A Top-Down Approach

to Metabolic Networks Functioning

Enzo Marinari

(***Sapienza***, **Rome, Italy**)

KITP Santa Barbara, March 2014

Summary.

An introduction to Flux Balance Analysis, FBA

(at genome scale, i.e. at large scale, you better do not need too many details).

Towards Statistical Mechanics: the Von Neumann approach.

Thermodynamic consistency and free energies.

Growth laws (Hwa) constrain allocation of resources when mass balancing: CAFBA

First paper from our group about Von Neumann and MinOver for metabolic networks: Identifying essential genes in Escherichia coli from a metabolic optimization principle, C. Martelli, A. De Martino, E.M., M. Marsili and I. Pérez Castillo, PNAS **106** (2009) 2607.

Thermodynamic consistency and free energies: A Scalable Algorithm to Explore the Gibbs energy Landscape of Genome-scale Metabolic Networks, D. De Martino, M. Figliuzzi, A. De Martino and E.M., PLoS Comp. Bio 8(6): e1002562 (2012);

Counting and correcting thermodynamically unfeasible flux cycles in genome-scale metabolic networks, D. De Martino, F. Capuani, M. Mori, A. De Martino, and E.M., Metabolites **2013** (2013) 946, preprint `arXiv: 1310.3693`.

An application to a very different system: Energy metabolism and glutamate-glutamine cycle in the brain: a stoichiometric modeling perspective, F. A. Massucci, M. Di Nuzzo, F. Giove, B. Maraviglia, I. Perez Castillo, E.M. and A. De Martino, BMC Systems Biology 7 (2013) 103 `arXiv: 1310.6556`.

CAFBA: Constrained Allocation Flux Balance Analysis, M. Mori, A. De Martino, T. Hwa and E.M. in writing.

People in Sapienza@Rome involved:

Andrea De Martino

Fabrizio Capuani

Daniele De Martino

Matteo Figliuzzi (now at Univ. PMC, Paris)

Matteo Mori

Collaboration on CAFBA with Terry Hwa, San Diego, and his group.

Cellular metabolism at genome scale: constraint based models. Minimal assumptions about the steady state.

Viable configuration of reaction fluxes + test for thermodynamic feasibility. This can be computationally very demanding.

Predict chemical activity and response to perturbations of cells without relying on kinetic details (many parameters, imprecise data available).

Specify minimal constraints to describe the reaction network.

1. flux vectors need to satisfy mass balance conditions.
2. the direction of each reaction should guarantee decrease in Gibbs energy (unfeasible cycle could plague the network).

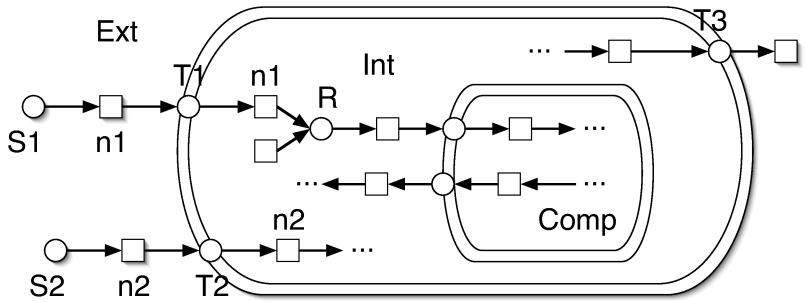
1. Von Neumann: More flexible approach than usual FBA. 2.Fast and scalable algorithm (a relaxation procedure) to reconstruct the Gibbs energy + remove unfeasible reaction cycles.

Problems can arise even when using at best estimates of chemical potentials in physiological conditions.

Information on feasible Gibbs energy ranges: exploit the patterns of reactions interconnections encoded in the stoichiometry to narrow the experimental bounds.

Requesting thermodynamic consistency of the model is also useful for estimating the metabolite concentration.

Find the landscape of Gibbs free energies compatible with a given vector of reaction directions using all stoichiometric information via heuristics inspired by perceptron learning.



Scheme of a cellular metabolic network.

Ext and Int: exterior and the interior of the cell.

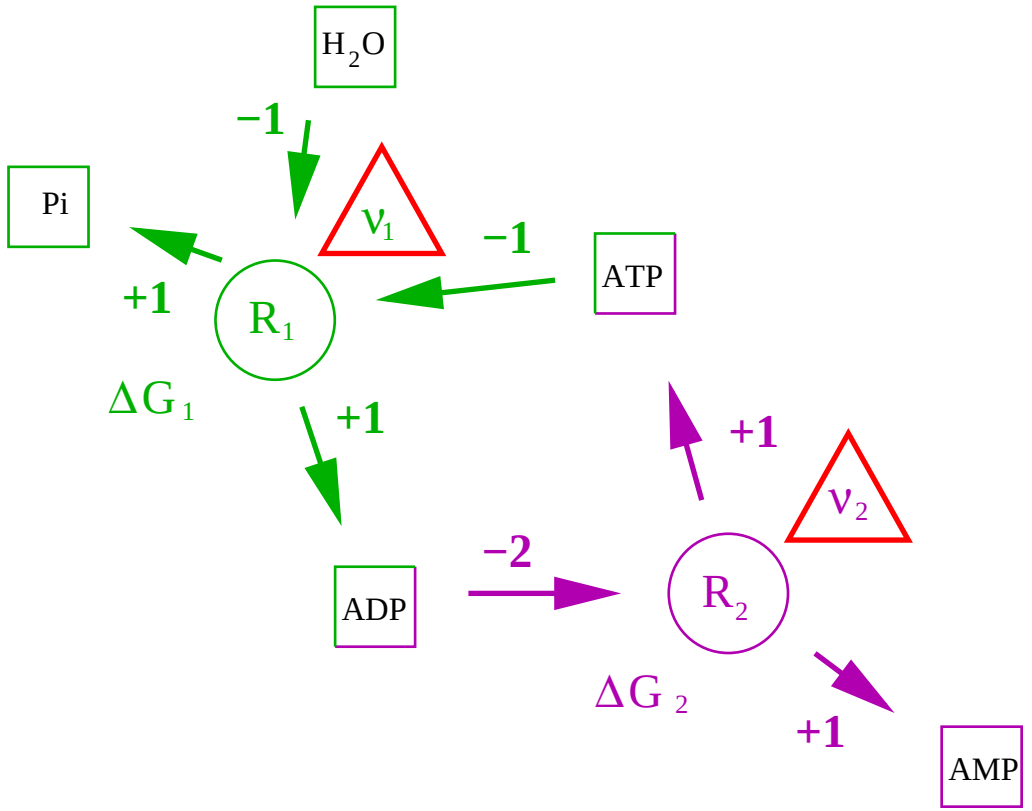
Reactions (resp. metabolites) are denoted by circles (resp. squares).

S1 and S2: fluxes supplying nutrients n1 and n2 to the environment.

T1, T2 and T3: membrane transport reactions by which metabolites are taken in or expelled from the cell.

R: intracellular reaction.

Comp: cellular compartment.



$2\ ADP \longleftrightarrow ATP + AMP$ (a priori reversible)

$ATP + H_2O \longrightarrow ADP + Pi$ (a priori irreversible)

Cell metabolic network: set of interconnected chemical reactions coupled with a set of transport processes.

Bacteria: membrane transport mechanisms by which nutrients are brought into the cell and intracellular reactions by which they are degraded and new biochemical species are produced.

Eukaryotes: also account for the transport of metabolites into and out of each compartment, i.e. for the cell's actual geometric structure.

"Flux analysis":

class of constraint based approaches
in the study of biochemical reaction networks.

Calculation of the reaction flux configurations $\{\nu\}$
compatible
with stoichiometric and thermodynamic constraints.

N reactions, M chemical species

$M \times N$ *stoichiometric matrix* $\boldsymbol{\Xi}$ (sparse matrix)

ξ_i^μ stoichiometric coefficient: species μ in reaction i

Convention. Products: $\xi_i^\mu > 0$. Substrates: $\xi_i^\mu < 0$

Reaction directionality (all in principle reversible, but under physiological conditions some may occur in one direction only). We treat physiologically reversible reactions as two separate processes.

$\boldsymbol{\Xi}$ includes external supply fluxes for the nutrients.

E.coli: $N \simeq 1100$, $M \simeq 700$. *S.cerevisiæ*: $N \simeq 1500$, $M \simeq 900$

$\boldsymbol{\nu} \geq \mathbf{0}$ N-dimensional vector of reaction fluxes

Time evolution of $\mathbf{c}$, metabolite concentrations:

$$\dot{\mathbf{c}} = \boldsymbol{\Xi}\, \boldsymbol{\nu}$$

Kinetics: as matter of principle one could start from the dependence of fluxes from parameters like rate constants k, obtaining

$$\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{c}, \mathbf{k}, \ldots)$$

and solving the dynamical system for the concentrations: this is impossible for large-scale networks.

Standard modeling routes: $\dot{\mathbf{c}} = \mathbf{0}$ (timescale separation between chemical processes and genetic regulation).

$\boldsymbol{\Xi\nu} = \mathbf{0}$: Kirchhoff-type mass-balance conditions.

$N > M \Rightarrow$ set of solutions has dimension $N - \mathrm{rank}(\boldsymbol{\Xi})$: equivalent feasible flux states of the network. Uniform sampling is impossible.

