Constraint-based modeling of cellular networks
WS 2023/2024

Master of Bioinformatics
University of Potsdam

Prof. Dr. Zoran Nikoloski

30.11.2023

Last lectures

Flux Balance Analysis

Based on the idea that organisms exposed to long-term evolutionary pressure optimize growth

Biomass reaction – with flux associated to cell growth

stoichiometry of biomass precursors from experiments

$$\alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_n S_n \rightarrow 1\ biomass + \alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_n S_n$$

Formalized as a linear program

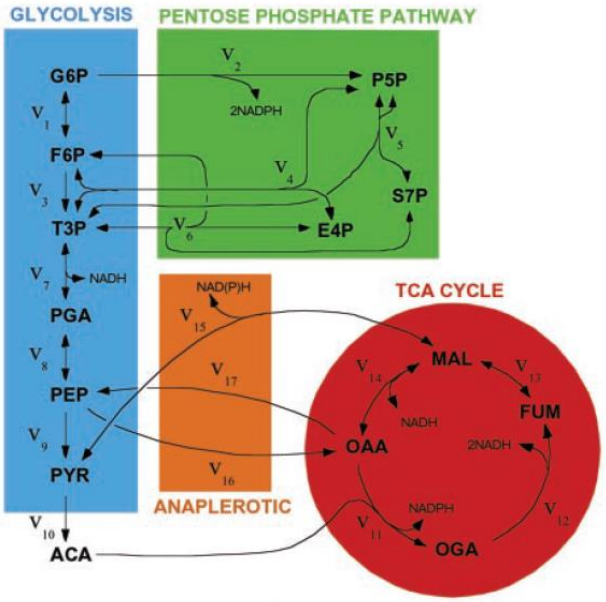
solved with the simplex method

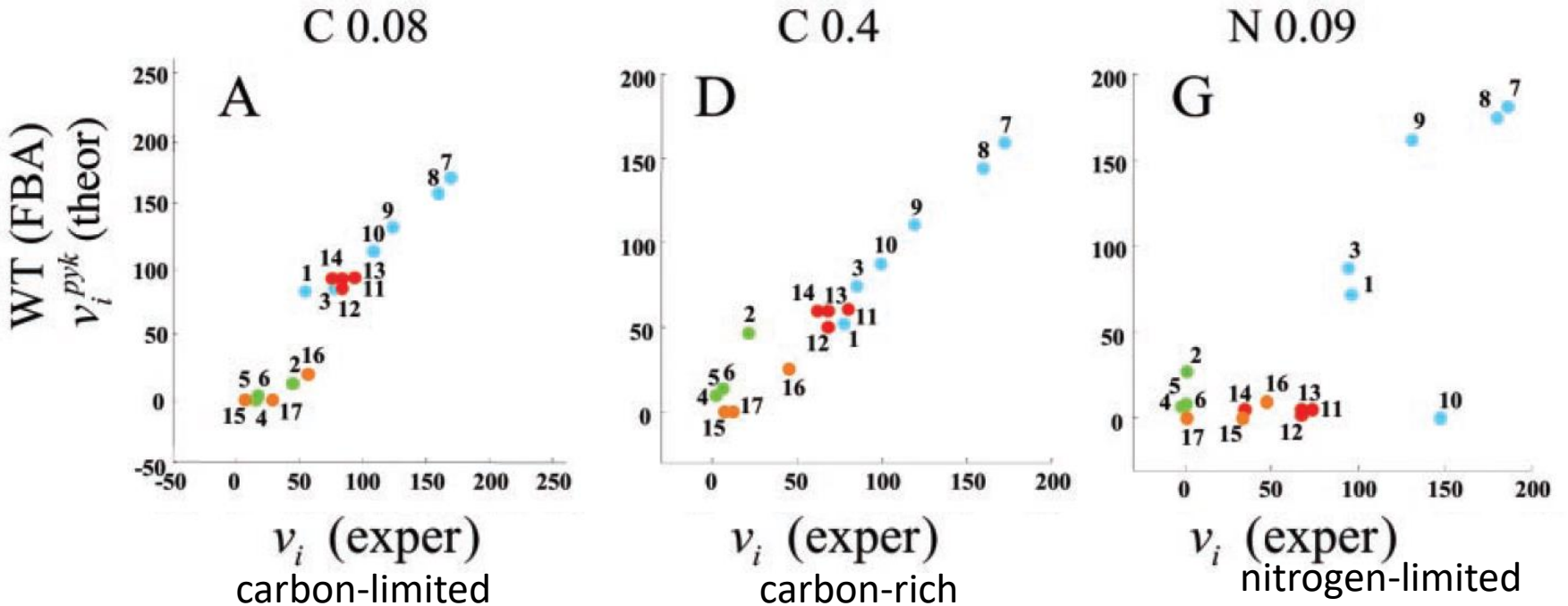
Last lectures

Flux Balance Analysis

Based on the idea that organisms exposed to long-term evolutionary pressure optimize growth

validation (Segre et al., *PNAS* 2002)





Performance of mutants

Flux Balance Analysis

applicable for organisms under long-term evolutionary pressure optimize growth

Allows to easily simulate knock-out (null) mutants

Performance of mutants

Flux Balance Analysis

applicable for organisms under long-term evolutionary pressure optimize growth

Allows to easily simulate knock-out (null) mutants

Example.

Suppose that a gene g codes for an enzyme which catalyzes a reaction r

Knock-out of the gene g is simulated by setting $v_r = 0$ and solving the FBA LP (on the board)

Performance of mutants

Knock-out of the gene g is simulated by setting $v_r = 0$ and solving the FBA LP (on the board)

