# *Constraint-Based Modelling Of Perturbed Organisms: A ROOM for improvement*

*Tomer Shlomi[1], Omer Berkman[1, 2] and Eytan Ruppin [* 1]*

*[1]School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel and [2]Dept. of Computer Science, The Academic College of Tel-Aviv Yaffo, Tel-Aviv 64044, Israel*

**ABSTRACT**

Regulatory On-Off Minimization (ROOM) is a model for predicting the behavior of metabolic networks in response to gene knockouts. It is based on minimizing the number of significant flux changes (hence, on/off) with respect to the wild-type. ROOM outperforms a previously suggested model, Minimization Of Metabolic Adjustment (MOMA), whose minimization metric is Euclidian. Furthermore, ROOM shows its ability to correctly identify alternative pathways for reactions associated with the knocked-out genes, thus strengthening its biological plausibility. ROOM outperforms MOMA in predicting intracellular fluxes and gene knockout lethality in mutated *E. coli* and the *S. cerevisiae* strains, respectively.

**Supplementary data:** www.cns.tau.ac.il/∼shlomito/room

**Contact:** ruppin@post.tau.ac.il

**Keywords:** Metabolic networks, FBA, MOMA, ROOM

## INTRODUCTION

The study of metabolic networks has attracted a lot of attention in recent years. Much of the research has concentrated on building mathematical models of cell metabolism. In this paper we focus on flux analysis using steady state constraint-based modelling. In constraint-based modelling, stoichiometric, thermodynamic, flux capacity, energy preservation and possibly other constraints are used to limit the space of possible flux distributions attainable by the metabolic network. Assuming optimal behavior of the network, various optimization criteria such as growth or energy maximization are applied, with the aim of achieving a biologically meaningful description of the metabolic state of the system. Flux Balance Analysis (FBA) is a specific constraint-based method which has been successfully used for predicting growth, uptake rates and by-product secretion, among others (Reed *et al.*, 2003; Förster *et al.*, 2003).

A perhaps more challenging task is predicting the lethality and phenotypes of organisms after gene knockouts. As was recently observed in Segre *et al.* (2002), assuming optimal behavior for a knocked-out organism may not adequately reflect its true behavior. Instead, Segre *et al.* (2002) suggested that the knocked-out organism adjusts by minimizing the changes of its flux distribution in accordance with the Minimization Of Metabolic Adjustment (MOMA) approach. This method minimizes the Euclidian norm of the flux differences between the metabolic networks of the knocked-out strain and the wild-type. MOMA was reported to successfully predict *E. coli* lethality and internal metabolic flux values of the knocked-out network.

Accepting the assumption that instead of optimizing growth, as in FBA, the mutated strain optimizes its flux proximity to the wild type, we argue that the Euclidian metric on which MOMA is based does not adequately reflect the true biological metabolic adjustment of the metabolic network after the knockout. This is since the Euclidian metric may prohibit large modification in single fluxes. Such large modifications may be required for re-routing metabolic flux through alternative pathways and are actually observed at times experimentally (Emmerling *et al.*, 2002) as discussed in the Results Section. For example, when a knocked-out enzyme is ”backed up” by an isoenzyme, the most reasonable adjustment of the perturbed network is through an alternative pathway consisting of this isoenzyme alone. In general, when an alternative short pathway can compensate for the lost reactions, it is reasonable that this pathway will be used to re-route the metabolic flux.

We therefore propose that proximity minimization takes place under a distance metric different from MOMA. Our new method, Regulatory On-Off Minimization (ROOM) simply minimizes the total number of significant flux changes. Specifically, ROOM finds a flux distribution for a perturbed strain, that satisfies stoichiometric constraints (mass balance), thermodynamical and flux capacity constraints, while minimizing the total number of significant flux changes from the respective fluxes of the wild-type strain. ROOM improves intracellular metabolic flux prediction for an *E. coli* mutated strain, with respect to MOMA’s predictions. ROOM also outperforms MOMA’s lethality predictions in a large-scale gene

*To whom correspondence should be addressed.

deletion experiment in the Yeast.

The underline heuristic behind ROOM's distance metric is based mainly on the view that: A. The genetic regulatory changes that are required for realizing flux changes, after gene knockouts, are minimized by the cell attempting to minimize its adaptation cost. B. That such regulatory changes can be parsimoniously described by boolean on/off dynamics (de Jong, 2002). ROOM implicitly accounts for regulatory changes by identifying significant flux changes.

## METHODS

We use the metabolic network model of the *E. coli* (MG1655) (Reed *et al.*, 2003). The stoichiometric matrix contains 536 metabolites and 953 reactions. Modifications of the model for the *E. coli* JM101 and PB25 strains were adapted from Segre *et al.* (2002). We also use the metabolic network model of the Yeast (Förster *et al.*, 2003), which contains 828 metabolites and 1433 reactions.

