J Ind Microbiol Biotechnol (2015) 42:339–348
DOI 10.1007/s10295-014-1554-9



SIMB
Society for Industrial Microbiology and Biotechnology

# Applications of genome-scale metabolic network model in metabolic engineering

**Byoungjin Kim · Won Jun Kim · Dong In Kim · Sang Yup Lee**

Received: 21 October 2014 / Accepted: 19 November 2014 / Published online: 3 December 2014


**Abstract** Genome-scale metabolic network model (GEM) is a fundamental framework in systems metabolic engineering. GEM is built upon extensive experimental data and literature information on gene annotation and function, metabolites and enzymes so that it contains all known metabolic reactions within an organism. Constraint-based analysis of GEM enables the identification of phenotypic properties of an organism and hypothesis-driven engineering of cellular functions to achieve objectives. Along with the advances in omics, high-throughput technology and computational algorithms, the scope and applications of GEM have substantially expanded. In particular, various computational algorithms have been developed to predict beneficial gene deletion and amplification targets and used to guide the strain development process for the efficient production of industrially important chemicals. Furthermore, an *Escherichia coli* GEM was integrated with a pathway prediction algorithm and used to evaluate all possible routes for the production of a list of commodity chemicals in *E. coli*. Combined with the wealth of experimental data produced by high-throughput techniques, much effort has been exerted to add more biological contexts into GEM through the integration of omics data and regulatory network information for the mechanistic understanding and improved prediction capabilities. In this paper, we review the recent developments and applications of GEM focusing on the GEM-based computational algorithms available for microbial metabolic engineering.

**Keywords** Genome-scale metabolic network · Systems metabolic engineering · Gene knock-out prediction · Gene amplification prediction · Metabolic pathway prediction · Integrated genome-scale model

## Introduction

Genome-scale metabolic network model (GEM) has become an essential tool for systems-level understanding of metabolism and its applications in metabolic engineering. GEM is reconstructed through the extensive collection and curation of the biological information on gene annotation and functions, metabolites, metabolic reactions, enzymes and their interactions within a given organism. It is the most comprehensive and a standardized representation of our knowledge on metabolic signature of microorganisms. Currently, more than 100 GEMs are available and the number of organisms with reconstructed GEMs is rapidly growing due to the advances in both high-throughput technologies and the tools for automatic data collection and draft network reconstruction [32, 87].

Computational algorithms such as constraint-based flux balance analysis (FBA) are essential to systematically analyze reconstructed GEMs and have been used to understand the objectives and functions of metabolic network and (adaptive) evolutionary process [17, 36, 40, 70]. Constraint-based GEM modeling is also proven to be a powerful tool to predict the phenotypic properties under genetic and environmental perturbations. In particular, the value of constraint-based GEM modeling was well appreciated

Special Issue: Metabolic Engineering.

B. Kim · W. J. Kim · D. I. Kim · S. Y. Lee (✉)
Department of Chemical and Biomolecular Engineering (BK21 Plus Program), BioProcess Engineering Research Center, Bioinformatics Research Center, Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
e-mail: leesy@kaist.ac.kr

**Table 1** Computational tools for genome-scale reconstruction, strain design, omics data integration, and pathway prediction

| Application | Tools |
|---|---|
| GEM reconstruction | Model SEED [32], Pathway Tools [42], SuBliMinaL [85], MBA [39], GLAMM [4], RAVEN [1], Path2Models [9] |
| Strain design | |
| Knockout | MOMA [78], ROOM [81], OptKnock [10], RobustKnock [86], OptSwap [95], OptGene [66], GDLS [53], EMILiO [93], OptFlux [73] |
| Amplification | FSEOF [15], FVSEOF with GR [65] |
| Knockout, knockdown, amplification | OptStrain [67], OptReg [68], OptForce [72], k-OptForce [16], OptORF [44], CosMos [20] |
| Omics data integration | |
| Transcriptome | GIMME [5], iMAT [82], GIM$^3$E [76], E-Flux [18], PROM [13], MADE [38], tFBA [90], RELATCH [45], TEAM [19], AdaM [89], GX-FBA [60], mCADRE [92], FCGs [43], EXAMO [75], TIGER [37] |
| Proteome | GIMMEp [6] |
| Pathway prediction | BNICE [29], Cho et al. [14], RetroPath [11], PathPred [59], DESHARKY [74], BioPath [94], XTMS [12], GEM-Path [56] |

in the field of systems metabolic engineering where the identification of targets for genetic alteration at the system-wide context is crucial to achieve the desired metabolic phenotype with minimal experimental trial and error. Many computational tools have been developed to predict promising gene deletion targets for increased production of target compounds [78, 86]. The earlier algorithms for identifying gene knockout targets have successfully been employed for the enhanced production of desired products upon knocking out the predicted gene targets. Depending on the algorithms, different mathematical formulations with optimization methods have been developed to handle two cellular objectives of cell growth and maximal production of target compound, which are often compromised by each other. In addition to the prediction of gene knockout targets, the algorithms have been expanded to predict the up- and down-regulation gene targets, allowing more exhaustive search of phenotypic space [28, 55, 65, 68, 72]. These computational tools have been successfully employed for the metabolic engineering of *Escherichia coli*, *Saccharomyces cerevisiae*, *Corynebacterium glutamicum,* and many others for the production of a range of natural and non-natural compounds including ethanol [22, 49], butanol [48, 71], succinic acid [46], lactic acid [26], lycopene [15], amino acids [64], vanillin [8], and 1,4-butanediol [94].