Performance of mutants

Flux Balance Analysis

applicable for organisms under long-term evolutionary pressure optimize growth

Allows to easily simulate knock-out (null) mutants

More involved cases

isoenzymes – enzymes catalyzing the same reaction

enzyme complexes – enzymes composed of multiple units

can be modeled with the help of **gene-product-reaction (GPR) associations**

Performance of mutants

isoenzymes – enzymes catalyzing the same reaction

all genes for the isoenzymes should be knocked-out to simulate a knock-out mutant

enzyme complexes – enzymes composed of multiple units

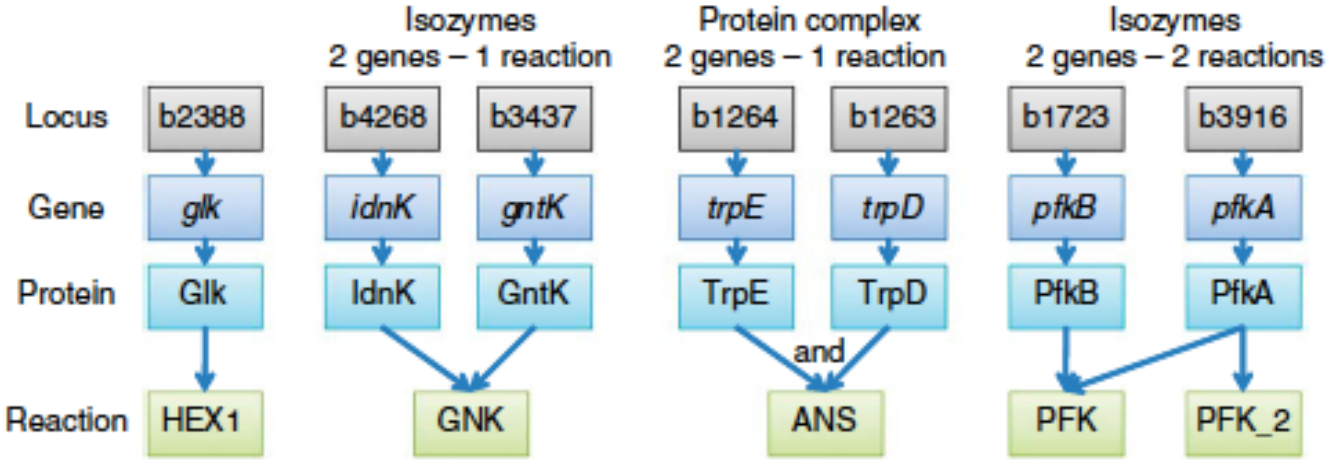
at least one genes for the protein units should be knocked-out to similate a knock-out mutant

can be modeled with the help of **gene-product-reaction (GPR) associations**

represented in a Boolean format

Performance of mutants

Examples of GPR associations and their representation in Boolean format are shown for *Escherichia coli* (Thiele I., et al.,2010)

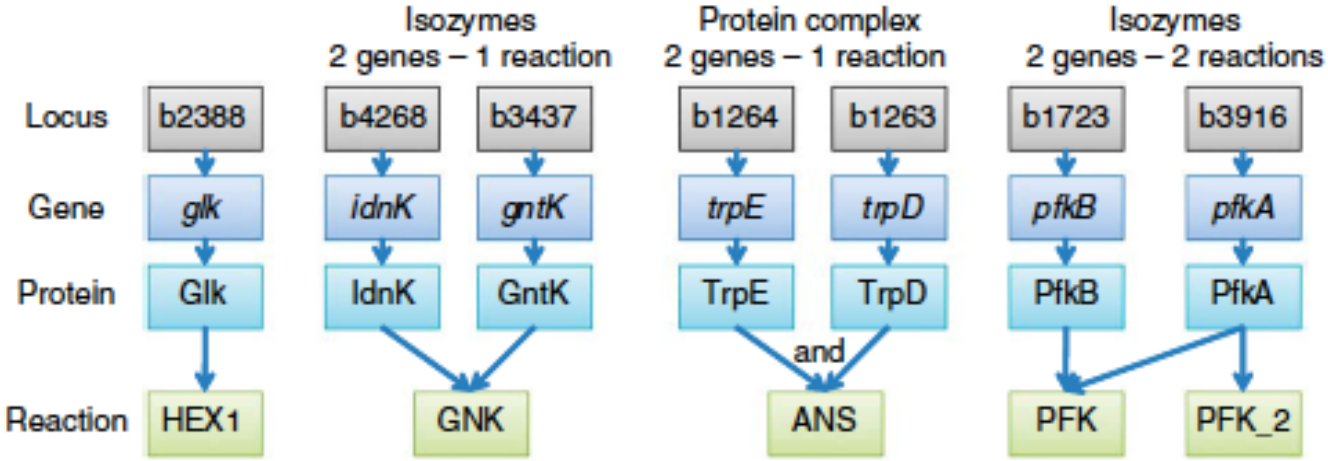


Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)

Performance of mutants

Knock-outs can be simulated by using the Boolean rules

if the expression evaluates to FALSE, the flux through reaction is set to zero



Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)

Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?

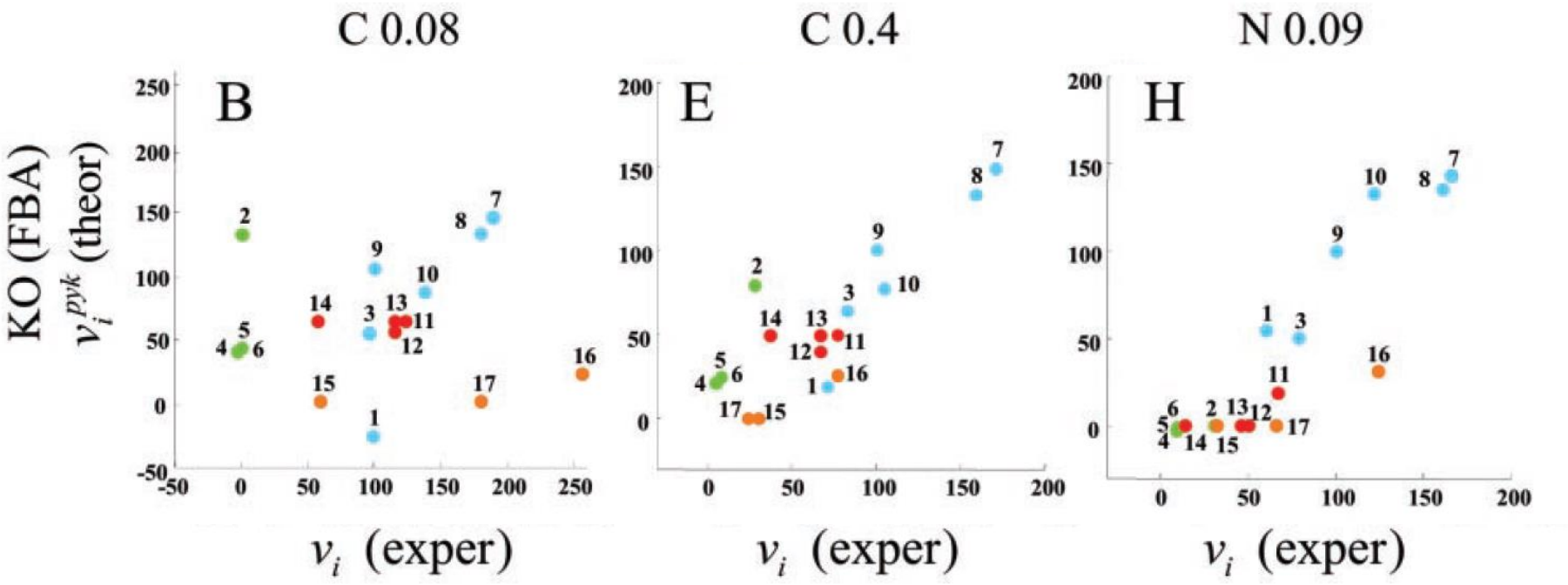
Which hypotheses can be invoked to accurately simulate mutants?

