Research Article

Constraint-Based Strain Design using Continuous Modifications (CosMos) of Flux Bounds Finds New Strategies for Metabolic Engineering

Cameron Cotten[1,2] and Jennifer L. Reed*[1,2]

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, USA

[2]Great Lakes Bioenergy Research Center, University of Wisconsin – Madison, Madison, WI, USA

*Corresponding Author:

Jennifer L. Reed

Assistant Professor

1415 Engineering Dr., Madison, WI 53706

Email: reed@engr.wisc.edu

phone: (608) 262-0188

fax: (608) 262-5434

Key words: Constraint-based modeling, enzyme expression, flux balance analysis, OptKnock, strain design

Submitted: 09-Oct-2012
Revised: 20-Feb-2013
Accepted: 02-Apr-2013

Abbreviations: ATP – Adenosine triphosphate, FBA – Flux Balance Analysis, MOMA – Minimization of Metabolic Adjustment, ROOM – Regulatory on-off minimization, RELATCH – Relative Change, gDW – Gram dry weight, FVA – Flux Variability Analysis, FRD – Fumarate reductase, ASPK – aspartate kinase, CYSDS - Cysteine desulfhydrase, CYSS – Cysteine synthase, ASPT - aspartase, AKGDH – 2-oxogulatarate dehydrogenase, ICL – isocitrate lyase, MALS – malate synthase, ASPTA - aspartate transaminase, ADSL1r - adenylsuccinate lyase, PPA - inorganic diphosphatase, PPKr - Propionate kinase, NDPK1 - nucleoside-diphosphate kinase (ATP:GDP), R15BPK - ribose-1,5-bisphosphokinase, ASAD - aspartate-semialdehyde dehydrogenase, ADK – Adenylate Kinase, CHORS – Chorismate Synthase, CHORM – Chorismate mutase, GLNS – glutamate synthase, DHBS - 2,3-dihydroxybenzoate adenylate synthase, GLCS1 - glycogen synthase (ADPGlc), DHBD - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, DDPA - 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase, PPND -prephenate dehydrogenase, PPNDH - prephenate dehydratase, QUINDH - Quinate dehydrogenase, OP – overproducing, WT – wild type, PPC – Phosphoenolpyruvate carboxylase, MDH – Malate dehydrogenase, ACONT – Aconitase, ICDHyr – Isocitrate dehydrogenase, GLUDy – Glutamate dehydronase, CS – Citrate synthase, IPMD - 3-isopropylmalate dehydrogenase, IPPMIa – 3-isopropylmalate dehydratase, LEUTAi – leucine transaminase (irreversible), AICART – phosphoribosylaminoimidazolecarboxamide formyltransferase, ACLS –acetolactate synthase, KARA1 – ketol-acid reductoisomerase (2,3-dihydroxy-3-methylbutanoate), HSDy – homoserine dehydrogenase (NADPH)

Abstract

In recent years, a growing number of metabolic engineering strain design techniques have employed constraint-based modeling to determine metabolic and regulatory network changes which are needed to improve chemical production. These methods use systems-level analysis of metabolism to help guide experimental efforts by identifying deletions, additions, downregulations, and upregulations of metabolic genes that will increase biological production of a desired metabolic product. In this work, we propose a new strain design method (CosMos) that provides strategies for deletions, downregulations, and upregulations of fluxes that will lead to production of desired products. The method is conceptually simple and easy to implement, and can provide additional strategies over current approaches. We found that the method was able to find strain design strategies that required fewer modifications and had larger predicted yields than strategies from previous methods in example and genome-scale networks. Using CosMos, we identified modification strategies for producing a variety of metabolic products, compared strategies derived from *Escherichia coli* and *Saccharomyces cerevisiae* metabolic models, and examined how imperfect implementation may affect experimental outcomes. This study gives a powerful and flexible technique for strain engineering and examines some of the unexpected outcomes that may arise when experimentally implementing strategies.

1 Introduction

In recent years, a growing number of metabolic engineering strain design techniques have employed constraint-based modeling to determine metabolic and regulatory network changes which are needed to improve chemical production [1-8]. Constraint-based models have an advantage in that they are independent of complex rate laws associated with enzyme kinetics. As a result, these models solve sets of linear equations rather than differential equations and can give provable optimal solutions. Constraint-based strain design methods are available that can predict flux distributions for a chosen gene or reaction deletion (such as FBA, MOMA, ROOM, FBA with grouping reaction constraints, and RELATCH) [9-15]. However, in metabolic engineering it is not always obvious what genes or reactions should be added or deleted, and so other constraint-based methods use bilevel optimization to identify the network changes that are needed to improve chemical production.

Most bilevel optimization methods to date use only binary (yes/no) decisions for metabolic interventions, such as deletions of genes or addition of exogenous pathways. The first of these was OptKnock, which gives reaction deletions that couple biomass production to desired product formation [2]. Genetic and regulatory information was added to OptKnock with the introduction of OptORF, which considers only the reaction deletions that are possible given the complex relationship between genes and reactions [3]. Two additional approaches (RobustKnock [13] and objective function tilting [14]) have been developed to avoid strategies that can have either high or low production at the maximum growth rate. OptStrain is a sequential approach that selects exogenous reactions to add which improve the maximum theoretical yield, and then applies OptKnock to the augmented network [8]. These two steps were recently implemented simultaneously in SimOptStrain, and the simultaneous approach can suggest higher production strategies compared to the sequential approach [4]. For smaller networks, other approaches based on analysis of elementary flux modes [16] can also identify deletions to couple biomass production to chemical production [17].

OptReg and OptForce expand the types of modifications that can be considered by including up- and downregulation as strain design options. However, these up- and downregulations remained binary decisions, which change the upper and lower flux bounds to preselected alternate values [5, 6]. An important difference between OptForce and previous methods (OptKnock, OptORF, and OptStrain) is that previous methods assume that designed strains will adaptively evolve to increase their biomass and production yields. OptForce instead guarantees that the desired product will be produced at the predicted rate, assuming that levels of flux up and downregulations are achieved experimentally, irrespective of adaptive evolution and the cell's objective. OptForce is a type of min/max optimization problem, in which the minimum amount of product that the cell must produce is increased by strain design interventions. An alternative approach using elementary flux modes is FluxDesign, which finds correlations between modes and identifies targets to be upregulated or downregulated [18].