Entering into the post-omics era, GEM has evolved to harness the wealth of sequence and omics data [3, 21, 34, 54]. Omics data produced by high-throughput technologies provide an accurate snapshot of cellular status under a given metabolic environment; however, they are inherently enormous in scale, incomplete, noisy and complex to interpret. Integration of omics data into the stoichiometric representation of GEM not only allows the analysis of cellular metabolism in a context-specific manner but provides a valuable framework for the collection of the ever-expanding omics data (Fig. 2). For their efficient integration, computational tools have been developed to map genome-wide gene expression and proteomics data onto GEM (Table 1); these tools attempt to improve the prediction accuracy of metabolic phenotypes by utilizing omics data as additional constraints. The *E. coli* GEM was expanded to account for all the biochemical reactions involved in the synthesis of transcriptional and translational machinery and protein complex formation so that the model could predict cell growth, nutrient uptake, product secretion, metabolic fluxes and gene expression levels. [50, 51]. Recently, an *E. coli* GEM was also integrated with a search algorithm for heterologous metabolic pathway (GEM-Path) to generate synthetic pathways for a given substrate and product [56]. Indeed, more and more innovative improvement and applications of GEM are being reported. In this paper, we review the recent progress in GEM reconstruction and computational tools with underlying principles (Fig. 1). Also, the strategies for integrating omics data, underlying regulatory network, and pathway prediction algorithms to GEM are highlighted. Future prospects are discussed towards further improvement of GEM and simulation algorithms.

## Genome-scale metabolic network

The *E. coli* metabolic network is a representative and the best validated GEM, largely due to the availability of comprehensive experimental data and its relatively simple network structure. Since its first report in 2000, *E. coli* GEM has continuously been updated over the past decade [23, 40]. The iAF1260 model published in 2007 accounts for 1260 open reading frame corresponding to 29 % of all the annotated genes (4,325 genes) and contains 1,039 metabolites and 1,387 metabolic reactions [24]. Using conditional essentiality analysis and new data, the model was further updated to the most recent version of iJO1366. The iJO1366 model accounts for 1,366 genes corresponding 32 % of all the annotated genes, 1,136 unique metabolites, and 1,473 metabolic reactions [62]. These recent models were reconstructed considering element and charge balanced reactions, thermodynamic consistency and gene-reaction-protein associations. All these features have been found to be critical to improve the accuracy of the prediction of cellular

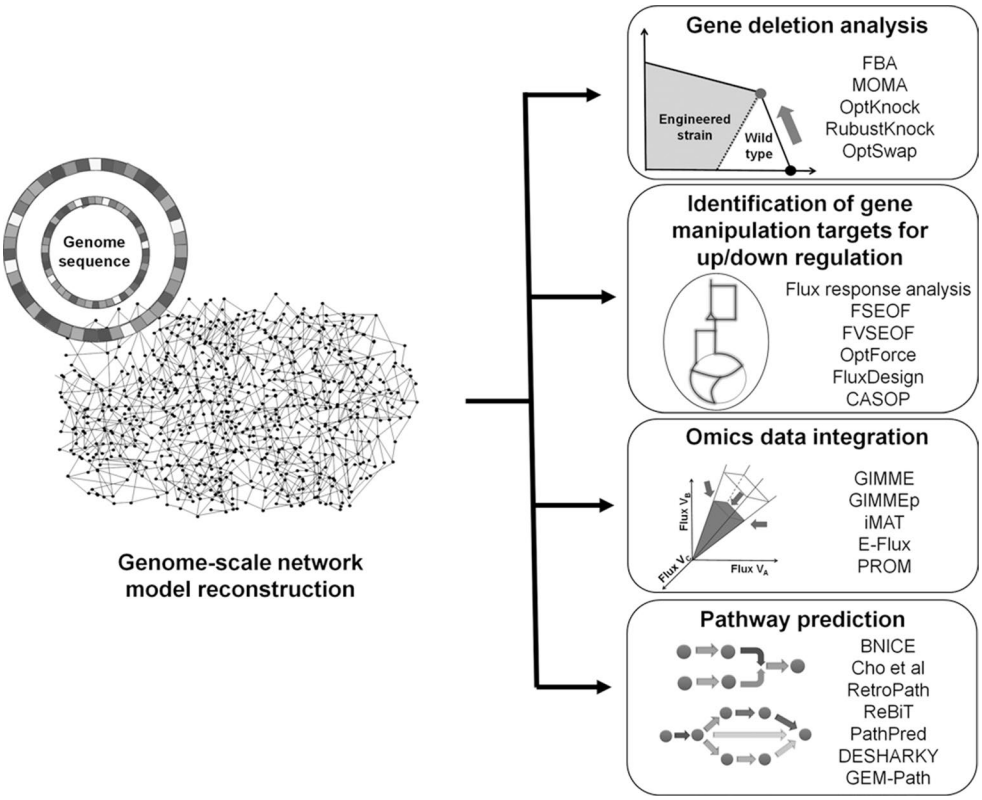


**Fig. 1** Applications of genome-scale metabolic network model. Using various algorithms, the genome-scale metabolic model can be employed for predicting gene manipulation targets, integrating omics data, and for predicting metabolic pathways for the production of a target compound

