Bioinformatics, 31(17), 2015, 2844–2851
doi: 10.1093/bioinformatics/btv217
Advance Access Publication Date: 25 April 2015
Original Paper

Systems biology

Genome-scale strain designs based on regulatory minimal cut sets

Radhakrishnan Mahadevan,[1,2,*] Axel von Kamp[3] and Steffen Klamt[3,*]

[1]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON, M5S3E5, Canada, [2]Institute of Biomaterials and Biomedical Engineering, Toronto, ON, M5S 3G9, Canada and [3]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, D-39106, Germany

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on December 2, 2014; revised on March 31, 2015; accepted on April 16, 2015

Abstract

Motivation: Stoichiometric and constraint-based methods of computational strain design have become an important tool for rational metabolic engineering. One of those relies on the concept of constrained minimal cut sets (cMCSs). However, as most other techniques, cMCSs may consider only reaction (or gene) knockouts to achieve a desired phenotype.

Results: We generalize the cMCSs approach to constrained *regulatory* MCSs (cRegMCSs), where up/downregulation of reaction rates can be combined along with reaction deletions. We show that flux up/downregulations can virtually be treated as cuts allowing their direct integration into the algorithmic framework of cMCSs. Because of vastly enlarged search spaces in genome-scale networks, we developed strategies to (optionally) preselect suitable candidates for flux regulation and novel algorithmic techniques to further enhance efficiency and speed of cMCSs calculation. We illustrate the cRegMCSs approach by a simple example network and apply it then by identifying strain designs for ethanol production in a genome-scale metabolic model of *Escherichia coli.* The results clearly show that cRegMCSs combining reaction deletions and flux regulations provide a much larger number of suitable strain designs, many of which are significantly smaller relative to cMCSs involving only knockouts. Furthermore, with cRegMCSs, one may also enable the fine tuning of desired behaviours in a narrower range. The new cRegMCSs approach may thus accelerate the implementation of model-based strain designs for the bio-based production of fuels and chemicals.

Availability and implementation: MATLAB code and the examples can be downloaded at http://www.mpi-magdeburg.mpg.de/projects/cna/etcdownloads.html.

Contact: krishna.mahadevan@utoronto.ca or klamt@mpi-magdeburg.mpg.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Recent advances in genome-wide characterization of cellular systems have allowed the opportunity to catalogue a significant fraction of the metabolic reactions in the cell. These advances coupled with the development of metabolic modelling methods have enabled the construction of detailed models for many industrially relevant microbial hosts including *Escherichia coli* and *Saccharomyces cerevisiae* (McCloskey *et al.*, 2013; Monk *et al.*, 2014; Nookaew *et al.*, 2011). In parallel, the increasing price and volatility associated with petroleum-based feedstocks have stimulated the use of biological processes for renewable chemical synthesis. Together, these two factors further have motivated the development of computational methods to guide the engineering of metabolism for renewable chemicals synthesis.

In the past, several algorithms have been developed for computational strain design, most of them relying on stoichiometric and constraint-based models. The latter includes the series of bilevel optimization algorithms such as OptKnock (Burgard *et al.*, 2003), OptStrain (Pharkya *et al.*, 2004), OptGene (Patil *et al.*, 2005), OptReg (Pharkya and Maranas, 2006), CosMos (Cotten and Reed, 2013) and OptORF (Kim and Reed, 2010), which have been recently reviewed (Zomorrodi *et al.*, 2012). In all of these methods, the inner optimization problem involves the formulation of a genome-scale metabolic model with growth rate maximization as the objective, whereas the outer optimization involves the identification of the specific deletions/additions (represented as binary variables) that lead to the maximization of the flux to the target chemical. Typically, duality theory is used to convert the bilevel optimization problem into a single level mixed integer linear optimization problem with both the primal and dual version of growth maximization linear programs. Another method based on mixed integer linear optimization is OptForce (Ranganathan *et al.*, 2010), which does not rely on a cellular objective function and uses instead a reference flux distribution typically obtained from ^{13}C labelling experiments. By solving a series of Mixed Integer Linear Programming (MILPs), OptForce first identifies all reaction pairs, triples, quadruples, etc. that must be upregulated or downregulated or deleted to achieve a desired production threshold. In the second stage, these MUST sets are used to identify a minimal intervention strategy that can lead to desired production levels. Hence, OptForce has been used to identify lower-order combinations of flux changes required to obtain a desired phenotype and was recently extended to k-OptForce to include kinetics of a subset of metabolic reactions (Chowdhury *et al.*, 2014). Finally, we have previously developed an integer-free optimization approach that considers up/downregulation based on successive linear programming and was shown to be highly efficient (EMILiO) (Yang *et al.*, 2011). However, identifying strategies with minimal number of interventions is still a challenge.

Model-based strain designs based on optimization approaches have been experimentally validated for lactate, butanediol, malonyl-CoA and fatty acid production in *E.coli*, and vanillin production in *S.cerevisiae* highlighting the value of *in silico* strain designs for metabolic engineering (Brochado *et al.*, 2010; Fong *et al.*, 2005; Ranganathan *et al.*, 2012; Xu *et al.*, 2011). However, solutions of bilevel optimization problems for genome-scale models for cases with more than four or so modifications are usually computationally prohibitive, even more if one wants to enumerate alternate intervention strategies.

Other methods for computational strain design are based on pathway-centric approaches, which also do not involve a cellular objective function. These methods analyze and design the space of metabolic behaviours by means of elementary modes (EMs; Schuster *et al.*, 2000; Trinh *et al.*, 2009). Examples are minimal metabolic functionality (Trinh, 2012; Trinh *et al.*, 2008, 2011), SMET (Flowers *et al.*, 2013), FluxDesign (Melzer *et al.*, 2009) and CASOP (Hädicke and Klamt, 2010). However, these methods cannot be applied to genome-scale networks due to the combinatorial explosion of EMs in large networks.

