Modeling with a View to Target Identification in Metabolic Engineering: A Critical Evaluation of the Available Tools

Jo Maertens
BIOMATH, Dept. of Applied Mathematics, Biometrics, and Process Control, Ghent University, Ghent 9000, Belgium

LIMAB, Dept. of Biochemical and Microbial Technology, Ghent University, Ghent 9000, Belgium

Peter A. Vanrolleghem
modelEAU, Dépt. de Génie Civil, Pavillon Pouliot, Université Laval, Québec, QC, Canada G1K7P4

DOI 10.1002/btpr.349
Published online January 5, 2010 in Wiley InterScience (www.interscience.wiley.com).

The state of the art tools for modeling metabolism, typically used in the domain of metabolic engineering, were reviewed. The tools considered are stoichiometric network analysis (elementary modes and extreme pathways), stoichiometric modeling (metabolic flux analysis, flux balance analysis, and carbon modeling), mechanistic and approximative modeling, cybernetic modeling, and multivariate statistics. In the context of metabolic engineering, one should be aware that the usefulness of these tools to optimize microbial metabolism for overproducing a target compound depends predominantly on the characteristic properties of that compound. Because of their shortcomings not all tools are suitable for every kind of optimization; issues like the dependence of the target compound's synthesis on severe (redox) constraints, the characteristics of its formation pathway, and the achievable/desired flux towards the target compound should play a role when choosing the optimization strategy.
 Biotechnol. Prog., 26: 313–331, 2010
Keywords: target identification, metabolic engineering, stoichiometric network analysis, stoichiometric modeling, dynamic metabolic modeling, multivariate statistics

Introduction

The well-established chemical synthesis routes face, although the era of the oil-based society has not come to an end yet, more and more competition from industrial biotechnological alternatives for the production of an increasing number of compounds, due to, e.g., environmental concerns and the increasing scarcity of oil. Whereas in the past micro-organisms were typically used for the production of stereochemical[1] and complex molecules, e.g., antibiotics,[2,3] nowadays they even become an interesting alternative for many bulk chemicals. To develop an industrial biotechnological process that can compete with the more mature chemical synthesis routes, there are four critical development phases: the choice of the favourite micro-organism, metabolic engineering, scaling-up, and downstream processing. The second phase in the development of an economically viable industrial biotechnological process is the optimization of the micro-organism itself using a wide range of both experimental and mathematical techniques.

To this end, due to the complexity of microbial metabolism, more and more metabolomic, proteomic, transcriptomic, and genomic data are collected,[4–7] which appear to be valuable to steer the process of genetic engineering with a view to the overproduction of a target compound. Indeed, these data help to elucidate the flux distribution, to determine the flux controlling reactions, and to yield insight in the regulation of metabolism.

In addition to these experimental techniques, mathematical methods are developed and commonly applied to interpret and to extract information from this pile of data and to identify genetic targets for the overproduction of a target compound (Table 1). In this context steady-state[20] and dynamic metabolic modeling,[21] multivariate statistics,[22–24] graph theory,[25] and neural networks are used to unravel microbial behavior.

Finally, the development of genetic toolboxes consisting of promoter libraries[22,26,23] and strategies for gene knock-outs, knock-ins, knock-downs, and knock-ups,[27] and the advent of functional genomics[28,29] have allowed the directed improvement of cellular properties in view of optimizing the production host. After some iterative rounds of genetic modification and host evaluation results this hotchpotch of techniques into the development of a host with improved performance.

Such a systematic approach is obvious as the vast variety of biochemical pathways micro-organisms dispose of, in order to fulfil their growth and reproduction requirements under a wide range of environmental conditions, renders them hard to fathom. A thorough understanding of the regulation of microbial processes is however a conditio sine qua non for the rational design of bioprocesses, as a disturbance in one part of metabolism can trigger a series of reactions on

Correspondence concerning this article should be addressed to J. Maertens at Jo.Maertens@biomath.UGent.be.

all levels of regulatory control and in all parts of metabolism. Indeed, in complex metabolic networks it is often a futile avocation to ad hoc predict the impact, both qualitatively and quantitatively, of a genetic intervention.[30] Hence, the popularity of models for metabolic engineering purposes. A concise overview and a critical evaluation of the popularly used tools in this development phase will be given below (Figure 1; Table 2).

Construction of a Stoichiometric Model

The reconstruction of genome-scale metabolic network models relies on assembling various sources of information about all the biochemical reactions in the network (Figure 2). A variety of data sources can be used to synthesise a list of chemical reactions that form an organism's metabolic network. The principal data sources are biochemistry, genomics, physiology, and indirect information, and in silico data. To this end, numerous databases can be consulted to collect the necessary data:

Genome annotation

Since extensive biochemical information is only available on a few organisms, reconstruction relies heavily on the annotated genome sequence. ORFs are identified on the genomic sequence, then assigned a function. This annotation can be achieved using experimental methods (gene cloning and expression or gene knock-out) or more commonly by comparing the sequence homology to genes of known function in other organisms. In silico annotation methods typically lead to the functional assignment of 40–70% of the identified ORFs on a freshly sequenced microbial genome. New and improved methods continue to be developed for genome annotation. For example, functions of gene products may be inferred from protein–protein interactions, transcriptomics, phylogenetic profiles, protein fusion, and operon clustering (Table 3). A genome-scale metabolic network can be reconstructed from the annotated sequence.[46–48] To produce high-quality, well-curated reconstructions one still has to manually verify all the components and links in a network, since there are often subtle differences even between related organisms. Many Web resources are available for this purpose (Table 3).

Table 1. Target Identification Relying on Metabolic Modeling

Model-Based Optimization Method	Production Host	Target Compound
Elementary modes	*E. coli*	L-methionine[8]
	C. glutamicum	L-methonine[8]
Flux balance analysis	*E. coli*	succinic acid[9]
	M. tuberculosis	mycolic acid[10]
	S. cerevisiae	succinic acid[11]
	S. cerevisiae	glycerol[11]
	S. cerevisiae	vanillin[11]
	E. coli	lycopene[12]
	E. coli	L-threonine[13]
	E. coli	L-valine[14]
	E. coli	succinic acid[15]
Partial least squares	*E. coli*	phenylalanine[16]
	Trichoderma sp.	cellulase[17]
Dynamic metabolic modeling	*E. coli*	carnitine[18,19]

Publicly available sources of sequence data

There are several publicly available databases that contain genomic data (Table 3). The comprehensive microbial resource (CMR) provides tools for the analysis of 63 annotated genome sequences, both individually and collectively. Another database that maintains many microbial genomes is the genomes on-line database (GOLD) site.

Biochemical data

Direct biochemical information is the most reliable source for the presence of a reaction in an organism. Biochemical data also yields the reaction stoichiometry and whether a reaction is reversible. Collections of biochemical data on an organism's metabolism is often found in review articles and more recently in whole volumes that are focused on the biology of a single organism.

Enzyme commission numbers

E.C. numbers are used to systematically characterize enzymatic reactions (http://www.chem.qmul.ac.uk/iubmb/enzyme/). They have been established to unambiguously classify reactions, which is needed because so many enzymes have ambiguous and duplicate names across organisms. A classification scheme similar to the E.C. system is being developed for transport reactions. Unfortunately, there is no similar system for genes, which have the same problem of ambiguous and duplicate names. Thus, the curation of gene annotation information for a reconstruction can be quite laborious.

Protein databases

Swiss-Prot (http://us.expasy.org/sprot/) is a very useful source for examining particular protein or reaction assignments in detail and is considered a standard for biochemical information because it is so well-curated. It contains literature references, sequences, functional assignments, and other

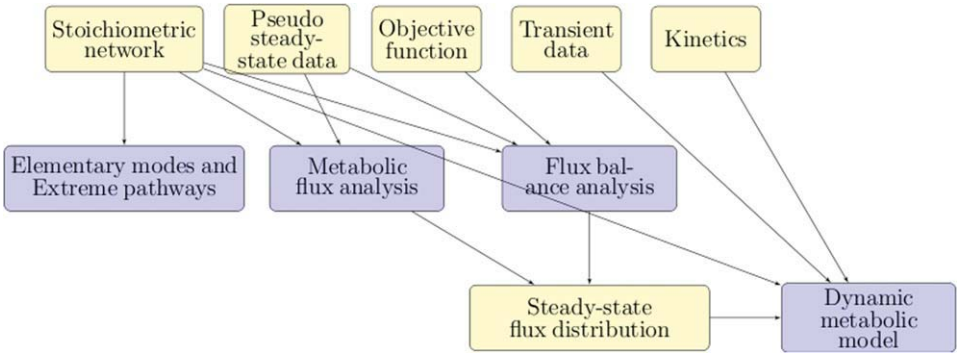


Figure 1. Modeling with a view to target identification in metabolic engineering.

Table 2. Overview of the Model-Based Approaches to Identify Genetic Engineering Targets

	Characteristic	New Advances	Similarities	Differences
Elementary modes & extreme pathways	Convex analysis[31,32] Pathway analysis Underdetermined systems	Differentiates between irreversible reactions and reversible reactions	Stoichiometry Insight Relies on assumptions on the P/O-ratio	No data
Metabolic flux analysis	(Over)determined systems Lump reactions	Identification of principal nodes Error identification Balancing	Relies on assumptions on the P/O-ratio stoichiometry Insight	Data (HPLC)
Flux balance analysis (FBA)	Underdetermined systems[33]	Goal function Linear/nonlinear programming[34]	Relies on assumptions on the P/O-ratio stoichiometry Insight	Prediction
Constrained FBA	Energy balance (EBA)[35] Genetic regulation (ROOM, MOMA)[36,37]	Additional constraints		No assumptions on P/O-ratio
Carbon modeling	Iterative solution method[38] Able to resolve parallel pathways,… Isotopomer balances[38] Cumomer balances[39] Elementary metabolite units[40]		Stoichiometry Insight	Iterative solution methods Data (NMR, MS) Multiple traces: C, H, and O Reduces computational cost
Approximative and mechanistic models	Dynamic data Overparameterized[41] Constant enzyme levels Kinetics are introduced Genetic regulation[42,43] Focus on a part of the metabolism[21]	Kinetics[21]	Identifiability issues Insight/Prediction stoichiometry	Kinetics Studied out of context
Cybernetic modeling	Goal function[44] Variable metabolite and enzyme levels Concept of limited resources Cell modeling	Variable metabolite and enzyme levels Concept of "limited resources" Controller steers cellular processes[45]	Identifiability issues Insight/Prediction	Kinetics Resolution sufficient ?

useful information, all specific to the organism being examined. If the presence of a protein in an organism is uncertain, but a page is found for it on Swiss-Prot, one can be fairly sure that the protein has been characterized and that literature references are available. TrEMBL contains new entries to Swiss-Prot that have not yet been curated.

Gene–protein–reaction (GPR) associations

When associating genes to reactions, and vice verse, it is important to remember that not all genes have a one-to-one relationship with their corresponding enzymes or metabolic reactions. Genes may encode subunits of a protein which catalyse one reaction. One example is the fumarate reductase. There are four subunits, FrdA, FrdB, FrdC, and FrdD, without which the enzyme (a protein complex) will not be able to catalyse the reaction. Conversely, there are genes that encode so-called promiscuous enzymes that can catalyse several different reactions, such as transketolase I in the pentose phosphate pathway. Such reactions typically involve similar chemical transformations of structurally related molecules. These examples highlight the need to keep track of associations between genes, proteins, and reactions.

Organism-specific sources of information

Several biological databases that integrate genomic and biochemical data for a particular organism are becoming available. One of the earliest of such databases is the *E. coli* encyclopedia (EcoCyc) database.[49] Comprehensive Yeast Genome Database (CYGD), Yeast Protein Database (YPD), and *Saccharomyces* Genome Database (SGD) are some examples for yeast.

