Co-evolution of strain design methods based on flux balance and elementary mode analysis

CrossMark

Daniel Machado [a,*], Markus J. Herrgård [b]

[a] Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal
[b] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark

ARTICLE INFO

ABSTRACT

More than a decade ago, the first genome-scale metabolic models for two of the most relevant microbes for biotechnology applications, *Escherichia coli* and *Saccaromyces cerevisiae*, were published. Shortly after followed the publication of OptKnock, the first strain design method using bilevel optimization to couple cellular growth with the production of a target product. This initiated the development of a family of strain design methods based on the concept of flux balance analysis. Another family of strain design methods, based on the concept of elementary mode analysis, has also been growing. Although the computation of elementary modes is hindered by computational complexity, recent breakthroughs have allowed applying elementary mode analysis at the genome scale. Here we review and compare strain design methods and look back at the last 10 years of *in silico* strain design with constraint-based models. We highlight some features of the different approaches and discuss the utilization of these methods in successful *in vivo* metabolic engineering applications.

Contents

1. Introduction

Computational modeling has emerged as a fundamental tool for unraveling the complexity of biological processes. There are currently many different mathematical formalisms that can be used to model biochemical reaction networks (Machado et al., 2011). Among these formalisms, the constraint-based modeling approach has become widely adopted for large-scale modeling of metabolism (Bordbar et al., 2014). Constraint-based models have been used for a multitude of applications from guiding biological discovery to the improvement of industrial bioprocesses (McCloskey et al., 2013).

Constraint-based models can be used to simulate the cellular phenotype at steady-state using different methods. The most common approach, flux balance analysis (FBA), is a linear programming formulation that relies on the maximization of a cellular objective, such as growth or ATP generation, to determine the steady-state flux distribution through a metabolic network (Orth et al., 2010). Other methods, typically used for simulation of mutant strains, are based on principles of minimization of metabolic and regulatory adjustments (MOMA, ROOM) (Segrè et al., 2002; Shlomi et al., 2005). These kinds of methods are usually classified as biased, since they rely on the assumption of some evolutionary optimization principle to determine a biologically
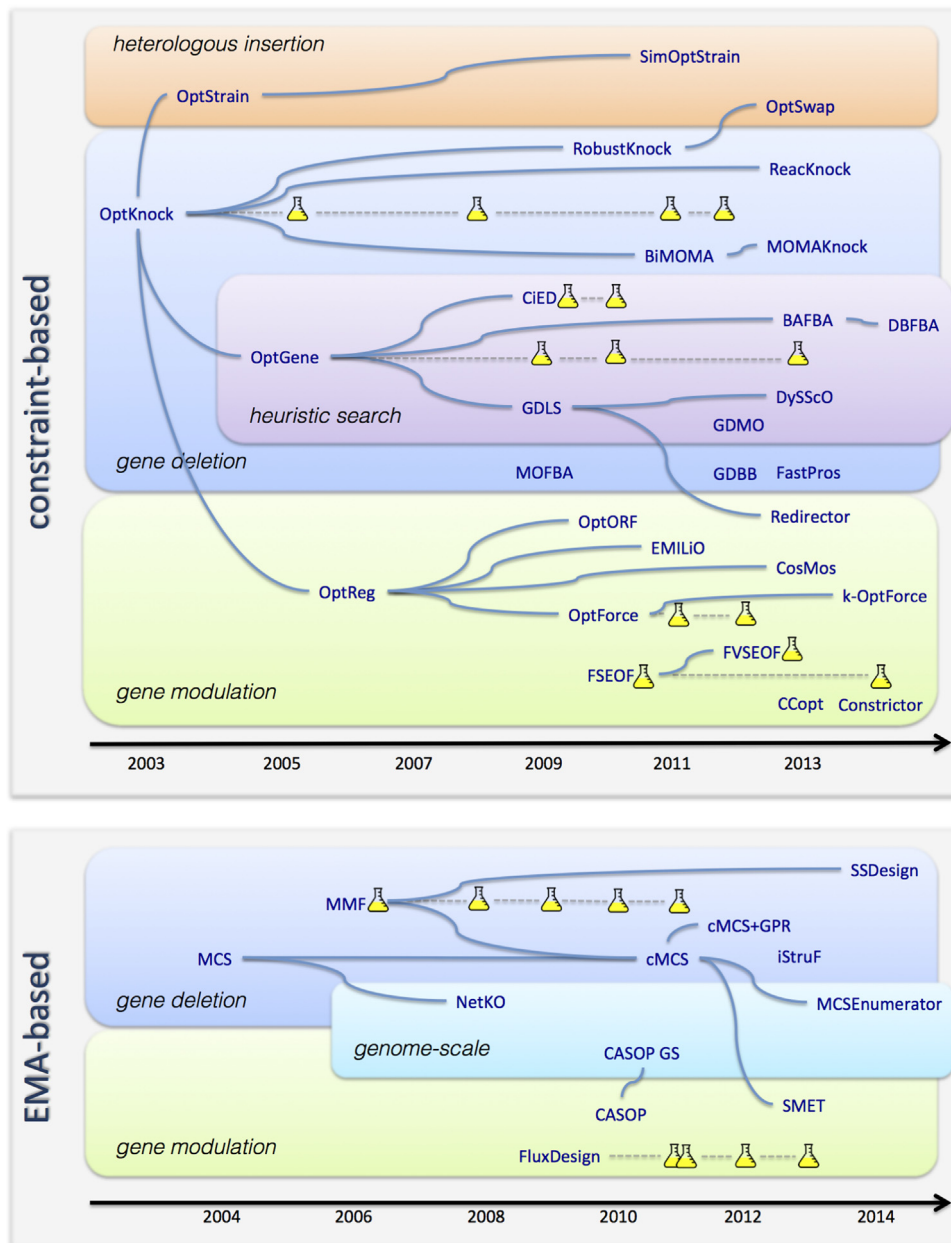
Fig. 1. Chronological perspective of the evolution of strain design methods using constraint-based analysis and elementary mode analysis (EMA). Connections represent common features between methods, not necessarily a direct extension of the previous method. The shake flask symbol represents experimental applications of the respective methods.

meaningful and physicochemically valid steady-state flux distribution.