An alternative approach, which, in contrast, to the aforementioned methods, seeks to enumerate all relevant intervention strategies, is based on minimal cut sets (MCSs). MCSs represent reaction/gene knockout sets eliminating all EMs with undesired phenotypes (e.g. low product yield) (Klamt, 2006; Klamt and Gilles, 2004). As an extension, constrained MCSs (cMCSs) were introduced to allow also the consideration of desired phenotypes to be maintained in the network (Hädicke and Klamt, 2011). Previously, calculation of (constrained) MCSs required the calculation of the EMs of the metabolic network in a pre-processing step; they are used to specify desired and undesired metabolic behaviours from which the resulting (c)MCSs can then be determined (Hädicke and Klamt, 2011). However, a recent study showed the equivalence between the MCSs of the primal and EMs of a dual problem (Ballerstein *et al.*, 2012). In principle, this advance enables the direct identification of MCSs using the network stoichiometry and flux capacity constraints without requiring the enumeration of all EMs of the network in a pre-processing step. However, enumeration of all MCSs in large-scale networks was still infeasible due to the large number of cut sets (or, equivalently, of EMs in the dual). More recently, the duality approach was used to identify the smallest MCSs in genome-scale metabolic networks as shortest EMs in the dual (von Kamp and Klamt, 2014). Once the smallest MCSs deleting all undesired phenotypes (the latter represented by linear inequalities) are identified, additional inequality constraints describing the desired metabolic behaviour (e.g. high product yield with some minimal biomass synthesis) can be used to filter the valid cMCSs from the calculated MCSs. The findings in von Kamp and Klamt (2014) highlighted the ability of these cMCSs to identify a large number of (partially novel) deletion strategies that are not typically found by growth-coupled bilevel optimization problems. However, one of the limitations of cMCSs is that they are limited to identifying only deletion modifications. Given the recent advances in synthetic biology (Keasling, 2012) that provide an unparalleled opportunity to manipulate and fine-tune expression of metabolic enzymes in cells, there is a need to consider graded regulation of metabolic reaction rates in strain design methods.

Consequently, herein we present an extended cut-set approach allowing the calculation of (constrained) *regulatory* MCSs where up/downregulation of metabolic reaction rates can be combined along with reaction (or gene) deletions. We present a modification by which up/downregulations can be treated as virtual cuts allowing their direct integration in the existing algorithmic framework of MCSs. As the generalized intervention strategies may lead to vastly enlarged search spaces in genome-scale networks, we developed strategies to (optionally) preselect suitable candidates of reactions to be considered for regulation. We illustrate this approach by a small example and then by identifying strain designs that lead to ethanol production in a genome-scale metabolic model of *E.coli*. The results clearly show that using the combined regulation and deletions allowed in constrained regulatory MCS, one can find strategies that are significantly smaller relative to the conventional cMCSs involving only knockouts.

We also present new algorithmic techniques to enhance efficiency and speed of cut set-calculations. For example, the currently employed two-step approach in which first all MCSs are enumerated in the dual before the admissible cMCSs (that keep desired behaviours feasible) are filtered in the primal system can be integrated in one step bringing several advantages.

2 Methods

2.1 Metabolic networks, cut sets and desired and undesired phenotypes

We are given a metabolic (stoichiometric) reaction network with m metabolites and n reactions represented by its stoichiometric matrix $\mathbf{N}$. The set Irrev contains the indices of irreversible reactions. We

assume that the network is in steady state meaning that the net reaction rates $\mathbf{r} = (r_1, r_2, \ldots, r_n)^T$ fulfil

$$\mathbf{Nr} = \mathbf{0},\ r_i \geq 0\ \forall i \in \text{Irrev.} \quad (1)$$

Undesired (target) flux vectors (undesired phenotypes) can be defined by a system of t linear inequalities

$$\mathbf{Tr} \leq \mathbf{t} \quad (2)$$

with an appropriate $t \times n$ matrix $\mathbf{T}$ and $t \times 1$ vector $\mathbf{t}$. These inequalities can, for instance, be used to describe flux distributions having a product yield below a minimum desired value. Similarly, the inequality system

$$\mathbf{Dr} \leq \mathbf{d} \quad (3)$$

with appropriate $d \times n$ matrix $\mathbf{D}$ and $d \times 1$ vector $\mathbf{d}$ is used to represent desired phenotypes, for instance, flux distributions with a high product yield coupled with some biomass synthesis. Note that the pair **D/d** will usually contain the negated version of constraint (2) implying that desired flux vectors violate (2). Hence, the desired space will be a subset of the complement of the undesired space (e.g. solutions with high product yield) but not necessarily the full complement since other constraints (e.g. some minimum biomass yield) might be added to **D/d** to characterize desired solutions. Therefore, flux vectors with high product but low biomass yield are usually neither contained in the undesired nor in the desired space; these irrelevant solutions are thus allowed to be deleted or to survive. Note also that all flux bounds in the system need to be integrated in both **T** and **D**.

An MCS is a support-minimal set of reaction knockouts, such that no flux vector **r** in the reduced network can fulfil (1) and (2) anymore, i.e. the undesired phenotypes are blocked. A constrained MCS (cMCS) fulfils in addition that there still exists a rate vector **r** satisfying (1) and (3). The set of cMCSs is always a subset of the MCSs.

In the following, we want to generalize this approach by allowing not only knockouts but also up/downregulation of certain reaction rates. Here, an up/downregulation refers to setting a certain lower (upregulation) or upper (downregulation) boundary for the respective rate. Combinations of knockouts and up/downregulations that block the undesired phenotypes will be called *regulatory MCS* (RegMCS) and the subset of RegMCSs that keep some desired phenotypes feasible are the *constrained regMCS* (cRegMCSs). From these definitions, it is clear that the (c)MCSs are a subset of the (c)RegMCSs. Importantly, whenever we use the term 'cut set(s)' (non-abbreviated), we refer to unspecific sets of constrained/unconstrained or/and regulatory/non-regulatory MCS. For example, if we say 'cut-set algorithm', we mean an algorithm that can compute any type of cut sets.

