10

CONSTRUCTION AND APPLICATIONS OF GENOME-SCALE *IN SILICO* METABOLIC MODELS FOR STRAIN IMPROVEMENT

Sang Yup Lee[1,2], Jin Sik Kim[1], Hongseok Yun[1], Tae Yong Kim[1], Seung Bum Sohn[1], and Hyun Uk Kim[1]

[1]*Department of Chemical and Biomolecular Engineering (BK21 Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, KAIST, Daejeon 305-701, Korea*

[2]*Department of Bio and Brain Engineering, BioProcess Engineering Research Center and Bioinformatics Research Center, KAIST, Daejeon 305-701, Korea*

10.1 INTRODUCTION TO SYSTEMS BIOTECHNOLOGY

Since the first genome sequence of a microorganism was finished in 1995, a number of projects for sequencing microbial genomes have been completed [1]. Currently, the complete sequences of more than 300 genomes are available in various databases [2]. The processes of sequencing and annotating microbial genomes have now become more routine, which resulted in the continued introduction of complete genome sequences of new microorganisms to the life science and biotechnology community. In addition, breakthroughs in studying biological systems at transcriptomic, proteomic, and other omic levels have enabled the researchers to generate and analyze high-throughput data

Systems Biology and Synthetic Biology Edited by Pengcheng Fu and Sven Panke

for the better characterization of the organisms of interest [3]. Furthermore, computational (*in silico*) tools for modeling and simulation of biological systems on large or genome scale have been developed and used for deciphering the characteristics of metabolic, regulatory, and signaling networks [4]. With such advances in experimental and computational techniques, microorganisms can be systematically engineered to be suitable for various industrial applications that fall into a new paradigm of research called ''systems biotechnology'' [3].

Systems biotechnology aims at improving the biotechnological processes by systems-level optimization of cellular metabolism, regulations and signaling circuits, and mid- to down-stream processes altogether [3]. Understanding basic genotype–phenotype relationship in an organism is important, but it is not sufficient to understand and control the entire behavior of the organism. For this reason, high-throughput technologies have been indispensable tools as they allow the expression of genes to be monitored on global scale at transcriptional and translational level. One of the high-throughput techniques that has helped make this progress is transcriptomics, which allows the analysis of mRNA expression levels of the entire genes using DNA microarray. Proteomics allows analysis of the protein contents in an organism or a given sample. Metabolomics and fluxomics, which quantitatively profile the metabolites and fluxes, respectively, in the cell, also occupy an important portion of the omics research to carry out systems biotechnology research. By combining all the information generated from these omics disciplines, it will be possible to model an organism at the systems level (although not complete yet) and perform a systematic analysis of large-scale data using bioinformatics for a better understanding of how that system works and how it can be best adjusted for our applications [5,6].

Analysis of the *in silico* metabolic network can be used as a powerful approach for the identification of drug targets and targets for the improvement of microbial performance suitable for industrial applications such as production of useful materials [7–9]. *In silico* model is a mathematical representation of the biological system in interest and allows researchers to perform experiments on a computer to predict physiological behaviors much faster and economically than the actual experiments. Recently, various approaches for the construction of reliable metabolic network model have been suggested [1,10].

In this chapter, we describe the recent developments and trends in systems biotechnology research based on the *in silico* genome-scale metabolic models. Various strategies are described for the reconstruction of genome-scale metabolic network. Thereafter, we will review their applications with specific examples from the metabolic engineering perspectives. Readers are recommended to read Chapter 7 in parallel, which presents the state-of-the-art review on building the constraints-based metabolic models and their use in flux balance analysis (FBA).

10.2 DATABASES AND TOOLS FOR THE RECONSTRUCTION OF METABOLIC NETWORKS

From the last decade, unprecedentedly large amounts of information have been accumulated from experiments in genomics and other omics research projects.

As a result, many different databases and related applications have been developed for researchers to use to extract suitable information for the analysis of pathways and the reconstruction of genome-scale metabolic networks. The databases and applications commonly used for the systems biotechnology research are listed in Table 10-1. These databases are mainly used for the retrieval and analysis of sequences, protein analysis, functional annotation of genes and sequences, metabolic pathways, and other information needed for the reconstruction of metabolic networks. Here we shall focus on the effective construction and analysis of the *in silico* genome-scale metabolic networks using the information present in the databases.

Databases such as DDBJ, EMBL, and NCBI Entrez contain information regarding the DNA, RNA, and protein sequences and other related information [11–13]. Along with these databases, the controlled vocabularies, such as Gene Ontology (GO), are used for standardizing the results of genome annotations. Other databases contain information on metabolic networks such as reactions and network maps, tools for comparative analysis, and various information on enzymes, metabolites, and other biomolecules. For example, the automatic annotation tools import raw genome sequences and find proper open reading frames (ORFs) and gene candidates by applying gene finding algorithms. The databases for protein profiles and motifs are very helpful in enhancing the quality of genome annotation and in predicting the detailed functions of proteins by taking advantage of the conserved domains found in the proteins [14–16].

Reconstruction of metabolic pathways is mostly based on the information from metabolic databases [17–21]. Most of these databases provide graphical references or metabolic maps for users to find the metabolic information such as gene names, enzyme commission (EC) numbers, and reactions that are highly interlinked within the frame of metabolic network. KEGG is one of the most widely used metabolic resources and provides various data on the genomes, pathways, compounds, and controlled vocabularies. The pathway maps supplied by KEGG can be used as a backbone for the reconstruction of the networks. The BioSilico database integrates components of heterogeneous metabolic databases such as LIGAND, ENZYME, and BioCyc for easy querying and comparison of metabolic information present in multiple databases [18].

10.3 *IN SILICO* MODELING AND SIMULATION OF GENOME-SCALE METABOLIC NETWORK

The first step for the reconstruction of genome-scale metabolic model is the analysis of genome information in the databases. The availability of the annotation results from the completely sequenced genomes for many organisms makes it possible to reconstruct the *in silico* models on a genome scale. Thus, the automatic annotation process, which uses the reference databases and relevant information to identify potential ORFs, is the first step for the reconstruction of *in silico* metabolic model. However, as many shortcomings become obvious in the reconstruction process [22], the automatic annotation process appears to be insufficient and various complementary processes

Table 10-1 Databases and tools useful for the reconstruction of genome-scale metabolic network

Database	Availability	Brief Description
Resources of sequences and genomic information		
DDBJ [11]	http://www.ddbj.nig.ac.jp/	DNA Database of Japan
EMBL [12]	http://www.ebi.ac.uk/embl/	Europe's primary nucleotide sequence resource
Entrez [13]	http://www.ncbi.nlm.nih.gov/sites/gquery	The integrated, text-based search and retrieval system used at NCBI
COG [30]	http://www.ncbi.nlm.nih.gov/COG/	Clusters of Orthologous Groups
Controlled vocabularies and ontology		
GO [86]	http://www.geneontology.org/	A controlled vocabulary to describe gene and gene product attributes in any organism
KO [20]	http://www.genome.jp/kegg/ko.html	KEGG Orthology
Protein sequences, motifs, and profiles		
InterPro [14]	http://www.ebi.ac.uk/interpro/	A database of protein families, domains, and functional sites
PROSITE [16]	http://www.expasy.org/prosite/	A database of protein families and domains
Metabolic databases and tools		
BioCyc [17]	http://biocyc.org/	A collection of pathway/genome databases
BioSilico [18]	http://biosilico.kaist.ac.kr/	Integrated metabolic databases
BRENDA [19]	http://www.brenda-enjymes.info/	The comprehensive enzyme information system
KEGG [20]	http://www.genome.ad.jp/kegg/	Kyoto Encyclopedia of Genes and Genomes
Pathway tools [25]	http://bioinformatics.ai.sri.com/ptools/	A software system for pathway analysis of genomes and for creating Pathway/Genome Databases (PGDBs)
PATIKA [21]	http://www.patika.org/	Pathway Analysis Tools for Integration and Knowledge Acquisition
Gene annotation and comparative genomics tools		
Glimmer [88]	http://www.cbcb.umd.edu/software/glimmer/	A system for finding genes in microbial DNA, especially the genomes of bacteria and archaea.

Table 10-1 (*Continued*)

Database	Availability	Brief Description
ERGO [90]	http://ergo.integratedgenomics.com/	Accommodation of data integration, providing the tools to support comparative analysis of genomes
STRING [35]	http://string.embl.de/	Search tool for the retrieval of interacting genes/proteins

are required for the validation of the constructed metabolic models. Recently developed genome-scale *in silico* metabolic models are listed in Table 10-2.

10.3.1 Reconstruction Using the Known Pathways and Enzymes

The common method for the reconstruction of genome-scale metabolic network has been the utilization of information obtained from the previously constructed biochemical pathways, related sequences, and proteins [1,10,23]. Such information is mostly derived from the sequence-based search. The major advantage of metabolic reconstruction using the sequence-based comparison is that proper function of the genes can be quickly assigned. However, the presence of multiple relationships between genes and metabolic reactions can cause an erroneous assignment of genes on the metabolic map [9]. For example, imprecise annotations may occur for the homologues within the metabolic network, which hampers the accurate assignment of specific metabolic functions to the ORFs. Therefore, advanced curating methods have been introduced to eliminate the limitation of sequence-based annotation method [10].

