V16 The Double Description method: Theoretical framework behind EFM and EP / Integration Algorithms

Double Description Method Revisited

Komei Fukuda[1] and Alain Prodon[2]

[1] Institute for Operations Research, ETHZ, CH-8092 Zürich, Switzerland
[2] Department of Mathematics, EPFL, CH-1015 Lausanne, Switzerland

in „Combinatorics and Computer Science Vol. 1120“ edited by Deza, Euler, Manoussakis, Springer, 1996:91

BMC Bioinformatics

BioMed Central

Research article

Open Access

Computation of elementary modes: a unifying framework and the new binary approach

Julien Gagneur[†1] and Steffen Klamt[*†2]

Address: [1]Cellzome AG, Meyerhofstr. 1, 69117 Heidelberg, Germany and [2]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany

Email: Julien Gagneur - julien.gagneur@cellzome.com; Steffen Klamt* - klamt@mpi-magdeburg.mpg.de

* Corresponding author †Equal contributors

Published: 04 November 2004

BMC Bioinformatics 2004, 5:175 doi:10.1186/1471-2105-5-175

This article is available from: http://www.biomedcentral.com/1471-2105/5/175

Received: 28 June 2004
Accepted: 04 November 2004

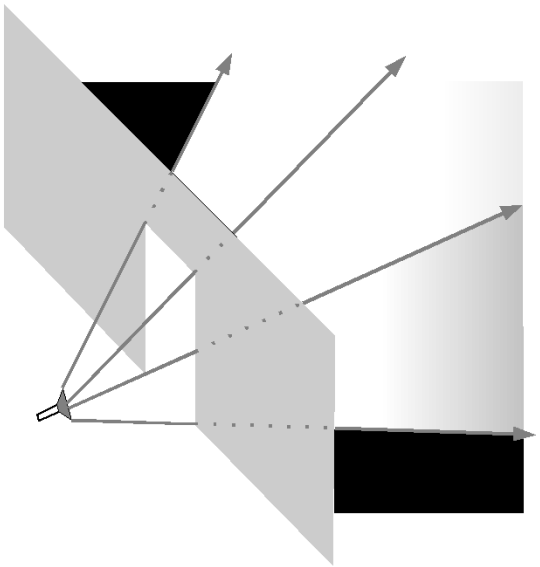
Double Description Method (1953)

The Double Description method is the basis for simple & efficient algorithms for the task of **enumerating extreme rays**.

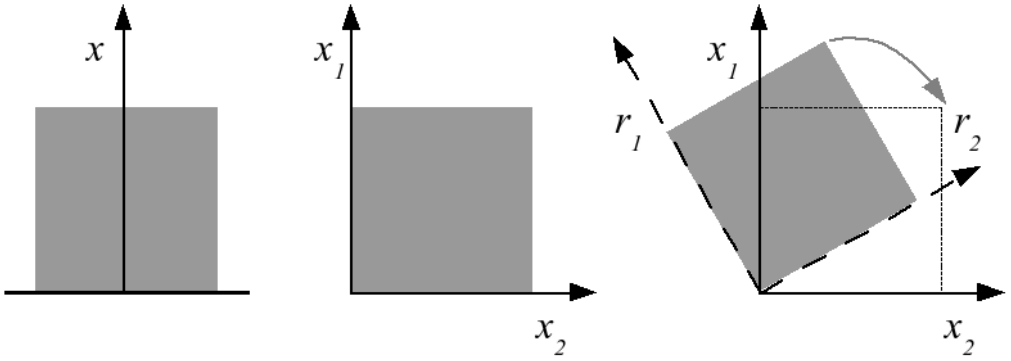
For example, it serves as a framework for popular methods to compute **elementary flux modes** and **extreme pathways**.

Analogy with Computer Graphics problem:

How can one efficiently describe the space in a dark room that is lighted by a torch shining through the open door?



Review: Duality of Matrices



This is the duality

Left: all points above the dividing line (the shaded area) fulfill the condition $x \geq 0$.
Middle: the points in the grey area fulfill the conditions $x_1 \geq 0$ and $x_2 \geq 0$.

But how could we describe the points in the grey area on the right side in a correspondingly simple manner?
Obviously, we could define a new coordinate system (r_1, r_2) as a new set of generating vectors.
But we could also try to transform this area back into the grey area of the middle panel and use the old axes x_1 and x_2.

In 2D, this transformation can be obviously best performed by multiplying all vectors inside the grey area by a two-dimensional **rotation matrix**.

The Double Description Method

A pair (**A,R**) of real matrices **A** and **R** is said to be a **double description pair** or simply a ***DD* pair** if the relationship

A x ≥ **0** if and only if **x** = **R** λ for some λ ≥ **0**

holds. The column size of **A** has to be equal to the row size of **R**, say *d.*

For such a pair, the set *P(***A***)* represented by **A** as $P(\mathbf{A}) = \{\mathbf{x} \in \Re^d : \mathbf{A}\mathbf{x} \geq 0\}$

is simultaneously represented by **R** as $\{\mathbf{x} \in \Re^d : \mathbf{x} = \mathbf{R}\boldsymbol{\lambda} \quad \text{for some} \quad \boldsymbol{\lambda} \geq 0\}$

A subset *P* of $\Re^d$ is called **polyhedral cone** if *P* = *P(***A***)* for some matrix **A**, and **A** is called a **representation matrix** of the polyhedral cone *P(***A***).*

Then, we say **R** is a **generating matrix** for *P*.

Each column vector of a generating matrix **R** lies in the cone *P* and every vector in *P* is a nonnegative combination of some columns of **R**.

The Double Description Method

Theorem 1 (Minkowski's Theorem for Polyhedral Cones)
For any $m \times n$ real matrix **A**, there exists some $d \times m$ real matrix **R** such that (**A**,**R**) is a *DD* pair, or in other words, the cone $P(\mathbf{A})$ is generated by **R**.

The theorem states that every polyhedral cone admits a generating matrix.

The nontriviality comes from the fact that the row size of **R** is finite. If we allow an infinite size, there is a trivial generating matrix consisting of all vectors in the cone.

Herrmann Minkowski
1864-1909

Also the converse is true:

Theorem 2 (Weyl's Theorem for Polyhedral Cones)
For any $d \times n$ real matrix **R**, there exists some $m \times d$ real matrix **A** such that (**A**,**R**) is a *DD* pair, or in other words, the set generated by **R** is the cone $P(\mathbf{A})$.

Herrmann Weyl
1885-1955

The Double Description Method

Task: how does one construct a matrix **R** from a given matrix **A**, and the converse?

These two problems are computationally equivalent.
Farkas' Lemma shows that (**A**,**R**) is a *DD* pair if and only if *(*$\mathbf{R}^T$,$\mathbf{A}^T$*)* is a *DD* pair.

A more appropriate formulation of the problem is to require the minimality of **R***:*
find a matrix **R** such that no proper submatrix is generating *P(***A***)*.

A minimal set of generators is **unique up to positive scaling** when we assume the regularity condition that the cone is **pointed**, i.e. the origin is an extreme point of *P(***A***).*

Geometrically, the columns of a minimal generating matrix are in 1-to-1 correspondence with the **extreme rays** of **P***.*

Thus the problem is also known as the **extreme ray enumeration problem**.

No efficient (polynomial) algorithm is known for the general problem.

Double Description Method: primitive form

Suppose that the $m \times d$ matrix **A** is given and let $P(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} \geq 0\}$

(This is equivalent to the situation at the beginning of constructing EPs or EFMs where **S** is given.)

The *DD* method is an **incremental** algorithm to construct a $d \times m$ matrix **R** such that (**A**,**R**) is a *DD* pair.

Let us assume for simplicity that the cone *P*(**A**) is pointed.

Let **K** be a subset of the row indices {1,2,...,*m*} of **A** and let $\mathbf{A}_\mathbf{K}$ denote the submatrix of **A** consisting of rows indexed by **K**.

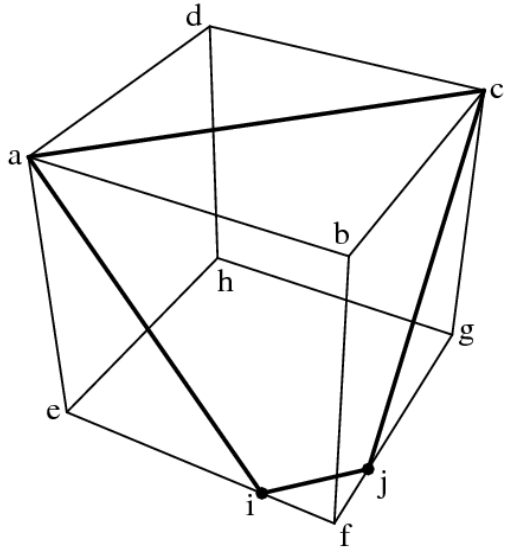
Suppose we already found a generating matrix **R** for $\mathbf{A}_\mathbf{K}$, or equivalently, ($\mathbf{A}_\mathbf{K}$,**R**) is a *DD* pair. If **A** = $\mathbf{A}_\mathbf{K}$, we are done.

Otherwise we select any row index *i* not in **K** and try to construct a *DD* pair ($\mathbf{A}_{\mathbf{K}+\mathbf{i}}$, **R**') using the information of the *DD* pair ($\mathbf{A}_\mathbf{K}$,**R**).

