Computational Methods in Metabolic Engineering for Strain Design

Matthew R. Long^{1,*}, Wai Kit Ong^{1,2,*}, and Jennifer L. Reed^{1,2}

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

²Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI

*These authors contributed equally to this work

Corresponding Author:

Jennifer L. Reed (reed@engr.wisc.edu);

1415 Engineering Dr, Madison, WI 53706

Keywords: metabolic engineering, strain design, flux balance analysis, constraint-based model; genome-scale model

Abstract

Metabolic engineering uses genetic approaches to control microbial metabolism to produce desired compounds. Computational tools can identify new biological routes to chemicals and the changes needed in host metabolism to improve chemical production. Recent computational efforts have focused on exploring what compounds can be made biologically using native, heterologous, and/or enzymes with broad specificity. Additionally, computational methods have been developed to suggest different types of genetic modifications (e.g., gene deletion/addition or up/down regulation), as well as suggest strategies meeting different criteria (e.g., high yield, high productivity, or substrate co-utilization). Strategies to improve the runtime performances have also been developed, which allow for more complex metabolic engineering strategies to be identified. Future incorporation of kinetic considerations will further improve strain design algorithms.

Introduction

Cells can be used industrially to make a wide variety of products, including food and beverages, therapeutics, enzymes, commodity chemicals, and specialty chemicals. Many organisms can be manipulated to enhance chemical production or improve cellular properties by modifying genetics, media composition, and/or reactor operating conditions. Metabolic engineering primarily uses genetic approaches to adjust metabolic and regulatory networks in biocatalysts to enhance production of desired chemicals. These desired chemicals can be natural biological molecules (e.g., amino acids, antibiotics, and ethanol) or molecules that are not normally produced biologically (e.g., 1,4-butanediol [1] and 3,4-dihydroxybutyric acid [2]). Regardless of whether the molecule is natural or non-natural, product biosynthesis pathways need to be identified and the metabolic network needs to be manipulated to increase flux towards the product of interest. Metabolic engineering can involve increasing precursor production, recycling cofactors (e.g., NAD(P)H), improving substrate utilization, removing bottlenecks in biosynthesis pathways, eliminating feedback regulation, and/or improving tolerance to products and their intermediates.

To facilitate metabolic engineering efforts, a variety of computational tools have been developed to identify genetic strategies (i.e., strain designs) for enhancing chemical production. Many of these tools model the integration of product biosynthesis pathways with host metabolism to predict how metabolic and regulatory changes will affect chemical production yields, titers, and productivity. Genome-scale constraint-based models have been useful in metabolic engineering since they can account for the entire metabolic (and regulatory) capabilities of a cell and do not require detailed kinetic parameters to generate them. Constraint-based models can be quickly generated by a variety of software packages (reviewed in [3]) using an organism's genomic, biochemical, and physiological data. All constraint-based models use two types of constraints to define a metabolic solution space: (1) steady-state mass balances (which equate metabolite production and consumption rates), and (2) flux bounds (which set maximum and minimum limits for metabolic fluxes based on an enzyme's capacity and directionality) [4]. Here, we discuss recent advances in biosynthesis pathway identification algorithms and constraint-based modeling approaches that can be used in metabolic engineering for strain design (**Figure 1**).

Identifying Biosynthesis Pathways

Pathways Involving Known Reactions

A number of databases exist that describe the reactions that can be catalyzed by known enzymes, including KEGG, BioCyc, BRENDA, MetRxn, and SEED [5-9]. These databases can be used to develop organism-specific reaction networks (based on an organism's genome annotation), as well as comprehensive reaction networks (containing known biochemical reactions for all organisms). Native and non-native pathways that lead to product formation can be identified from these reaction networks using graphical and modeling approaches. While graphical approaches can identify shortest paths to a desired product, these paths might violate steady-state mass balances (indicating that reactions used might consume or produce a metabolite that cannot be synthesized or degraded, respectively) [10]. Methods that include steady-state mass balances can be used to find mass-balanced paths [10], and constraint-based models can identify paths that are the shortest and/or have the highest maximum yield [11, 12].