2.2 Representing up- and downregulation of reaction rates as cuts

In the following, we show that up/downregulations can formally be treated as cuts, i.e. by knocking out an (artificial) reaction. This extension will allow us to use available cut set algorithms to calculate not only (c)MCSs but also (c)RegMCSs. Each reaction i considered for regulation is modified so that it produces a pseudometabolite M_i (Fig. 1), which is then consumed by an introduced *demand reaction*. The rate e_i of this demand reaction is either constrained by a lower boundary $e_{i,\min}$ (in the case of upregulation; Fig. 1a) or by an upper boundary $e_{i,\max}$ (in the case of downregulation, Fig. 1b). These modifications force the same upper and lower bounds also for the rate r_i of reaction i thereby forcing the up/downregulation of the flux through the corresponding reaction. We then add a *slack reaction* (with rate s_i) that produces the pseudometabolite (for upregulation) or consumes it (for downregulation). This modification relaxes the constraints on the flux r_i (the rate is free in its original boundaries); however, if the slack reaction is cut (deleted), rate r_i becomes up- or downregulated, respectively. We can therefore apply the duality-based cut set calculation approaches (Ballerstein *et al.*, 2012; von Kamp and Klamt, 2014) to this extended network to identify deletions in 'normal' reactions (representing knockouts) as well as in the added slack reactions representing then up/downregulations (the introduced demand reactions are not allowed to be cut). If the up/downregulation of a specific reaction can contribute to the elimination of the unwanted behaviours, then the corresponding slack reaction is deleted. The calculation of (c)RegMCSs can thus be reduced to calculating (c)MCSs in the extended network. Importantly, the feasible flux space in the network is not changed by these modifications (unless the slack reaction is cut).

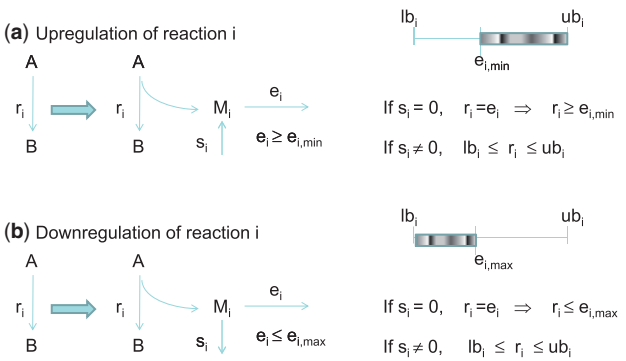


Fig. 1. Network modifications at a reaction i to find regulatory MCSs involving upregulation (a) or downregulation (**b**) of this reaction

2.3 Flux variability analysis-based identification of promising reactions for regulation

The approach introduced above requires specification of the levels of flux regulation (the $e_{i,\min}$ and $e_{i,\max}$ in Fig. 1), which are usually not known before. Moreover, the addition of pseudometabolites and demand and slack reactions for all reactions in a genome-scale model (with possibly different thresholds for the demand reactions) will quickly lead to a combinatorial explosion of integer variables leading to poor computational performance. In the past, algorithms such as FVSEOF (Park *et al.*, 2012) or OptForce (Ranganathan *et al.*, 2010) have been used to identify reaction subsets that are correlated to production of a target metabolite, we therefore use a flux variability analysis (FVA) -based preprocessing algorithm (Mahadevan and Schilling 2003) to (i) identify a subset of promising candidate reactions for up/downregulation and (ii) to determine meaningful flux boundaries (the $e_{i,\min}$ and $e_{i,\max}$) to be used for their up/downregulation. Accordingly, we perform FVA for (i) the 'wild type' [Equation (1) plus known flux boundaries (e.g. substrate uptake rate; see Section 3.2)] and for (ii) the 'desired space' [Equations (1) and (3) and known flux constraints] and identify reactions whose ranges were reduced most under the desired conditions and rank them based on the extent of range reduction. From this list, one may then choose (e.g. the top 10) candidates to be considered for regulation. Modifications in these reactions are then introduced as

described above (Fig. 1) with several equidistant boundaries for the demand reactions e_i; meaningful thresholds are derived from the FVA analysis in the desired system. Typically, we allow 8–10 reactions to be regulated (both up/down-regulation) at three levels. Using the algorithm of von Kamp and Klamt (2014), we then calculate the resulting cRegMCSs and identify three or four reactions that were most frequently regulated in the 'coarse' initial cRegMCSs. Subsequently, for each of these frequent reactions, we specify much finer regulation levels and then identify the 'fine' cRegMCSs with the least amount of interventions. See Section 3 for details on how we concretely proceeded in the case studies.

2.4 Algorithmic improvements to enhance efficiency of calculation of cMCSs

2.4.1 Direct enumeration of cMCSs

In genome-scale networks, cut set calculations will usually have to focus on the smallest intervention strategies. The two-step method for identifying smallest cMCSs in genome-scale networks described in von Kamp and Klamt (2014) identifies first the smallest (unconstrained) MCSs as shortest EMs in the dual system by solving a series of MILP problems. These MCSs are then filtered in a second step to keep only the valid cMCSs. This filtering is done by solving one linear optimization problem per MCS to check whether the desired phenotype is still feasible for the MCS [i.e. whether a solution **r** exist fulfilling (1) and (3) under the given cuts]. In the following, we formulate an alternative approach for the *direct* enumeration of cMCSs in one step by integrating the constraints for the undesired and desired phenotypes directly in one integrated MILP problem.

The central equation used by von Kamp and Klamt (2014) in the MILP problem for calculating the MCSs blocking the undesired flux vectors [obeying Equations (1) and (2)] reads as follows:

$$\begin{pmatrix} \mathbf{N}_{\text{rev}}^T & \mathbf{I}_{\text{rev}} & -\mathbf{I}_{\text{rev}} & \mathbf{T}_{\text{rev}}^T \\ \mathbf{N}_{\text{irr}}^T & \mathbf{I}_{\text{irr}} & -\mathbf{I}_{\text{irr}} & \mathbf{T}_{\text{irr}}^T \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{vp} \\ \mathbf{vn} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} 0 \\ \geq \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$\mathbf{t}^T\mathbf{w} \leq -c$$
$$\mathbf{u} \in \Re^m, \mathbf{vp}, \mathbf{vn} \in \Re^n, \mathbf{w} \in \Re^t, \mathbf{vp}, \mathbf{vn}, \mathbf{w} \geq \mathbf{0}, c > 0 \quad (4)$$