Currently, a number of databases and tools have been developed for systems biotechnology research. Among them, several resources have been developed to represent the biochemical reactions and pathways on a two-dimensional space. Representative resources are KEGG [20] and BioCyc [17,24]. These are the most easily accessible and widely used databases on genes, enzymes, metabolites, and biochemical reactions. In addition to these tools, numerous databases and tools for the analysis of metabolic pathways have been released (Table 10-1). The utilization of these resources helps to gather the information on biochemical reactions and their location on the metabolic map. For example, the PathoLogic software, part of Pathway Tools that also contains MetaCyc database, automatically reconstructs the metabolic pathways of any organism only if the annotation file is available as an input [25]. The core algorithm in this process is that the software matches the enzyme in the annotation file (input file) to the ones defined in the MetaCyc database by EC number or enzyme name. Then, the software graphically displays the metabolic pathways and the associated components including reactions, enzymes, substrates, and products. The initial version of the automatically reconstructed metabolic network can be used as a basic framework and can be upgraded by manual curation.

Table 10-2 Recently developed genome-scale *in silico* models

Organism	Year	Genome Size (kbp)	Metabolites (ea)	Reactions (ea)	Referred Strains or Species for Biomass Composition
Corynebacterium glutamicum [91]	2009	3,309	411	446	*C. glutamicum* ATCC 13032, *C. glutamicum* ATCC17965, *C. glutamicum* CGL2005, *C. glutamicum* CGL2022, *E. coli*
Pseudomonas putida iJP815 [92]	2008	6,182	886	877	*E. coli*
P. putida iJN746 [93]	2008	6,182	710	950	*E. coli*, *P. putida*
Clostridium acetobutylicum [94]	2008	4,132	422	552	*S. aureus* 292, *S. aureus* 49/1974, *S. aureus* 6571, S. *aureus* 8325, S. *aureus* ATCC 12600, *S. aureus* DSM 20233, *S. aureus* Duncan, *S. aureus* H1AA, *Staphylococcus aureus* NCTC, *S. aureus* U-71, *S. aureus* oxford 209p *E. coli*
C. acetobutylicum [95]	2008	4,132	479	502	*C. acetobutylicum* ATCC 824, *B. subtilis*
Acinetobacter baylyi [96]	2008	3,583	701	875	*A. calcoaceticus*, *Micrococcus cerificans* HO1-N, *A. sp.* MJT/F5/199A
S. cerevisiae iIN800 [44]	2008	12,069	1013	1446	*S. cerevisiae*
P. aeruginosa iMO1056 [97]	2008	6,264	760	883	*P. aeruginosa* PAO1, *E. coli*
Rhizobium etli iOR363 [98]	2008	6,159	371	387	No biomass equation
Neisseria meningitidis [99]	2007	2,272	471	496	*N. meningitidis* HB-1
E. coli iAF1260 [65]	2007	4,639	1039	2077	*E. coli*
Bacillus subtilis iYO844 [100]	2007	4,214	988	1020	*B. subtilis* RB50:pRF69
Mycobacterium tuberculosis iJN661 [101]	2007	4,412	828	939	*M. tuberculosis* H37Rv, *M. bovis*
M tuberculosis [102]	2007	4,412	739	849	*M. tuberculosis* H37Rv, *M. bovis* BCG
Geobacter sulfurreducens [103]	2006	3,814	541	523	*G. sulfurreducens* ATCC 51573, *E. coli*
M. succiniproducens MBEL55E [104]	2005	2,314	519	686	*M. succiniproducens*, *E. coli*

L. plantarum WCFS1 [38]	2005	3,308	658	762	*L. plantarum* WCFS1
S. aureus N315 [42]	2005	2,813	712	774	*S. aureus* 292, *S. aureus* 49/1974, *S. aureus* 6571, S. *aureus* 8325, S. *aureus* ATCC 12600, *S. aureus* DSM 20233, *S. aureus* Duncan, *S. aureus* H1AA, *Staphylococcus aureus* NCTC, *S. aureus* U-71, *S. aureus* oxford 209p *E. coli*
S. aureus N315 [45]	2005	2,813	571	641	*B. subtilis*
Lactococcus lactis [46]	2005	2,365	422	621	*L. lactis* NCDO 2118, *L. lactis* subsp *cremoris*, *L. lactis* subsp. *cremoris* NCDO763. *E. coli*
Helicobacter pylori [75]	2005	1,668	485	476	*H. pylori* NCTC 11638, *E. coli*
S. coelicolor A3(2) [22]	2005	8,667	500	971	*S. antibioticus* RIA-594, *S. clavuligerus*, *S. antibioticus*, *S. chrysomallus*, *S. roseoflavus*, *S. roseoflavus* var. *Roseofungini*, *E. Coli*, *S. typhimurium*
M. succiniproducens MBEL55E [59]	2004	2,314	352	373	*M. succiniproducens*, *E. Coli*
S. cerevisiae iND750 [105]	2004	12,069	646	1149	*S. cerevisiae*
S. cerevisiae iFF708 [77]	2003	12,069	584	842	*S. cerevisiae*
E. coli K-12 iJR904 GSM/GPR [106]	2003	4,639	625	931	*E. coli*
H. pylori [107]	2002	1,668	340	388	*H. pylori* NCTC 11638, *E. coli*
E. coli K-12 iJE660a GSM [43]	2000	4,639	438	627	*E. coli*

10.3.2 Reconstruction Using Controlled Vocabulary

The interactions among the molecules in the metabolic and regulatory networks are known to be highly complex and incompletely understood [21]. To understand this, abstractions on different levels are used to analyze the cellular processes more effectively and to deal with the complex network structure more easily. The abstractions can be utilized to construct and analyze the graphical representation of metabolic pathways [26].

Ontologies for the standardization of the vocabularies were used for the automatic annotation analyses in many projects. The sequence similarity can be directly related to the potential protein functions by utilizing the ontology [27]. However, the limitation of gene ontology is that it cannot be directly connected to cellular metabolism. This is compensated by the application of metabolism-based orthology concept such as KEGG orthology to the annotation process [28]. When the proper KEGG orthology term can be assigned to a gene, the associated metabolic pathways can be found by tracing back the hierarchical structure of KEGG orthology [28].

10.3.3 Completion of Reconstruction Using Phylogenetic Profiles and Contexts

As the amount of sequence data increases explosively, the noise in the data also increases; accumulation of incomplete and/or wrong sequences causes obvious problems during bioinformatic analyses [29]. Annotation and analysis based only on these resources can generate wrong results and result in incorrect interpretations. This limitation can be overcome by employing controlled vocabulary and large-scale phylogenetic trees. Bacteria share many functional components with a high degree of conservation in the components. As mentioned in Section 10.3.1, Clusters of Orthologous Groups (COGs) use the grouping of previously annotated genes based on the sequence homology [30]. There are many ways to construct phylogenetic trees [29,31]. Different from the sequence-based analysis, the genome-scale phylogenetic profiles use various components of the genome such as the metabolic profiles and the distribution of gene contents [32]. Especially, the highly conserved components such as transporters and proteins involved in signaling and carbon source utilization can be used to find the proper orthologous genes [33,34]. The molecular interactions and network can be identified by using databases for protein–protein interaction and metabolic context such as STRING [35]. Similar to STRING, the SEED genome annotation system is based on the fundamental principle that the value of genome analysis increases with the number of genomes available as a context for comparative analysis [36].

Various bioinformatic methods, such as genome context analysis that includes chromosomal gene clustering, protein fusions, occurrence profiles, and shared regulatory sites, can be employed to obtain further information [37]. For example, a draft *in silico* metabolic model of *Lactobacillus plantarum* showed that succinyl-CoA is involved in a reaction related to methionine biosynthesis. However, after the phylogenetic studies and pathway analysis of the *L. plantarum* metabolic network, it was concluded that succinyl-CoA is not produced due to the operation of a branched tricarboxylic acid (TCA) cycle and that the actual substrate is most likely acetyl-CoA [38].

In addition to these methods, several integrated programs have been developed. The recent version of PathoLogic provides the function called "Pathway Hole Filler," which employs genome context analysis to fill in missing genes using the candidate sequences from the database, and subsequently a Bayes classifier to evaluate the probability of how likely the candidate has the desired function for the missing genes in the newly reconstructed metabolic network [39].

10.3.4 Completion of Reconstruction Using the Information from Various Sources

When the metabolic network reconstruction is complete, it should be able to describe and predict various phenotypic characteristics of the organism reasonably well under different genotypic and environmental conditions. However, some metabolic data are missing, inconsistent and insufficient to fully represent the physiology of a particular organism. In particular, reaction reversibility, substrate specificity, isoenzyme functions, cofactor specificity, and absence of certain pathways can make reconstruction process difficult.

Updated and new knowledge on the metabolic pathways and their components can be obtained by a thorough examination of literature. For example, the initial reconstruction of metabolic model of *Streptomyces coelicolor* A3(2) suggested that valine dehydrogenase (E.C. 1.4.1.8) is an NADP-dependent enzyme. However, after thorough examination of literature, it was found to use NAD as the preferred cofactor [22,40]. In the case of *Staphylococcus aureus* N315, literature indicates that acetate can be transported by acetate permease [41]. This transport reaction was then added to the reconstruction model to allow proper representation of observed physiological behavior *in vivo* [42].

When all the possible inconsistencies are considered, the reconstructed model should be validated and tested to see whether mathematical methods, such as convex analysis and linear programming, can effectively represent the physiology of the organism under the various genetic and environmental conditions. If the results reasonably represent what are observed in actual experiments, the reconstruction of genome-scale metabolic model is said to be done. However, it should be emphasized that metabolic reconstruction is not truly complete but has to be upgraded continuously as new information and knowledge on metabolic pathways and their participating components are discovered.

