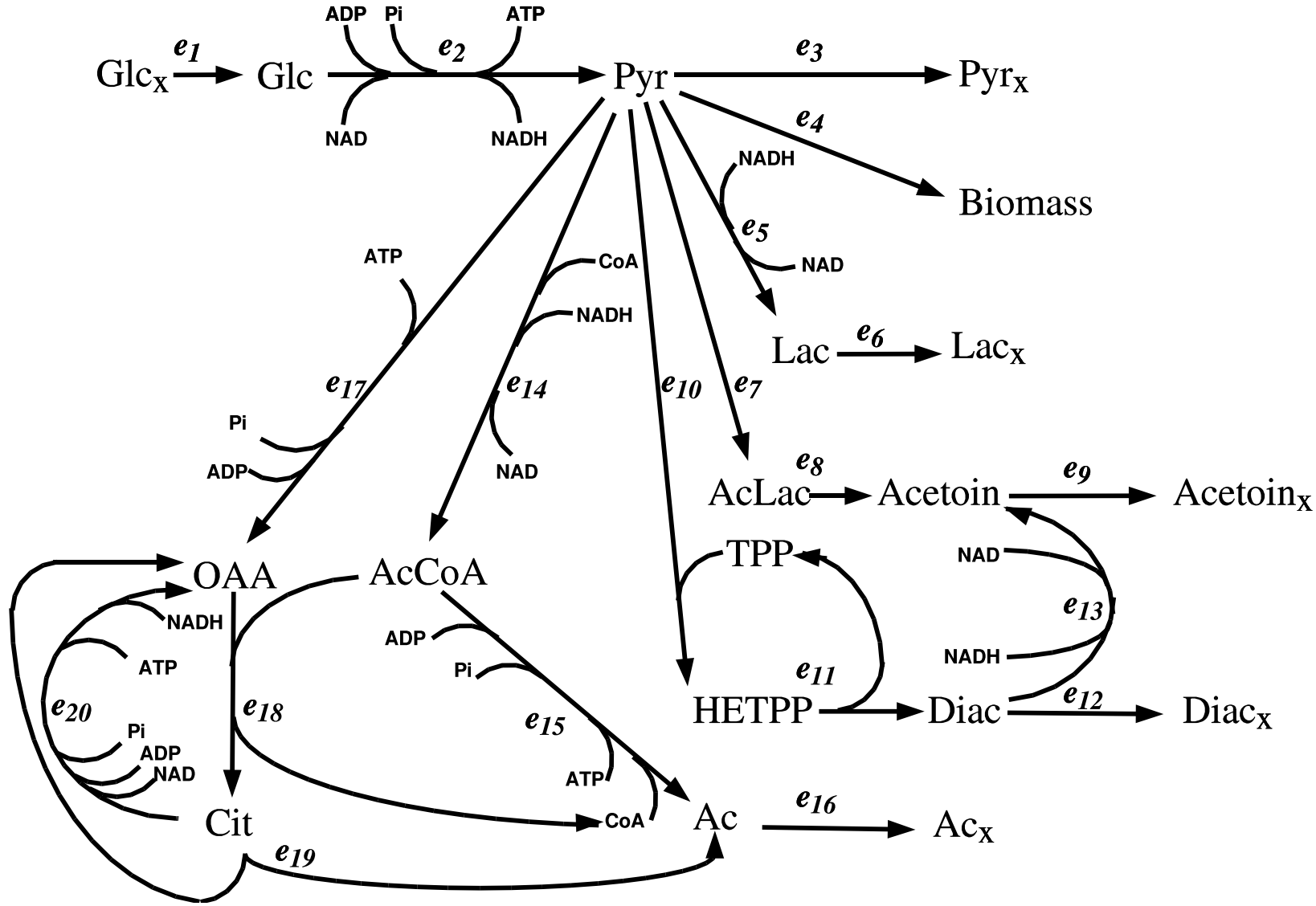
Why select $\hat{\mathbf{w}}$ from the many feasible values of $\mathbf{w}$?