Relevant configurations: maximize a objective function $\boldsymbol{\alpha} \cdot \boldsymbol{\nu}$.

$$\max_{\mathbf{0} \leq \boldsymbol{\nu} \leq \boldsymbol{\nu}_{max}} (\boldsymbol{\alpha} \cdot \boldsymbol{\nu}) \quad \text{subject to} \quad \boldsymbol{\Xi\nu} = \mathbf{0}$$

Select $\boldsymbol{\nu}_{max}$ on experimental input.

This is Flux Balance Analysis, FBA.

The vector $\boldsymbol{\alpha}$ contains the crucial biological assumptions, as much as possible from experimental evidence.

One possibility is to maximize biomass: a combination of different metabolites in precise stoichiometric proportions.

Other possibilities: maximize the total flux of all ATP-producing reactions, or minimize glucose consumption (a proxy for efficient nutrient usage) or the total flux of intracellular reactions (for maximal enzymatic efficiency).

All these approaches are highly correlated.

Very interesting: extension to biologically complex situations. Gene knock-outs that prevent the execution of certain reactions.

The Von Neumann approach and producibility.

Link between the network's structure and its productive capabilities.

A metabolite μ is *producible* from a given set of nutrients if

$$\exists\ \boldsymbol{\nu} \geq \mathbf{0} \text{ such that } \ \boldsymbol{\Xi\nu} \geq 0 \ \& \ \left[\boldsymbol{\Xi\nu}\right]_\mu > 0$$

(where $\left[\boldsymbol{\Xi\nu}\right]_\mu = \sum_{i=1}^{N} \nu_i \xi_i^\mu$). At least one flux vector exists allowing for a net production of μ irrespective of whether other metabolites are being also produced.

Nutrient usage may never exceed its supply.

Concentrations of producible metabolites can increase in a stationary flux state with the sole consumption of the nutrients, so that the cell is allowed to employ them for purposes other than metabolic (e.g. to form proteins, membranes, etc.).

Possibility to produce a metabolite: limited by conservation laws from the stoichiometry.

Stoichiometric matrix: separate **B** (input coefficients) from **A** (output coefficients): $\mathbf{A} - \mathbf{B} = \mathbf{\Xi}$.

Conserved pool G of metabolites:

$$\sum_{\mu \in G} (a_i^\mu - b_i^\mu) = 0 \qquad \forall i = 1, \cdots N.$$

Famili and Palsson, Biophys. J. **85** (2003) 16

Conserved moieties are abundant in real metabolic networks.

Robustness of the cellular production profile emerging in given nutrient conditions. Study:

$$\mathcal{V} = \{\boldsymbol{\nu} \text{ such that } \boldsymbol{\Xi\nu} \geq 0\} \quad \left(\sum_i \nu_i = N\right) .$$

In this case we can apply Von Neumann approach (originally devised to study economics), and even statistical sampling is feasible. A. De Martino and M. Marsili, J. Stat. Mech. (2005) L09003.

Given a constant $\rho > 0$, a flux vector $\boldsymbol{\nu}$ such that $\mathbf{A}\boldsymbol{\nu} \geq \rho\, \mathbf{B}\boldsymbol{\nu}$ describes a network state in which every species is being produced at a rate at least equal to ρ.

What is $\rho^\star$, the maximum value of ρ for which flux vectors satisfying

$$(\mathbf{A} - \rho\, \mathbf{B})\, \boldsymbol{\nu} \geq \mathbf{0}$$

exist?

$\rho^\star$: optimal productive performance.

$\rho^\star > 1$ optimal states expand. $\rho^\star < 1$ optimal states contract.

Flux vectors $\{\boldsymbol{\nu}\}$ satisfying the constraints at $\rho^\star$ are the optimal flux states of the system.

When A and B are random matrices, statistical mechanics, replica trick and the cavity approach can be applied.

$n \equiv N/M$. A critical value n_c separates contracting from expanding regimes.

Enhanced dilution increases $\rho^\star$ in the expanding phase.
Enhanced dilution decreases $\rho^\star$ in the contracting phase.

A single flux vector is optimal when $\rho = \rho^\star$.

In real biological systems, where **A** and **B** are real stoichiometric coefficients, the situation is radically different.

The existence of conserved metabolic pools, enforces the condition

$$\rho_\star = 1 \ .$$

Here at $\rho_\star$ one finds a number of different solutions.

The MinOver algorithm

W. Krauth and M. Mézard, "Learning algorithms with optimal stability in neural networks" J. Phys. A: Math. Gen. 20 (1987) L745

Start by setting all fluxes equal to random values or to zero.

REPEAT: Calculate the M functions $c^\mu = \sum_{i=1,N} \left(a_i^\mu - \rho\, b_i^\mu\right)\, \boldsymbol{\nu}_i$

and determine $\mu_0 = \arg\, \left(\min_\mu c^\mu\right)$ i.e. the index of the metabolite with the smallest value of c (if degenerate get one at random).

If $c^{\mu_0} \geq 0$ (but not all c are zero) we have a solution $\boldsymbol{\nu}$ for the fluxes at growth rate ρ.

Else update the fluxes:

$$\nu_i \to \nu_i' = \max\left(0, \nu_i + a_i^{\mu_0} - \rho b_i^{\mu_0}\right)$$

and repeat from REPEAT.

Application to the bacterium E. coli. Very good reconstruction of its metabolic network exists (J. L. Reed, T. D. Vo, C. H. Schilling and B. O. Palsson, Genome Biology **4** (2003) R54).

Main findings: see next slides (I start from the end...).

Ranges of variability of the fluxes in specified extracellular conditions agree well with the (limited) experimental data available.

Dynamically stiff variables (reactions with small ranges) correspond to *E.coli*'s phenomenologically essential genes: genes that are both necessary for the organism's survival and highly conserved across different bacterial species.

At least in some conditions metabolic networks may operate close to their optimal productive capacity.

Selection of the environment where the cell lives is an important step.

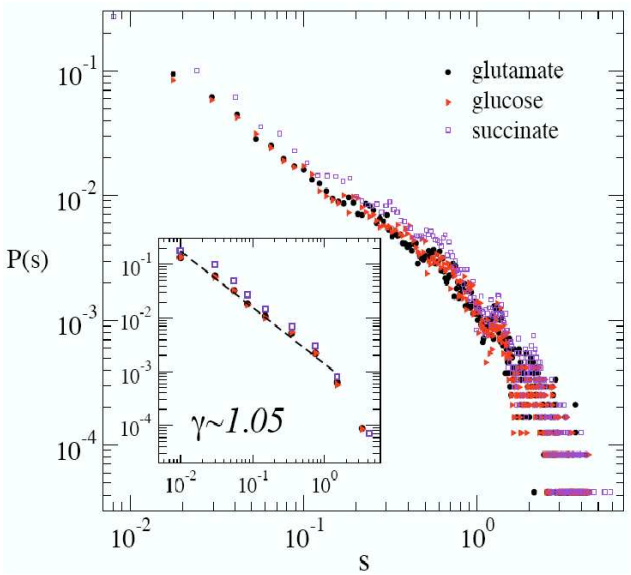
Select a set of external metabolites where uptake fluxes are applied.

Three distinct environmental conditions:

(i) isolated cell, without uptakes;
(ii) minimal environment, with uptakes on a restricted set of metabolites;
(iii) rich environment, with uptakes on all external metabolites.

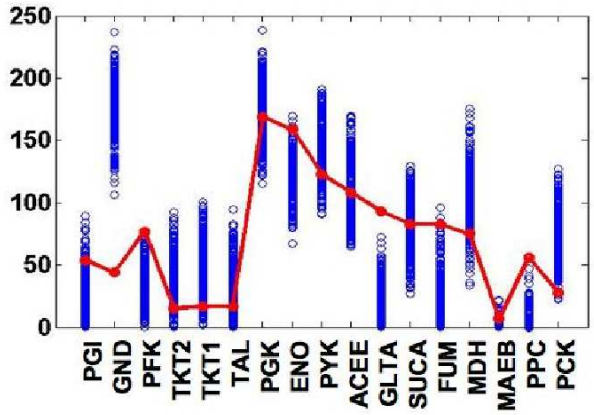
Calculate $\rho^\star$ applying MinOver at fixed ρ and then gradually increasing it.

$\rho^\star = 0.999 \pm 0.001$ independently of the environmental conditions: the state of optimal growth has constant fluxes.



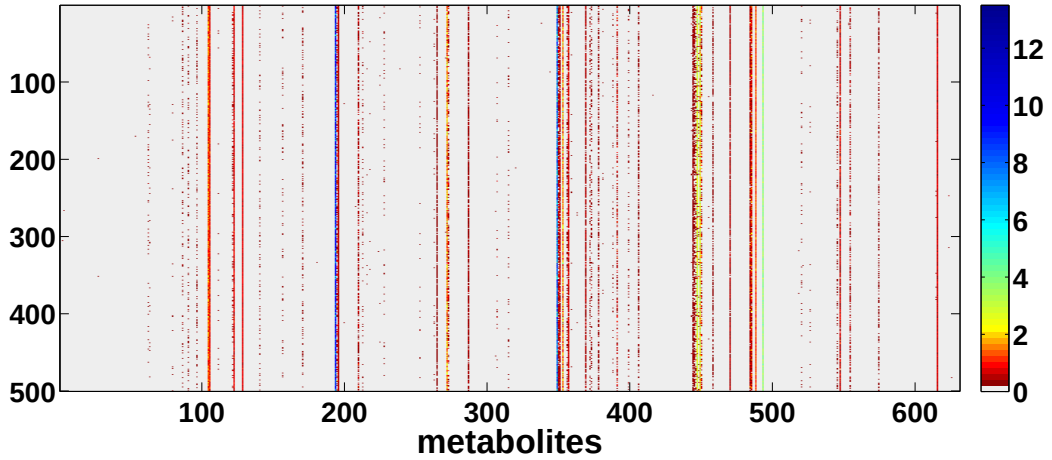
Flux distributions at $\rho^\star$ in a minimal environment.

