BMC Genomics

Robust mutant strain design by pessimistic optimization

Meltem Apaydin[1], Liang Xu[2], Bo Zeng[2] and Xiaoning Qian[1]*

Abstract

Background: Flux Balance Analysis (FBA) based mathematical modeling enables *in silico* prediction of systems behavior for genome-scale metabolic networks. Computational methods have been derived in the FBA framework to solve bi-level optimization for deriving "optimal" mutant microbial strains with targeted biochemical overproduction. The common inherent assumption of these methods is that the surviving mutants will always cooperate with the engineering objective by overproducing the maximum desired biochemicals. However, it has been shown that this optimistic assumption may not be valid in practice.

Methods: We study the validity and robustness of existing bi-level methods for strain optimization under uncertainty and non-cooperative environment. More importantly, we propose new pessimistic optimization formulations: P-ROOM and P-OptKnock, aiming to derive robust mutants with the desired overproduction under two different mutant cell survival models: (1) ROOM assuming mutants have the minimum changes in reaction fluxes from wild-type flux values, and (2) the one considered by OptKnock maximizing the biomass production yield. When optimizing for desired overproduction, our pessimistic formulations derive more robust mutant strains by considering the uncertainty of the cell survival models at the inner level and the cooperation between the outer- and inner-level decision makers. For both P-ROOM and P-OptKnock, by converting multi-level formulations into single-level Mixed Integer Programming (MIP) problems based on the strong duality theorem, we can derive exact optimal solutions that are highly scalable with large networks.

Results: Our robust formulations P-ROOM and P-OptKnock are tested with a small *E. coli* core metabolic network and a large-scale *E. coli* iAF1260 network. We demonstrate that the original bi-level formulations (ROOM and OptKnock) derive mutants that may not achieve the predicted overproduction under uncertainty and non-cooperative environment. The knockouts obtained by the proposed pessimistic formulations yield higher chemical production rates than those by the optimistic formulations. Moreover, with higher uncertainty levels, both cellular models under pessimistic approaches produce the same mutant strains.

Conclusions: In this paper, we propose a new pessimistic optimization framework for mutant strain design. Our pessimistic strain optimization methods produce more robust solutions regardless of the inner-level mutant survival models, which is desired as the models for cell survival are often approximate to real-world systems. Such robust and reliable knockout strategies obtained by the pessimistic formulations would provide confidence for in-vivo experimental design of microbial mutants of interest.

Keywords: Strain optimization, Pessimistic bi-level optimization, Stoichiometric models

*Correspondence: xqian@ece.tamu.edu
[1]Dept. of Electrical and Computer Engineering, Texas A&M University, 77843
College Station, USA
Full list of author information is available at the end of the article

Introduction

Whole-genome high-throughput profiling techniques have enabled the systematic study of biological systems at the genome scale [1, 2]. In particular, systems models and computational methods for analyzing and controlling genome-scale metabolic networks have greatly advanced the field of metabolic engineering [3]. With the better understanding of the systems behavior of microbial metabolism, metabolic engineering based on the metabolic network models can help predict metabolic phenotypes [4, 5] and derive engineering strategies for strain design by manipulating the native microbial pathways to produce chemicals of commercial and biomedical benefits [6–11].
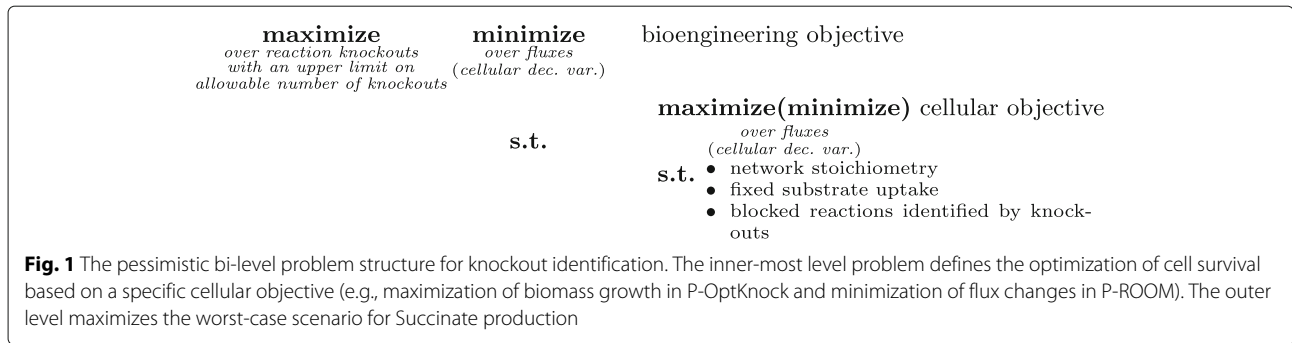
Mathematical modeling of metabolism often focuses on steady-state behavior, especially when long-term metabolic dynamics is of interest, as accurate reconstruction of genome-scale kinetic models is challenging when considering the large model space and parameter uncertainties. Constraint-based approaches based on the reaction stochiometry, notably Flux Balance Analysis (FBA) by Linear Programming (LP) formulations, study genome-scale dynamics by mass-balance equations at steady states to understand and predict macro-level microbial behavior in the presence of perturbation, for example caused by mutations or environmental changes [12–15]. By adding thermodynamic and flux capacity constraints, FBA often models steady-state behavior by assuming that the cell growth flux needs to be maximized based on biomass composition [16].

Many computational approaches have been proposed in this computational framework for *in silico* prediction of potentially feasible metabolic phenotypes and evaluation of theoretical limits of metabolic performance after knocking out certain genes or reactions for strain design [14, 16]. Metabolic engineering with such computationally derived knockout strategies has shown to be effective to achieve overproduction of biochemicals of interest [17]. Those strain optimization methods are generally modeled as bi-level optimization problems that seek for maximum overproduction of a desired biochemical at the outer level under the condition that mutant cells are still alive, modeled as the inner-level optimization problem. For example, OptKnock [17], ROOM (Regulatory On/Off Minimization) [18], and MOMAKnock [19] all fall under this category with different mutant cell survival models at the inner level detailed in the "Background" section.

These existing bi-level formulations for microbial strain design have inherent assumptions that nature will always cooperate with the human desire to select the mutants that serve the outer-level engineering objective the best, namely they assume that a cooperative environment exists between the outer-level (human) and inner-level

(microbial strain) decision makers. However, in practice, surviving mutant strains may not always fit the best with the engineering goal. In addition, the model assumptions for the cell survival at the inner level are often the approximations to the real-world scenarios, which may result in the derived knockout strategies not overproducing the predicted amount of desirable biochemicals. Therefore, the robustness and viability of these predicted knockout strategies may need to be re-evaluated. For example, if the perturbed strains do not cooperate with the knockout implementer, they may do very opposite to the engineering objective and these optimistic bi-level strategies may not work in practice. In [20], we have shown that OptKnock-derived knockout strategies may produce the desired chemicals at levels much lower than the expected ones, when the aforementioned two assumptions are relaxed.