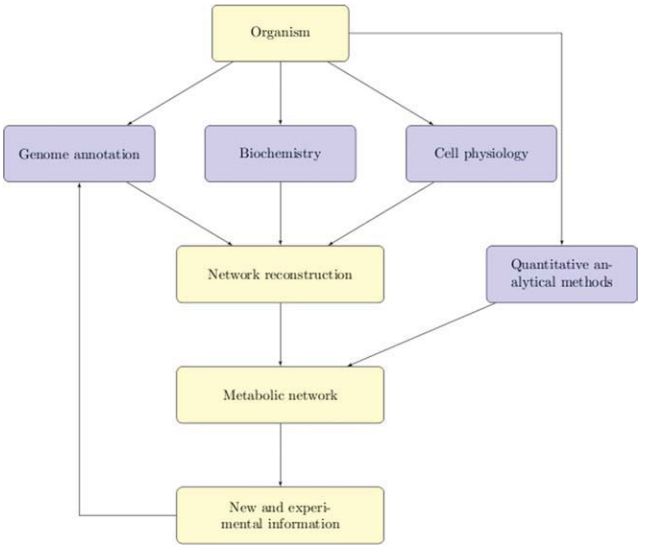


Figure 2. Construction of a genome-scale model.

Table 3. Popularly Used Databases for the Construction of a Stoichiometric Model

Comparative Genomic Databases	
SEED	http://seed-viewer.theseed.org/
GenDB	http://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb_info/
GeneQuiz	http://swift.cmbi.kun.nl/swift/genequiz/
STRING	http://string.embl.de/
DNA Sequence and Genome Annotation Databases	
EMBL	http://www.ebi.ac.uk/embl/
GenBank (NCBI)	http://www.ncbi.nlm.nih.gov/
TIGR	http://www.jcvi.org/cms/research/software/
CMR	http://cmr.jcvi.org/
Strain Specific Databases	
ECOCYC	http://ecocyc.org/
Colibri	http://genolist.pasteur.fr/Colibri/
GenProtEC	http://genprotec.mbl.edu/
CYGD	http://mips.gsf.de/genre/proj/yeast/
PyloriGene	http://genolist.pasteur.fr/PyloriGene/
Protein and Enzyme Databases	
BRENDA	http://www.brenda-enzymes.org/
ENZYME	
TransportDB	http://www.membranetransport.org/
UniProt	http://www.uniprot.org/
STRING	http://string.embl.de/
KEGG	http://www.genome.jp/kegg/
Pathway Databases	
KEGG	http://www.genome.jp/kegg/
Biocyc	http://biocyc.org/
UniPathway	http://www.grenoble.prabi.fr/obiwarehouse/unipathway

Stoichiometric Network Analysis

From the early stages of metabolic modeling, stoichiometric network models have been used to facilitate the choice of where to intervene genetically. The metabolic network comprises the metabolites and the reactions they are involved in, including formation, degradation, transport, and cellular utilization gathered from databases[49–51] and the literature.[52] For every metabolite a mass balance can be derived:

$$\frac{dx_i}{dt} = \sum_j s_{ij} r_j - b_i \tag{1}$$

where s_{ij} is the stoichiometric coefficient associated with flux r_j and b_i the net transport flux of metabolite x_i. Under pseudo steady-state conditions Eq. 1 will reduce to:

$$0 \cong \sum_j s_{ij} r_j - b_i \tag{2}$$

Eq. 2 can be rewritten in matrix notation:

$$\begin{bmatrix} 0 \\ b \end{bmatrix} \cong S \times R \tag{3}$$

where S is the stoichiometric matrix, R is the vector of metabolic fluxes, and b is the vector representing m transport fluxes over the cell membrane.

Despite success stories of stoichiometric metabolic model use to identify targets for modification, there have also been many false positive targets identified by these models. It is still unclear whether the well-established technique of stoichiometric modeling is fully apt to steer the process of metabolic engineering, since the kinetics and the regulation of the enzymatic reactions are not accounted for.[12,53]

Especially for the optimization of the production of metabolites in primary metabolism that are subject to severe (redox) constraints, stoichiometric modeling has been shown to be useful. It is less so for the optimization of minor pathways.[54]

Once the metabolic network model is built one can resort to stoichiometric network analysis, in the absence of data. Network analysis provides for the identification of elementary modes and extreme pathways as will be discussed later.

Elementary modes and extreme pathways

Network-based pathway analysis, e.g., identification of elementary modes (EMs) and extreme pathways (EPs) facilitates the assessment of network properties. Both of these methods use convex analysis, a branch of mathematics that enables the analysis of inequalities and systems of linear equations to generate a convex set of vectors that can be used to characterise all of the steady-state flux distributions of a biochemical network.[55]

This convex solution corresponds geometrically to a convex polyhedral cone in a n-dimensional space emanating from the origin. Within this cone lie all possible steady-state solutions. For convex cones one studies the extreme rays that correspond to the edges of the cone being half-lines emanating from the origin. These extreme rays are said to generate the cone and cannot be decomposed into a nontrivial convex combination of any other vectors residing in the flux cone. Every point within this cone can be written as a non-negative linear combination of the extreme pathways.[56]

The number of elementary modes can be greater than the number of extreme pathways, these elementary modes lie within the interior of the flux cone generated by the extreme pathways and are positive combinations of the extreme pathways. This creates a situation in which there is a redundancy in the pathway structure resulting often in a nonunique decomposition of a steady state flux distribution.[56] Both EMs and EPs have the following properties[55]:

1. There is a unique set of elementary modes/extreme pathways for a given network.

2. Each elementary mode/extreme pathway consists of the minimum number of reactions that is required to exist as a functional unit, i.e., a collection of reactions obeying to the steady-state equation (Eq. 3). If any reaction in an elementary mode/extreme pathway would be removed, the whole elementary mode/extreme pathway could not operate anymore as a functional unit. This property has been called genetic independence and non-decomposability.

Whereas elementary modes are the set of all routes through a metabolic network consistent with the latter properties, extreme pathways are the systemically independent subset of elementary modes (Figure 3); that is, no extreme pathway can be represented as a non-negative linear combination of any other extreme pathways.[55]

Pros and cons for the rational design of bioprocesses—One thus should be aware that EMs and EPs are related network-based approaches. Though, pretty similar, small differences do exist:

In contrast with EMs, EPs are not a set of all genetically independent reactions through a metabolic network, rather they are the edges of the high-dimensional convex solutions space and as such are the convex basic vectors. The elementary modes are a superset of the extreme pathways, including additional network pathways that meet the specific criteria. The number of EPs is less than or equals the number of

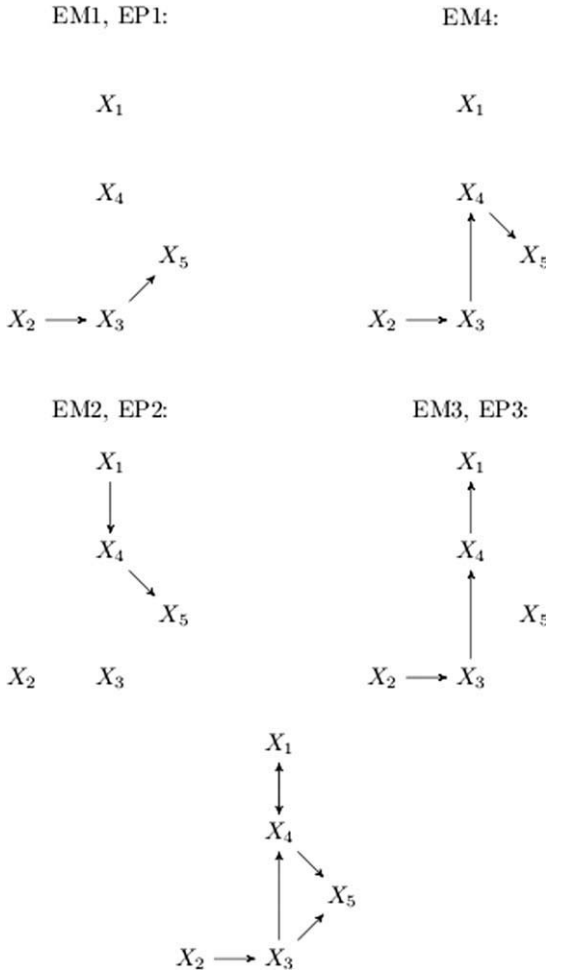


Figure 3. The three extreme pathways and four elementary modes of the stoichiometric network.

Note that the EM 4 is a non-negative linear combination of two extreme pathways, i.e., EP2 and EP3.

EMs. The EMs that are additional to the EPs are non-obvious highly complex combinations of the basic vectors (EPs).

Being (merely) basic vectors, extreme pathways might have to be added together to represent a particular flux distribution that cancels out a reversible exchange flux. Such occurrence complicates the full evaluation of network properties, such as pathway redundancy and product yields. Conversely, elementary modes have a much larger set of vectors to account for the absence of reversible exchange flux.

Both elementary modes and extreme pathways have been used to calculate product yields, to evaluate pathway redundancy, to determine correlated reaction sets, and to assess the effect of gene deletions.[55] Carlson et al. (2002)[57] and Kromer et al. (2006)[8] used elementary modes for rational design purposes and Carlson and Scrienc (2003),[58] Nookaew et al. (2007),[59] and Schwarts et al. (2007)[60] used the concept of elementary modes in combination with experimental data for network analysis.

Popularly used objective functions for EMs and EPs to steer the process of metabolic engineering with a view to overproduction of a target compound are typically based on reaction participation, which is a measure that links the occurrence of a reaction to the production of the target compound.[32] Another objective function focusses on the minimum length of EMs or EPs. This measure reflects the demand for cellular resources for enzyme synthesis to install a metabolic pathway.[32] The physiological interpretation of the results, see also Figure 4, and their computation for genome-scale models remain however challenging.[63] Though, progress has been made to deal with the latter.[64]

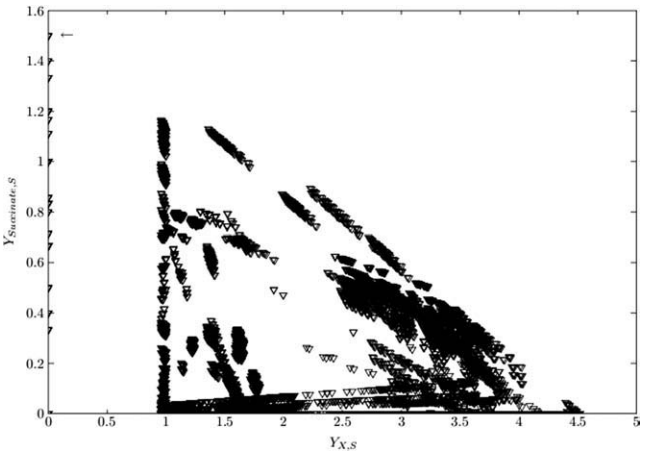


Figure 4. The 17528 elementary modes of the stoichiometric *E. coli* model of Lequeux et al. (2006)[61] represented as ▽s, calculated by using Metatool 5.0,[62] and presented in the $Y_{X,S}$, $Y_{succinate,S}$ space, with $Y_{X,S}$ and $Y_{succinate,S}$ the biomass $\left[\frac{c-mole}{mole}\right]$ and succinate $\left[\frac{mole}{mole}\right]$ yields on glucose, respectively.

The arrow indicates the EMs characterized by the optimal flux distribution, here with respect to maximal $Y_{succinate,S}$.