Once this basic procedure is described, we have an algorithm to construct a generating matrix **R** for *P*(**A**).

Geometric version of iteration step

The procedure can be understood geometrically by looking at the cut-section *C* of the cone $P(\mathbf{A}_K)$ with some appropriate hyperplane *h* in $\Re^d$ which intersects with every extreme ray of $P(\mathbf{A}_K)$ at a single point.



Such a cutsection is illustrated in the Figure.

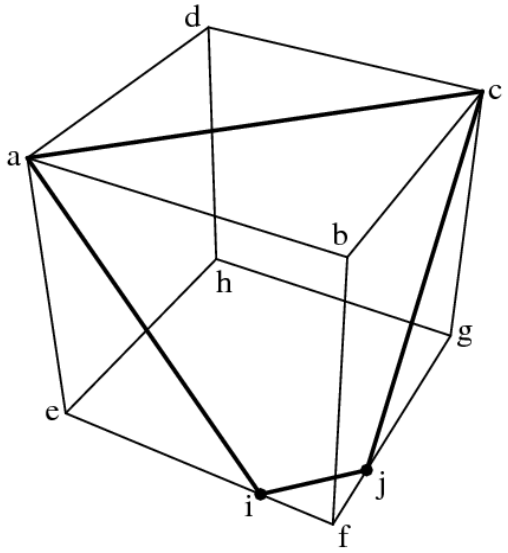
Here, *C* is the cube *abcdefgh.*

Geometric version of iteration step

The newly introduced inequality $\mathbf{A_i \cdot x} \geq 0$ partitions the space $\Re^d$ into three parts:

$$H_i^+ = \{\mathbf{x} \in \Re^d : \mathbf{A_i \cdot x} > 0 \}$$
$$H_i^0 = \{\mathbf{x} \in \Re^d : \mathbf{A_i \cdot x} = 0 \}$$
$$H_i^- = \{\mathbf{x} \in \Re^d : \mathbf{A_i \cdot x} < 0 \}$$

The intersection of H_i^0 with P and the new extreme points i and j in the cut-section C are shown in bold in the Figure.



Geometric version of iteration step

Let J be the set of column indices of the current generating matrix **R**.

The rays $\mathbf{r_j}$ $(j \in J)$ are then partitioned into three parts accordingly:

$$J^+ = \{j \in J : \mathbf{r_j} \in H_i^+ \}$$
$$J^0 = \{j \in J : \mathbf{r_j} \in H_i^0 \}$$
$$J^- = \{j \in J : \mathbf{r_j} \in H_i^- \}$$

We will call the rays indexed by J^+, J^0, J^- the **positive, zero, negative** rays with respect to *i,* respectively.

To construct a matrix **R'** from **R**, we generate new $|J^+| \times |J^-|$ rays lying on the *ith* hyperplane H_i^0

- by taking an appropriate positive combination of each positive ray $\mathbf{r_j}$ and each negative ray $\mathbf{r_{j'}}$ and
- by discarding all negative rays.

Geometric version of iteration step

The following lemma ensures that we have a *DD* pair ($\mathbf{A_{K+i}}$,$\mathbf{R}$‘), and provides the key procedure for the most primitive version of the *DD* method.

Lemma 3 Let *(*$\mathbf{A_K}$,$\mathbf{R}$*)* be a *DD* pair and let *i* be a row index of **A** not in **K**.

Then the pair ($\mathbf{A_{K+i}}$,$\mathbf{R}$‘*)* is a *DD* pair, where $\mathbf{R}$‘ is the $d \times |J'|$ matrix with column vectors $\mathbf{r}_j$ *(*$j \in J'$*)* defined by

$J' = J^+ \cup J^0 \cup (J^+ \times J^-)$, and

$\mathbf{r}_{jj'} = (\mathbf{A}_i \cdot \mathbf{r}_j) \cdot \mathbf{r}_{j'} - (\mathbf{A}_i \cdot \mathbf{r}_{j'}) \cdot \mathbf{r}_j$ for each $(j,j') \in J^+ \times J^-$

Proof omitted.

Finding seed *DD* pair

It is quite simple to find a *DD* pair ($\mathbf{A_K}$,**R**) when |**K**| = 1.
This can serve as the initial *DD* pair.

Another simple (and perhaps the most efficient) way to obtain
an initial *DD* form of *P* is by selecting a maximal submatrix $\mathbf{A_K}$ of **A**
consisting of **linearly independent** rows of **A**.

The vectors $\mathbf{r_j}$'s of matrix **R** are then obtained by solving the system of equations

$$\mathbf{A_K}\,\mathbf{R} = \mathbf{I}$$

where **I** is the identity matrix of size |**K**|.

As we have assumed rank(**A**) = *d,* i.e. $\mathbf{R} = \mathbf{A_K}^{-1}$ **,**
the pair ($\mathbf{A_K}$,**R**) is clearly a *DD* pair,
since $\mathbf{A_K}\cdot\mathbf{x} \geq \mathbf{0} \leftrightarrow \mathbf{x} = \mathbf{A_K}^{-1}\lambda,\ \lambda \geq \mathbf{0}$.

Primitive algorithm for DoubleDescriptionMethod

```
procedure DoubleDescriptionMethod(A);
begin
    Obtain any initial DD pair (A_K, R);
    while K ≠ {1, 2, ..., m} do
    begin
        Select any index i from {1, 2, ..., m} \ K;
        Construct a DD pair (A_{K+i}, R') from (A_K, R);
            /* by using Lemma 3 */
        R := R';    K := K + i;
    end;
    Output R;
end.
```

This algorithm is very primitive.
The straightforward implementation will be quite useless because the size of *J* increases extremely fast.

This is because many vectors $\mathbf{r}_{jj'}$ generated by the algorithm defined in Lemma 3 are unnessary.
We need to avoid generating redundant vectors!

To avoid generating redundant vectors, we will use the zero set or active set Z(**x**) which is the set of inequality indices satisfied by **x** in P(**A**) with equality.

Noting $\mathbf{A}_{i\cdot}$ the ith row of **A**, $Z(\mathbf{x}) = \{i : \mathbf{A}_{i\cdot}\mathbf{x} = 0\}$

Towards the standard implementation

Two distinct extreme rays **r** and **r'** of *P* are **adjacent**
if the minimal face of *P* containing both rays contains no other extreme rays.

Proposition 7. Let **r** and **r'** be distinct rays of *P.*

Then the following statements are equivalent
(a) **r** and **r'** are adjacent extreme rays,

(b) **r** and **r'** are extreme rays and the rank of the matrix $\mathbf{A}_{Z(\mathbf{r}) \cap Z(\mathbf{r}')}$ is $d - 2$,

(c) if **r"** is a ray with $Z(\mathbf{r}") \supset Z(\mathbf{r}) \cap Z(\mathbf{r}')$ then either $\mathbf{r}" \cong \mathbf{r}$ or $\mathbf{r}" \cong \mathbf{r}'$.

Towards the standard implementation

Lemma 8. Let $(\mathbf{A}_K,\mathbf{R})$ be a *DD* pair such that rank($\mathbf{A}_K$) = d and let i be a row index of **A** not in K.

Then the pair ($\mathbf{A}_{K+i}$, $\mathbf{R}'$) is a *DD* pair, where $\mathbf{R}'$ is the $d \times |J'|$ matrix with column vectors $\mathbf{r}_j$ ($j \in J'$) defined by

$J' = J^+ \cup J^0 \cup \text{Adj}$

Adj = $\{(j,j') \in J^+ \times J^- : \mathbf{r}_j$ and $\mathbf{r}_{j'}$ are adjacent in $P(\mathbf{A}_K)\}$ and

$\mathbf{r} = (\mathbf{A}_i\, \mathbf{r}_j)\, \mathbf{r}_{j'} - (\mathbf{A}_i \mathbf{r}_j)\, \mathbf{r}_j$ for each $(j,j') \in$Adj.

Furthermore, if **R** is a minimal generating matrix for $P(\mathbf{A}_K)$ then $\mathbf{R}'$ is a minimal generating matrix for $P(\mathbf{A}_{K+i})$.

Algorithm for standard form of double description method

This is now a straightforward variation of the *DD* method which produces a minimal generating set for P:

```
procedure DDMethodStandard(A)
begin
    Obtain any initial DD pair (A_K, R); such that R is minimal
    while K ≠ {1, 2, ..., m} do
    begin
        Select any index i from {1, 2, ..., m} \ K;
        Construct a DD pair (A_{K+i}, R') from (A_K, R);
            /* by using Lemma 8 */
        R := R';    K := K + i;
    end;
    Output R;
end.
```

To implement `DDMethodStandard`, we must check for each pair of extreme rays **r** and **r'** of $P(\mathbf{A}_K)$ with $\mathbf{A}_i\,\mathbf{r} > 0$ and $\mathbf{A}_i\,\mathbf{r}' < 0$ whether they are adjacent in $P(\mathbf{A}_K)$.

This completes our quick look at the Double Description method.

V16 – part II – applications of FBA and EFM

Review:

(1) The concept of metabolic networks required revising the traditional picture of separate biochemical pathways into a **densely-woven metabolic network**

(2) **Connectivity** of substrates in this network follows a **power-law** (Yeong&Barabasi).