### FBA

FBA uses Linear Programming (LP) to maximize an objective function under different constraints. In our model, we look for a steady state flux distribution ($v$) that maximizes growth rate under mass balance, thermodynamical and flux capacity constraints. The LP is formalized as follows:

$\max f^T v$,
s.t. $S \cdot v = 0$, $v_{min} \leq v \leq v_{max}$.

Here, mass balance constraints are imposed by a system of linear equations, where $S$ is a $m \times n$ stoichiometric matrix, in which $m$ is the number of metabolites and $n$ is the number of reactions. The vector $f$ is an objective function maximizing growth rate which is represented by a reaction that drains biomass components. Thermodynamic constraints that restrict directional flow of reaction, and capacity constraints are imposed by setting $v_{min}$ and $v_{max}$ as lower and upper bounds on flux values.

### MOMA

MOMA finds a solution that satisfies the same constraints as the FBA while minimizing the Euclidian distance from a wild type flux distribution (usually obtained previously by FBA). MOMA is formalized using Quadratic Programming (QP) as follows:

$\min(v - w)^T(v - w)$,
s.t. $S \cdot v = 0$, $v_{min} \leq v \leq v_{max}$,
$v_j = 0$, $j \in A$,

where $w$ is the wild type flux distribution and $A$ is a set of reactions associated with the deleted genes.

### ROOM

ROOM finds a flux distribution that satisfies the same constraints as the FBA while minimizing the number of significant flux changes. We account only for significant flux changes because of the inherent noise in biological systems and to achieve acceptable running time. We define a range $[w^l, w^u]$ around the vector $w$ for non-significant flux change. We turn to Mixed Integer Linear Programming (MILP) which can be formalized as:

$\min \sum_{i=1}^{m} y_i$,
s.t. $S \cdot v = 0$,
$v - y(v_{max} - w^u) \leq w^u$, (1)
$v - y(v_{min} - w^l) \geq w^l$, (2)
$v_j = 0$, $j \in A$, $y_i \in \{0, 1\}$,
$w^u = w + \delta|w| + \epsilon$, $w^l = w - \delta|w| - \epsilon$,

where, for each flux $i$, $1 \leq i \leq m$, $y_i = 1$ for a significant flux change in $v_i$, and $y_i = 0$ otherwise.

Indeed, when $y_i = 1$, inequalities (1) and (2) do not impose new constraints on $v_i$, where if $y_i = 0$ inequalities (1) and (2) constrain $v_i$ to the range defined above. The size of $\delta$ and $\epsilon$ influences the running time of the MILP solver, and very small values extremely increase the running time of the solver. In our experiments we have used $\delta$ in the range $[0.01, 0.1]$ and $\epsilon = 0.0001$.

## RESULTS

### Flux predictions

Measurements of metabolic fluxes in a *E. coli* - Pyruvate Kinase *pyk* knockout were reported in Emmerling *et al.* (2002). A collection of 17 intracellular fluxes from the central carbon metabolism of the *E. coli* were empirically determined by combining NMR spectroscopy in $^{13}$C labelling experiments and physiological data measurements (see Table ST1 in the Supp. data for a list of reactions with the corresponding measurements). The fluxes of both the wild-type strain (JM101) and the *pyk* mutant (PB25) were measured on 3 different growing conditions resulting in six sets of fluxes. Two glucose-limited conditions were measured (low and high concentration) and one nitrogen-limited condition.

In order to model the *pyk* mutant strain (PB25, *pykA::kan pykF::cat*), two reactions from Phosphoenolpyruvate (PEP) to Pyruvate (PYR) were constrained to zero. Flux v9 (Figure SF1 in the Supp. data) is composed of those 2 reactions and an additional reaction.

Segre *et al.* (2002) have compared FBA and MOMA predictions with this set of experimental fluxes. For the wild-type strain the FBA is a good predictor of the flux values in all three cases, while for the *pyk* mutant its prediction performance is poor for both the low-glucose and low-nitrogen growing conditions. MOMA's prediction for the mutated strain is much more accurate than that of the FBA.

Predictions obtained by applying ROOM for the *pyk* mutant were significantly more accurate than those obtained by either FBA or MOMA. For both the low