10.3.5 Simulation of Genome-Scale *In Silico* Metabolic Network

Once the genome-scale metabolic network is constructed from the genomic and other related information, computer-based experiments such as quantitative flux analysis, network topology analysis, and simulation can be performed to characterize the metabolic network under various conditions. There are two main strategies of quantitative *in silico* simulation of metabolic systems: static analysis and dynamic analysis.

Metabolic flux analysis (MFA), which utilizes stoichiometric matrices, has been employed for the large-scale analyses of metabolism (see Box 10-1). MFA calculates the intracellular flux distribution with an assumption of steady-state condition and does

BOX 10-1 VARIOUS MODELING APPROACHES

Metabolic Flux Analysis

MFA is a mathematical analysis of metabolic pathways in which metabolic fluxes are calculated by constructing a stoichiometric model of the biochemical reactions along with mass balances on intracellular metabolites [85]. Given a metabolic system, the mass conservation around metabolites can be expressed as

$$\frac{dc}{dt} = S \cdot v - b$$

where c is the concentration vector of metabolites, S is the $m \times n$ stoichiometric matrix in which m is the number of metabolites and n is the number of reactions, and v is the n-dimensional vector of intracellular fluxes. b is the concentration vector of metabolites that are diluted owing to biomass growth. Assuming the pseudo-steady or stationary state based on rapid turnover of most metabolites and dilution effects that are relatively small compared with the fluxes, we can simplify the kinetic model into a static representation. Unlike the dynamic approach, static model only considers the network's connectivity and capacity as time-invariant properties of the metabolic system.

$$S \cdot v = 0$$

The metabolic network can be classified as determined, overdetermined, and underdetermined systems if the degrees of freedom are zero, negative, and positive, respectively.

Different approaches are undertaken depending on the degree of freedom of the system. In general, two general methodologies of MFA have been practiced most widely: isotopomer balance analysis and flux balance analysis. The notable difference between these two methods is that the former is usually employed for overdetermined system whereas the latter is applicable to underdetermined system. For isotopomer balance analysis, ^{13}C carbon labeling measurements produce the flux data that can help solve the overdetermined system; it has been shown that the combination of information gathered from such isotopomer measurements using NMR and GC/MS and metabolite balancing enabled refined analysis of the metabolic fluxes. However, it should be mentioned that isotopomer analysis has so far been used for the analysis of small-scale metabolic networks because of the complicated mathematical formulation and limited availability of parameters. FBA allows determination of intracellular fluxes even for a large underdetermined system through linear optimization. Even though the accuracy of FBA can be thought as not as good as that achievable with isotopomer analysis, it generally gives satisfactory flux distribution under various genotypic and environmental conditions. Many successful examples are available in the literature, which report the use of FBA in various applications (see the text and Chapter 7).

Minimization of Metabolic Adjustment

MOMA [69] is based on the same constraints as the FBA. However, quadratic programming (QP) is used instead of linear programming to formalize the MOMA. The goal is to minimize the Euclidian distance from a wild-type flux distribution as follows:

$$\begin{aligned}&\text{Minimize } (v-w)^T(v-w)\\&\text{Subject to } S\cdot v=0, v_{\min}\le v\le v_{\max}\\&v_j=0,\quad j\in R\end{aligned}$$

where w is the wild-type flux distribution and R is a set of reactions related to the deleted genes.

Regulatory On/Off Minimization

ROOM [70] is based on the same constraints as FBA. The goal is to minimize the number of significant flux changes. A range $[w^l, w^u]$ around the vector w is defined for nonsignificant flux change. The mixed integer linear programming (MILP) can be formulated as

$$\begin{aligned}&\text{Minimize } \sum_{i=1}^{m} y_i\\&\text{Subject to } S\cdot v=0\\&v-y(v_{\max}-w^u)\le w^u\\&v-y(v_{\min}-w^l)\ge w^l\\&v_j=0,\quad j\in R,\quad y_i\in\{0,1\}\\&w^u=w+\delta|w|+\varepsilon, w^l=w-\delta|w|-\varepsilon\end{aligned}$$

where, for each flux i, $1\le i\le m$, $y_i=1$ for a significant flux change in v_i, and $y_i=0$ otherwise.

Optknock

The bilevel optimization framework, OptKnock, was introduced to propose reactions to be eliminated from the *E. coli* network for maximizing the production of simple compounds such as succinate, lactate, and 1,3-propanediol [71]. This is accomplished by calculating solutions that simultaneously optimize two objective functions, biomass formation and secretion of a target biochemical. This bilevel optimization algorithm is based on the fact that the overproduction of target biochemical can be achieved by altering the structure of the metabolic network through gene deletion such that the stoichiometry of the perturbed network forces production of the target metabolite while normal

biomass precursors are generated.

$$\begin{aligned}
&\text{Maximize } v_{\text{biochemical}}(\text{over } y_j)\\
&\text{Maximize } v_{\text{biomass}}(\text{over } v_j)\\
&\text{subject to } \sum_{j=1}^{M} S_{ij} v_j = 0, \quad \forall i \in N\\
&v_{\text{pts}} + v_{\text{glk}} = v_{\text{glucose uptake}}\\
&v_{\text{ATP}} \geq v_{\text{ATP maintenance}} \qquad \theta\\
&v_{\text{biomass}} \geq v_{\text{biomass}}^{\text{target}}\\
&v_j^{\min} y_j \leq v_j \leq v_j^{\max} y_j, \quad \forall j \in M\\
&\sum_{j=M} (1 - y_j) \leq K\\
&y_j \in \{0, 1\}, \quad \forall j \in M
\end{aligned}$$

where S_{ij} is the coefficient of metabolite i in reaction j, biomass formation is quantified as an aggregate reaction flux, v_{biomass}, draining biomass components in their appropriate biological ratios, and $v_{\text{ATP maintenance}}$ is the non-growth-associated minimum ATP requirement. The uptake rate of glucose $v_{\text{glucose\_uptake}}$ is fixed and encompasses both the phosphotransferase system, v_{pts}, and glucokinase reaction, v_{glk}. K is the number of allowable reactions to be eliminated. Binary variable, y_j, is one if a particular reaction is active, and zero otherwise. An active reaction has an upper bound, $v_j^{\max}$, and a lower bound, $v_j^{\min}$, obtained by maximizing and minimizing each flux subject to the constraints.

not require rate equations and kinetic parameters. The result is a flux map showing the distribution of anabolic and catabolic fluxes within the metabolic network. Among the various applications of MFA, two general ones are as follows. The first application field is to characterize the cell's physiology under genetic and environmental perturbations. MFA has been used to characterize the effects of acute metabolic perturbation, especially, gene deletion in the organism. It was also performed under the combination of rich and minimal media and aerobic and anaerobic conditions to predict which reactions are essential for the growth of the organism under these conditions [4,42–46]. The second application field is to improve the production of various products including commodity chemicals by overexpression/deletion of key metabolic pathways that already exist in the host organism or by introducing new routes of metabolism. Of course, MFA is used to identify the candidate target genes to be manipulated.

The MFA solution provides a snapshot of a certain pathway in a defined state, but is insufficient to predict the dynamic behavior of metabolism. Recently, this approach was extended to allow the prediction of dynamic behavior. Dynamic simulation of genome-scale network model can be performed using the differential equations

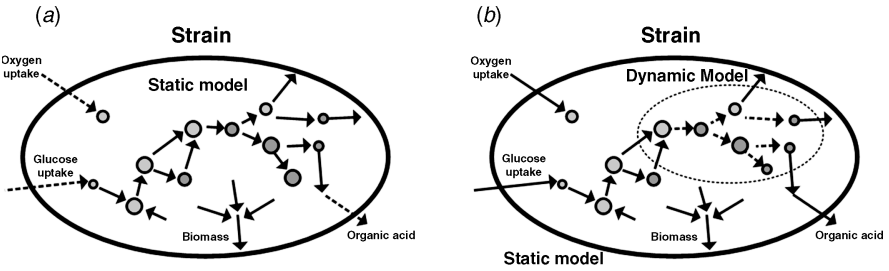

Figure 10-1 Concepts of dynamic flux balance analysis (a) and hybrid dynamic/static simulation (b). Fluxes represented by dashed arrows are given by kinetic equations.

representing the dynamic mass balances incorporating the reaction rate equations [47]. One of the major difficulties in the dynamic simulation of metabolic network is the lack of accurate kinetic equations and parameters for the reactions in the metabolic network. The parameters also tend to vary as the environmental conditions change. Therefore, the dynamic simulation of large-scale network requires many assumptions and is generally restricted to the small-scale network model. However, much effort has been devoted to solve this problem; a static simulation method was combined with a dynamic method, which is called dynamic flux balance analysis (DFBA) [48]. This method was developed to incorporate extracellular metabolite dynamics and substrate uptake kinetics within the flux balance analysis for extracellular glucose, acetate, and liquid- and gas-phase oxygen. Simple Michaelis–Menten kinetics and mass transfer kinetics are used to model the glucose uptake rate, oxygen uptake rate, and the acetate secretion rate (Fig. 10-1). When applied to the analysis of diauxic growth of *Escherichia coli* on glucose [48], the results from DFBA were qualitatively similar to the experimental observations. Yugi et al. [47] improved the DFBA method by introducing the dynamic methods (kinetics) to the rate-limiting steps of the metabolic reactions and the static methods (FBA) to the remaining reactions. In this method, the reactions expressed in the form of the static model require no prior information about kinetic equations and parameters or about the initial concentrations of metabolites (Fig. 10-1). This method was successfully used for the simulation of the erythrocyte model [47].