- It is the minimum norm solution. i.e. it minimizes

$$\sqrt{\sum \hat{w}_i^2}$$

- This ensures that zero flux is assigned to cycles with no overall stoichiometry.
- It assigns flux as evenly as possible over available modes (the ‘democratic’ option).
- When $\mathbf{v}_o$ changes smoothly, so does $\hat{\mathbf{w}}$. (Not the case for other extreme solutions, e.g. the ‘autocratic’ option.)

Flux analysis of lactic acid metabolism



Elementary modes analysis

- This reduced scheme has 191 modes.

Elementary modes analysis

- This reduced scheme has 191 modes.
- Only 83 modes consumed nutrients and metabolites observed to be used in the experiments.

Elementary modes analysis

- This reduced scheme has 191 modes.
- Only 83 modes consumed nutrients and metabolites observed to be used in the experiments.
- Flux was assigned to these 83 at different time points through a 48 h fermentation.

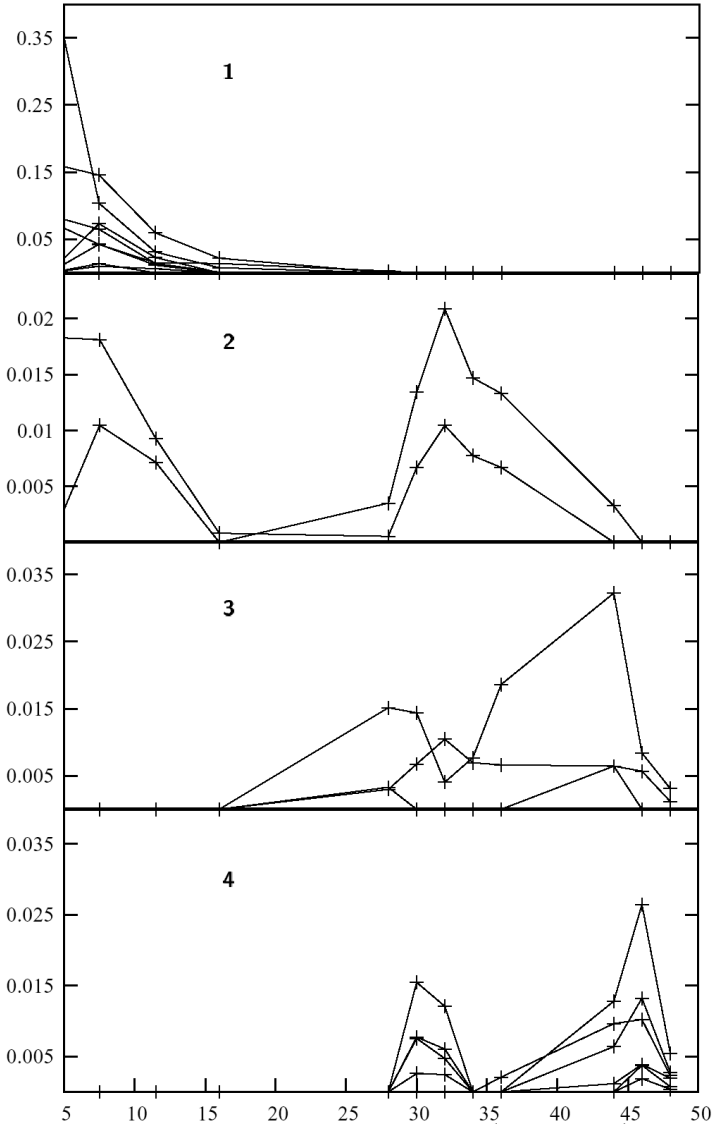
Elementary modes analysis

- This reduced scheme has 191 modes.
- Only 83 modes consumed nutrients and metabolites observed to be used in the experiments.
- Flux was assigned to these 83 at different time points through a 48 h fermentation.
- Modes with equivalent overall stoichiometries were grouped and their assigned flux summed to a single figure, giving 23 assigned fluxes.