In this paper, we innovate a computational framework to predict robust knockout strategies for overproduction of desirable chemicals with different inner-level mutant cell survival models. Specifically, we develop a pessimistic optimization formulation [21] that considers the uncertainty of the cell survival models at the inner level and the cooperation between the outer- and inner-level optimization decisions. Figure 1 illustrates our pessimistic bi-level optimization framework for identifying the knockouts to have the maximum overproduction of a desired chemical as a bioengineering objective under the worst-case or least-favorable scenario. The inner-level cellular model has a competing objective function to maintain the cell survival. According to the cellular model, different objective functions can be chosen. We study such pessimistic formulations with both the cell survival criteria of (1) minimization of significant flux changes with respect to wild-type flux values as in ROOM [18]; and (2) maximization of biomass growth as in OptKnock [17]. Based on the corresponding pessimistic formulations, which we call P-ROOM and P-OptKnock respectively, we will optimize the desired overproduction in the least favorable situation (i.e., the non-cooperative situation) and investigate the sensitivity of resulting knockouts with respect to cell response and uncertainty introduced by the inner-level cell survival models. Since the inner-level cellular model is linear given the outer-level binary decision variables, we can employ the LP duality theory and convert the pessimistic bi-level optimization problem to a single-level MIP problem by applying two dual transformations as described in Methods. Experimental results on both a core *E. coli* metabolic network and a genome-scale iAF1260 network [22] show that our pessimistic formulations can generate more robust knockout strategies compared to the existing bi-level optimization methods for strain design. Regardless of the choice of inner-level surrogate functions for cell survival, P-ROOM

Fig. 1 The pessimistic bi-level problem structure for knockout identification. The inner-most level problem defines the optimization of cell survival based on a specific cellular objective (e.g., maximization of biomass growth in P-OptKnock and minimization of flux changes in P-ROOM). The outer level maximizes the worst-case scenario for Succinate production

and P-OptKnock produce consistent and stable knockout strategies for the targeted biochemical overproduction as the formulations are specifically designed to take care of the model uncertainty.

Background

Before we present pessimistic optimization formulations to derive robust knockout strategies for strain design, we briefly summarize the background and important mathematical notations for FBA and the bi-level optimization models for deriving knockouts, including OptKnock [17] and ROOM [18].

FBA

FBA is an LP problem for the analysis of stoichiometric-based metabolic network models of involved reactions at the genome scale. A linear objective function is minimized or maximized subject to mass-balance, thermodynamic and capacity constraints, with respect to reaction fluxes in a vector form **v**. The LP formulation of FBA can be expressed as:

$$\max(\min)_{\mathbf{v}} \quad \mathbf{c}^T \mathbf{v}$$
$$\text{s.t.} \quad \mathbf{Sv} = \mathbf{0}, \quad \mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}.$$

Under the steady-state assumption, mass-balance constraints constitute a system of linear equations where the weighted sum of fluxes, based on stoichiometric coefficients given in a matrix form **S**, is 0. Thermodynamic and capacity constraints are defined as lower and upper bounds on reaction fluxes. In the FBA framework, maximization of biomass growth is often adopted as the objective function for modeling cell survival, where **c** becomes a vector with all values of 1 for the reactions corresponding to the biomass formation.

OptKnock

FBA enables efficient computational design of beneficial genetically engineered microbial strains. The pioneering work by the authors of [17], OptKnock, searches for potential genetic perturbations (e.g., gene or reaction

knockouts) for redirection of metabolic flux to overproduce desired biochemicals and maintain cellular growth. This bi-level optimization problem captures the engineering objective at the outer level (e.g., to maximize the overproduction) while the inner-level problem models a cellular fitness objective (e.g., maximization of biomass growth). The mathematical programming formulation of OptKnock can be expressed as follows:

$$\max_{z_j \in \{0,1\}, \forall j \in J} \quad v_{chemical}$$

$$\text{s.t.} \quad \max_{v_j \in \mathbb{R}, \forall j \in J} \quad v_{biom}$$

$$\text{s.t.} \quad \sum_{j \in J} S_{ij} v_j = 0, \quad \forall i \in I$$

$$v_{glc} = v_{glc\_uptake}, \quad v_{biom} \geq v_{biomass}^{target} \tag{1}$$

$$v_j^{min} z_j \leq v_j \leq v_j^{max} z_j, \quad \forall j \in J$$

$$\sum_{j \in J} (1 - z_j) \leq K,$$

where I and J are the sets of metabolites and reactions, respectively. In this model, v_j represents the flux of reaction j in J. Each element of the matrix **S** is the stoichiometric coefficient S_{ij} of metabolite i in reaction j. The outer-level binary decision variable z_j is 1 if the corresponding reaction flux v_j is active and 0 otherwise. The overproduction of a chemical of interest is maximized at the outer level allowing K possible knockout reactions. The inner-level cell survival model is based on steady-state analysis with FBA constraints. 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 for the reaction j, respectively. The glucose consumption rate v_{glc} is often set to a fixed value as v_{glc_uptake}, and $v_{biomass}^{target}$ is the minimum level of biomass production. As the biomass growth is a linear objective function of metabolic reaction fluxes, strong duality of the inner-level optimization helps to convert the original bi-level optimization problem into a Mixed Integer Linear Programming (MILP) problem,

which can be solved efficiently for large-scale metabolic networks [17, 23].

ROOM

Observed experimental flux values reported in [24] showed that maximizing biomass reaction flux as a surrogate function for the most probable physiological state of the metabolic model is indeed effective for wild-type strains as they have been exposed to long-term evolutionary pressure. However, it may not be valid for the engineered mutants without such a long-term progress. Therefore, it calls for other realistic cell survival models for mutants. ROOM [18] is one of such models, in which binary variables y_j's are introduced to capture significant (up/down regulated) or insignificant reaction flux changes. Instead of maximizing the biomass growth, ROOM [18] aims to minimize the number of significant flux changes. Relaxing the binary variables y_j's in the original MILP formulation leads to the LP variant of ROOM:

$$\max_{z_j \in \{0,1\}, \forall j \in J} v_{chemical}$$

$$\text{s.t.} \quad \min_{v_j \in \mathbb{R}, y_j \in \mathbb{R}, \forall j \in J} \sum_{j \in J} y_j$$

$$\text{s.t.} \quad \sum_{j \in J} S_{ij} v_j = 0, \forall i \in I$$

$$v_{glc} = v_{glc\_uptake}, \quad v_{biom} \geqslant v_{biomass}^{target}$$

$$v_j^{min} z_j \leqslant v_j \leqslant v_j^{max} z_j, \forall j \in J \qquad (2)$$

$$v_j - y_j(v_j^{max} - w_j) \leq w_j, \forall j \in J$$

$$v_j - y_j(v_j^{min} - w_j) \geq w_j, \forall j \in J$$

$$0 \leq y_j \leq 1, \forall j \in J$$

$$\sum_{j \in J}(1 - z_j) \leqslant K,$$

in which w_j is the wild-type flux values that can be solved by FBA. In this bi-level optimization model (2), the inner-level cellular objective is to minimize the flux changes. The inequalities with y_j's constrain the flux values not to deviate from the wild-type flux values. We again can employ the strong duality condition to convert the bi-level optimization problem into a MILP to solve (2).