There are also unbiased approaches to analyze feasible flux distributions in large-scale metabolic networks, including Monte Carlo sampling and metabolic pathway analysis (Lewis et al., 2012). Elementary mode analysis (EMA) is one of the most popular approaches for metabolic pathway analysis. It provides an unbiased description of the metabolic solution space in terms of minimal sets of reactions that operate in steady-state (Schuster and Hilgetag, 1994). These so-called elementary (flux) modes (EMs) are the basis for several methods to analyze the properties of metabolic networks, including robustness and fragility, as well as to calculate the theoretical yields of all metabolic routes (Trinh et al., 2009).

Both biased and unbiased methods have been used for strain design since the first genome-scale metabolic models of two industrially relevant microbes, *Escherichia coli* and *Saccharomyces*

cerevisiae, were published in the early 2000s (Edwards and Palsson, 2000; Förster et al., 2003). From a metabolic engineering perspective, such models can be used for computer-aided design of optimal genetic and culture condition manipulation strategies to improve the production of industrially relevant compounds. However, given the size of metabolic networks, the exhaustive analysis of multiple simultaneous genetic manipulations becomes computationally infeasible. In order to address this challenge, a variety of methodological solutions have been proposed (Fig. 1).

2. Constraint-based methods

The first systematic optimization-based method for strain design was the OptKnock approach introduced by Burgard et al. (2003). OptKnock is a bilevel optimization approach that determines reaction deletion strategies to couple the production of a
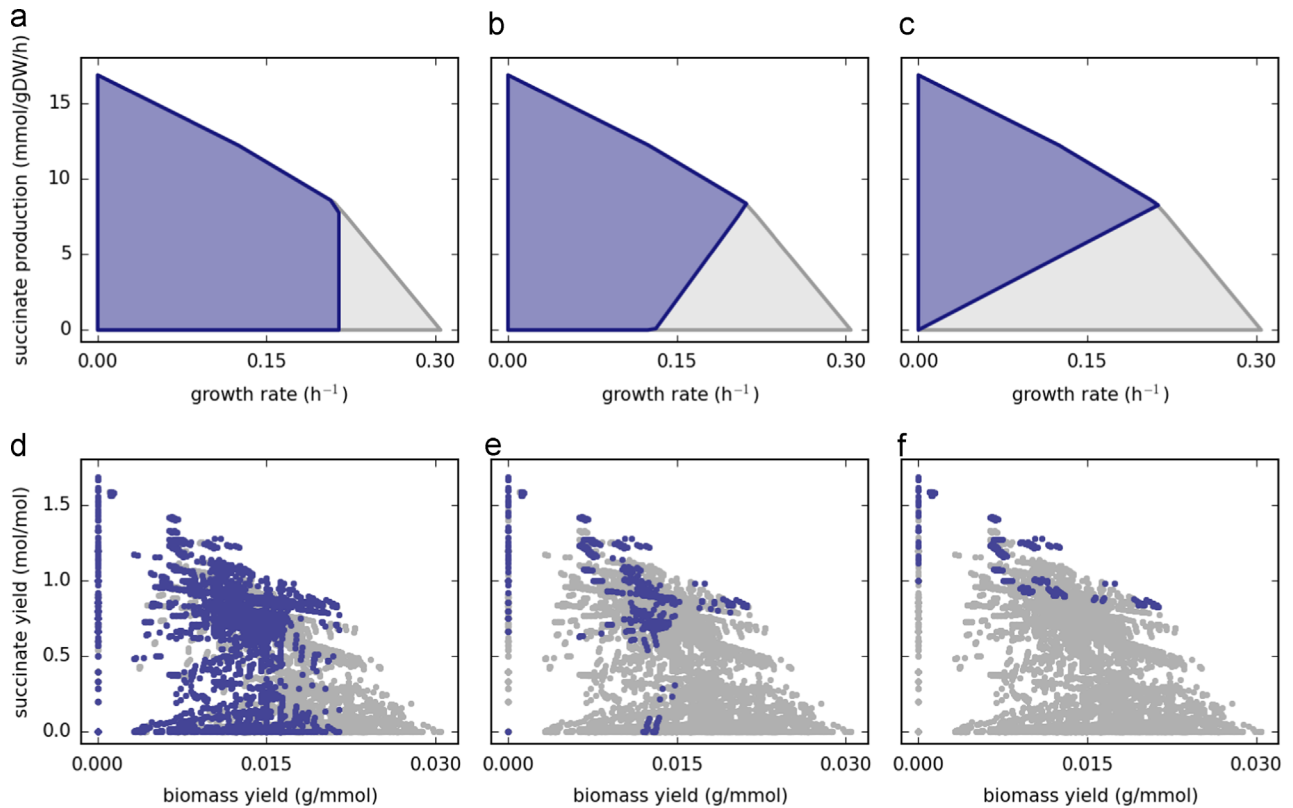
Fig. 2. Production envelopes for anaerobic succinate production from glucose based on the *E. coli* core model (Orth et al., 2009) for a maximum glucose uptake rate of 10 mmol/gDW/h (a–c) and the respective EM yield distribution for each solution space (d–f): (a) wild-type strain (light gray) vs. triple-deletion mutant (ACKr, ATPS4r, FUM) resulting in a design without growth-coupling (purple); (b) wild-type strain vs. triple-deletion mutant (ACALD, PYK, ME2) resulting in a partially growth-coupled design; (c) wild-type strain vs. double-deletion mutant (ACALD, LDH_D) resulting in a fully growth-coupled design; (d–f) EM distribution of the wild-type (light gray) and the respective mutant strains in (a–c) (purple). The EM yield locations correspond to vertices in the flux solution space at the maximal glucose uptake rate. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

desired compound with cellular growth (Burgard et al., 2003). In OptKnock, the outer optimization layer maximizes the product yield, while the inner layer optimizes for the cellular growth. Using duality theory, this bilevel optimization problem can be reformulated as a single mixed integer linear programming (MILP) problem. Growth-coupled designs represent mutant strains that are forced to carry flux through the target pathway as a requirement to achieve optimal growth rates. When such strains are subject to adaptive evolution they gradually evolve towards the optimal phenotype (Fong et al., 2005). Growth-coupled designs can be visualized by projecting the flux solution space onto the growth and production axes, forming the so-called production envelope (Fig. 2). Two different kinds of growth-coupled designs can be distinguished, which will herein be referred to as *partial* and *full* growth-coupling. In the first case, the strain is only forced to produce the target product at optimal growth rates (Fig. 2b), whereas in the latter case the strain is unable to grow without product formation (Fig. 2c). One limitation of OptKnock is the degeneracy in the solution of the inner problem, which can sometimes result in overly optimistic predictions and lead to strain designs that are not effectively growth-coupled (Fig. 2a).