(3) **Constraint-based modeling** (FBA) enables
to analyze the **capabilities** of cellular **metabolism** including e.g.

- its capacity to predict **deletion phenotypes**
- the ability to calculate the **relative flux values** of metabolic reactions, and
- the capability to identify properties of **alternate optimal growth states**
  in a wide range of simulated environmental conditions

Open questions

- what parts of metabolism are involved in **adaptation** to environmental conditions?
- is there a central essential metabolic **core**?
- what role does **transcriptional regulation** play?

Central metabolism of *E.coli* characterized by EFMs

Catabolic part: substrate uptake reactions, glycolysis, pentose phosphate pathway, TCA cycle, excretion of by-products (acetate, formate, lactate, ethanol)

Anabolic part: conversions of precursors into building blocks like amino acids, to macromolecules, and to biomass.

Table 1 **Number and distribution of elementary flux modes.**

Selection*		Glucose	Acetate	Glycerol	Succinate	Sum
-	N	27,099	598	11,332	4,249	43,279
Growth only	$N(\mu, \neq ATP)$	73.1%	58.7%	78.6%	76.3%	74.6%
ATP only	$N(\neq \mu, ATP)$	3.2%	5.0%	2.4%	2.4%	3.0%
Growth and ATP	$N(\mu, ATP)$	6.6%	2.0%	5.1%	4.2%	5.9%
No growth/ATP	$N(\neq \mu, \neq ATP)$	17.1%	34.3%	13.9%	17.1%	16.5%
Aerobic growth	$N(\mu, O_2)$	73.1%	60.7%	83.6%	80.5%	76.4%
Anaerobic growth	$N(\mu, \neq O_2)$	6.6%	0.0%	0.0%	0.0%	4.1%

*We denote the number of elementary flux modes simultaneously meeting a set of conditions, $C_1, \ldots, C_n$, by $N(C_1, \ldots, C_n)$. These conditions include, for example, the situation where cells can grow, which is abbreviated by μ. Excess energy production in the form of ATP (ATP), the substrate metabolized (S_k for the k-th substrate) and oxygen uptake (O_2) are specified accordingly. The operator ' $\neq$ ' indicates that certain fluxes must not occur. The total number of modes includes one futile cycle without substrate uptake.

Stelling et al. Nature 420, 190 (2002)

Jörg Stelling
ETH Zürich

Metabolic network topology ↔ phenotype

Question:

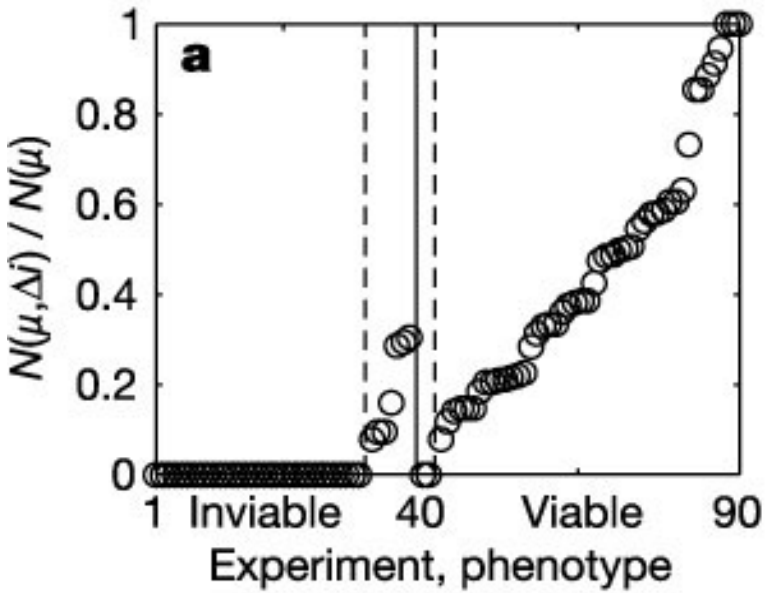
Can the total number of EFMs for given conditions be used as quantitative measure of metabolic flexibility?

Δi : deletion mutant of gene *i*

μ : ability to grow

N : number of EFMs enabling wild-type or deletion mutants in *E. coli* to grow

Shown are results for 90 deletions of different individual genes relative to the situation for wild-type.



Answer: Yes, for more than 90% of single gene deletions,
the number of EFMs for the mutant strain was correctly associated with the growth phenotype.

Stelling et al. Nature 420, 190 (2002)

EFM-based robustness analysis

The # of EFMs qualitatively indicates whether a mutant is viable or not, but does not describe quantitatively how well a mutant grows.

Define **maximal biomass yield** Y^{max} as the optimum of:

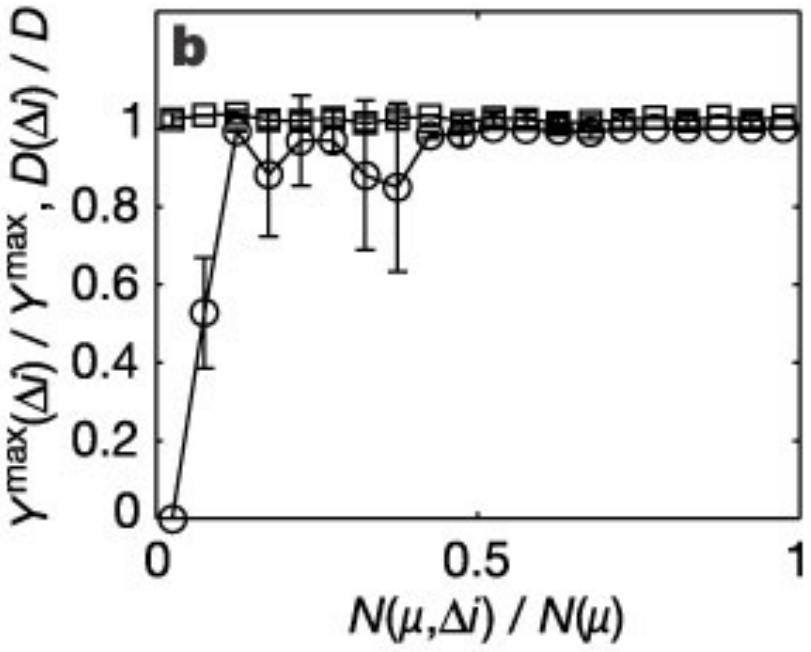
$$Y_{i,X/S_i} = \frac{e_i^{\mu}}{e_i^{S_k}}$$

e_i is the single reaction rate (growth and substrate uptake) in EFM i selected for utilization of substrate S_k.

Thus, Y^{max} selects the EFM where most substrate medium is converted into biomass ($0 \leq Y^{max} \leq 1$).

Stelling et al. Nature 420, 190 (2002)

EFM-based robustness analysis



X-axis: fraction of elementary modes operational in the mutants.

Y-axis: Open squares: relative network diameter $D(\Delta i)$ / D (is essentially constant)

Open circles: maximal growth yield of the mutant $Y^{max}(\Delta i)$ (open circles)

→ Central metabolism of *E.coli* behaves in a highly **robust** manner.

Even mutants with significantly reduced metabolic flexibility (> 15% or so) show a growth yield similar to wild type.

Stelling et al. Nature 420, 190 (2002)

Distribution of fluxes in *E.coli*

Global organization of metabolic fluxes in the bacterium *Escherichia coli*

E. Almaas[1], B. Kovács[1,2], T. Vicsek[2], Z. N. Oltvai[3] & A.-L. Barabási[1]

[1]*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA*
[2]*Biological Physics Department and Research Group of HAS, Eötvös University, H-1117 Budapest, Hungary*
[3]*Department of Pathology, Northwestern University, Chicago, Illinois 60611, USA*

Nature 427, 839 (2004)

Aim: understand principles that govern the use of individual reactions under different growth conditions.

Stoichiometric matrix for *E.coli* strain MG1655 containing 537 metabolites and 739 reactions was taken from Palsson *et al*.

Apply FBA to characterize solution space
(all possible flux states under a given condition).

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} v_j = 0$$

v_j is the flux of reaction j and S_{ij} is the stoichiometric coefficient of reaction j.

Optimal states

Denote the **mass** carried by reaction *j* producing (consuming) metabolite *i* by

$$\hat{v}_{ij} = \left|S_{ij}\right| v_j$$

Observation:

Fluxes vary widely: e.g. the dimensionless flux of the succinyl coenzyme A synthetase reaction is 0.185, whereas the flux of the aspartate oxidase reaction is 10.000 times smaller, 2.2×10^{-5}.

Use FBA to compute flux states that optimize cell growth on various substrates.