Note that, before setting up (4), the stoichiometric matrix **N** and the matrix **T** used for specifying the undesired phenotypes in inequality (2) have been split to two submatrices containing only the reversible ($\mathbf{N}_{\text{rev}}$ and $\mathbf{T}_{\text{rev}}$) and irreversible ($\mathbf{N}_{\text{irr}}$ and $\mathbf{T}_{\text{irr}}$) reactions, respectively. The matrices $\mathbf{I}_{\text{rev}}$ and $\mathbf{I}_{\text{irr}}$ represent matrices containing identity submatrices for the reversible or irreversible reactions, respectively. MCSs are then identified by determining solutions to (4) having the smallest number of non-zero entries in the variables **vp** and **vn** (these non-zero entries indicate knockouts in the corresponding reactions). Binary indicator variables **zp** and **zn** are introduced (together with the constraint $zp_i + zn_i \leq 1$) to represent the non-zero entries in **vp** and **vn** efficiently and to pose the objective function of the resulting MILP optimization problem:

$$\text{minimize} \sum_{i=1}^{q} (zp_i + zn_i)$$
$$s.t. \quad (4).$$

Once an MCS has been found, exclusion constraints (integer cuts) must be added to the MILP to ensure that this MCS and supersets thereof will not be found in later iterations (cf. von Kamp and Klamt, 2014).

To ensure that the calculated cut sets maintain some solutions of the desired space, we incorporate the corresponding constraints (1) and (3) for the desired flux vectors **r** directly in (4):

$$\begin{pmatrix} \mathbf{N}_{\text{rev}}^T & \mathbf{I}_{\text{rev}} & -\mathbf{I}_{\text{rev}} & \mathbf{T}_{\text{rev}}^T & 0 \\ \mathbf{N}_{\text{irr}}^T & \mathbf{I}_{\text{irr}} & -\mathbf{I}_{\text{irr}} & \mathbf{T}_{\text{irr}}^T & 0 \\ 0 & 0 & 0 & 0 & \mathbf{N} \\ 0 & 0 & 0 & 0 & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{vp} \\ \mathbf{vn} \\ \mathbf{w} \\ \mathbf{r} \end{pmatrix} \begin{matrix} = \\ \geq \\ = \\ \leq \end{matrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mathbf{d} \end{pmatrix}$$
$$\mathbf{t}^T\mathbf{w} \leq -c$$
$$\mathbf{u} \in \Re^m, \mathbf{vp}, \mathbf{vn}, \mathbf{r} \in \Re^n, \mathbf{w} \in \Re^t, \mathbf{d} \in \Re^d, \mathbf{vp}, \mathbf{vn}, \mathbf{w}, \mathbf{r}_{Irrev} \geq \mathbf{0}, c > 0 \quad (5)$$

By (5), we thus demand infeasibility of the undesired behaviours (by finding some minimal solutions for their dual representation) while ensuring existence of a solution for the (primal) desired behaviours. As an additional requirement to (5), we need to express that reaction rate r_i in **r** is zero if a cut has been made in reaction *i*:

$$r_i \leq (1 - zp_i - zn_i)ub_i \text{ and } r_i \geq (1 - zp_i - zn_i)lb_i. \quad (6)$$

Here, lb_i and ub_i are the lower and upper bounds of reaction *i*. For this integrated approach, it is thus mandatory to have finite lower and upper boundaries for all reaction rates, which is, however, not a limitation since one may set large absolute flux ranges if the bounds are not known. Alternatively, one may use FVA to determine these ranges for the desired conditions, ideally under consideration of thermodynamic consistency (Muller and Bockmayr, 2013). Solving the MILP with the same objective function ('minimize $\sum_{i=1}^{q} zp_i + zn_i$'), now subject to (5) and (6), will directly deliver the cMCSs in one step. Clearly, the dimension of this integrated MILP is considerably larger than the one spanned by (4), but it brings two major advantages. First, MCSs need not to be checked anymore for feasibility of desired behaviours. However, as the LPs required for testing the found MCSs for feasibility of desired behaviours are rather fast, only a minor speed enhancement can be expected here. The second advantage weights more: since a lower number of solutions (cMCSs) will be computed by the integrated MILP, lower number of integer cuts need to be integrated to exclude previously found solutions [cf. the case study in von Kamp and Klamt (2014): only 3–6% of the found MCSs are eventually kept as cMCSs]. We observed that this can significantly enhance the relative speed of the integrated algorithm compared with the standard approach, especially at later iterations.

2.4.2 Additional revisions for enhancing computational efficiency

To increase the performance of the whole enumeration procedure, the metabolic network is usually compressed by removing reactions that are blocked in steady state and by combining reactions that operate with a fixed ratio into reaction subsets. Both these techniques have been applied also in von Kamp and Klamt (2014). In addition, two new techniques are applied here to reduce the search space during MCSs enumeration (which can be applied to both the integrated and the non-integrated MILP): it was already mentioned in von Kamp and Klamt (2014) that reactions being essential for the desired phenotypes should be removed from the set of deletable reactions as they cannot be part of any admissible cMCS (although they can occur in the MCSs). For example, if growth is a desired functionality then reactions essential for the production of a biomass component should be removed from the set of reaction knockout candidates. This idea can be carried further by determining also

reaction pairs that cannot be deleted simultaneously for the desired conditions and by excluding them and their supersets from the search space [via constraints of the form shown by Equation (20) in von Kamp and Klamt (2014)]. In principle, this method could be extended to even reaction triples or higher combinations, but pairs can usually be very quickly calculated and there is likely to be a trade-off between the reduction of the search space and the complication of the MILP through additional constraints.

Another technique is to add constraints ensuring that only one of the potentially many slack reactions linked to the *same* regulated metabolic reaction can be knocked-out at a given time. The rationale behind this restriction is that only one regulatory constraint (one up- or downregulation) can be active on a given reaction in a cRegMCS because any additional regulatory constraint would be either contradictory or redundant. Both techniques mentioned above do not affect the resulting cMCSs and are only used to tune the performance.