The introduction of OptKnock has laid the foundation for a diversity of bilevel methods for rational strain design that have been improving over the years. One of such extensions, Robust-Knock, uses a max–min strategy to account for the degeneracy in FBA solutions, leading to strain designs that are effectively growth-coupled (Tepper and Shlomi, 2010). An alternative, simpler to implement, approach is to apply objective tilting in the OptKnock formulation (Feist et al., 2010). ReacKnock is a reformulation of OptKnock that differs in the transformation applied to convert the

bilevel problem to a single level MILP. A comparison of the two methods shows that higher product yields and faster computations can be obtained with ReacKnock (Xu et al., 2013b). The BiMOMA and MOMAKnock approaches replace the simulation layer in OptKnock with MOMA (Segrè et al., 2002), making it possible to find designs that do not need to couple cellular growth to metabolite production (Kim et al., 2011; Ren et al., 2013).

Although MILP formulations can be used to find a globally optimal solution, their computational cost can, in the worst case, increase exponentially with the number of reaction deletions. A common approach to deal with complex optimization problems is to use heuristic optimization strategies that do not guarantee finding globally optimal solutions, but often find sufficiently good solutions with a reasonable computational cost. Patil and co-workers implemented OptGene, an optimization method based on genetic algorithms, that allows accounting for larger numbers of deletions without an increase in computational cost (Patil et al., 2005). This formulation was later extended to also support simulated annealing as an optimization method (Rocha et al., 2008). Besides increased efficiency for finding reasonably good strain design strategies, these approaches are also more flexible, allowing the implementation of non-linear objective functions. Heuristic optimization methods also allow the utilization of different simulation strategies for the inner problem, such as MOMA and ROOM (Segrè et al., 2002; Shlomi et al., 2005) without the need to significantly change the overall optimization approach.

Other methods using nature-inspired metaheuristics to solve strain design optimization problems include CiED, based on evolutionary algorithms (Fowler et al., 2009), BAFBA, implementing a bees algorithm (Choon et al., 2012), and DBFBA, combining

the latter with differential evolution (Choon et al., 2014). Since heuristic methods do not guarantee a globally optimal solution, it is advantageous to experiment with multiple optimization strategies. The hybrid method GDLS combines an heuristic global search approach with a local search method for more efficiently scanning the solution space of genetic designs (Lun et al., 2009).

Gene deletions represent only a subset of the genetic manipulations that can be performed *in vivo* in order to redirect metabolic flux towards a target metabolite. In many applications it is necessary to up-regulate pathways that would become production bottlenecks, or to down-regulate a pathway where a full deletion would otherwise be lethal. OptReg is an extension of OptKnock that accounts for up/down-regulation of gene expression as well as gene deletions (Pharkya and Maranas, 2006). In order to account for the effects of transcriptional regulation, Kim and Reed introduced OptORF, which includes transcriptional regulatory constraints as part of its formulation (Kim and Reed, 2010). With this method it is also possible to simulate manipulations directly at the gene level rather than at the reaction level. Other approaches that account for gene modulation targets include EMILiO (Yang et al., 2011) and CosMos (Cotten and Reed, 2013). These two methods simulate the effect of gene up-regulation by forcing an increase in the flux of the respective reaction, which may not always be realistic as increasing gene or protein expression *in vivo* is not guaranteed to increase the flux through the corresponding reaction. The Redirector method takes a different approach by modeling up-regulations as negative weights in the cellular objective, reflecting the trade-off between cellular growth and the cost of amplifying the desired pathways (Rockwell et al., 2013).

All the methods mentioned so far rely on the simulation of a particular flux distribution for the proposed mutant strain. However, these simulations may not realistically reflect the metabolic response to the proposed genetic perturbations. This discrepancy between *in silico* and *in vivo* response can result in misleading design strategies. A more reliable alternative is to directly compare the metabolic fluxes of the wild-type and the desired mutant strain, and observe which fluxes must necessarily change in order to obtain the desired phenotype. This kind of approach has been implemented in FSEOF (later extended to FVSEOF) using an iterative search method that scans the production envelope to detect flux changes (Choi et al., 2010; Park et al., 2012). The OptForce method allows finding optimal intervention sets of minimal size by first calculating flux variability ranges in a manner similar to FSEOF and then using bilevel optimization to find the minimal set (Ranganathan et al., 2010).

The majority of methods described above focus on predicting which manipulations of native genes would result in increased production of a desired metabolite. However, in actual cell factory development project the focus is often on expression of heterologous genes to enable introducing new biosynthetic capabilities to the host organism. OptStrain, one of the first extensions of OptKnock, was developed to account for heterologous gene expression by inserting reactions that are retrieved from an universal metabolic reaction database (Pharkya et al., 2004). SimOptStrain improves OptStrain by performing additions and deletions simultaneously, resulting in novel design solutions that would not be found otherwise (Kim et al., 2011). The OptSwap method presents the option of replacing native oxidoreductase enzymes with heterologous versions with different co-factor specificities (NADH or NADPH), resulting in growth-coupled designs with significantly higher yields (King and Feist, 2013).

There are few other constraint-based approaches worth mentioning, which are not so commonly referenced and do not fall into the categories described above. These include GDBB, implementing a truncated branch and bound technique (Egen and Lun, 2012);

CCopt, based on probabilistic flux bounds (Yousofshahi et al., 2013); FastPros, with a new iterative design based on shadow-price information (Ohno et al., 2013); Constrictor, with combinatorial search of down-regulation targets (Erickson et al., 2014); and multi-objective search methods based on Pareto front analysis such as MOFBA (Oh et al., 2009) and GDMO (Costanza et al., 2012). An in-depth analysis of each of these methods is out of the scope of this review.