A variety of constraint-based modeling approaches can identify biosynthesis pathways involving native and/or non-native pathways by analyzing organism-specific reaction networks augmented with all known reactions. Flux balance analysis (FBA) can identify mass-balanced pathways with the highest

yields [12]. OptStrain [13] and SimOptStrain [14] can be used to identify non-native reactions that should be added (and native reactions that should be deleted) to improve product formation. Both OptStrain and SimOptStrain try to minimize the number of added reactions and couple product and biomass formation using deletions. OptStrain performs the addition and deletion steps sequentially, and as a result can miss higher yielding production strategies [14]. SimOptStrain was therefore developed to simultaneously consider additions and deletions. These approaches can identify pathways involving known reactions; however, they cannot identify potentially novel pathways for which enzymes might need to be discovered or engineered.

Pathways Involving Novel Reactions

There are many commercially desirable compounds with no known biological synthesis route. In these situations, new routes need to be found along with enzymes that can catalyze any new reactions. A variety of de novo pathway methods have been developed to identify these novel reaction pathways and potential enzymes that could catalyze them. In general, these methods apply Biochemical Reaction Operators (BROs) [15]—generic reaction rules that act on a reaction center—to starting or target metabolites to enumerate potential reactions that could occur. After enumerating all possible de novo pathways of interest, most methods filter or rank these potential pathways, and some methods propose enzymes that could potentially catalyze the desired reactions (**Table 1**).

Novel reactions (and pathways) can be evaluated based on thermodynamic and enzyme considerations (which provides confidence an enzyme could be found or engineered to catalyze the novel reaction). Thermodynamic analysis usually uses Group Contribution Methods [16] to estimate Gibbs free energies of reactions ($\Delta_r G'$) and filters out infeasible pathways (e.g., if $\Delta_r G' \geq 25$ kJ/mol). Reactions can also be scored based on the structural similarities between known and novel substrates (e.g., Tanimoto coefficient) and/or molecular simulations of the reactions. Since many enzymes can act on multiple substrates, enzymes with structurally similar substrates to the novel substrates may be able to catalyze the novel reaction.

A detailed review of novel pathway design methods was recently published [17] and a summary of the different methods is provided in **Table 1**. BNICE [18] was the foundational method, which developed BROs from the KEGG database. BNICE ranks pathways based on pathway length, thermodynamic feasibility, maximum yield, and maximum activity [19]. BNICE has been further extended to identify potential enzymes to catalyze novel reactions based on substrate similarity and molecular simulations [20]. A related method by Cho et al. [21] uses a unique ranking system which incorporates binding site covalence, chemical similarity, and organism specificity. RetroPath [22] uses molecular signatures to automatically filter the applicable BROs to those with similar substrates or products. RetroPath applies reactions to a target compound until a known metabolite is reached and also uniquely includes gene compatibility, enzymatic performance, and compound toxicity in its pathway ranking metric. DESHARKY [23] incorporates the growth potential of the cell with consideration to cellular load for enzyme production.

These de novo pathway design methods have been applied to design pathways for a variety of products. Recently, GEM-Path [15] was used to identify products that could be made within a limited number of steps from native *Escherichia coli* metabolites. In contrast to many of the other methods, filtering is done at each step to reduce the number of possible paths that are considered. In this way, adding one reaction step to a genome-scale model of *E. coli* resulted in 120 non-native compounds, significantly less than alternative methods that would report all feasible and infeasible compounds. Ultimately, pathways to produce 20 large volume compounds were reported [15]. Yim et al. used a de novo pathway design tool to identify and implement the first biological route to synthesize 1,4-butanediol [1]. Pathways were

ranked based on theoretical yield, pathway length and types, and thermodynamic feasibility. After identifying top candidates, enzymes catalyzing the necessary biosynthesis reactions were heterologously expressed and host metabolism was modified through gene knockouts, yielding strains capable of producing up to 18 g/L. These examples illustrate that new routes to desired chemicals can be found computationally using known and novel reactions