In this work, we describe a new method called Continuous Modifications (CosMos) that incorporates flux up/down regulation and reaction knockouts in a more flexible manner than OptForce. In OptForce and our method, up and downregulations are represented by changes to the upper and lower bounds on reaction fluxes. In OptForce, changes to bounds are chosen ahead of time using values determined by flux variability analysis [19]. In contrast, CosMos can choose changes to bounds continuously without preselecting changes to bounds or relying on flux variability analysis, leading to the identification of more metabolic engineering strategies that guarantee chemical production. We compared strategies found using OptForce and CosMos on a simple illustrative network and for the case of succinate production in *E. coli* to illustrate differences between the two approaches, and showed that continuous changes to bounds can provide additional solutions compared to OptForce which require fewer modifications and give higher yields. We then applied CosMos to identify overproduction strategies for a

variety of products in *E. coli* and *Saccharomyces cerevisiae* that guarantee product formation. In addition, we propose an approach that uses random sampling to compare strategies based on the likelihood that experimental implementation will still result in product formation even if optimal flux modifications are not achieved.

2 Methods

2.1 Predicting Minimal Production Levels

Constraint-based modeling finds steady-state distributions of metabolic fluxes (*v*) that satisfy mass balances equations and upper and lower flux bounds [20]. Since a unique solution often does not exist, an objective function is often chosen which gives flux distributions that are optimal in some biologically-relevant sense. In our case, we want to identify how changes to the upper and lower flux bounds would affect the minimum production of a compound of interest. This problem can be formulated as

$$\min_{v} \sum_j c_j v_j \quad (1)$$

$$\sum_j S_{ij} v_j = 0 \qquad \forall i \in I \quad (2)$$

$$v_j \geq LB_j + \alpha_j \qquad \forall j \in J \quad (3)$$

$$v_j \leq UB_j - \beta_j \qquad \forall j \in J \quad (4)$$

$$\alpha_j \beta_j \geq 0 \qquad \forall j \in J \quad (5)$$

where J is the set of reactions in the biological network and I is the set of metabolites in the network. The fluxes (*v)* are chosen so that they minimize product secretion (defined in eq. 1 using the vector *c*), while also satisfying mass balance constraints (eq. 2), that use the stoichiometric matrix *S,* and lower (eq. 3) and upper (eq. 4) flux bounds. The *LB* and *UB* are bounds used to limit the value for each flux in the absence of any flux bound modifications. Modifications to some lower and upper bounds can be made using α and β, respectively. These flux bound modifications could be implemented experimentally by up- and downregulating expression of metabolic genes, altering enzyme kinetics, and removing feedback inhibition. In this optimization problem, these flux bound modifications are fixed parameters whose values must be chosen either *a priori* or by another optimization (detailed below). Metabolic network reconstructions for *E. coli* (iJO1366) [21] and yeast (*S. cerevisiae,* iMM904) [22] were used in this work to define *S*, *LB*, and *UB*. Lower bounds for biomass production (0.1 hr^{-1} for *E. coli* and 0.05 hr^{-1} for yeast) and ATP maintenance (3.15 and 1 mmol/gDW/hr for *E. coli* and yeast, respectively) were enforced to ensure viability of engineered strains. Additional constraints were used on the exchange fluxes to reflect a glucose aerobic conditions by setting the maximum uptake rates for glucose and oxygen to be -10 and -60 mmol/gDW/h, respectively for *E. coli* and -5 and -60 mmol/gDW/h, respectively for *S. cerevisiae*. The oxygen uptake rates were chosen so that glucose was the limiting nutrient. Flux bound modifications (α and β) that result in a non-zero minimal production (found using eq. 1-5) can essentially guarantee production of the desired target among all the possible steady-state flux distributions that the organism can choose. However, finding values of α and β that guarantee a minimum non-zero production is not obvious for many networks and products.

To systematically determine values for α and β corresponding to optimal flux up and downregulation, we embed the above minimization problem (eq. 1-5) into a bilevel optimization problem. These flux bound modifications can be chosen continuously, so we call this problem Continuous Modifications (CosMos). This is in contrast to previous methods, which only allow for binary changes to the flux bounds (e.g., a flux goes to zero or within some pre-defined change in bounds). The bilevel formulation of CosMos is given here:

$$\max_{\alpha,\beta,\delta^{\alpha},\delta^{\beta}} \sum_j c_j v_j - \varepsilon\left(\delta_j^{\alpha} + \delta_j^{\beta}\right) \quad (6)$$

s.t. $\min_{v} \sum_j c_j v_j$ (7)

s.t. $\sum_j S_{ij} v_j = 0 \quad \forall i \in I$ (8)

$v_j \geq LB_j + \alpha_j \quad \forall j \in J$ (9)

$v_j \leq UB_j - \beta_j \quad \forall j \in J$ (10)

$\alpha_j \leq \alpha_j^{\max} \delta_j^{\alpha} \quad \forall j \in J$ (11)

$\beta_j \leq \beta_j^{\max} \delta_j^{\beta} \quad \forall j \in J$ (12)

$\alpha_j, \beta_j \geq 0 \quad \forall j \in J$ (13)

$\delta_j^{\alpha}, \delta_j^{\beta} \in \{0,1\} \quad \forall j \in J$ (14)

The δ variables are binary variables that are one when a particular flux is modified and zero otherwise. More details including the full formulation of CosMos, including the single level mixed integer formulation, can be found in Supplementary Information. The CosMos problem can be seen as a generalization of previous strain design techniques. For example, OptKnock chooses reactions to delete, and solutions to OptKnock correspond to all solutions of CosMos that use $\alpha_j=LB_j$ and/or $\beta_j=UB_j$, since selecting these values for α and β blocks flux through the reaction. Similarly, α and β can be chosen *a priori* to give the up and downregulations that are used by OptForce. However, CosMos differs from OptForce in that it can identify additional modification strategies due to the continuous nature of each potential modification. We did not allow modifications of transport reactions or non-metabolic reactions, such as exchange reactions, ATP maintenance and the biomass reaction. In addition, we chose upper limits on α (α^{max}) and β (β^{max}) so that choices of α and β could not force a flux bound to change sign. Integer cuts can also be added to find additional solutions [23].