2.5 Implementation

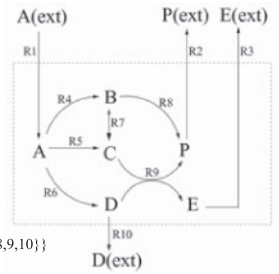
The calculation of cRegMCSs has been integrated as a new functionality in the *CellNetAnalyzer* package, a MATLAB toolbox for biological network analysis (Klamt *et al.*, 2007). The implementation uses the IBM ILOG CPLEX Optimization Studio for solving the respective MILP and LP problems. Arbitrary intervention problems can be defined by providing the (i) stoichiometric network, (ii) the matrices (**D**,**T**) and vectors (**d**,**t**) specifying the desired and undesired flux vectors and (iii) the list of reactions considered for regulation together with their respective flux levels (or the number of intervals to be used for each regulated reaction). *CellNetAnalyzer* can be downloaded from the following website: http://www.mpi-magdeburg.mpg.de/projects/cna. A separate package containing the data and script files needed for running the examples discussed herein can be downloaded from http://www.mpi-magdeburg.mpg.de/projects/cna/etcdownloads.html.

3 Results

We applied the cRegMCSs approach (i) to a small illustrative network used previously in Hädicke and Klamt (2011) and (ii) to a genome-scale metabolic model of *E.coli* (iAF1260; Feist *et al.*, 2007) where we calculate intervention strategies based on regulatory MCSs for ethanol overproduction as a realistic application.

3.1 cRegMCSs for a small example network

The small example network (Fig. 2) is used to illustrate the differences in the type of manipulations identified by cRegMCSs compared with the traditional cMCSs as described in Hädicke and Klamt (2011). Here, the metabolite P is designated as the product and we intend to identify intervention strategies enforcing greater than 60% yield of P on substrate A. Using deletions alone, four cMCSs can be identified [cf. also (Hädicke and Klamt, 2011)]. These strategies are mainly intuitive and suggest the deletion of alternative pathways leading to the production of the competing metabolites D and E. Calculating the cRegMCSs with our new approach (reactions R4, R5, R7, and R8 were considered for upregulation; for the regulatory levels $e_{i,\min}$ and $e_{i,\max}$ we considered equidistant values of 1,2,... 10) significantly expands the number of strategies found to 53, including the purely deletion-based cMCSs identified earlier. The cRegMCSs suggest in addition strategies combining knockouts and regulations or even those that are purely based on regulation of one, two or three fluxes (Fig. 2). We also



Constraints: 1≤R1≤10
Undesired Space:
$Y^{P/A}$ = R2/R1≤0.6 (→R2-0.6R1≤0)
Desired Space:
$Y^{P/A}$ = R2/R1>0.6 (→0.6R1-R2<0)

cMCS with deletions alone:
MCS1-4={R6},{R3,R10},{R9,R10},{R5,R7,R10}

cRegMCS with deletions and upregulations in R4,R5,R7,R8:
MCS1-4={R6},{R3,R10},{R9,R10},{R5,R7,R10}
With regulation of one reaction:
MCS5-16={R4≥{7,8,9,10}},{ R5≥{7,8,9,10}},{ R8≥{7,8,9,10}}
MCS17-20={R10 & 1 from R8≥{3,4,5,6},
MCS21-24={R7,R10 & 1 from R4≥{3,4,5,6}}
With regulation of two reactions:
MCS25-42={1 from R4≥{1,2,3,4,5,6} & 1 from R5≥{1,2,3,4,5,6} such that R4+R5≥{7,8,9,10}}
MCS43-46={R7≥1,R8≥6},{R7≥2,R8≥6},{R7≥2,R8≥5},{R7≥3, R8≥4}
MCS47-50={R7 and one from {R5≥1,R8≥6}, {R5≥2,R8≥5}, {R5≥3,R8≥4},{R5≥2,R8≥6}}
With regulation of three reactions:
MCS51-53={R5≥1,R7≥2,R8≥4},{R5≥1,R7≥1,R8≥5}, {R5≥2,R7≥1,R8≥4}

Fig. 2. Illustration of cRegMCSs in a small example network. We assume that all reactions can be knocked out, whereas the flux levels of reactions R4,R5,R7 and R8 are allowed to be upregulated with flux minimum thresholds in [1,2,... 10]

notice that there are some redundant cRegMCSs. For example, from the four MCS5-MCS8 = R4 ≥ {7,8,9,10} we only need to keep {R4>=≥7} and discard the three others (R4 ≥ {8,9,10}) since the latter are only special cases of {R4 ≥ 7} and thus redundant. In the Supplementary Text S1, we explicitly list the 24 non-redundant cRegMCSs. Afterwards one may furthermore group cRegMCSs into *classes of intervention strategies* in each of which all cRegMCSs use the same combinations of reaction knockouts and regulations but with different thresholds for the regulated fluxes. For example, MCS51–MCS53 are not redundant to each other but belong to one and the same class (up-regulation of R5, R7 and R8). In the Supplementary Material, we also list the 13 classes of the (24 non-redundant) cRegMCSs.

These results demonstrate the largely extended space of suitable intervention strategies when taking not only deletions but also graded regulation of metabolic fluxes into account. For a condensed representation, redundant cRegMCSs may be dropped and classes of cRegMCSs easily identified in a postprocessing step.

3.2 Ethanol production in genome-scale metabolic network of *E.coli*

To further evaluate the cRegMCS method and to understand how the computational complexity changes with the scale of the network, we applied this method to identify strain designs in a genome-scale network of *E.coli* enforcing growth-coupled production of ethanol. We used the iAF1260 model (Feist *et al.*, 2007) with a glucose uptake rate of 18.5 mmol/gDW/h^{-1} and an ATP maintenance demand of 8.39 mmol/gDW/h^{-1} under anaerobic conditions (Portnoy *et al.*, 2010). We aimed to find cRegMCSs for anaerobic conditions that guarantee an ethanol yield greater than 1.4 mol/mol glucose while still enabling growth with a rate greater than 0.5 h^{-1} [these values are used to specify the constraints of desired and undesired behaviors in Equations (2) and (3)]. This approach resembles the cMCSs analysis done in von Kamp and Klamt (2014) with the difference that we choose a larger minimum growth rate (von Kamp and Klamt used 0.001 h^{-1}) to enable sufficient biomass yields. As the search space for flux deletions combined with flux regulations becomes quite complex, we use FVA in a preprocessing step as described in Section 2 to identify 10 reactions that had the largest reduction in their flux range when comparing the desired versus the entire flux space. For these