Engineering Host Metabolism

Once a biosynthesis pathway has been identified and the necessary enzymes expressed in the host, other aspects of native host metabolism may need to be adjusted to provide sufficient precursors, ATP, redox carriers, and cofactors. Experimentally, there are three common genetic approaches for modifying a strain's metabolism to overproduce a target metabolite: gene knockout, upregulation, and downregulation. As such, various computational strain design algorithms have been developed to improve production of a target metabolite using one or a combination of the above genetic manipulation techniques. Many of these algorithms rely on 'gene perturbation' constraint-based modeling tools [4], such as FBA or minimization of metabolic adjustment (MOMA), to predict chemical production rates for a set of genetic manipulations [12, 24]. For example, FBA can be used to predict chemical production rates at the maximal growth rate or predict minimal production rates (e.g., to consider worst case scenarios) for a given knockout. The algorithms discussed below usually embed a gene perturbation tool in them so that production rates from different genetic manipulation strategies can be compared to identify the top strategies (**Figure 2**). The following sections discuss recent advances for identifying gene deletion and gene up/down regulation to improve chemical production.

Modifications Involving Deletions

A number of methods exist to predict the effects of gene deletions [12, 24-27], which can be used in an exhaustive fashion (e.g., evaluate all possible triple deletions) to identify high production strategies [28, 29]; however, strain design algorithms are computationally more efficient than exhaustive searches. The first strain design method involving knockouts is OptKnock [30], which is a bi-level optimization framework to identify optimal reaction deletion strategies that couple cellular growth with target metabolite production. OptKnock's outer problem selects reactions to delete while the inner problem uses FBA to maximize cellular growth given those deletions and returns the corresponding chemical production rate to the outer problem. OptKnock identifies those deletions with the highest chemical production at the maximum growth rate. Since OptKnock was published, a variety of related approaches have been developed to (1) prevent strategies which could have both high and low production at the maximum growth rate [31, 32], (2) use MOMA instead of FBA to predict knockout growth and production rates [14, 33], (3) use gene deletions instead of reaction deletions [33-36], and (4) use regulatory models to find metabolic and regulatory gene deletions [35].

Most of these methods optimize only one criterion (i.e., maximum chemical production); however, other criteria may also be important. In addition to chemical production, OptGene can also find strains that maximize the product of growth and production rates (an indicator of productivity) [33]. DySScO uses dynamic FBA simulations of a bioreactor to compare strains based on predicted yields, titers, and volumetric productivity [37]. Two recent methods (SIMUP and FOCAL) can be used to design strains for substrate co-utilization [38, 39], and SIMUP designed strains were shown experimentally to co-utilize both glucose and xylose [38]. A general framework to include multiple criteria (in addition to chemical production) that strain designs must satisfy has also been proposed [40]. Such criteria can include the degree and type of coupling between growth rate and chemical production, and the minimum (or maximum) allowable levels of product (or by-product) secretion rate. For example, *E. coli* strains that must produce ethanol at any growth rate were identified [40]. Constrained minimal cut sets (cMCSs) can

also be used to identify deletions that eliminate an undesired functionality while retaining a desired functionality (e.g., achieve high product yields while eliminating undesired by-product formation) [41]. While this approach has generally only been applied to small reaction networks, recent advances allow cMCSs to be calculated for genome-scale networks [42].