On close to two decades: scaling form $P(s) \sim s^{-\gamma}$ with exponent close to -1 followed by a cutoff. Ok with experiments. The flux histogram is very stable in different solutions.



Comparison of reaction fluxes predicted by the Von Neumann scheme with 17 fluxes measured in [Emmerling M *et al.*, J. Bacteriol. **184** (2002) 152] and analyzed in [Segrè D, Vitkup D, and Church GM Proc Natl Acad Sci USA **99** (2002) 15112]. The different reactions (in no specific order) are on the horizontal axis, their corresponding fluxes (relative to the glucose uptake) on the vertical axis. Red markers are for experimental values, blue ones theoretical predictions. Data points are for 300 solutions. Simplified medium, not coinciding with the experimental situation.

$\rho^\star = 1$. In each Von Neumann solution, as opposed to FBA, some metabolites exist with a non-zero c^μ.



c^μ for different metabolites in 500 different solutions. White: $c^\mu = 0$. Colored marks: net production of the metabolite in that solution. Mass balance holds for most metabolites, but some are consistently unbalanced while for some it depends on the solution. This may also signal incompleteness of the stoichiometric data.

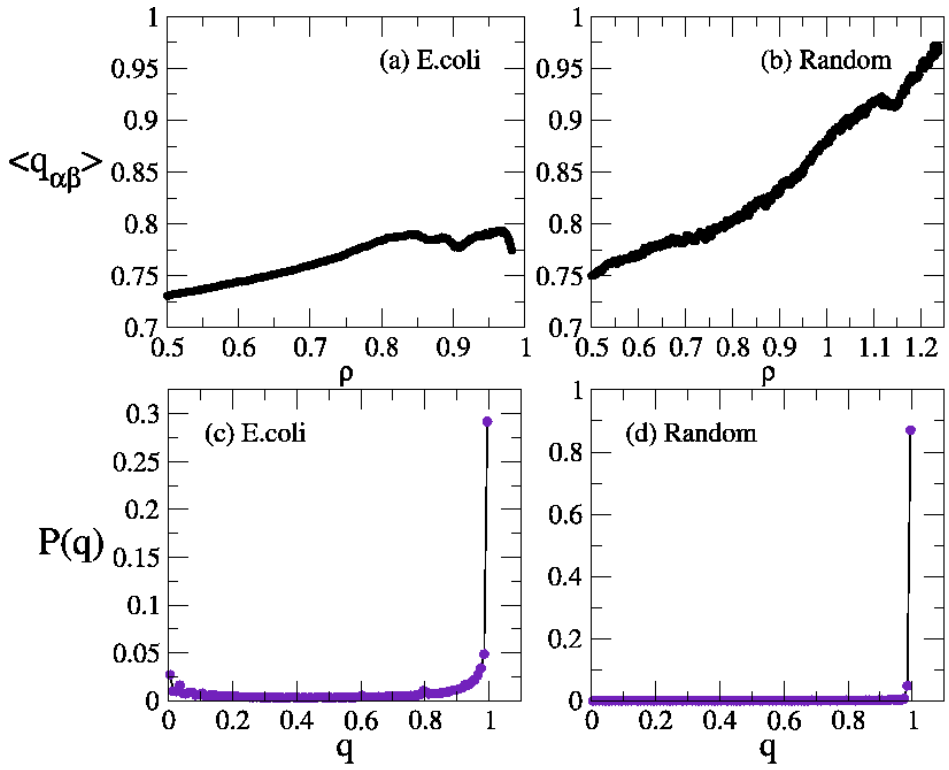
To get further insight on flux states at $\rho^\star$ and on the shape of the solution space we monitor the mean overlap between different solutions.

Given two solutions α and β at fixed ρ, we define their overlap $q_{\alpha\beta}$ as:

$$q_{\alpha\beta} = \frac{2}{N}\sum_{i=1}^{N}\frac{s_{i\alpha}s_{i\beta}}{s_{i\alpha}^2 + s_{i\beta}^2} = \frac{1}{N}\sum_{i=1}^{N} q_{\alpha\beta}^{(i)} \ .$$

$q_{\alpha\beta} = 1$ when $\mathbf{s}_\alpha$ and $\mathbf{s}_\beta$ coincide, and becomes smaller when their difference increase.

Technicalities: a flux is taken to be zero whenever its value is below a threshold ϵ, to take into account the fact that there is a loss of information about relative fluctuations between different solutions in small fluxes (we take $\epsilon = 10^{-5}$). Also the overlap between null fluxes must be defined by consistency to be equal to one.



(a,b) Mean overlap between 500 different solutions in *E. coli* and in a random metabolic network; the last point on the abscissa is $\rho^\star$.

(c,d) Overlap histogram $P(q)$ at $\rho^\star$ in E.coli and in the random network. Note the different scales of the y axes in the lower panels.

Artificial metabolic networks: $\langle q_{\alpha\beta} \rangle \to 1$ as $\rho \to \rho^\star$.

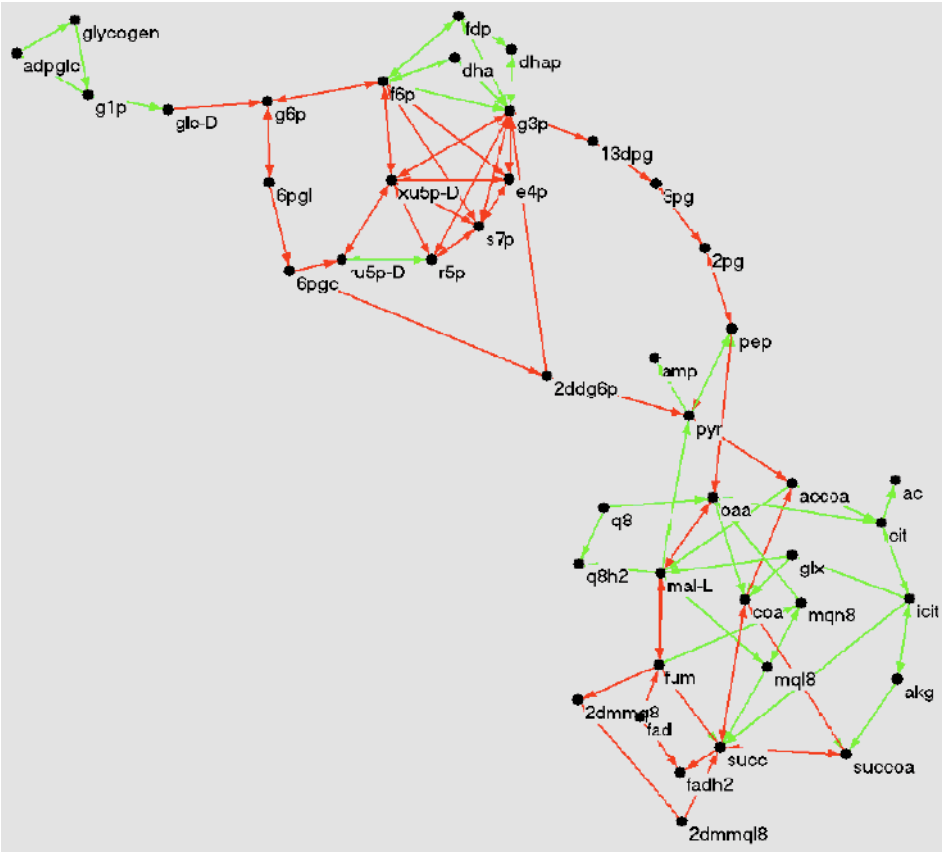
Overlap histogram: δ-peak at $q = 1$ (mass increases as $\rho \to \rho^\star$).
Solution at $\rho^\star$ is unique.

E. coli: the volume of solutions stops contracting when ρ reaches roughly 0.8 from below. At $\rho^\star$ multiple solutions survive.

Histogram: only about the 30% of reactions have an overlap close to 1. In other terms roughly 30% of the variables are *frozen* (i.e. assume the same value on all solutions of the constrained optimization problem), while the remaining are *free*.

Expect that for purely structural reasons reaction chains are entirely frozen when the first reaction of the chain is.

Confirmed by the map of frozen/free fluxes in *E. coli*'s central metabolism, next slide.



E. coli's central metabolism. Nodes: metabolites. Arrow joining two nodes: reaction converting one into the other. Red (green) links are for frozen (free) reactions, with overlap larger (smaller) than 0.9.

Biological significance of frozen fluxes?

Notion of essentiality: combines phenomenological relevance (a gene is essential if knocking it out the cell dies) with evolutionary retention (the presence of the gene in different species).