3. EMA-based methods

Methods based on EMA have been evolving in parallel with the constraint-based methods described previously. In general, elementary mode (EM) enumeration is computationally expensive due to the combinatorial explosion in the number of EMs as a function of network size. On the other hand, since EMs fully describe the steady-state solution space, once the EM set is enumerated, the impact of structural changes in the metabolic network regarding specific phenotypes such as a desirable metabolite production can be assessed very efficiently. For instance, the solution space of a deletion mutant is precisely described by the subset of EMs that do not contain the respective reaction(s).

In 2004, Klamt and Gilles introduced the concept of minimal cut sets (MCS). These are the minimal sets of reaction deletions required to fully block an undesired function in a metabolic network (Klamt and Gilles, 2004). MCSs and EMs are dual properties of metabolic networks, and one can be computed from the other. From a metabolic engineering perspective, this approach can be used to block the production of undesired by-products including common overflow metabolites such as acetate, ethanol, or glycerol, in order to route flux towards a more desirable product.

As an alternative to MCSs, Trinh and co-workers developed the method of minimal metabolic functionality (MMF), a greedy search approach for determining reaction deletions that reduce a metabolic network to its most efficient pathways (Trinh et al., 2006). One advantage of MMF is that a prioritized list of interventions is given, sorted by their impact on network flexibility (i.e. the number of pathways that can carry a given flux).

One limitation of the MCS and MMF approaches is that the elimination of undesired functionality from the network can result in the elimination of desired functionality as well. To address this problem, Hädicke and Klamt introduced constrained minimal cut sets (cMCS) (Hädicke and Klamt, 2011). This is a generalization of MCS to account for a set of desired EMs that must be preserved when the undesired EMs are eliminated. The authors show that other strain-design methods such as MMF, OptKnock and Robust-Knock can be implemented as particular cases of cMCS. The SSDesign method is a reformulation of MMF that also includes additional constraints for preserving a minimum set of EMs with desired properties during the elimination steps (Toya et al., 2014).

Two other methods to find optimal deletions include an extension of cMCS with gene-protein-reaction (GPR) constraints, which allows a direct analysis of gene (rather than reaction) deletions (Jungreuthmayer and Zanghellini, 2012), and the iStruF method that searches for reaction deletions based on a flux estimation method named *structural fluxes* (Soons et al., 2013). An application of MCS for gene essentiality prediction at the genome-scale (NetKO) was performed by decomposing the network into so-called *pathway fragments* (Imielinski and Belta, 2008).

EMA-based method have also been used to determine gene up/down-regulation targets that redirect metabolic flux towards the target product. One of the first EMA-based methods to implement this kind of interventions was FluxDesign (Melzer et al., 2009).

This method computes the correlations between fluxes through each reaction and the target reaction in the set of EMs. Strongly positive or negative correlations suggest, respectively, overexpression or deletion targets. This method can be considered to be analogous to the FSEOF method introduced above.

CASOP, from the same authors of cMCS, is another method that accounts for gene modulation to maximize, not only yield, but also productivity (Hädicke and Klamt, 2010). This method searches for manipulation strategies with an optimal trade-off between product yield and network flexibility. Here, the network flexibility is used as a measure of network capacity (i.e. the total flux carried by a pathway), which is interpreted as an indicator for the specific production rate.

As mentioned earlier, one of the main limitations of EMA is that the number of EMs grows exponentially with the size of the metabolic network. Since this makes full EM enumeration impractical at the genome scale, a more feasible alternative might be to compute a subset of the full EM set by random sampling (Kaleta et al., 2009; Machado et al., 2012). This kind of approach was implemented in CASOP-GS, which extends CASOP to the genome scale (Bohl et al., 2010).

The issue of computational feasibility was also addressed by the MILP approach introduced by de Figueredo et al. that allows computing the *K-shortest* EMs at the genome-scale. Extending this work, von Kamp and Klamt recently introduced MCSEnumerator, a method to compute the smallest MCSs in genome-scale models by exploiting the duality between MCSs and EMs (von Kamp and Klamt, 2014). With this method the authors were able to enumerate an unprecedented number of intervention strategies, including an exhaustive enumeration of up to 7 simultaneous deletions using a genome-scale model of *E. coli*. The design strategies found by MCSEnumerator included strain design strategies that were not previously found with OptKnock (Feist et al., 2010).

4. Applications

Strain design methods are usually published with practical case studies proposing specific strain design strategies. Most of the case studies have used *E. coli* as the model organism as the genome-scale metabolic models for this organism have continued to be the most predictive of all models. The most popular target products considered in the case studies were succinate (20 cases), ethanol (12), lactate (9), acetate (5), fumarate (5) and glycerol (5). Unfortunately, one of the main drawbacks of most methods is the lack of experimental validation of the proposed strain designs, as most of the designs are only evaluated *in silico* by comparison with previously explored experimental design strategies that had proven to be successful. Only four of the methods reviewed herein (CiED, MMF, FSEOF, FVSEOF) have included experimental validation of the design suggestions in their original publication. In some cases (OptORF, Redirector, CASOP, SSDesign) a comparison of the selected manipulation targets with previously published strain designs is presented. However, this does not allow us to make conclusions about quantitative predictions, since the *in silico* and previously experimentally validated designs are only partially the same. Despite this apparent gap between computational work and experimental validation, there are examples of utilization of these methods in successful applications (Table S1).

OptKnock was used to find growth-coupled designs for lactate production in *E. coli* (Fong et al., 2005). Three proposed knock-out designs were implemented *in vivo* and subjected to adaptive evolution with selection for increased growth rate. It was observed that all the strains evolved in the direction of the predicted phenotype of increased L-lactate production. OptKnock was also used for identifying strain designs that increase the respiratory

rate of *G. sulfurreducens* (Izallalen et al., 2008), the production of 1,4-butanediol in *E. coli* (Yim et al., 2011), and the production of 2,3-butanediol in yeast (Ng et al., 2012).