Modifications Involving Up/Down Regulation

While significant focus has been placed on algorithms to identify knockout strategies, other types of genetic manipulations are often needed to improve product formation. One of the first strain design algorithms that considered flux up- and down-regulation in addition to knockouts was OptReg [43]. OptReg identified manipulations needed to achieve 99.8% of the maximum theoretical ethanol yield in *E. coli* at the maximum growth rate. Since OptReg's publication, many other algorithms have emerged that: (1) identify the minimum set of flux bound modifications needed to guarantee a minimum production yield of the target metabolite at all growth rates [44, 45], (2) include flux bounds as variables instead of parameters [44, 46], (3) identify coupling of the target metabolite with cellular growth through analysis of flux bounds using flux variability analysis (FVA) [47, 48], (4) use an iterative local search scheme with target enzyme manipulation added to the biomass objective function [49], and (5) identify genes whose overexpression is needed to overcome transcriptional regulatory effects [35].

Approaches for Improving Runtime Performance

Many of the above algorithms (summarized in **Supplementary Table S1**) are formulated as mixed-integer linear programming problems which are computationally expensive to solve to optimality. For example, OptKnock can take up to 10 hours to find a triple knockout for acetate production in *E. coli*, and the runtimes increase exponentially when more knockouts are allowed [34]. To improve runtime performance, different approaches have been used including: using genetic algorithms [33, 36] and iterative linear programs [46, 49]; adjusting truncated branch and bound options [50]; reformulating problem objective and bounds [14] or using Karush-Kuhn-Tucker conditions [46, 51]; and restricting allowable knockouts [34, 52]. While these methods improve runtime performance, they can sometimes come at the expense of missing more optimal solutions.

Conclusions

Many of the computational approaches described above have been used to successfully develop new production strains and/or have identified strategies that were consistent with previous metabolic engineering studies. These computational approaches primarily rely on structural information (i.e., reaction stoichiometry) to predict production phenotypes. As such, they cannot distinguish between different steps in a linear pathway to find which ones are rate limiting. In addition, while deleting a gene will eliminate flux through associated reactions, it is currently difficult to predict how gene expression changes will impact fluxes using stoichiometric models. While other approaches (e.g., kinetic models, ensemble models, and metabolic control analysis) can overcome some of these limitations, they are often applied to models containing only a small number of reactions and pathways. Methods for incorporating kinetic information into genome-scale models and strain design algorithms are needed and such efforts have recently begun [53-57]. A key challenge will be how to parameterize kinetic rate laws used in the models and how to handle uncertainties in parameter estimates. The development of such models will lead to improved strain design algorithms that will facilitate the rapid construction, evaluation, and improvement of metabolically engineered strains.

Acknowledgement

This work was funded by the Office of Science (BER), U.S. Department of Energy (DE-SC008103) and the U.S. Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure Captions

Figure 1. Overview of Strain Design Algorithms. Computational methods have been developed to identify pathways to enable production of desired compounds using known native (solid black and orange arrows) and known non-native (dashed green arrows) reactions, as well as de novo reactions (dotted blue arrows) that have not been known to occur biologically. Additional host metabolic changes can be suggested by a variety of constraint-based modeling algorithms, which can involve deletions (red cross) and up/down regulation (orange arrows). The example de novo pathways and metabolic modifications shown are for illustrative purposes and were reported previously for 1,4-butanediol and succinate production [1, 44, 45]. Chemical structures were generated using Open Babel [58].

Figure 2. Flux Prediction Methods used by Strain Design Algorithms. Strain design algorithms use different approaches to predict chemical production rates in response to host metabolism modifications. FBA can be used to predict fluxes that maximize growth rate (μ); optimize a linear combination of other fluxes, e.g., product or by-product formation; optimize a flux ratio, e.g., product formation (v_p) to growth rate; or a combination of these. MOMA can be used instead to predict fluxes (v) that are closest to the un-perturbed or parental flux distribution (w) by minimizing the Euclidean distance between the two states. The Venn diagrams indicate what flux prediction methods and their objectives are used by the different strain design algorithms. MCS Enumerator [42] uses an optimization framework to find minimal cut sets and is not shown.