2.2 Sampling of Sub-Optimal Modifications

Because desired modifications to flux bounds would be difficult to implement exactly experimentally, we examined the effects of sub-optimal flux bound modifications on the guaranteed minimum production rate. After optimal values for α and β were derived from CosMos, random sub-optimal flux bound modifications were generated by drawing a random number for each individual modification from a uniform distribution between zero and one, and multiplying these random numbers by the optimal α and β values. The product minimization problem (eq. 1-5) was then solved with these values, and the minimum chemical production given these sub-optimal flux bound modifications was recorded to give a distribution of production values. For each CosMos strategy, we sampled 3,000 different sub-optimal flux bound modifications to evaluate how many would still yield high production even if the optimal flux bound modifications could not be achieved experimentally.

3 Results

In this study, we compared our method to previous strain design methods, found up- and downregulation strategies for producing a variety of metabolic products, and examined the effects of implementing these strategies imperfectly. Unless otherwise noted, the reported results used the iJO1366 [21] or iMM904 [22] metabolic models for *E. coli* and *S. cerevisiae*, respectively. We found that our method could identify new strategies over existing approaches for an example network and a genome-scale network. We compared optimal strategies for the overproduction of a variety of products in *E. coli* and *S. cerevisiae*, and found that generally production of products in *E. coli* required fewer

metabolic changes. We also found that there were significant differences between strategies that were equivalent in CosMos (i.e. they had the same number of modifications and yields) when they were implemented imperfectly, and we recommend using this type of evaluation to further aid in the selection of strategies for experimental implementation.

3.1 Comparisons to Existing Approaches

We constructed a simple example to illustrate the differences between solutions from existing approaches (OptForce and OptKnock) and CosMos (Figure 1). An example network was constructed that had a branch between an energetically favorable reaction (producing 3 ATP molecules) and two alternative reactions that produced a target chemical of interest, T (Figure 1A). One of these alternative reactions produced ATP, while the other did not. An uptake rate of 100 for A and minimal biomass fluxes of 11.5 and 57.5 were enforced for the OptForce overproducing (OP) strain and the wildtype strain (WT), respectively. A minimal biomass production rate of 11.5 was also used in CosMos. Additionally, the production of T was at least 90 when calculating the flux bounds for the OP strain in OptForce. To guarantee non-zero production of T, CosMos found two solutions that each up regulate one of the two reactions that produce T. The CosMos solution with the highest minimal production suggests upregulating the ATP producing reaction by 97.7 (Figure 1B). The alternate solution with lower guaranteed production involves upregulating the non-ATP producing reaction by 90.

CosMos finds strategies that still allow flux through the biomass reaction, and in this example gives higher product formation than two existing approaches, OptForce and OptKnock. OptForce is unable to find a solution for this network that involves gene up or downregulation, because the flux bounds that are identified in an OP strain for the two alternative T producing reactions have the same lower bound as the WT strain (Figure 1C). Up and downregulations are represented by an exchange of the WT flux bound (lower or upper, respectively) for the OP flux bound. For the OP strain, the two alternative reactions to produce T have the same lower bound of zero, since either alternative reaction can be used. As a result, OptForce cannot consider increasing the lower bound from zero for either alternative reaction. This issue arises due to the presence of alternate pathways for generating the product of interest, which is common in biological networks. Another method, OptKnock (which maximizes production at the maximum biomass production), proposed deleting the reaction converting B into R so that maximal biomass production requires production of T. This mutant strain should gradually evolve to produce more biomass and more T. At the best possible growth rate, T would be produced at a rate of 90.9 (Figure 1D). This OptKnock strategy has lower production than the best CosMos strategy and would require adaptive evolution to achieve the desired production.

We also compared results from OptForce to results from CosMos using the *E. coli* iAF1260 genome-scale network [24]. We determined strategies for succinate production from glucose in anaerobic conditions using CosMos and compared these to previously published results from OptForce [6]. Using OptForce, Ranganathan *et al.* found up and downregulation strategies that give 98% theoretical yield of succinate while still allowing biomass production. They detailed a number of strategies that use combinations of 10 up and down flux modifications to reach 98% maximum theoretical yields. Using CosMos, we were able to find strategies to give 99.6% of the maximum theoretical yield using only two modifications. An example CosMos strategy that gives 99.6% yield was upregulation of isocitrate lyase and one of the two fumarate reductase reactions (FRD2 or FRD3, Figure 1F). This result was similar to the example network, since the two fumarate reductase reactions are alternative ways to produce succinate and both have very small lower bounds in the OP strain. As a result, CosMos was able to generate modification strategies that required fewer modifications than were needed in OptForce.

CosMos also has the advantage of being a single optimization problem. In OptForce, two iterations of finding flux bounds are necessary, and an additional step of finding MUST sets is required. These MUST sets are the sets of reactions that have to be modified in order to meet overproduction goals, and they

are derived in a separate step to reduce the number of reactions that are considered by OptForce. The authors suggest finding 5 MUST sets, but more MUST sets can be found. These higher order MUST sets are progressively more computationally expensive to compute, and it is not clear how many higher order sets are needed to find good solutions. Thus, CosMos is easier to implement and is not very computationally expensive. Although information was not given on resource requirements in the original OptForce publication, the algorithm solved in less than 5 minutes for most products in both *S. cerevisiae* (iMM904) and *E. coli* (iJO1366).

3.2 CosMos Flux Modification Strategies for *E. coli*

CosMos was used to find flux modification strategies in *E. coli* for multiple products to demonstrate its flexibility and utility. Using glucose as a carbon source in aerobic conditions, multiple strategies to guarantee production of chorismate, ethanol, succinate, and a variety of L-amino acids (aspartate, histidine, leucine, phenylalanine, valine, and alanine) were found. Typical run times were in the range of 5 to 30 minutes, except products that required more modifications (alanine and ethanol) which needed over 60 minutes for a single solution. In all cases, strategies were found that guaranteed production at the maximum theoretical yield (after accounting for biomass and ATP maintenance). CosMos strategies generally involved flux bound changes which eliminate or lower flux through degradation pathways and which increase flux through reactions whose products must lead to target chemical production. An example was cysteine, in which CosMos increased cysteine production by downregulating flux through aspartate kinase (ASPK) and cysteine desulfhydrase (CYSD) and increasing cysteine synthase (CYSS) flux (Figure 2). When ASPK and CYSD are downregulated, increasing CYSS flux must increase cysteine production.