Methods

The above optimistic bi-level optimization methods effectively model the interacting objectives of the outer-level knockout implementer and the inner-level microbial cells. However, the effectiveness of these optimistic formulations depends on the inherent assumption that the outer-level engineering objective and inner-level cellular fitness function behave cooperatively by selecting the inner-level solutions in favor of the outer-level optimization problem. In practice, when the inner-level problem either does not faithfully reflect cellular fitness objectives or has non-unique solutions, there is no guarantee that the cell response will be in cooperation with the engineering objective to maximize the desirable biochemical product formation as we recently found in [20]. The main questions that we would like to address here are:

1. *How robust are the derived knockout strategies based on these optimistic bi-level optimization formulations?*
2. *Does the robustness depend on how the inner-level optimization problem approximates mutant cell objectives?*
3. *Can we formulate new optimization problems that derive more robust solutions?*

We answer the first two questions by allowing the inner-level problem to take non-optimistic solutions as we have done in [20] to evaluate the knockout strategies derived by OptKnock and ROOM. More importantly, we propose a novel pessimistic bi-level optimization framework in this paper to derive robust knockout strategies, which consider the uncertainty introduced by the inner-level models from the aforementioned non-cooperative and non-unique issues.

We now present our pessimistic optimization formulations for deriving robust mutant strains and provide the detailed derivation of the optimization solutions to our pessimistic models, in which we maximize the desired overproduction for the worst-case scenario due to uncertainty and non-cooperative environment with different inner-level mutant cell survival models.

Pessimistic bi-level optimization

As discussed, due to the inner-level model uncertainty and non-cooperative issues, the existing optimistic bi-level strain optimization methods may not be able to produce reliable results when there exist non-unique solutions to the inner-level problem. In order to overcome this limitation, we propose a pessimistic optimization framework to optimize for the worst-case scenario under uncertainty and non-cooperative environment. The mathematical programming formulation of pessimistic bi-level optimization can be represented in the following general form [21]:

$$\max_{\mathbf{x} \in \mathbb{X}} \min_{\mathbf{y} \in Y(\mathbf{x})} \mathbf{F}(\mathbf{x}, \mathbf{y})$$

$$\text{s.t.} \quad \mathbf{G}(\mathbf{x}) \leq 0$$

$$Y(\mathbf{x}) = \arg\min(\max)_{\mathbf{y}}\{\mathbf{f}(\mathbf{x}, \mathbf{y}) : \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq 0, \mathbf{y} \in \mathbb{Y}\},$$

$$(3)$$

where $\mathbf{x}$ and $\mathbf{y}$ are the outer-level and inner-level decision variables with the corresponding feasible sets $\mathbb{X}$ and $\mathbb{Y}$. In the above formulation, $\mathbf{F}$ represents the outer-level objective function, $\mathbf{F} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$, and $\mathbf{f}$ represents inner-level objective function, $\mathbf{f} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$. $\mathbf{G}$ and $\mathbf{g}$ represent the inequality constraint functions at the outer and inner levels, respectively. $Y(\mathbf{x})$ denotes the set of optimal solutions to the inner-level problem for a given $\mathbf{x}$, which may contain multiple elements. Solving such a pessimistic bi-level optimization problem is computationally very difficult. Even a linear optimistic bi-level problem, where both the outer-level and the inner-level problems are linear programs, is theoretically NP-hard [25]. In our study, to solve the more difficult pessimistic bi-level optimization problem, we first employ the LP duality theory based on the fact that the inner-most level problem is a linear program. It converts the pessimistic bi-level model into a standard bi-level problem, which enables us to fully make use of existing bi-level optimization algorithms. For example, we transform the standard bi-level model into a single level mixed integer programming (MIP) problem using the strong duality theory, which can be solved through the commercial MIP solver CPLEX. As those operations are rather simple, we make a challenging pessimistic bi-level problem practically solvable even for large-scale cases.

We mention that for P-ROOM (4), there are $|J|$ binary variables in the outer level, $2|J|$ continuous variables and $O(|I| + 5|J|)$ constraints in the inner level. By taking the dual of the inner-most problem, we convert (4) into a standard bi-level problem (7), whose inner-level problem has $O(|I| + 7|J|)$ variables and $O(|I| + 7|J|)$ constraints. Comparing to the traditional bi-level ROOM model (2), which has the identical outer-level problem, and $2|J|$ continuous variables and $O(|I| + 5|J|)$ constraints in the inner level, (7) does not have a much larger or more complicated structure. Hence, it can be expected that the computational complexity of (7) will not be drastically more than that of the traditional ROOM model. Moreover, our numerical study shows that the single level MIP reformulation of (7) for a larger-scale network can be computed by CPLEX in a reasonable time, which indicates the efficiency of our proposed solution strategy.

Inspired by this pessimistic optimization framework, we propose pessimistic models for mutant strain optimization under model uncertainty. We note that the notations appearing in the following pessimistic formulations have the same definitions as in the "Background" section.

Pessimistic strain optimization I: P-ROOM

In the following, we propose the pessimistic formulation based on the original ROOM model (2) in the same *minimax* flavor as in (3):

$$\max_{z_j, \forall j \in J} \min_{v_j, y_j \in Y(z_j), \forall j \in J} v_{chemical}$$

$$\text{s.t.} \sum_{j \in J} (1 - z_j) \leqslant K, \ z_j \in \{0, 1\}, \forall j \in J$$

$$Y(z_j) = \operatorname{argmin}_{v_j, y_j} \left\{ \sum_{j \in J} y_j : \sum_{j \in J} S_{ij} v_j = 0, \forall i \in I, v_{glc} = v_{glc\_uptake}, \right.$$

$$v_{biom} \geqslant v_{biomass}^{target}, \ v_j^{min} z_j \leqslant v_j \leqslant v_j^{max} z_j, \ v_j - y_j \left(v_j^{max} - w_j \right) \leq w_j,$$

$$\left. v_j - y_j \left(v_j^{min} - w_j \right) \geq w_j, \ 0 \leq y_j \leq 1, \forall j \in J \right\}.$$

$$(4)$$

Originally introduced in [21], the ϵ-approximation extension of the above pessimistic formulation is flexible when model uncertainty needs to be considered for bi-level decision making. Specifically, the ϵ-approximation of the pessimistic problem is to allow a proportional gap of ϵ for the inner-level objective function value from the optimal value that the actual knockout solutions would take. When $\epsilon = 0$, we have the original pessimistic formulations assuming that the inner-level models are faithful, but inner-level decision variables may act against outer-level engineering objectives. Higher ϵ values reflect that the inner-level mutant strains have a higher tolerance level for inner-level model uncertainty when a non-cooperative decision is taken against the outer-level engineering objectives. Such an ϵ-approximation of the pessimistic problem can be written as follows:

$$\max_{z_j : z_j \in \{0,1\}, \forall j \in J, \sum_{j \in J} (1 - z_j) \leqslant K} \min_{v_j, y_j, \forall j \in J} v_{chemical}$$

$$\text{s.t.} \sum_{j \in J} S_{ij} v_j = 0, \forall i \in I, v_{glc} = v_{glc\_uptake}, v_{biom} \geqslant v_{biom}^{target}$$

$$v_j^{min} z_j \leqslant v_j \leq v_j^{max} z_j, \ v_j - y_j \left(v_j^{max} - w_j \right) \leq w_j, \forall j \in J$$

$$v_j - y_j \left(v_j^{min} - w_j \right) \geq w_j, \ 0 \leq y_j \leq 1, \forall j \in J$$