224 **Tables**225 **Table 1.** De novo pathway prediction algorithms.

Name	Pathway Scoring and Filtering	Enzyme Identification
BNICE [18]	Pathway length, thermodynamics, maximum yield, maximum activity	Retrofit and computationally engineer an enzyme for a de novo reaction using substrate similarity and molecular simulations [20]
DESHARKY [23]	Pathway length, RNA polymerase and ribosomal capacity, cellular growth rate	Not available
Cho et al. [21]	Pathway length, thermodynamics, substrate and binding site similarity, organism specificity	Not available
PathPred/E-zyme [59, 60]	Substrate similarity	Assigns third-tier E.C. number
SimPheny Biopathway Predictor [1]	Pathway length, thermodynamics, maximum yield	Not available
RetroPath [22]	Thermodynamics, substrate similarity, enzyme promiscuity, enzyme kinetics, gene compatibility, compound toxicity, cellular growth rate, expected yield	Specific enzyme from substrate similarity
GEM-Path [15]	Pathway length, thermodynamics, enzyme promiscuity, substrate similarity	Specific enzyme and organism from enzyme matching and substrate similarity

226

227

References (1,15,20,38,40,57 special interest (●) or outstanding interest (●●).)

1. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, et al.: **Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol.** *Nat Chem Biol* 2011, **7**:445-52.
- ** The authors developed a novel biological route to produce 1,4-butanediol. The authors used both de novo pathway and strain design algorithms to develop *E. coli* strains capable of producing 1,4-butanediol up to 18 g/L.
2. Martin CH, Dhamankar H, Tseng HC, Sheppard MJ, Reisch CR, Prather KL: **A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy-gamma-butyrolactone.** *Nat Commun* 2013, **4**:1414.
3. Hamilton JJ, Reed JL: **Software platforms to facilitate reconstructing genome-scale metabolic networks.** *Environmental microbiology* 2014, **16**:49-59.
4. Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.** *Nat Rev Microbiol* 2012, **10**:291-305.
5. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG.** *Nucleic Acids Res* 2014, **42**:D199-205.
6. Sreenath TL, Polasa H: **Effect of N-beta-phenyl ethyl anthranilic acid on macromolecular synthesis during outgrowth of Bacillus subtilis hcr-9 spores.** *Microbios* 1985, **44**:281-3.
7. Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, et al.: **BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA.** *Nucleic Acids Res* 2013, **41**:D764-72.
8. Kumar A, Suthers PF, Maranas CD: **MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases.** *BMC Bioinformatics* 2012, **13**:6.
9. Henry CS, DeJongh M, Best AA, Frybarger PM, Lindsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models.** *Nat Biotechnol* 2010, **28**:977-82.
10. Pey J, Prada J, Beasley JE, Planes FJ: **Path finding methods accounting for stoichiometry in metabolic networks.** *Genome Biol* 2011, **12**:R49.
11. Burgard AP, Vaidyaraman S, Maranas CD: **Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments.** *Biotechnol Prog* 2001, **17**:791-7.
12. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245-8.
13. Pharkya P, Burgard AP, Maranas CD: **OptStrain: a computational framework for redesign of microbial production systems.** *Genome Res* 2004, **14**:2367-76.
14. Kim J, Reed JL, Maravelias CT: **Large-scale bi-level strain design approaches and mixed-integer programming solution techniques.** *PLoS One* 2011, **6**:e24162.
15. Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM: **Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path.** *Metab Eng* 2014, **25**:140-58.
- ** Using a filtered de novo synthesis algorithm, the authors identified multiple pathways for the production of 20 high volume products in *E. coli*. This is the first study to develop a parameter for the promiscuity of each known enzyme and identify reaction-organism associations for finding the best candidate enzymes.
16. Mavrovouniotis ML: **Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution.** *Biotechnol Bioeng* 1990, **36**:1070-82.