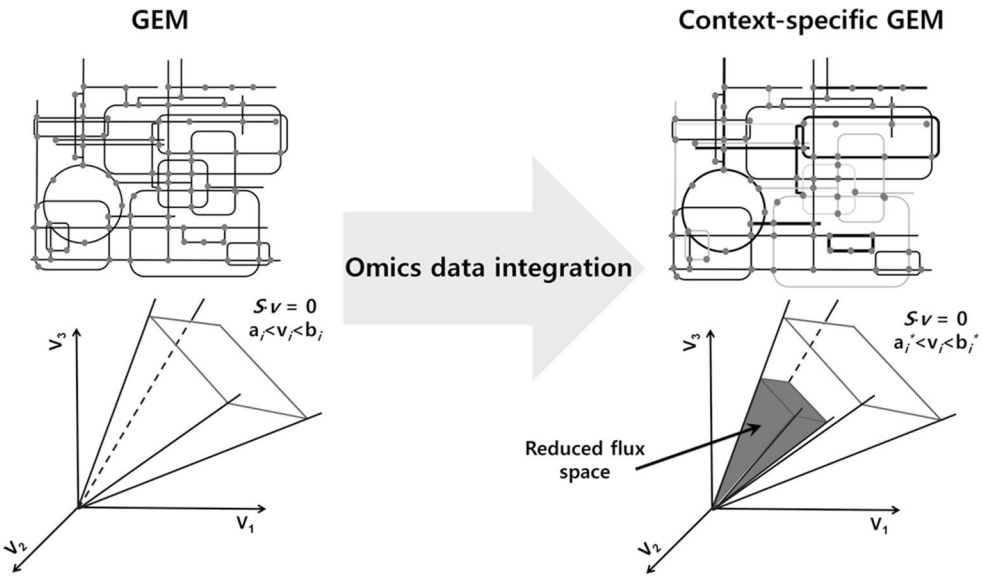


**Fig. 2** The consequence of omics data integration into genome-scale network model. Omic data integration provides a context-specific genome-scale metabolic network model (*S*) under a given genetic and environmental condition. The ranges ($a_i$ and $b_i$) of the fluxes ($v_i$) given as constraints are modified to reflect the transcriptome and proteomic data ($a_i^*$ and $b_i^*$) reducing the space of possible flux distribution

phenotype and gene essentiality [24]. Even further, taking advantage of a large set of genome sequences available for various *E. coli* strains, the GEMs for 55 *E. coli* strains were used to investigate the variations in gene, reaction and metabolite contents, and the capabilities to adapt to different nutritional environments among the strains [40].

Among eukaryotic microbes, *S. cerevisiae* is the best-characterized model microbial organism. Since the first report of a compartmentalized GEM for *S. cerevisiae* (iFF708) in 2003, the number of metabolites and reactions were substantially increased from 584 to 1,353 and 1,145 to 1,566, respectively, in the recent *S. cerevisiae* GEM iTO977 published in 2013 [27, 30, 63]. The complex nature of the metabolic network of *S. cerevisiae* due to the large genomic contents and cellular compartments initiated an effort to consolidate a large pool of common knowledge from various fields into a consensus

model. A community-based approach emerged to efficiently reconcile the discrepancies among the models by standardizing nomenclature for metabolites, curation protocol, and the format of model representation. Through the extensive manual curations and continuous updates with newly generated experimental data, the consensus GEM reached the version 7.0 in 2013 [33]. Compared with the previous consensus model, the Yeast 7.0 contained the updates on the fatty acid, glycerophospholipid, glycerolipid metabolism and significantly increased number of metabolites (2,218) and reactions (3,493). However, it should be noted that the increase in the number of metabolites and reactions reflects our expanded knowledge on the details of a part of the metabolic network and does not necessarily leads to the improved prediction capability. More parallel effort needs to be exerted on developing simulation algorithms suitable for the expanded model and compartmentalization, together with physiologically relevant constraints. Nonetheless, realizing the benefits of such community-driven approach for a consensus model, concerted efforts for the construction of the consensus models for *E. coli* [62], *Salmonella typhimurium* LT2 [88], Chinese hamster ovary cell [79] and Homo sapiens [35] have been initiated and produced the early versions of the consensus GEMs.

In addition to the two model microorganisms, more than 100 GEMs covering a wide range of microorganisms, including *Bacillus subtilis* [31], *Clostridium acetobutylicum* [48], *Clostridium beijerinckii* [57], *Corynebacterium glutamicum* [80], *Lactococcus lactis* [25], *Mannheimia succiniciproducens* [46], *Pichia pastoris* [83], *Pseudomonas putida* [69], *Synechococcus* sp. [91], and others, have been reported; in 2013 alone, GEMs for more than 10 microorganisms were newly reconstructed [52, 58]. To facilitate the data-driven reconstruction of GEM, several data mining and curation tools have also been developed (Table 1). These tools can automatically generate draft GEMs from genome sequences and the bioinformatic data available in the public databases base such as KEGG and MetaCyc. Although the automatically reconstructed GEMs need to go through the extensive validation and manual curation processes, these tools would accelerate the reconstruction of new GEMs and be essential to improve the quality and completeness of the GEMs to the levels of the *E. coli* and *S. cerevisiae* GEMs.