OptGene has been used to optimize the production of multiple products in yeast. Deletion targets for non-growth-coupled production of sesquiterpene (Asadollahi et al., 2009) and vanillin (Brochado et al., 2010) were found using MOMA for simulation of the mutant phenotypes. On the other hand, using FBA for simulation, a growth-coupled design strategy for succinate production was obtained with OptGene. The growth-coupled mutant was further optimized by adaptive laboratory evolution, followed by a second round of metabolic engineering, resulting in a final 30-fold improvement in succinate titer (Otero et al., 2013).

Another evolutionary optimization-based approach, CiED, was used in combination with MOMA to find gene deletion strategies to increase the intracellular pools of malonyl-CoA (Fowler et al., 2009) and NADPH (Chemler et al., 2010) in *E. coli*. These are two important precursors for the production recombinant natural products. The resulting strains showed not only an increase in intracellular concentrations of these precursors, but also increased production titers of multiple flavonoids (naringenin, eriodictyol, leucocyanidin, and (+)-catechin).

OptForce was also used to address the increase of intracellular malonyl-CoA concentration for production of heterologous compounds in *E. coli*. A combination of gene overexpression and deletions resulted in a 4-fold increase in malonyl-CoA concentration relative to the wild-type strain (Xu et al., 2011). This method was also used for fatty acid production in *E. coli*, in a study where 39% of the maximum theoretical yield was reached (Ranganathan et al., 2012).

FSEOF and FVSEOF were respectively used to find gene up-regulation targets for lycopene (Choi et al., 2010) and putrescine (Park et al., 2012) production in *E. coli*. Recently, FSEOF was used to find gene up-regulation targets for actinorhodin production in *S. coelicolor*, resulting in a 52-fold increase in product titers (Kim et al., 2014).

Of the EMA-based approaches, only MMF and FluxDesign seem to have been used in practical applications so far. MMF has been used in multiple applications in *E. coli*. In its original publication, this method was used to reach increased yields of biomass on glucose (Trinh et al., 2006). Later, MMF was used to improve the production of ethanol from a mixture of glucose and xylose (Trinh et al., 2008) and from glycerol (Trinh and Srienc, 2009). In both cases the experimentally obtained yields (0.49 and 0.45 g/g) were remarkably close to the theoretical predictions (0.36–0.51 and 0.50 g/g, respectively). Note that these designs require a high number of deletions, compared to methods such as OptKnock, with a total of 8 and 9 gene deletions, respectively. MMF was also used to obtain designs for increased production of isobutanol (Trinh et al., 2011) and diapolycopendial (Unrean et al., 2010). However, in these cases the match between the predicted and experimental yields was not was not as good as in the case of ethanol production.

FluxDesign was used to find gene up-regulation and deletion targets for the production of lysine in *C. glutamicum* (Neuner and Heinzle, 2011; Becker et al., 2011). In one case a high yield of 0.55 (g/g glucose) was reached (Becker et al., 2011). This method was also used to improve isobutanol production in *B. subtilis* (Li et al., 2012), and poly-hydroxyalkanoates production in *P. putida* (Poblete-Castro et al., 2013). In the latter study, the experimental results confirmed the design suggested by FluxDesign and disproved a different design that had been proposed with OptKnock.

Finally, it is important to highlight the numerous model-guided applications that do not use targeted optimization methods like the ones described so far. We performed a literature survey,

expanding a previous survey by Xu et al. (2013a) (Table S1). It can be observed that the production titers and yields of many products have been improved with model-guided simulations without the use of methods for optimized search of genetic modifications. Lycopene titers in *E. coli* were increased up to 8.5-fold using a model-guided strategy that included combinatorial simulations of single, double and triple deletion mutants (Alper et al., 2005a,b). Other applications of combinatorial simulation of gene deletions in *E. coli* have resulted in the development of high producing strains, including a 7-fold increase in succinate titer (Lee et al., 2005), a 7.4-fold increase in molar yield of 3HP production from glycerol (Tokuyama et al., 2014), and a high production yield of L-valine (0.378 g/g) from glucose (Park et al., 2007). Instead of simulating all gene deletions, one can also combine biological intuition on potential manipulation candidates with computational analysis to observe the flux response of the target reaction with respect to the given candidates (flux response analysis). This method was used to select gene manipulations in *E. coli* that achieved a high yield of L-threonine (0.393 g/g) from glucose (Lee et al., 2007). This method has also be used to determine optimal feeding strategies in fed-batch cultivations (Park et al., 2011).

5. Discussion

Strain design approaches based on flux balance and elementary mode analysis have been co-evolving independently. In both cases, the genetic designs have been extended to include not only gene deletions but also gene up/down-regulation and addition of heterologous genes.

The number of constraint-based methods that have been suggested for strain design is significantly higher than the number of EMA-based methods. This is a likely consequence of the computational limitations of the latter, requiring access to high performance computing resources. Given the recent advances that have raised the power of EMA to the genome scale (de Figueiredo et al., 2009; von Kamp and Klamt, 2014) it is possible that EMA-based approaches will be increasingly explored. Constraint-based approaches have limitations in their predictive power due to the potential biases introduced by using phenotype simulation to predict metabolic flux distributions. Recent methods have begun to circumvent this problem by comparing flux ranges instead of simulating particular solutions for each genetic modification (Ranganathan et al., 2010; Choi et al., 2010).

With the decreasing bias of constraint-based methods, the increasing ability of EMA-based methods to compute with larger networks, and the support for more types of genetic modifications for both methods, the gap between these approaches is diminishing. However, there are still fundamental differences between the methods that may lead to different results. Most constraint-based methods search for genetic modifications that shape the solution space in a way that favors the coupling between growth and production of the target compound at the point of optimality (i.e. partial growth-coupling) (Fig. 2b). Reaching a suitable shape may require multiple, possibly unintuitive, simultaneous modifications. During *in silico* simulations, the effect of single modifications on product flux is negligible until a suitable combination is found, making it difficult to implement an iterative strain engineering process where only a few genetic modifications at a time are introduced.