17. Shin JH, Kim HU, Kim DI, Lee SY: **Production of bulk chemicals via novel metabolic pathways in microorganisms.** *Biotechnol Adv* 2013, **31**:925-35.
 18. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ: **Exploring the diversity of complex metabolic networks.** *Bioinformatics* 2005, **21**:1603-9.
 19. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate.** *Biotechnol Bioeng* 2010, **106**:462-73.
 20. Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U: **Integrating computational methods to retrofit enzymes to synthetic pathways.** *Biotechnol Bioeng* 2012, **109**:572-82.
- ** The authors used molecular dynamics simulations to identify the binding and catalytic efficiency for a proposed promiscuous enzyme and further engineered the enzyme to better catalyze the desired reaction. This is the first study to combine de novo pathway synthesis with enzyme modeling and design.
21. Cho A, Yun H, Park JH, Lee SY, Park S: **Prediction of novel synthetic pathways for the production of desired chemicals.** *BMC Syst Biol* 2010, **4**:35.
 22. Carbonell P, Planson AG, Fichera D, Faulon JL: **A retrosynthetic biology approach to metabolic pathway design for therapeutic production.** *BMC Syst Biol* 2011, **5**:122.
 23. Rodrigo G, Carrera J, Prather KJ, Jaramillo A: **DESHARKY: automatic design of metabolic pathways for optimal cell growth.** *Bioinformatics* 2008, **24**:2554-6.
 24. Segre D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks.** *Proc Natl Acad Sci U S A* 2002, **99**:15112-7.
 25. Shlomi T, Berkman O, Ruppin E: **Regulatory on/off minimization of metabolic flux changes after genetic perturbations.** *Proc Natl Acad Sci U S A* 2005, **102**:7695-700.
 26. Park JM, Kim TY, Lee SY: **Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses.** *Proc Natl Acad Sci U S A* 2010, **107**:14931-6.
 27. Kim J, Reed JL: **RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations.** *Genome Biol* 2012, **13**:R78.
 28. Alper H, Jin YS, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli.** *Metab Eng* 2005, **7**:155-64.
 29. Park JH, Lee KH, Kim TY, Lee SY: **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 U S A* 2007, **104**:7797-802.
 30. Burgard AP, Pharkya P, Maranas CD: **Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization.** *Biotechnol Bioeng* 2003, **84**:647-57.
 31. Tepper N, Shlomi T: **Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways.** *Bioinformatics* 2010, **26**:536-43.
 32. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO: **Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli.** *Metab Eng* 2009, **12**:173-86.
 33. Patil KR, Rocha I, Forster J, Nielsen J: **Evolutionary programming as a platform for in silico metabolic engineering.** *BMC Bioinformatics* 2005, **6**:308.
 34. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM: **Large-scale identification of genetic design strategies using local search.** *Mol Syst Biol* 2009, **5**:296.
 35. Kim J, Reed JL: **OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains.** *BMC Syst Biol* 2010, **4**:53.