Stoichiometric Modeling

In the presence of data, one can resort to steady-state modeling, e.g., metabolic flux analysis and flux balance analysis (Figure 5). Eq. 3 can be rewritten as:

$$0 \cong \begin{bmatrix} S_{in} & 0 & 0 \\ S^c_{ex} & -I^c_{ex} & 0 \\ S^m_{ex} & 0 & -I^m_{ex} \end{bmatrix} \begin{bmatrix} r_{in} \\ b^c_{ex} \\ b^m_{ex} \end{bmatrix} \tag{4}$$

where r_{in} represents the intracellular fluxes, b^c_{ex} and b^m_{ex} the net transport fluxes to be calculated and measured, respectively. S_{in}, S^c_{ex}, and S^m_{ex} are the corresponding stoichiometric matrices and I represents a unity matrix. This equation can be rewritten as:

$$0 \cong \underbrace{\begin{bmatrix} S_{in} & 0 \\ S^c_{ex} & -I^c_{ex} \\ S^m_{ex} & 0 \end{bmatrix}}_{W_c} \underbrace{\begin{bmatrix} r_{in} \\ b^c_{ex} \end{bmatrix}}_{a_c} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ -I^m_{ex} \end{bmatrix}}_{W_m} \underbrace{\begin{bmatrix} b^m_{ex} \end{bmatrix}}_{a_m} \tag{5}$$

the solution of which is:

$$a_c \cong -W^{\#}_c W_m a_m + \text{null space}(W_c) f \tag{6}$$

with $W^{\#}_c$ the pseudo inverse of matrix W_c, with the null space defined as the set of linear independent basis vectors R_n that fulfil the equation:

$$W_c R_n = 0 \tag{7}$$

and f an arbitrary vector that reflects the indeterminacy of the system and with as many elements as there are columns

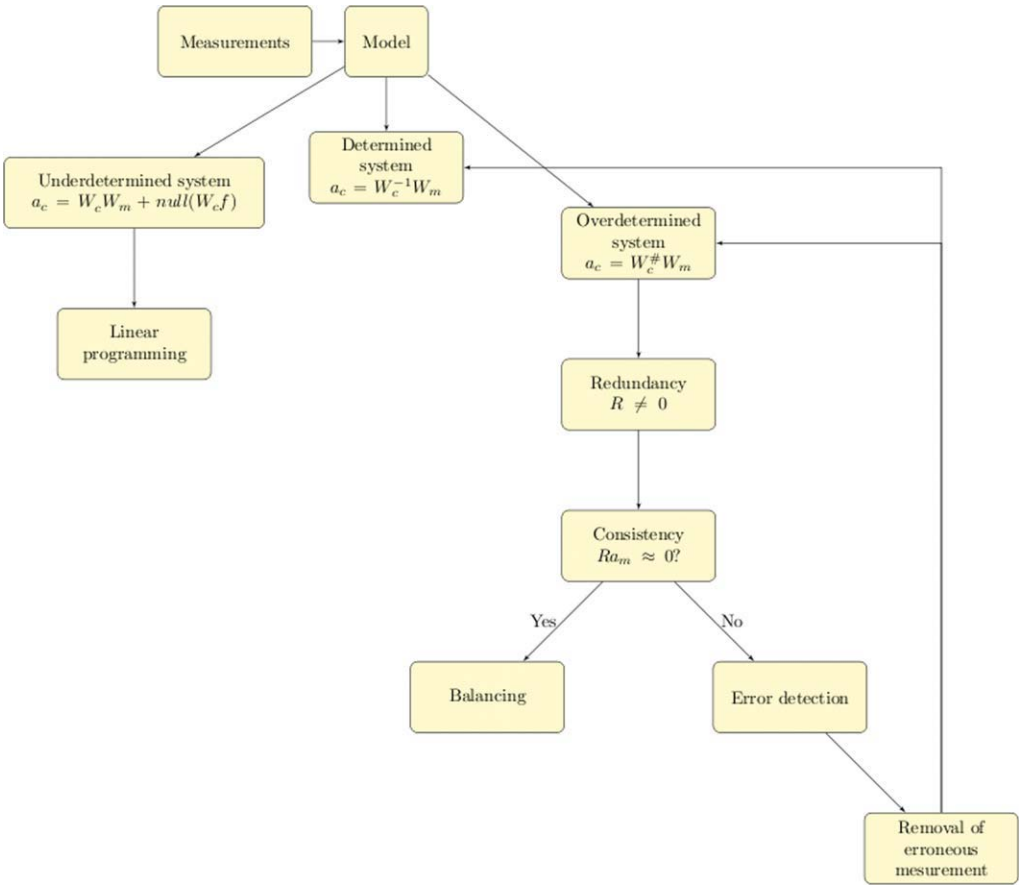


Figure 5. Stoichiometric modeling.

in the null space of W_c. The number of independent null space vectors is equal to:

$$n - \text{rank}(W_c) \tag{8}$$

with n the number of fluxes to be calculated.

Metabolic flux analysis

If Eq. 8 = 0 the system is determined and has one unique solution:

$$a_c \cong -W_c^{\#} W_m a_m \tag{9}$$

In the case of a determined system without extra measurements (i.e., not redundant), $W_c^{\#}$ is simply the inverse of W_c. Substituting Eq. 9 in Eq. 5 now yields:

$$W_m a_m + W_c\left(-W_c^{\#} W_m a_m\right) = 0 \tag{10}$$

Or rewritten, since Eq. 7:

$$(W_m - W_c^{\#} W_c W_m) a_m = 0 \tag{11}$$

When the system is (partially) overdetermined, the redundant measurements, which are specified by the so-called redundancy matrix: $W_m - W_c^{\#} W_c W_m$ in Eq. 11, can be used for statistical testing and error analysis. van der Heijden et al. (1994)[65] introduced a method for error detection and analysis which allows stating whether a model is consistent. If this is not the case, this error analysis method can be used to identify erroneous measurements. Removing these erroneous measurements from the data set improves the chance of a consistent result. An overview is given by Lequeux et al. (2006).[61]

Pros and cons for the rational design of bioprocesses—Though metabolic flux analysis (MFA) merely yields a snapshot of the metabolic state in a particular condition, it might be of some significance to steer the process of metabolic engineering because it allows identifying principal nodes. These principal nodes, which are characterized by significant changes in flux partitioning under different conditions, should be regarded as potential bottlenecks.[20]

It should be clear that due to the large variety of metabolic pathways, e.g., parallel pathways, reversible reactions, and cycles, the system is in general under determined (Eq. 8 > 0). For example, genome-scale models have been constructed, that are typically useful for the design of minimal media, e.g., for *Escherichia coli* (931 reactions),[52] *Saccharomyces cerevisiae* (1,175 reactions),[66] *Helicobacter pylori* (388 reactions),[67] and *Neisseria meningitidis* (496 reactions).[68] Such genome scale models contain all known reactions, formation, degradation, transport, and cellular utilization gathered from databases and the literature.

Starting from such genome-scale model the modeller can opt/has to reduce the metabolic network in order to get an identifiable system by incorporating as much knowledge, e.g., prior knowledge about the flux size, and data as possible. Examples of such data are: metabolomic data, as these data yield thermodynamic information $\Delta_r G'^{o}$ and consequently information about the reversibility and irreversibility of certain reactions,[69,70] labeled metabolomic data as these data yield information on split ratios,[39,71,72] and transcriptomic data,[61,73–75] through the incorporation of additional

Table 4. Objective Functions Used in FBA

Objective Function	Mathematical Analogue	Underlying Assumption
Maximize biomass[44,76]	$\max \frac{v_{\text{biomass}}}{v_{\text{substrate}}}$	The cell optimizes its biomass yield
Maximize ATP yield[33]	$\max \frac{v_{\text{ATP}}}{v_{\text{substrate}}}$	The cell functions maximally energetic efficient
Minimize substrate consumption[77]	$\max \frac{v_{\text{substrate}}}{v_{\text{biomass}}}$	The cell uses the substrate most efficiently to produce biomass
Minimize reaction steps[32]	$\min \sum_{i=1}^{n} v_i, v_i = \{0,1\}$	The cell optimally uses the cellular resources for enzyme synthesis to install the network
Maximize ATP per reaction step[34]	$\min \frac{v_{\text{ATP}}}{\sum_{i=1}^{n} v_i}, v_i = \{0,1\}$	The cell produces as much ATP as possible
Minimize redox potential[78]	$\min \frac{v_{\text{NADH}}}{v_{\text{biomass}}}$	The cell uses the available energy as efficient as possible
Minimum norm[68]	$\min \sum_{i=1}^{n} v_i^2$	The cell functions with maximal enzymatic efficiency for cellular growth

constraints for the metabolic network, e.g., presence of an enzymatic conversion, to reduce the uncertainty about the obtained flux distribution. The consequences of the choice metabolic network on the calculated flux distribution is however not that well studied.

Flux balance analysis

If Eq. $8 > 0$ the system is under determined and no unique solution exists. The question then is which of the feasible metabolic states is manifested under that condition. Flux balance analysis (FBA)[33,76] postulates that a metabolic system exhibits a metabolic state that is optimal under some criteria. This objective is often expressed as a linear combination of the fluxes contained in R. The model can then be formulated as a linear or nonlinear programming problem as follows:

$$\max J \tag{12}$$

subject to:

$$0 \cong S \times R - b \tag{13}$$

$$\alpha_i \leq r_i \leq \beta_i \tag{14}$$

where J is the objective function and the boundaries α_i and β_i represent known constraints on the minimum and maximum flux values. Popularly used objective functions are listed in Table 4.

Pros and cons for the rational design of bioprocesses—Flux balance analysis allows to rapidly evaluate the effects of certain genetic modifications, i.e., by adding an additional constraint ($r_i = 0$) to the metabolic network the effect of a gene knock-out can be evaluated. The effect of a gene knock-in on the other hand can be evaluated by adding the reaction stoichiometry to the set of model equations. To automate this quest for the optimum a whole set of local and global optimization algorithms, e.g., simplex,[79] genetic algorithms,[80] SIMPSA,[81] … is available to in silico evaluate the effect of multiple gene knock-outs and knock-ins.

Though many objective functions have been used, the optimization of ATP production and the optimization of growth have been shown to comply best with experimental observations[34] in many micro-organisms. One should however be aware that the cell's goal may be different under certain conditions. The applications of FBA have been many (Table 1) and the in silico metabolic constraint predictions can be used to optimize the behavior of interesting mutants.

Tools like OptKnock,[82] which identifies multiple gene deletion combinations that maximally relate growth objectives with the production of the desired compound, and OptGene,[11] which facilitates the search for multiple gene deletion combinations using genetic algorithms for complex metabolic networks, were developed. However, the genetic toolbox to modify cellular metabolism to oblige a cell to produce the compound of interest consists of much more than knock-out procedures.[22]

Despite success stories of FBA use to identify targets for modification, there have also been many false positive targets identified by these models. It is still unclear whether the well-established technique of stoichiometric modeling is fully apt to steer the process of metabolic engineering, since the kinetics and the regulation of the enzymatic reactions are not accounted for.[12,53] Especially for the optimization of the production of metabolites in primary metabolism that are subject to severe (redox) constraints, stoichiometric modeling has been shown to be useful. It is less so for the optimization of minor pathways.[54]

FBA heavily relies on prior knowledge (through the constraints introduced in Eq. 14), but at present the knowledge on the regulatory mechanisms is still lacking and fragmentary.[83,84] For example, the determination of intracellular fluxes depends heavily on the correctness of the assumed stoichiometry, the determination of fluxes in complex networks often requires the inclusion of NADH and NADPH balances, which are subject to controversial debate.[85] The assumption on the energy metabolism, i.e., Y_{ATP}, stoichiometry of the oxidative phosphorylation (P/O-ratio). These stoichiometric parameters often originate from WT strains and continuous cultures and may not hold true in cases of highly engineered strains.[86] In addition, in some cases no unique optimum exists and consequently many metabolic states may result in the same optimal behavior.[87]

To reduce the solution space genetic regulation, which yields additional constraints, can be taken into account as well.[88] These constraints are introduced in a Boolean manner, i.e., a reaction can be present (1) or absent (0). The genetic regulation is however much more subtle than the on/off regulation of gene expression so one may wonder, though good results have been obtained to predict diauxic growth,[89] whether equally good results would have been obtained in the case of simultaneous substrate utilization.

Another option to reduce the solution space is the incorporation of energetic constraints, next to constraints on certain reaction rates. To this end, Beard et al. (2002)[35] and Nagrath et al. (2007)[90] introduced energy balances to ensure the thermodynamic feasibility of the solution, by making use of the first law of thermodynamics (conservation of energy).