Running CosMos over multiple iterations using the integer cut method gives multiple strategies for maximum production of a variety of products that were tested (Table 1 and Supporting Information). In most cases, only a small number of modifications were needed to guarantee maximum theoretical production. Some products proved to be considerably more difficult, such as ethanol and alanine, requiring more than 15 modifications to guarantee production. When solving for multiple strategies using integer cuts to eliminate solutions that have already been found, generally the same reactions were used to force production of a product or its intermediates, and different reactions to downregulate were chosen to prevent the degradation of these products. For example, in 6 strategies for the overproduction of histidine, flux was forced through histidine biosynthesis by upregulating the first step of the histidine pathway catalyzed by adenine phosphoribosyltransferase. However, all 6 strategies downregulated different reactions that would have directed flux away from the pathway.

3.3 Comparisons with Previous Modification Strategies

Modification strategies produced by CosMos are in some cases similar to previously-reported modification strategies in experimental studies. Previously-reported genetic strategies were found for overproducing isoleucine and were compared to strain modifications suggested by CosMos. One previous experimental strategy in *E. coli* involved the upregulation of *ilvADEG* genes, which convert pyruvate and threonine to isoleucine, and *thrABC* genes, which convert aspartate to threonine [25]. CosMos found a strategy that requires the upregulation of two fluxes in isoleucine and threonine biosynthesis (associated with *ilvG* and *thrB*, respectively) and downregulation of one flux in threonine biosynthesis (shown as thick and dashed arrows, respectively, in Figure 3). CosMos called for the downregulation of a reaction associated with *asd*, which consumes 4-phospho-aspartate, to prevent excess aspartate from being converted into lysine. Downregulation of *asd* or *thrA* are both alternative CosMos solutions, since the two genes are responsible for neighboring steps in the pathway. In this example, CosMos was able to correctly determine that threonine and isoleucine biosynthesis fluxes need to increase. While CosMos suggested three fluxes to modify, the previous study upregulated genes that were associated with over ten reactions. In this case it may be necessary to additionally modify adjacent fluxes in linear pathways, if other reactions become bottlenecks after upregulating reactions

suggested by CosMos. In addition, CosMos identified that production of intermediates in threonine biosynthesis need to be controlled so that their excess production does not lead to lysine production.

Genetic strategies for producing cysteine and succinate have also appeared in the literature and are similar to CosMos strategies. Mutants of *E. coli* deficient in cysteine desulfhydrase overproduce cysteine when serine acetyltransferase is overexpressed on a plasmid [26]. CosMos recommends downregulation of cysteine desulfhydrase and aspartate kinase fluxes, and upregulation of cysteine synthase flux. Upregulation of cysteine synthase and serine acetyltransferase are equivalent computationally since they are adjacent enzymes in the cysteine biosynthesis pathway. Although downregulating aspartate kinase (ASPK) was not needed experimentally, it was needed computationally to prevent production cysteinyl-glycine by the model. For succinate, CosMos found that upregulating isocitrate lyase flux (part of the glyoxylate shunt) along with fumarate reductase flux could be a simple method to guarantee succinate production (Figure 1E). It had previously been suggested by FBA that use of the glyoxylate shunt could increase succinate production, and de-repression of the glyoxylate shunt by deletion of the *iclR* repressor leads to increased succinate production in experiments [27]. Comparisons between CosMos results and previous experimental studies indicate that CosMos can identify fluxes whose modifications improve production and that additional fluxes (found by CosMos) may need adjustment experimentally to guarantee product formation.

3.4 Production Strategies Comparisons Across Organisms

We also solved CosMos to find flux modification strategies for *S. cerevisiae* and compared the results to *E. coli* strategies. For three of the products compared (phenylalanine, leucine, and valine), the best strategies used the same number of flux modifications in *E. coli* and yeast (two, two and three modifications, respectively). However, for other products, strategies generally required more modifications in yeast than *E. coli*. Histidine required three modifications in yeast and only two in *E. coli*. For aspartate and ethanol, a strategy was found in *E. coli*, but no strategy could be found in the yeast model. Since the flux distribution with zero production proved to be the optimal flux distribution, it was proven that no viable strategy could be found for ethanol and aspartate in yeast without modifying transport fluxes or fixing the directions of fluxes *a priori*. The *S. cerevisiae* ethanol result is surprising since the microbe readily produces ethanol; however, the strain does not necessarily have to produce ethanol, even in anaerobic conditions. It is important to note that the simulations were run for aerobic conditions.

Significantly more modifications were needed to guarantee alanine production in *E. coli* compared to yeast (45 versus 8, respectively). Similarly, 22 modifications were needed to guarantee ethanol production in *E. coli*, and production could not be guaranteed in yeast. It appears that the large number of modifications needed is primarily due to the fact CosMos was not allowed to force non-zero flux through reversible reactions, giving less control over reversible reactions when designing an overproduction strategy. In the alanine production strategies found for *E. coli*, none of the reactions that were upregulated produce alanine since all alanine-producing reactions were reversible. Instead, a large number of reactions were modified to indirectly force alanine production. A similar problem was observed for ethanol production in both organisms since the alcohol dehydrogenase reaction was also reversible. When CosMos was instead allowed to force non-zero flux through reversible reactions, only five modifications were needed to guarantee alanine production in *E. coli*, which is similar in number to the eight reaction modifications needed to guarantee production of alanine in yeast. While it is possible to allow CosMos to force flux through reversible reactions (by changing α^{max} and β^{max}, see supplementary information for details) it may be more difficult to implement such strategies experimentally since flux is needed in just one direction. Other reasons for larger numbers of modifications can be due to an increase in the number of degradation pathways for products and their intermediates.