36. Costanza J, Carapezza G, Angione C, Lio P, Nicosia G: **Robust design of microbial strains.** *Bioinformatics* 2012, **28**:3097-104.
37. Zhuang K, Yang L, Cluett WR, Mahadevan R: **Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity.** *DySScO strategy for strain design.* *BMC Biotechnol* 2013, **13**:8.
38. Gawand P, Hyland P, Ekins A, Martin VJ, Mahadevan R: **Novel approach to engineer strains for simultaneous sugar utilization.** *Metab Eng* 2013, **20**:63-72.
- ** The authors developed an approach (SIMUP) for designing knockout mutants that must co-utilize two sugars (glucose and xylose) in order to grow. Experiments were then conducted on three strains and two strains co-utilized both substrates to grow.
39. Tervo CJ, Reed JL: **FOCAL: an experimental design tool for systematizing metabolic discoveries and model development.** *Genome Biol* 2012, **13**:R116.
40. Tervo CJ, Reed JL: **Expanding metabolic engineering algorithms using feasible space and shadow price constraint modules.** *Metab Eng Commun* 2015, **1**:1-11.
- * The authors developed two formulations (FaceCon and ShadowCon) that can be included into other strain design algorithms to identify strains that meet additional criteria. Examples were provided for knockout strains that require ethanol production to grow, strains with high ethanol production and no formate production, and strains with different degrees of coupling between ethanol and biomass production.
41. Hadicke O, Klamt S: **Computing complex metabolic intervention strategies using constrained minimal cut sets.** *Metab Eng* 2011, **13**:204-13.
42. von Kamp A, Klamt S: **Enumeration of smallest intervention strategies in genome-scale metabolic networks.** *PLoS Comput Biol* 2014, **10**:e1003378.
43. Pharkya P, Maranas CD: **An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems.** *Metab Eng* 2006, **8**:1-13.
44. Cotten C, Reed JL: **Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering.** *Biotechnol J* 2013, **8**:595-604.
45. Ranganathan S, Suthers PF, Maranas CD: **OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions.** *PLoS Comput Biol* 2010, **6**:e1000744.
46. Yang L, Cluett WR, Mahadevan R: **EMILiO: a fast algorithm for genome-scale strain design.** *Metab Eng* 2011, **13**:272-81.
47. Choi HS, Lee SY, Kim TY, Woo HM: **In silico identification of gene amplification targets for improvement of lycopene production.** *Appl Environ Microbiol* 2010, **76**:3097-105.
48. Park JM, Park HM, Kim WJ, Kim HU, Kim TY, Lee SY: **Flux variability scanning based on enforced objective flux for identifying gene amplification targets.** *BMC Syst Biol* 2012, **6**:106.
49. Rockwell G, Guido NJ, Church GM: **Redirector: designing cell factories by reconstructing the metabolic objective.** *PLoS Comput Biol* 2013, **9**:e1002882.
50. Egen D, Lun DS: **Truncated branch and bound achieves efficient constraint-based genetic design.** *Bioinformatics* 2012, **28**:1619-23.
51. Xu Z, Zheng P, Sun J, Ma Y: **ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network.** *PLoS one* 2013, **8**:e72150.
52. Ohno S, Shimizu H, Furusawa C: **FastPros: screening of reaction knockout strategies for metabolic engineering.** *Bioinformatics* 2014, **30**:981-7.
53. Cotten C, Reed JL: **Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models.** *BMC bioinformatics* 2013, **14**:32.

54. Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T: **Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model.** *Bioinformatics* 2010, **26**:i255-60.
55. Khodayari A, Zomorodi AR, Liao JC, Maranas CD: **A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data.** *Metab Eng* 2014, **25**:50-62.
56. Chakrabarti A, Miskovic L, Soh KC, Hatzimanikatis V: **Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints.** *Biotechnol J* 2013, **8**:1043-57.
57. Chowdhury A, Zomorodi AR, Maranas CD: **k-OptForce: integrating kinetics with flux balance analysis for strain design.** *PLoS Comput Biol* 2014, **10**:e1003487.
- ** By combining stoichiometric models with available kinetic descriptions, the authors developed a method that suggests both reaction flux and enzymatic parameter changes to achieve production of a desired compound. Analysis of L-serine production using k-OptForce highlights the importance of substrate-level inhibition in strain design
58. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR: **Open Babel: An open chemical toolbox.** *J Cheminform* 2011, **3**:33.
59. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S, Kanehisa M: **PathPred: an enzyme-catalyzed metabolic pathway prediction server.** *Nucleic Acids Res* 2010, **38**:W138-43.
60. Yamanishi Y, Hattori M, Kotera M, Goto S, Kanehisa M: **E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs.** *Bioinformatics* 2009, **25**:i179-86.

Figure 1

Figure 2