Is *OptKnock* a reliable strategy for desirable mutants?

Meltem Apaydin[1]
[1]Dept. of Electrical & Computer Engineering
Texas A&M University
College Station, TX, USA
ma193746@tamu.edu

Bo Zeng[2]
[2]Dept. of Industrial Engineering
University of Pittsburg
Pittsburg, PA, USA
bzeng@pitt.edu

Xiaoning Qian[1,3]
[3]Center for Bioinformatics & Genomic Systems Engineering
Texas A&M University
College Station, TX, USA
xqian@ece.tamu.edu

Abstract

Flux balance analysis (FBA) has enabled the development of computational methods for predicting optimal knockout strategies to genetically engineer microbial strains for desirable behavior, such as optimal biochemical overproduction for alternative energy sources. Many of these existing methods are based on bi-level optimization formulations to maximize the desired biochemical overproduction at the outer-level while modeling cell survival as the inner-level optimization problem, for example, by maximizing cell growth as in the seminal *OptKnock*.

Nevertheless, optimal knockout strategies derived in such a bi-level optimization framework may be heavily depending on the closeness and robustness of the inner-level optimization model in capturing actual cell survival states. We investigate how reliable the knockout strategies derived by *OptKnock* are, considering two critical but overlooked factors: (i) the surviving mutant of the inner-level optimization model may not be well-defined, i.e., it is not unique; and (ii) we cannot guarantee that the nature always cooperates with the human desire to select the microbial strains that produce maximum biochemical products among surviving mutants. We present our study in a core *E. coli* metabolic network and show that the knockout solutions from *OptKnock* could be of arbitrarily poor performance. Then, we revamp *OptKnock* through a novel pessimistic bi-level optimization framework, which considers the non-unique and non-cooperative issues and potential modeling errors. Through computing pessimistic knockout solutions and benchmarking with those from *OptKnock*, we observe that they are more reliable and perform significantly better. We believe that the proposed pessimistic bi-level optimization framework will help identify more practical and robust knockout strategies.

1 Introduction

Metabolic engineering of microbial strains has been studied extensively for targeted biochemical overproduction that may benefit human society, for example, in different energy-related and pharmaceutical applications [6, 8, 2, 10, 9]. Due to the demanding experimental cost and time to test different microbial strains *in vivo*, computational methods based on genome-scale dynamic analysis at steady states, such as Flux Balance Analysis (FBA) in metabolic networks, have been developed for *in silico* prediction of useful knockout strategies for beneficial mutants [15, 5]. Many of existing algorithms are based on bi-level optimization [13, 14, 3, 12, 7, 4], such as *OptKnock* [3, 7, 4].

All these bi-level optimization formulations have inherent assumptions that the nature will always cooperate with the human desire to select the mutants that serve the human society the best. However, in practice, it is often the case that the knockout mutant that survives may not always fit the best with the human desire. In addition, these computational methods always depend on certain model assumptions of cell survival, which are approximations

to the real-world situations. Therefore, it is natural to ask how robust these derived knockout strategies are. In this work, we investigate the robustness of the seminal *OptKnock* and show that it is necessary to have pessimistic bi-level optimization in metabolic engineering to derive practical knockout strategies.

2 *OptKnock*: Model and Evaluation

2.1 The Bi-level *OptKnock* Model

FBA [15, 5] has enabled *in silico* genome-scale manipulation of metabolism. Based on such a steady-state dynamic model constrained by stoichiometry for balanced production and consumption fluxes, computational methods based on bi-level optimization have demonstrated their potential for predicting beneficial genetically engineered microbial strains [13, 14, 3, 12, 7, 4]. The arguably best-known method in this category—*OptKnock*—searches for potential mutations (gene or reaction knockouts for example) that can accomplish optimal biochemical overproduction (the outer-level optimization) and, at the same time, optimize or maintain cellular survival objectives (the inner-level optimization), modeled by the maximum biomass growth at steady states. The mathematical programming formulation of *OptKnock* is as follows [3]:

$$
\begin{aligned}
\max_{y_j} \quad & \nu_{chemical} \\
\text{s.t.} \quad & \max_{\nu_j} \quad v_{biom} \\
& \text{s.t.} \quad \sum_j S_{ij}\nu_j = 0, \forall i; \\
& \qquad \nu_{glc} = \nu_{glc\_uptake}; \quad \nu_{biom} \geqslant \nu_{biom}^{target} \\
& \qquad \nu_j^{min}.y_j \leqslant \nu j \leqslant \nu_j^{max}.y_j, \forall j; \\
& \sum_j (1 - y_j) \leqslant K; \qquad y_j = \{0, 1\}\,.
\end{aligned}
$$

In this model, $\mathbf{y}$ and ν denotes the outer-level binary decision variables for possible knockout strategies and inner-level continuous decision variables for resulting reaction fluxes, respectively. The goal of *OptKnock* is to maximize desired biochemical production target $v_{chemical}$, allowing K possible knockout reactions. The inner-level optimization problem is to maximize the flux value for biomass production v_{biom}, modeling cell survival objectives. This model is again based on steady-state analysis with FBA stoichiometry constraints that the weighted sum of fluxes based on stoichiometric coefficients S for each metabolite is 0: $\sum_j S_{ij}v_j = 0$. In addition, depending on the availability of nutrients or the maximal fluxes that can be supported by enzymatic pathways [13], v_j^{min} and v_j^{max} are the lowest and highest possible reaction fluxes respectively for the reaction v_j. The glucose consumption rate v_{glc} is often set to a fixed value v_{glc_uptake}. As the biomass growth is a linear objective function of metabolic reaction fluxes, the strong duality of the inner-level optimization helps to convert the original bi-level optimization problem into a Mixed Integer Linear Programming (MILP) , which can be solved efficiently for large-scale metabolic networks [3, 7, 4].

Table 1: Knockout strains derived by *OptKnock* on the core *E.coli* metabolic network

K	Knockouts	Ideal Succi	Biomass
3	g6p → 6pg + nadph, mal → pyr + co2 + nadph, nadh → nadph	110.179	6.972
4	g6p → 6pg + nadph, mal → pyr + co2 + nadph, nadh → nadph, 3pg + glu → ser + akg + nadh	123.314	5
5	g6p → 6pg + nadph, 3pg + glu → ser + akg + nadh, nadh → nadph, glyc → glyc(ext), ac(ext) →	129.786	5.470

Following [3], we have derived the optimal knockout strategies for a core *E. coli* metabolic network model [1] with 74 chemicals and 75 reactions. Succinate is set as the targeted bioproduction, the glucose uptake rate is set at a fixed value of 100 mmol/gDW.hr and the minimum biomass is set as 5 mmol/gDW.hr. The experiments are based on the aerobic condition for *E. coli* metabolic model. We allow K =3 to 5 knockouts. Table 1 gives the knockout strategies from *OptKnock*. It is clear that the predicted knockout strains give high Succinate production rates. However, as pointed out in [12], the inner-level biomass rates of the derived mutants are very close to the minimally allowed rate in all three cases, which implies that they may not survive easily in practice. Given that, we are motivated to further investigate the robustness of *OptKnock* solutions.