Relations to biochemical reality of the underlying hypotheses

Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?



Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?

Knock-out mutants generated artificially in the lab are not generally subjected to the same evolutionary pressure that have shaped the wild type

Such knock-outs do not have regulatory mechanisms to regulate fluxes towards optimal growth

Before going to other hypotheses, we ask:

How are the feasibility spaces of mutants and wild type related?

Performance of mutants

Before going to other hypotheses, we ask:

How are the feasibility spaces of mutants and wild type related?

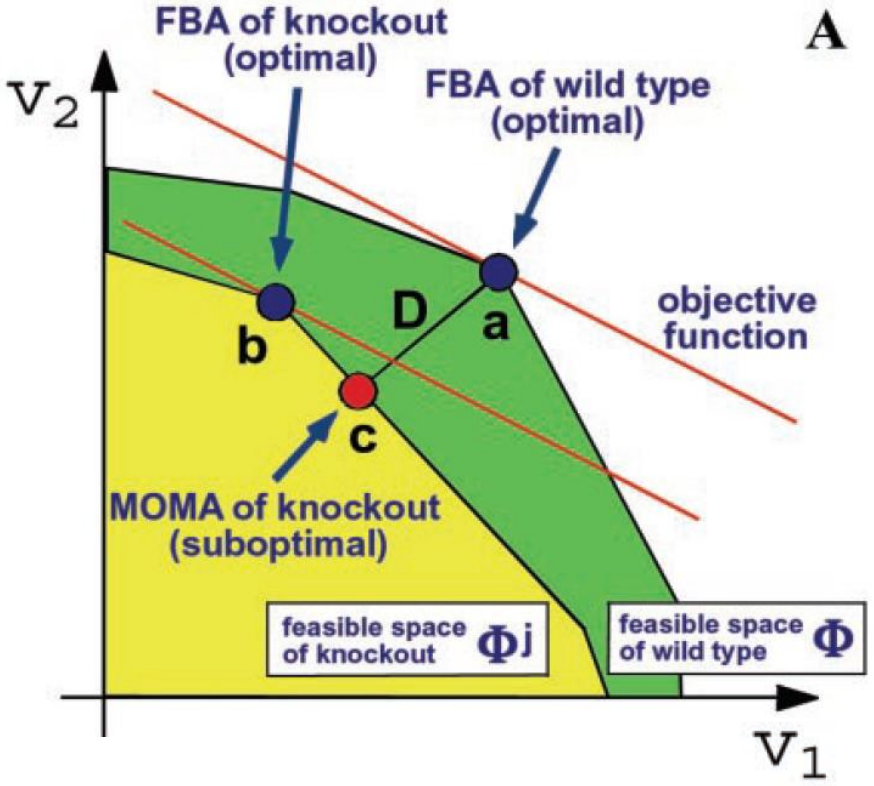
Feasibility space of mutant is properly contained in the feasibility space of the wild type!

assuming all lower and upper boundaries on flux capacities are the same

Performance of mutants

Before going to other hypotheses, we ask:

How are the feasibility spaces of mutants and wild type related?



Minimization of metabolic adjustment (MOMA)

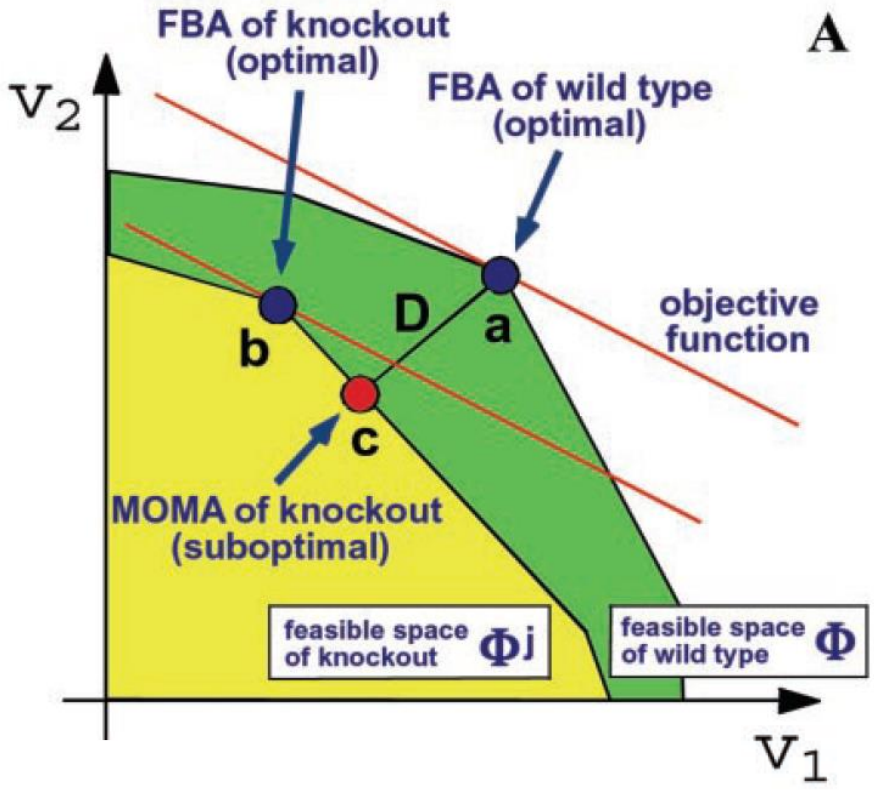
MOMA relaxes the assumption of optimal growth flux states for gene deletions