3.5 Sampling of Sub-Optimal Flux Modifications

While CosMos gives precise levels of up- and downregulation needed to guarantee minimal production, these exact flux modifications might be difficult to achieve experimentally. To compare how robust different strategies would be to experimental implementation errors, we determined the minimal production levels (by solving eq. 1-5) if sub-optimal flux modifications were used instead of the optimal values found by CosMos. Sampling of sub-optimal up- and downregulation strategies demonstrated some of the effects of increasing the number of modifications and elucidated differences between strategies that appeared equivalent according to CosMos. Five strategies (listed in Table 2) for production of leucine (requiring two modifications) and valine (requiring three modifications) were each sampled randomly 3,000 times. For each sample, random values for each α and β were drawn from a uniform distribution between zero and the optimal value found by CosMos, and the minimum production was then found using these sub-optimal flux bounds. Sub-optimal samples of strategies for leucine guaranteed some production in 96.9% of the samples (Figure 4A), while sub-optimal samples of strategies for valine guaranteed some production in fewer samples, with the best strategy (V1) guaranteeing production in 79% of its samples and only 58.3% of the samples in the worst strategies (V4 and V5) (Figure 4B).

As shown in Figure 4, sampling analysis of sup-optimal flux modifications can be used to identify differences between seemingly equivalent strategies for a given product (i.e., strategies have same number of flux modifications and the same predicted guaranteed yields). Of the five strategies that were examined for the production of leucine, two strategies (L4 and L5) were significantly better than the others when considering sub-optimal flux modifications. The two better strategies gave guaranteed production values that were closer to the theoretical maximum production rate with greater frequency, while the other strategies had a near uniform distribution of production rates, which appears as a straight line on a survival plot (Figure 4A). The five strategies examined for valine production also had varying results, despite being equivalent solutions to CosMos. Based on these results, it is important to further examine seemingly equivalent strategies more closely when considering which strategy to implement experimentally.

4 Conclusions

In this study, we formulated and solved a strain design problem that identified flux deletion, upregulation and downregulation strategies to guarantee production of a variety of metabolic products. We found that different products required varying degrees of intervention to guarantee production, with histidine only requiring 2 modifications and alanine requiring more than 30. Most often, these strategies identified upregulation of reactions to produce the desired chemical and downregulation of pathways that degrade the products and intermediates leading to its formation. Comparing results from *E. coli* and *S. cerevisiae*, we found that strategies generally required more modifications to guarantee production in *S. cerevisiae*. We further demonstrated that CosMos is able to find additional strategies for producing chemicals than existing approaches in both example and genome-scale metabolic networks.

CosMos identifies which fluxes need to change and by how much to guarantee production; however, these required flux changes might be difficult to achieve experimentally with great accuracy. Binary decisions in previous methods, such as the deletion or insertion of a gene, could be executed experimentally with certainty (gene was either deleted/inserted or not); however, the engineering suggestions proposed by our CosMos are continuous. Since it is difficult to engineer specific changes to flux bounds precisely, it is important to examine solutions that are similar to our desired strategy but not optimal. As a way to compare proposed strategies we examined the effects of sub-optimal implementation of CosMos solutions on chemical production, and demonstrated that the up- and downregulation strain design solutions differ in their sensitivity to modified flux bounds. When random

sub-optimal implementations of our solutions were evaluated, we found that some strategies yielded higher production rates more frequently than other strategies, despite having equal production rates if implemented exactly as proposed by CosMos. We believe that this approach can be applied when multiple up- and downregulation strategies exist to help aid in the selection of a strategy to implement experimentally.

CosMos in its current implementation does not favor one type of modification (upregulation, downregulation or deletion) over another. Since some modifications such as deletions or upregulations might be easier to implement experimentally, future work could consider modifying the outer objective function in CosMos to penalize more for types of modifications that might be more difficult to implement. In addition, the approach could be extended to also consider flux changes from a current metabolic state rather than all possible states. Finally, genetic information could be added to better represent up and downregulation when multiple reactions are catalyzed by a single enzyme.

CosMos expands the suite of computational strain design approaches by suggesting new strategies for guaranteeing chemical production. We believe this systems-level approach, along with the sensitivity analysis to modified flux bounds, will help better enable computer aided-design of engineered microbes, thereby reducing the time and costs needed for strain development.

Acknowledgements

The authors wish to acknowledge James Leutdke, Jeff Linderoth, and Christos Maravelias for helpful discussions. This work was funded by the United States Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). CC is also supported by a fellowship from the 3M Foundation. CC implemented the models and approach, performed the analysis, analyzed the data, and drafted the manuscript. JLR conceived of the study, participated in its design and coordination, and helped to analyze the data and draft the manuscript.

Conflict of Interest Statement

The authors declare no commercial or financial conflict of interest.

REFERENCES

[1] Asadollahi, M. A., Maury, J., Patil, K. R., Schalk, M.*, et al.*, Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. *Metabolic engineering* 2009, *11*, 328-334.
[2] Burgard, A. P., Pharkya, P., Maranas, C. D., Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering* 2003, *84*, 647-657.
[3] Kim, J., Reed, J. L., OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC systems biology* 2010, *4*, 53.
[4] Kim, J., Reed, J. L., Maravelias, C. T., Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PloS one* 2011, *6*, e24162.
[5] Pharkya, P., Maranas, C. D., An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic engineering* 2006, *8*, 1-13.
[6] Ranganathan, S., Suthers, P. F., Maranas, C. D., OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS computational biology* 2010, *6*, e1000744.
[7] Rocha, I., Maia, P., Evangelista, P., Vilaca, P.*, et al.*, OptFlux: an open-source software platform for in silico metabolic engineering. *BMC systems biology* 2010, *4*, 45.
[8] Pharkya, P., Burgard, A. P., Maranas, C. D., OptStrain: a computational framework for redesign of microbial production systems. *Genome research* 2004, *14*, 2367-2376.
[9] Orth, J. D., Thiele, I., Palsson, B. O., What is flux balance analysis? *Nature biotechnology* 2010, *28*, 245-248.
[10] Segre, D., Vitkup, D., Church, G. M., Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* 2002, *99*, 15112-15117.
[11] Shlomi, T., Berkman, O., Ruppin, E., Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102*, 7695-7700.
[12] Kim, J., Reed, J. L., RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome biology* 2012, *13*, R78.
[13] Tepper, N., Shlomi, T., Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* 2010, *26*, 536-543.
[14] Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J.*, et al.*, Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metabolic engineering* 2010, *12*, 173-186.
[15] Park, J. M., Kim, T. Y., Lee, S. Y., Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses. *Proceedings of the National Academy of Sciences of the United States of America* 2010, *107*, 14931-14936.
[16] Trinh, C. T., Wlaschin, A., Srienc, F., Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied microbiology and biotechnology* 2009, *81*, 813-826.
[17] Klamt, S., Gilles, E. D., Minimal cut sets in biochemical reaction networks. *Bioinformatics* 2004, *20*, 226-234.
[18] Melzer, G., Esfandabadi, M. E., Franco-Lara, E., Wittmann, C., Flux Design: In silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC systems biology* 2009, *3*, 120.
[19] Mahadevan, R., Schilling, C. H., The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering* 2003, *5*, 264-276.