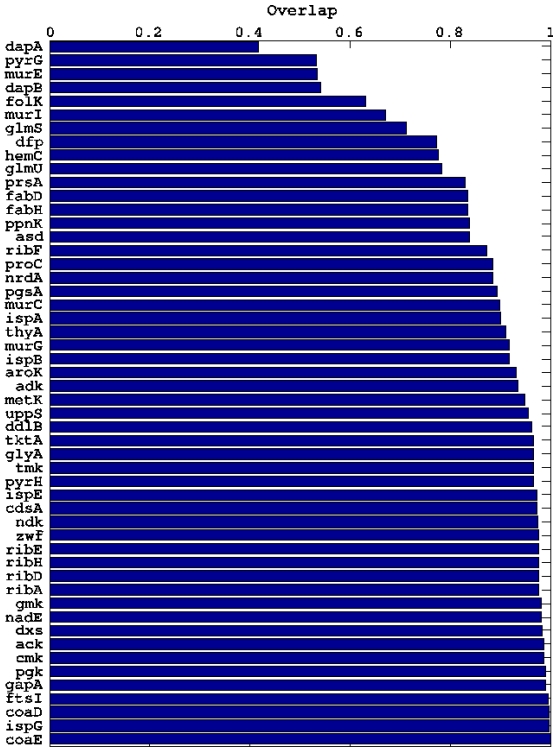
We correlate $q_{\alpha\beta}^{(i)}$ with the essentiality of the corresponding genes

Gerdes et al., J. Bacteriol. 185 (2003) 5673: 55 essential genes of *E. coli* involved in metabolism (also present in 80% of 32 different bacterial genomes).

We have linked 52 of them to reactions in the reconstructed network.

Next figure: 43 of these genes correspond to reactions with overlap larger than 0.8 (only 7 genes relate to reactions with an overlap significantly smaller than 80%).

Frozen fluxes may carry an evolutionary significance.



Essential genes (vertical axis) versus overlap of the corresponding reactions in the reconstructed metabolic network of *E. coli* (horizontal axis).

Now thermodynamical consistency.

$$G = E - P\,V - T\,S$$

does not increase spontaneously in an open system at constant T, P.

Let $u_i = \text{sign}\{\nu_i\} = \pm 1$, ν_i flux of reaction i ($+1$ is for a forward reaction, -1 for a backward one). Gibbs energy change ΔG_i induced by the reaction must be such that $u_i \Delta G_i \leq 0$.

$S_{\alpha i}$: stoichiometric coefficient of metabolite α in reaction i: $S_{\alpha i} > 0$ product, $S_{\alpha i} < 0$ substrate.

$\mu = \{\mu_\alpha\}$. μ_α Gibbs energy per mole of species α:

$$\Delta G_i = \left(S^T\right)\ \mu_\alpha$$

$$X_i \equiv -u_i \sum_{\alpha=1}^{M} S_{\alpha i}\ \mu_\alpha \geq 0\ \ \forall i$$

We look for $\boldsymbol{\mu}$.

For fixed u_i solution space for $\boldsymbol{\mu}$ is convex.

Relaxation methods: iterations where variables are updated and violated inequalities get fixed.

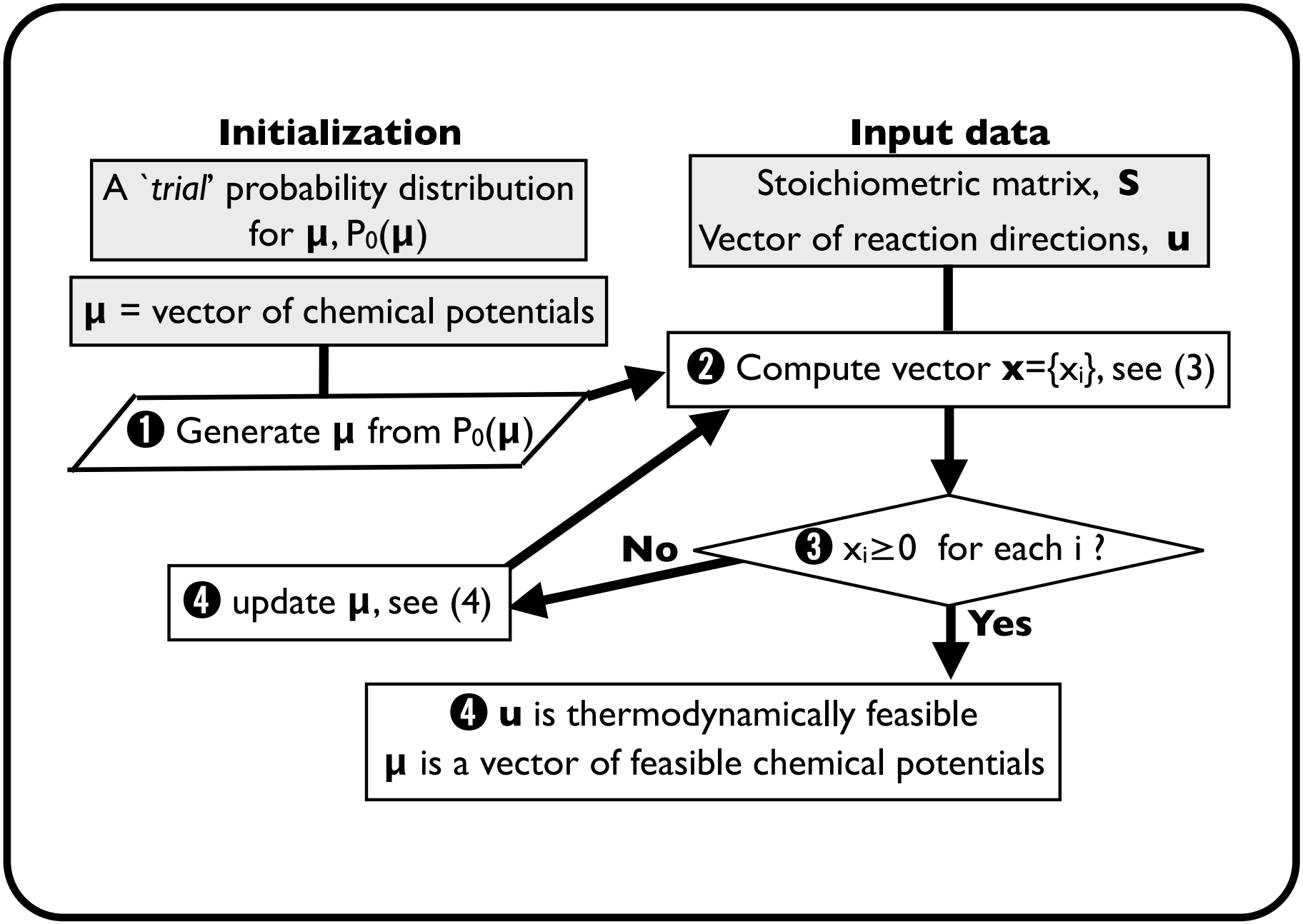
MinOver (Krauth Mézard 1987 for perceptron learning) is also used for computing fluxes in Von Neumann approach.

Start from trial probability distribution

$$P_0(\boldsymbol{\mu}) = \prod_{\alpha=0}^{M} P(\mu_\alpha)$$

P_0^α uniform over a given μ range.

P_0^α contains prior biochemical information (i.e. centered around experimental value with width connected to experimental errors and such to span a few orders of magnitude).


Initialization
A `trial' probability distribution for μ, P0(μ)
μ = vector of chemical potentials
❶ Generate μ from P0(μ)
Input data
Stoichiometric matrix, S
Vector of reaction directions, u
❷ Compute vector x={xi}, see (3)
❸ xi≥0 for each i ?
No
Yes
❹ update μ, see (4)
❹ u is thermodynamically feasible
μ is a vector of feasible chemical potentials

The algorithm is based on the following steps:

1. Generate a chemical potential vector $\boldsymbol{\mu} = \{\mu_\alpha\}$ from $P_0(\boldsymbol{\mu})$.
2. Compute $\mathbf{x} = \{x_i\}$ and $i_0 = \arg\ \min_i x_i$ (i.e., i_0 is the index of the least satisfied constraint).
3. If $x_{i_0} \geq 0$ then $\boldsymbol{\mu}$ is a thermodynamically consistent chemical potential vector for $\mathbf{u}$; exit (or go to 1 to obtain a different solution).
4. If $x_{i_0} < 0$, update $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} \ \rightarrow \ \boldsymbol{\mu} - \lambda u_{i_0} \mathbf{S}_{i_0}$$

(where $\lambda > 0$ is a constant and $\mathbf{S}_j$ is the j-th column of matrix $\boldsymbol{S}$), go to 2 and iterate.

The $-\lambda u_{i_0} \boldsymbol{S}_{i_0}$ drives the adjustment of chemical potentials: at every iteration the least satisfied constraint gets improved.

If a solution exists, convergence to a solution is guaranteed $\forall \lambda > 0$ (and the time of convergence depends on λ).