$$\sum_{j \in J} y_j \leq \varphi(\mathbf{v}^*, \mathbf{y}^*)(1 + \epsilon),$$

$$(5)$$

where $\varphi(\mathbf{v}^*, \mathbf{y}^*)$ is the inner-level function value changing with respect to the inner-level decision variables $\mathbf{v}^*$ and $\mathbf{y}^*$. As the inner-level problem is LP, we can convert the problem into its dual representation by the strong duality theorem as in [17, 26], which is:

$$\max_{glc, u_i, \forall i \in I, \mu, a_j, \forall j \in J} glc.v_{glc\_uptake} + \mu_{biom} v_{biomass}^{target} + \sum_{j \in J} v_j^{min} \mu_{min, j} z_j -$$

$$\sum_{j \in J} v_j^{max} \mu_{max, j} z_j - \sum_{j \in J} \mu_{max2, j} w_j + \sum_{j \in J} \mu_{min2, j} w_j - \sum_{j \in J} a_j$$

$$\text{s.t.} \sum_{i \in I} S_{i, glc} u_i + glc + \mu_{min, glc} - \mu_{max, glc} - \mu_{max2, glc} + \mu_{min2, glc} = 0$$

$$\sum_{i \in I} S_{i, biom} u_i + \mu_{biom} + \mu_{min, biom} - \mu_{max, biom} - \mu_{max2, biom}$$

$$+ \mu_{min2, biom} = 0$$

$$\sum_{i \in I} S_{i,j} u_i + \mu_{min,j} - \mu_{max,j} - \mu_{max2,j} + \mu_{min2,j} = 0, \forall j \in J, j$$

$$\neq \text{biom, glc}$$

$$\mu_{max2,j} \left(v_j^{max} - w_j \right) - \mu_{min2,j} \left(v_j^{min} - w_j \right) - a_j \leq 1, \forall j \in J \quad (6)$$

$$glc, u_i \in \mathbb{R}, \forall i \in I \quad \mu_{biom}, \mu_{min,j}, \mu_{max,j}, \mu_{max2,j}, \mu_{min2,j}, a_j$$

$$\geq 0, \forall j \in J$$

where u_i denotes the dual variable associated with the mass-balance constraint, $\sum_j S_{ij} v_j = 0$ for metabolite i, glc is the dual variable associated with the glucose uptake constraint, μ_{biom} is the dual variable associated with the minimum biomass threshold constraint, and $\mu_{min,j}$ and $\mu_{max,j}$ are the dual variables for the two directions of the inequality constraint $v_j^{min} z_j \leq v_j \leq v_j^{max} z_j$ for each reaction j. Finally, $\mu_{min2,j}$ and $\mu_{max2,j}$ are the dual variables associated with the constraints $v_j - y_j \left(v_j^{min} - w_j \right) \geq w_j$ and $v_j - y_j \left(v_j^{max} - w_j \right) \leq w_j$, denoting the flux changes with respect to wild-type flux values, respectively, and a_j is the dual variable associated with the upper bound constraint on y_j. Note that the constraint $v_j^{min} z_j \leq v_j \leq v_j^{max} z_j$ links the outer-most level binary decision variables z_j and inner-level decision variables v_j.

By aggregating the inner-level constraints to the outer level with those introduced by its dual in (6) with the adopted ϵ-approximation, the final pessimistic formulation P-ROOM aims to solve the following max-min problem;

$$\max_{z_j : z_j \in \{0,1\}, \forall j \in J, \sum_{j \in J}(1 - z_j) \leq K} \quad \min_{v_j, y_j, glc, u_i, a_j, \mu, \forall i \in I, \forall j \in J} \quad v_{chemical}$$

$$\text{s.t.} \quad \sum_{j \in J} S_{ij} v_j = 0, \forall i \in I$$

$$v_{glc} = v_{glc\_uptake}, v_{biom} \geq v_{biomass}^{target}, v_j^{min} z_j \leq v_j$$

$$\leq v_j^{max} z_j, \forall j \in J$$

$$v_j - y_j \left(v_j^{max} - w_j \right) \leq w_j, v_j - y_j \left(v_j^{min} - w_j \right) \geq w_j, 0$$

$$\leq y_j \leq 1, \forall j \in J$$

$$\left(glc.v_{glc\_uptake} + \mu_{biom} v_{biomass}^{target} + \sum_{j \in J} v_j^{min} \mu_{min,j} z_j \right.$$

$$- \sum_{j \in J} v_j^{max} \mu_{max,j} z_j$$

$$\left. - \sum_{j \in J} \mu_{max2,j} w_j + \sum_{j \in J} \mu_{min2,j} w_j - \sum_{j \in J} a_j \right)(1 + \epsilon)$$

$$- \sum_{j \in J} y_j \geq 0$$

$$\sum_{i \in I} S_{i,glc} u_i + glc + \mu_{min,glc} - \mu_{max,glc} - \mu_{max2,glc}$$

$$+ \mu_{min2,glc} = 0$$

$$\sum_{i \in I} S_{i,biom} u_i + \mu_{biom} + \mu_{min,biom} - \mu_{max,biom}$$

$$- \mu_{max2,biom} + \mu_{min2,biom} = 0$$

$$\sum_{i \in I} S_{i,j} u_i + \mu_{min,j} - \mu_{max,j} - \mu_{max2,j} + \mu_{min2,j}$$

$$= 0, \forall j \in J, j \neq \text{biom, glc}$$

$$\mu_{max2,j} \left(v_j^{max} - w_j \right) - \mu_{min2,j} \left(v_j^{min} - w_j \right) - a_j \leq 1, \forall j \in J \quad (7)$$

$$v_j, glc, u_i \in \mathbb{R}, \forall j \in J, \forall i \in I,$$

$$\mu_{biom}, \mu_{min,j}, \mu_{max,j}, \mu_{max2,j}, \mu_{min2,j}, a_j, y_j \geq 0, \forall j \in J$$

In order to solve the max-min problem above, we use the strong duality theorem again to transform the minimization problem into its dual maximization representation, resulting in its equivalent single-level MIP as:

$$\max_{z_j, \gamma, p_j, q_j, x, c_j, t, s_j, r_j} \quad \gamma_{glc} v_{glc\_uptake} + \gamma_{biom} v_{biomass}^{target} + \sum_{j \in J} v_j^{min} x_{min,j} z_j -$$

$$\sum_{j \in J} v_j^{max} x_{max,j} z_j - \sum_{j \in J} x_{max2,j} w_j + \sum_{j \in J} x_{min2,j} w_j$$

$$- \sum_{j \in J} c_j - \sum_{j \in J} q_j$$

$$\text{s.t.} \quad \sum_{j \in J}(1 - z_j) \leq K,$$

$$\sum_{j \in J} S_{i,j} s_j = 0, \forall i \in I, \, s_{glc} = v_{glc\_uptake}$$

$$s_{biom} \geq v_{biomass}^{target}, \, v_j^{min} z_j \leq s_j \leq v_j^{max} z_j, \forall j \in J$$

$$s_j - r_j \left(v_j^{max} - w_j \right) \leq w_j, \, s_j - r_j \left(v_j^{min} - w_j \right) \geq w_j, \, 0 \leq r_j \leq 1, \forall j \in J$$

$$\sum_{i \in I} S_{i,chem} \gamma_i + x_{min,chem} - x_{max,chem} - x_{max2,chem}$$