[20] Price, N. D., Reed, J. L., Palsson, B. O., Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature reviews. Microbiology* 2004, *2*, 886-897.
[21] Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A*., et al.*, A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Molecular systems biology* 2011, *7*, 535.
[22] Mo, M. L., Palsson, B. O., Herrgard, M. J., Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology* 2009, *3*, 37.
[23] Burgard, A. P., Vaidyaraman, S., Maranas, C. D., Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnology progress* 2001, *17*, 791-797.
[24] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M*., et al.*, A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology* 2007, *3*, 121.
[25] Hashiguchi, K., Matsui, H., Kurahashi, O., Effects of a feedback-resistant aspartokinase III gene on L-isoleucine production in Escherichia coli K-12. *Bioscience, biotechnology, and biochemistry* 1999, *63*, 2023-2024.
[26] Awano, N., Wada, M., Kohdoh, A., Oikawa, T*., et al.*, Effect of cysteine desulfhydrase gene disruption on L-cysteine overproduction in Escherichia coli. *Applied microbiology and biotechnology* 2003, *62*, 239-243.
[27] Wang, Q., Chen, X., Yang, Y., Zhao, X., Genome-scale in silico aided metabolic analysis and flux comparisons of Escherichia coli to improve succinate production. *Applied microbiology and biotechnology* 2006, *73*, 887-894.

Table 1 – *E. coli* Modification Strategies Proposed by CosMos for iJO1366

Strategy	Increase Lower Bound	Decrease Upper Bound	Deletions	Min. Production Rate[a]
Succinate-1	AKGDH, ASPT	FUM		14.17
Succinate-2	ASPT, ICL	FUM		14.13
Succinate-3	ASPT, MALS	FUM		14.13
Aspartate-1	ASPK, ASPT, ASPTA, ADSL1r, PPA, PPKr			15.96
Aspartate-2	ASPT, ASPTA, NDPK1, R15BPK, ASAD		ADK1	15.96
Aspartate-3	ASPK, ASPT, ASPTA, ADSL1r PPA, ACS			15.96
Chorismate-1	CHORS, CHORM, GLNS	DHBS		4.87
Chorismate-2	CHORS, CHORM, GLCS1	PPA		4.87
Chorismate-3	DHBD, GLNS, DDPA	PPND, PPNDH	QUINDH	4.87

Overproduction strategies recommended by CosMos. Deletions are changes of bounds that force flux to zero (e.g., $\beta_j=UB_j$). Reaction abbreviations match those used in iJO1366: ASPT - aspartase, AKGDH – 2-oxogulatarate dehydrogenase, ICL – isocitrate lyase, MALS – malate synthase, ASPTA - aspartate transaminase, ADSL1r - adenylsuccinate lyase, PPA - inorganic diphosphatase, PPKr - Propionate kinase, NDPK1 - nucleoside-diphosphate kinase (ATP:GDP), R15BPK - ribose-1,5-bisphosphokinase, ASAD - aspartate-semialdehyde dehydrogenase, ADK1 – Adenylate Kinase, CHORS – Chorismate Synthase, CHORM – Chorismate mutase, GLNS – glutamate synthase, DHBS - 2,3-dihydroxybenzoate adenylate synthase, GLCS1 - glycogen synthase (ADPGlc), DHBD - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, DDPA - 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase, PPND -prephenate dehydrogenase, PPNDH - prephenate dehydratase, QUINDH - Quinate dehydrogenase [a] Units for minimum production rates are mmol/gDW/hr. The maximum theoretical production rates for succinate, aspartate and chorismate after accounting for biomass production and ATP maintenance are 14.87, 16.35 and 4.99 mmol/gDW/hr, respectively.

Table 2 – Comparison of CosMos Strategies for Leucine and Valine Production

Strategy	Increase Lower Bound	Decrease Upper Bound	Relative Performance
Leucine-1 (L1)	IPMD	IPPMIa	Low
Leucine-2 (L2)	LEUTAi, IPMD		Low
Leucine-3 (L3)	IPMD, AICART		Low
Leucine-4 (L4)	LEUTAi	CS	High
Leucine-5 (L5)	LEUTAi	GLNS	High
Valine-1 (V1)	ACLS	LEUTAi, KARA1	High
Valine-2 (V2)	ACLS	LEUTAi, IPMD	Medium
Valine-3 (V3)	ACLS, IPPMIa	LEUTAi	Medium
Valine-4 (V4)		KARA1, IPMD, GLNS	Low
Valine-5 (V5)	IPPMIa	KARA, HSDy	Low

Overproduction strategies tested in sampling analysis. Reaction abbreviations match those used in iJO1366: IPMD - 3-isopropylmalate dehydrogenase , IPPMIa – 3-isopropylmalate dehydratase, LEUTAi – leucine transaminase (irreversible), AICART – phosphoribosylaminoimidazolecarboxamide formyltransferase, CS – Citrate synthase , GLNS – glutamate synthase, ACLS –acetolactate synthase, ASPT - aspartase, KARA1 – ketol-acid reductoisomerase (2,3-dihydroxy-3-methylbutanoate), HSDy – homoserine dehydrogenase (NADPH). Relative performance refers to how often sampled sub-optimal flux modifications still result in high production.