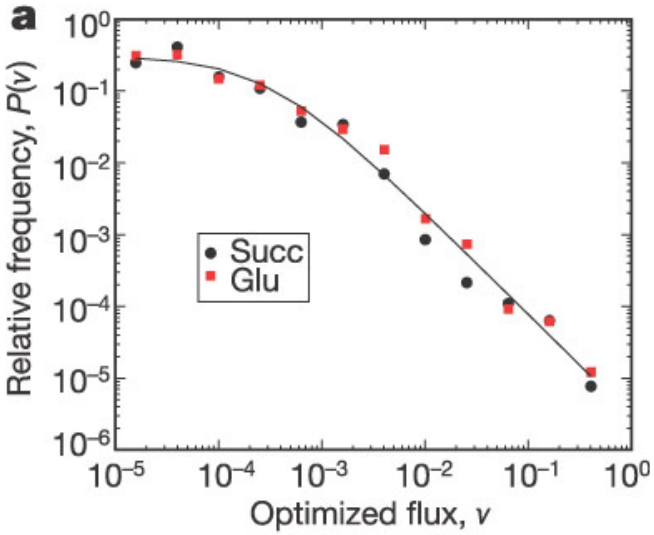
Focus on active (non-zero flux) reactions of *E.coli*.

Compare growth on glutamate- or succinate-rich substrate media.

Overall flux organization of *E.coli* metabolic network

a, Flux distribution for optimized biomass production on succinate (black) and glutamate (red) substrates.

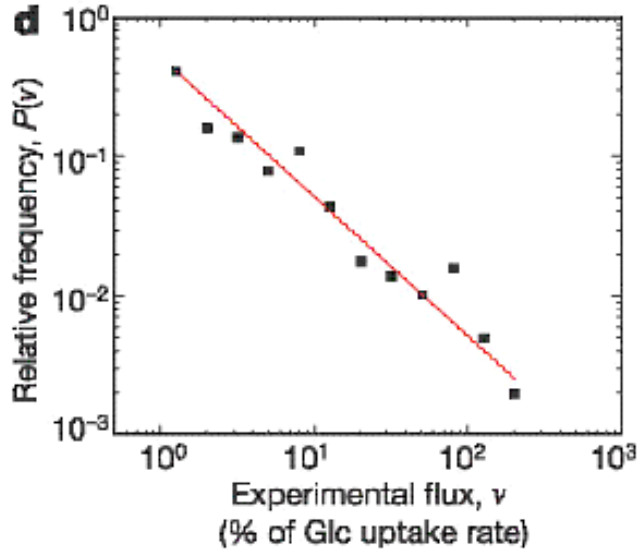
Solid line : power-law fit



d, Experimentally determined fluxes for reactions of the central metabolism of *E. coli.*

Clear **power-law** behaviour.
Best fit with $P(v) \propto v^{-\alpha}$ with $\alpha = 1$.

Both computed and experimental flux distribution show wide spectrum of fluxes.



Almaar et al., Nature 427, 839 (2004)

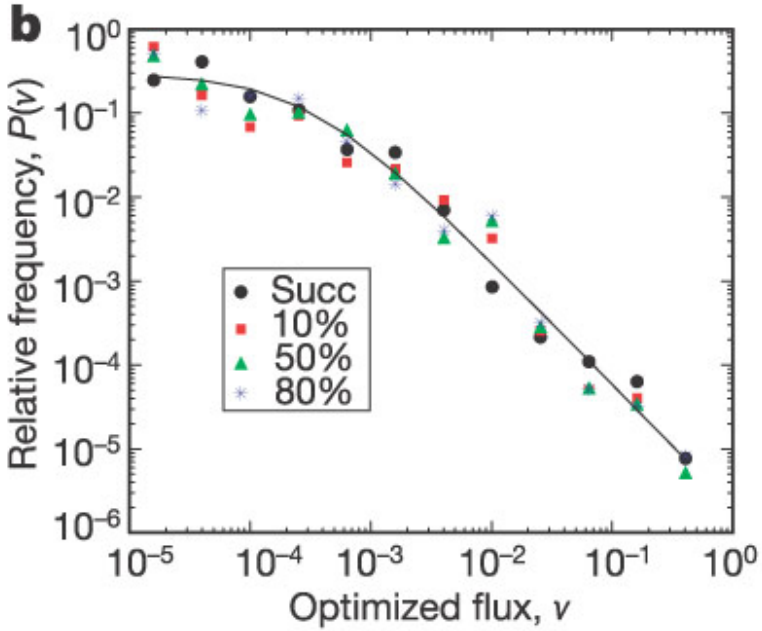
Response to different environmental conditions

Is the flux distribution independent of environmental conditions?

Black: Flux distribution for optimized biomass on pure succinate substrate.

Red / **green** / **blue** :
Flux distributions when an additional 10%, 50%, or 80% of randomly chosen subsets of the 96 input channels (substrates) are added to succinate.

The flux distribution was averaged over 5,000 independent random choices of uptake metabolites.



→ **Yes, the flux distribution is independent of the external conditions.**

Almaar et al., Nature 427, 839 (2004)

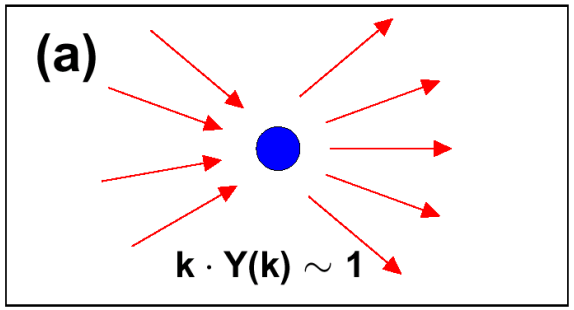
Use scaling behavior to determine local connectivity

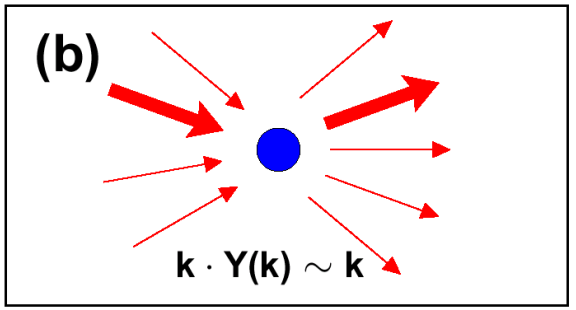
The observed flux distribution is compatible with two different potential local flux structures:

(a)a **homogenous local organization** would imply that all reactions producing (consuming) a given metabolite have **comparable fluxes**

(b) a more delocalized „**high-flux backbone** (HFB)“ is expected if the local flux organisation is heterogenous such that each metabolite has a dominant source (consuming) reaction.

$$Y(k,i) = \sum_{j=1}^{k} \left[\frac{\hat{v}_{ij}}{\sum_{l=1}^{k} \hat{v}_{il}} \right]^2$$





$$k \times \left(k \left(\frac{v}{k \cdot v} \right)^2 \right) = 1$$

All fluxes v_{ij} are the same, say v.

$$k \times \left(\left(\frac{v_{max}}{v_{max}} \right)^2 \right) = k$$

One flux dominates -> replace sum by this flux v_{max}.

Almaar et al., Nature 427, 839 (2004)

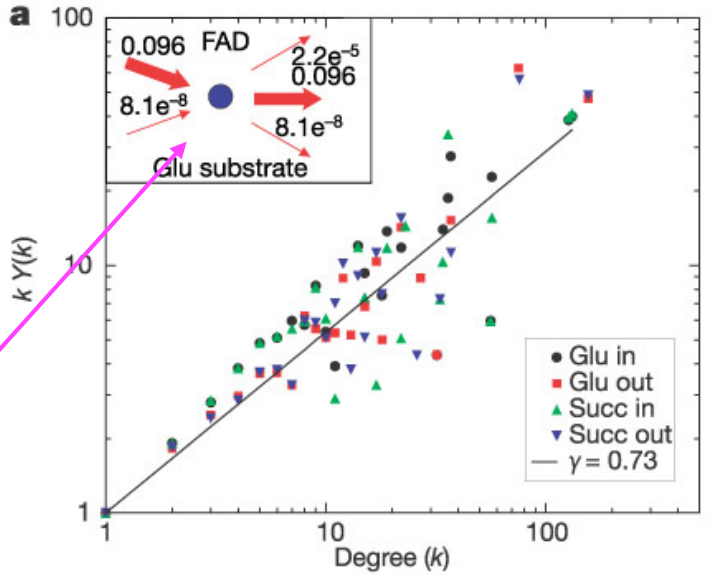
Characterizing the local inhomogeneity of the flux net

FBA-computed $kY(k)$ as a function of k, averaged over all metabolites shows linear dependence $k \times Y(k) \propto k^{0.73}$ with slope 0.73.
This is true for incoming and outgoing reactions.

→ an **intermediate behavior** is found between the two extreme cases discussed before.

→ the large-scale inhomogeneity observed in the overall flux distribution is also valid at the level of the individual metabolites.

The more reactions consume (produce) a given metabolite, the more likely a single reaction carries most of the flux, see inset (FAD).



Inset shows non-zero mass flows producing (consuming) FAD on a glutamate-rich substrate.

