Elementary Mode Analysis
A review

May 2006, MIT

Orhan Karsligil

Major Approaches to Metabolism Modeling

(a) Interaction-based

C
A ———— B

Static models

No stoichiometry

No parameters

(b) Constraint-based

$A + B \rightleftharpoons C$

Static models

Stoichiometry

No parameters

Flux distribution

Flux 2
Flux 3
Flux 1

(c) Mechanism-based

$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C$

Dynamic models

Stoichiometry

Kinetic parameters

Concentration

Time

Current Opinion in Microbiology
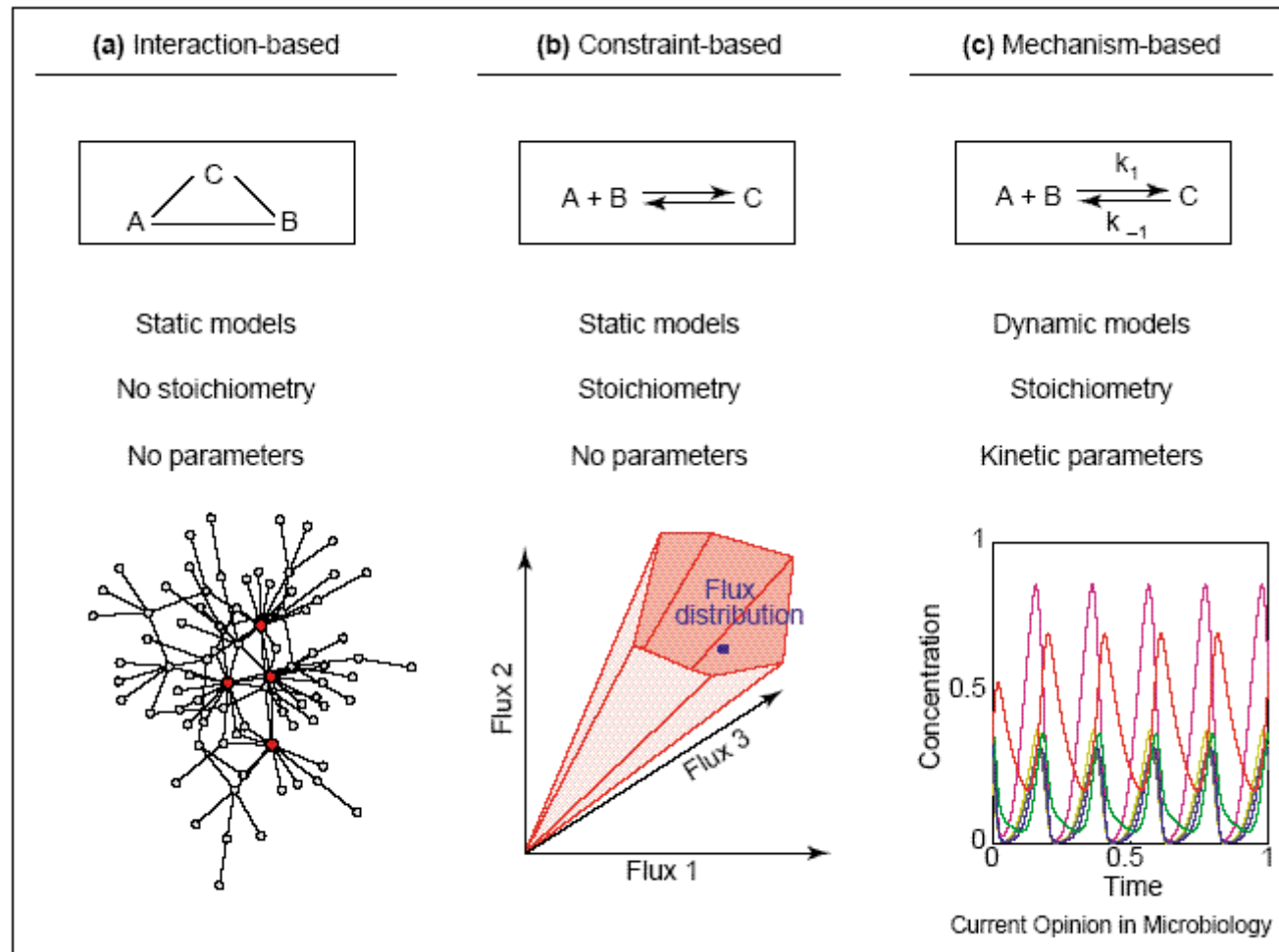
Steady State Flux Analysis