## Prediction of gene deletion and amplification targets

An important application of GEM is the prediction of the effects of genetic perturbations on cellular metabolic phenotypes. Several different algorithms based on constraint-based FBA on GEM have been developed for the prediction of gene deletion targets for the maximal production of target product (Table 1). In constraint-based FBA, the gene knockout simulation is performed by setting the corresponding flux carried by the gene product (enzyme) to zero. Then, the prediction algorithms enable its effects on the production of the target compound and biomass formation to be systematically determined at genome-scale. The objective of maximizing the product formation rate can be different from maximizing the cell growth rate, which is probably the objective of the host organism [77]. The prediction tools differ from one another in how to apply objective functions and impose the proper constraints in gene knockout simulations, resulting in different formulation of the optimization problem and the outcome of the predictions.

The minimization of metabolic adjustment (MOMA) is one of the early algorithms developed for gene knockout simulation. MOMA is based on the hypothesis that the metabolic state of mutant is best represented by the minimal flux redistribution upon genetic perturbations with respect to the flux distribution of the wild-type strain [78]. The flux distribution of a mutant is determined by seeking the closest point in the flux space of the mutant to the optimal state of the wild-type strain and, in most of the cases, the predicted flux distribution in the mutant is sub-optimal for cell growth. MOMA was applied to identify gene deletion targets for the production of the cytosolic human superoxide dismutase (hSOD) in *Pichia pastoris* [61]. The alcohol dehydrogenase *adh2* was identified as a deletion target and the production of hSOD was increased by 20 % in the *adh2* knockout strain without compromising the cell growth. MOMA has also been successfully used to predict gene knockout targets for the enhanced production of lycopene [2], L-valine [64] and polylactic acid in *E. coli* [41]. In the study on developing an *E. coli* strain overproducing L-valine, the triple knockout of the *aceF*, *pfkA,* and *mdh* genes predicted by a sequential knockout simulation increased the L-valine production yield close to 39 g L-valine per 100 g glucose fed [64].

OptKnock is another popular gene knockout simulation algorithm that identifies a set of deletion targets that maximizes both the production of compounds of interest and cell growth using a bilevel optimization algorithm [10]. OptKnock was used to develop strategies for the metabolic engineering of *E. coli* for the production of 1,4-butanediol (BDO) and lactate [26, 94]. RobustKnock, a derivative of OptKnock, was developed to overcome the alternative optima problem of OptKnock caused by the presence of competing pathways [86]. RobustKnock searches gene deletion targets by maximizing the guaranteed minimal production of target chemical, whereas OptKnock simply searches for a set of gene deletion targets leading to the maximal production at a given biomass yield. For more specific target identification, OptSwap uses RobustKnock algorithm with an additional function of optimization of the cofactor specificities of oxidoreductases. Using *E. coli* GEM iJO1366, OptSwap proposed the strategies for the modification of cofactor specificities of oxidoreductases and gene knockout targets for the production of L-alanine, succinate, acetate,

and D-lactate [95]. Cofactor modification analysis (CMA) is another constraint-based flux balance analysis tool that finds targets for the switch of cofactor specificity to achieve improved production rate of a desired product and high cellular growth rate simultaneously. It identifies targets by monitoring changes in flux distribution patterns in response to altered cofactor specificity in metabolic network model [47].

Together with gene deletion, increasing metabolic fluxes by the overexpression of the relevant genes has been one of the most frequently employed engineering strategies for the enhanced production of desired compounds. Several strategies such as flux response analysis [84], flux scanning based on enforced objective flux (FSEOF) [15] and flux variability scanning based on enforced objective flux (FVSEOF) with grouping reaction (GR) constraints [65] have been developed to predict gene amplification targets (Table 1). Flux response analysis allows systematic examination of the effects of altering particular fluxes on the distribution of other metabolic fluxes, allowing identification of those fluxes that increase the production rate of the target product [84]. For example, the phosphoenolpyruvate carboxylase (*ppc*) was identified as an amplification target for fumaric acid overproduction using flux response analysis; the overexpression of the *ppc* gene increased the fumaric acid titer by 2.8-fold. Similarly, FSEOF allows identification of the intracellular fluxes that increased together with the enforced objective flux (e.g., product formation rate) as the gene amplification targets [15]. In a study on the metabolic engineering of *E. coli* for the production of lycopene, the overexpression of the *idi* and *mdh* genes predicted by FSEOF increased the lycopene production by 2.7-fold compared to the control strain [15]. In order to overcome the limitations of FSEOF such as alternative optima and the lack of thorough consideration of the physiological state, FVSEOF with GR constraints was developed. The GR constraints are derived from omics data and incorporated as additional constraints to reflect the physiological status of cells [65], which allowed more accurate simulation results. This algorithm was successfully employed for the identification of *glk*, *acnA*, *acnB*, *ackA*, and *ppc* genes as amplification targets; individual amplification of one of these genes resulted in the enhanced production of putrescine in *E. coli* up to 20.5 % [65].