On the other hand, EMA-based methods try to eliminate (or decrease the utilization of) EMs that do not produce the target compound. A fully growth-coupled design is obtained if such EMs are completely eliminated (Fig. 2f). However, each modification alters the EM distribution in the solution space even if the production envelope is unaltered. Since the impact of a modification is measured by the number of affected EMs, there are no "silent" modifications in EMA-based design. This allows utilizing an iterative strain engineering process where only a few modifications at a time are introduced. While this is an advantage of EM-based approaches, the resulting designs tend to contain a larger number of modifications compared to constraint-based designs.

The lack of experimental validation hampers a critical comparison of the different methods. The cases of successful applications that have been published so far seem to be concentrated around a few well established methods. The successful results show that strain design methods are indeed a useful tool for guiding metabolic engineering applications. However, it is not straightforward to make a quantitative assessment of the expected results, since the predictions are based on yields and in many cases only the final product titers are reported.

Growth-coupled designs obtained with FBA seem to result in a good match between *in silico* and *in vivo* results after adaptation of the mutant strains (Fong et al., 2005; Otero et al., 2013). This is consistent with the observation that FBA is suitable to predict the phenotype of evolved strains (Lewis et al., 2010). For MOMA-based predictions there can be some discrepancy between predicted and experimental yields (Asadollahi et al., 2009), which is probably true for non-growth-coupled strain designs in general. In these cases, as well as for designs based on gene modulation, the proposed interventions should be regarded as qualitative, rather than quantitative, predictions. Employing these types of predictions *in vivo* will require testing multiple alternative predictions in a combinatorial fashion to find the optimal design.

So far, EMA-based designs have been computed based on core metabolic models. Although these models have a limited scope compared to genome-scale models, this may not reduce their applicability in metabolic engineering. A possible reason is that, in most strain designs, the intervention targets are enzymes in central metabolic pathways. With designs obtained with core models, the system-wide impact of the proposed genetic modifications can be simulated using genome-scale models. For instance, the growth-coupled design shown in Fig. 2c does not result in growth coupling of succinate production in the respective genome-scale model (Feist et al., 2007). This is due to an alternative pathway for acetaldehyde production from acetyl-CoA using 2-amino-3-oxobutanoate and allo-threonine as intermediates. This alternative pathway can be eliminated with an additional knockout of one of the steps. Hence, the designs obtained with core models are often subsets of more complex designs that can be obtained with genome-scale models. Succinate overproducing mutants have been successfully obtained by deletion of the central acetaldehyde dehydrogenase without eliminating the alternative acetaldehyde production pathway (Sánchez et al., 2006; Jantama et al., 2008). Hence, it is unlikely that most of the carbon flux would be rerouted through this pathway in the mutant strain, but if this strain is subjected to adaptive evolution under growth selection, the alternative pathway may get activated thus reducing succinate production. Experimental results show that EMA-based predictions using core models are highly accurate for central carbon metabolism (Trinh et al., 2008; Trinh and Srienc, 2009). However, the predictive power of EMA-based designs seems to decrease when heterologous pathways are considered, indicating limitations in these heterologous pathways themselves as opposed to limitations in central metabolic precursor generation (Unrean et al., 2010; Trinh et al., 2011).

One of the most important problems in bioprocess engineering is the trade-off between yield, titer, and productivity, which are often conflicting objectives. Since constraint-based models can only make yield-based predictions, other parameters can only be targeted indirectly (Patil et al., 2005; Hädicke and Klamt, 2010). The recent DySSco framework addresses this problem by integrating dynamic

flux balance analysis with GDLS (Zhuang et al., 2013). Another solution for this problem is the integration of kinetic and constraint-based modeling, in order to calculate absolute flux rates. This kind of integration was recently implemented in k-OptForce (Chowdhury et al., 2014). Similarly, the recent method SMET combines EM analysis with ensemble modeling to identify rate limiting steps (Flowers et al., 2013).

With the algorithmic improvements and the increasing availability of computational power, the evolution of strain design methods will certainly become less driven by computational efficiency, and more importantly so by the biological plausibility of their results. Recent impressive developments in synthetic biology methods for genome engineering such as the *CRISPR/Cas9* system (Sander and Joung, 2014) will allow implementing more complex designs *in vivo*. Systematic evaluations of *in silico* strain design predictions from different methods will hopefully become the norm rather than the exception in the future.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.meteno.2015.04.001.

References

Alper, H., Jin, Y.-S., Moxley, J., Stephanopoulos, G., 2005a. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in escherichia coli. Metabol. Eng. 7 (3), 155–164.

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005b. Construction of lycopene-overproducing e. coli strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. 23 (5), 612–616.

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. Metabol. Eng. 11 (6), 328–334.

Becker, J., Zelder, O., Häfner, S., Schröder, H., Wittmann, C., 2011. From zero to hero—design-based systems metabolic engineering of *Corynebacterium glutamicum* for ʟ-lysine production. Metabol. Eng. 13 (2), 159–168.

Bohl, K., de Figueiredo, L.F., Hädicke, O., Klamt, S., Kost, C., Schuster, S., Kaleta, C., 2010. Casop gs: computing intervention strategies targeted at production improvement in genome-scale metabolic networks. In: Lecture Notes in Informatics, pp. 71–80.

Bordbar, A., Monk, J.M., King, Z.A., Palsson, B.O., 2014. Constraint-based models predict metabolic and associated cellular functions. Nat. Rev. Genet. 15 (2), 107–120.

Brochado, A.R., Matos, C., Møller, B.L., Hansen, J., Mortensen, U.H., Patil, K.R., 2010. Improved vanillin production in Baker's yeast through in silico design. Microb. Cell Fact. 9 (1), 84.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84 (6), 647–657.

Chemler, J.A., Fowler, Z.L., McHugh, K.P., Koffas, M.A., 2010. Improving nadph availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. Metabol. Eng. 12 (2), 96–104.

Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M., 2010. In silico identification of gene amplification targets for improvement of lycopene production. Appl. Environ. Microbiol. 76 (10), 3097–3105.

Choon, Y.W., Mohamad, M.S., Deris, S., Chong, C.K., Chai, L.E., Ibrahim, Z., Omatu, S., 2012. Identifying gene knockout strategies using a hybrid of bees algorithm and flux balance analysis for in silico optimization of microbial strains, Distributed Computing and Artificial Intelligence. Springer, Berlin, Heidelberg, pp. 371–378.