Biochemical reaction $\quad aA + cC \rightarrow eE + hH \quad (v_i)$

$$
\text{Metabolites} \quad
\begin{array}{c}
A \\ B \\ C \\ D \\ E \\ F \\ G \\ H
\end{array}
\left[
\begin{array}{ccccccccc}
\cdot & \cdot & \cdot & \cdot & -a & \cdot & \cdot & \cdot & \cdot \\
 & & & & 0 & & & & \\
 & & & & -c & & & & \\
 & & & & 0 & & & & \\
 & & & & +e & & & & \\
 & & & & 0 & & & & \\
 & & & & 0 & & & & \\
\cdot & \cdot & \cdot & \cdot & +h & \cdot & \cdot & \cdot & \cdot
\end{array}
\right]
$$

v_i (above matrix)

Reactions

Biochemical Reactions

Set of Reactions

$$
\begin{cases}
A(ext) \rightarrow A & R1 \\
B(ext) \leftrightarrow B & R2 \\
P \rightarrow P(ext) & R3 \\
A \rightarrow B & R4 \\
A \rightarrow C & R5 \\
A \rightarrow D & R6 \\
B \leftrightarrow C & R7 \\
B \rightarrow P & R8 \\
C + D \rightarrow P & R9
\end{cases}
\longrightarrow
$$

	R1	R2	R3	R4	R5	R6	R7	R8	R9
A(ext)	-1	0	0	0	0	0	0	0	0
B(ext)	0	-1	0	0	0	0	0	0	0
P(ext)	0	0	1	0	0	0	0	0	0
A	1	0	0	-1	-1	-1	0	0	0
B	0	1	0	1	0	0	-1	-1	0
C	0	0	0	0	1	0	1	0	-1
D	0	0	0	0	0	1	0	0	-1
P	0	0	-1	0	0	0	0	1	1

Flux vector and Mass Balance Equation

$$\frac{d}{dt}\begin{bmatrix} A(ext) \\ B(ext) \\ P(ext) \\ A \\ B \\ C \\ D \\ P \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_{R1} \\ v_{R2} \\ v_{R3} \\ v_{R4} \\ v_{R5} \\ v_{R6} \\ v_{R7} \\ v_{R8} \\ v_{R9} \end{bmatrix}$$

Vector of
metabolites'
concentrations

Fluxes' vector

$$\frac{d}{dt}X = S.v$$

Steady State Assumption

$$\frac{d}{dt}X = 0 \qquad \text{thus} \qquad S.v = 0$$

- The nullspace of S is the whole set of vectors which fulfill this equality.

- This set contains all the fluxes' configurations in steady-state.