## Prediction of target genes to be up- or down-regulated

Beyond the identification of gene deletion and amplification targets, there have been several approaches which can identify up- or down-regulation gene targets (Table 1). A gene knock-out prediction tool OptKnock was extended to OptReg [68] and OptForce [72] to identify gene targets for deletion and up-/down-regulations by comparative analysis of the possible metabolic flux range in wild-type and engineered strains. In OptForce, the flux ranges are calculated using flux variability analysis in which each flux is iteratively maximized and minimized. If the flux range of a certain pathway in the wild-type strain substantially deviates from that of production strain, the corresponding gene is predicted to be up- or down-regulation targets and the gap between the two flux ranges indicates the degree of required modification. OptForce was demonstrated to recapitulate the previously reported engineering strategies for succinic acid production, and also suggested the additional genetic manipulation strategies to be tested.

CosMos uses a similar approach, but it differs from OptForce in that it allows continuous changes in flux ranges instead of fixing them to the values determined by flux variability analysis [20]. When compared with OptForce for succinate production in *E. coli*, CosMos suggested new strategies which required fewer modifications and gave higher succinate yield. Recently, k-OptForce was developed by incorporating known kinetic information of metabolic reactions into the OptForce platform [16]. In a benchmark study on the overproduction of L-serine in *E. coli*, k-OptForce identified key regulatory bottlenecks that OptForce failed to predict and eliminated unnecessary genetic interventions predicted by OptForce [16]. k-OptForce also suggested genetic interventions leading to increased production of the target compound by alleviating the substrate-level inhibition of key enzymes, which was emphasized as an example demonstrating the benefit of integrating enzyme kinetic information into the stoichiometric model. In addition to the algorithms introduced above, several other algorithms have been developed (Table 1). The effectiveness of applying these algorithms in actual strain development studies needs to be seen in the future.

## Omics integrated genome-scale models

Omics data can provide most relevant snapshots of the biological and metabolic status of microorganisms under certain genotypic and environmental conditions, and can be used as constraints to narrow down the solution space of flux distribution (Fig. 2). GEM is a powerful platform to which various biological information can be integrated thanks to the simplicity in its formulation, no requirement for kinetic parameters, and well-established analytical methods. However, the noise, incompleteness and complexity of omics data pose significant challenges in their integration into GEM. For the efficient mapping of large omics data set into GEM, various tools have been developed (Table 1). Gene Inactivity Moderated by Metabolism and Expression (GIMME) generates a context-dependent GEM by the integration of transcriptome data [5]. In GIMME, a set of reactions corresponding to the mRNA transcript levels below the specified thresholds is temporarily set to zero

and the model is tested for a given objective function. If the model fails to achieve the cellular objective (cell growth or ATP production and etc.), a set of deleted reaction(s) is restored to meet the constraints using linear programming. Later on, GIMME was expanded to incorporate proteomic (GIMMEp) and metabolomic data (GIM$^3$E) [6, 76]. Unlike GIMME, integrative Metabolic Analysis Tool (iMAT) does not require an objective function [82]. Instead, gene expression data is first discretized (−1, 0 and 1) and mapped into GEM to divide the reactions into two groups: high and low expression subsets. Then, iMAT solves a mixed integer linear problem under the stoichiometric and thermodynamic constraints to find a steady-state flux distribution that maximizes the number of reactions the activities of which are consistent with the corresponding expression state. Its key advantage over GIMME is no requirement of a user-specified objective function that GEM is assumed to meet.

Whereas GIMME and iMAT integrate gene expression data by translating gene expression level to discrete values, E-Flux incorporates gene expression data by using them to set a continuous range of flux values [7, 18]. E-Flux constrains the upper and lower bounds of fluxes according to the expression levels of the corresponding genes. Probabilistic Regulation of Metabolism (PROM) integrates metabolic and gene expression data utilizing a large set of microarray data and pre-determined transcriptional regulatory network structure [13]. PROM takes a probabilistic approach to describe the status of metabolic reactions by analyzing the gene states and gene-transcription factor (TF) interactions. The integrated PROM model for *E. coli* was built upon 136 TFs regulating the expression of 708 metabolic genes through 1,773 interactions. By systematically analyzing the gene expression levels in relation to the state of TFs, the probabilities of the genes to be on are determined. The calculated probabilities are then, similarly to E-Flux, used to constrain the flux ranges through the reactions controlled by the target genes. Because the transcriptional regulatory network structure is incorporated into PROM, it can also be used to identify transcriptional regulatory factors to be engineered for the production of target compounds and desired complex phenotype.

In a recent study, a pool of the integrated models including GIMME, iMAT, MADE and E-Flux were evaluated with respect to their capabilities to predict the flux distributions in *E. coli* growing aerobically in batch and chemostat and *S. cerevisiae* growing in glucose-limited chemostat [54]. Interestingly, no integrated model consistently performed well under all three cases, and often, the predictions by simple constraint-based FBA were as good as or better than those produced from the integrated models. These results do not undermine the efforts for integrated GEM, but at the same time, they highlight the gaps in our knowledge on the correlation of gene expression and underlying determinants of metabolic fluxes. Once these gaps are filled by the rapidly advancing omics technologies and algorithms for GEM applications, the accuracy of GEM will be substantially improved leading to more reliable target prediction for genetic manipulations including gene deletion, up- and down-regulation.