$$+ x_{min2,chem} = 1$$

$$\sum_{i \in I} S_{i,glc} \gamma_i + \gamma_{glc} + x_{min,glc} - x_{max,glc} - x_{max2,glc}$$

$$+ x_{min2,glc} = 0$$

$$\sum_{i \in I} S_{i,biom} \gamma_i + \gamma_{biom} + x_{min,biom} - x_{max,biom} - x_{max2,biom}$$

$$+ x_{min2,biom} = 0$$

$$\sum_{i \in I} S_{i,j} \gamma_i + x_{min,j} - x_{max,j} - x_{max2,j} + x_{min2,j} = 0, \forall j \in J, j$$

$$\neq \text{biom,glc,chem}$$

$$x_{max2,j} \left(v_j^{max} - w_j \right) - x_{min2,j} \left(v_j^{min} - w_j \right) - c_j - t \leq 0, \forall j \in J$$

$$\sum_{j \in J} S_{ij} p_j = 0, \forall i \in I$$

$$(1 + \epsilon) v_{glc\_uptake} t + p_{glc} = 0$$

$$- v_j^{max}(1 + \epsilon) t z_j - p_j \leq 0, \forall j \in J$$

$$v_j^{min}(1 + \epsilon) t z_j + p_j \leq 0, \forall j \in J$$

$$(1 + \epsilon) v_{biomass}^{target} t + p_{biom} \leq 0$$

$$- w_j(1 + \epsilon) t - p_j - \left(v_j^{max} - w_j \right) q_j \leq 0, \forall j \in J$$

$$w_j(1 + \epsilon) t + p_j + \left(v_j^{min} - w_j \right) q_j \leq 0, \forall j \in J$$

$$q_j - (1 + \epsilon) t \leq 0, \forall j \in J$$

$$s_j, \gamma_{glc}, \gamma_i, p_j \in \mathbb{R}, \forall i \in I, \forall j \in J$$

$$\gamma_{biom}, x_{min,j}, x_{max,j}, x_{max2,j}, x_{min2,j}, c_j, t, q_j, r_j \geq 0, z_j$$

$$\in \{0,1\}, \forall j \in J.$$

$$(8)$$

Since some knockout strategies z_j may not have corresponding feasible solutions to the inner problem of (7), causing the corresponding dual solutions to be unbounded, new continuous decision variables s_j and r_j are added here in order to enforce the feasibility of the inner primal problem of (7) to avoid such degenerated cases. The bilinear terms in the objective function and the constraints in (8) can be further linearized by using the commonly adopted big-M method. This single-level MILP problem can be solved efficiently for large-scale problems by available solvers.

Pessimistic strain optimization II: P-OptKnock

In OptKnock [17], the inner-level mutant survival model is approximated by the maximum biomass growth condition. In order to optimize for the overproduction of the target biochemicals in worse-case scenario, we can similarly write the corresponding pessimistic formulation as what we have derived for P-ROOM:

$$
\max_{z_j, \forall j \in J} \quad \min_{v_j \in Y(z_j)} \quad v_{chemical}
$$

$$
\text{s.t.} \quad \sum_{j \in J}(1 - z_j) \leqslant K
$$

$$
z_j \in \{0,1\}, \forall j \in J
$$

$$
Y(z_j) = \arg\max \left\{ v_{biom} : \sum_{j \in J} S_{ij}v_j = 0, \forall i \in I, \right.
$$

$$
v_{glc} = v_{glc\_uptake},
$$

$$
\left. v_{biom} \geqslant v_{biomass}^{target}, \quad v_j^{min}z_j \leqslant v_j \leqslant v_j^{max}z_j, \forall j \in J \right\}.
$$

$$(9)$$

By adopting the same ϵ-approximation to the pessimistic problem in (9) to incorporate the inner-level model uncertainty, we have the P-OptKnock formulation as:

$$
\max_{z_j: z_j \in \{0,1\}, \forall j \in J, \sum_{j \in J}(1-z_j) \leqslant K} \quad \min_{v_j, \forall j \in J} \quad v_{chemical}
$$

$$
\text{s.t.} \sum_{j \in J} S_{ij}v_j = 0, \forall i \in I, v_{glc} = v_{glc\_uptake},
$$

$$
v_{biom} \geqslant v_{biomass}^{target}
$$

$$
v_j^{min}z_j \leqslant v_j \leqslant v_j^{max}z_j, \forall j \in J
$$

$$
v_{biom} \geq \varphi(\mathbf{v}^*)(1 - \epsilon),
$$

$$(10)$$

where $\varphi(\mathbf{v}^*)$ is the inner-most level function value of a given inner-level decision variable $\mathbf{v}^*$. Similar to the approach we have employed for P-ROOM, we can convert this optimization problem (10) to a single-level MIP. We note that the single-level MIP equivalent of the pessimistic formulation P-OptKnock is indeed similar to the work in [26] when the tolerance level ϵ is chosen as 0. However, we emphasize that our proposed pessimistic framework is more general and can be extended to different bi-level strain optimization formulations that employ various cell-survival models.
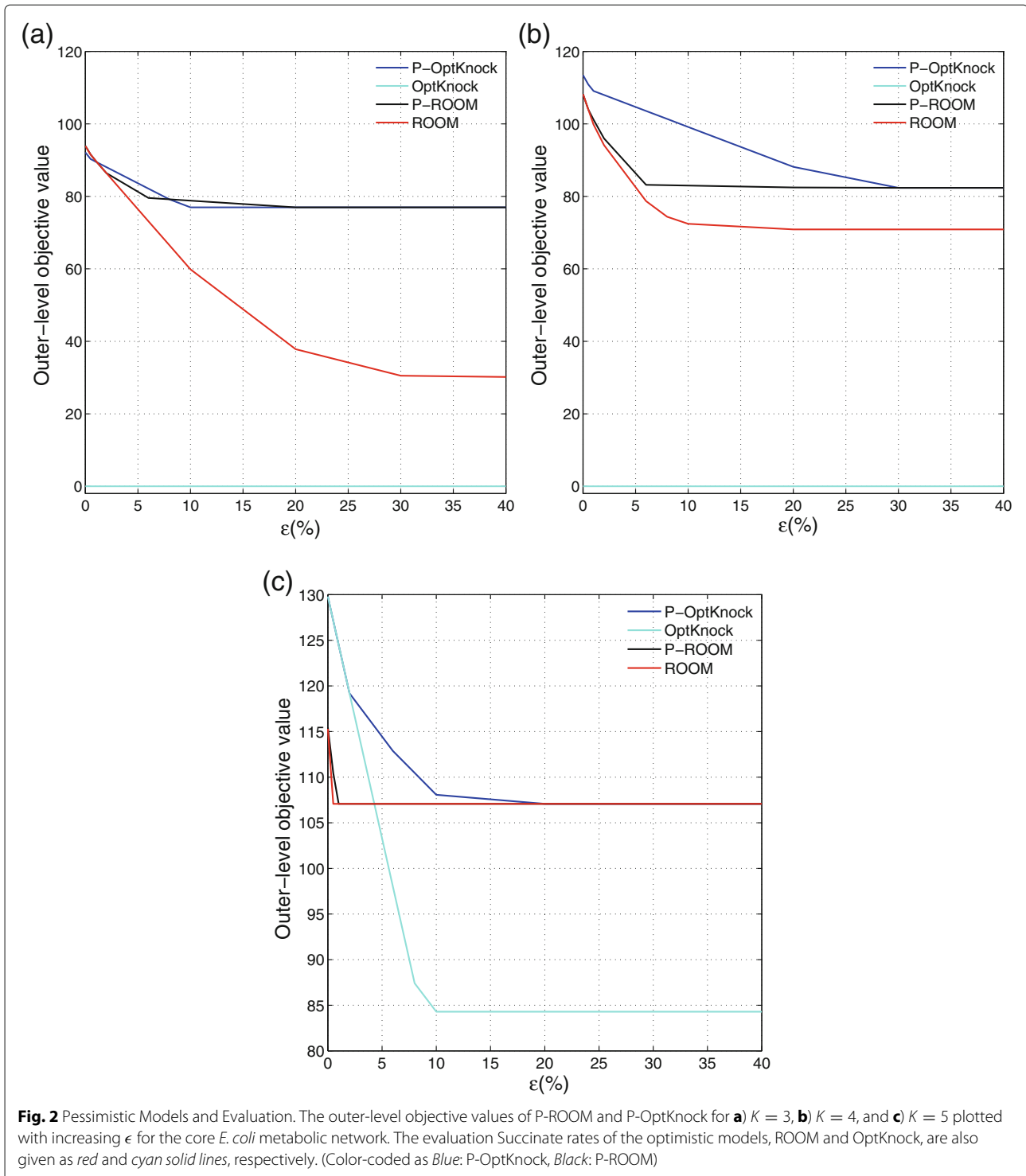
Results and discussion