Choon, Y.W., Mohamad, M.S., Deris, S., Illias, R.M., Chong, C.K., Chai, L.E., Omatu, S., Corchado, J.M., 2014. Differential bees flux balance analysis with optknock for in silico microbial strains optimization. PLoS One 9 (7), e102744.

Chowdhury, A., Zomorrodi, A.R., Maranas, C.D., 2014. k-optforce: integrating kinetics with flux balance analysis for strain design. PLoS Comput. Biol. 10 (2), e1003487.

Costanza, J., Carapezza, G., Angione, C., Lió, P., Nicosia, G., 2012. Robust design of microbial strains. Bioinformatics 28 (23), 3097–3104.

Cotten, C., Reed, J., 2013. Constraint-based strain design using continuous modifications (cosmos) of flux bounds finds new strategies for metabolic engineering. Biotechnol. J. 8 (5), 595–604.

de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J., 2009. Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25 (23), 3158–3165.

Edwards, J., Palsson, B., 2000. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. 97 (10), 5528–5533.

Egen, D., Lun, D.S., 2012. Truncated branch and bound achieves efficient constraint-based genetic design. Bioinformatics 28 (12), 1619–1623.

Erickson, K.E., Gill, R.T., Chatterjee, A., 2014. Constrictor: constraint modification provides insight into design of biochemical networks. PloS One 9 (11), e113820.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. Mol. Syst. Biol. 3 (1).

Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.Ø., 2010. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metabol. Eng. 12 (3), 173–186.

Flowers, D., Thompson, R.A., Birdwell, D., Wang, T., Trinh, C.T., 2013. Smet: systematic multiple enzyme targeting—a method to rationally design optimal strains for target chemical overproduction. Biotechnol. J. 8 (5), 605–618.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., Palsson, B.O., 2005. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol. Bioeng. 91 (5), 643–648.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø, Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13 (2), 244–253.

Fowler, Z.L., Gikandi, W.W., Koffas, M.A., 2009. Increased malonyl coenzyme a biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. Appl. Environ. Microbiol. 75 (18), 5831–5839.

Hädicke, O., Klamt, S., 2010. Casop: a computational approach for strain optimization aiming at high productivity. J. Biotechnol. 147 (2), 88–101.

Hädicke, O., Klamt, S., 2011. Computing complex metabolic intervention strategies using constrained minimal cut sets. Metabol. Eng. 13 (2), 204–213.

Imielinski, M., Belta, C., 2008. Exploiting the pathway structure of metabolism to reveal high-order epistasis. BMC Syst. Biol. 2 (1), 40.

Izallalen, M., Mahadevan, R., Burgard, A., Postier, B., Didonato Jr, R., Sun, J., Schilling, C.H., Lovley, D.R., 2008. Geobacter sulfurreducens strain engineered for increased rates of respiration. Metabol. Eng. 10 (5), 267–275.

Jantama, K., Haupt, M., Svoronos, S.A., Zhang, X., Moore, J., Shanmugam, K., Ingram, L., 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* c that produce succinate and malate. Biotechnol. Bioeng. 99 (5), 1140–1153.

Jungreuthmayer, C., Zanghellini, J., 2012. Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. BMC Syst. Biol. 6 (1), 103.

Kaleta, C., de Figueiredo, L., Behre, J., Schuster, S., 2009. EFMEvolver: computing elementary flux modes in genome-scale metabolic networks. In: Lecture Notes in Informatics, vol. 157, pp. 179–189.

Kim, J., Reed, J., 2010. Optorf: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst. Biol. 4 (1), 53.

Kim, J., Reed, J.L., Maravelias, C.T., 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One 6 (9), e24162.

Kim, M., Sang, Yi, J., Kim, J., Kim, J.-N., Kim, M.W., Kim, B.-G., 2014. Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in streptomyces coelicolor a3. Biotechnol. J. (2).

King, Z.A., Feist, A.M., 2013. Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strainsoptswap. Ind. Biotechnol. 9 (4), 236–246.

Klamt, S., Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20 (2), 226–234.

Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U., Lee, S.Y., 2007. Systems metabolic engineering of escherichia coli for ʟ-threonine production. Mol. Syst. Biol. 3, 1.

Lee, S.J., Lee, D.-Y., Kim, T.Y., Kim, B.H., Lee, J., Lee, S.Y., 2005. Metabolic engineering of escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl. Environ. Microbiol. 71 (12), 7880–7887.

Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., et al., 2010. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6 (1).

Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10 (4), 291–305.

Li, S., Huang, D., Li, Y., Wen, J., Jia, X., 2012. Rational improvement of the engineered isobutanol-producing bacillus subtilis by elementary mode analysis. Microb. Cell Fact. 11, 101.

Lun, D.S., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., Church, G.M., 2009. Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5 (1).

Machado, D., Costa, R.S., Rocha, M., Ferreira, E.C., Tidor, B., Rocha, I., 2011. Modeling formalisms in systems biology. AMB Express 1 (1), 1–14.

Machado, D., Soons, Z., Patil, K.R., Ferreira, E.C., Rocha, I., 2012. Random sampling of elementary flux modes in large-scale metabolic networks. Bioinformatics 28 (18), i515–i521.

McCloskey, D., Palsson, B.Ø, Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol. Syst. Biol. 9, 1.

Melzer, G., Esfandabadi, M.E., Franco-Lara, E., Wittmann, C., 2009. Flux design: in silico design of cell factories based on correlation of pathway fluxes to desired properties. BMC Syst. Biol. 3 (1), 120.

Neuner, A., Heinzle, E., 2011. Mixed glucose and lactate uptake by Corynebacterium glutamicum through metabolic engineering. Biotechnol. J. 6 (3), 318–329.

Ng, C.Y., Jung, M.-y., Lee, J., Oh, M.-K., 2012. Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering. Microb. Cell Fact. 11 (1), 68.

Oh, Y.-G., Lee, D.-Y., Lee, S.Y., Park, S., 2009. Multiobjective flux balancing using the nise method for metabolic network analysis. Biotechnol. Prog. 25 (4), 999–1008.