Constrained Flux Analysis: The Goal
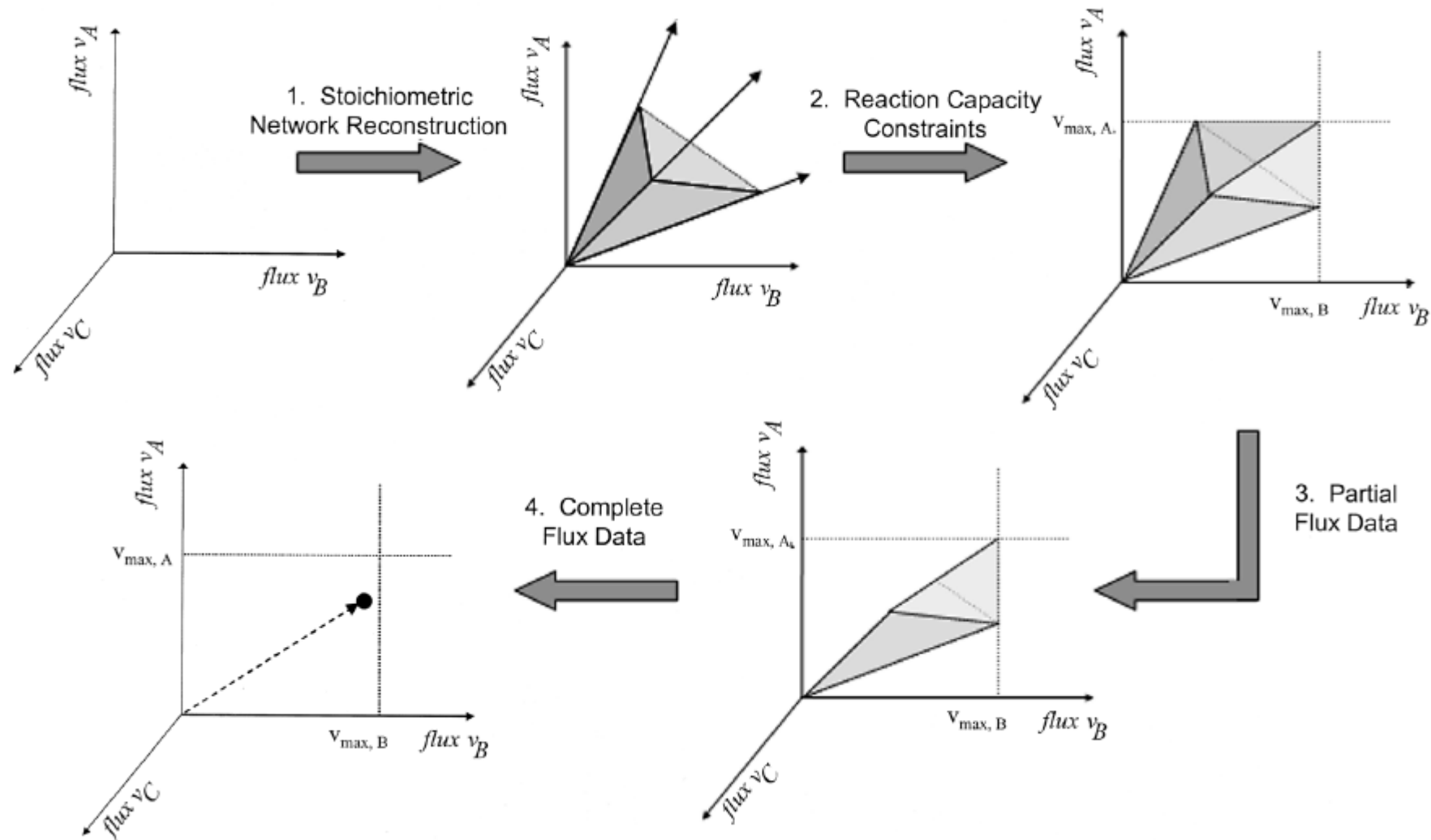
Bernhard Ø. Palsson et al. BIOTECHNOLOGY AND BIOENGINEERING, VOL. 86, NO. 3, MAY 5, 2004

Definitions:

Metabolic networks composed of q reactions and m metabolites:

Stoichiometric matrix: N_{mxq}

Flux distribution: $e=\{e_1,e_2,...,e_q\}$
each element describes the net rate
of the corresponding reaction

The pathway: $P(e)$ where $e_i \neq 0$
identified by the utilized reactions

Conditions behind EFM & EP

- [1] <u>Pseudo steady-state:</u> **Ne=0**. (metabolite balancing equation).