## Prediction of novel biosynthetic pathways

The efforts to expand the repertoire of bio-based chemicals have produced computational algorithms capable of predicting synthetic routes to the production of natural and non-natural compounds (Table 1). The two key elements in pathway prediction are the reactions rules describing the patterns of the chemical transformation and heuristics ranking the predicted pathways. Biochemical network integrated computational explorer (BNICE) uses the third-level Enzyme Commission (EC) system to classify the reactions according to the similarity of the biochemical transformation pattern [29]. BNICE could account for ~50 % of the reactions in KEGG database with the 86 reaction rules. BNICE also takes into account reaction thermodynamics and the entries in BNICE are not limited to the chemicals from a specific database such as KEGG LIGAND facilitating the prediction of novel synthetic pathways. Various prediction algorithms, including the pathway prediction system developed by Cho et al. [14], RetroPath [11], PathPred [59], DESHARKY [74] and Biopathway Predictor (BioPath), have been developed using their own reaction rules and the heuristics for ranking and pathway search algorithms (Table 1) [94].

Metabolic Tinker first compiles compounds and reactions from CHEBI and Rhea database to generate the universal reaction network in the form of hyper-graph [54]. Then, Metabolic Tinker searches for synthetic metabolic pathways between two given compounds based on chemical similarity and thermodynamic feasibility using graph search heuristics. The latest web-based pathway analysis platform, XTMS presents a unique strategy of molecular signature and extended metabolic space [12]. A molecular signature containing information about the connectivity of each atom through chemical bonds up to a predetermined diameter $d$ is used to enumerate all the chemical entries and derive reaction rules. Then, XTMS extends the metabolic network of an organism such as *E. coli* to generate extended metabolic space for a given $d$. As $d$ decreases, the promiscuity of reactions increases resulting in a lager extended metabolic space. When $d = 12$ was given, XTMS extended a metabolic network consisting of 6,093 metabolites connected through 6,078 reactions (MetaCyc database) to an extended metabolic space containing 27,743

reactions. In an example, XTMS proposed natural and heterologous routes in *E. coli* for the production of raspberry ketone.

GEM-Path is a prediction algorithm that is most tightly integrated into GEM among the various algorithms reported [56]. While many algorithms suffer from the explosion of synthetic pathways and depend on the ranking heuristics for filtering infeasible pathways, GEM-Path performs pathway integration with a given GEM at each step of retrosynthetic pathway assembly. For example, the predicted intermediates are modeled, using constraint-based GEM analysis, to determine the compatibility with the metabolic context of a given host organism in terms of condition-specific cofactor utilization and regeneration, substrate utilization, oxygen requirement, energy balance, thermodynamic feasibility, and others. The pathways leading to the intermediates that could not be coupled with cell growth are filtered out. This approach was devised to efficiently reduce the search space and ensure growth-coupled production of target compounds. With respect to the reaction promiscuity, GEM-Path analyzes the similarity of native and non-native substrates based on the molecular fingerprints representing the characteristics of a particular bond pattern within a molecule and determines a promiscuity score for the pairs of substrates. If the score is below a certain value, the same reaction is assigned to both substrates. GEM-Path was tested for the growth-coupled production of 20 different chemicals under various growth conditions on different carbon sources.

## Concluding remarks

With the availability of the complete genome sequences for increasing number of organisms, the reconstruction and applications of GEM has become an essential practice in metabolic engineering. Various algorithms have been developed to predict metabolic states and characteristics of strains under genetic and environmental perturbations, and consequently to suggest metabolic engineering strategies for the enhanced production of target compounds. The reconstruction of GEM, which was extensively dependent on the manual curation of available gene annotation data for an organism, now can be done in much easily using various computational tools. Many different algorithms for the simulation of GEMs have been developed for the identification of gene knockout, gene amplification and gene up- or down-regulation targets. Much advance has been made on integrating multiple omics data with GEM with an objective of more accurate simulation of metabolic states through the consideration of regulatory information in flux analysis. The integration of the rapidly expanding omics data with GEM will expand our understanding of the operation and regulation of metabolic networks and improve the prediction capability as well. It is expected that the GEM and its simulation using various algorithms will play increasingly important roles in developing strains capable of efficiently producing chemicals, fuels and materials.

**Acknowledgments** This work was supported by the Intelligent Synthetic Biology Center through the Global Frontier Project (2011-0031963) of the Ministry of Science, ICT & Future Planning through the National Research Foundation of Korea.