For all reactions the reaction potential can be calculated:

$$\Delta\mu = S^{\mathrm{T}}\mu \tag{15}$$

wherein μ represents a m-dimensional vector of chemical potentials, $\Delta\mu$ a n-dimensional vector of reaction potentials, and S the stoichiometric matrix. The energy balance, is then given by:

$$K^{\mathrm{T}}\Delta\mu = 0 \tag{16}$$

with K the null space matrix of S. This constraint ensures that the sum of reaction potentials around any cycle of reactions equals zero.[35]

The second law of thermodynamics yields on the other hand that a reaction only can take place in the direction of negative Gibbs free energy of that reaction, $\Delta_r G$, which is given by:

$$\Delta_r G = \Delta_r G^o + RT \ln \prod x_{M_i}^{s_{M_i j}} \tag{17}$$

with x_{M_i} the concentration of metabolite M_i, $s_{M_i,j}$ its stoichiometric coefficient for reaction j, and $\Delta_r G^o$ the standard Gibbs free energy of the reaction.[70,91] The Gibbs free energies of reaction and the flux direction can then be used to identify thermodynamically feasible ranges for the Gibbs free energies of reaction and for the concentration of nonmeasured metabolites.[70]

However, it is not because a micro-organism has the genetic potential that it will just like that perform optimally, i.e., artificially created mutants are generally not subject to the same evolutionary pressure that shaped the wild type.[12,36,87] The method of minimization of metabolic adjustment (MOMA) attempts to deal with this issue. Instead of maximizing biomass production the knock-out mutant, KO, is believed to initially remain as close as possible to the wild type optimum, *WT*, in terms of flux values.[36] The objective function then becomes:

$$\min D(R_{\mathrm{WT}}, R_{\mathrm{KO}}) \tag{18}$$

with

$$D(R_{\mathrm{WT}}, R_{\mathrm{KO}}) = \sqrt{\sum_i^n (r_{WT} - r_{KO})^2} \tag{19}$$

Another constraint-based method for predicting the metabolic steady state after gene knock-outs is Regulatory on/off minimization (ROOM).[37] This method aims to minimise the number of significant flux changes with respect to the wild type. Whereas MOMA provides more accurate predictions for the mutant directly after the genetic perturbation, ROOM and FBA more successfully predict the mutant's behavior after adaptation to the new genotype.[37]

An intracellular pseudo steady state can also be assumed under certain dynamic conditions, because of the relatively small time constants of cellular processes, e.g., mass action and the metabolic adaptation to novel conditions, in comparison with processes affecting the observed environmental conditions. Dynamic FBA, using an instantaneous objective function, can then be used. For example, to predict the diauxic shift in *Escherichia coli*.[89,92,93] Lee et al. (2008)[94] introduced integrated dynamic FBA (idFBA) that dynamically simulates cellular phenotypes arising from integrated networks. This framework uses an integrated stoichiometric reconstruction of signaling, metabolic, and regulatory processes. Pseudo steady state conditions are assumed for fast reactions and slow reactions are incorporated into the stoichiometric formalism in a time-delayed manner. This pseudo steady state assumption thus allows the use of FBA for monitoring and control of complex bioprocesses.

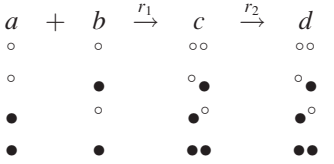
Carbon modeling

The increasing realization that, using the aforementioned techniques, large parts of the metabolic network, e.g., parallel metabolic pathways, metabolic cycles, and bidirectional reaction steps cannot be resolved, has boosted the application of carbon modeling. The data collected, measured by ^{1}H-NMR, ^{13}C-NMR, or MS instruments, during a carbon labeling experiment can then help to elucidate the flux distribution.

These carbon models make use of the concept of isotopomers. The isotopomer distribution in a metabolite pool with n carbon atoms can unambiguously be characterized by 2^n numbers, each representing the relative amount of one specific isotopomer. The labeling pattern of these isotopomers is binary encoded: labeled and nonlabeled carbon atoms are represented as ones and zeros, respectively. For a two-carbon molecule c the isotopomer distribution vectors becomes:

$$I_{\mathrm{c}} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} c_{00} \\ c_{01} \\ c_{10} \\ c_{11} \end{pmatrix} \tag{20}$$

With $\sum_{i=0}^{2^n-1} c_{(i)} = 1$. The fate of isotopomers is encoded in isotopomer mapping matrices (IMMs) that hold information on which product isotopomers evolve from which substrate isotopomers in a biochemical reaction.[95] One IMM is defined for each pair of substrate and product molecules in a biochemical conversion. Isotopomer distribution vectors (IDVs) of product molecules can be calculated from the IDVs of substrate molecules by simple matrix multiplication.[38] Consider the following reaction pathway, r_1 converts the one-carbon molecules a and b into c and r_2 converts c into d:



(21)

where labeled and nonlabeled carbon atoms are represented by the filled spheres (●) and the empty spheres (○), respectively. In matrix notation the isotopomer balance for metabolite c then becomes:

$$\frac{d\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix}}{dt} = r_1 \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}}_{IMM_{A>C}} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \otimes \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{IMM_{B>C}} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} - r_2 \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{22}$$

Assuming an isotopic and metabolic stationary state Eq. 22 can be rewritten as:

$$\begin{aligned} c_0 &: a_0 * b_0 r_1 - c_0 r_2 = 0 \\ c_1 &: a_0 * b_1 r_1 - c_1 r_2 = 0 \\ c_2 &: a_1 * b_0 r_1 - c_2 r_2 = 0 \\ c_3 &: a_1 * b_1 r_1 - c_3 r_2 = 0 \end{aligned} \tag{23}$$

The measured multiplet patterns which are determined by the ^{13}C isotopes in the molecule can be compared with the simulated signals that can be computed from the isotopomer distributions (Figure 6). However, since the relationship between the labeling state and the intracellular flux distribution is complex and nonlinear it is practically impossible to find an analytical expression for the intracellular reaction rates as functions of measurement data. Hence, the flux distribution has to be computed by minimizing the sum of squares of the difference between the measurements and the simulated data.[38]

Different iterative numerical solution approaches have been proposed to solve the isotopomer balance equations due to their apparent nonlinear structure and high dimensionality.[38] However, after a suitable variable transformation, the equations can be solved explicitly. Hereto, the concept of cumomers has been introduced. Cumomers, or cumulated isotopomer fractions, are the sum of certain isotopomer factions of a metabolite. The weight 0 (Eq. 24), 1 (Eqs. 25 and 26), and 2 (Eq. 27) cumomers of metabolite c are defined as:

$$c_{xx} = \sum_{i,j=0}^{1} c_{ij} = 1 \tag{24}$$

$$c_{1x} = \sum_{i=0}^{1} c_{1j} \tag{25}$$

$$c_{x1} = \sum_{i=0}^{1} c_{i1} \tag{26}$$

$$c_{11} = c_{11} \tag{27}$$

Or rewritten in matrix notation:

$$\begin{pmatrix} c_{xx} \\ c_{x1} \\ c_{1x} \\ c_{11} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_{00} \\ c_{01} \\ c_{10} \\ c_{11} \end{pmatrix} \tag{28}$$

By summing up the equations of all isotopomers belonging to the a certain cumomer the cumomer balances are computed:

$$\begin{aligned} c_{xx} &: a_x * b_x r_1 - c_{xx} r_2 \\ c_{x1} &: a_x * b_1 r_1 - c_{x1} r_2 \\ c_{1x} &: a_1 * b_x r_1 - c_{1x} r_2 \\ c_{11} &: a_1 * b_1 r_1 - c_{11} r_2 \end{aligned} \tag{29}$$

For the metabolic network, this system of cumomer balances can then be rewritten in matrix formulation:

$$\begin{aligned} 1 &= {}^0x \\ 0 &= {}^1A(v)\,{}^1x + {}^1b(v) \\ 0 &= {}^2A(v)\,{}^2x + {}^2b(v, {}^1x) \\ 0 &= {}^3A(v)\,{}^3x + {}^3b(v, {}^1x, {}^2x) \end{aligned} \tag{30}$$

with ${}^ix$ a vector containing the cumomers belonging to weight i. In this way, the system can be solved for each weight level independently. Hence, a cascade of linear equations can be solved.

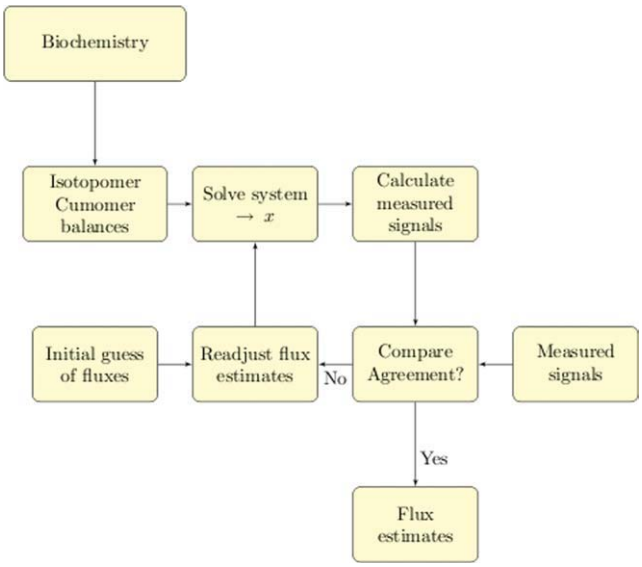


Figure 6. Solving a carbon model.

In general, a large number of isotopomer or cumomer balances needs to be solved.[39] To deal with this issue, Antoniewicz et al. (2007)[40] introduced the concept of elementary metabolite units (EMU). Such an EMU of a compound is a moiety comprising any distinct subset of the compound's atoms. The developed framework is based on a highly efficient decomposition method that identifies the minimum number of information needed to simulate isotopic labeling within a reaction network, using the knowledge of atomic transitions occurring in the network. Hence, it requires significantly less computation time, enabling the analysis of labeling by multiple tracers, e.g., ^2H, ^{13}C, and ^{18}O.

Pros and cons for the rational design of bioprocesses—To compute the flux distribution, carbon modeling requires a nonlinear optimization, which minimizes a quadratic error function. In such cases, one might, however, get trapped in local minima, and consequently no guaranty can be given that the global optimum has been found. In addition, the dependence of the solution on the initial settings (chosen rates) should be carefully checked, e.g., by performing Monte Carlo simulations. The consistency check of the ultimate solution remains rather primitive, because in general merely extracellular exchange rates are used to verify the calculated flux distribution.[96,97]

The determination of intracellular fluxes depends heavily on the correctness of the assumed stoichiometry, the determination of fluxes in complex networks often requires the inclusion of NADH and NADPH balances, which are subject to controversial debate.[85] The assumption on the energy metabolism, i.e., Y_{ATP}, stoichiometry of the oxidative phosphorylation (P/O-ratio). These stoichiometric parameters often originate from WT strains and continuous cultures and may not hold true in cases of highly engineered strains.[86] Using ^{13}C tracer experiments flux analysis can be performed on the basis of only well established stoichiometric equations and measurements of the labeling state of the intracellular metabolites.[85,86]

One of the main assumptions of carbon modeling is that the network stoichiometry is complete. However, this assumption may not be as trivial as it seems, e.g., many

enzymes have a much broader substrate range than considered[98] and the modelled stoichiometric network is only a gathering of the most important metabolic reactions. Metabolite channelling is another issue which should be carefully considered in this context. This is a mechanism whereby the product of an enzymatic reaction is transferred to the next enzymatic reaction without mixing with the metabolite pool.[98]

In addition, despite the increased information content of the gathered data the parameter sensitivity and identifiability may remain low. Hence, the need for optimal experimental design.[99] Several software tools for carbon modeling are (freely) available, e.g.[100,101]

Because of the cost of labeled substrates there is a clear tendency towards miniaturization[102] and reducing the duration of the labeling experiment. Hence, the interest in isotopic and/or metabolic instationary labeling experiments. In stead of studying the isotopic steady state, which takes a long time before it is established, the transient can be studied both under metabolic stationary and instationary conditions.

Initially, the labeling pattern of the amino acids was measured, the aforementioned evolutions also resulted in the measurement of the labeling pattern of metabolites in the glycolysis and TCA cycle. In this case, one has to be careful since many non-obvious reactions do intervene, e.g., synthesis and degradation of macromolecular compounds (storage compounds and protein turn) and transamination reactions have to be taken into account.[103] In addition, these reactions significantly increase the time necessary to reach isotopic stationary state for these glycolytic and TCA cycle metabolites.

The value of such isotopic instationary data is huge, e.g., the quality of the parameter estimates of kinetic models will be significantly improved because the information content of the data collected during ^{13}C-labeling experiments both under stationary and instationary metabolic state is much larger than when no labeling is used.[104,105] These calculations are however computationally quite intensive and methods to render these calculations more feasible will have to be developed, e.g.[106]

Mechanistic and Approximative Modeling

Because of the aforementioned limitations of stoichiometric modeling, kinetic equations have been introduced in metabolic models. The general form of the mass balances of extracellular and intracellular metabolites is now given by Eq. 31 and Eq. 32, respectively:

$$\frac{dx_{S_i}}{dt} = D\left(x^0_{S_i} - x_{S_i}\right) - \frac{x_X}{\rho_X}\sum_j s_{M_i j} r_j \tag{31}$$

$$\frac{dx_{M_i}}{dt} = \sum_j s_{M_i j} r_j - \mu x_{M_i} \tag{32}$$

with x_{M_i} and x_{S_i} the concentration of an intracellular metabolite M_i and an extracellular metabolite S_i, respectively, $s_{M_i j}$ is the stoichiometric coefficient of metabolite M_i in reaction j, r_j the rate of reaction j, ρ_X the specific weight of biomass, x_X the biomass concentration, D the dilution rate, x^0_S the concentration of an extracellular metabolite S in the feed, and μ the specific growth rate. Note that x_S is expressed per reactor volume whereas x_M is expressed per cell volume. The term μx_M in the mass balances of the intracellular metabolites represents the dilution effect due to growth.