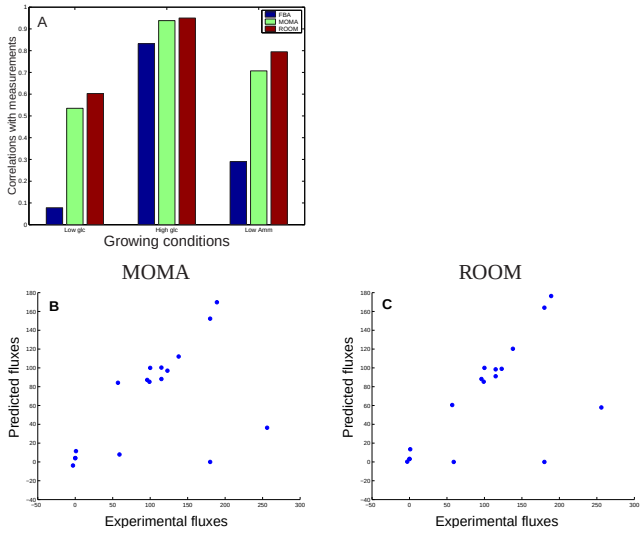


**Fig. 1.** (A) Pearson correlation between FBA, MOMA and ROOM flux predictions and experimental measurements for a *pyk* knockout strain under three different growing conditions. (B,C) A comparison of MOMA and ROOM flux predictions with experimental measurements for a *pyk* knockout strain under the low glucose growing condition. Fluxes are expressed in percent of glucose uptake rate.

glucose and the nitrogen-limited experiments, ROOM's correlations with the experimental flux data were 12% higher than MOMA's (Figure 1A). Figure 1B,C compares MOMA and ROOM flux prediction accuracy.

Furthermore, for the high-glucose growing condition, ROOM found a feasible solution in which only the fluxes on an alternative pathway for the knocked-out reactions were significantly modified (pathway v16-v14-v15 in Figure SF1 in the Supp. data). Indeed, in the high-glucose growing condition, experimental data show that the above 3 fluxes have the largest flux deviations between wild-type and mutated strains (Emmerling *et al.*, 2002). MOMA predicts smaller flux changes along this pathway, reflecting the fact that minimization of Euclidian distance may not allow large changes.

### Lethality predictions

A Large-scale evaluation of In-Silico gene deletion for the *Saccharomyces cerevisiae* was reported in Förster *et al.* (2003). FBA was used to compute maximized growth rate for 555 knockouts. A gene is considered lethal if the respective mutant maximal growth rate is below 5% of the wild-type growth rate. FBA correctly predicts lethality for 89.6% of the genes. We examined both MOMA and ROOM on this model which offers an opportunity to study a relatively large knockout dataset.

Genes predicted as lethal by FBA will also be predicted lethal by both MOMA and ROOM since the maximum growth rate computed by FBA can only decrease in MOMA and ROOM. Running MOMA over the FBA-viable-predicted genes, gave significantly worse prediction than FBA. Specifically, out of the FBA-viable-predicted genes, MOMA falsely predicts 50 genes to be lethal (Table ST3, Supp. data). In contradiction, the results of running ROOM on the set of FBA-viable-predicted genes completely agree with the FBA's predictions.

We turn to closely examine the 50 viable genes correctly predicted by ROOM (and FBA) and falsely predicted lethal by MOMA. As in the *E. coli* high-glucose experiment, for all 50 genes, ROOM finds a short alternative pathway to replace the zero constrained reactions. It is MOMA's failure to recognize these alternative pathways for each of the 50 genes which leads it to a wrong solution (this time a lethal one).

## EXTENSIONS

As discussed earlier, short alternative pathways are sometimes used by mutated strains to re-route flux. It thus seems reasonable to explicitly include alternative pathways or other structural network properties, such as distance from the deleted reaction, in constraint based models. To do this, a variant of ROOM can assign different costs to significant flux change in different reactions. Specifically, the objective function can be changed to minimize $\sum_{i=1}^{m} c_i y_i$, where $c_i$ denotes the cost for a change in flux $i$. For example, to account for alternative pathways, each reaction that is on an alternative pathway can be assigned a cost that is correlated with the length of the pathway. Other reactions will be assigned a higher cost. In fact, we found that assigning higher costs to significant flux changes from zero using the ROOM variant, results in slightly better predictions. This presumably reflects a higher adaptation cost that is required for on-regulating a gene which is not expressed in the wild-type strain.

**Acknowledgments:** We are grateful to Daniel Segre for his help.

## REFERENCES

Emmerling,M, Dauner,M, Ponti,A, Fiaux,J, Hochuli,M, Szyperski,T, Wuthrich,K, Bailey,J.E and Sauer,U (2002) Metabolic fluxresponses to pyruvate kinase knockout in Escherichia coli. *J Bacteriol.*, **184(1)**, 64–154.

Förster,J, Famili,I, Palsson,B.O and Nielsen,J (2003) Large-scale evaluation of in silico genedeletions in Saccharomyces cerevisiae. *OMICS*, **7(2)**, 193–202.

de Jong,H (2002) Modeling andsimulation of genetic regulatory Systems: A literature review. *J. Comput Biol*, **9(1)**, 69–105.

Reed,J.L and Palsson,B.O (2003) Thirteen Years of Building Constraint-Based InSilico Models of Escherichia coli. *J. Bacteriol*, **185(9)**, 2692–2699.

Segre,D, Vitkup,D and Church,G.M (2002) Analysis ofoptimality in natural and perturbed metabolic networks. *PNAS*, **99(23)**, 15112–15117.