We can prove it: suppose that a solution $\boldsymbol{\mu}^*$ exists

$$-u_i(\boldsymbol{S}_i \cdot \boldsymbol{\mu}^\star) \geq c \qquad \forall i \ ,$$

with $c > 0$ constant. We have that after ℓ steps

$$\begin{aligned} \boldsymbol{\mu}(\ell) \cdot \boldsymbol{\mu}^\star &= \boldsymbol{\mu}(\ell-1) \cdot \boldsymbol{\mu}^\star - \lambda s_{i_0(\ell-1)} (\boldsymbol{S}_{i_0(\ell-1)} \cdot \boldsymbol{\mu}^\star) \\ &\geq \boldsymbol{\mu}(\ell-1) \cdot \boldsymbol{\mu}^\star + \lambda c \\ &\geq \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}^\star + \ell \lambda c \ , \end{aligned}$$

and

$$\boldsymbol{\mu}(\ell) \cdot \boldsymbol{\mu}(\ell) \leq \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(0) + \ell \lambda^2 A \, ,$$

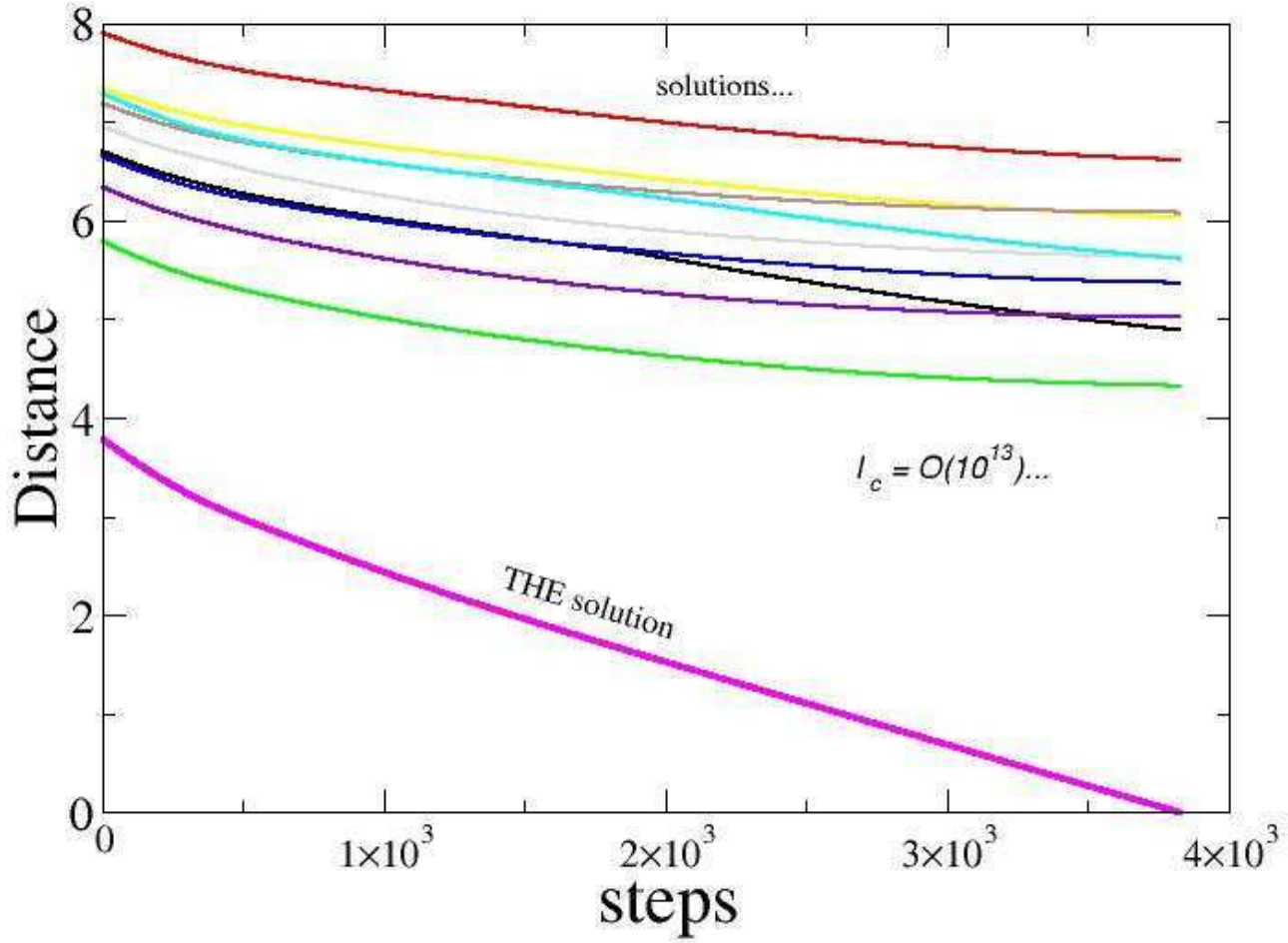
with

$$A = \max_i \sum_\alpha (S_{\alpha,i})^2 \, .$$

This implies

$$d(\ell) \equiv \frac{\boldsymbol{\mu}(\ell) \cdot \boldsymbol{\mu}^\star}{|\boldsymbol{\mu}(\ell)||\boldsymbol{\mu}^\star|} \geq \frac{\boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}^\star + \ell \lambda c}{|\boldsymbol{\mu}^\star| \sqrt{|\boldsymbol{\mu}(0)|^2 + \ell \lambda^2 A}} \, .$$

By Cauchy-Schwarz $d(\ell) \leq 1$: imposing $d(\ell_c) = 1$ one finds an upper bound on the number of steps needed to converge.


solutions...
$l_c = O(10^{13})$...
THE solution
Distance
steps
0
2
4
6
8
0
1×10^3
2×10^3
3×10^3
4×10^3

Starting from a random vector if a solution exists we find it. Starting from different random vectors allows sampling the solutions (if there are many).

Many random starts: obtain a set of solutions of correlated $\{\mu_\alpha\}$.

The reinforcement steps build up the correlations.

Nature of solution space: unbounded cone passing through the origin. To obtain boundness one can use different approaches:

1. can clamp some μ_α keeping them fixed;
2. can assign a fixed range o variability to some or all μ_α :

$$\mu_\alpha^{\min} < \mu_\alpha < \mu_\alpha^{\max} ;$$

3. add global constraint (for example fix the potential for external metabolites in uptakes).

Important questions:

1. Minimal amount of a priori information needed to bound the solution space?

Here: we look at solutions close to prior biochemical information.

Our approach works "better" than relaxation and penalty methods even under noisy and inconsistent biochemical priors.

Identify and remove loops: an extension of the algorithm

A generic assignment of reaction directions can be unfeasible, i.e. such that $-u_i \sum_{\alpha=1}^{M} S_{\alpha i}\, \mu_\alpha \geq 0 \;\; \forall i$ does not have any non-trivial solution.

Farkas-Minkowski theorem: this happens if and only if there is at least one unfeasible loop, i.e. it $\exists$ a set $\mathcal{L}$ of reactions for which $\exists \{k_i > 0\}$ constant such that

$$\sum_{i \in \mathcal{L}} k_i u_i S_{\alpha,i} = 0 \qquad \forall \alpha \; .$$

In this case MinOver does not converge: the least satisfied constraint cycles along the loop. There is a problem in the network reconstruction (work in progress in analyzing a statistical theory that helps solving this problem: S. Colabrese, A. De Martino, D. De Martino, E.M.).

A simple way to correct an unfeasible set of reaction directions:

1. run MinOver for (large) number of iteration steps T. Keep track of the last K unsatisfied constraints, with K large;
2. select the reactions that appear more frequently in this set and search for a loop among these reactions;
3. if you find a loop, change the direction of one of its reactions.

All codes available from `http://chimera.roma1.infn.it/SYSBIO/`

The red blood cell metabolic network

Flux configurations from MC sampling of FBA solutions (Price, Schellenberger and Palsson 2004) and from Von Neumann (A. De Martino, Granata, EM, Martelli, Van Kerrebroeck 2010).

They are similar:

in FBA all reactions are bidirectional, all others are forward;

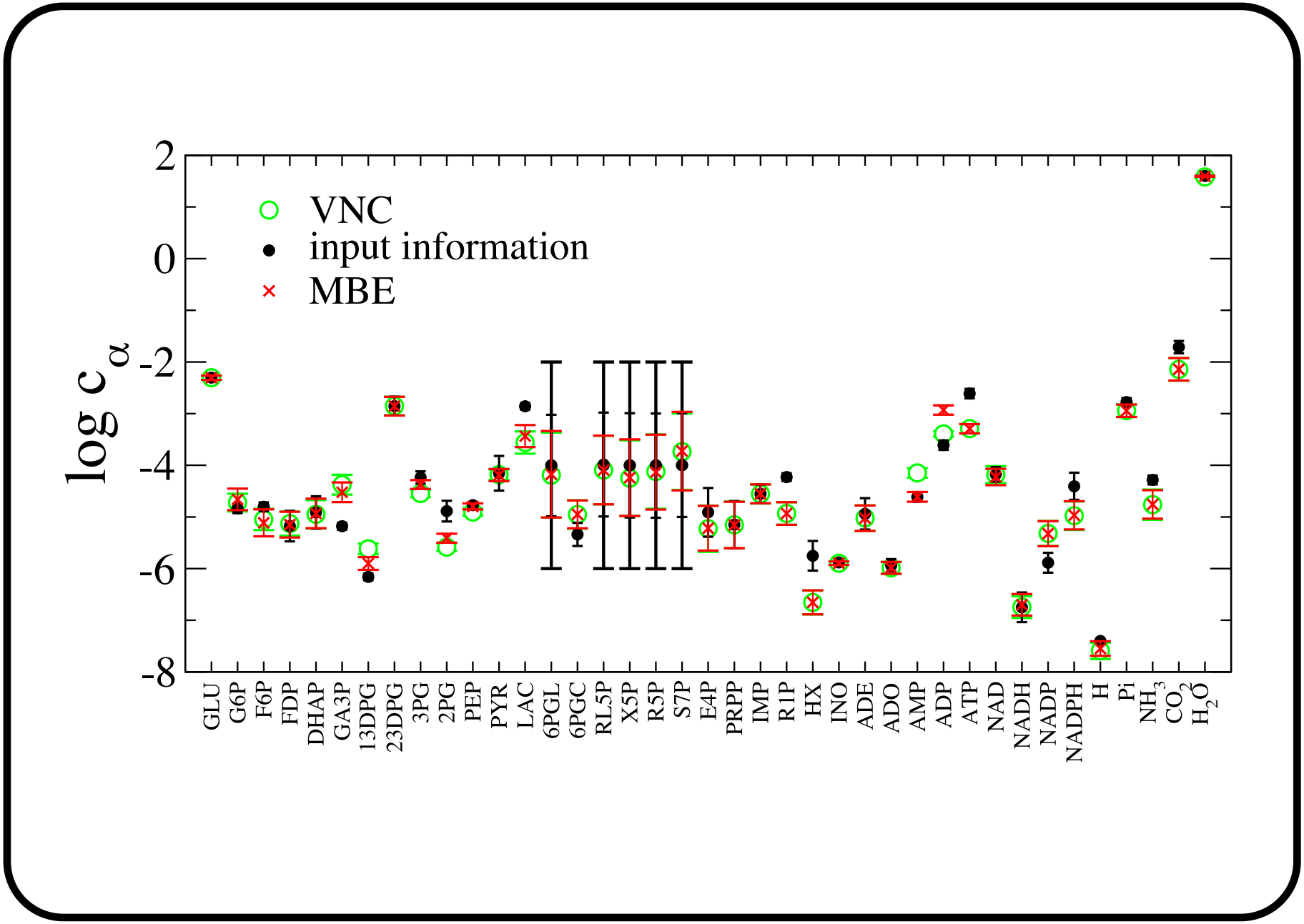
in VN one reaction is bidirectional (as in FBA) and two (that are bidirectional in FBA) are backward, all others are forward.