A mutant is likely to initially display a suboptimal flux distribution

intermediate between

WT optimum and

mutant optimium



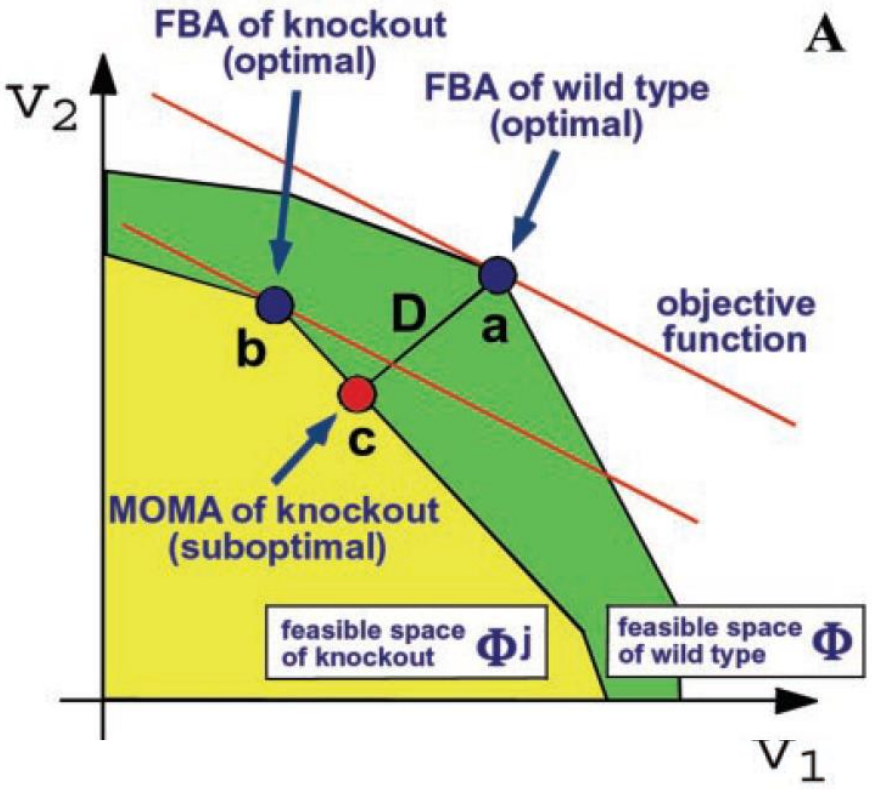
Minimization of metabolic adjustment (MOMA)

MOMA is based on the hypothesis that a mutant initially remains close to the wild type in terms of flux values

This hypothesis is to be tested!

The math behind differs from LP used in FBA.

Involves distance minimization in flux space



Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?

Which hypotheses can be envoked to accurately simulate mutants?

Relations to biochemical reality of the underlying hypotheses

MOMA hypothesis

MOMA searches for a point in the mutant feasible space that has a minimum distance from a given vector v

The goal is to find a vector w in the knock-out feasible space such that the Euclidean distance

$$\sqrt{\sum_{i=1}^{m}(w_i - v_i)^2}$$

is minimized.

Clearly, $\sqrt{\sum_{i=1}^{m}(w_i - v_i)^2}$ is minimized when $\sum_{i=1}^{m}(w_i - v_i)^2$ is minimized

$$\sum_{i=1}^{m}(w_i - v_i)^2$$

MOMA hypothesis

Closer look at

$$\sum_{i=1}^{m}(w_i - v_i)^2$$

$$\sum_{i=1}^{m}(w_i - v_i)^2 = \sum_{i=1}^{m} {w_i}^2 - 2w_i v_i + {v_i}^2 =$$

$$= \sum_{i=1}^{m} {w_i}^2 - 2\sum_{i=1}^{m} v_i w_i + \sum_{i=1}^{m} {v_i}^2$$

Parts in red denote constants

The last, constant term can be removed

it does not affect the minimization

MOMA hypothesis

Closer look at

$$\sum_{i=1}^{m}(w_i - v_i)^2$$

$$\sum_{i=1}^{m}(w_i - v_i)^2 = \sum_{i=1}^{m} {w_i}^2 - 2w_i v_i + {v_i}^2 =$$

$$= \sum_{i=1}^{m} {w_i}^2 - 2\sum_{i=1}^{m} v_i w_i + \sum_{i=1}^{m} {v_i}^2$$

Note that $\sum_{i=1}^{m} v_i w_i = v^T w$

MOMA hypothesis

How could we re-write $\sum_{i=1}^{m} {w_i}^2$?

$$\sum_{i=1}^{m} {w_i}^2 = w^T I_m w$$

Example.

for $m = 3$,

$$\sum_{i=1}^{3} {w_i}^2 = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

Quadratic programming (QP)

Quadratic programming problem can be formulated as

$$\min_x \frac{1}{2} x^T Q x + c^T x$$

s.t.

$$Ax \leq b$$

Note that the constraints are linear

Objective is quadratic and composed of

quadratic part and

linear part

Quadratic programming (QP)

What kind of matrix can Q be?

One can inspect the eigenvalues of Q

What is an eigenvalue and eigenvector?

$$Qx = \lambda x$$
$$(Q - \lambda I)x = 0$$

Has a non-zero solution for x if and only if

$$\det(Q - \lambda I) = 0$$

Quadratic programming (QP)

What is an eigenvalue and eigenvector?

$$Qx = \lambda x$$

Has a non-zero solution for x if and only if

$$\det(Q - \lambda I) = 0$$

Example

$$Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\det(Q - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 3 - 4\lambda + \lambda^2$$

$$\lambda_1 = 1, \lambda_2 = 3$$

Solve for x by substitution in $(Q - \lambda I)x = 0$!

Quadratic programming (QP)

What kind of matrix can Q be?

One can inspect the eigenvalues of Q

If all eigenvalues of Q are positive, the matrix is **positive definite**

If some eigenvalues are zero and the others are positive, the matrix is **positive semidefinite**

If the matrix is not positive semidefinite it can be

negative semidefinite (some zero, all other negative) or

indefinite (eigenvalues of mixed signes)

Quadratic programming (QP)

What kind of matrix can Q be?