In this section, we first evaluate the knockout solutions by optimistic bi-level optimization methods on a core *E. coli* metabolic network [27] when the outer- and inner-level decision makers are not cooperative. We then test our robust strain optimization methods, namely P-ROOM and P-OptKnock, derived in the "Methods" section, on the core network and a large *E. coli* metabolic network, iAF1260 [22], for overproduction of Succinate. The models are optimized using the MILP solver CPLEX 12.6.3 [28].

Succinate overproduction on AntCore metabolic network

We derive knockout solutions for a core *E. coli* metabolic network model [27] with 74 chemicals and 75 reactions for Succinate overproduction. The network model is from the OptKnock package [17], in which the glucose uptake is set at a fixed value of 100mmol/gDW.hr and the minimum biomass is set as 5 mmol/gDW.hr. Since the glucose-uptake rate is fixed, the Succinate yield is equivalent to the corresponding flux value considering steady-state stoichiometry constraint. The wild-type flux distribution is computed by maximizing the biomass in the FBA framework. In this set of experiments, we set the allowable knockout numbers $K = 3$, 4, and 5. All the reported experiments are based on the aerobic condition.

We first evaluate the robustness of two optimistic bi-level programs OptKnock (1) and ROOM (2) by computing the performance of least favorable Succinate overproduction under uncertainty for the derived knockouts $\mathbf{z}^*$ by solving (1) and (2). As discussed in the "Methods" section, we introduce a parameter ϵ as the model tolerance level to capture how faithful the inner-level mutant survival model is and whether the mutants will cooperate with the outer-level overproduction objectives. By allowing the gap ϵ between the realistic responses and the optimal inner-level objective function value, the increase in ϵ reflects the higher model tolerance by approximating the inner-level model uncertainty and non-cooperativeness. The pessimistic Succinate overproduction rate evaluations of optimistic knockout solutions are derived by finding the worst-case solutions of reaction flux values in ϵ-approximation sets.

The computed evaluation values for different ϵ values ($0 \leq \epsilon \leq 0.4$) are plotted in cyan and red lines in Fig. 2 with different numbers of allowed knockouts. It is clear that with higher model uncertainty (larger ϵ values), both ROOM and OptKnock strategies lead to smaller Succinate overproduction compared to the predicted optimistic

Fig. 2 Pessimistic Models and Evaluation. The outer-level objective values of P-ROOM and P-OptKnock for **a**) $K = 3$, **b**) $K = 4$, and **c**) $K = 5$ plotted with increasing ϵ for the core *E. coli* metabolic network. The evaluation Succinate rates of the optimistic models, ROOM and OptKnock, are also given as *red* and *cyan solid lines*, respectively. (Color-coded as *Blue*: P-OptKnock, *Black*: P-ROOM)

overproduction rates. We also note that the Succinate rates drop very quickly even with small ϵ, which implies that these optimistic knockout strategies are not robust. Finally, when the inner-level mutant survival models are not as faithful as desired, the derived knockout strategies may not be effective at all. For example, when $K = 3$ and 4,

the knockouts suggested by OptKnock are too "optimistic" as the corresponding evaluation Succinate rates go to 0 as shown in Fig. 2a and b. Compared with the results from ROOM, this validates that the minimum flux change model may be a better mutant cell survival model than the maximum biomass growth model adopted in

OptKnock. These experiments bring out that the cellular and engineering objectives may behave in opposite directions and the optimistic knockout strategies may not work in practice.

With the same ϵ-approximation model uncertainty, we now compare the corresponding derived knockout strategies based on our proposed pessimistic models: P-OptKnock and P-ROOM. Figure 2 illustrates the ϵ-optimal Succinate overproduction rates of the derived knockout solutions by P-OptKnock and P-ROOM in blue and black lines, respectively. Obviously, our pessimistic strategies derive more robust solutions, for which the worst-case or least favorable Succinate rates for different ϵ values are consistently on top of the rates from optimistic strategies. The least favorable Succinate rates decrease monotonically with increasing ϵ corresponding to higher model uncertainty. More interestingly, as ϵ increases, both P-OptKnock and P-ROOM converge to the stable Succinate overproduction rates. In fact, they derive the same knockout reactions as provided in Table 1. This clearly shows that our pessimistic formulations can produce consistent and robust mutant strains with respect to the inner-level mutant survival model uncertainty. As a side note, when $K = 5$, the pessimistic Succinate values produced by P-ROOM and its optimistic counterpart are almost the same for different tolerance levels, probably because the surviving mutants are more restricted with the increasing number of knockout reactions. This also shows that the minimum flux change is a better mutant survival model.

In Table 1, the succinate overproduction rates are given by the pessimistic formulations P-ROOM and P-OptKnock for the tolerance values ϵ in which both

models became stable ($\epsilon = 0.4$) as shown in Fig. 2. Although the pessimistic models provide different Succinate values for different ϵ values in Fig. 2, the knockout strategies and the Succinate values for both P-ROOM and P-OptKnock are identical when they reach stability. When $K = 3$, one of the knockouts suggested by P-ROOM and P-OptKnock involves competing byproduct metabolism pathways for succinate such as 6-Phospho-D-gluconate (6pg) and Ribulose 5-phosphate (ru5p). With $K = 4$, the reaction decomposing succinate (suc) is also eliminated. As more knockouts are allowed, the resulting succinate production can be further increased. We should also note that the removal of an important reaction for Trans-hyrogenation (nadh↔nadph) may cause significant reduction in biomass flux value. For the knockout strategies suggested by ROOM and OptKnock, when assuming the optimistic environment, the succinate rates are larger than the values suggested by pessimistic knockouts as expected since optimistic formulation provides an upper bound for the pessimistic formulation. From all the experiments with different K's, we demonstrate that, when we formulate the mutant optimization problem as pessimistic optimization by incorporating the model uncertainty, both P-OptKnock and P-ROOM can suggest consistent and robust knockouts, making the mutant strain optimization problem more reliable, accurate, and independent from the choice of inner-level surrogate functions for mutant cell objectives.