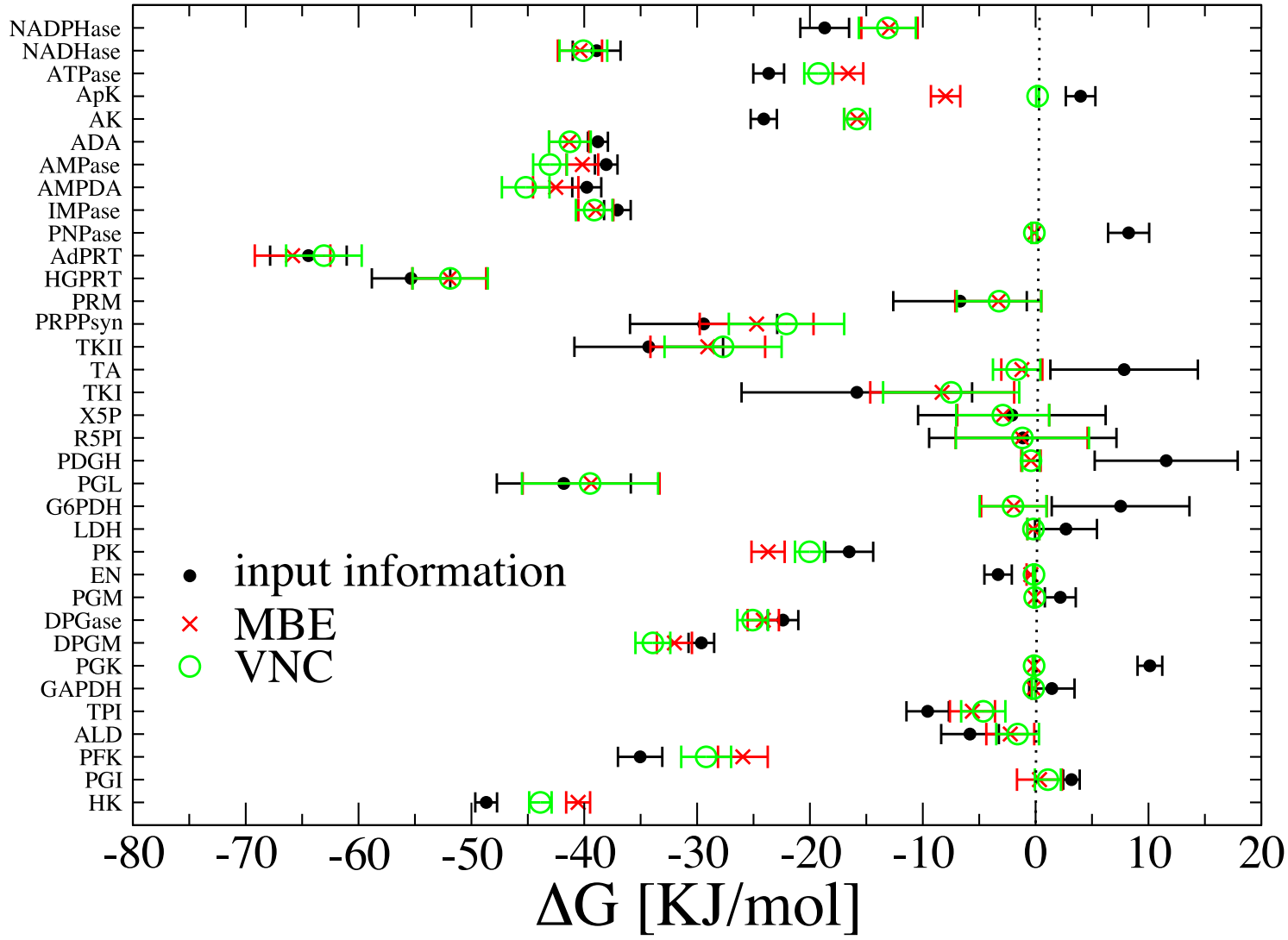
Both sets of solutions turn out to be thermodynamically feasible.

Concentrations are computed from (dilute solution approximation)

$$\log C_\alpha = \frac{\mu_\alpha - \mu_\alpha^{(0)}}{RT}$$

Water is clamped.


VNC
input information
MBE
log c α
2
0
-2
-4
-6
-8
GLU
G6P
F6P
FDP
DHAP
GA3P
13DPG
23DPG
3PG
2PG
PEP
PYR
LAC
6PGL
6PGC
RL5P
X5P
R5P
S7P
E4P
PRPP
IMP
R1P
HX
INO
ADE
ADO
AMP
ADP
ATP
NAD
NADH
NADP
NADPH
H
Pi
NH3
CO2
H2O


NADPHase
NADHase
ATPase
ApK
AK
ADA
AMPase
AMPDA
IMPase
PNPase
AdPRT
HGPRT
PRM
PRPPsyn
TKII
TA
TKI
X5P
R5PI
PDGH
PGL
G6PDH
LDH
PK
EN
PGM
DPGase
DPGM
PGK
GAPDH
TPI
ALD
PFK
PGI
HK
input information
MBE
VNC
-80
-70
-60
-50
-40
-30
-20
-10
0
10
20
ΔG [KJ/mol]

On growth laws. See for example Scott et al., Science 330, 2010 (Hwa group).

In bacteria cells under steady state exponential growth the rate of cell proliferation and the level of gene expression are intimately intertwined.

The existence of intrinsic constraints governing the allocation of resources towards protein synthesis has been recently established.

Growth rate affects the expression of individual genes which in turn affects the RNA over protein ratio r.

One needs to understand the cell allocation to ribosome synthesis (where most of the RNA belongs: for example in E. coli close to 85%), i.e. the value of r at different growth rates.

For exponentially growing E. coli cell the RNA/protein ratio r is linearly correlated with the growth rate λ

$$r(\lambda) = r_0 + \frac{\lambda}{\mathbf{k}_t} .$$

Valid for fast to moderately slow growth.

Universal: valid for other microbes.

Linear correlation is expected if the ribosomes are growth limiting; translation at a constant rate.

$\mathbf{k}_t$ is proportional to the rate of protein synthesis. Proportional to the translational capacity of the organism.

Scott et al., Science 2010

Supplementary Figures

Figure S1

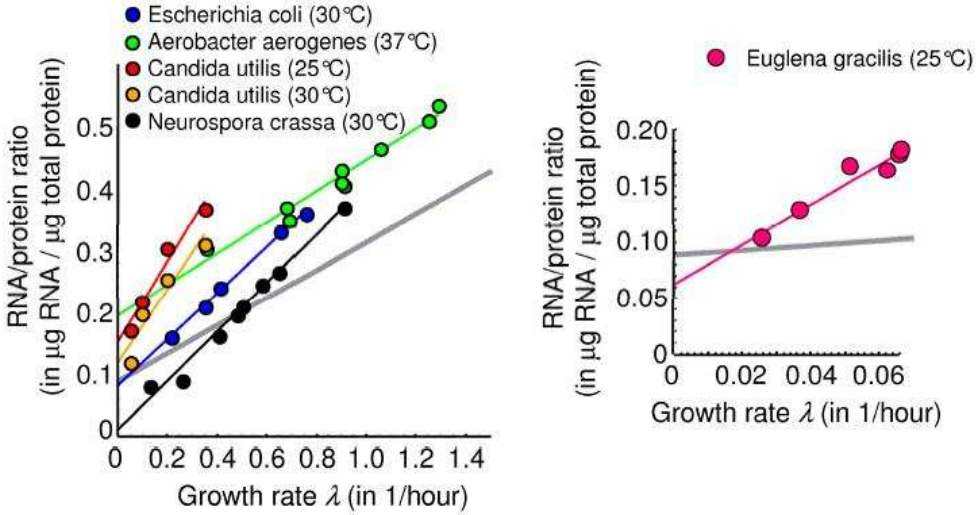


Fig. S1: Linear correlation between RNA/protein ratio and growth rate in various microbes. The linear relation between the RNA/protein ratio and the growth rate is evident in a number of bacteria studied, and in exponentially growing unicellular eukaryotes. *Left*: *Escherichia coli* (blue, 30°C; Ref. (*36*)), *Aerobacter aerogenes* (green, 37°C; Ref. (*37*)), *Candida utilis* (red, 25°C and orange, 30°C; Ref. (*38*)), *Neurospora crassa* (black, 30°C; Ref. (*39*)). *Right*: *Euglena gracilis* (magenta, 25°C; Ref. (*40*)).