Elementary modes analysis

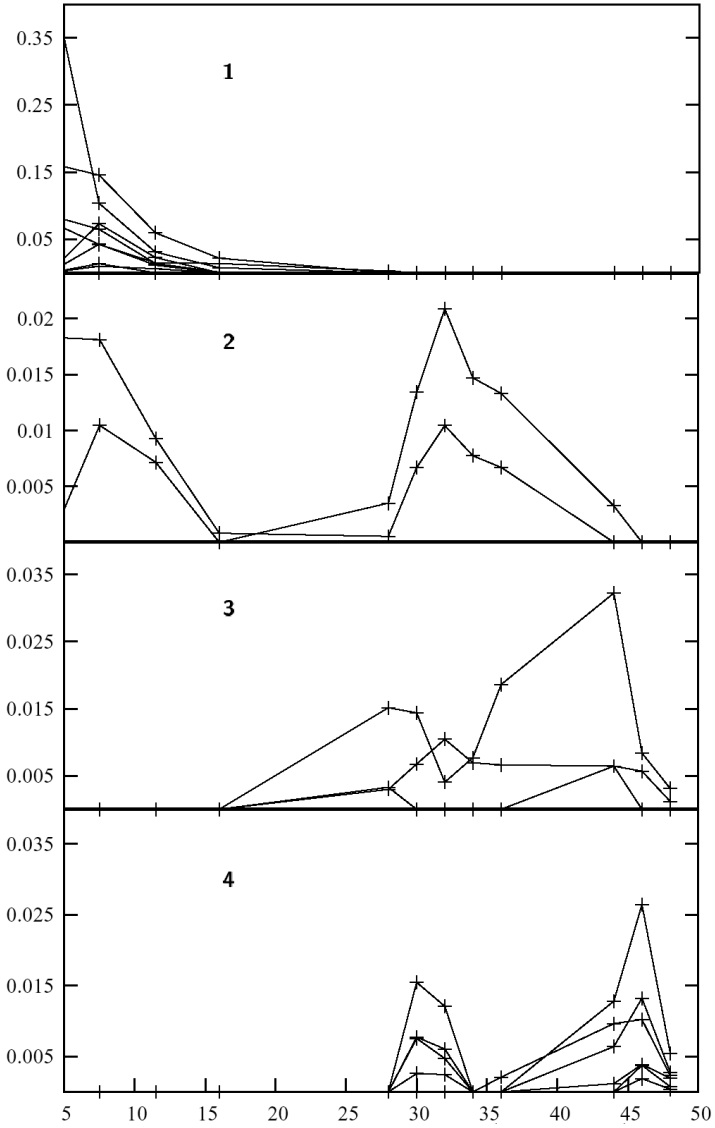
- This reduced scheme has 191 modes.
- Only 83 modes consumed nutrients and metabolites observed to be used in the experiments.
- Flux was assigned to these 83 at different time points through a 48 h fermentation.
- Modes with equivalent overall stoichiometries were grouped and their assigned flux summed to a single figure, giving 23 assigned fluxes.
- The time–dependence of these fluxes defined 4 groups.