- [2] <u>Feasibility:</u> rate **$e_i \geq 0$** if reaction i is irreversible.

- [3] <u>Non-decomposability:</u> there is no vector **v** (unequal to the zero vector and to **e**) fulfilling [1] and [2] and that **P(v)** is a proper subset of **P(e)**.

Additional Conditions behind EP

- [4] 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
  - No internal reaction can have a negative flux
  - Exchange fluxes can be reversible,but each metabolite can participate in only one exchange flux.

- [5] Systemic independence:
  - The set of EPs in a network (configured properly by [4]) is the minimal set of EFMs that can describe all feasible steady-state flux distributions.
  - The EPs represent a convex basis in this network.
  - The reconfiguration [4] ensures that the set of EPs is unique.

Relevant objective functions

Minimize:

- ATP production
- nutrient uptake
- redox production
- metabolite production

Maximize:

- biomass production (i.e. growth)
- the Euclidean norm of the flux vector

Types of objective functions

- For basic exploration and probing of solution space
- To represent likely physiological objectives
- To represent bioengineering design objectives

α- Spectrum

- Given measured fluxes calculate the minimum and maximum flux rates for each flux rate

The α-spectrum is calculated using linear programming to maximize and minimize the participation of each extreme pathway in a given steady-state flux distribution.