For comparison, the grey line corresponds to the solid line drawn in Fig. 1A for various *E. coli* strains grown at 37°C.

One can gradually inhibit translation by exposing cells to translation-inhibiting antibiotics.

A new correlation among r and λ emerges:

$$r = r_{\max} - \frac{\lambda}{\mathbf{k}_n}$$

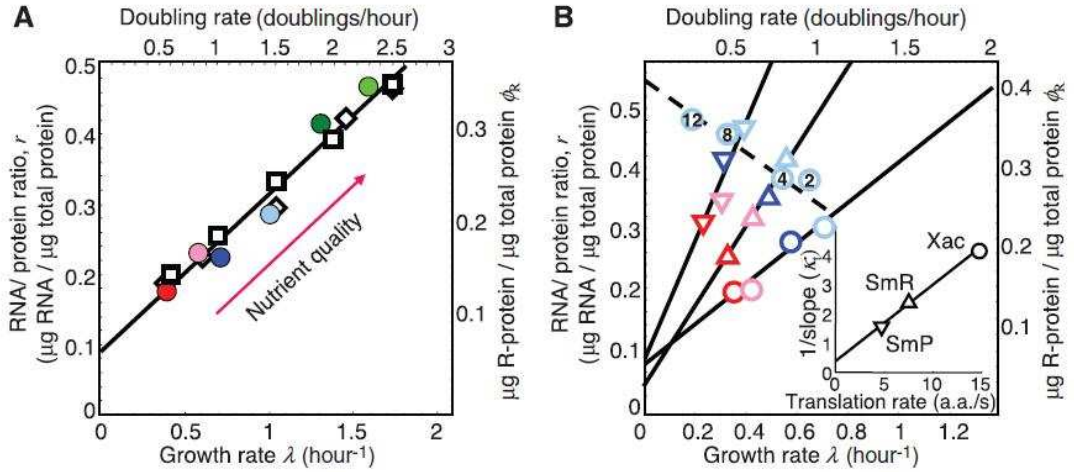
(see figure).

Increase of r with increasing degree of translational inhibition, for a given nutrient source.

Valid for all nutrient sources.

$\mathbf{k}_n$ has a strong positive correlation with the growth rate of cells in drug free medium: nutritional capacity.

Scott et al., Science 2010 (see dashed line)



Strain / Medium	EQ2	Xac	SmR	SmP	Xac in cAA+glc Cm conc. (µM)	
M63+glyc	●	○	△	▽	2	(2)
M63+gluc	●	○	△	▽	4	(4)
cAA+glyc	●	○	△	▽	8	(8)
cAA+gluc	●	○	△	▽	12	(12)
RDM+glyc	●	Historical data: □ Strain B/r; Ref. (10) ◇ Strain 15τ-bar; Ref. (12)				
RDM+gluc	●					

Fig. 1. Correlation of the RNA/protein ratio r with growth rate λ for various strains of *E. coli*. (**A**) Comparison among *E. coli* strains grown in minimal medium: Strain B/r [(*10*), squares], 15τ-bar [(*12*), diamonds], and EQ2 (this work, solid circles). The growth rate is modulated by changing the quality of nutrients as indicated in the key at lower left. The fraction of total protein devoted to ribosome-affiliated proteins (ϕ_R) is given by the RNA/protein ratio as $\phi_R = \rho \cdot r$ (table S1). (**B**) The RNA/protein ratio for a family of translational mutants SmR (triangles) and SmP (inverted triangles) and their parent strain Xac (circles) (*27*), grown with various nutrients (see key at lower left) (table S2). Translational inhibition of the parent Xac strain via exposure to sublethal doses of chloramphenicol (circled numbers; see legend table) gave RNA/protein ratios similar to those of the mutant strains grown in medium with the same nutrient but without chloramphenicol (light blue symbols). Dashed line is a fit to Eq. 2. Inset: Linear correlation of κ_t values obtained for the Xac, SmR, and SmP strains (table S2) with the measured translation rate of the respective strains (*14*) ($r^2 = 0.99$).

Problems with FBA.

A major one: overflow metabolism.

Overflow metabolism is the preference of several unicellular species to sustain high rate growth, even in presence of oxygen, by selecting anaerobic pathways over aerobic ones (Crabtree effect in yeast).

E. coli undergoes a major metabolism switch upon increasing the growth yield, passing from mainly aerobic to mainly anaerobic energetics.

This cannot be explained by FBA without ad hoc constructions.

1. Regulatory scenario for the proteome organization

plus

2. Flux Balance (steady state and optimization)

leads to CAFBA, Constrained Allocation Flux Balance Analysis.

Proteome is organized in sectors.

M, R, Q: different proteome sectors.

Mass fractions adjust with growth rate λ.

M: metabolic proteins;
R: ribosomal proteins;
Q: growth independent proteins.

ϕ_M, ϕ_R, ϕ_Q: mass fractions of different proteome sectors.

$$\phi_M + \phi_R + \phi_Q = 1 \, .$$

Since ϕ_Q does not depend on λ

$$\phi_M + \phi_R = 1 - \phi_Q \ = \text{constant} \, .$$

In other terms:

$$\phi_M(\lambda) + \phi_R(\lambda) = \mathbf{K} \, .$$

Experimental evidence:

$$\phi_R(\lambda) = \phi_R^0 + \frac{\lambda}{\mathbf{k}_t}$$

$\phi_R(\lambda)$ is found to vary linearly with λ. There is an offset.

$\mathbf{k}_t$: "translational efficiency".

$\mathbf{k}_t \simeq 5.9/h$ (in absence of translation limiting antibiotics).

$\phi_R \sim 0.06$.

Both do not depend much on environmental conditions.

A robust regulatory mechanism enforces this behavior.

We assume that the proteome fraction has the form

$$\phi_i = \phi_i^0 + w_i \left|\nu_i\right| \ ,$$

where

ν_i is the flux of reaction i,

ϕ_i^0 is the mass fraction of protein i (catalyzing reaction i) present when the reaction is not active (not important: it goes in the constant),

and w_i is a weight.

We consider the w_i and the ϕ_i^0 of regulatory origin.

The absolute value of the fluxes takes care of negative fluxes in reversible reactions.

We divide the metabolic M sector in two sectors: transporters T and enzymes E.

$$\phi_T(\lambda) = \sum_{i \in \text{transport}} w_i \, |\nu_i|$$

$$\phi_E(\lambda) = \sum_{i \in \text{internal reactions}} w_i \, |\nu_i|$$

This is very useful since in our model nutrient limitation is modeled by an increase of the weight w_i of the relative transporters. It will be useful to change separately the weights of the two sectors.

So, we are introducing a proteome allocation constraint of the form

$$\phi_T\left(\{\nu\}\right) + \phi_E\left(\{\nu\}\right) + \phi_R\left(\lambda\right) = \mathbf{K} \; .$$

(all constant value have been included in $\mathbf{K}$).

Let $S_{\mu,i}$ be the stoichiometric matrix; l_i and u_i respectively the lower and the upper bound for the i-th flux; $\mathbf{c}$ a vector prescribing the linear combination of fluxes we will optimize via the quantity $\mathbf{c} \cdot \nu$ typically the biomass.

CAFBA is defined by the following constrained optimization problem: find

$$\{\nu_{\rm opt}\} = \max_{\{\nu\}} (\mathbf{c} \cdot \nu) = \max_{\{\nu\}} (\lambda) \ ,$$

under the constraint $\sum_i S_{\mu,i}\nu_i = 0$ (FBA constraint, steady state) and $l_i \leq \nu_i \leq u_i \ , \ \ \forall i$, bounds for fluxes, and

$$\phi_T (\{\nu\}) + \phi_E (\{\nu\}) + \phi_R (\lambda) = \mathbf{K} \ .$$

(CAFBA constraint)

We select external conditions as simple as possible: growth rate maximization in glucose minimal medium.