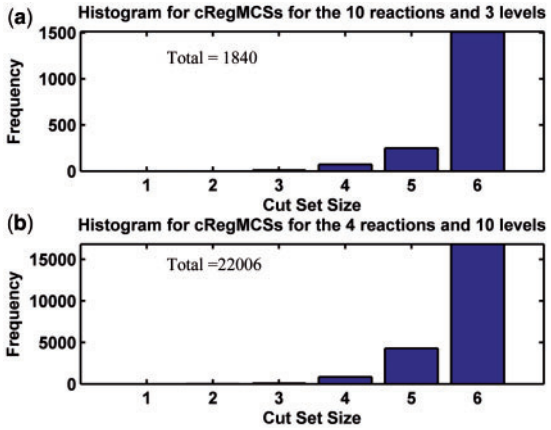


Fig. 3. Size distribution of the cRegMCSs up to size 6 for the genome-scale *E.coli* model enforcing ethanol yield above a desired threshold (1.4 mol ethanol/mol glucose). (**a**) Histograms for the case with deletions and regulation of 10 reactions at 3 levels. (**b**) Histograms for the case with deletions and regulation of 4 reactions at 10 levels computed in this paper. Note that all cRegMCSs involve at least one regulation; cRegMCSs using only deletions (= cMCSs) require at least seven knockouts

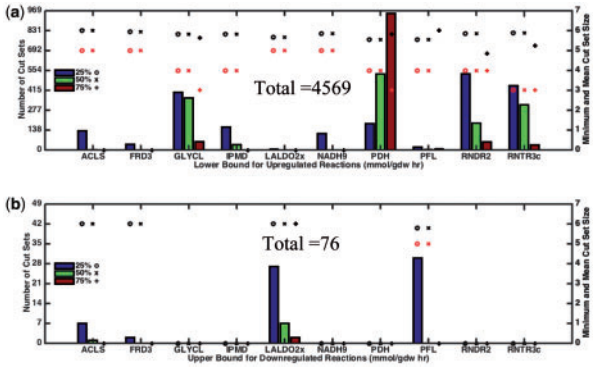


Fig. 4. Analysis of the 1840 cRegMCSs for the case with 10 regulated reactions at 3 levels up to size 6 for the genome-scale *E.coli* model enforcing ethanol yield above a desired threshold. Number of the cut sets involving each of the 10 regulated reactions at different levels is shown. The minimum (red symbols) and the mean (black symbols) cut-set size (right side *y* axis) for each reaction at the three different levels of regulation [25% (o), 50% (x), 75% (+) of the range] is shown. (**a**) cRegMCSs involving reactions with upregulated flux. (**b**) cRegMCSs involving reactions which with downregulated flux. Note that some of the cut sets include both up- and downregulation and will appear in both the plots. ACLS: acetolactate synthase; FRD3: fumarate reductase; GLYCL: glycine cleavage system; IPMD: 3-isopropylmalate dehydrogenase; LALDO2x: D-lactaldehyde:NAD+ 1-oxido-reductase; NADH9: NADH dehydrogenase; PDH: pyruvate dehydrogenase; PFL: pyruvate formate lyase; RNDR2: ribonucleoside-diphosphate reductase (GDP); RNTR3c: ribonucleoside-triphosphate reductase (CTP) (flavodoxin)

10 reactions, we considered three levels for up- and downregulation. Accordingly, four auxiliary reactions (one demand and one slack reaction for upregulation and the same for downregulation) for each level were incorporated resulting in 120 ($10 \times 3 \times 4$) additional reactions. The three levels for each reaction were chosen, so that the flux range calculated by FVA for the wild-type space would be discretized into four equal intervals.

We then calculated the cRegMCSs for the extended network and identified 1840 designs with ≤ 6 modifications as shown in Figure 3a. In contrast, for the case without regulation (only cuts) with same parameters, the minimal number of deletions required in the calculated cMCSs was seven. There were 11 cRegMCSs of size 3, all of them involving upregulation of PDH. We further analyzed these 1840 designs and identified the frequency of regulation of each of the 10 reactions considered for regulation (Fig. 4). The upregulation of PDH at 75% level is involved in over 850 designs followed by GLYCL (glycine cleavage), RNDR2 and RNTR3c as reactions that were upregulated in a large number of designs. Upregulation of PDH as a potential strategy for ethanol production was seen in three of the four cases identified using OptReg on an earlier model of *E.coli* suggesting the importance of this strategy (Pharkya and Maranas, 2006). It is also surprising that several reactions appear as upregulation and downregulation targets, although only one is mostly dominating. In the case of the four most frequently regulated reactions, only upregulation of these reactions was used in the cRegMCSs.

To test the sensitivity of the cRegMCSs with respect to the chosen flux levels for regulation, we recalculated the cRegMCSs with the same 10 reactions for regulation but varied the predefined levels. Overall, the results appear to be robust against smaller changes in the regulation levels, in particular with respect to the distribution of the regulated reactions in the cRegMCSs (see Supplementary Text S2 and Fig. S2 in the Supplementary Material).

In the next step, we considered the four most relevant reactions for regulation identified in the first step (GLYCL, PDH, RNDR2 and RNTR3) with a finer discretization of 10 instead of three flux levels and identified 22 006 designs of which over 4000 designs had fewer than five modifications (Fig. 3b). The distribution of these designs for the four reactions at the 10 levels is shown in Figure 5. This histogram shows again the importance of upregulation of the PDH reaction in obtaining strain designs, as upregulating PDH between 55% and 82% of the FVA range is predicted to occur in over 3000 strain designs suggesting that regulation of the PDH reaction within a range might force ethanol production. Such a strategy can be rationalized by the fact that the upregulation of PDH results in production of Acetyl-CoA and subsequently ethanol (via aldehyde dehydrogenase) as opposed to lactate production via lactate dehydrogenase.