$$Max\ \alpha_i \text{ subject to } \mathbf{v} = \mathbf{P} \cdot \boldsymbol{\alpha},\ i = 1 \ldots n_p, 0 \le \alpha_i \le 1$$
$$Min\ \alpha_i \text{ subject to } \mathbf{v} = \mathbf{P} \cdot \boldsymbol{\alpha},\ i = 1 \ldots n_p, 0 \le \alpha_i \le 1$$
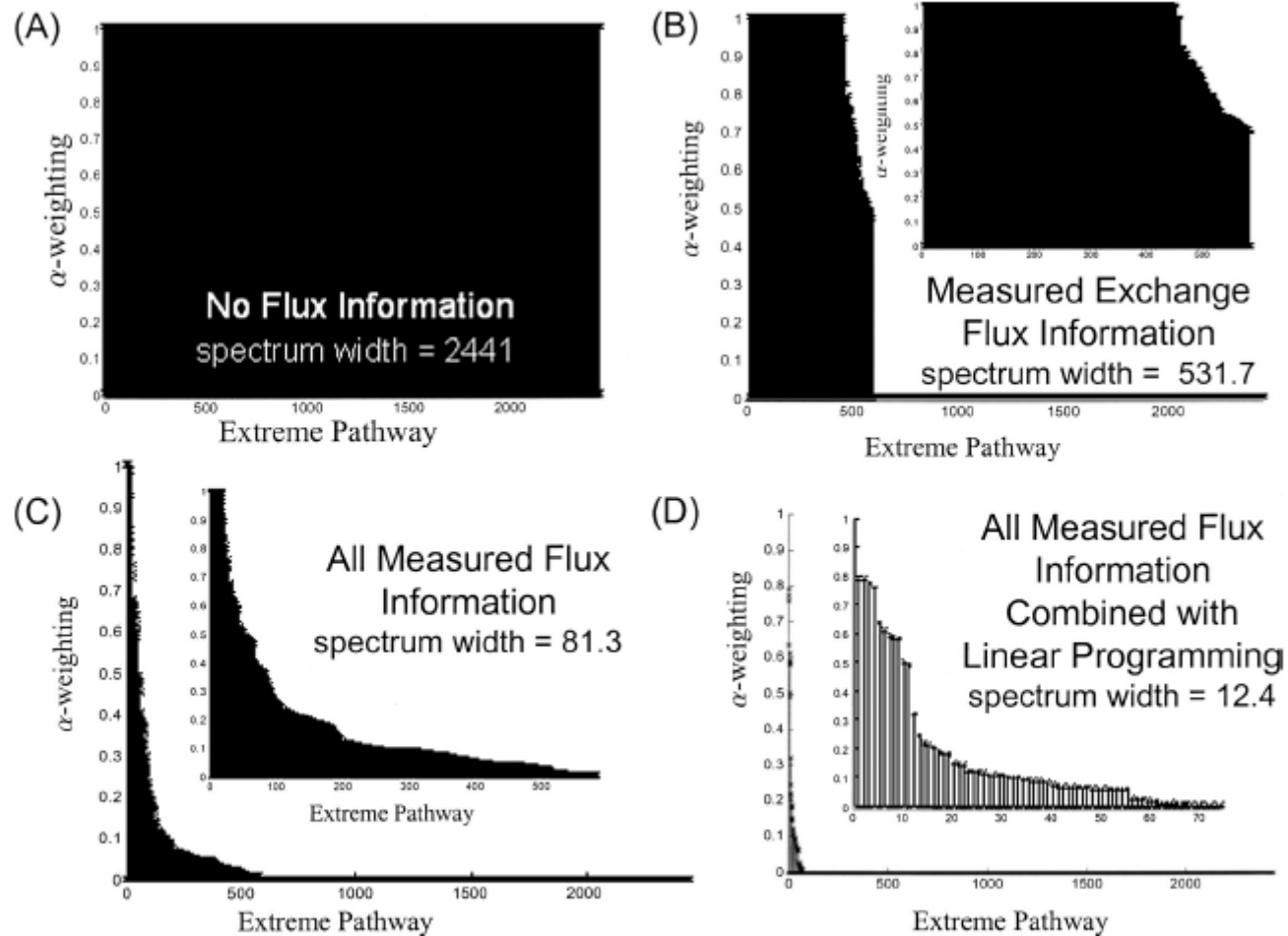
α- Spectrum
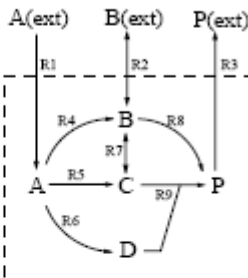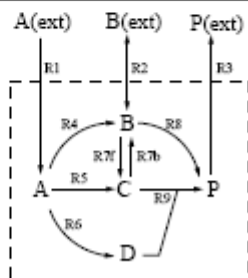
Figure 3. Narrowing the α-spectrum using experimental data. Varying extents of flux knowledge for the low-growth rate (D = 0.09 h⁻¹) glucose-limited case with a 10% error rate incorporated. (A) The α-spectrum with no flux data. (B) The α-spectrum with measured exchange flux data. (C) The α-spectrum with measured internal and exchange flux data. (D) The α-spectrum with all fluxes known based on linear optimization.

The Strategy

- Include as much data as possible
  - Known Fluxes
  - Capacities
  - Objective Functions (optimization)
  - Possible min/max ranges

- Reduce the feasibility space

EFM vs EP (example)

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	Reactions								
	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		Reactions								
	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