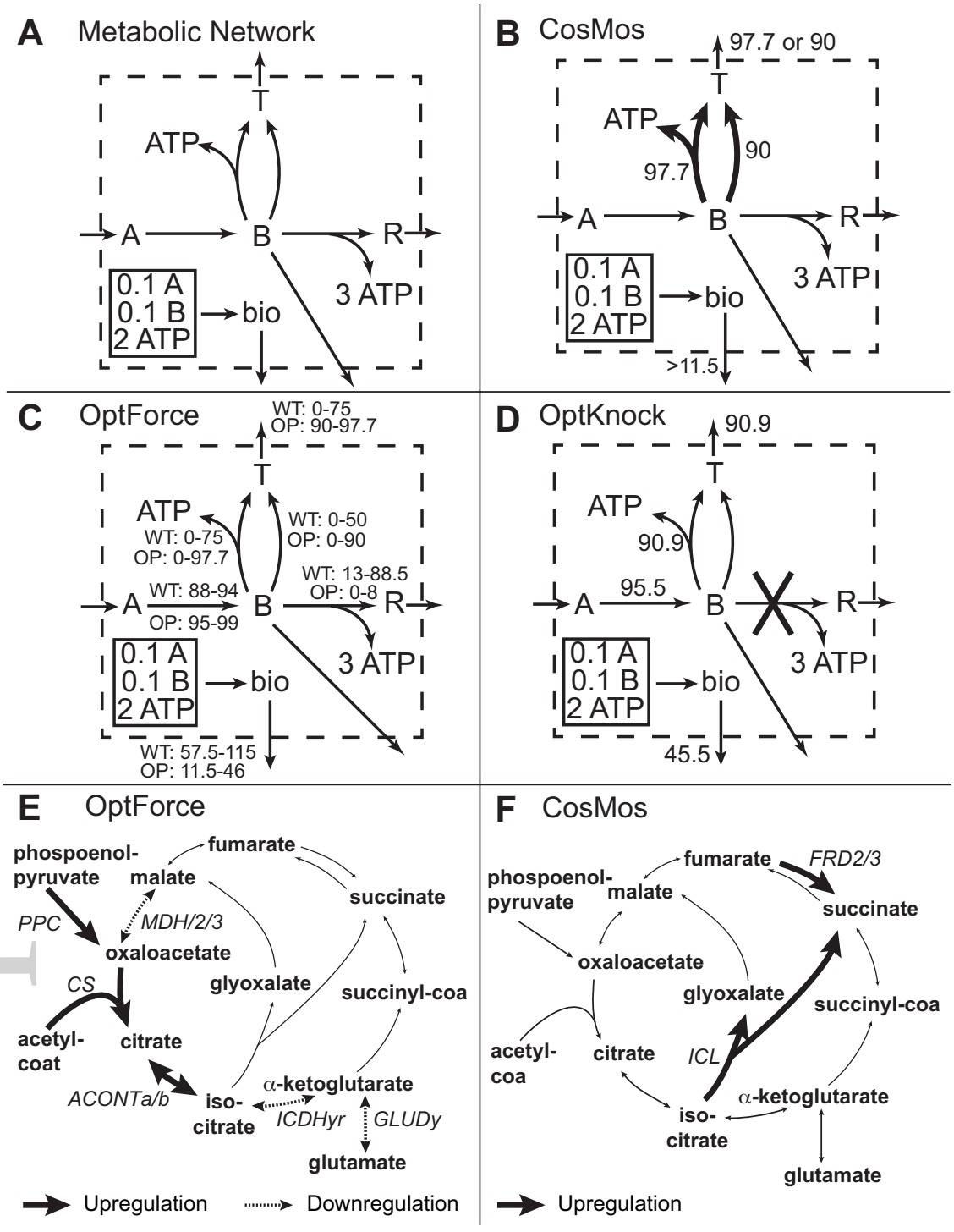

Figure 1 – Comparison of OptForce, OptKnock, and CosMos

A small network where CosMos gives recommendations for flux upregulation and downregulation but OptForce cannot. System boundaries are indicated by the dotted line, and any metabolite (except ATP) can be transferred out of the system freely. (A) The network consists of substrate A, intermediate B, by-product R, target chemical T, and ATP. Since production of R leads to the most ATP production, this would be the organism's preferred product to maximize biomass production. (B) Application of CosMos to the network to guarantee production of T. CosMos chose upregulation (thick arrows) of the ATP generating reaction by 97.7 to give the maximum theoretical production of T. An alternate solution was

also found involving upregulation of the non-ATP generating reaction by 90. (C) In OptForce, flux variability analysis finds the bounds for the wildtype strain (WT) network and overproducing (OP) strain (with production of T greater than 90). No upregulation strategies are found by OptForce because the lower limits for both T generating reactions in the OP network are zero. (D) OptKnock suggests deleting the B to R reaction (indicated by the X) and the resulting flux distribution for this strain at the maximum growth rate has production of T; however, at other growth rates the production of T can be zero. (E) Citric acid cycle modifications suggested by OptForce to produce succinate from glucose in an anaerobic environment (five additional modifications outside the citric acid cycle are not shown). The OptForce strategy was reported previously [6]. Many reactions both proximal and distal to succinate must be modified, and flux is directed through citrate into the glyoxylate shunt, and away from glutamate and malate. (B) CosMos suggested a simpler strategy for succinate involving the upregulation of the glyoxylate shunt (ICL reaction) and fumarate reductase (FRD2 or FRD3 reactions). Reaction abbreviations: PPC - phosphoenolpyruvate carboxylase, CS - citrate synthase, MDH/2/3 – malate dehydrogenase, ACONTa/b – aconitase a and b, ICL – isocitrate lyase, ICDHyr - isocitrate dehydrogenase (NADP), FRD2/3 – fumarate reductase, GLUDy - glutamate dehydrogenase