In most of the cases (29 of the 36 cases where at least one cRegMCS was found), the strain designs are predicted to have either two or three modifications (Fig. 5) highlighting again the potential of cRegMCSs to obtain design strategies with much fewer interventions. In particular, there were two cRegMCSs of size 2, both of which involved the upregulation of PDH and GLYCL. The identification of smaller cRegMCSs with finer level of regulation suggests the dependence of cRegMCSs on the level of regulation and indicates the importance of the second step with finer discretization of regulated levels. Finally, removal of redundant cRegMCSs and their subsequent partitioning into intervention classes (where the level of regulation is neglected) resulted in 1830 non-redundant cRegMCSs grouped in 1792 cRegMCS classes (for regulation of 10 reactions at 3 levels) and in 20 804 non-redundant cRegMCSs grouped in 7464 classes (for regulation of 4 reactions at 10 levels), respectively. These results show that, even when removing redundancies and neglecting the distinct levels of regulation, the number of cRegMCSs is an order of magnitude larger than the number of corresponding cMCSs highlighting further the value of incorporating regulation in generating a diversity of strategies. Although deregulation of PDH for synthesizing 1,4, butanediol has been shown to be a viable strategy for making acetyl-CoA anaerobically (Yim *et al.*, 2011), it is not clear yet whether it can likewise be employed for ethanol production. Regardless, the results from the cRegMCSs analysis suggest a number of hypotheses that could be tested experimentally in the future.

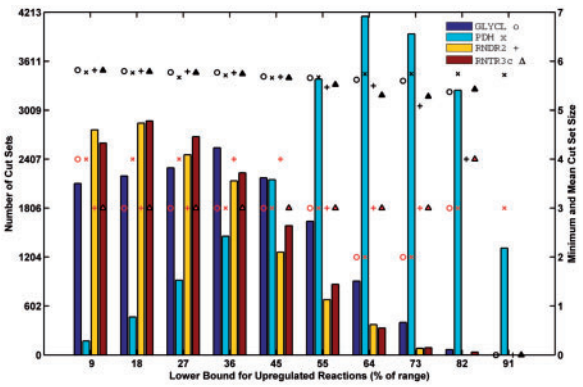


Fig. 5. Analysis of the 22 006 cut sets for the case with four regulated reactions at 10 levels for the core *E.coli* model. Number of the cut sets for each of the four regulated reactions at different levels is shown. The minimum (red symbols) and the mean (black symbols) (right side *y* axis) for each of the four regulated reactions GLYCL (o), PDH (x), RNDR2 (+), RNTR3c (Δ) is shown. Since the four reactions were not downregulated in the previous analysis, only cut sets with upregulation were considered

4 Discussion

In this study, we have generalized the MCS-based framework to *regulatory* MCSs allowing the consideration of reaction/gene knockouts and up- and downregulation of reaction fluxes. Using a small example and a realistic strain design problem in a genome-scale network, we have shown that cRegMCSs, in comparison to cMCSs involving only reaction deletions, can lead to (i) a lower number of required modifications and (ii) an increased number of possible intervention strategies. For example, in the case of genome-scale network of *E.coli*, using cRegMCSs we were able to identify several strategies with only two or three modifications and 333 strategies that have fewer than or equal to five modifications compared with the case where the cMCSs requires at least seven deletions. Thus, our cRegMCSs approach increases the choices available to metabolic engineers for experimental implementation and may thus accelerate the implementation of model-based strain designs for the bio-based production of fuels and chemicals. Another advantage of cRegMCSs is that they enable the fine tuning of desired behaviours in a narrower range which could not be attained by cMCSs alone. For example, in the toy network in Figure 2, the cMCSs involving exclusively knockouts can only redirect the complete flux to P (the yield $Y^{P/A}$ is always 1). An intermediate level of $0.6 \leq Y^{P/A} \leq 0.8$ cannot be attained but is possible with some suitable cRegMCSs.

Although there are previously developed methods that focus on the identification of reactions for regulation such as OptReg, OptForce and the more recently developed k-OptForce, all of them involve a bilevel optimization problem. Although the OptForce approach shares similarities with the cRegMCSs approach leading to partially overlapping results, the nature of the algorithm and of the solutions produced are significantly different. In contrast to OptForce requiring two optimization steps, cRegMCSs are calculated in a single-stage MILP algorithm and all intervention sets up to a given size can be directly enumerated. In addition, although OptForce indicates the direction of flux change relative to the reference flux distribution, it does not specify an optimal level for up/downregulation. Furthermore, cRegMCSs guarantee a high level of product synthesis irrespective of the optimality of growth in the mutant strain. On the other hand, OptForce allows direct consideration of phenotype (metabolic flux) data for the wild type; this narrower specification may help to obtain smaller (because more directed) intervention strategies.

Although gene or reaction knockouts can normally easily be implemented, it will be less straightforward to achieve, e.g. elevation of a metabolic flux as suggested by cRegMCSs analysis. At least for irreversible reactions, overexpression or/and control of the activity of associated metabolic enzymes will be the first approach and novel tools of molecular and synthetic biology enabling the facile regulation of metabolic enzymes [e.g. ribosome binding site variants for fine tuning expression (Farasat *et al.*, 2014; Salis *et al.*, 2009) and RNA devices for programming expression (Carothers *et al.*, 2011) and other tools (Stephanopoulos, 2012; Way *et al.*, 2014)] could provide key technologies in this context. Another (more indirect) possibility could be to enhance availability of metabolites consumed in a reaction to be upregulated. Nevertheless, the ability to precisely upregulate a flux as predicted in the cRegMCSs is still challenging. Therefore, one may select cRegMCSs that are not sensitive to the level of regulation. Additionally, cRegMCSs could also be used as a tool for preselecting promising intervention strategies, which are then analyzed in more detail in combination with kinetic modelling frameworks such as Ensemble modelling (Flowers *et al.*, 2013; Tran *et al.*, 2008) or ORACLE framework (Miskovic and Hatzimanikatis, 2011; Wang and Hatzimanikatis, 2006) to eventually identify the most robust strategies. In any case, cRegMCSs analysis reveals combinations of flux constraints that will lead to fulfilment of a given intervention goal and multiple experimental solutions may exist to implement these flux constraints.

A potential limitation of the new method is the largely enhanced solution space, which forces cRegMCSs approach to focus on a subset of reactions considered for regulation. In this work, we used an FVA-based method to identify most promising candidates, other methods such as EMILiO (Yang *et al.*, 2011) could be applied to identify these reactions as well. In the future, it would be desirable to develop algorithms that can calculate cRegMCSs without explicit specification of the flux levels of the regulated reactions (i.e. the relevant flux boundaries are determined directly by the algorithm), which would also help to avoid calculation of redundant cRegMCS. Ideally, the algorithm would also select the relevant reactions for regulation automatically *and* then determine the respective flux boundaries to obtain smallest intervention strategies. Another possible extension to cRegMCSs would be the incorporation of kinetics in a similar manner as the recently developed k-OptForce algorithm (Chowdhury *et al.*, 2014).