## References

1. Agren R, Liu LM, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLoS Comput Biol 9(3):e1002980
2. Alper H, Jin YS, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab Eng 7(3):155–164
3. Anna SB, Jason AP (2012) Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 3:299
4. Bates JT, Chivian D, Arkin AP (2011) GLAMM: genome-linked application for metabolic maps. Nucleic Acids Res 39:W400–W405
5. Becker SA, Palsson BØ (2008) Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 4(5):e1000082
6. Bordbar A, Mo ML, Nakayasu ES, Schrimpe-Rutledge AC, Kim YM, Metz TO, Jones MB, Frank BC, Smith RD, Peterson SN et al (2012) Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. Mol Syst Biol 8:558
7. Brandes A, Lun DS, Ip K, Zucker J, Colijn C, Weiner B, Galagan JE (2012) Inferring carbon sources from gene expression profiles using metabolic flux models. PLoS One 7(5):e36947
8. Brochado AR, Matos C, Moller BL, Hansen J, Mortensen UH, Patil KR (2010) Improved vanillin production in baker's yeast through *in silico* design. Microb Cell Fact 9:84
9. Buchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, Keller R, Mittag F, Schubert M, Glont M, Golebiewski M et al (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. BMC Syst Biol 7:116
10. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84(6):647–657
11. Carbonell P, Planson AG, Fichera D, Faulon JL (2011) A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Syst Biol 5:112
12. Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon JL (2014) XTMS: pathway design in an extended metabolic space. Nucleic Acids Res. doi:101093/nar/gku362
13. Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 107(41):17845–17850
14. Cho A, Yun H, Park JH, Lee SY, Park S (2010) Prediction of novel synthetic pathways for the production of desired chemicals. BMC Syst Biol 4:35
15. Choi HS, Lee SY, Kim TY, Woo HM (2010) *In silico* identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097–3105

16. Chowdhury A, Zomorrodi AR, Maranas CD, Beard DA (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. PLoS Comput Biol 10(2):e1003487
17. Chung BKS, Lee DY (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. BMC Syst Biol 3:117
18. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput Biol 5(8):e1000489
19. Collins SB, Reznik E, Segre D (2012) Temporal expression-based analysis of metabolism. PLoS Comput Biol 8(11):e1002781
20. Cotten C, Reed JL (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. Biotechnol J 8(5):595–604
21. Dave L, Kieran S, Warwick BD, Ettore M, Catherine LW, Douglas BK, Pedro M, Neil S (2012) Improving metabolic flux predictions using absolute gene expression data. BMC Syst Biol 6:73
22. Dikicioglu D, Pir P, Onsan ZI, Ulgen KO, Kirdar B, Oliver SG (2008) Integration of metabolic modeling and phenotypic data in evaluation and improvement of ethanol production using respiration-deficient mutants of *Saccharomyces cerevisiae*. Appl Environ Microb 74(18):5809–5816
23. Edwards JS, Palsson BØ (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 97(10):5528–5533
24. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (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
25. Flahaut NA, Wiersma A, van de Bunt B, Martens DE, Schaap PJ, Sijtsma L, Dos Santos VA, de Vos WM (2013) Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. Appl Microbiol Biotechnol 97(19):8729–8739
26. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BØ (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 91(5):643–648
27. Forster J, Famili I, Fu P, Palsson BØ, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 13(2):244–253
28. Hadicke O, Klamt S (2010) CASOP: a computational approach for strain optimization aiming at high productivity. J Biotechnol 147(2):88–101
29. Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ (2005) Exploring the diversity of complex metabolic networks. Bioinformatics 21(8):1603–1609
30. Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP (2012) Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. BMC Syst Biol 6:55
31. Henry CS, Zinner JF, Cohoon MP, Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. Genome Biol 10(6):R69
32. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28(9):977–982
33. Hnin WA, Susan AH, Larry PW (2013) Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. Ind Biotechnol 9(4):215–228
34. Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale models of metabolism. Mol BioSyst 9(2):167–174
35. Ines T, Neil S, Ronan MTF, Andreas H, Swagatika S, Maike KA, Hulda H, Monica LM, Ottar R, Miranda DS et al (2013) A community-driven global reconstruction of human metabolism. Nat Biotechnol 31(5):419–425
36. Jan S, Richard Q, Ronan MTF, Ines T, Jeffrey DO, Adam MF, Daniel CZ, Aarash B, Nathan EL, Sorena R et al (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6(9):1290–1307
37. Jensen PA, Lutz KA, Papin JA (2011) TIGER: toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst Biol 5:147
38. Jensen PA, Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 27(4):541–547
39. Jerby L, Shlomi T, Ruppin E (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol Syst Biol 6:401
40. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci USA 110(50):20338–20343
41. Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol Bioeng 105(1):161–171
42. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L et al (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform 11(1):40–79
43. Kim HU, Kim WJ, Lee SY (2013) Flux-coupled genes and their use in metabolic flux analysis. Biotechnol J 8(9):1035–1042
44. Kim J, Reed JL (2010) OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 4:53
45. Kim J, Reed JL (2012) RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol 13(9):R78
46. Kim TY, Kim HU, Park JM, Song H, Kim JS, Lee SY (2007) Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. Biotechnol Bioeng 97(4):657–671
47. Lakshmanan M, Chung BKS, Liu CC, Kim SW, Lee DY (2013) Cofactor modification analysis: a computational framework to identify cofactor specificity engineering targets for strain improvement. J Bioinf Comput Biol 11(6):1343006
48. Lee J, Yun H, Feist AM, Palsson BØ, Lee SY (2008) Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. Appl Microbiol Biotechnol 80(5):849–862
49. Lee KY, Park JM, Kim TY, Yun H, Lee SY (2010) The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. Microb Cell Fact 9:94
50. Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimpe-Rutledge AC, Smith RD, Adkins JN, Zengler K et al (2012) *In silico* method for modelling metabolism and gene product expression at genome scale. Nat Commun 3:929
51. Lerman JA, Chang RL, Hyduke DR (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 9:693
52. Lewis NE, Nagarajan H, Palsson BØ (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. Nat Rev Microbiol 10(4):291–305
53. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM (2009) Large-scale identification of genetic design strategies using local search. Mol Syst Biol 5:296