Almaar et al., Nature 427, 839 (2004)

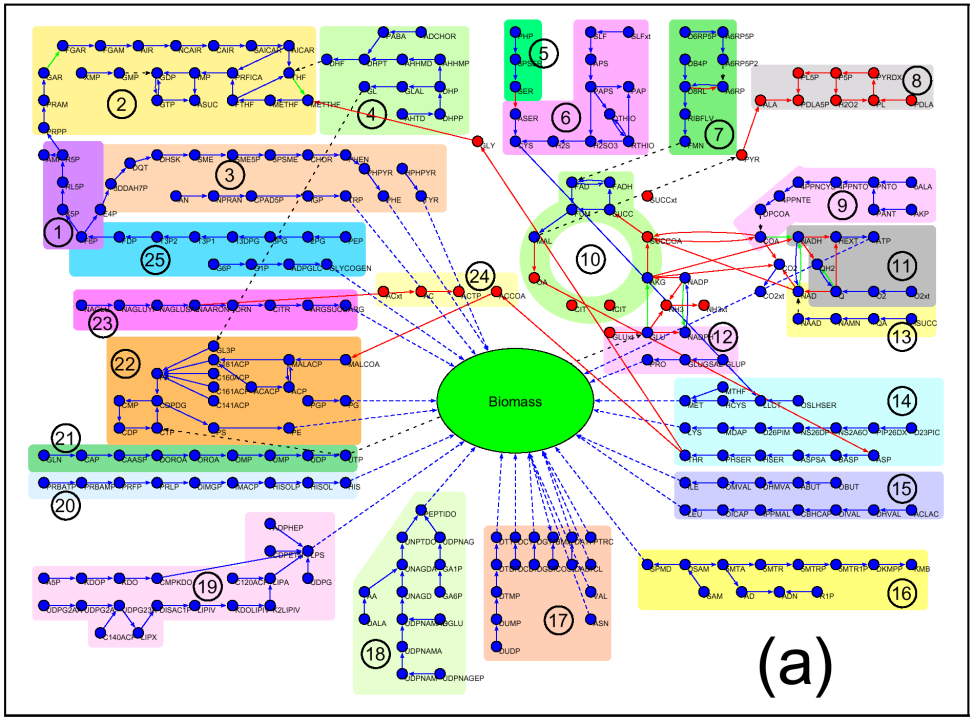
Clean up metabolic network

Use simple algorithm that removes for each metabolite systematically all reactions but the one providing the largest incoming (outgoing) flux distribution.

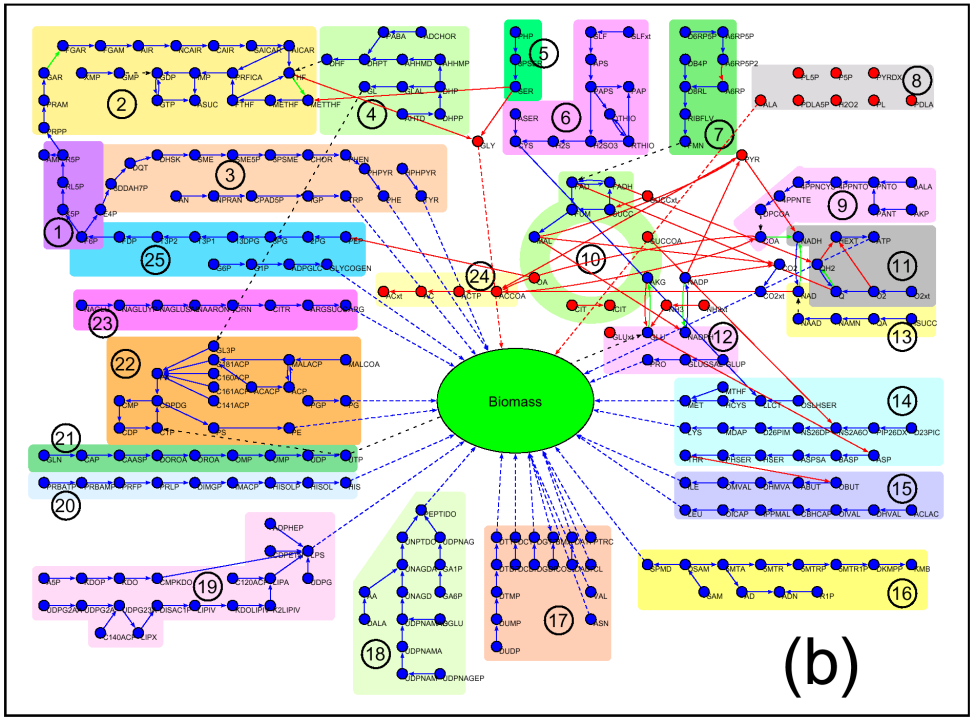
This algorithm uncovers the „**high-flux-backbone**“ of the metabolism.

Almaar et al., Nature 427, 839 (2004)

High-flux backbone of *E.coli* metabolic network



glutamate rich medium



succinate rich medium

Directed links: Metabolites A and B are connected with an arc from A to B if the reaction with maximal flux consuming A is the reaction with maximal flux producing B.
Shown are all metabolites that have at least one neighbour after completing this procedure.

Background colours : known biochemical pathways.

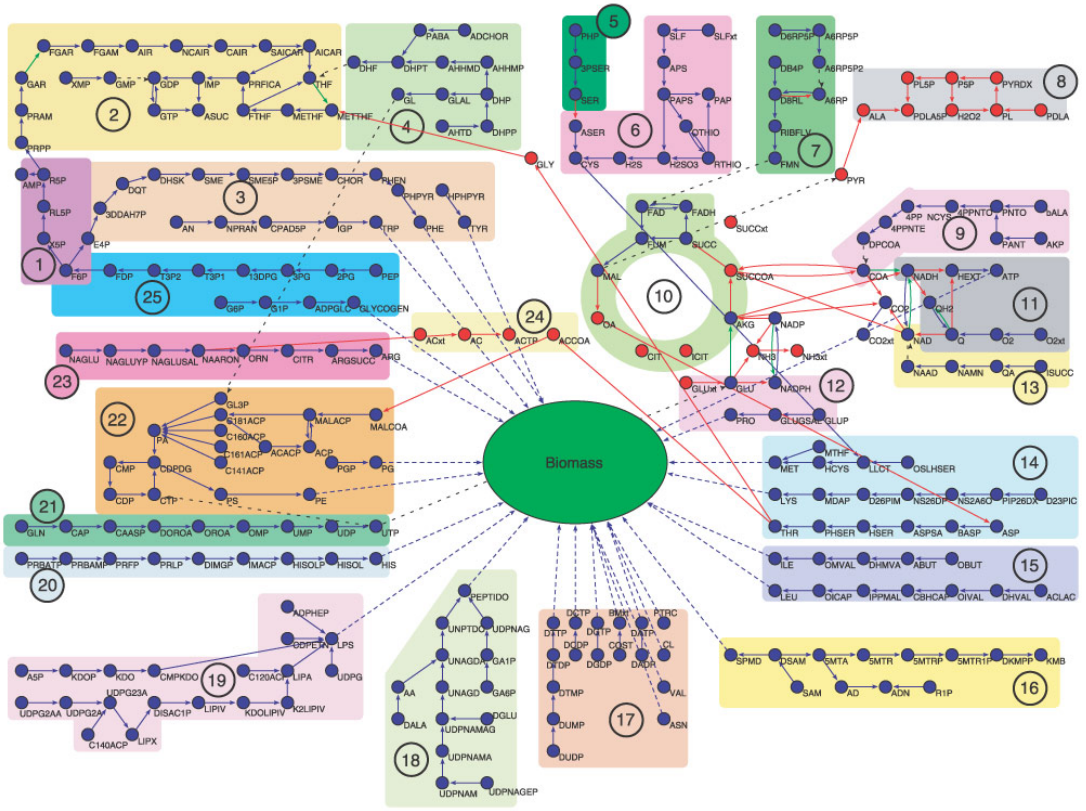
Almaar et al., Nature 427, 839 (2004)

FBA-optimized high-flux backbone on glutamate-rich medium

Blue colored **Metabolites** (vertices) have at least one neighbour in common in glutamate- and succinate-rich substrates.

Red colored nodes have no common neighbors („rewiring“)