The hybrid method reduced the cost for the development of large-scale *in silico* models as well as the number of experiments for the identification of kinetic properties for dynamic simulation.

There are several software programs available for performing analyses and simulations of genome-scale *in silico* metabolic network. MetaFluxNet is a software package for the modeling and simulation of metabolic reaction networks focusing on MFA [49,50] (Table 10-3). It also provides the management of metabolic information and supports the systems biology markup language (SBML) and the metabolic flux analysis markup language (MFAML) [50] for the exchange of metabolic models. Simpheny (Genomatica, San Diego, CA) is a commercial software program for the construction and simulation of *in silico* genome-scale metabolic models [51]. Simpheny allows construction of *in silico* cells from their molecular components and simulation of the complete biochemical reaction network of a cell. Simpheny can be used for the prediction of various phenotypic characteristics based on FBA. General

Table 10-3 Useful softwares for the analyses of genome-scale *in silico* metabolic network

Application	Web Site Address	Reference	Note (Usability)
MetaFluxNet	http://mbel.kaist.ac.kr/lab/mfn/	[49,50]	MetaFluxNet is a powerful software package for the *in silico* modeling and simulation of metabolic network using metabolic flux analysis. It supports for the generation and management of metabolic model using MFAML
GEPASI	http://www.gepasi.org/	[52]	GEPASI is a dynamic modeling software to construct and optimize network models with kinetic parameters
COPASI	http://www.copasi.org	[53]	COPASI provides tools for metabolic model generation, time course simulation, and metabolic control analysis. COPASI improved the GEPASI
BioSPICE	http://biospice.sourceforge.net/	[54,55]	BioSPICE provides an integration framework/workbench to integrate various tools according to their purpose

Pathway Simulator (GEPASI) has been widely used for the dynamic simulation and metabolic control analysis (MCA) of the metabolic network [52]. GEPASI contains several predefined kinetic models for the easy construction of dynamic simulation model. The Complex Pathway Simulator (COPASI) is an application for the simulation and analysis of biological networks [53]. It was developed based on the dynamic simulation tool, GEPASI. COPASI provides various tools including model generator, stochastic simulation tool, metabolic control analysis, and elementary mode analysis. It also supports the SBML format for the effective description of parameters of the kinetic equations. BioSPICE is an integrated system of the systems biology workbench (SBW) that allows the sharing of computational codes in various tools. It can be used to develop metabolic and genetic models using the common software framework [54,55].

10.4 ITERATIVE *IN SILICO* MODEL DEVELOPMENT

Since the *in silico* metabolic network cannot truly represent the real cell, it needs to be improved by iterative process. This process involves creating the metabolic network model, obtaining experimental data, comparing the predicted outcomes with experimental data, and resolving inconsistencies in the results to update the model. Hypotheses based on the results of *in silico* analysis can be tested by experiments, from which the model can be updated and improved based on the experimental results.

Both biochemical and genetic engineering experiments as well as computational tests are parts of the iterative process. As additional data for an organism become available, such as gene expression data and metabolic profiles, new biological information will be discovered to further refine and improve the *in silico* model.

In silico microbial models have been found to correctly predict experimentally observed behaviors of microbes *in vitro* 70–80 percent of the time [43,56,57]. Despite the relatively good agreement between the model predictions and actual experimental results, it is the 20 percent "failure" rate that is of most interest to us. These *in silico* "failures" point to areas of the model in which current knowledge on the organism is lacking (such as unknown pathways in the reconstruction, unaccounted-for regulatory interactions, etc.). These gaps in information must be filled in through new biological discovery. It is through the iterative process of model construction, testing, validation, and revision that new information on the organism can be discovered for filling in those gaps that will refine and improve the *in silico* model of the organism. By this iterative process, the most comprehensive and predictive *in silico* model of the organism can be built.

The current strategy for this process is to involve both experiments and the mathematical modeling/simulation in a feedback and iterative fashion (Fig. 10-2). The feedback approach is based on the prediction of genetic and metabolic modifications that can be compared with the experimental results, leading to a more rational strategy for the reconstruction of *in silico* model. Palsson et al. [58] showed that FBA could be used to predict what the eventual effects of genetic modifications would be on the global host cell physiology. The ability of a constraints-based model of *E. coli* describing genetic modifications was examined by subjecting them to adaptive evolution under different growth conditions.

Lee et al. [59] used this iterative approach by integrating genome and fluxome information in the characterization of a relatively less studied bacterium *Mannheimia succiniciproducens*. The genome was used to construct the genome-scale *in silico* metabolic map of *M. succiniciproducens*, and flux analysis was used to calculate the succinic acid yields and flux distributions under various conditions. It was found from the genome-scale flux analysis that carboxylation of phosphoenolpyruvate to oxaloacetate by PEP carboxykinase is the most important anaplerotic pathway leading to the efficient production of succinic acid by the reductive tricarboxylic acid cycle and menaquinone system [59]. In this iterative process, the proteome reference map of *M. succiniciproducens* was established by 2-DE coupled with mass spectrometry [60], and the results obtained were used to fine-tune the *in silico* metabolic network. The *in silico* metabolic network thus improved can be used to design new experiments for flux profiling and consequently for characterizing the metabolic characteristics under various environmental conditions.

10.5 METABOLIC ENGINEERING BASED ON THE *IN SILICO* MODEL FOR THE ENHANCED PRODUCTION OF VARIOUS BIOPRODUCTS

After the valid model is constructed, many *in silico* experiments can be carried out to quantify flux distributions under the numerous conditions of interest. These *in silico*

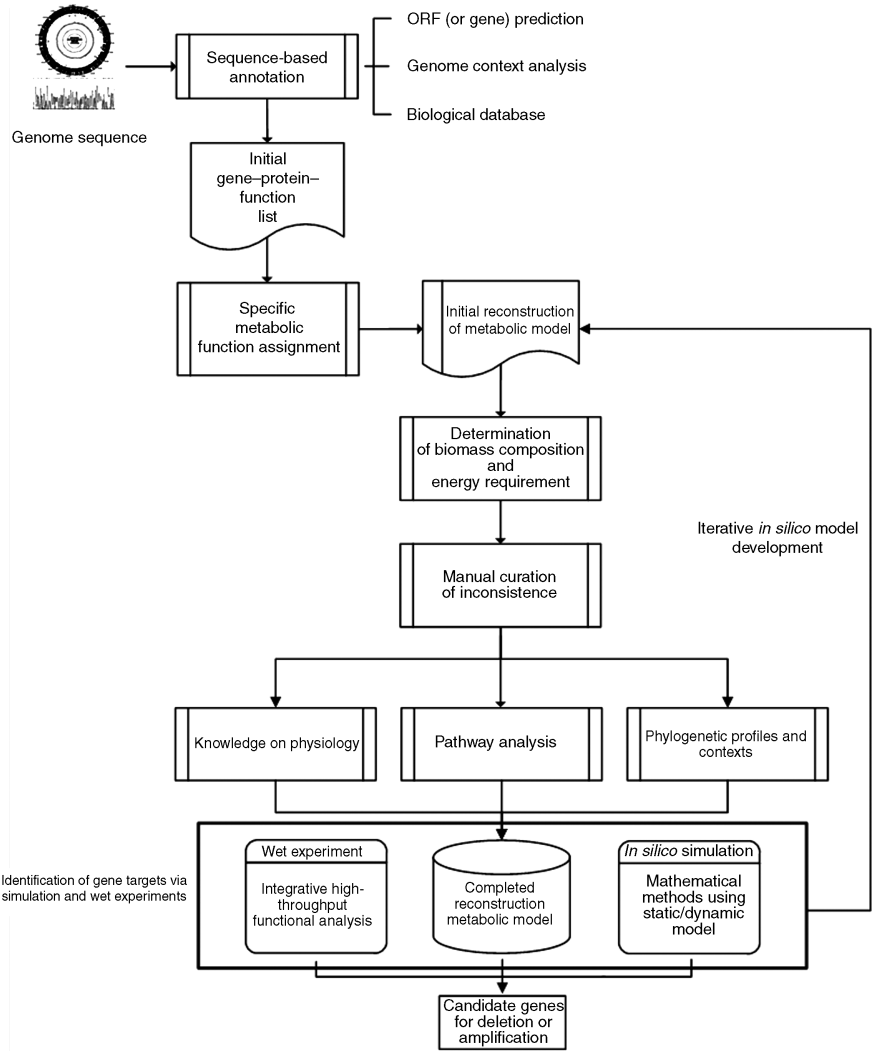

Figure 10-2 Flow chart for identifying gene targets by combining computational modeling/simulation and high-throughput experimental analyses. The outcomes of these analyses evolve during the iterations to allow identification of new gene targets.

experiments make it possible to decipher the metabolic and physiological changes of the cells under various genetic and/or environmental conditions, and consequently establish a more rational metabolic engineering strategy to achieve desired goals. Furthermore, plausible targets for genetic modifications can be identified to improve the strain's performance through the comparative study of responses observed under various genetic and environmental perturbations (Fig. 10-3).