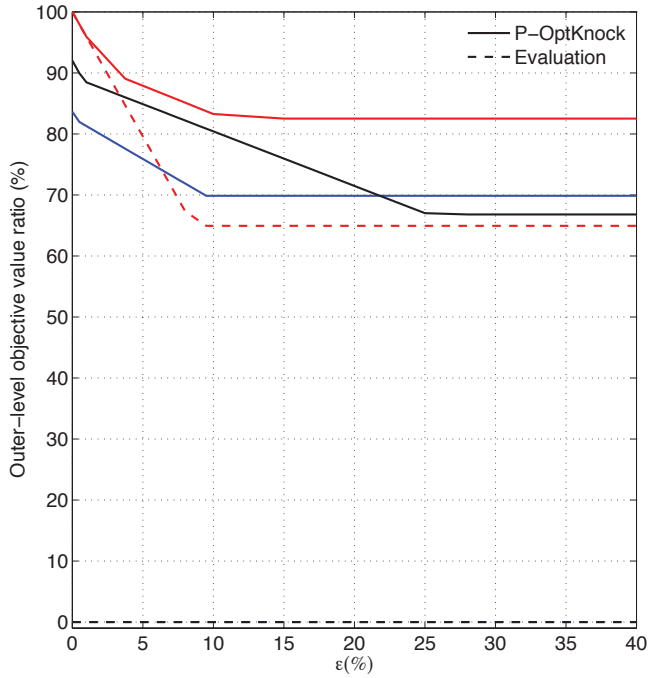

Figure 1: The ratio (%) between outer-level objective values of *P-OptKnock* and "optimistic" *OptKnock* plotted as solid lines with increasing ε (Blue: $K = 3$, Black: $K = 4$, Red: $K = 5$). The ratios for pessimistic Succinate rate evaluation of *OptKnock* also are given as dashed lines (Note that the rates are 0 for both K =3 and 4.)

2.2 How reliable are *OptKnock* strategies?

The bi-level structure in [3, 11] clearly provides an effective approach to model the interacting objectives of the outer-level knockout implementer (i.e. human) and the inner-level microbial cells. Nevertheless, two critical factors have been overlooked, which may seriously affect the performance of *OptKnock* solutions in practice:
(i) For a given outer-level knockout strategy $\mathbf{y}^*$, the surviving mutant output from the inner-level optimization model is not well-defined, i.e., it may not give unique reaction fluxes.
(ii) Given non-unique surviving mutants, we cannot guarantee that the nature always select the mutant that produces the maximum desirable biochemical products.
To evaluate the impact of those two factors on *OptKnock* solutions, we have performed the following experiment: For a given K, we compute an optimal solution $\mathbf{y}^*$ to *OptKnock* and the corresponding inner-level optimal value $\theta(\mathbf{y}^*)$. Then, we formulate and solve the following *Evaluation* problem:

$$
\begin{aligned}
\min_{\nu_j} \quad & \nu_{chemical} \\
\text{s.t.} \quad & \sum_j S_{ij}\nu_j = 0, \forall_i; \\
& \nu_{glc} = \nu_{glc\_uptake}; \\
& \nu_{biom} \geqslant \nu_{biom}^{target}; \\
& \nu_j^{min}.y_j \leqslant \nu_j \leqslant \nu_j^{max}.y_j, \forall_j; \\
& v_{biom} \geqslant \theta(\mathbf{y}^*(1-\varepsilon)).
\end{aligned}
$$

Obviously, this *Evaluation* problem computes the performance of the least favorable surviving mutant for the given $\mathbf{y}^*$. For the derived *OptKnock* mutants reported in Table 1, we found that: When K =3 and 4, the derived "optimal" mutants actually have 0 Succinate rates, if the cells do not cooperate with the knockout implementer (see Figure 1 when $\varepsilon = 0$). It highlights that nature may do very opposite to human desire and *OptKnock* strategies may not work in practice.

3 Pessimistic Bi-level Optimization

3.1 Pessimistic OptKnock

OptKnock is based on the regular Optimistic Bi-level Optimization (OBO), which assumes that the inner-level decision variables are always cooperative by selecting the inner-level solutions in favor of the outer-level optimization problem. However, in many practical scenarios such as designing mutant strains in metabolic engineering, it is often the case that the inner-level decision variables behave non-cooperatively by taking a solution against the desire of the outer-level optimization. For such cases, it would be more reasonable to revamp *OptKnock* model by considering its Pessimistic Bi-level Optimization (PBO) version for modeling and computing reliable solutions. In the following, we provide such a pessimistic formulation, i.e., *P-OptKnock*:

$$\begin{aligned}
&\max_{y_j} \quad \min_{\nu_j \in S(y_j)} \quad \nu_{chemical} \\
&\text{s.t.} \ \sum_j (1-y_j) \leqslant K; \qquad y_j = \{0,1\} \\
&\qquad S(y_j) = \text{argmax}\{v_{biom} : \sum_j S_{ij}\nu_j = 0, \forall i; \\
&\qquad\qquad \nu_{glc} = \nu_{glc_uptake}; \\
&\qquad\qquad \nu_{biom} \geqslant \nu_{biom}^{target};\ \nu_j^{min}.y_j \leqslant \nu_j \leqslant \nu_j^{max}.y_j, \forall j\}.
\end{aligned}$$

Computing a PBO problem is recognized very challenging. Nevertheless, a recent solution scheme developed in [16] provides us a capable tool to address this challenge. Specifically, by replicating variables and constraints in the inner-level and introducing a new constraint in the inner-level, we obtain a tight bi-level relaxation to *P-OptKnock*:

$$\begin{aligned}
&\max_{y_j, \overline{\nu}} \quad \nu_{chemical} \\
&\text{s.t.} \quad \sum_j (1-y_j) \leqslant K,\ y_j = \{0,1\},\ \sum_j S_{ij}\overline{\nu}_j = 0, \forall i \\
&\quad \overline{\nu}_{glc} = \nu_{glc_uptake},\ \overline{\nu}_{biom} \geqslant \nu_{biom}^{target},\ \nu_j^{min}.y_j \leqslant \overline{\nu}_j \leqslant \nu_j^{max}.y_j, \forall j, \\
&\quad \nu_j \in \text{argmin}\{\nu_{chem} : \sum_j S_{ij}\nu_j = 0, \forall i \\
&\qquad\qquad \nu_{glc} = \nu_{glc_uptake},\ \nu_{biom} \geqslant \nu_{biom}^{target},\ \nu_j^{min}.y_j \leqslant \nu_j \leqslant \nu_j^{max}.y_j, \forall j\} \\
&\qquad\qquad \nu_{biom} \geqslant \overline{\nu}_{biom}(1-\varepsilon)\},
\end{aligned}$$

where ε is introduced to reflect the cell response or modeling error. By varying ε, we can investigate the sensitivity of $\mathbf{y}$ with respect to such type of errors. According to [16], the last constraint ensures that (when $\varepsilon = 0$) ν_j is an optimal solution of the inner-level, i.e. $\nu_j \in S(y_j)$. The aforementioned bi-level problem can be computed by its single-level mixed integer program (MIP) reformulation through Karush-Kuhn-Tucker (KKT) conditions. Then, using the *Relaxation-and-Correction* scheme in [16], the original *P-OptKnock* can be readily solved.

3.2 Experimental results

We have derived *P-OptKnock* knockout strains on the same core *E. coli* network when we achieve stable knockout solutions with increasing ε, given in Table 2. It is clear that the derived *P-OptKnock* mutants have much higher biomass rates compared to *OptKnock* knockout strains in Table 1 when K =3 and 4, though the Succinate rates are lower than the corresponding values if the cooperative assumption holds. This indicates that *P-OptKnock* knockout mutants are more reliable and have a higher chance of surviving in practice. When K =5, we find that both *P-OptKnock* and *OptKnock* knockout mutants are similar as the surviving mutants are more restricted with the increasing number of knockout reactions.

Furthermore, Figure 1 shows the ratio of *P-OptKnock* optimal Succinate rates and the pessimistic Succinate rate evaluations of *OptKnock* with respect to the "optimistic" *OptKnock* Succinate rates for different numbers of knockouts and different levels of relaxation reflected by the ε value ($0 \leqslant \varepsilon < 0.4$), taking care of realistic cell responses and modeling errors.