One can inspect the eigenvalues of Q

If all eigenvalues of Q are positive, the matrix is **positive definite**

If some eigenvalues are zero and the others are positive, the matrix is **positive semidefinite**

If the matrix is not positive semidefinite it can be

negative semidefinite (some zero, all other negative) or

indefinite (eigenvalues of mixed signes)

Quadratic programming (QP)

What kind of matrix can Q be?

One can inspect the eigenvalues of Q

$x^T Q x$ is convex if Q is positive semidefinite

local optimum coincides with the global optimum

$x^T Q x$ is strictly convex if Q is positive definite

local optimum coincides with the global optimium and x is unique optimizer

Quadratic programming (QP)

Different approaches to solve QP problems

- interior point
- active set
- augmented Lagrangian
- conjugate gradient
- extensions to simplex algorithm

For QPs for which Q is positive definite and contain only equality constraints, the problem reduces to solving a system of linear equations

MOMA hypothesis

Closer look at

$$\sum_{i=1}^{m}(w_i - v_i)^2$$

$$\sum_{i=1}^{m}(w_i - v_i)^2 = \sum_{i=1}^{m} {w_i}^2 - 2w_i v_i + {v_i}^2 =$$

$$= \sum_{i=1}^{m} {w_i}^2 - 2\sum_{i=1}^{m} v_i w_i + \sum_{i=1}^{m} {v_i}^2$$

Can be turned into $w^T I_m w + (-2v^T)w + const$

Analogous with the QP objective function $\frac{1}{2}x^T Q x + c^T x$

MOMA formulation

Provided a steady-state flux distribution of a wild type v, MOMA is given by the following QP problem

$$\min_{w} w^T I_m w + (-2v^T)w$$

s.t.

$$Nw = 0$$
$$w_j = 0$$
$$\forall i, 1 \leq i \neq j \leq m, w_i^{min} \leq w_i \leq w_i^{max}$$

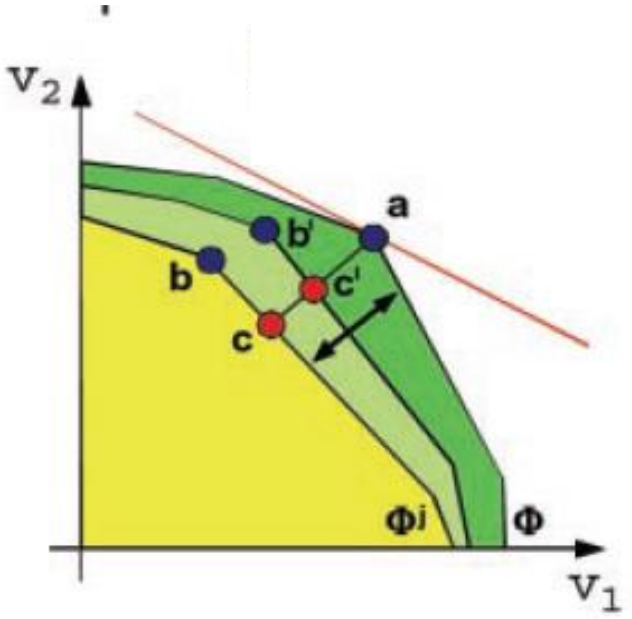
Note that biomass optimization does not enter explicitly the formulation of MOMA

MOMA robustness analyses

Partial knock-outs

by modifying upper bounds

recalculating optimum for the mutant and the MOMA projection



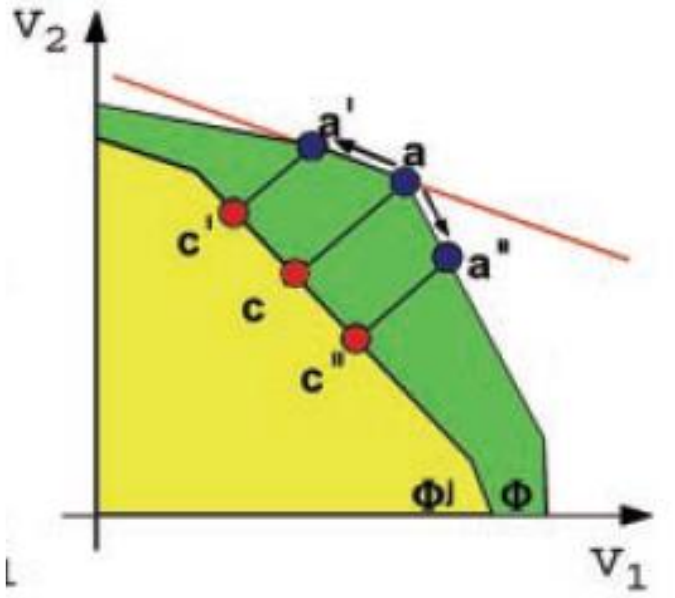
MOMA robustness analyses

Sampling-based

sample around the FBA optimum

determine MOMA projections

allows to study alternative optima to MOMA



Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?

Which hypotheses can be invoked to accurately simulate mutants?

Relations to biochemical reality of the underlying hypotheses

Findings based on MOMA

Knock-outs are simulated by fixing reaction fluxes to zero if

all isoenzymes are deleted and

(erroneously) all units in a protein complex are deleted

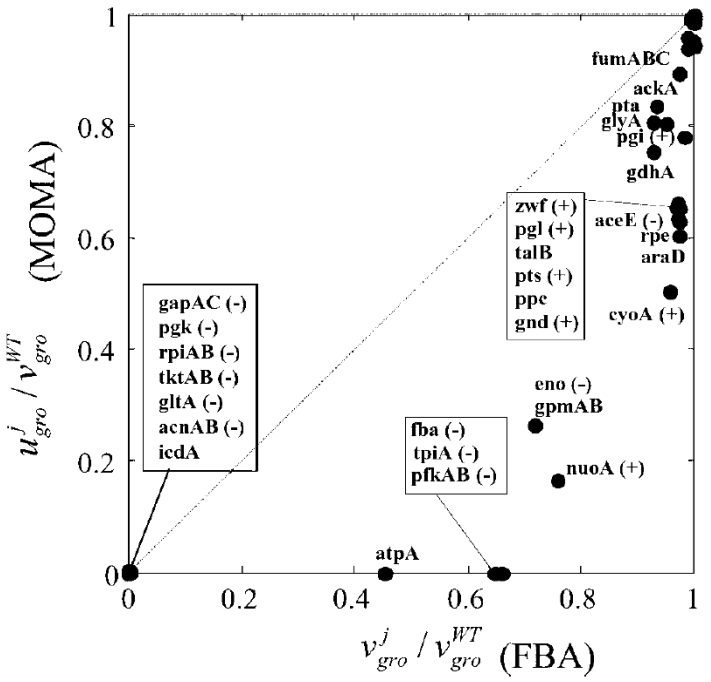
MOMA naturally results in smaller yields than predicted by FBA