Currently, there are a large number of microorganisms that are used industrially for the production of bioproducts. Although these microorganisms do produce the desired

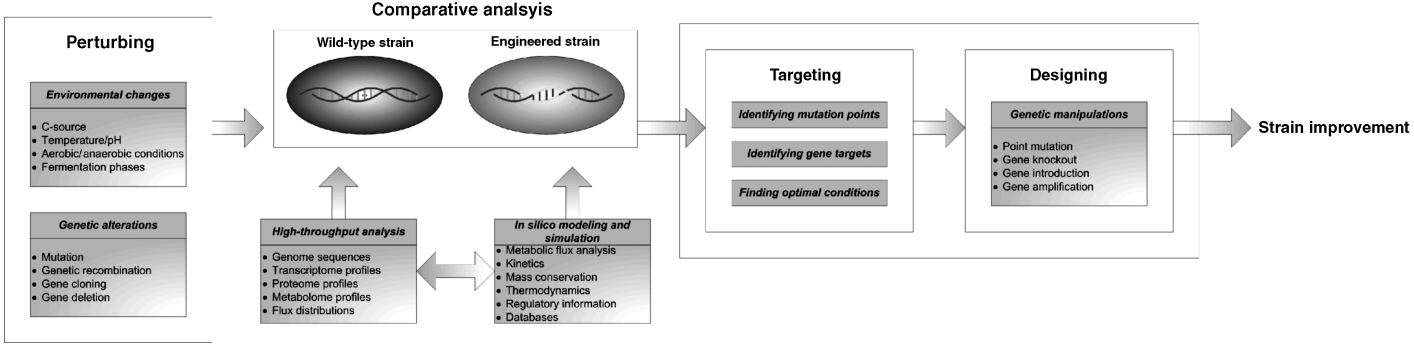

Figure 10-3 Procedure for the improvement of strains. Environmental and genetic changes significantly affect the cellular physiology. Wild-type and engineered strains can be comparatively analyzed by combining high-throughput technologies and *in silico* modeling and simulation. This comparative analysis would lead to identification of new gene targets to be introduced, gene knockout or amplification targets, and optimal production conditions. This newly generated knowledge can then be used to design experiments for genetic manipulations to bring about the desired phenotypes.

bioproducts, they do not naturally produce them to the concentrations and productivity high enough for commercialization. Additionally, the biological networks of microorganisms are robust enough to resist many changes introduced to them. Therefore, many combinatorial experiments including genetic manipulations (gene amplification and knockout), regulatory modification, and cultivation experiments need to be carried out to understand the metabolic characteristics and improve the phenotype to a desired level good enough for industrial applications. Here, *in silico* metabolic modeling and simulation can be used to overcome the impossibility of carrying out these many combinatorial experiments.

In silico organisms have been constructed to generate more knowledge about the cell and tackle the aforementioned problems. *E. coli*, the most well known and widely used bacterium, has been used for the production of a wide variety of bioproducts ranging from primary and secondary metabolites to biopolymers [61–64]. The *in silico E. coli* metabolic network has been expanded to contain up to 2077 reactions with 1039 metabolites [65]. However, baker's yeast, *Saccharomyces cerevisiae*, has been a model organism for understanding cellular physiology and compartmentalized intracellular biochemical behavior of a eukaryotic cell. Genome-based yeast model has the biochemical network of 1446 biochemical reactions and 1013 metabolites covering cytosolic and mitochondrial and transport reactions [44].

Obviously, gene manipulation is a very essential tool for strain improvement for the production of industrially valuable bioproducts. However, it is not possible to try every possible combination of gene targets as it is very time consuming and laborious. This is where FBA comes into play. FBA has most widely been exploited to quantitatively analyze the metabolic system thanks to its capability to predict the phenotypic behavior under various genetic and/or environmental conditions, and its applicability to genome-scale metabolic models [66,67]. Herein, strategies for the identification of gene knockout and addition targets as well as the combinatorial deletion, amplification, and regulation are described.

10.5.1 Identifying Gene Knockout and Addition Targets

Identifying the target genes for metabolic engineering to enhance the production of certain products is not always easy because of the large number of genes to be considered in the organism. Also, there is no guarantee that the identified single or even multiple target genes will enhance the production of the desired product due to the robustness of the biological network against changes to be made. At the initial stage, the potential target genes can be found through comparative analysis. The main obstacle to obtaining a rational solution to the problem of introducing genetic modifications is the lack of a reliable, global, metabolic model that captures stoichiometric, kinetic, and regulatory effects of the modifications on metabolite interconversions and metabolic flux distributions through the cellular reaction network. As a result, strain improvement has conventionally been achieved by random approaches whereby the target genes to be knocked out or amplified were intuitively selected rather than systematically. Consequently, the unexpected outcomes were often obtained. However, the genome-scale *in silico* metabolic model has changed a

paradigm by enabling systemic approaches for strain improvement. Such genome-scale model has been simulated by means of linear optimization with a particular objective function such as maximization of cellular growth rate or production rate of certain metabolite of industrial value. Although the optimal value obtained by linear programming does not exactly describe the actual state of the cellular physiology, this methodology is still worthy to consider as it provides an overall picture of the cell metabolism, particularly carbon and energy distribution.

Raman et al. [68] employed FBA to search drug targets from the mycolic acid pathway of *Mycobacterium tuberculosis*, an important human pathogen. Mycolic acid constitutes the protective layer of this pathogen, and the inhibition of its biosynthesis has been the drug target due to its essentiality in cell growth, survival, and pathogenicity. Based on this biochemical background, a comprehensive model of mycolic acid biosynthetic system was built, and FBA was performed to identify essential genes by systematically knocking out the genes. Those genes that, when knocked out, resulted in a zero value for the objective function, the maximization of mycolic acid production in this case, were considered as drug targets as the pathogens cannot survive without mycolic acids. Candidate drug targets were further screened by homology search of these genes against the human genome to ensure that the host system does not possess the similar genes, which may be unexpectedly targeted by the drug, leading to adverse effects. This study is a nice example of how *in silico* analyses can be applied to the drug development process.

Since FBA does not account for the physiological changes caused by genotypic mutation, the simulation results may deviate from the experimental data. This has led to the development of a new algorithm called minimization of metabolic adjustment (MOMA) (see Box 10-1). This method attempts to determine more realistic flux distributions in knockout mutants by minimizing the changes in the flux distribution of the mutant with respect to the wild type instead of maximizing the biomass formation in the mutant [69]. This method takes into account that the mutant strain is not optimized for the production of metabolites because it has not had a chance to fine-tune its new metabolic network through evolution. This framework can be used to identify target genes to be knocked out to present a phenotype that is closest to the wild type (Fig. 10-4). It was found that this suboptimal profile actually lies between the wild-type and the mutant optimals. In one study using MOMA, *in silico* single- and multiple-gene knockout experiments were performed to systematically identify the gene targets and, ultimately, increase the lycopene yield [61]. This strategy can be used to guide the choice of gene knockout targets. This method yielded a triple knockout mutant that produced less than 40 percent more lycopene compared with an engineered overproducing *E. coli* strain. This study demonstrates the value of system optimization using MOMA for the strain improvement.

Another method that is similar to MOMA is the regulatory on/off minimization (ROOM) method (see Box 10-1). MOMA is based on the minimization of the changes in the metabolic fluxes in the mutant strain from the wild type. There may be one or two fluxes that require huge changes to compensate for the effects the mutation puts on the system. ROOM, however, minimizes all the fluxes with respect to the wild type regardless of any other factors. This method is based on the assumption that the system chooses to minimize its adaptation cost through regulation of the fluxes to maintain the

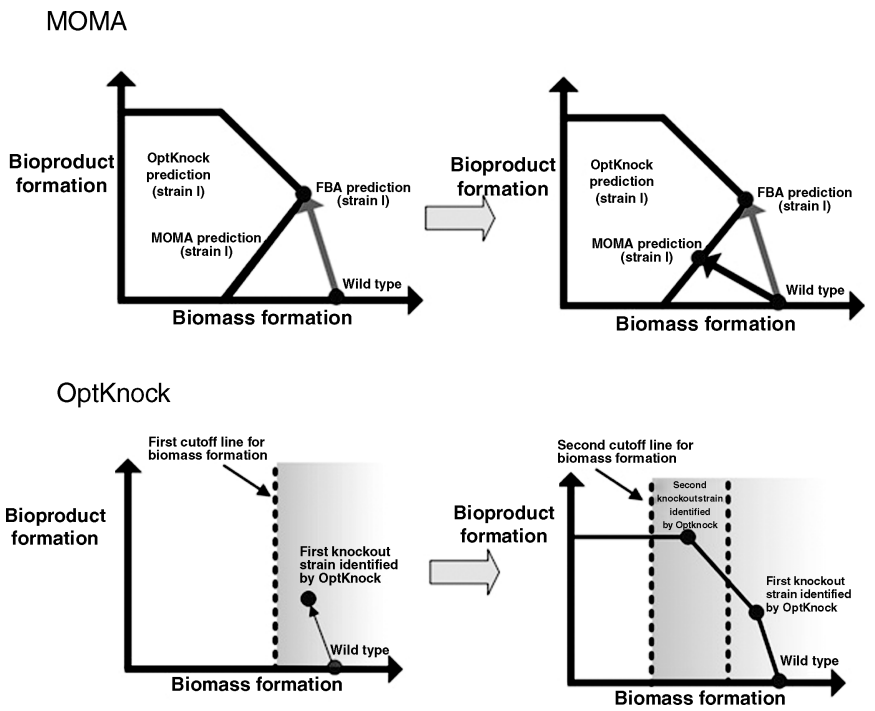

Figure 10-4 Graphical representation of the principles of MOMA and OptKnock. MOMA utilizes quadratic programming to find a metabolic state, in which artificially generated mutants try to minimize the redistribution of intracellular fluxes compared to the optimal flux distribution of wild type. Consequently, MOMA identifies a suboptimal metabolic state of the mutant that lies somewhere between the optimal state of the mutant and wild type in the altered solution space. This approach is based on the assumption that artificially generated mutant cannot immediately redistribute its fluxes toward the optimal growth rate since it has not undergone evolutionary pressure for an enough period of time as wild type had. OptKnock is a framework that suggests gene(s) to be knocked out for the enhanced production of bioproducts by considering both cell growth rate and objective metabolite production rate. This approach only considers the optimal production rate of the strain whose biomass formation rate is greater than the predetermined cutoff line. As a result, it will suggest a mutant genotype that allows faster growth only when it simultaneously produces a metabolite at faster rate.

wild-type stoichiometric and thermodynamic constraints. While both methods do not maximize biomass for the mutant strain, ROOM, by constraining the fluxes to "run in parallel" to the wild type, implicitly gives results under the maximum growth rates. ROOM was found to give similar or better prediction compared with FBA or MOMA in knockout experiments eight out of nine times [70].