It is clear from both the table and figure that *P-OptKnock* achieves Succinate overproduction with more robust knockout solutions. Compared with *OptKnock* solutions, which may give arbitrarily low Succinate overproduction if the cooperative assumption is not valid, *P-OptKnock* can derive more practical and robust knockout strains in practice.

Table 2: Knockout strains derived by *P-OptKnock* on the core *E.coli* metabolic network

K	Knockouts	Succi	Biomass
3	g6p → 6pg + nadph, oac + accoa → cit, ac → ac(ext)	76.971	14.362
4	g6p → 6pg + nadph, oac + accoa → cit, suc → fum + fadh2 , ac(ext) →	82.362	14.362
5	g6p → 6pg + nadph, mal → pyr + co2 + nadph, 3pg + glu → ser + akg + nadh, fadh2 + 0.5o2 → 2atp, nadh → nadph	107.075	5

4 Conclusions

We investigate the robustness of *OptKnock* solutions for designing knockout mutants for biochemical overproduction and observe that its knockout solutions could be of arbitrarily poor performance. Then, we propose and compute a novel pessimistic bi-level optimization framework *P-OptKnock* to derive reliable knockout strains. By benchmarking both *OptKnock* and *P-OptKnock* on a core *E. coli* metabolic network, we demonstrate that the pessimistic bi-level optimization solutions are indeed more reliable and has the promising potential of identifying practical and robust knockout strategies.

References

[1] M R Antoniewicz, D F Kraynie, L A Laffend, GL Joanna, K K Joanne, and Gregory S. Metabolic flux analysis in a nonstationary system: Fed-batch fermentation of a high yielding strain of e. coli producing 1,3-propanediol. *Genome Res*, 9(3):277–292, 2007.

[2] C Broa, B Regenberga, J Förster, and J Nielsen. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metabolic Engineering*, 8:102–111, 2006.

[3] AP Burgard, P Pharkya, and CD Maranas. OptKnock: A bi-level programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, 84:647–657, 2003.

[4] A Chowdhury, AR Zomorrodi, and CD Maranas. Bilevel optimization techniques in computational strain design. *Computers and Chemical Engineering*, 72:363–372, 2015.

[5] JS Edwards and BO Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. In *Proc Natl Acad Sci USA*, volume 97, pages 5528–5533, 2000.

[6] MA Haro and V de Lorenzo. Metabolic engineering of bacteria for environmental applications: Construction of Pseudomonas strains for biodegradation of 2-chlorotoluene. *Journal Biotechnol*, 85:103–113, 2001.

[7] J Kim, JL Reed, and GM Church. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS ONE*, 6(9):e24162, 2011.

[8] Jing Lu, Con Sheahan, and Pengcheng Fu. Metabolic engineering of algae for fourth generation biofuels production. *Energy Environ Sci*, 4:2451–2466, 2011.

[9] J M Luengo, B Garcia, A Sandoval, G Naharro, and E R Olivera. Bioplastics from microorganisms. *Current Opinion in Microbiology*, 6:251–260, 2003.

[10] K Ohta, D S Beall, J P Mejia, K T Shanmugam, and L O Ingram. Metabolic engineering of Klebsiella oxytoca M5A1 for ethanol production from xylose and glucose. *Applied and Environmental Microbiology*, 57:2810–2815, 1991.

[11] Priti Pharkya, Anthony P Burgard, and Costas D Maranas. Exploring the overproduction of amino acids using the bilevel optimization framework optknock. *Biotechnology and bioengineering*, 84(7):887–899, 2003.

[12] S Ren, B Zeng, and X Qian. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics*, 14(Suppl 2):S17, 2013.

[13] D Segre, D Vitkup, and GM Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 99:15112–15117, 2002.

[14] T Shlomi, O Berkman, and E Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA*, 102:7695–7700, 2005.

[15] A Varma and BO Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology*, 12:994–998, 1994.

[16] B. Zeng. Easier than we thought-a practical scheme to compute pessimistic bilevel optimization problem. Technical report, University of Pittsburgh, available in optimization-online.org, 2015.