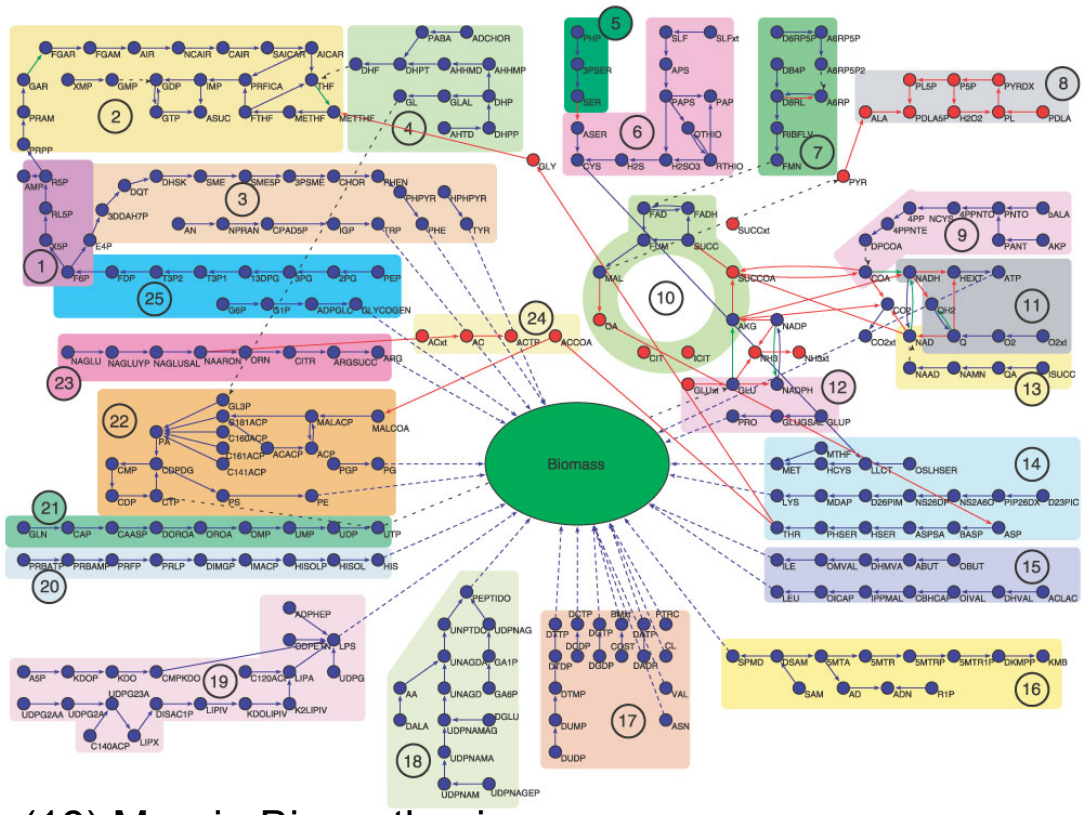
Reactions (lines) are coloured
blue if they are identical in glutamate- and succinate-rich substrates,
green if a different reaction connects the same neighbour pair, and
red if this is a new neighbour pair („rewiring“).



Black dotted lines indicate where the disconnected pathways, e.g., folate biosynthesis (4), would connect to the cluster through a link that is not part of the HFB.
Thus, the red nodes and links highlight the predicted changes in the HFB when shifting *E. coli* from glutamate- to succinate-rich media.
Dashed lines indicate links to the biomass growth reaction.

Almaar et al., Nature 427, 839 (2004)

FBA-optimized high-flux backbone on glutamate-rich medium



(1) Pentose Phospate
(2) Purine Biosynthesis
(3) Aromatic Amino Acids
(4) Folate Biosynthesis
(5) Serine Biosynthesis
(6) Cysteine Biosynthesis
(7) Riboflavin Biosynthesis
(8) Vitamin B6 Biosynthesis
(9) Coenzyme A Biosynthesis
(10) TCA Cycle
(11) Respiration
(12) Glutamate Biosynthesis
(13) NAD Biosynthesis
(14) Threonine, Lysine and Methionine Biosynthesis
(15) Branched Chain Amino Acid Biosynthesis
(16) Spermidine Biosynthesis
(17) Salvage Pathways
(18) Murein Biosynthesis
(19) Cell Envelope Biosynthesis
(20) Histidine Biosynthesis
(21) Pyrimidine Biosynthesis
(22) Membrane Lipid Biosynthesis
(23) Arginine Biosynthesis
(24) Pyruvate Metabolism
(25) Glycolysis

Almaar et al., Nature 427, 839 (2004)

Interpretation

Only a few pathways appear disconnected.

This indicates that although these pathways are part of the HFB, their end product is only the second-most important source for another HFB metabolite.

Groups of individual **HFB reactions largely overlap**
with traditional biochemical partitioning of cellular metabolism ☺

Almaar et al., Nature 427, 839 (2004)

How sensitive is the HFB to changes in the environment?

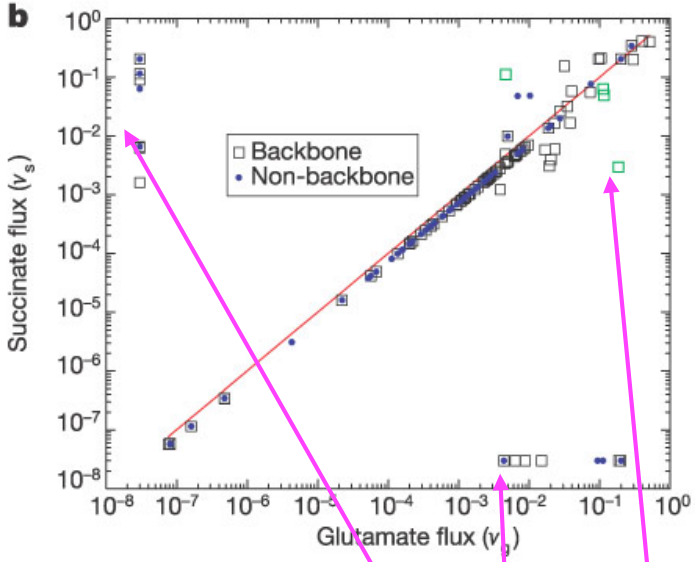
Fluxes of individual reactions on glutamate-rich and succinate-rich medium.

Black squares: reactions belonging to the HFB,
blue dots : remaining reactions

Green squares : reactions in which the direction of the flux is reversed.

Reactions with **negligible** flux changes follow the diagonal (solid line).

Some reactions are turned off in only one of the conditions (shown close to the coordinate axes).



Only reactions in the high-flux territory undergo noticeable differences!

Type I: reactions turned on in one conditions and off in the other.

Type II: reactions remain active but show an orders-in-magnitude shift in flux under the two different growth conditions.

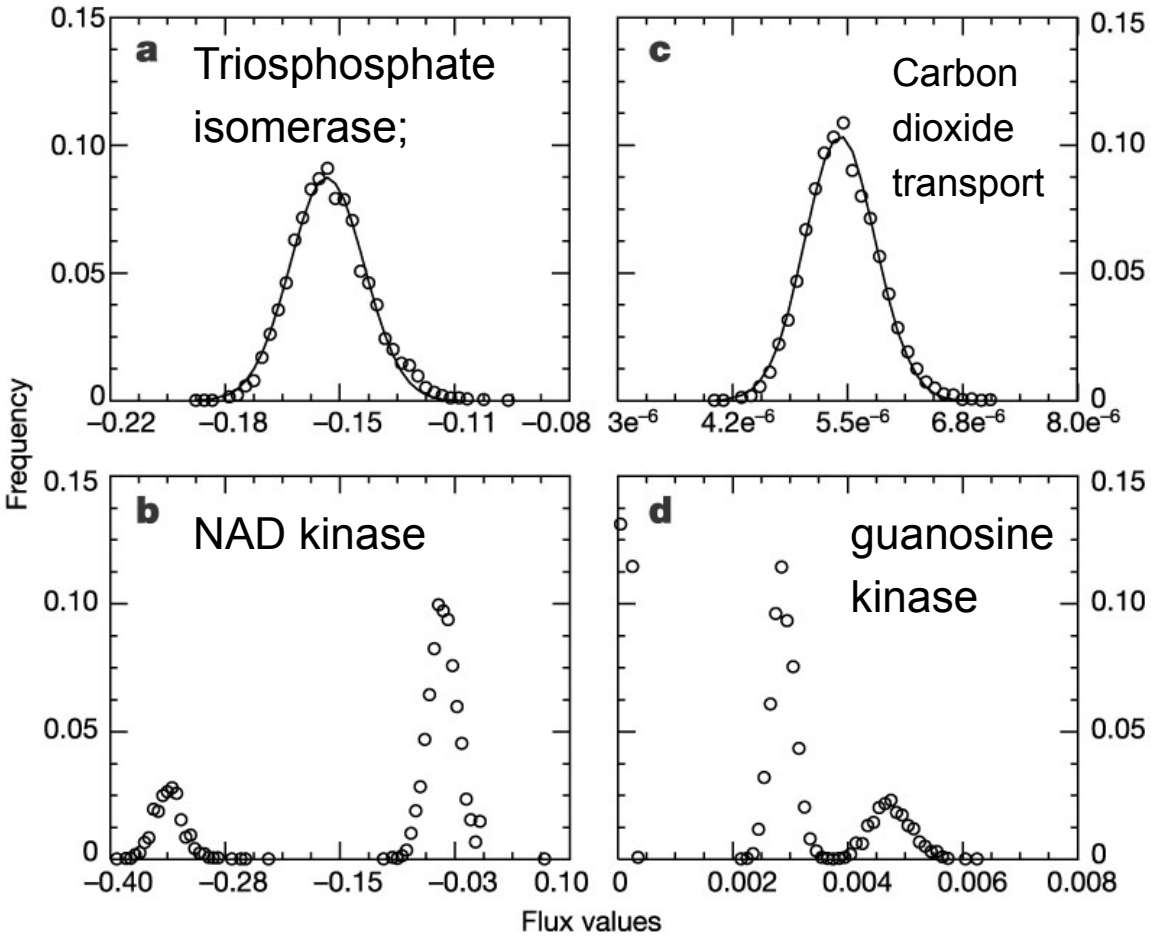
Almaar et al., Nature 427, 839 (2004)

Flux distributions for individual reactions

Shown is the flux distribution for 4 selected *E. coli* reactions on a 50% random medium.

Reactions with small fluxes have **unimodal/gaussian distributions** (a and c).
Shifts in growth-conditions only lead to small changes of their flux values.

Off-diagonal reactions have **multimodal distributions** (b and d), showing several discrete flux values under diverse conditions.