The OptKnock method is another approach of identifying the knockout targets. This method identifies genes to be knocked out for bioproduct overproduction while considering the cell's needs as well (Fig. 10-4) (see Box 10-1). This approach was applied to lactate production in *E. coli* under anaerobic conditions where lactate production was maximized as an objective function in addition to the biomass objective function. This resulted in a coupling of lactate production with the formation of biomass [71]. The OptKnock method was also employed by the same research group

to optimize the production of amino acids. Additional constraints, such as ammonia and oxygen transport by the cell, were introduced to eliminate alternative solutions. The OptKnock method is especially well suited for the study of amino acid production system because the metabolic reactions for amino acid production are highly regulated. Although OptKnock does not consider regulatory networks, it is satisfactorily acceptable because it considers the global effects of any changes made. It is because of this global consideration on the cell that less intuitive strategies need to be formulated by using this method [72]. In the production of amino acids, OptKnock suggested a number of knockout strategies that could enhance the production of various amino acids. The results of the study showed an increase in the amino acid production compared to the current strains used in industry. For example, an alanine yield of 91.5 percent from glucose could be achieved, which is much higher than that (45–55 percent) currently achieved in industry [73].

Lee et al. [74] compared the metabolism of *M. succiniciproducens*, a succinic acid overproducer, with that of *E. coli* to engineer an *E. coli* strain to overproduce succinic acid (Fig. 10-5). Several candidate genes for deletion were identified in *E. coli*. From

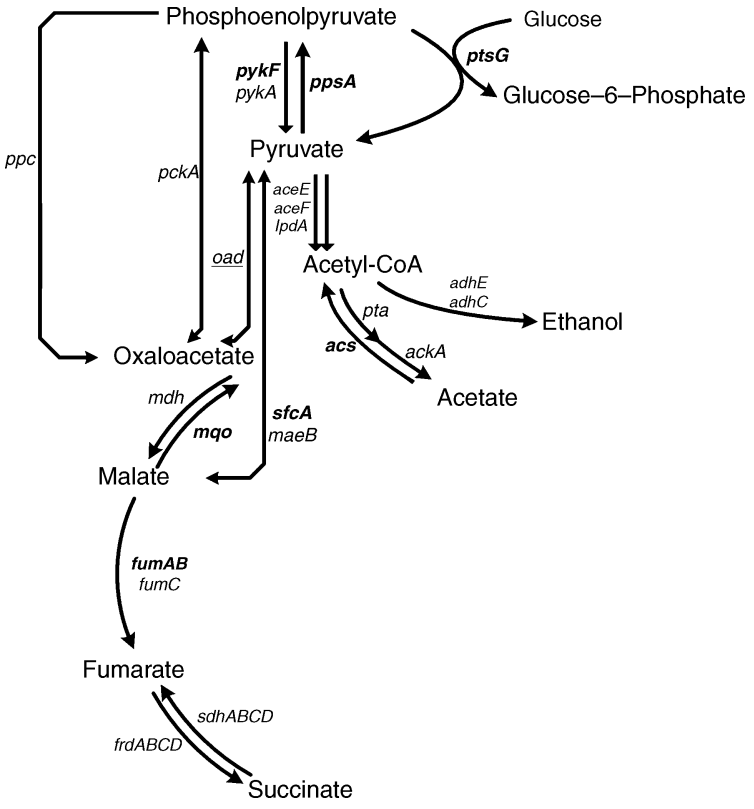

Figure 10-5 Comparison of the central metabolic pathways related to succinic acid formation in *M. succiniciproducens* and *E. coli*. Underlined genes represent those present only in *M. succiniciproducens* while those in boldface represent genes only present in *E. coli*.

this comparative genome analysis and using flux analysis to reverse engineer the metabolic network, succinic acid overproduction by *E. coli* could be achieved. During this process, our understanding on the general fermentative metabolic pathways in *E. coli* could be broadened. The fluxes to pyruvate and other acids were found to be the knockout targets for redirecting metabolic pathways toward enhanced succinic acid production. This example shows the effectiveness of combining comparative genomics, metabolic flux prediction, gene knockout, and fermentation toward strain development.

In contrast to gene knockout, FBA has also been used to identify genes to be amplified for enhanced metabolite production. FBA on poly(3-hydroxybutyrate) (PHB) producing *E. coli* predicted that the Entner–Doudoroff (ED) pathway, which was known to be inactive under normal culture conditions, was active during the production of PHB from glucose [64] (Fig. 10-6). This prediction was validated by actual experiments with a mutant *E. coli* strain defective in the activity of 2-keto-3-deoxy-6-phosphogluconate aldolase (Eda), a key enzyme in the ED pathway. Low PHB accumulation in the *eda* mutant strain compared to its parent strain could be restored when the *eda* gene was overexpressed in the *eda* mutant *E. coli* strain [64]. Also, the overexpression of the target genes (*fba* and *tpiA*) identified by FBA allowed enhanced production of PHB [75]. Therefore, MFA allows not only the knockout targets but also amplification targets to achieve enhanced metabolite production.

10.5.2 Combining the Deletion, Amplification, and Regulation of the Target Genes

The *in silico* genome-scale metabolic model can be used to further enhance the production of useful materials by combining the strategies of gene deletion, amplification, and regulation. Bro et al. [76] employed the genome-scale *in silico* model for the metabolic engineering of *S. cerevisiae* to improve the ethanol production. To increase the ethanol yield and reduce the yield of glycerol, an unnecessary by-product, a number of strategies were simulated using the previously reconstructed genome-scale model of *S. cerevisiae* [77]. Before they actually perform the simulations with the model, a few modifications were made to the model including incorporation of the necessary reactions. For example, those reactions catalyzed by xylose reductase and xylitol dehydrogenase for xylose metabolism were added to reflect actual experimental conditions as microorganisms were cultivated on the mixture of glucose and xylose. They then performed a gene insertion analysis by adding reactions one at a time from a pool of 3800 biochemical reactions that are derived from the LIGAND database [23]. The results of simulation by linear programming were scored based on the improvement of growth and ethanol yield and decreased glycerol yield. Consequently, the best-scored strategy, which predicted to improve the ethanol yield by 10 percent, but completely block the glycerol formation, was chosen for the actual experiment. According to the suggested strategy, they constructed a *S. cerevisiae* mutant, in which NADP-dependent glyceraldehydes-3-phosphate dehydrogenase (GAPN) was overexpressed, and achieved a 40 percent reduced glycerol yield with 3 percent increase in ethanol yield without affecting the specific growth rate. In a later study, the increased ethanol yield was also achieved with a GAPN expressing strain containing

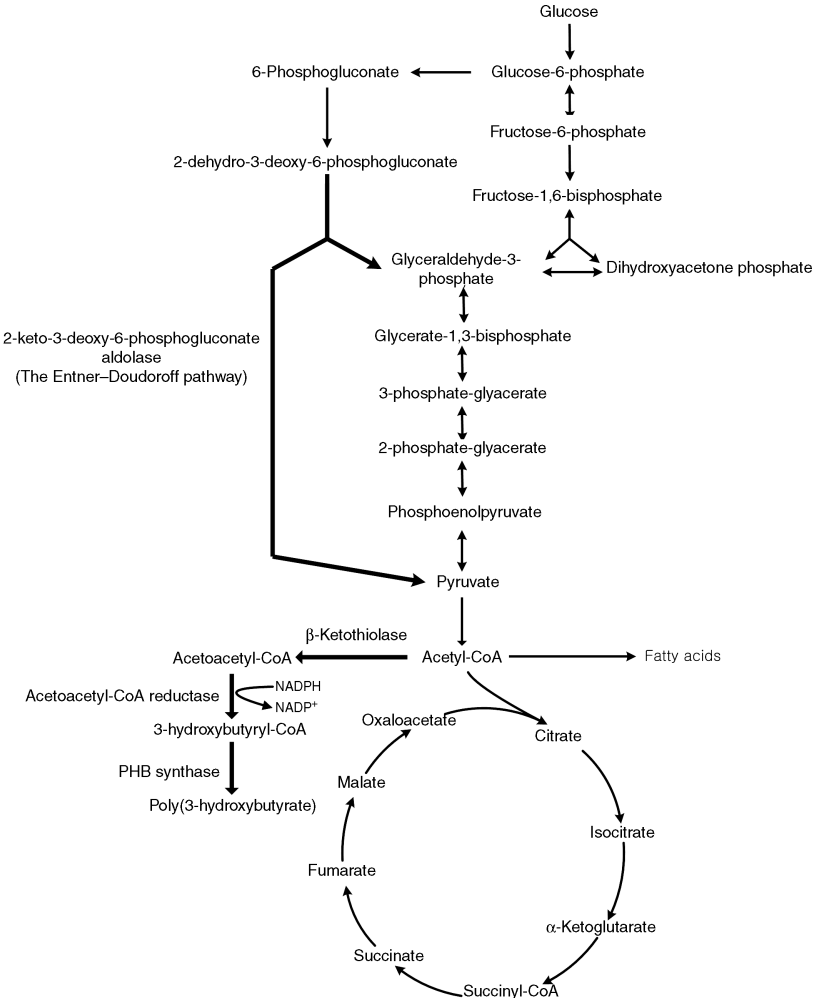

Figure 10-6 Metabolic network of *E. coli* for the production of PHB. The ED and PHB producing pathways are indicated with thick arrows. The ED pathway had been known to be inactive under the normal growth of *E. coli* using glucose as a carbon source. However, the simulation results of the *E. coli* model by FBA showed that the ED pathway is active. Consequently, overexpression of the corresponding enzyme, 2-keto-3-deoxy-6-phosphogluconate aldolase, in *E. coli* led to the improved production yield of PHB, and thus validated the simulation results of FBA.