Succinate overproduction on iAF1260 network
We have shown that the derived knockout strategies based on optimistic methods may not be robust and

Table 1 Knockout strains derived by pessimistic and optimistic models on the core *E.coli* metabolic network

	K	Knockouts	Succinate production
P-ROOM	3	6pg→ru5p + co2 + nadph, oac + accoa→cit, ac→ac(ext)	76.97
	4	g6p→6pg + nadph, oac + accoa→cit, suc↔fum + fadh2, ac→ac(ext)	82.36
	5	g6p→6pg + nadph, mal→pyr + co2 + nadph, 3pg + glu→ser + akg + nadh, fadh2 + 0.5o2→2atp, nadh↔nadph	107.07
P-OptKnock	3	6pg→ru5p + co2 + nadph, oac + accoa→ cit, ac→ac(ext)	76.97
	4	g6p→6pg + nadph, oac + accoa→cit, suc↔fum + fadh2, ac→ac(ext)	82.36
	5	g6p→6pg + nadph, mal→pyr + co2 + nadph, 3pg + glu→ser + akg + nadh, fadh2 + 0.5o2→2atp, nadh↔nadph	107.07
ROOM	3	dhap↔gap, g6p→6pg + nadph, fadh2 + 0.5o2→2atp	93.92
	4	dhap↔gap, g6p→6pg + nadph, fadh2 + 0.5o2→2atp, glyc(ext)→	108.22
	5	g6p→6pg + nadph, mal→pyr + co2 + nadph, 3pg + glu→ser + akg + nadh, fadh2 + 0.5o2→2atp, nadh↔nadph	115.58
OptKnock	3	g6p→6pg + nadph, mal→pyr + co2 + nadph, nadh↔nadph	110.179
	4	g6p→6pg + nadph, mal→pyr + co2 + nadph, 3pg + glu→ser + akg + nadh, nadh↔ nadph	123.314
	5	g6p→6pg + nadph, 3pg + glu→ser + akg + nadh, nadh↔nadph, glyc↔glyc(ext), ac(ext)→	129.786

often "over-optimistic" by the experiments on the core *E. coli* metabolic network, while our proposed methods P-ROOM and P-OptKnock based on pessimistic optimization for mutant strain design have achieved robust and more practical knockout strategies. We further test P-ROOM and P-OptKnock to derive mutant strains for the overproduction of Succinate on a large *E. coli* metabolic network model, iAF1260 [22], which has 1658 metabolites and 2936 reactions including the pseudo reactions required for the computational model, also from the OptKnock package [17]. The glucose uptake rate is fixed at 100 mmol/gDW.hr, and the minimum biomass value is also set to 5 mmol/gDW.hr. The reported experiments are based on the anaerobic environment. We allow $K = 3$ knockout reactions on this large network.

Figure 3 provides the ϵ-optimal Succinate flux rates, which are the outer-level objective function values based on two pessimistic models of P-ROOM and P-OptKnock at different ϵ values. When there is no model uncertainty ($\epsilon = 0$), P-ROOM clearly provides higher Succinate rate than the value P-OptKnock suggests. This again suggests that modeling mutant cell survival by minimization of flux changes plays a role in favor of engineering objectives with improved targeted productions. At higher ϵ values, P-ROOM and P-OptKnock reach stability, and they derive the same knockout reactions as provided in Table 2 as we also observed in the experiments on the *E. coli* core metabolic network. If we consider the fact that the inner-level cell survival models are approximate

Fig. 3 Pessimistic Models. The outer-level objective values of P-ROOM and P-OptKnock for $K = 3$ plotted as *solid lines* with increasing ϵ for a large-scale iAF1260 *E. coli* metabolic network. (Color-coded as *Blue*: P-OptKnock, *Black*: P-ROOM)

to the real-world systems, incorporating this model uncertainty with ϵ-approximation in our formulated pessimistic mutant strain optimization methods, P-ROOM and P-OptKnock, has a key role to achieve robust knockout solutions that are less affected by these approximate cell survival models.

Table 2 provides the predicted knockouts by P-ROOM and P-OptKnock when $K = 3$. The minimum and maximum Succinate flux rates are also given for the mutant and also wild-type strains. The Succinate flux rates of pessimistic models are given for two different ϵ values: $\epsilon = 1$ denoting the highest tolerance level for the inner-level model uncertainty, and $\epsilon = 0$ denoting 0 tolerance level. In Table 2, both P-ROOM and P-OptKnock predict three identical knockouts that yield a minimum production rate of the Succinate that is higher than the predicted production rate in the wild-type strains. It validates that our pessimistic mutant strain optimization methods guarantee a higher Succinate production rate than that of wild-type strains even in the worst-case scenario. Furthermore, minimum and maximum Succinate rates given the predicted knockouts define a larger range for the OptKnock model, which may indicate the inner-level mutant cell survival model by biomass maximization is less faithful than ROOM's minimum flux change model.

We have also included the derived knockouts by ROOM and OptKnock and corresponding minimum and maximum succinate rates in Table 2. The mutant strain with the knockouts obtained by ROOM produces 0 minimal succinate flux value with the highest inner-level model uncertainty. However, P-ROOM succeeds in getting higher production rates even when $\epsilon = 1$. The minimal succinate rate when $\epsilon = 0$ in ROOM is greater than the minimal succinate rate for P-ROOM which is 8.09. This is because the reported pessimistic knockouts are identified when the model uncertainty is taken as $\epsilon = 1$. We note that the mutant with the pessimistic knockouts identified by P-ROOM when $\epsilon = 0$ actually produces the minimal succinate rate as 31.36. The knockouts captured by OptKnock are same as the ones identified by pessimistic formulations, leading to the same mutant with the same succinate production rates as P-OptKnock. Pessimistic formulations identified the succinate dehydrogenase reaction (SUCDi) as one of the suggested knockouts, which directly consumes succinate (succ), and the reaction pyruvate formate lyase (PFL). These have been reported as effective knockout strategies in [17, 29].

As demonstrated by the results from both the small core *E. coli* metabolic network and large iAF1260 network, our pessimistic formulations for strain optimization produce robust and stable knockout solution strategies under model uncertainty. Especially when comparing the results from P-ROOM and P-OptKnock, both models derive the
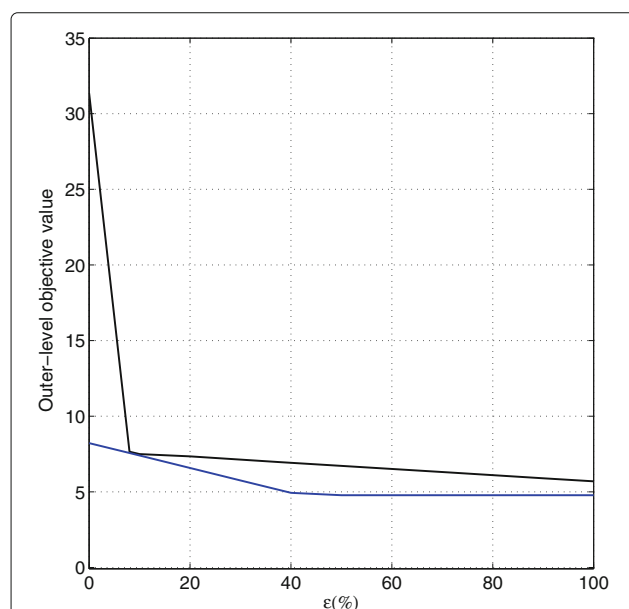
Table 2 Knockout strains derived by pessimistic and optimistic models on the iAF1260 *E.coli* metabolic network