The combinatorial complexity of cRegMCSs calculation also guided us to further improve the cut-set enumeration algorithm. Specifically, the new approach for calculating smallest cMCSs in large-scale networks directly integrates constraints of the desired phenotypes thereby avoiding preliminary MCSs that will later be discarded. In the *E.coli* genome-scale network, we observed that the integrated algorithm is more efficient than the standard algorithm proposed by von Kamp and Klamt (2014). However, relative performance of the algorithms will also depend on structure and properties of the given intervention problem.

The sheer number of strain designs identified by cRegMCSs points to another important requirement, namely, methods for the prioritization of these strains for experimental implementation. Some of the results obtained in this study suggest that upregulation of certain reaction fluxes or enzyme activities such as PDH, which occur more often at multiple levels, might be a more robust strategy in comparison to other designs. In the future, it will be valuable to screen cRegMCSs to prioritize strategies that guarantee robust performance (Yang *et al.*, 2015). However, the detailed evaluation of the impact of regulation of these reaction fluxes or strain designs in general require additional systematic evaluation for measuring the ability

of strain designs to make the desired product in the presence of perturbations in the regulated reactions. The use of such metrics will be important to maximize the chances of success during the experimental implementation and validation of these strain designs.

Funding

This work was supported by Alexander von Humboldt foundation [to R.M.]; BioFuelNet [to R.M.]; the German Federal Ministry of Education and Research [projects CYANOSYS II (FKZ 0316183D) and CASCOO (FKZ: 031A180B) to S.K.] and the Federal State of Saxony-Anhalt [to S.K.] (Research Center 'Dynamic Systems: Biosystems Engineering').

Conflict of Interest: none declared.

References

Ballerstein,K. *et al.* (2012) Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, **28**, 381–387.

Brochado,A.R. *et al.* (2010) Improved vanillin production in baker's yeast through in silico design. *Microb. Cell Fact.*, **9**, 84.

Burgard,A.P. *et al.* (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, **84**, 647–657.

Carothers,J.M. *et al.* (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. *Science*, **334**, 1716–1719.

Chowdhury,A. *et al.* (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.*, **10**, e1003487.

Cotten,C. and Reed,J.L. (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnol. J.*, **8**, 595–604.

Farasat,I. *et al.* (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.*, **10**, 731.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Flowers,D. *et al.* (2013) SMET: systematic multiple enzyme targeting—a method to rationally design optimal strains for target chemical overproduction. *Biotechnol. J.*, **8**, 605–618.

Fong,S.S. *et al.* (2005) In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.*, **91**, 643–648.

Hädicke,O. and Klamt,S. (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J. Biotechnol.*, **147**, 88–101.

Hädicke,O. and Klamt,S. (2011) Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab. Eng.*, **13**, 204–213.

Keasling,J.D. (2012) Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.*, **14**, 189–195.

Kim,J. and Reed,J.L. (2010) OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.*, **4**, 53.

Klamt,S. (2006) Generalized concept of minimal cut sets in biochemical networks. *BioSystems*, **83**, 233–247.

Klamt,S. and Gilles,E.D. (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics*, **20**, 226–234.

Klamt,S. *et al.* (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.*, **1**, 2.

Mahadevan,R. and Schilling,C.H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.

McCloskey,D. *et al.* (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.*, **9**, 661.

Melzer,G. *et al.* (2009) Flux Design: In silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Syst. Biol.*, **3**, 120.

Miskovic,L. and Hatzimanikatis,V. (2011) Modeling of uncertainties in biochemical reactions. *Biotechnol. Bioeng.*, **108**, 413–423.

Monk,J. *et al.* (2014) Optimizing genome-scale network reconstructions. *Nat. Biotechnol.*, **32**, 447–452.

Muller,A.C. and Bockmayr,A. (2013) Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, **29**, 903–909.

Nookaew,I. *et al.* (2011) Genome-scale metabolic models of *Saccharomyces cerevisiae*. *Methods Mol. Biol.*, **759**, 445–463.

Park,J.M. *et al.* (2012) Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst. Biol.*, **6**, 106.

Patil,K.R. *et al.* (2005) Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, **6**, 308.

Pharkya,P. and Maranas,C.D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.*, **8**, 1–13.

Pharkya,P. *et al.* (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.*, **14**, 2367–2376.

Portnoy,V.A. *et al.* (2010) Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655. *Appl. Environ. Microbiol.*, **76**, 6529–6540.

Ranganathan,S. *et al.* (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.*, **6**, e1000744.

Ranganathan,S. *et al.* (2012) An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. *Metab. Eng.*, **14**, 687–704.

Salis,H.M. *et al.* (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.

Schuster,S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18**, 326–332.

Stephanopoulos,G. (2012) Synthetic biology and metabolic engineering. *ACS Synth. Biol.*, **1**, 514–525.

Tran,L.M. *et al.* (2008) Ensemble modeling of metabolic networks *Biophys. J.*, **95**, 5606–5617.

Trinh,C.T. (2012) Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.*, **95**, 1083–1094.

Trinh,C.T. *et al.* (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.*, **74**, 3634–3643.

Trinh,C.T. *et al.* (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.*, **81**, 813–826.

Trinh,C.T. *et al.* (2011) Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Appl. Environ. Microbiol.*, **77**, 4894–4904.

von Kamp,A. and Klamt,S. (2014) Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput. Biol.*, **10**, e1003378.

Wang,L. and Hatzimanikatis,V. (2006) Metabolic engineering under uncertainty. I: framework development. *Metab. Eng.*, **8**, 133–141.

Way,J.C. *et al.* (2014) Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell*, **157**, 151–161.

Xu,P. *et al.* (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.*, **13**, 578–587.

Yang,L. *et al.* (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metab. Eng.*, **13**, 272.

Yang,L. *et al.* (2015) Characterizing metabolic pathway diversification in the context of perturbation size. *Metab. Eng.*, **28**, 114–122.

Yim,H. *et al.* (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.*, **7**, 445–452.

Zomorrodi,A.R. *et al.* (2012) Mathematical optimization applications in metabolic networks. *Metab. Eng.*, **14**, 672–686.