xylose reductase and xylitol dehydrogenase, cultured on the mixture of glucose and xylose. This study is another good example of genome-scale *in silico* model for the hypothesis-driven metabolic engineering capable of predicting various strategies with acceptable accuracies. Moreover, it provides deeper insight into the metabolic characteristics because it shows how cofactors are linked with one another in different parts of the metabolic network. All benefits would lead to the more efficient way of a desired strain development.

Random mutagenesis, such as transposon mutagenesis and overexpression libraries, takes opposite approach to the systematical *in silico* analysis by randomly mutating the host organism and, thereby, producing a wide range of mutants. Since a large number of mutants must be screened for their improved phenotypes, smart screening system is essential. This approach is particularly beneficial and complementary to the *in silico* analysis because it can create mutants that cannot be predicted with the current techniques [78]. When coupled with global *in silico* metabolic analysis, this method becomes powerful for identifying targets for strain improvement [61]. Furthermore, it allows the dissection of critical subnetworks within the cell and a deeper understanding of that network, such as regulatory networks. By investigating how product formation correlates with these regulatory networks, putative molecular interactions may be inferred and examined in subsequent perturbations.

The metabolism of a living organism is controlled by not only mass balances but also various regulatory mechanisms such as transcriptional, translational, and allosteric regulations. By the incorporation of regulatory mechanisms, conditional activation and inactivation of metabolic networks can be mimicked, and optimal metabolic distributions can be obtained for different environmental conditions. So far, this has been achieved by incorporating transcriptional regulation into an *E. coli* MFA model based on Boolean logic [79–81]. With the inclusion of the transcriptional regulatory mechanisms, the accuracy of the MFA results increased to match experimental data better. Although this Boolean rule has been successfully combined with stoichiometric analysis, there is an inherent limitation in this method as the gene expression is somewhat stochastic and is not distinctive on-and-off type phenomenon in the real biological system [82,83]. In this context, probabilistic graphical models have been employed to model such regulatory networks, but its integration with a genome-scale metabolic model remains to be an open problem [83,84].

10.6 CONCLUSIONS AND FUTURE PROSPECTS

In this chapter, we have described the processes for the reconstruction of genome-scale *in silico* metabolic network using the genomic information and the applications of these models. These genome-scale metabolic networks are being applied to various fields. When combined with metabolic engineering, the genome-scale network can be utilized as a fundamental platform to identify key steps of bioproduct production under different conditions.

However, the construction of the metabolic network is by no means complete. Limitations on the network from the insufficient knowledge on the genetic characteristics of the genome create missing information such as gaps in the network. To complement the incompleteness of the model, experimental data should be sufficiently supported. In the post genomic era, the high-throughput omics technologies including transcriptomics, proteomics, fluxomics using ^{13}C-labeling flux analysis, and metabolomics can be efficiently used to validate the genome-scale model at least in a qualitative manner. For instance, simulation results (fluxes of biochemical reactions) of a genome-scale model can be compared with the transcriptome data

to confirm whether the corresponding genes are expressed in the transcriptome profile. Likewise, proteome and metabolome profiles can be compared with the simulation results and used to generate further constraints. At present, these constraints are rather on-and-off type (e.g., the flux is set to be zero if there is no transcription of the gene encoding the enzyme carrying out that reaction). It is expected that an efficient algorithm will be developed for integrating the changing levels of various omics data in a quantitative manner during MFA. Such upgrade of modeling and simulation based on the integration of omics data will reveal the metabolic and regulatory characteristics more realistically and help designing strategies for the future experiments aiming at strain improvement. Eventually, all these efforts will lead to the development of virtual cell factory that can be used to tailor-design strains that are capable of producing various useful materials for human life.

ACKNOWLEDGMENTS

Our work described in this chapter was supported by a grant from the Korean Ministry of Education, Science and Technology (Korean Systems Biology Research Grant, M10309020000-03B5002-00000) and by the Brain Korea 21 Project. Furthermore supports through the LG Chem Chair Professorship, IBM SUR program, and Microsoft are greatly appreciated.

REFERENCES

1. Bansal AK. Bioinformatics in microbial biotechnology—a mini review. *Microb Cell Fact* 2005;4:19.
2. Liolios K, Tavernarakis N, Hugenholtz P, Kyrpides NC. The Genomes OnLine Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res* 2006;34:D332–D334.
3. Lee SY, Lee DY, Kim TY. Systems biotechnology for strain improvement. *Trends Biotechnol* 2005;23;349–358.
4. Borodina I, Nielsen J. From genomes to *in silico* cells via metabolic networks. *Curr Opin Biotechnol* 2005;16:350–355.
5. Ishii N, Robert M, Nakayama Y, Kanai A, Tomita M. Toward large-scale modeling of the microbial cell for computer simulation. *J Biotechnol* 2004;113:281–294.
6. Hocquette JF. Where are we in genomics? *J Physiol Pharmacol* 2005;56(Suppl 3):37–70.
7. Koffas M, Stephanopoulos G. Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Curr Opin Biotechnol* 2005;16:361–366.
8. Rajasethupathy P, Vayttaden SJ, Bhalla US. Systems modeling: a pathway to drug discovery. *Curr Opin Chem Biol* 2005;9:400–406.
9. Smid EJ, Molenaar D, Hugenholtz J, de Vos WM, Teusink B. Functional ingredient production: application of global metabolic models. *Curr Opin Biotechnol* 2005;16:190–197.
10. Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol* 2005;13:550–558.

11. Tateno Y, Saitou N, Okubo K, Sugawara H, Gojobori T. DDBJ in collaboration with mass-sequencing teams on annotation. *Nucleic Acids Res* 2005;33:D25–D28.
12. Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K, Browne P, van den Broek A, Castro M, Cochrane G, Duggan K, Eberhardt R, Faruque N, Gamble J, Diez FG, Harte N, Kulikova T, Lin Q, Lombard V, Lopez R, Mancuso R, McHale M, Nardone F, Silventoinen V, Sobhany S, Stoehr P, Tuli MA, Tzouvara K, Vaughan R, Wu D, Zhu W, Apweiler R. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res* 2005;33:D29–D33.
13. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2005;33:D54–D58.
14. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McDowall J, Mitchell A, Nikolskaya AN, Orchard S, Pagni M, Ponting CP, Quevillon E, Selengut J, Sigrist CJ, Silventoinen V, Studholme DJ, Vaughan R, Wu CH. InterPro, progress and status in 2005. *Nucleic Acids Res* 2005;33: D201–D205.
15. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res* 2004;32:D138–D141.
16. Hulo N, Sigrist CJ, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A. Recent improvements to the PROSITE database. *Nucleic Acids Res* 2004;32:D134–D137.
17. Krummenacker M, Paley S, Mueller L, Yan T, Karp PD. Querying and computing with BioCyc databases. *Bioinformatics* 2005;21:3454–3455.
18. Hou BK, Kim JS, Jun JH, Lee DY, Kim YW, Chae S, Roh M, In YH, Lee SY. BioSilico: an integrated metabolic database system. *Bioinformatics* 2004;20:3270–3272.
19. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 2004;32: D431–D433.
20. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004;32:D277–D280.
21. Demir E, Babur O, Dogrusoz U, Gursoy A, Nisanci G, Cetin-Atalay R, Ozturk M. PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* 2002;18:996–1003.
22. Borodina I, Krabben P, Nielsen J. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res* 2005;15:820–829.
23. Goto S, Okuno Y, Hattori M, Nishioka T, Kanehisa M. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* 2002;30:402–404.
24. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. Expansion of the BioCyc collection of pathway/ genome databases to 160 genomes. *Nucleic Acids Res* 2005;33:6083–6089.
25. Karp PD, Paley S, Romero P. The Pathway Tools software. *Bioinformatics* 2002;18(Suppl 1): S225–S232.
26. Demir E, Babur O, Dogrusoz U, Gursoy A, Ayaz A, Gulesir G, Nisanci G, Cetin-Atalay R. An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics* 2004;20:349–356.