EFM vs EP

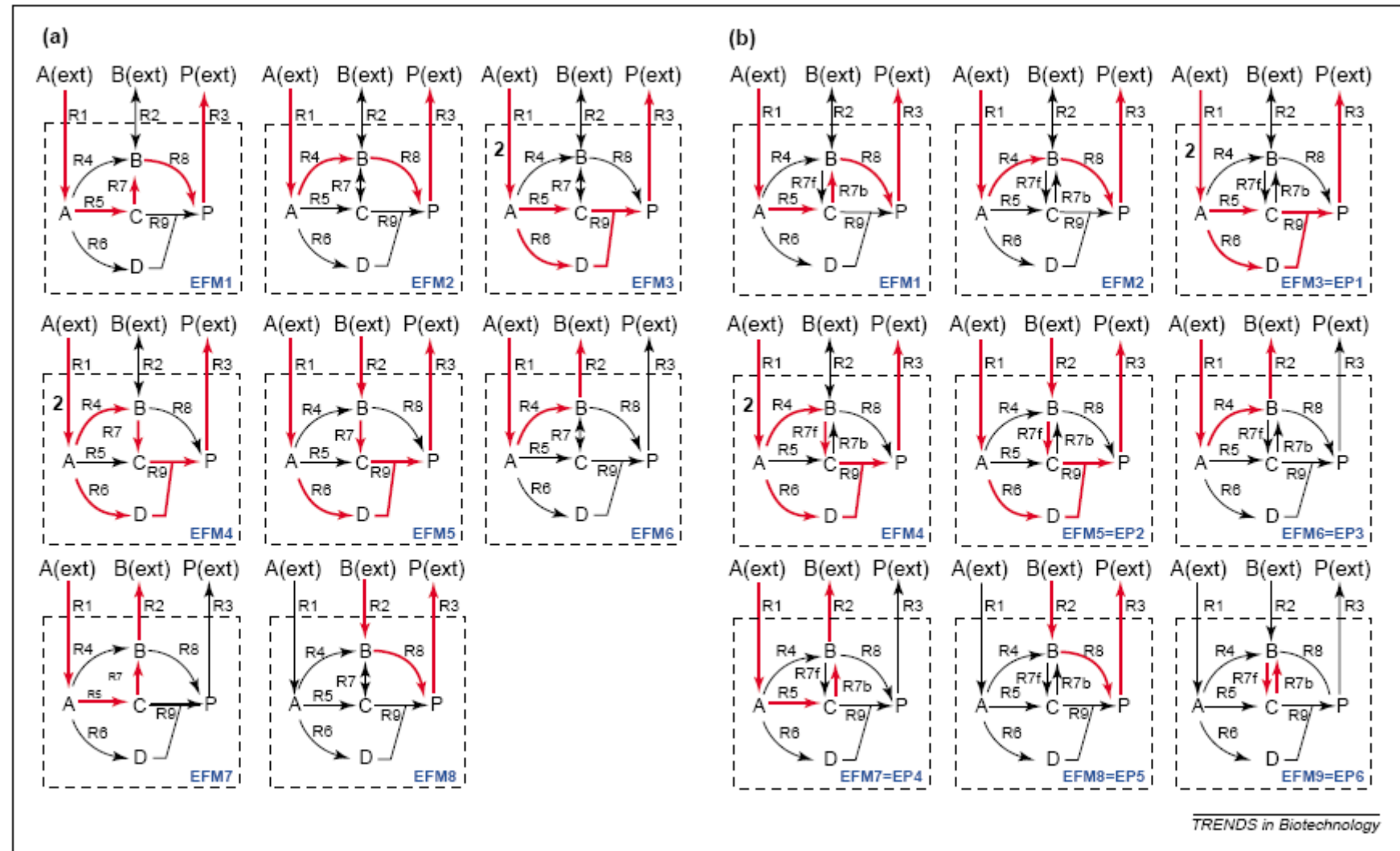Fig. 1. Graphical representation of (a) the elementary flux modes (EFMs) in network N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

EFM vs EP

Table 2. Case study on the applicability of elementary flux modes and extreme pathways