Mode assignment



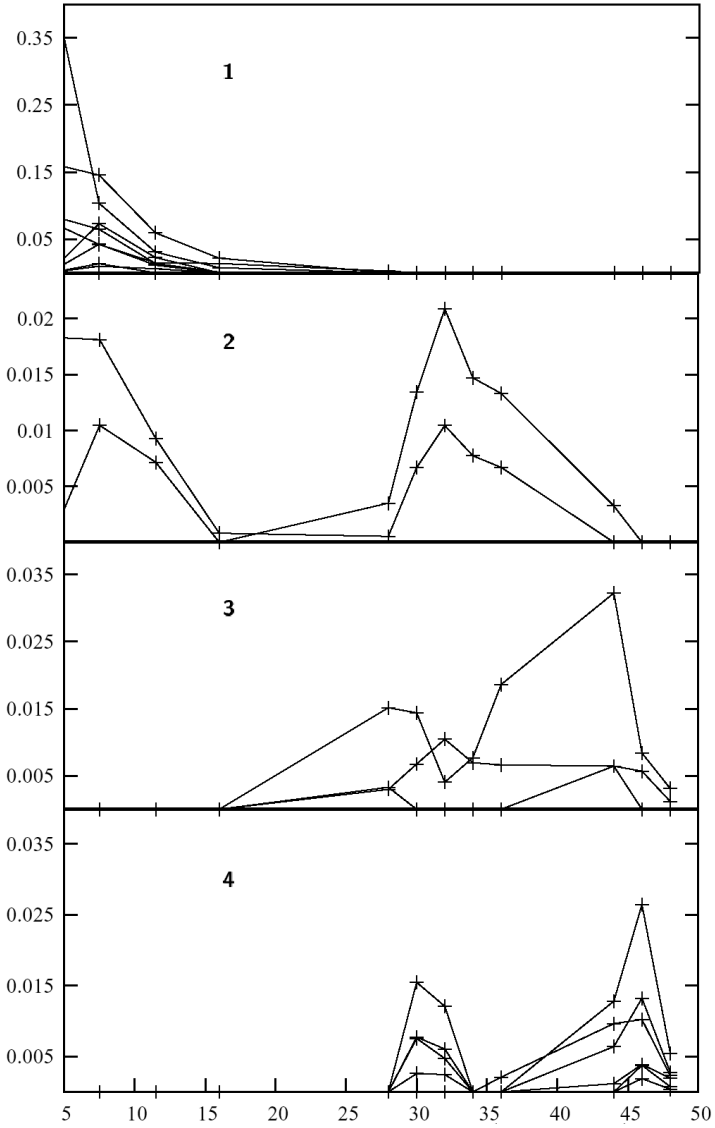
Stoichiometry	Peak flux mol.hr^{-1}.(Mol. Org.)$^{-1}$
Group 1	
Glc $\longrightarrow$ 2 Lac	0.45
1/2 O_2 + Glc $\longrightarrow$ 2 Biomass	0.16
1/2 O_2 + Glc $\longrightarrow$ 2 CO_2 + Acetoin	0.086
O_2 + Glc $\longrightarrow$ 2 Ac + 2 CO_2	0.077
1/2 O_2 + Glc $\longrightarrow$ 2 Pyr	0.073
Pyr $\longrightarrow$ 5/6 Lac + 1/2 CO_2	0.042
3/4 O_2 + Glc $\longrightarrow$ Diac + 2 CO_2	0.014
3/4 O_2 + Glc $\longrightarrow$ Cit	0.01
3 O_2 + Glc $\longrightarrow$ 6 CO_2	0.01
Glc + 6 Pyr $\longrightarrow$ 4 Lac + 2 Cit	0.0003

Mode assignment



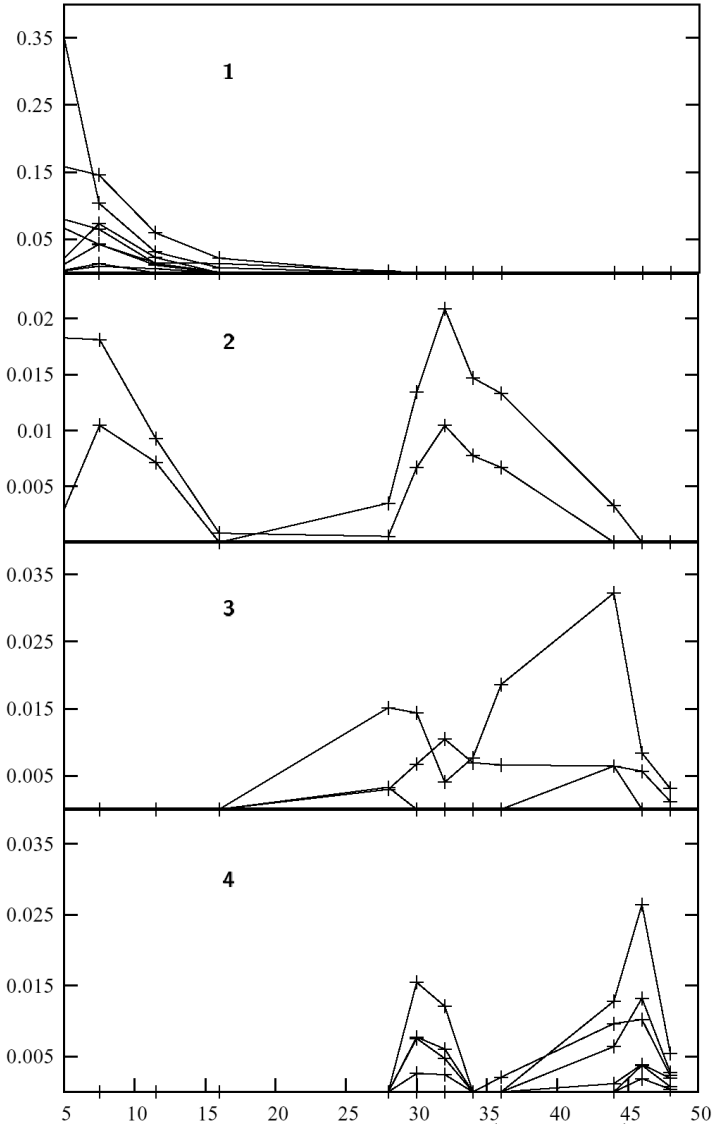
Stoichiometry	Peak flux $mol.hr^{-1}.(Mol.\ Org.)^{-1}$
Group 2	
Lac 1/4 + O_2 $\longrightarrow$ Biomass	0.021
Pyr $\longrightarrow$ Biomass	0.018