Largest difference for

Fructose-1,6-bisphosphate aldolase (fba)

Triose phosphate isomerase (tpi)

Phosphofructokinase (pfk)



Findings based on MOMA

Knock-outs are simulated by fixing reaction fluxes to zero if

all isoenzymes are deleted and

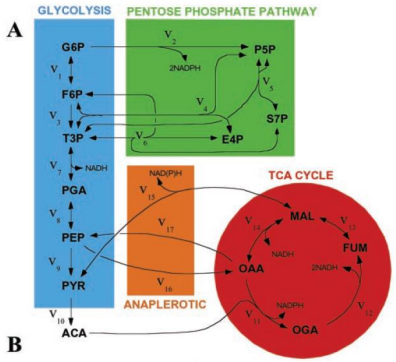
(erroneously) all units in a protein complex are deleted

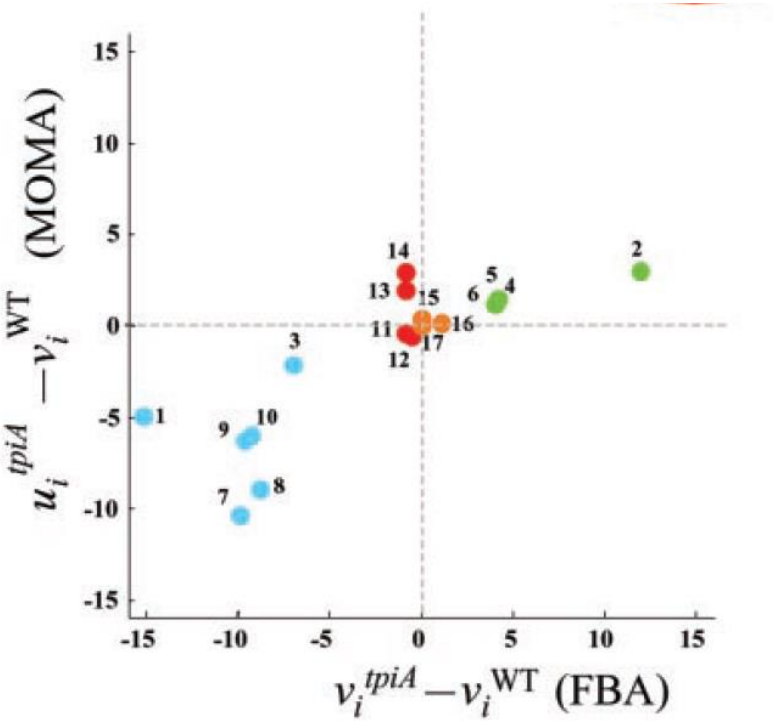
MOMA naturally results in smaller yields than predicted by FBA

Taking tpi as example

Largest difference in

Pentose phosphate pathway





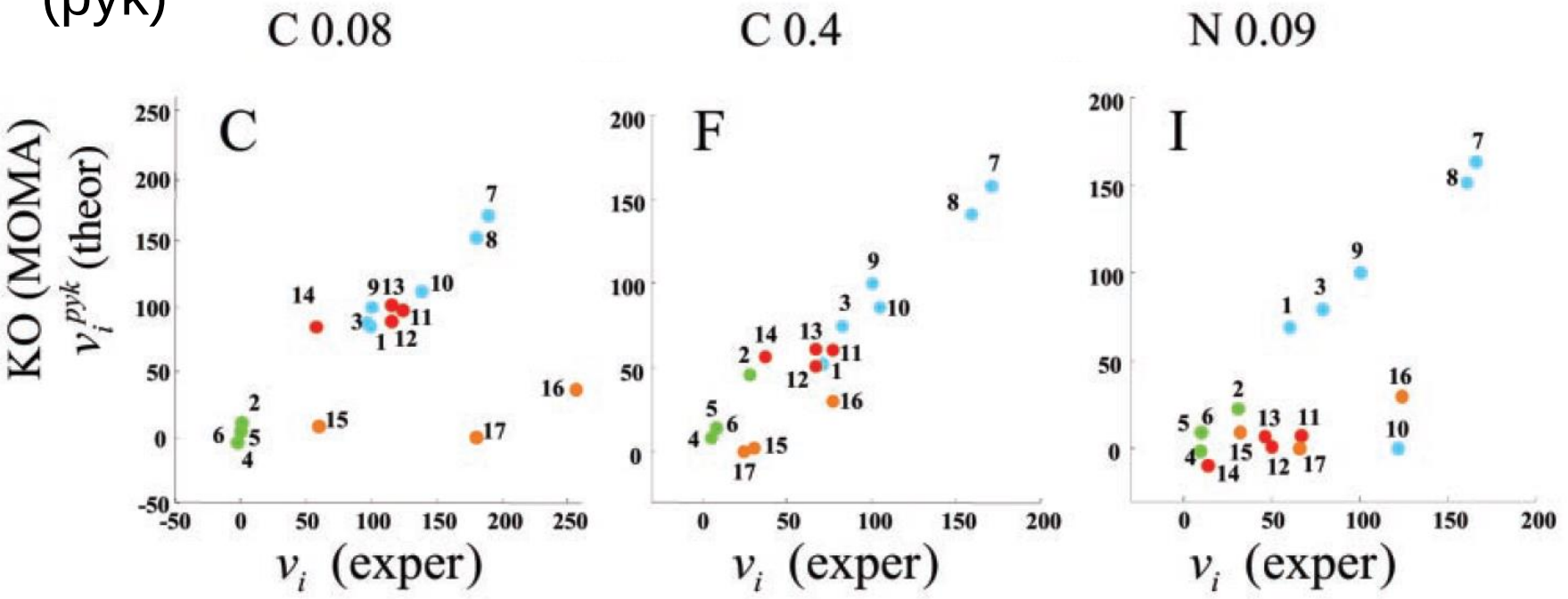
Findings based on MOMA

Knock-outs are simulated by fixing reaction fluxes to zero if

all isoenzymes are deleted and

(erroneously) all units in a protein complex are deleted

Better fit to experimental data from pyruvate kinase knock-out (pyk)



Performance of mutants

Question

Is FBA suitable for simulation of flux distributions for knock-out mutants?