In mechanistic dynamic metabolic modeling, one can resort to complex mechanistic equations determined in vitro to describe the rate equations r_j in Eqs. 31 and 32,[21,41,107,108] e.g., the popularly used Michaelis Menten kinetics. These complex in vitro derived kinetics typically contain many hard to estimate parameters (K_{PTS,a_i}, $r_{\max,\mathrm{PTS}}$, and $n_{\mathrm{PTS,G6P}}$):

$$r_{\mathrm{PTS}} = r_{\max,\mathrm{PTS}} \frac{x_{\mathrm{GLU}} x_{\mathrm{PEP}}}{x_{PYR}\left(K_{\mathrm{PTS},a_1} + K_{\mathrm{PTS},a_2}\frac{x_{\mathrm{PEP}}}{x_{\mathrm{PYR}}} + K_{\mathrm{PTS},a_3} x_{\mathrm{GLU}} + x_{\mathrm{GLU}}\frac{x_{\mathrm{PEP}}}{x_{\mathrm{PYR}}}\right)\left(1 + \frac{x_{\mathrm{GSP}}^{n_{\mathrm{PTS,G6P}}}}{K_{\mathrm{PTS},a_4}}\right)}$$

In approximative modeling, one can resort to nonmechanistic kinetics to describe the rate equations r_j in Eqs. 31 and 32, e.g., the generalized mass action type power law approximation (GMA) (Eq. 33),[109] the loglinear approximation (Eq. 34),[110,111] the linear in metabolite and enzyme levels (Eq. 35),[112] and the linlog approximation (Eq. 36).[113,114]

$$\ln\left(\frac{r_j}{J^0}\right) = \ln\left(\frac{x_{\mathrm{E}}}{x^0_{\mathrm{E}}}\right) + \sum_{i=1}^{n} \varepsilon^0_{M_i} \ln\left(\frac{x_{M_i}}{x^0_{M_i}}\right) \tag{33}$$

$$\frac{r_j}{J^0} - 1 = \ln\left(\frac{x_{\mathrm{E}}}{x^0_{\mathrm{E}}}\right) + \sum_{i=1}^{n} \varepsilon^0_{M_i} \ln\left(\frac{x_{M_i}}{x^0_{M_i}}\right) \tag{34}$$

$$\frac{r_j}{J^0} - 1 = \left(\frac{x_{\mathrm{E}}}{x^0_{\mathrm{E}}} - 1\right) + \sum_{i=1}^{n} \varepsilon^0_{M_i} \left(\frac{x_{M_i}}{x^0_{M_i}} - 1\right) \tag{35}$$

$$\frac{r_j}{J^0} = \left[\frac{x_{\mathrm{E}}}{x^0_{\mathrm{E}}}\right]\left(1 + \sum_{i=1}^{n} \varepsilon^0_{M_i} \ln\left(\frac{x_{M_i}}{x^0_{M_i}}\right)\right) \tag{36}$$

where the superscript ^{0} stands for the operating point and x_{E} the enzyme concentration, ε^0_{Mi} an elasticity coefficient, and J^0 the steady-state flux. The applied equations are not as complex as mechanistic rate equations and contain less parameters to approximate the true kinetics. The rationale behind this approach is that metabolic redesign does not require detailed mechanistic models because of the concept of homeostasis, which implies that the micro-organism keeps its intracellular metabolite levels approximately constant.[115–117] In other words, the extrapolation range of the kinetic metabolic model does not need to be very large, as far as metabolite levels are concerned. This reasoning suggests that one can safely apply approximative kinetic equations instead of the detailed mechanistic ones that are valid over a wide range of concentration levels.

The metabolic regulation of enzyme activity is, however, not always that easily to be captured using the aforementioned kinetic structures, e.g., to describe the biphasic effect of ATP on phosphoenolpyruvate carboxykinase (PCK) a fuzzy-logic approach was used by Lee et al. (1999).[118]

Pros and cons for the rational design of bioprocesses—Using mechanistic and approximative models, metabolic control analysis can be used to identify the response and flux control coefficients of the metabolic network[110] and consequently the metabolic engineering targets. Such a flux control coefficient is a measure of how a change in the level of an enzyme affects the steady-state flux through that particular pathway:

$$C_i^F = \frac{\partial J^0}{\partial x_{E_i}} \frac{x_{E_i}}{J} = \frac{\partial \ln J^0}{\partial \ln x_{E_i}} \tag{37}$$

Hence, it is a measure of the degree of control exerted by this enzyme on the steady-state flux.[119]

One should however be aware that these symbolically[110,119] or numerically[120] determined coefficients are merely local properties and consequently the envisaged flux change may result in a completely different metabolic state, for which these local findings may be no longer valid. However, a whole set of local and global optimization algorithms, e.g., simplex,[79] genetic algorithms,[80] SIMPSA,[81] …, is available to in silico evaluate the modification of enzymatic properties and levels. These findings may then be used to knock-out and knock-in genes, or to directedly evolve enzymes in order to alter their affinity for the substrate and/or to modify product inhibition characteristics.

The enormous variety of well regulated metabolic pathways impedes a thorough understanding of the regulation of microbial processes on the metabolomic, proteomic, transcriptomic, and genomic level in a qualitative and quantitative way. Such understanding would be beneficial for the rational design of bioprocesses, as a genetic or environmental disturbance in one part of metabolism can trigger a series of reactions on all levels of regulatory control and in all parts of metabolism.[30] Hence, in many applications, e.g., metabolic engineering, "whole cell modeling" is probably the way to go.[121,122]

However, since the knowledge about the transcriptional and translational regulation is still fragmentary, the state of the art dynamic metabolic models typically focus on the metabolome, assuming constant proteomic levels. In view of the extrapolation capacity of these models this is a drawback. Hence, in order to comply with the assumption of steady-state proteome, data for parameter identification have to be collected during a relatively short period after perturbation, typically within 0.2–180 s.[108,123–125]

Thus, the calibration of such kinetic models requires highly dynamic experiments starting from well-defined cultures, with constant proteomic levels. Hereto, well-defined chemostat or batch cultures, which are in a pseudo steady-state, are typically perturbed. Then the metabolic state of the cell prior to the perturbation, i.e., pseudo steady-state flux distribution, can be (unambiguously) determined by making use of a pseudo steady-state model. However, because kinetic metabolic models typically zoom in on a limited part of the microbial metabolism the evolution of the flux distribution under the applied dynamic conditions is more uncertain.

Because such kinetic metabolic models do not consider the whole metabolic network (as they typically zoom in on only a limited part of the microbial metabolism) they typically contain a number of fluxes toward parts of the metabolism which one is not primarily interested in. However, when such a model contains two or more of those reactions, that convert metabolites into nonconsidered cellular fractions. This will result in the untractable disappearance of carbon out of the model, which will create some uncertainty about the flux distribution under the dynamic conditions, because only indirect, secondary information about the dynamic evolution of the flux distribution after a perturbation of the metabolism is collected. Only having the information of the evolution of unlabeled metabolite concentrations is insufficient for these aims. Thus, in contrast to steady-state modeling, where mass balances are essential to verify the accuracy of the calculated fluxes, this check is not performed in most dynamic metabolic models,[21,107,108,126] as the size of the out fluxes is not exactly known. It should, however, be clear that modeling the whole metabolism would be a daunting task as well, because when a perturbation passes through the metabolic network it broadens and dampens out and the information content of data collected further on in the network is limited.

To reduce this uncertainty the cofactors might be used as "closure terms," e.g., the generation of NADPH, might be a good indicator for the flux through the pentose phosphate cycle. However, it should be clear that these closure terms are weak as cofactors intervene in many reactions, which are also perturbed during a pulse experiment. In addition, modeling these cofactors dynamically is not easy at all because this approach is hampered, for instance, by the inability to explain the short-term reduction in the pool size of the adenine nucleotides (AXP) after a glucose pulse.[107,108] At present, it is still unclear what is/are the cause(s) of this reduction (adaptation would only be responsible for 15% of this gap,[127] formation of adenylated compounds, e.g., ADP-glucose, excretion of cAMP, …).

Therefore, some researchers opt to describe the evolution of the cofactors as time dependent functions,[107] which results in a model that is no longer useful for extrapolation. Not taking the cofactors into account "mechanistically" thus results in a limited usefulness of the resulting model. Then, also assumptions have to be made about the evolution of the flux distributions during the transient but it is questionable whether these hold.

To reduce the uncertainty, one could gather a lot of data both under steady-state and dynamic conditions, e.g., by perturbing the microbial cells with different substrates. However, such efforts have been limited thus far.[123,128]

The use of dynamic labeling data[129] allows as well to reduce the degrees of freedom related to the metabolic fluxes, also under dynamic conditions. However, the huge variety of biochemical pathways will render such an exercise tricky, as the chosen metabolic network will influence the calculated flux distribution.[130]

In addition, one should be aware that a lot of challenges still remain in the field of analytical methods, since the accurate determination of the intracellular metabolites is a considerable task as well, due to, e.g., leakage and their low concentrations.[131–134] For example, whereas the expected (equilibrium) ratio of the concentrations of glucose-6-phosphate [G6P] and fructose-6-phosphate [F6P], i.e., $\frac{[F6P]}{[G6P]} \cong 0.25$,[107,128,135,136] Bucholz et al. (2001)[137] found for this ratio $\frac{[F6P]}{[G6P]} \cong 0.88$.

Another issue is that the state of the art dynamic metabolic models either rely on in vitro determined kinetic equations or are based on approximative kinetics[107,114,136] and the consequences of a potentially erroneous model structure are not well known. With respect to the in vitro determined kinetic equations it is doubtful whether the kinetics are valid under in vivo conditions, as these kinetics are obtained using purified enzymes studied out of context.[136,138]

The variety of well—regulated metabolic pathways also impedes a thorough understanding of the regulation of microbial processes, e.g., the relative importance of the flux through pyruvate oxidase compared to the flux through pyruvate dehydrogenase is not that clear.[139,140] Another example is the jumble of reactions around the PEP-pyruvate-oxaloacetate node. Their regulation and importance under one or the other condition is still not that well studied.[141–143] The inability to properly describe the dynamics of phosphoenolpyruvate (PEP) during the observation window of a perturbation experiment,[10,107] even though it is a key metabolite in the primary metabolism, is the perfect illustration that setting up a metabolic model in a proper way will be demanding both for modellers and for experimentalists.

To deal with the limited predictive power of dynamic models, a number of researchers have attempted to incorporate genetic regulation in their kinetic model.[42,43] One should however be aware that these examples are merely academic examples, because or these models consider only a small sub network, e.g., oxygen regulation of cytochrome production in *Escherichia coli*,[43] neglecting all interactions between the small network and the rest of the cellular network, which is thus studied completely out of context, or do consider the whole cell but lack sufficient resolution, e.g., diauxic growth on glucose and lactose,[42] rendering them equally unfit to really identify metabolic engineering targets. At present mechanistic information is not available for large networks and hence a mechanistic description of the genetic regulation is intractable and impossible. Models, whether they are approximative or mechanistic, can be useful to identify bottlenecks[110,115,120,144] in metabolism and consequently could steer the process of metabolic engineering. However, since enzyme levels are not taken into account nor the influence of a genetic intervention on the metabolism, it should be clear that the extrapolation power of such models remains limited.

In addition, one should be aware that the metabolic data gathered during typically only one perturbation experiment are highly correlated, this is both due to the nature of some of the metabolic conversions, e.g., pseudo equilibrium reactions (isomerization, …), the multiple interconversion possibilities between different metabolites, and the way the perturbation propagates throughout the metabolic network. These highly correlated metabolite data renders the identification of the model parameters a truly daunting. Because, for the poorly identifiable model,[41] many (completely) different parameter sets will give almost identical fits to the limited data, in terms of information content, available to identify the numerous model parameters, i.e., the equifinalty problem.[145] This will most likely lead to erroneous model predictions.