Mode assignment



Stoichiometry	Peak flux $mol.hr^{-1}.(Mol.\ Org.)^{-1}$
Group 3	
Lac 1/2 + O_2 $\longrightarrow$ Ac + CO_2	0.032
Lac 1/4 O_2 $\longrightarrow$ Pyr	0.01
1/4 O_2 + Pyr $\longrightarrow$ Ac + CO_2	0.006

Mode assignment



Stoichiometry			Peak flux $mol.hr^{-1}.(Mol.\ Org.)^{-1}$
Group 4			
Lac + 3/2 O_2	$\longrightarrow$	3 CO_2	0.026
Lac + 3/4 O_2	$\longrightarrow$	1/3 Cit + CO_2	0.013
Lac + 5/12 O_2	$\longrightarrow$	1/3 Cit + 1/3 Ac + 1/3 CO_2	0.010
Lac + 1/4 O_2	$\longrightarrow$	$CO_2$1/2 Acetoin	0.0039
5/4 O_2 + Pyr	$\longrightarrow$	3 CO_2	0.0038
1/2 O_2 + Pyr	$\longrightarrow$	1/3 Cit + CO_2	0.0019

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;
- Analysis of metabolic capabilities;

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;
- Analysis of metabolic capabilities;
- A tool for aiding metabolic engineering;

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;
- Analysis of metabolic capabilities;
- A tool for aiding metabolic engineering;
- A functional interpretation of flux patterns;

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;
- Analysis of metabolic capabilities;
- A tool for aiding metabolic engineering;
- A functional interpretation of flux patterns;
- Identification of metabolic subsystems;

Conclusion

Structural analysis can offer:

- Valid routes through metabolic networks;
- Analysis of metabolic capabilities;
- A tool for aiding metabolic engineering;
- A functional interpretation of flux patterns;
- Identification of metabolic subsystems;
- Prediction of possible co–regulation/co–expression patterns..

Acknowledgements

- **Berlin:** Stefan Schuster, Claus Hilgetag, Thomas Pfieffer & Ferdinand Moldenhauer
- **Oxford:** Mark Poolman, John Woods, Bhushan Bonde
- **Heidelberg/Freiburg:** Thomas Dandekar
- **Minnesota:** Friedrich Srienc, Ross Carlson
- **Mumbai:** K V Venkatesh

Further details

References:

- S Schuster, T Dandekar & D A Fell, Trends in Biotechnology, 17, 53-60 (1999).
- S Schuster, D A Fell & T Dandekar, Nature Biotechnol. 18, 326-332 (2000)

Algorithm:

- http: //mudshark.brookes.ac.uk/algorithm.pdf

Programs:

- ScrumPy: http://mudshark.brookes.ac.uk/ScrumPy/
- Metatool: http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/