	Knockouts		min $v_{succinate}$	max $v_{succinate}$
ROOM	-		0	0
FBA	-		0	0.372
P-ROOM	• mlthf + nadp↔methf + nadph	$\epsilon = 0$	8.09	8.09
	• coa + pyr→accoa + for	$\epsilon = 1$	5.69	40.08
	• q8 + succ→fum + q8h2			
P-OptKnock	• mlthf + nadp↔methf + nadph	$\epsilon = 0$	8.23	8.23
	• coa + pyr→accoa + for	$\epsilon = 1$	4.78	80.52
	• q8 + succ→fum + q8h2			
ROOM	• g6p + nadp↔6pgl + h + nadph	$\epsilon = 0$	31.36	31.36
	• h2o + pser-L→pi + ser-L	$\epsilon = 1$	0	76.143
	• q8 + succ→fum + q8h2			
OptKnock	• mlthf + nadp↔methf + nadph	$\epsilon = 0$	8.23	8.23
	• coa + pyr→accoa + for	$\epsilon = 1$	4.78	80.52
	• q8 + succ→fum + q8h2			

same stable knockout solutions regardless of the inner-level surrogate functions for mutant cell survival. Such consistency is critical when considering the translation of computationally derived results into in vivo experiments in practical metabolic engineering.

Conclusions

In this paper, we have proposed a new pessimistic optimization framework to identify optimal knockout strategies for maximum targeted bio-production under model uncertainty. We specifically have investigated two cell survival models and presented two corresponding pessimistic models to derive robust knockout reactions to achieve maximum biochemical overproduction: (1) P-ROOM under the minimum flux change condition; and (2) P-OptKnock under the maximum biomass condition. The experiments on both the core *E. coli* metabolic network [27] and large-scale iAF1260 network [22] have demonstrated that the pessimistic models for strain optimization derive robust and stable metabolic strain perturbation strategies through genome-scale steady-state stoichiometric modeling. Formulating the strain optimization problem based on pessimistic optimization considering the model uncertainty leads to consistent and reliable solutions regardless of the inner-level mutant survival models, which can provide high confidence for future in vivo experimental design of promising microbial mutants that benefit the human society.

Availability of data and materials
The C++ source code and relevant network data are available upon request.

About this supplement
This article has been published as part of *BMC Genomics* Volume 18 Supplement 6, 2017: Selected articles from the International Conference on Intelligent Biology and Medicine (ICIBM) 2016: genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-18-supplement-6.

Authors' contributions
Conceived and designed the experiments: MA, LX, BZ, XQ. Designed and Implemented the algorithm: MA, LX, BZ, XQ. Performed the experiments: MA. Analyzed the results: MA, LX, BZ, XQ. Wrote the paper: MA, LX, BZ, XQ. Supervised the study: BZ, XQ. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Dept. of Electrical and Computer Engineering, Texas A&M University, 77843 College Station, USA. [2]Dept. of Industrial Engineering, University of Pittsburgh, 15260 Pittsburgh, USA.

Published: 3 October 2017

References

1. Barrett CL, Kim TY, Kim HU, Palsson BØ, Lee SY. Systems biology as a foundation for genome-scale synthetic biology. Curr Opin Biotechnol. 2006;17(5):488–92.
2. Esvelt KM, Wang HH. Genome-scale engineering for systems and synthetic biology. Mol Syst Biol. 2013;9(1):641.
3. Palsson BØ. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge: Cambridge University Press; 2015.
4. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol Syst Biol. 2013;9(1):661.
5. Terzer M, Maynard ND, Covert MW, Stelling J. Genome-scale metabolic networks. Wiley Interdiscip Rev Syst Biol Med. 2009;1(3):285–97.
6. Chotani G, Dodge T, Hsu A, Kumar M, LaDuca R, Trimbur D, Weyler W, Sanford K. The commercial production of chemicals using pathway engineering. Biochim Biophys Acta Protein Struct Mol Enzymol. 2000;1543(2):434–55.
7. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol. 2012;8(6):536–46.
8. Broa C, Regenberga B, Förster J, Nielsen J. *In silico* aided metabolic engineering of *saccharomyces cerevisiae* for improved bioethanol production. Metab Eng. 2006;8:102–11.
9. Lu J, Sheahan C, Fu P. Metabolic engineering of algae for fourth generation biofuels production. Energy Environ Sci. 2011;4:2451–66.
10. Luengo JM, Garcia B, Sandoval A, Naharro G, Olivera ER. Bioplastics from microorganisms. Curr Opin Microbiol. 2003;6:251–60.
11. Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO. Metabolic engineering of *klebsiella oxytoca m5a1* for ethanol production from xylose and glucose. Appl Environ Microbiol. 1991;57:2810–1815.
12. Burgard AP, Maranas CD. Probing the performance limits of the Escherichia coli metabolic network subject to gene additions or deletions. Biotech Bioeng. 2001;74(5):364–75.
13. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci. 2002;99(23):15112–7.
14. Varma A, Palsson BØ. Metabolic flux balancing: Basic concepts, scientific and practical use. Nat Biotechnol. 1994;12:994–8.
15. Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. Brief Bioinform. 2006;7(2):140–50.
16. Edwards JS, Palsson BØ. The escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci. 2000;97(10):5528–33.
17. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotech Bioeng. 2003;84(6):647–57.
18. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci USA. 2005;102(21):7695–700.
19. Ren S, Zeng B, Qian X. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC Bioinforma. 2013;14(2):1.
20. Apaydin M, Zeng B, Qian X. A reliable alternative of OptKnock for desirable mutant microbial strains. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 2016. p. 573–576. doi:10.1109/BHI.2016.7455962.
21. Zeng B. Easier than we thought-a practical scheme to compute pessimistic bilevel optimization problem. 2015. Available at SSRN: https://ssrn.com/abstract=2658342 or http://dx.doi.org/10.2139/ssrn.2658342.
22. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. 2007;3(1):121.
23. Kim J, Reed JL, Maravelias CT. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS ONE. 2011;6(9):24162.
24. Burgard AP, Maranas CD. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotech Bioeng. 2003;82(6):670–7.
25. Deng X. Complexity issues in bilevel linear programming. In: Multilevel Optimization: Algorithms and Applications. US: Springer; 1998. p. 149–64.
26. Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 2010;26(4):536–43.
27. Antoniewicz MR, Kraynie DF, Laffend LA, Gonzalez-Lergier J, Kelleher JK, Stephanopoulos G. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of E. coli producing 1, 3-propanediol. Metab Eng. 2007;9(3):277–92.
28. IBM ILOG CPLEX Optimizer 12.6.3. 2015, IBM Corporation.
29. Stols L, Donnelly MI. Production of succinic acid through overexpression of NAD (+)-dependent malic enzyme in an Escherichia coli mutant. Appl Environ Microbiol. 1997;63(7):2695–701.