Problem/example	Elementary flux modes (network N1)	Extreme pathways (network N2)
Recognition of operational modes: routes for converting exclusively A to P	Analysis yields four genetically independent routes (EFM1–EFM4)	The set of EPs does not contain all genetically independent routes. Searching for EPs leading from A to P via B, no pathway would be found (but compare with EFM1 and EFM2)
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 suboptimal (R3/R1 = 0.5)[a]	One would only find the suboptimal EP1, not the optimal routes EFM1 and EFM2
Analysis of network flexibility (structural robustness, redundancy): relative robustness of exclusive growth on A or B	Four pathways convert A to P (EFM1–EFM4), whereas for B only one route (EFM8) exists. When one of the internal reactions (R4–R9) fails, for production of P from A two pathways will always 'survive'. By contrast, removing reaction R8 already impedes the production of P from B alone	Only one EP exists for producing P by substrate A alone, and one EP for synthesizing P by (only) substrate B. One might suggest 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
Relative importance of single reactions: relative importance of reaction R8	R8 is essential for producing P by substrate B, whereas for A there is no structurally 'favoured' reaction (R4–R9 all occur twice in EFM1–EFM4). However, considering the optimal modes EFM1,2, one recognizes the importance of R8 also for growth on A	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, as can be easily verified, without this reaction it is impossible to obtain optimal yields (one P per A; EFM1 and EFM2)
Enzyme subsets and excluding reaction pairs: suggest regulatory structures or rules	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 (of course, in an arbitrary composable steady-state flux distribution they might occur together)	The EPs pretend R4 and R8 to be an excluding reaction pair – but they are not (EFM2). The enzyme subsets would be correctly identified. However, one can construct simple examples where the EPs would also pretend wrong enzyme subsets (not shown)
Pathway length: shortest/longest pathway for production of P from A	The shortest pathway from A to P needs two internal reactions (EFM2), the longest requires four (EFM4)	Both the shortest (EFM2) and the longest (EFM4) pathway from A to P are not contained in the set of EPs
Removing a reaction and mutation studies: effect of deleting R7	All EFMs not involving the specific reactions build up the complete set of EFMs in the new (smaller) sub-network. If R7 is cut away, EFMs 2, 3, 6 and 8 would 'survive', hence the mutant is viable	Analyzing a subnetwork implies that the EPs must be newly computed. For example, when cutting away reaction R2 the EFM2 (not contained in the original set of EPs) would become an EP. For this reason mutation studies cannot be performed easily
Constraining reaction reversibility: effect of R7 limited to B → C	For the case of R7, all EFMs but EFM1 and EFM7 'survive', because the latter ones utilize R7 with negative rate	In general, the set of EPs must be recalculated: compare the EPs in network N2 (R2 reversible) and N4 (R2 irreversible)

[a]Unlike analyzing the EFMs, linear optimization as used for example in Flux Balance Analysis (FBA, e.g. [8,30]) finds only a single and, in many cases, a nonunique solution for pathways with optimal product yield. In terms of linear programming, the set of EFMs contains all basic feasible solution. But note, it is generally difficult to consider further constraints (such as fixing different rates to values unequal to zero) when analyzing optimality by EFMs.

What is a Null Space?

$$Ax=b$$

where **A** is a mxn matrix

x is a nx1 vector and

b is a mx1 vector

If m=n and det|A|≠0 then there will a unique solution	If m>n then it is an overdetermined system. Projection methods are used (least squares).	If m<n then it is an underdetermined system.

What is Null Space? Underdetermined Systems

- If Ax = b is consistent and A has full column rank then Ax = b has a unique solution

- If Ax = b is consistent and A does not have full column rank then Ax = b has infinitely many solutions.

- If Ax = b is consistent then there is exactly one solution in the row space of A and it is the solution with smallest norm. This solution is the projection onto row(A) of any solution. To find it solve $AA^T y = b$ and set $x = A^T y$

What is Null Space?

For an underdetermined system:

$$Ax=b$$

$$x=A^T(AA^T)^{-1}b$$

but also

$$Ar=0$$

so the full solution is:

$$x=A^T(AA^T)^{-1}b+rz$$

where r is the Null Space of matrix A

What is Null Space?

- Multiplication with a matrix is a transformation.

- If this transformation is from a higher dimension to a lower one (n to m) then some vectors in n dimensions will be transformed to null. The space spanned by these vectors is called the Null Space.