Almaar et al., Nature 427, 839 (2004)

Summary

Metabolic network use is **highly uneven** (power-law distribution) at the global level and at the level of the individual metabolites.

Whereas most metabolic reactions have low fluxes, the overall activity of the metabolism is dominated by several reactions with very high fluxes.

E. coli responds to changes in growth conditions by reorganizing the rates of selected fluxes predominantly within this high-flux backbone.

Apart from minor changes, the use of the other pathways remains unaltered.

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

The Activity Reaction Core and Plasticity of Metabolic Networks

Eivind Almaas[1,2], Zoltán N. Oltvai[3*], Albert-László Barabási[2,4]

PLoS Computational Biology | www.ploscompbiol.org 0557 December 2005 | Volume 1 | Issue 7 | e68

The same authors as before used FBA to examine utilization and relative flux rate of each metabolite in various simulated environmental conditions for *E.coli, H. pylori* and *S. cerevisae*:
For each system they considered 30.000 randomly chosen combinations where each uptake reaction is assigned a random value between 0 and 20 mmol/g/h.

→ **adaptation** to different conditions occurs by 2 mechanisms:
(a) **flux plasticity**: changes in the fluxes of already active reactions.
E.g. changing from glucose- to succinate-rich conditions alters the flux of 264 *E.coli* reactions by more than 20%

(b) less often, adaptation includes **structural plasticity**, turning on previously zero-flux reactions or switching off active pathways.

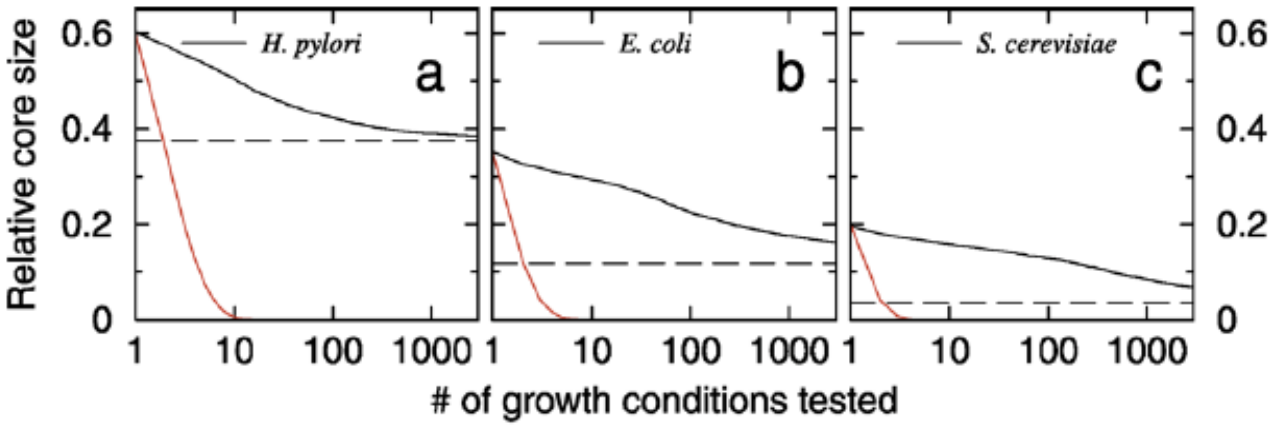
Emergence of the Metabolic Core

The two adaptation mechanisms enable a group of reactions that are not subject to structural plasticity to be active under all environmental conditions.

Are these **core reactions** randomly distributed?

If typically a fraction q of the metabolic reactions were active under a specific growth condition,
we would expect for n distinct conditions an overlap of at least q^n reactions.
This converges quickly to 0.

Emergence of the Metabolic Core

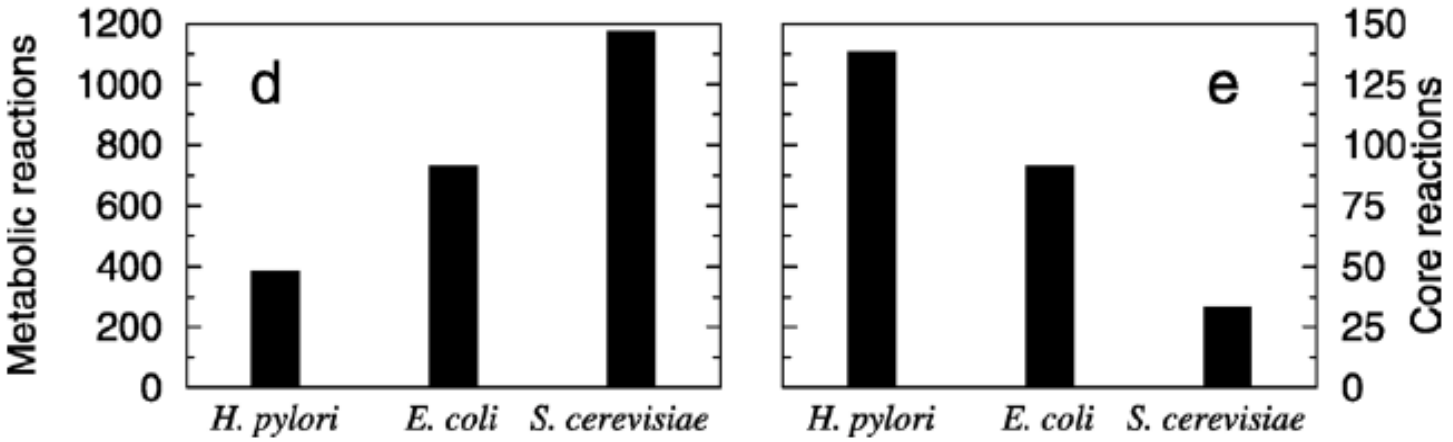


(a–c) Average relative size of the number of reactions that are always active as a function of the number of sampled conditions (black line).

As the number of conditions increases, the curve converges to a constant marked by the dashed line, identifying the **metabolic core** of an organism.

Red line : number of reactions that are always active if activity is randomly distributed in the metabolic network.
The fact that it converges to zero indicates that the real core represents a collective network effect, forcing a group of reactions to be active in all conditions.

Emergence of the Metabolic Core



The number of metabolic reactions (d) and the number of metabolic core reactions (e) in the 3 studied organisms.

As the complexity of the organism increases (the prokaryote *H. pylori* has fewest reactions, the prokaryote *E. coli* has more, and the eukaryote *S. cerevisiae* has most), the number of core reactions decreases.

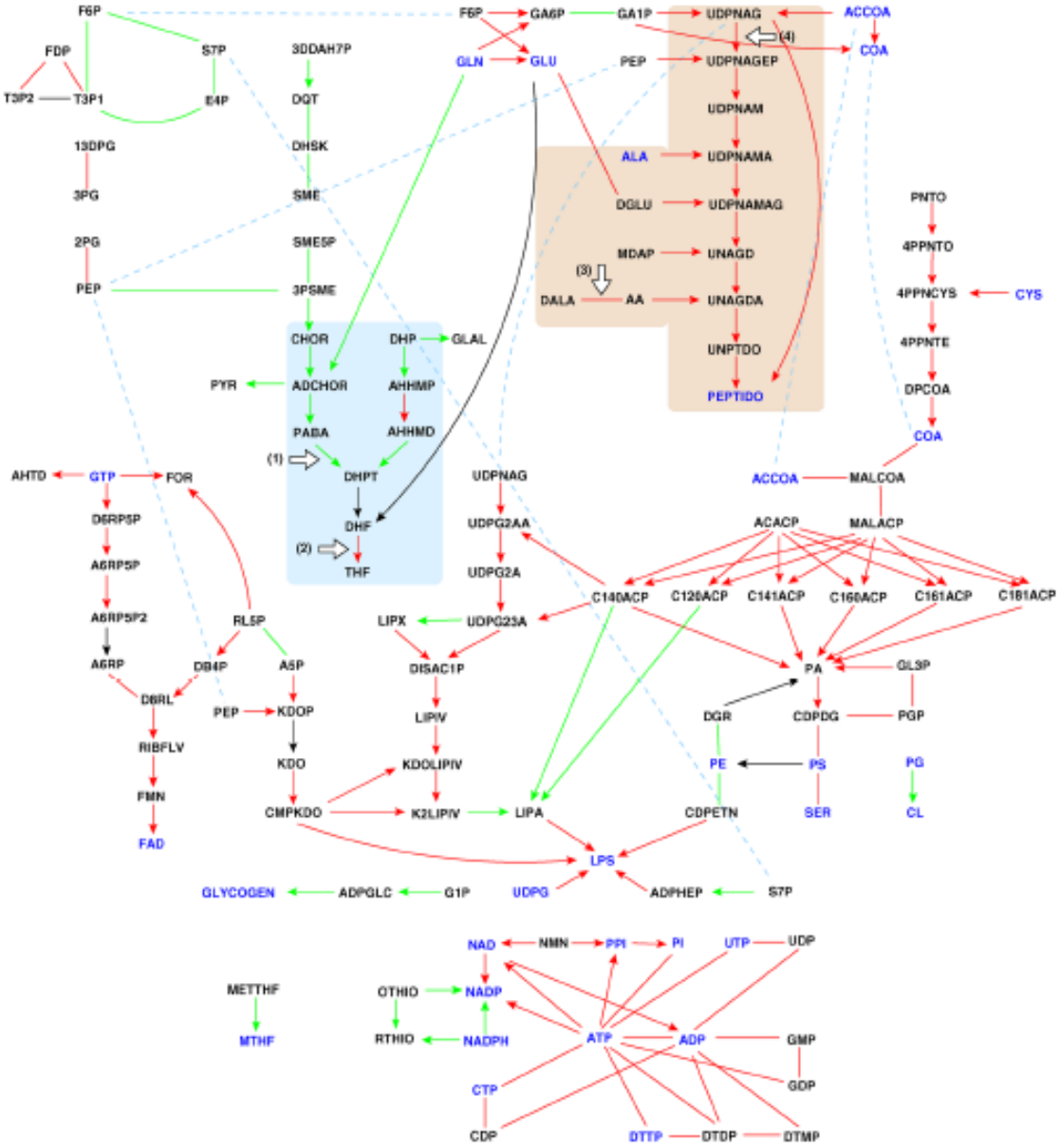
Complex organisms have more flexible metabolic networks.
Fewer reactions are always on.