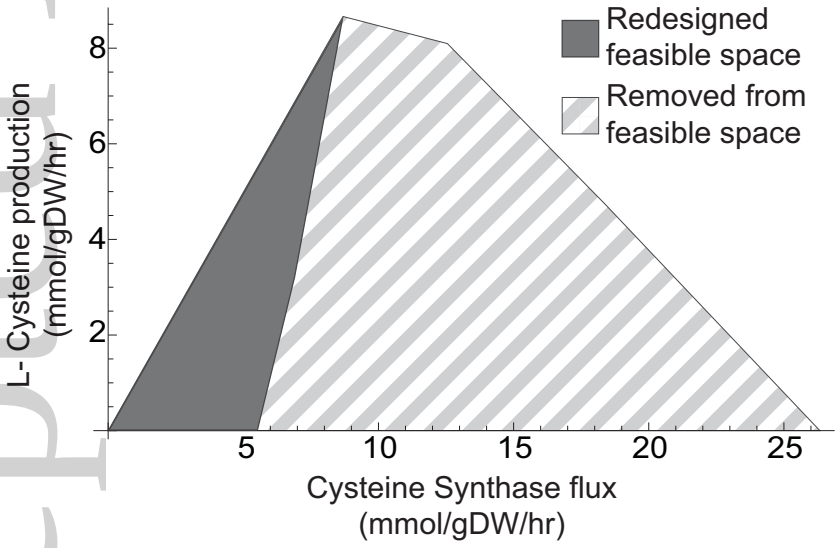

Figure 2 – Feasible Flux Space for Parental and Modified Network

For a glucose uptake rate of 10 mmol/gDW/hr, CosMos suggested downregulating aspartate kinase (ASPK) and cysteine desulfhydrase (CYSD) and upregulating cysteine synthase (CYSS) by 20.65, 28.36 and 8.67 mmol/gDW/hr, respectively to guarantee the maximum theoretical production of cysteine (8.67 mmol/gDW/hr). Without CYSD and ASPK downregulation, the union of the dark grey and hatched light grey areas shows the feasible combinations of CYSS flux and cysteine production. All feasible combinations of CYSS flux and cysteine production after downregulation of ASPK and CYSDS are shown in dark grey, while the hatched area is no longer feasible due to the downregulations. It is easy to see that upregulation of CYSS by 8.67 flux units will give the maximum theoretical production. Given this amount of CYSS upregulation, no smaller value of cysteine production is possible.

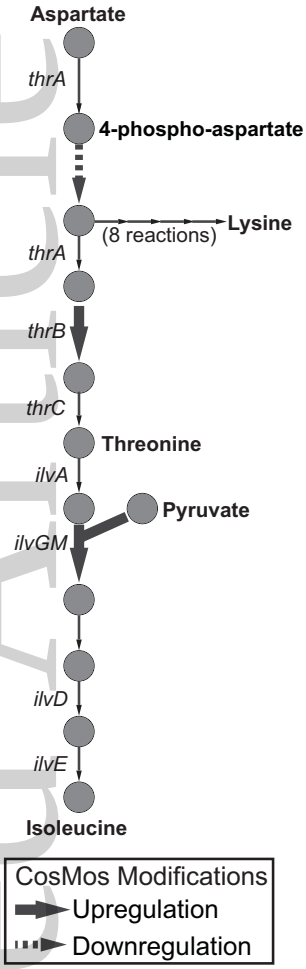

Figure 3 – Comparison to previous strategy for isoleucine production

The isoleucine biosynthesis pathway shows a strategy implemented experimentally that produced isoleucine. In this previous strategy, all of the named genes were upregulated compared to the wild type, except for *ilvA*, which had its feedback inhibition removed as another way to increase flux. Thicker arrows indicate fluxes should be upregulated according to CosMos, while the dotted arrows indicate what should be downregulated. The CosMos strategy shown guarantees 43 mol of isoleucine per 100 mol of glucose under aerobic conditions.

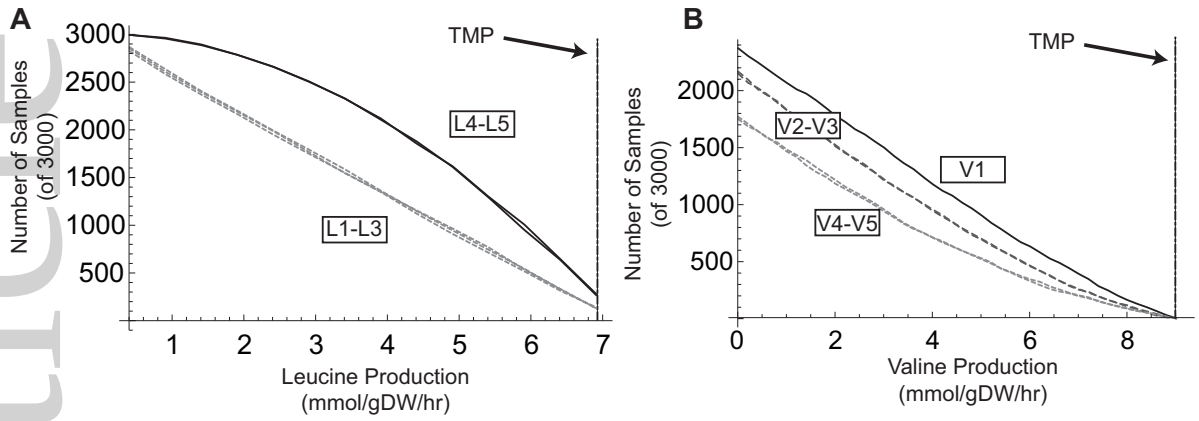

Figure 4 – Sampling of Sub-Optimal Flux Modifications

Survival plots of production after sampling of sub-optimal flux modifications. Optimal up- and downregulation values (α and β) were found for each CosMos strategy that produces the theoretical maximum production (TMP). For each strategy, 3,000 random values for each α and β were drawn from a uniform distribution between zero and the optimal value. Survival plots were generated by showing how many of the 3,000 randomly sampled sub-optimal flux modifications would guarantee production at a level greater than or equal to the corresponding value on the x-axis. (A) Survival plots for leucine strategies reported in Table 2 which require two modifications to achieve theoretical maximum production. Three strategies (L1-L3) have a uniform distribution of production rates when sub-optimal flux modifications were used, but two strategies (L4 and L5) gave production rates that were frequently closer to the theoretical maximum. (B) Survival plots for valine strategies reported in Table 2 which require three modifications. Here, one strategy (V1) is clearly superior than the other four since a larger fraction of samples were able to achieve a given production rate.