Ohno, S., Shimizu, H., Furusawa, C., 2013. Fastpros: screening of reaction knockout strategies for metabolic engineering. Bioinformatics, btt672, http://dx.doi.org/10.1093/bioinformatics/btt672.

Orth, J., Fleming, R., Palsson, B., 2009. Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. In: Bock, A., Curtiss, I.R., Kaper, J., Karp, P., Neidhardt, F., Nystrom, T., Slauch, J., Squires, C., Ussery, D. (Eds.), EcoSal—Escherichia coli and Salmonella: Cellular and Molecular Biology. ASM Press, Washington, DC, pp. 56–99.

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248.

Otero, J.M., Cimini, D., Patil, K.R., Poulsen, S.G., Olsson, L., Nielsen, J., 2013. Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. PloS One 8 (1), e54144.

Park, J.H., Kim, T.Y., Lee, K.H., Lee, S.Y., 2011. Fed-batch culture of Escherichia coli for l-valine production based on in silico flux response analysis. Biotechnol. Bioeng. 108 (4), 934–946.

Park, J.H., Lee, K.H., Kim, T.Y., Lee, S.Y., 2007. Metabolic engineering of escherichia coli for the production of l-valine based on transcriptome analysis and in silico gene knockout simulation. Proc. Natl. Acad. Sci. 104 (19), 7797–7802.

Park, J.M., Park, H.M., Kim, W.J., Kim, H.U., Kim, T.Y., Lee, S.Y., 2012. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. BMC Syst. Biol. 6 (1), 106.

Patil, K., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinf. 6 (1), 308.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. Optstrain: a computational framework for redesign of microbial production systems. Genome Res. 14 (11), 2367–2376.

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

Poblete-Castro, I., Binger, D., Rodrigues, A., Becker, J., Martins dos Santos, V.A., Wittmann, C., 2013. In-silico-driven metabolic engineering of pseudomonas putida for enhanced production of poly-hydroxyalkanoates. Metabol. Eng. 15, 113–123.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. Optforce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6 (4), e1000744.

Ranganathan, S., Tee, T.W., Chowdhury, A., Zomorrodi, A.R., Yoon, J.M., Fu, Y., Shanks, J.V., Maranas, C.D., 2012. An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. Metabol. Eng. 14 (6), 687–704.

Ren, S., Zeng, B., Qian, X., 2013. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC Bioinf. 14 (Suppl 2), S17.

Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., Rocha, I., 2008. Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinf. 9 (1), 499.

Rockwell, G., Guido, N.J., Church, G.M., 2013. Redirector: designing cell factories by reconstructing the metabolic objective. PLoS Comput. Biol. 9 (1), e1002882.

Sánchez, A.M., Bennett, G.N., San, K.-Y., 2006. Batch culture characterization and metabolic flux analysis of succinate-producing Escherichia coli strains. Metabol. Eng. 8 (3), 209–226.

Sander, J.D., Joung, J.K., 2014. Crispr-cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32 (4), 347–355.

Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst. 2 (02), 165–182.

Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99 (23), 15112–15117.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. 102 (21), 7695–7700.

Soons, Z.I., Ferreira, E.C., Patil, K.R., Rocha, I., 2013. Identification of metabolic engineering targets through analysis of optimal and sub-optimal routes. PLoS One 8 (4), e61648.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26 (4), 536–543.

Tokuyama, K., Ohno, S., Yoshikawa, K., Hirasawa, T., Tanaka, S., Furusawa, C., Shimizu, H., 2014. Increased 3-hydroxypropionic acid production from glycerol, by modification of central metabolism in Escherichia coli. Microb. Cell Fact. 13 (1), 1–11.

Toya, Y., Shiraki, T., Shimizu, H., 2014. Ssdesign: Computational metabolic pathway design based on flux variability using elementary flux modes. Biotechnol. Bioeng, http://dx.doi.org/10.1002/bit.25498.

Trinh, C.T., Carlson, R., Wlaschin, A., Srienc, F., 2006. Design, construction and performance of the most efficient biomass producing E. coli bacterium. Metabol. Eng. 8 (6), 628–638.

Trinh, C.T., Li, J., Blanch, H.W., Clark, D.S., 2011. Redesigning escherichia coli metabolism for anaerobic production of isobutanol. Appl. Environ. Microbiol. 77 (14), 4894–4904.

Trinh, C.T., Srienc, F., 2009. Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. Appl. Environ. Microbiol. 75 (21), 6696–6705.

Trinh, C.T., Unrean, P., Srienc, F., 2008. Minimal escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. Appl. Environ. Microbiol. 74 (12), 3634–3643.

Trinh, C.T., Wlaschin, A., Srienc, F., 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl. Microbiol. Biotechnol. 81 (5), 813–826.

Unrean, P., Trinh, C.T., Srienc, F., 2010. Rational design and construction of an efficient E. coli for production of diapolycopendioic acid. Metabol. Eng. 12 (2), 112–122.

von Kamp, A., Klamt, S., 2014. Enumeration of smallest intervention strategies in genome-scale metabolic networks. PLoS Comput. Biol. 10 (1), e1003378.

Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., Chen, M., 2013a. Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. Appl. Microbiol. Biotechnol. 97 (2), 519–539.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., Koffas, M.A., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-coa. Metabol. Eng. 13 (5), 578–587.

Xu, Z., Zheng, P., Sun, J., Ma, Y., 2013b. Reacknock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. PLoS One 8 (12), e72150.

Yang, L., Cluett, W.R., Mahadevan, R., 2011. Emilio: a fast algorithm for genome-scale strain design. Metabol. Eng. 13 (3), 272–281.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., et al., 2011. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat. Chem. Biol. 7 (7), 445–452.

Yousofshahi, M., Orshansky, M., Lee, K., Hassoun, S., et al., 2013. Probabilistic strain optimization under constraint uncertainty. BMC Syst. Biol. 7 (1), 29.

Zhuang, K., Yang, L., Cluett, W.R., Mahadevan, R., 2013. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. Dyssco strategy for strain design. BMC Biotechnol. 13 (1), 1–15.