Metabolic Core of *E.coli:* The constantly active reactions form a tightly connected cluster!

Shown are all reactions that are found to be active in each of the 30,000 investigated external conditions.

Blue: Metabolites that contribute directly to biomass formation,
Red (**green**): core reactions (links) catalyzed by essential (or nonessential) enzymes.
Black-colored links: enzymes with unknown deletion phenotype.

Blue dashed lines: multiple appearances of a metabolite (to simplify the plot), links with arrows: unidirectional reactions.

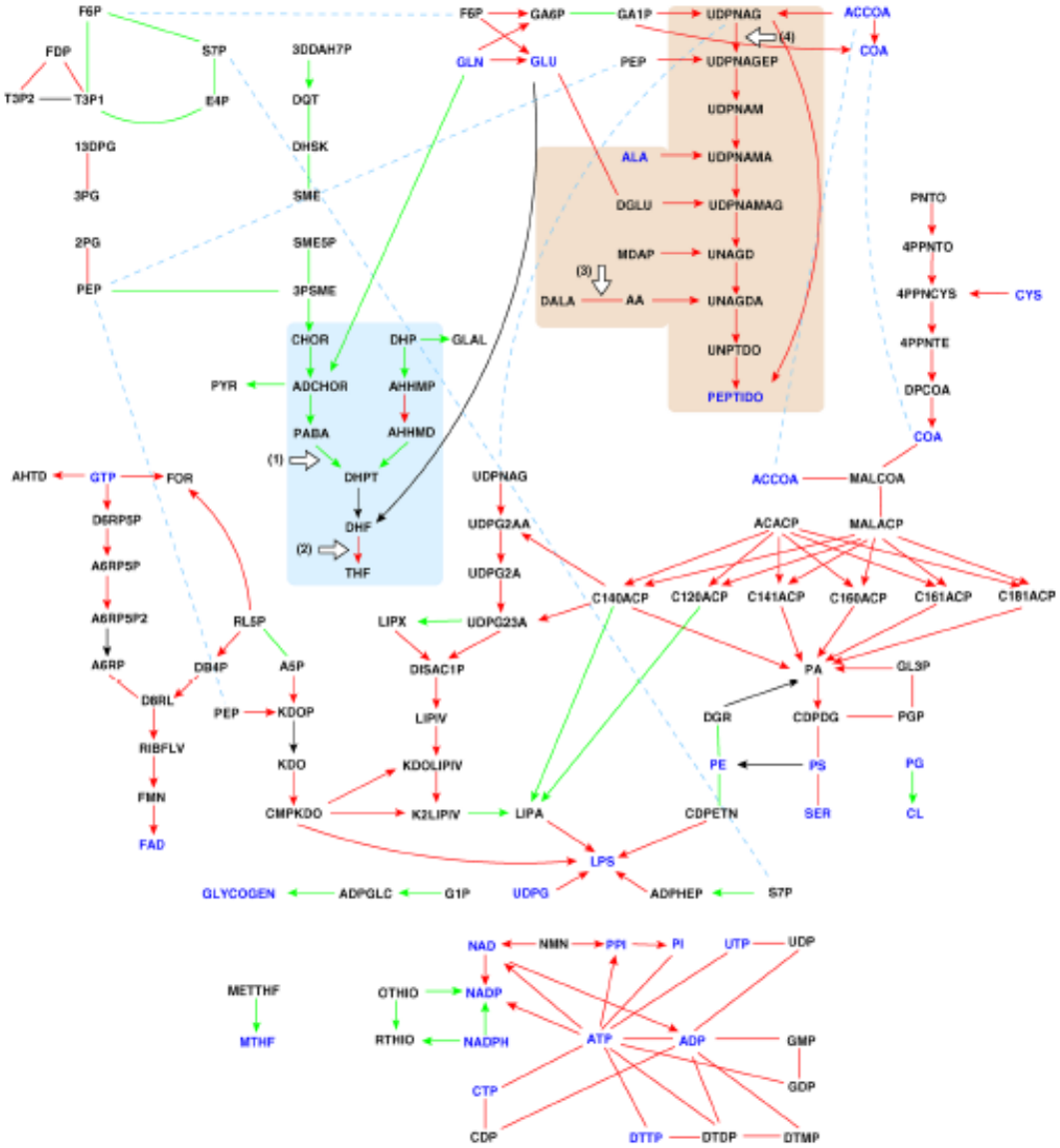


Metabolic Core of *E.coli:* The constantly active reactions form a tightly connected cluster!

20 out of the 51 metabolites necessary for biomass synthesis are not present in the core.
This indicates that they are produced (or consumed) in a growth-condition-specific manner.

Blue and brown shading: folate and peptidoglycan biosynthesis pathways

White numbered arrows denote current **antibiotic targets** inhibited by: (1) sulfonamides, (2) trimethoprim, (3) cycloserine, and (4) fosfomycin.



Metabolic Core Reactions

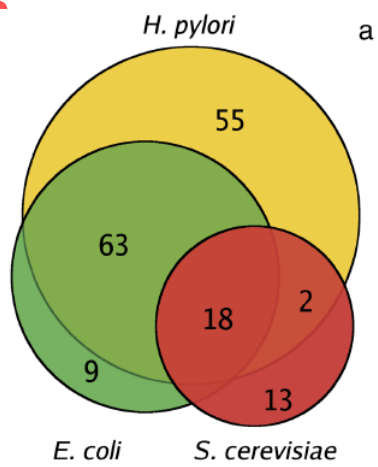
The metabolic cores contain 2 types of reactions:

(a) reactions that are essential for biomass production under all environment conditions (81 of 90 in *E.coli*)

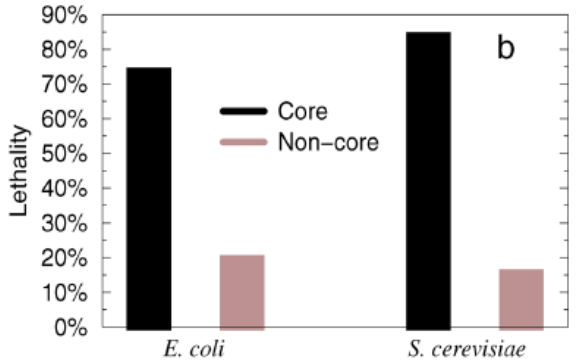
(b) reactions that assure optimal metabolic performance.

Characterizing the Metabolic Cores

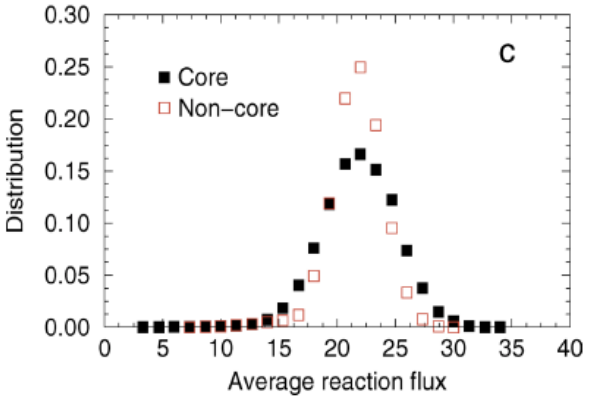
(A) Number of overlapping metabolic reactions in the metabolic core of *H. pylori, E. coli,* and *S. cerevisiae*.
The metabolic cores of simple organisms (*H. pylori* and *E.coli*) overlap to a large extent.
The largest organism (*S.cerevisae*) has a much larger reaction network that allows more flexbility → the relative size of the metabolic core is much lower.



(B) The fraction of metabolic reactions catalyzed by essential enzymes in the cores (black) and outside the core in *E. coli* and *S. cerevisiae*.
→ **Reactions of the metabolic core are mostly essential ones.**



(C) One could assume that the core represents a subset of high-flux reactions. This is apparently not the case.
The distributions of average metabolic fluxes for the core and the noncore reactions in *E. coli* are very similar.



Summary

- Adaptation to environmental conditions occurs via structural plasticity and/or flux plasticity.

Here: a surprisingly stable **metabolic core** of reactions was identified that are tightly connected to eachother.

- the reactions belonging to this core represent **potential targets** for antimicrobial intervention.