Experimental evidence indicates that, in many cases, growth rate after deletion drops in comparison to that of wild type, but then gradually increases and reaches nearly that of wild type

MOMA provides one way to explain these changes; however

it implies numerous small changes in all fluxes

rather than few large changes in few fluxes

(of same value)

These characteristics reflect the used Euclidean norm!

Regulatory on/off minimization (ROOM)

Question

Which hypotheses can be envoked to accurately simulate mutants?

ROOM is based on the hypotheses that

- The genetic regulatory changes recquired for realizing flux changes after gene knock-outs are minimized by the cell

Minimizes adaptation costs

- The regulatory changes can be parsimoniously described by Boolean on/off dynamics

Assigns fixed cost to each regulatory change, regardless of its magnitude

ROOM formulation

ROOM employs a different metric

minimizes large enough flux changes

cast as a MILP

Let $y_i = 1$ denote that reaction i shows significant flux change and $y_i = 0$, otherwise

Let v_i^u and v_i^l denotes the upper and lower bounds for significant changes

These bounds are set based on a given wild type flux distribution v and two parameters

δ - determining the relative changes and

ϵ – sensitivity bound

ROOM formulation

ROOM employs a different metric

minimizes large enough flux changes

cast as a MILP

$$\min \sum_{i=1}^{m} y_i$$

s.t.

$$Nw = 0$$

$$w_j = 0$$

$$\forall i, 1 \leq i \leq m, w_i^{min} \leq w_i \leq w_i^{max}$$

$$\forall i, 1 \leq i \leq m, y_i \in \{0,1\}$$

$$w_i - y_i(w_i^{max} - v_i^u) \leq v_i^u$$

$$w_i - y_i\left(w_i^{min} - v_i^l\right) \geq v_i^l$$

$$v_i^u = v_i + \delta|v_i| + \epsilon$$

$$v_i^l = v_i - \delta|v_i| - \epsilon$$

ROOM formulation

ROOM employs a different metric

minimizes large enough flux changes

cast as a MILP

Let $y_i = 1$

$$w_i - w_i^{max} + v_i^u \leq v_i^u$$
$$w_i - w_i^{min} + v_i^l \geq v_i^l$$

Flux can vary between w_i^{min} and w_i^{max}

Let $y_i = 0$

$$w_i \leq v_i^u$$
$$w_i \geq v_i^l$$

Flux can vary between bounds v_i^l and v_i^u determined by the wild-type

Is there a guarantee for significant flux change?

Regulatory on/off minimization (ROOM)

Like MOMA, ROOM also does not explicitly consider optimization of biomass

At a branch point, ROOM predicts taking one path (**linearity hypothesis**) rather than letting flux through both branches

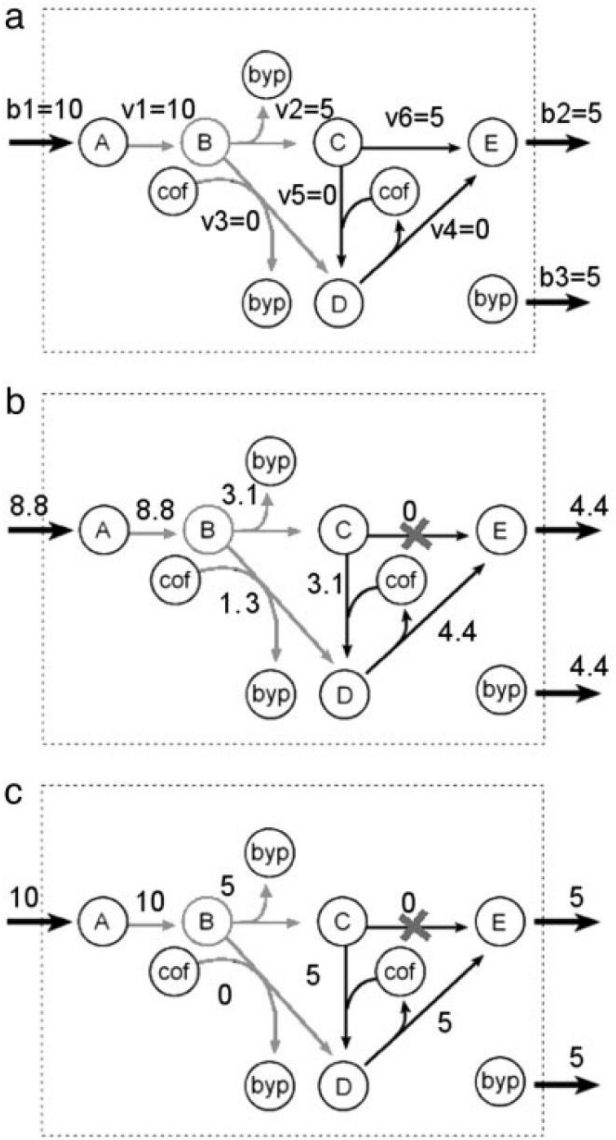
as a result of the different metric used

Regulatory on/off minimization (ROOM)