Here $\phi_T \propto$ glucose intake flux (this is the only nutrient that gets transported), i.e.

$$\phi_T \; w_{glc} \, |\nu_{glc}| \; .$$

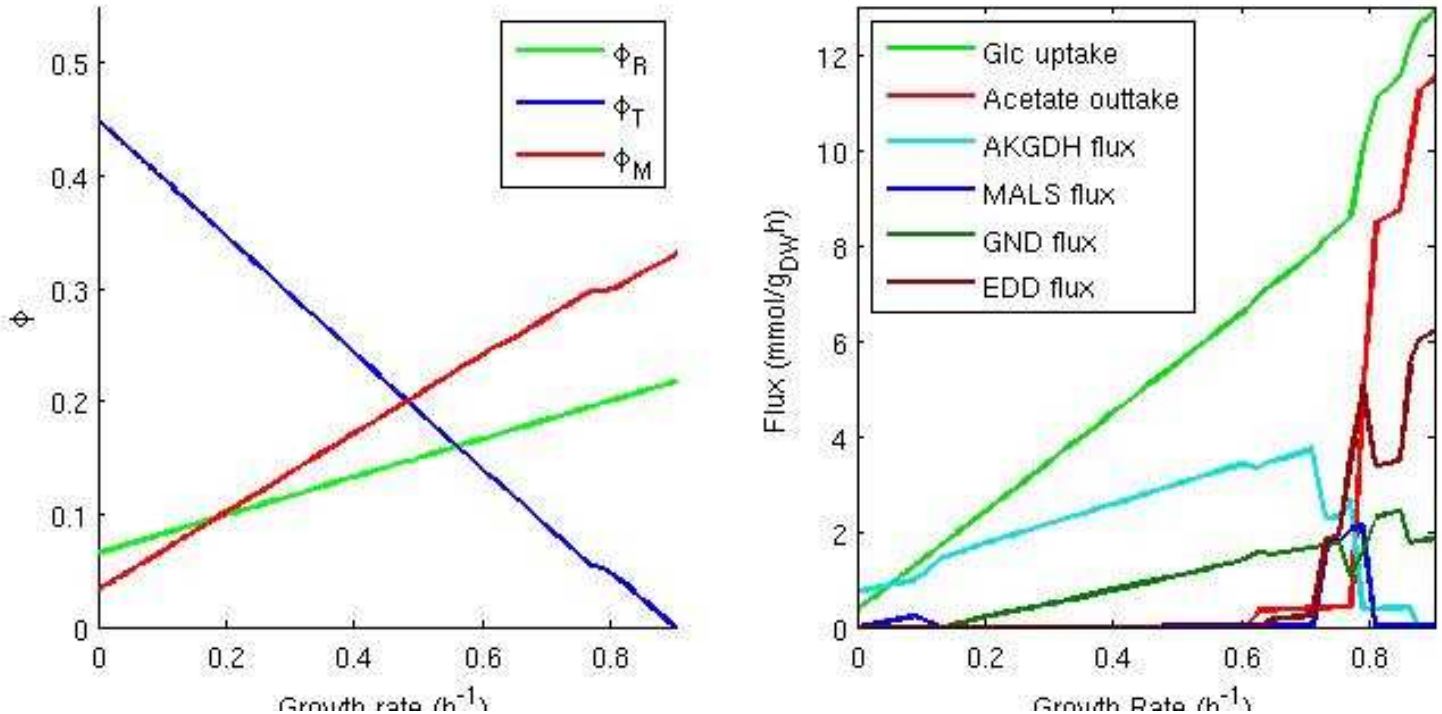
For ϕ_E, the enzymatic sector, the simplest (and first) choice is to set

$$w_i = w_E \;\; \forall i \in \text{internal reactions.}$$

This is natural if we assume that this dependence has a regulatory origin. For example several enzymes are co-regulated.

We fix w_E by assuming that the maximum growth rate (achieved for zero cost for transport, W_T=w_{glz}=0) equals a reference value

$$\lambda_{\max} = 0.9/h.$$


ϕ_R
ϕ_T
ϕ_M
ϕ
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
Growth rate (h^{-1})
Glc uptake
Acetate outtake
AKGDH flux
MALS flux
GND flux
EDD flux
Flux (mmol/g_{DW}h)
12
10
8
6
4
2
0
0
0.2
0.4
0.6
0.8
Growth Rate (h^{-1})

Most striking feature: at $\lambda \sim 0.75/h$ production of acetate (absent in FBA) starts, and it is present at larger growth rates.

Switch from anaerobial to aerobial metabolism.

TCA cycle is predominant at low growth rates.

At high growth rate a different pathway is activated, and acetate is secreted.

The transport proteome fraction decreases with λ, basically linearly.

Maximum growth rate is for null transport rate since $w_{glc} = 0$.

If nutrients are scarce the cell invests more of its resources in enzymes favoring transport.

See coarse grained model, Molenaar et al, Molecular Systems Biology 5:323 (2009).

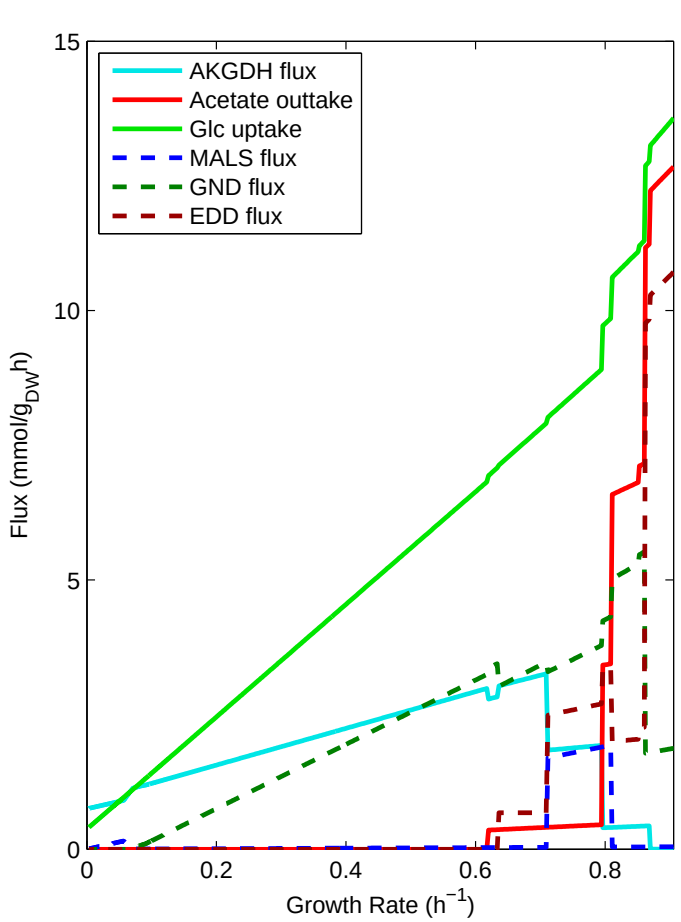
Right panel of the former figure: for our simple choice of constant weights the optimal flux configurations do not always vary smoothly with the growth rate.

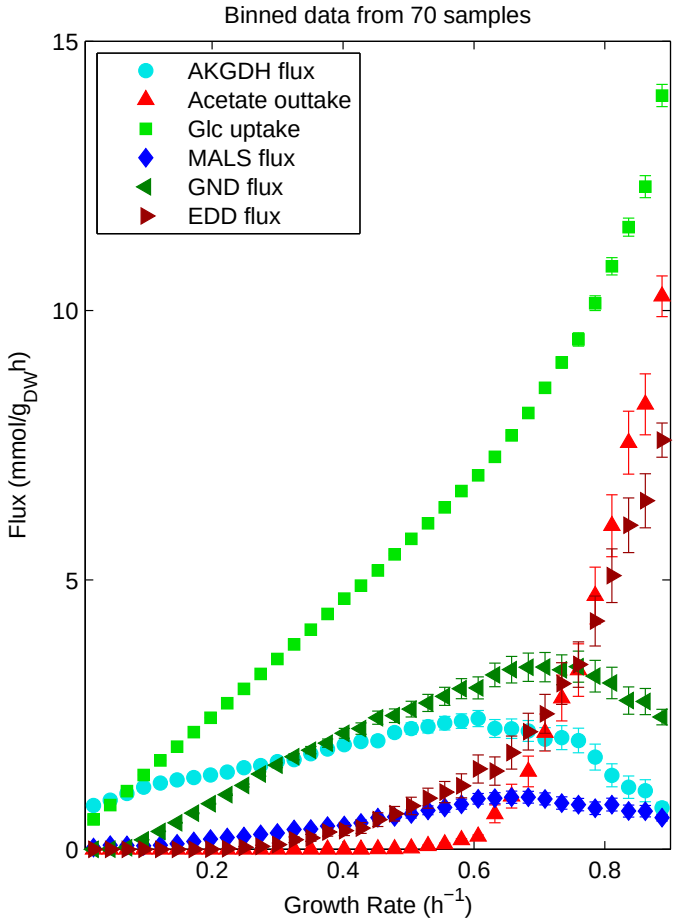
What is actually measured in experiments are the total fluxes concerning an ensemble of bacteria: one has to consider an average of the fluxes.

We obtain such an average by averaging over CAFBA solutions with randomly sampled weights w_i for the E-sector.

We extract independently the logarithm of the weights from the same uniform distribution, such that the weights span one order of magnitude.

Now we get a smooth pattern with the same features than before.


AKGDH flux
Acetate outtake
Glc uptake
MALS flux
GND flux
EDD flux
Flux (mmol/g$_{DW}$h)
Growth Rate (h^{-1})
0
5
10
15
0
0.2
0.4
0.6
0.8


Binned data from 70 samples
AKGDH flux
Acetate outtake
Glc uptake
MALS flux
GND flux
EDD flux
Flux (mmol/g$_{DW}$h)
Growth Rate (h^{-1})
0
5
10
15
0.2
0.4
0.6
0.8

Because of the CAFBA constraint to maximize the growth rate λ is equivalent to minimize $\phi_T(\{\nu\}) + \phi_E(\{\nu\})$.

Let us consider the simple case where in the transport section we only have the glucose intake flux.

We can fix λ and:

1. Minimize ϕ_T. Equivalent to standard FBA.

2. Minimize ϕ_E. Equivalent to FBA with a bound on total flux in the E-enzymatic sector.

The first solution is the canonical FBA solution for a given glucose flux ν_{glc}. It has maximum biomass yield, and no overflow metabolism.

The second solution presents overflow metabolism.

CAFBA interpolates between the two solutions.