54. Machado D, Herrgard M (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol 10(4):e1003580
55. Melzer G, Esfandabadi ME, 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 3:120
56. Campodonico MA, Andrews BA, Asenjo JA, Palsson BØ, Feist AM (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. Metab Eng 25:140–158
57. Milne CB, Eddy JA, Raju R, Ardekani S, Kim P-J, Senger RS, Jin Y-S, Blaschek HP, Price ND (2011) Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Syst Biol 5:130
58. Monk J, Nogales J, Palsson BØ (2014) Optimizing genome-scale network reconstructions. Nat Biotechnol 32(5):447–452
59. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S, Kanehisa M (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res 38:W138–W143
60. Navid A, Almaas E (2012) Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. BMC Syst Biol 6:150
61. Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Russmayer H, Pflugl S, Ask M, Haberhauer-Troyer C et al (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. Metab Eng 24:129–138
62. Orth JD, Conrad TM, Na J, Lerman JA (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. Mol Syst Biol 7:535
63. Österlund T, Nookaew I, Bordel S (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Syst Biol 7:36
64. Park JH, Lee KH, Kim TY, Lee SY (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 USA 104(19):7797–7802
65. Park JM, Park HM, Kim WJ, Kim HU, Kim TY, Lee SY (2012) Flux variability scanning based on enforced objective flux for identifying gene amplification targets. BMC Syst Biol 6:106
66. Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinform 6:308
67. Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367–2376
68. Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8(1):1–13
69. Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VA (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol 4:10
70. Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform 10(4):435–449
71. Ranganathan S, Maranas CD (2010) Microbial 1-butanol production: identification of non-native production routes and *in silico* engineering interventions. Biotechnol J 5(7):716–725
72. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6(4):e1000744
73. Rocha I, Maia P, Evangelista P, Vilaca P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst Biol 4:45
74. Rodrigo G, Carrera J, Prather KJ, Jaramillo A (2008) DESHARKY: automatic design of metabolic pathways for optimal cell growth. Bioinformatics 24(21):2554–2556
75. Rossell S, Huynen MA, Notebaart RA (2013) Inferring metabolic states in uncharacterized environments using gene-expression measurements. PLoS Comput Biol 9(3):e1002988
76. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR (2013) GIM$^3$E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics 29(22):2900–2908
77. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. Science 336(6081):601–604
78. Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99(23):15112–15117
79. Selvarasu S, Ho YS, Chong WPK, Wong NSC, Yusufi FNK, Lee YY, Yap MGS, Lee D-Y (2012) Combined *in silico* modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnol Bioeng 109(6):1415–1429
80. Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H (2009) Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. Microb Cell Fact 8:43
81. Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci USA 102(21):7695–7700
82. Shlomi T, Cabili MN, Herrgard MJ, Palsson BØ, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotechnol 26(9):1003–1010
83. Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for *in silico* analysis of heterologous protein production. Biotechnol J 5(7):705–715
84. Song CW, Kim DI, Choi S, Jang JW, Lee SY (2013) Metabolic engineering of *Escherichia coli* for the production of fumaric acid. Biotechnol Bioeng 110(7):2025–2034
85. Swainston N, Smallbone K, Mendes P, Kell DB, Paton NW (2011) The SuBliMinaL toolbox: automating steps in the reconstruction of metabolic networks. J Integr Bioinform 8:186
86. Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26(4):536–543
87. Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5(1):93–121
88. Thiele I, Hyduke DR, Steeb B, Fankam G, Allen DK, Bazzani S, Charusanti P, Chen F-CC, Fleming RM, Hsiung CA et al (2011) A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella Typhimurium* LT2. BMC Syst Biol 5:8
89. Topfer N, Jozefczuk S, Nikoloski Z (2012) Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. BMC Syst Biol 6:148
90. van Berlo RJP, de Ridder D, Daran JM, Daran-Lapujade PAS, Teusink B, Reinders MJT (2011) Predicting metabolic fluxes using gene expression differences as constraints. IEEE ACM Trans Comput Biol 8(1):206–216

91. Vu TT, Hill EA, Kucek LA, Konopka AE, Beliaev AS, Reed JL (2013) Computational evaluation of *Synechococcus sp*. PCC 7002 metabolism for chemical production. Biotechnol J 8(5):619–630
92. Wang YL, Eddy JA, Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst Biol 6:153
93. Yang L, Cluett WR, Mahadevan R (2011) EMILiO: a fast algorithm for genome-scale strain design. Metab Eng 13(3):272–281
94. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R et al (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 7(7):445–452
95. Zachary AK, Adam MF (2013) Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. Ind Biotechnol 9(4):236–246