Cybernetic Modeling

At present, one can not see the wood for the trees as the knowledge on the regulatory mechanisms is lacking and fragmentary.[83,84] To partially circumvent this knowledge gap, the cybernetic framework can be used, since microbial species, that is, those that have undergone the process of evolution, strive to regulate their metabolism in an optimal manner.[87,146] This reasoning is the rationale of the cybernetic school of thought: a micro-organism tries to optimize its behavior, e.g., with respect to growth or substrate uptake. This is achieved by allocating the limited resources a micro-organism disposes of to these competing enzymes yielding the optimal performance.[147–149] To this end, cybernetic variables were introduced into kinetic models with the aim of substituting the unknown mechanistic details of the cell's regulatory architecture by an objective function that incorporates the fact that the metabolism of a micro-organism operates with a specific overall goal, such as the optimization of growth.

Initially, the value of the cybernetic approach was demonstrated using relatively simple examples, typically situated in the domain of bioprocess control. In these cases, some lumped pathways competed with each other for the available resources, e.g., simultaneous and sequential substrate utilization[44] and single-substrate growth.[150–152] Then the cybernetics units could readily be identified. A cybernetic unit is a cluster of enzymes that compete with each other for the same pool of resources.

Over time more challenging "proofs of principle" were chosen, e.g., in view of metabolic engineering of a production host,[45,149,153] and the model's complexity increased. More complex networks, without lumping were considered,[154,155] but then a jumble of cybernetic units could be identified and the corresponding cybernetic variables had to be derived from the control laws. As a result, the choice of the cybernetic units became less straightforward, even quite arbitrary, and the library of cybernetic units had to be extended (convergent, divergent, linear, and cycles).[148,149]

To overcome this, a more general framework was developed (Figure 7), based on the principles of optimal control theory.[45] Optimal control theory is a mathematical optimization method for deriving control policies. It aims to find a control law for a given system such that a certain optimality criterion is achieved. In general, such a control problem includes a gain function and a cost function relating state and control variables. An optimal controller is a set of differential equations describing the paths of the control variables that maximise the performance function. Rephrasing this in the context of a micro-organism, the cost becomes, e.g., the pool of amino acids a micro-organism needs to invest for the production of a particular enzyme, and the cell's gain could be merely growth.

From a cybernetic point of view the microbial system can be represented by following set of differential equations[45]:

$$\frac{dx_{S_i}}{dt} = D\left(x^0_{S_i} - x_{S_i}\right) - \frac{x_X}{\rho_X}\sum_j s_{M,ij} r_j v_j \tag{38}$$

$$\frac{dx_{M_i}}{dt} = \sum_j s_{M,ij} r_j v_j - \mu x_{M_i} \tag{39}$$

$$\frac{dx_{E_i}}{dt} = \alpha^\star_{E_i} + \alpha_{E_i} u_i - \beta_{E_i} x_{E_i} - \mu x_{E_i} \tag{40}$$

with exception of α^* which represents the constitutive synthesis rate of enzyme E_i, α the synthesis rate of enzyme E_i,

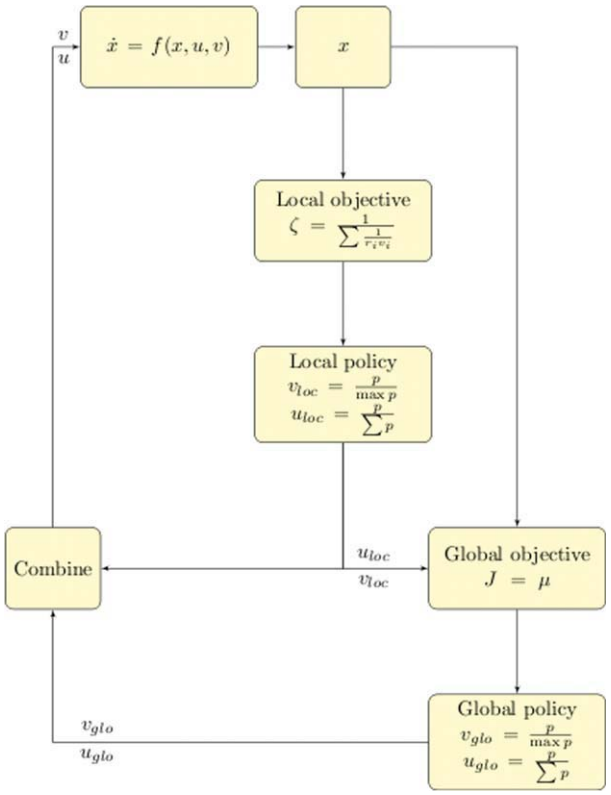


Figure 7. The cybernetic framework.

On a local control level compete the various reactions/enzymes of an EM with each other for the limited resources, which are allocated in such a way that a steady throughput is ensured (harmonic mean). On a global control level the various EMs compete with each other for the available resources. On a global control level the resources are allocated to the various EMs in order to maximise growth.

and β the specific degradation rate of enzyme E_i all variables and parameters have been introduced in "Mechanistic and Approximative Modeling" Section. This system is subject to regulatory control both at the level of enzyme synthesis and enzyme activity. These inputs are accounted for by introducing the control vectors u and v in Eqs. 39 and 40, respectively, which specify how the resources are allocated among the various alternatives/enzymes.

For clarity, the cybernetic variable controlling enzyme activity, v, will be discarded. It is assumed that the cell allocates its resources in such a way that the performance function J is maximized. This can be described by optimal control theory:

$$\begin{aligned} &\max J \\ &\text{subject to} : \dot{x} = f(x,u) \end{aligned} \tag{41}$$

Computing the optimal control is numerically quite demanding. Assuming however that regulatory decisions are made at each instant based on the projected system response over a short time interval Δt, the system can be approximated by linearization (working point u^0):

$$\Delta\dot{x} = A\Delta x + B_u \Delta u + f\left(x(t), u^0\right) \tag{42}$$

$$A = \left.\frac{\partial f(x,u)}{\partial x}\right|_{x,u^0} \tag{43}$$

$$B_u = \left.\frac{\partial f(x,u)}{\partial u}\right|_{x,u^0} \tag{44}$$

The change in model performance ΔJ over the system's planning window Δt then becomes:

$$\Delta J = q\Delta x(t+\Delta t) - \frac{1}{2}\int_t^{t+\Delta t} \left(u^T \sigma_u u\right) d\tau \tag{45}$$

$$q = \frac{\partial \phi(x(t))}{\partial x} \quad \Delta J = J(t+\Delta t) - J(t). \tag{46}$$

in which the function $\phi(x(t))$ is the metabolic objective function of the system and σ_u a parameter that scales the cost associated with resource investment. The values for u and v are obtained by solving this constrained optimal control problem. By solving the Hamiltonian and deriving the Karash- Kun-Tucker conditions,[156] one finally finds for the control law ruling enzyme synthesis:

$$u_i = \frac{\max(p_i, 0)}{\sum_{i=1}^{n} \max(p_i, 0)}$$

with p_i the return on investment for resources allocated to the *ith* enzyme. Akin, the control law ruling enzyme activity can be derived, yielding:

$$v_i = \frac{\max(p_i, 0)}{\max_n(p_n)}$$

Young (2005)[45] introduced a global and a local control level in the developed framework to ensure a steady behavior. On a local control level Young (2005)[45] opted for EMs as cybernetic units (Figure 7). As elementary modes appear to be useful to understand cellular objectives for the overall metabolic network,[157] the choice for the EMs as local control level seems quite obvious. However, the choice of the associated objective function is less so. Young (2005)[45] opted for the optimization of a harmonic mean flux J:

$$\zeta = \frac{\sum_{i=1}^{n} \xi_i}{\sum_{i=1}^{n} \frac{\xi_i}{r_i v_i}} \tag{47}$$

with n the number of reactions involved in the elementary mode, r_i the rate of reaction i, v_i the cybernetic variable controlling enzyme activity, and ξ_i the flux through reaction i in the elementary mode. This objective function aims at a steady throughput through the EM, and consequently accumulation or depletion of certain metabolites is avoided. However, its biological foundation seems difficult to grasp.

On a global control level the various EMs compete with each other for the available resources (Figure 7). The objective function used for this control level is maximisation of growth ($J = \mu$).

Pros and cons for the rational design of bioprocesses—When applying cybernetic models the same armamentum of optimization techniques can be used as for mechanistic and approximative models. However, because enzyme levels are considered to be variable the extrapolation potential and hence the predictive power of those models is in theory much larger. In addition, the genetic regulation, which is so difficult to be captured by both flux balance analysis models

and mechanistic models, is relatively successfully captured by means of the nonlinear programming, which allows to capture the subtleties of genetic regulation, in stead of the on/off regulation of gene expression typically used in flux balance analysis models.

Cybernetic models consider both metabolome and proteome. They apply principles of control theory with the aim of substituting the unknown mechanistic details of the cell's regulatory architecture by an objective function by supposing that the metabolism of a micro-organism operates with a specific overall goal. Such models are therefore thought to have more extrapolation power. Although the approach thus seems appealing, given the present lack of knowledge and detailed experimental omics data and the aforementioned problems linked to mechanistic modeling, there still remain some issues unresolved: (i) it is still unclear to what extent unknown regulatory mechanisms can be captured by the framework, (ii) the robustness of the approach is unclear, e.g., although cybernetic models are said to be able to properly describe steady-state multiplicity,[158,159] real experimental evidence to support such a claim is lacking, (iii) though the cybernetic approach is a minimalistic approach, contrary to mechanistic models containing complex kinetics with a large number of (unidentifiable) parameters,[41] the incorporation of enzymes and the parameters for enzyme synthesis and degradation results in many parameters that are difficult to estimate, and (iv) for even relatively small networks the number of EMs is huge, e.g., for the metabolic network of[61] 17528 EMs are calculated, which use glucose as carbon source. Which EMs to choose, remains a question hard to answer.

Multivariate Statistics

Finally, multivariate statistics, principal component analysis (PCA) and partial least squares (PLS),[160,161] are more and more used in the field of metabolism studies[22–24] to interpret and to extract information from the pile of metabolomic, transcriptomic, and genomic data. By applying these methods, targets can be identified from data only in view of further improving production hosts. Especially the use of partial least squares seems promising. The objective in PLS modeling is to find a few "new" variables, *X*-scores, in such a way that the information in the dependent variables *Y* can be predicted as good as possible.

In fact, this projection method decomposes variables of high collinearity into one-dimensional variables, i.e., an input score vector t and an output score vector u, which allows PLS to handle many and correlated predictor variables.[161] The vectors t_1 and u_1 are defined as[162]:

$$t_1 = E_0 w_1 \tag{48}$$

$$u_1 = F_0 c_1 \tag{49}$$

where E_0 is the standardised data matrix from X and F_0 is the standardised data matrix from Y.[163] The aim of this data pretreatment is to focus on the (relevant) biological information by emphasising different aspects in the data, for instance, the value of a variable relative to its average value and to reduce the influence of disturbing factors, e.g., measurement noise.[163] Hence, the regression formulae for components t_1 and u_1 are given by:

$$E_0 = t_1 p_1^T + E_1 \tag{50}$$

$$F_0 = u_1 q_1^T + F_1 \tag{51}$$

where p_1 and q_1 are the loading vectors, and E_1 and F_1 are residual matrices. The linear relationship between t_1 and u_1 is calculated by:

$$u_1 = b_1 t_1 + r_1 \tag{52}$$

where b_1 is the regression coefficient and r_1 is the residual vector. If t_1 and u_1 cannot explain the data within a specified precision or do not contain enough information, E_0 and F_0 will be replaced by the residual matrices E_1 and F_1. Consequently, the next latent variable vectors t_2 and u_2 are calculated by:

$$t_2 = E_1 w_2 \tag{53}$$

$$u_2 = F_1 c_2 \tag{54}$$

The regressions for components t_2 and u_2 are therefore calculated by:

$$E_1 = t_2 p_2^T + E_2 \tag{55}$$

$$F_1 = u_2 q_2^T + F_2 \tag{56}$$

This iterative procedure is repeated by using the regression residual terms obtained at the previous iteration on both the inputs and outputs at each step. The decomposition of E_0 and F_0 by score vectors is defined by:

$$E_0 = \sum_{h=1}^{m} t_h p_h^T + E \tag{57}$$

$$F_0 = \sum_{h=1}^{m} u_h q_h^T + F \tag{58}$$

where p and q are loading vectors, E and F are residuals. For the number of m terms, a cross-validation method can be applied or a threshold variance of E can be used as stopping criterion.[160]

In PLS one can calculate a similar kind of regression coefficients as one does in multiple linear regression. These regression coefficients relate matrix X directly to Y:

$$Y = XB + \varepsilon \tag{59}$$

Both regression coefficients and loading weights can be used to study the system. Note that these regression coefficients are not independent unless the number of partial least squares regression components equals the number of *X*-variables. By studying the loading weights, one can see how important the variable is in each latent variable. A large positive or negative weight value indicates that the corresponding X variable is highly correlated with the values in the score matrix U and hence with matrix Y. Correlations between variables can be verified by looking at the loading weights.[160,164]

van der Werf et al. (2005)[24] applied PLS regression to link metabolite levels to the microbial phenotype, i.e., by ordering the importance of the metabolites by virtue of the weight factors, metabolites that contributed most to the phenotype of interest could be identified.