27. Zehetner G. OntoBlast function: from sequence similarities directly to potential functional annotations by ontology terms. *Nucleic Acids Res* 2003;31:3799–3803.
28. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 2005;21:3787–3793.
29. Snel B, Huynen MA, Dutilh BE. Genome trees and the nature of genome evolution. *Annu Rev Microbiol* 2005;59:191–209.
30. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 2001;29:22–28.
31. Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. *Nat Genet* 1999;21:108–110.
32. Hong SH, Kim TY, Lee SY. Phylogenetic analysis based on genome-scale metabolic pathway reaction content. *Appl Microbiol Biotechnol* 2004;65:203–210.
33. Eisen JA, Wu M. Phylogenetic analysis and gene functional predictions: phylogenomics in action. *Theor Popul Biol* 2002;61:481–487.
34. Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 2003;4:275–284.
35. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 2005;33:D433–D437.
36. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005;33:5691–5702.
37. Osterman A, Overbeek R. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 2003;7:238–251.
38. Teusink B, van Enckevort FH, Francke C, Wiersma A, Wegkamp A, Smid EJ, Siezen RJ. *In silico* reconstruction of the metabolic pathways of *Lactobacillus plantarum*: comparing predictions of nutrient requirements with those from growth experiments. *Appl Environ Microbiol* 2005;71:7253–7262.
39. Green ML, Karp PD. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* 2004;5:76.
40. Navarrete RM, Vara JA, Hutchinson CR. Purification of an inducible L-valine dehydrogenase of *Streptomyces coelicolor* A3(2). *J Gen Microbiol* 1990;136:273–281.
41. Somerville GA, Said-Salim B, Wickman JM, Raffel SJ, Kreiswirth BN, Musser JM. Correlation of acetate catabolism and growth yield in *Staphylococcus aureus*: implications for host-pathogen interactions. *Infect Immun* 2003;71:4724–4732.
42. Heinemann M, Kummel A, Ruinatscha R, Panke S. *In silico* genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol Bioeng* 2005;92:850–864.
43. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 2000;97:5528–5533.

44. Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanark S, Nielsen J, Bhumiratana S. The genome-scale metabolic model ilN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC syst Biol* 2008;2:71.
45. Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol* 2005;5:8.
46. Oliveira AP, Nielsen J, Forster J. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol* 2005;5:39.
47. Yugi K, Nakayama Y, Kinoshita A, Tomita M. Hybrid dynamic/static method for large-scale simulation of metabolism. *Theor Biol Med Model* 2005;2:42.
48. Mahadevan R, Edwards JS, Doyle FJ, 3rd, Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002;83:1331–1340.
49. Lee DY, Yun H, Park S, Lee SY. MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics* 2003;19:2144–2146.
50. Lee SY, Woo HM, Lee D-Y, Choi HS, Kim TY, Yun H. Systems-level analysis of genome-scale microbial metabolisms under the integrated software environment. *Biotechnol Bioproc Eng* 2005;10:425–431.
51. Mahadevan R, Burgard AP, Famili I, Dien SV, Schilling CH. Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals. *Biotechnol Bioproc Eng* 2005;10:408–417.
52. Mendes P. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput Appl Biosci* 1993;9:563–571.
53. Vallabhajosyula RR, Chickarmane V, Sauro HM. Conservation analysis of large biochemical networks. *Bioinformatics* 2006;22:346–353.
54. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *Omics* 2003;7:355–372.
55. Garvey TD, Lincoln P, Pedersen CJ, Martin D, Johnson M. BioSPICE: access to the most current computational tools for biologists. *Omics* 2003;7:411–420.
56. Famili I, Forster J, Nielsen J, Palsson BO. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci USA* 2003;100:13134–13139.
57. Thiele I, Vo TD, Price ND, Palsson BO. Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J Bacteriol* 2005;187:5818–5830.
58. Palsson BO, Ibarra RU, Edwards JS. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 2002;420:186–189.
59. Lee SY, Hong SH, Kim JS, In YH, Choi SS, Rih JK, Kim CH, Jeong H, Hur CG, Kim JJ. The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat Biotechnol* 2004;22:1275–1281.
60. Lee JW, Lee SY, Song H, Yoo JS. The proteome of *Mannheimia succiniciproducens*, a capnophilic rumen bacterium. *Proteomics* 2006;6:3550–3566.
61. Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 2005;23:612–616.

62. Lee SY, Hong SH, Moon SY. *In silico* metabolic pathway analysis and design: succinic acid production by metabolically engineered *Escherichia coli* as an example. *Genome Inform* 2002;13:214–223.
63. Hong SH, Lee SY. Metabolic flux distribution in metabolically engineered *Escherichia coli* strain producing succinic acid. *J Microbiol Biotechnol* 2000;10:496–501.
64. Hong SH, Park SJ, Moon SY, Park JP, Lee SY. *In silico* prediction and validation of the importance of the Entner–Doudoroff pathway in poly(3-hydroxybutyrate) production by metabolically engineered *Escherichia coli*. *Biotechnol Bioeng* 2003;83:854–863.
65. Feist Am, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Paisson BØ. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121.
66. Patil KR, Akesson M, Nielsen J. Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* 2004;15:64–69.
67. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 1994;60:3724–3731.
68. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* 2005;1:e46.
69. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 2002;99:15112–15117.
70. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* 2005;102:7695–7700.
71. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643–648.
72. Pharkya P, Burgard AP, Maranas CD. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* 2003;84:887–899.
73. Ikeda M. Amino acid production processes. *Adv Biochem Eng Biotechnol* 2003;79:1–35.
74. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol* 2005;71:7880–7887.
75. Park SJ. Metabolic engineering for the production of poly(3-hydroxyalkanoate) in recombinant *Escherichia coli*, Dissertation, Korea Advanced Institute of Science and Technology, Daejeon, Korea, 2003.
76. Bro C, Regenberg B, Forster J, Nielsen J. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* 2006;8:102–111.
77. Forster J, Famili I, Fu P, Palsson BO, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 2003;13:244–253.
78. Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K, Ikeda M. A novel methodology employing Corynebacterium glutamicum genome information to generate a new L-lysine-producing mutant. *Appl Microbiol Biotechnol* 2002;58:217–223.
79. Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002;277:28058–28064.
80. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004;429:92–96.

81. Herrgard MJ, Lee BS, Portnoy V, Palsson BO. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* 2006;16:627–635.

82. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 2005;6:451–464.

83. Sun N, Zhao H. Genomic approaches in dissecting complex biological pathways. *Pharmacogenomics* 2004;5:163–179.

84. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science* 2004;303:799–805.

85. Lee SY, Papoutsakis ET. *Metabolic Engineering*. New York: Marcel Dekker, 1999.

86. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32:D258–D261.

87. Hoersch S, Leroy C, Brown NP, Andrade MA, Sander C. The GeneQuiz Web server: protein functional analysis through the Web. *Trends Biochem Sci* 2000;25:33–35.

88. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999;27:4636–4641.

89. Riley ML, Schmidt T, Wagner C, Mewes HW, Frishman D. *The PEDANT genome database in 2005*. *Nucleic Acids Res* 2005;33:D308–D310.

90. Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, Selkov E, Jr., Liolios K, Joukov V, Kaznadzey D, Anderson I, Bhattacharyya A, Burd H, Gardner W, Hanke P, Kapatral V, Mikhailova N, Vasieva O, Osterman A, Vonstein V, Fonstein M, Ivanova N, Kyrpides N. The ERGO genome analysis and discovery system. *Nucleic Acids Res* 2003;31:164–171.

91. Kjeldsen KR, Nielsen J. *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng* 2009;102:583–597.

92. Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VA. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol* 2008;4:e1000210.

93. Nogales J, Palsson BØ, Thiele I. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst Biol* 2008;2:79.

94. Senger RS, Papoutsakis ET. Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. *Biotechnol Bioeng* 2008;101:1036–1052.

95. Lee J, Yun H, Feist AM, Palsson BØ, Lee SY. Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl Microbiol Biotechnol* 2008;80:849–862.

96. Durot M, Le Fèvre F, de Berardinis V, Kreimeyer A, Vallenet D, Combe C, Smidtas S, Salanoubat M, Weissenbach J, Schachter V. Iterative reconstruction of a global metabolic model of *Acinetobacter baylyi* ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst Biol* 2008;72:85.

97. Oberhardt MA, Puchałka J, Fryer KE, Martins dos Santos VA, Papin JA. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J Bacteriol* 2008;190:2790–2803.

98. Resendis-Antonio O, Reed JL, Encarnación S, Collado-Vides J, Palsson BØ. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput Biol* 2007;3:1887–1895.

99. Baart GJ, Zomer B, de Haan A, van der Pol LA, Beuvery EC, Tramper J, Martens DE. Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biol* 2007;8:R136.

100. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* 2007;282:28791–28799.

101. Jamshidi N, Palsson BØ. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol* 2007;1:26.

102. Beste DJ, Hooper T, Stewart G, Bonde B, Avignone-Rossa C, Bushell ME, Wheeler P, Klamt S, Kierzek AM, McFadden J. GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol* 2007;8:R89.

103. Mahadevan R, Bond DR, Butler JE, Esteve-Nuñez A, Coppi MV, Palsson BO, Schilling CH, Lovley DR. Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Appl Environ Microbiol* 2006;72:1558–1568.

104. Kim TY, Kim HU, Park JM, Song HH, Kim JS, Lee SY. Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. *Biotechnol Bioeng* 2007;97:657–671.

105. Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 2004;14:1298–1309.

106. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 2003;4:R54.

107. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol* 2002;184:4582–4593.