FBA flux distribution

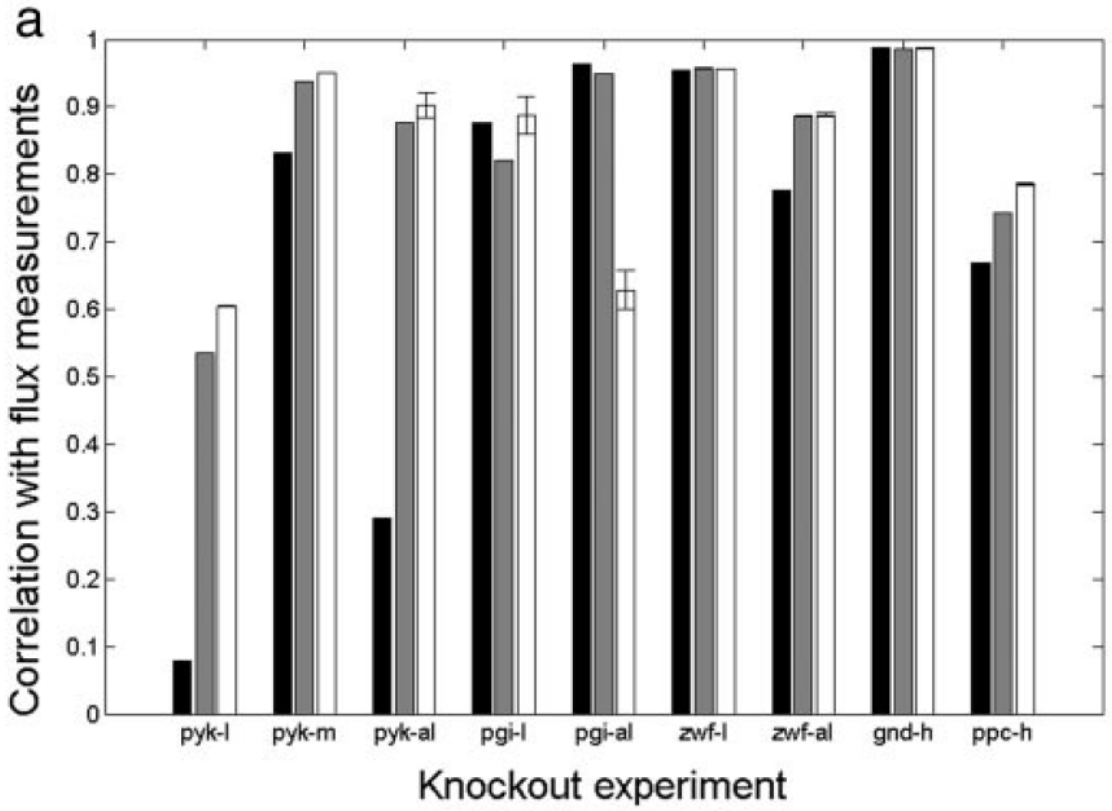
MOMA solution

ROOM solution



Findings based on ROOM

ROOM improves correlation between fluxes in comparison to MOMA

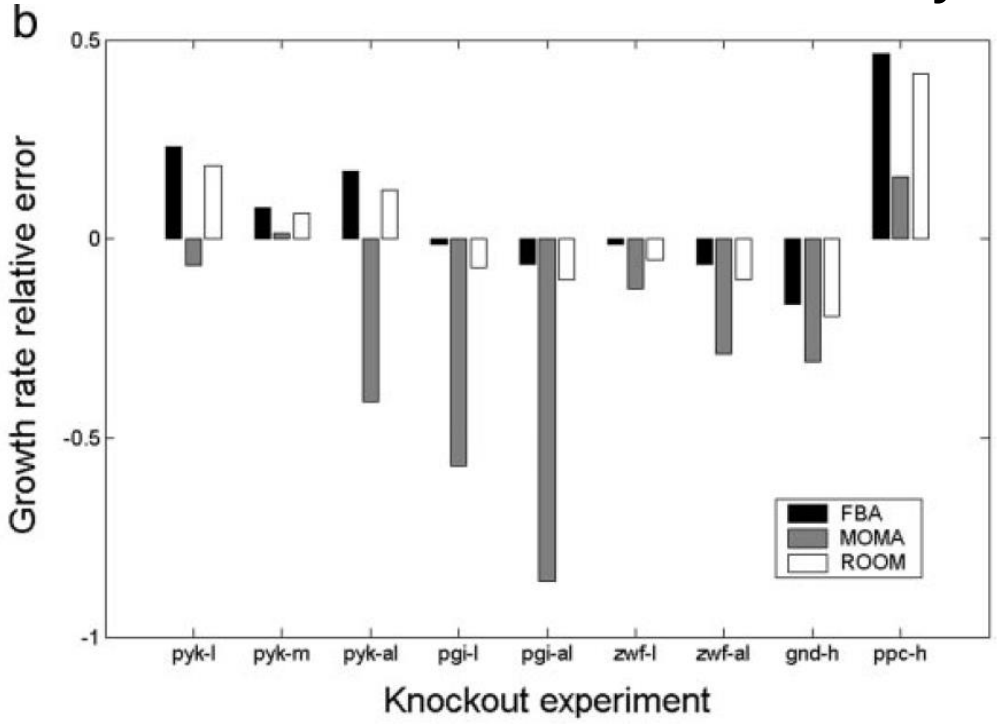


Findings based on ROOM

Simulations show that ROOM results in higher optimal biomass than MOMA

thus matching the **postadaptation states**

(after the transient, better modeled by MOMA)



Findings based on ROOM

Lethality prediction

(based on threshold of 5% from optimal biomass)

Table 2. Lethality predictions by FBA, MOMA, and ROOM

	FBA	MOMA	ROOM
True-positive	449	399	449
False-positive	64	60	62
True-negative	23	27	25
False-negative	19	69	19
Positively predicted genes	96%	85%	96%
Negatively predicted genes	26%	31%	29%
Overall prediction	85.0%	76.7%	85.4%

A true-positive is a case in which the gene deletion is lethal both experimentally and computationally. A false-positive case occurs when the gene deletion is computationally nonlethal, but the mutant cannot grow experimentally. A true-negative case occurs when the gene deletion is nonlethal both experimentally and computationally. A false-negative case occurs when the gene deletion is computationally lethal, but the mutant can grow experimentally. The positively predicted genes comprise the fraction of experimentally nonlethal genes that are correctly predicted, and the negatively predicted genes are the fraction of experimentally lethal genes that are correctly predicted. The overall prediction is the percent of true results of all predictions.

Recap

FBA is not suitable for simulation of flux distributions for knock-out mutants

Two hypotheses to simulate mutants

- minimization of metabolic adjustment (MOMA)
- regulatory on/off minimization (ROOM)
- QP vs. MILP
- alternative optima

The quest for distance functions which explain the difference between wild type and mutants is still an open problem

- **thermodynamics**
- **metabolite concentrations**
- **change in biomass composition**