Pros and cons for the rational design of bioprocesses—Partial least squares regression is a useful tool to analyze the numerous and highly correlated data that are typically collected in the context of metabolic engineering. In this context metabolite[24] and transcript levels are linked to the microbial phenotype and subsequently their importance can be evaluated.

However, though the establishment of correlations between product formation and process variables, intracellular concentrations, etc. may be relatively easy, it is not that easy to remedy the identified bottlenecks. For example, though the concentration levels (low or high) of certain metabolites can be linked with increased product formation, PLS does not yield information about how to obtain the desired concentration levels. van der Werf et al. (2005)[24,54] relatively successfully applied this method to select targets in view of optimization. This selection was based on the calculated regression coefficients. However, since these regression coefficients are in general dependent (unless the number of PLSR components equals the number of *X*-variables) their confidence intervals according to the traditional statistical interpretation are infinite.[160] Hence, the robustness against false positive targets is limited. Martens and Martens[165] used jack-knifing to estimate standard errors and confidence intervals (During cross-validation the variation in the parameters can be used to compute their standard deviation and subsequently their confidence interval).

Because of its properties, PLS is also popularly used in the context of process analytical technology (PAT) to monitor and control complex bioprocesses.[166] This PAT framework aims to design, develop, and operate processes consistently, in order to ensure a predefined quality at the end of the manufacturing process.

For interpretation purposes, better tools do exist in comparison with partial least squares, e.g., stepwise multivariate linear regression. However, the typical data collected in the context of metabolic engineering are often highly correlated which leads to multicollinearity problems, e.g., biased parameter estimates and dangerously unstable predictions. Hence, partial least squares is popularly used.[22–24]

Another disadvantage of such an approach is that these models are completely data driven and consequently do not use the state of the art knowledge. To deal with this issue some have attempted to combine partial least squares and mechanistic models, e.g., Lee et al. (2005).[167] Such examples are for the moment however scarce.

Thus, though relationships can be established between, for instance, metabolite pool sizes and a process parameter, it remains unclear how to modify the cell with a view to improving process performance.

Conclusions

It should be clear from the earlier sections that despite the vast lack of knowledge about the cell's regulatory architecture, the application of both experimental techniques and mathematical methods steadily yields valuable information about the microbial metabolism. For instance, stoichiometric network analysis is a useful tool to rapidly evaluate the possible impact of knock-outs and knock-ins on the process performance.

In the presence of data, the use of constraints-based flux balance analysis seems more promising to steer the process of metabolic engineering. One should however be aware that it is hard to mathematically predict/capture the many subtleties of the microbial regulation. Hence, the applied environmental conditions and the (adaptation to a perturbed) genetic background will determine the success of such an approach. Furthermore, it should be clear that the objective function should be chosen with great care, since cellular optimal behavior does not always coincide with growth maximization. Especially for the optimization of the production of metabolites in primary metabolism that are subject to severe (redox) constraints, stoichiometric modeling has been shown to be useful. It is less so for the optimization of minor pathways.

Carbon modeling allows to obtain a more detailed insight in the flow of carbon throughout the metabolic network. In many metabolic engineering applications a detailed understanding of the flow of carbon is a prerequisite to directedly modify the cell. In this context, the gathering of isotopic instationary data both under steady-state and dynamic conditions will boost model-based metabolic engineering, since the information content of such data is huge. More attention should however be devoted to the validity check of the ultimately obtained flux distribution. The application of these tools may result in new insights about the metabolic network to be chosen.

Dynamic metabolic models might be useful tools to optimize microbial metabolism as well, as these models do incorporate kinetics and the regulation of enzymatic reactions and are able to identify the bottlenecks in a metabolic pathway. Especially for the optimization of minor pathways and metabolites that are not subject to severe (redox) constraints, dynamic metabolic modeling has been shown to be useful. However, the drawbacks of this approach are still numerous. Models relying on in vitro derived mechanistic equations are overparameterized for the available data, nowadays typically collected during only one perturbation experiment. The alternative, approximative modeling is no deus ex machina either as in order to collect informative data for parameter identification it might be necessary to radically perturb the cell, probably way beyond the metabolite range for which approximative kinetic formats yield an adequate description of the true kinetics. In addition, these dynamic metabolic models, both mechanistic and approximative ones, zoom in on a limited part of the metabolism, which impedes mass balance checks during transient conditions. Moreover, the behavior of cofactors is not yet modelled in a mechanistic manner, since, for instance, the pool size of the adenine nucleotides inexplicably changes during the transition from a glucose-limited to a glucose-abundant culture. Despite the rise of exchange tools like the systems biology markup language (SBML),[168] one thus should be aware that the "plug and play" character of such model(s) (structures) remains limited. In addition there is still a long way to go before genetic regulation will be properly addressed.

At present the only feasible way to model this multilevel regulation appears to be the cybernetic framework, since detailed mechanistic knowledge about this regulation is lacking, which renders such a mechanistic approach completely intractable for large metabolic networks. The nonlinear programming that steers the allocation of the limited resources the cell disposes of, seems to successfully capture this genetic and metabolic regulation. Although the approach seems appealing, given the present lack of knowledge,

detailed experimental omics data, and some of the aforementioned problems linked to "conventional" kinetic metabolic modeling, there still remain some issues unresolved, which will require further research.

Despite the potential of the afore-mentioned modeling tools the usefulness of data driven techniques, such as partial least squares regression, to optimize microbial metabolism, is clear. However, though, the establishment of correlations between product formation and process variables, intracellular concentrations, etc. may be relatively easy, it is not that easy to remedy the identified bottlenecks.

Despite the increasing ability of these tools to describe cellular metabolism as a whole, the contemporary lack of knowledge about the functioning of the cell is still limiting the use and usefulness of many of them to steer the process of metabolic engineering, e.g., at present a lot of mathematically relevant questions remain unanswered, e.g., which network and objective function to choose? Consequently, the search for tools which are useful to help unravel the complex regulation of microbial metabolism has not come to an end yet. In addition, more labeled and unlabeled intracellular metabolomic, proteomic, transcriptomic, and genomic data gathered under both steady-state and dynamic conditions will be imperative to fully exploit the potential of these techniques and to reduce the uncertainty on the identified genetic engineering targets.

Acknowledgments

The authors thank the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for financial support in the framework of the SBO-project 040125. The first author was research assistant of the Fund for Scientific Research-Flanders (2003-2007).

Literature Cited

1. Ui S, Mimura A, Ohkuma M, Kudo T. Formation of a chiral acetoinic compound from diacetyl by *Escherichia coli* expressing meso-2,3-butanediol dehydrogenase. *Lett Appl Microbiol*. 1999;28:457–460.
2. Cameron DC, Cooney CL. A novel fermentation: the production of R(-)-1,2-propanediol and acetol by *Clostridium thermosaccharolyticum*. *Bio/Technology*. 1986;4:651–654.
3. Takiguchi Y, Mishima H, Okuda M, Terao M, Aoki A, Fukuda R. Milbemycins, a new family of macrolide antibiotics: fermentation, isolation and physico-chemical properties. *J Antibiot*. 1980;33:1120–1127.
4. De Mey M, Lequeux G, Maertens J, De Maeseneire S, Soetaert W, Vandamme E. Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms. *Anal Biochem*. 2006;352:198–203.
5. Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 2004;101:10205–12012.
6. Kell DB. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol*. 2004;7:296–307.
7. Palsson BO. *In silico* biology through "omics". *Nature Biotechnol*. 2002;20:649–650.
8. Kromer JO, Wittman C, Schroder H, Heinzle E. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab Eng*. 2006;8:353–369.
9. Lee SY, Homg 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.
10. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLOS Comput Biol*. 2005;1:e46.
11. Patil KR, Rocha I, Forster J, Nielsen J. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*. 2005;6:308.
12. Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng*. 2005;7:155–164.
13. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol*. 2007;3:149.
14. Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome and *in silico* gene knockout simulation. *Proc Natl Acad Sci USA*. 2007;104:7797–7802.
15. Cox SJ, Shalel LS, Sanchez AM, Lin H, Peercy B, Bennet GN, San KY. Development of a metabolic network design and optimization framework incorporating implement constraints: a succinate production case study. *Metab Eng*. 2006;8:46–57.
16. van Luijk N, Rochat K, Muilwijk B, van der Werff-van der Vat B, Bijlsma S, Coulier L, Jellema R, Hankemeier T, van der Werf MJ. Strain inprovement with seven-league boots: application of metabolomics to improve phenylalanine production by *Escherichia coli*. 2008. Availavle at: www.tno.nl/downloads/KvL_FBI_Poster_democoli_metabolomics.pdf.
17. van der Werf MJ. TNO: microbial biotechnology, 2008. Available at: www.tno.nl/downloads\%5CMicrobial\%20Pro duction\%20Proceses1.pdf, www.tno.nl/downloads\%5CMicro bial\%20Production\%20Proceses1.pdf.
18. Cánovas M, Bernal V, Sevilla A, Iborra JL. Role of wet experiment design in data generation: from *in vivo* to *in silico* and back. In Silico *Biol*. 2007;7:S3–S16.
19. Sevilla A, Schmid JW, Mauch K, Iborra JL, Reuss M, Cánovas. Model of central and trimethylammonium metabolism for optimizing L-carnitine production by *E. coli*. *Metab Eng*. 2005;7:401–425.
20. van Gulik WM, de Laat WT, Vinke JL, Heijnen JJ. Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G. *Biotechnol Bioeng*. 2000;68:602–618.
21. Rizzi M, Baltes M, Theobald U, Reuss M. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: *II*. Mathematical model. *Biotechnol Bioeng*. 1997;55:592–608.
22. De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol*. 2007;7:34.
23. Jonsson J, Norberg T, Carlsson L, Gustafsson C, Wold S. Quantative sequence-activity models (QSAM)-tools for sequence design. *Nucleic Acids Res*. 1993;21:733–739.
24. van der Werf MJ, Jellema RH, Hankemeier T. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J Ind Microbiol Biot*. 2005;32:234–252.
25. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA*. 2005;102:2685–2689.
26. Hammer K, Mijakovic I, Jensen PR. Synthetic promoter libraries - tuning of gene expression. *Trends Biotechnol*. 2006;24:53–55.
27. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA*. 2000;97:6640–6645.
28. Garoff H. Using recombinant DNA techniques to study protein targeting in the eucaryotic cell. *Annu Rev Cell Biol*. 1985;1:403–445.
29. Hieter P, Boguski M. Functional genomics: it's all how you read it. *Science*. 1997;278:601–602.
30. Bailey JE, Sburlati AR, Hatzimanikatis V, Lee K, Renner WA, Tsai PA. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng*. 1996;52:109–121.

31. Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol*. 2000;18:326–332.
32. Papin JA, Price ND, Palsson BO. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res*. 2002;12:1889–1900.
33. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of system stoichiometric constraints. *Am J Physiol Regul Integr Comp Physiol*. 2001;280:695–704.
34. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol*. 2007;3:119.
35. Beard DA, Liang S-D, Qian H. Energy balance for analysis of complex metabolic network. *Biophys J*. 2002;83:79–86.
36. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*. 2002;99:15112–15117.
37. 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.
38. Schmidt K, Nielsen J, Villadsen J. Quantitative analysis of metabolic fluxes in *Escherichia coli* using two-dimensional NMR spectroscopy and complete isotopomer models. *J Biotechnol*. 1999;71:175–190.
39. Wiechert W, Mollney M, Isermann N, Wurzel M, de Graaf AA. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng*. 1999;66:69–85.
40. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): A novel framework for modeling isotopomer distributions. *Metab Eng*. 2007;9:68–86.
41. Degenring D, Froemel C, Dikta G, Takors R. Sensitivity analysis for the reduction of complex metabolism models. *J Process Control*. 2004;14:729–745.
42. Wong P, Gladney S, Keasling JD. Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol Prog*. 1997;13:132–143.
43. Peercy BE, Cox SJ, Shalel-Levanon S, San K-Y, Bennet G. A kinetic model of oxygen regulation of cytochrome production in *Escherichia coli*. *J Theor Biol*. 2006;242:547–563.
44. Kompala DS, Ramkrishna D, Tsao GT. Cybernetic modeling of microbial growth on multiple substrates. *Biotechnol Bioeng*. 1984;26:1272–1281.
45. Young JD. A System-Level Mathematical Description of Metabolic Regulation Combining Aspects of Elementary Flux Mode Analysis with Cybernetic Control Laws. PhD thesis, Purdue University, USA, 2005.
46. Francke C, Siezen RJ, Teusink B. Reconstucting the metabolic network of a bacterium from its genome. *Trends Microbiol*. 2005;13:550–558.
47. DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M, Best A. Towards the automated genetaion of genome-scale metabolic networkds in the seed. *BMC Bioinformatics*. 2007;8:139.
48. Suthers PF, Dasika MS, Kumar VS, Denisov F, Glass JI, Maranas CD. A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, *i*PS189. *PLOS Comput Biol*. 2009;5:1–14.
49. Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, Peralta-Gil M, Karp PD. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res*. 2005;33:D334–D337.
50. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1999;27:29–34.
51. Schomburg I, Hofmann O, Bansch C, Chang A, Schomburg D. Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. *Gene Funct Dis*. 2000;3–4:109–118.
52. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12. *Genome Biol*. 2003;4:54.
53. Stephanopoulos GN, Aristidou AA, Nielsen J. *Metabolic Engineering*. New York: Academic Press, 1998.
54. van der Werf MJ. Towards replacing closed with open target selection strategies. *Trends Biotechnol*. 2005;23:11–16.
55. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. *Trends Biotechnol*. 2004;22:400–405.
56. Schilling CH, Letscher D, Palsson BO. Theory of the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*. 2000;203:229–248.
57. Carlson R, Fell D, Scrienc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol Bioeng*. 2002;79:121–134.
58. Carlson R, Scrienc F. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall fluxes. *Biotechnol Bioeng*. 2003;86:149–162.
59. Nookaew I, Meechai A, Thammarongtham, C, Laoteng K, Ruanglek V, Cheevadhanarak S, Nielsen J, Bhumiratana S. Identification of flux regulation coefficients from elementary flux modes: a systems biology tool for analysis of metabolic networks. *Biotechnol Bioeng*. 2007;97:1535–1549.
60. Schwarts JM, Gauguin C, Nacher J, de Daruvar A, Kanehisa M. Observing metabolic functions at the genome scale. *Genome Biol*. 2007;8:R123.
61. Lequeux G, Johansson L, Maertens J, Vanrolleghem PA, Lidén G. MFA for overdetermined systems reviewed and compared with RNA expression data to elucidate the difference in shikimate yield between carbon- and phosphate-limited continuous cultures of *E. coli* W3110. shik1. *Biotechnol Prog*. 2006;22:1056–1070.
62. von Kamp A, Schuster S. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*. 2006;22:1930–1931.
63. Wiback SJ, Palsson BO. Extreme pathway analysis of human red blood cell metabolism. *Biophys J*. 2002;83:808–818.
64. Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*. 2008;24:2229–2235.
65. van der Heijden RTJM, Romein B, Heijnen JJ, Hellinga C, Luyben KCAM. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. *Biotechnol Bioeng*. 1994;43:11–20.
66. 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.
67. 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.
68. Baart GJE, Zomer B, de Haan A, van der Pol LA, Beuvery C, Tramper J, Martens DE. Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biol*. 2007;8:R316.
69. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. *Biophys J*. 2007;92:1792–1805.
70. Kummel A, Panke S, Heineman M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol*. 2006;2:2006–2034.
71. Dauner M, Bailey JE, Sauer U. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. *Biotechnol Bioeng*. 2001;76:144–156.
72. van Winden WA, van Dam JC, Ras C, Kleijn RJ, Vinke JL, van Gulik WM, Heijnen JJ. Metabolic-flux analysis of *Saccharomyces cerevisiae* CEN. PK113–7D based on mass isotopomer measurements of ^{13}C-labeled primary metabolites. *FEMS Yeast Res*. 2005;5:559–568.
73. Akesson M, Forster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. *Metab Eng*. 2004;6:285–293.
74. Cakir T, Kirdar B, Onsan ZI, Ulgen KO, Nielsen J. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. *BMC Syst Biol*. 2007;1:18.
75. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*. 2007;3:101.

76. Edwards JS, Palsson B. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA*. 2000;97:5528–5533.
77. Oliveira AP, Nielsen J, Forster J. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol*. 2005;5:39.
78. Knorr AL, Jain R, Srivastava R. Bayesian-based selection of metabolic objective functions. *Bioinformatics*. 2007;23:351–357.
79. Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1965;7:308–313.
80. Srinivas M, Patnaik LM. Genetic algorithms: a survey. *Computer*. 1994;27:17–26.
81. Kvasnicka V, Pospichal J. A hybrid of simplex method and simulated annealing. *Chemom Intell Lab*. 1997;39:161–173.
82. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003;30:887–899.
83. Gombert AK, Nielsen J. Mathematical modeling of metabolism. *Curr Opin Biotechnol*. 2000;11:180–186.
84. Koffas M, del Cardayre S. Evolutionary metabolic engineering. *Metab Eng*. 2005;7:1–3.
85. Schmidt K, Marx A, de Graaf AA, Wiechert W, Sahm H, Nielsen J, Villadsen J. ^{13}C tracer experiments and metabolite balancing for metabolite flux analysis: Comparing two approaches. *Biotechnol Bioeng*. 1998;58:254–257.
86. Wittmann C, Heinzle E. Application of MALDI-TOF MS to lysine-producing *Corynebacterium glutamicum* a novel approach for metabolic flux analysis. *Eur J Biochem*. 2001; 268:2441–2455.
87. Fong SS, Joyce AR, Palsson BO. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res*. 2005;15:1365–1372.
88. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady state solution space. *J Theor Biol*. 2003;221:309–325.
89. Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem*. 2002;277:28058–28064.
90. Nagrath D, Avila-Elchiver M, Berthiaume F, Tilles AW, Messac A, Yarmush ML. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng*. 2007;35:863–885.
91. De Noronha Pissarr P, Nielsen J. Thermodynamics of metabolic pathways for penicillin production: analysis of thermodynamic feasibility and free energy charges during fed-batch cultivation. *Biotechnol Prog*. 1997;13:156–165.
92. Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. *Brief Bioinformactics*. 2006;7:140–150.
93. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*. 2002;83:1331–1340.
94. Lee JM, Gianchandani EP, Eddy JA, papin JA. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLOS Comput Biol*. 2008;4:1–20.
95. Schmidt K, Carlsen M, Nielsen J, Villadsen J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol Bioeng*. 1997;1997:831–840.
96. Grotkjaer T, Cristakopoulos P, Nielsen J, Olsson L. Comparative metabolic network analysis of two xylose fermenting recombinant *Saccharomyces cerevisiae* strains. *Metab Eng*. 2005;7:437–444.
97. Kleijn RJ, Geertman J-MA, Nfor BK, Ras C, Schipper D, Pronk JT, Heijnen JJ, Maris AJ, van Winden WA. Metabolic flux analysis of a glycerol-overproducing *Saccharomyces cerevisiae* strain based on GC-MS, LC-MS and NMR-derived ^{13}C-labeling data. *FEMS Yeast Res*. 2007;7:216–231.
98. Van Winden WA, Verheijen PJT, Heijnen JJ. Possible pitfalls of flux calculations based on ^{13}C labeling. *Metab Eng*. 2001:3:151–162.
99. Mollney M, Wiechert W, Kownatzki D, De Graaf AA. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnol Bioeng*. 1999;66:86–103.
100. Zamboni N, Fischer E, Sauer U. FiatFlux - a software for metabolic flux analysis from ^{13}C-glucose experiments. *BMC Bioinformatics*. 2005;6:209.
101. Quek L-E, Wittmann C, Nielsen LK, Kromer JO. OpenFLUX: efficient modeling software for ^{13}C-based metabolic flux analysis. *Microb Cell Fact*. 2009;8:25.
102. Nanchen A, Schicker A, Sauer U. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microb*. 2006;72:1164–1172.
103. Grotkjaer T, Akesson M, Christensen B, Gombert AK, Nielsen J. Impact of transamination reactions and protein turnover on labeling dynamics in ^{13}C-labeling experiments. *Biotechnol Bioeng*. 2004;86:209–216.
104. Noh K, Wahl A, Wiechert W. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metab Eng*. 2006;8:554–577.
105. Wahl SA, Noh K, Wiechert W. ^{13}C labeling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics*. 2008;9:152.
106. Zhao Z, Kuijvenhoven K, Ras C, van Gulik WM, Heijnen JJ, Verheijen PJT, van Winden WA. Isotopic non-stationary ^{13}C gluconate tracer method for accurate determination of the pentose phosphate pathway split-ratio in *Penicilium chrysogenum*. *Metab Eng*. 2008;10:178–186.
107. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng*. 2002;79:53–73.
108. Vaseghi S, Baumeister A, Rizzi M, Reuss M. *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metab Eng*. 1999;1:128–140.
109. Savageau MA. *Biochemical System Analysis*, 1st ed. Reading: Addison-Wesley Publishing Company; 1976.
110. Hatzimanikatis V, Bailey JE. MCA has more to say. *J Theor Biol*. 1996;182:233–242.
111. Hatzimanikatis V, Bailey JE. Effects of spatiotemporal variations on metabolic control: approximative analysis using (log) linear kinetic models. *Biotechnol Bioeng*. 1997;54:91–104.
112. Westerhoff HV, Van Dam K. *Thermodynamics and Control of Biological Free Energy Transduction*. Amsterdam: Elsevier; 1987.
113. Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng*. 2005;91:534–545.
114. Visser D, Heijnen JJ. Dynamic simulation and metabolic redesign of a branched pathway using lin-log kinetics. *Metab Eng*. 2003;5:164–176.
115. Fell D. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem J*. 1992;286:313–330.
116. Stephanopoulos G, Simpson T. Flux amplification in complex metabolic networks. *Chem Eng Sci*. 1997;52:2607–2627.
117. Thomas S, Fell DA. The role of multiple enzyme activation in metabolic flux control. *Adv Enzyme Regul*. 1998;38:65–85.
118. Lee B, Yen J, Yang L, Liao JC. Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnol Bioeng*. 1999;62:722–729.
119. Yang YT, Bennet GN, San KY. Genetic and metabolic control. *Eur J Biotechnol*. 1998;1:3.
120. Mauch K, Arnold S, Reuss M. Dynamic sensitivity analysis for metabolic systems. *Chem Eng Sci*. 1997;52:2589–2598.
121. Tomita M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol*. 2001;19:205–210.
122. Tomita M, Takahashi K, Matsuzaki Y, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA III. E-cell: software environment for whole-cell simulation. *Bioinformatics*. 1999;15: 72–84.
123. Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnol Bioeng*. 1997;55:305–316.
124. Theobald U, Mailinger W, Reuss M, Rizzi M. *In vivo* analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Anal Biochem*. 1993;214:31–37.
125. Visser D, van Zuylen GA, van Dam JC, Eman MR, Proll A, Ras C, Wu L, van Gulik WM, Heijnen JJ. Analysis of *in vivo* kinetics of

glycolysis in aerobic *Saccharomyces cerevisiae* by application of glucose and ethanol pulses. *Biotechnol Bioeng*. 2004;88:157–167.
126. Gao J, Gorenflo VM, Scharer JM, Budman HM. Dynamic metabolic modeling for a MAB bioprocess. *Biotechnol Prog*. 2007;23:168–181.
127. Kresnowati MTAP, van Winden WA, Almering MJH, ten Pierick A, Ras C, Knijnenburg TA, Daran-Lapujade P, Pronk JT, Heijnen JJ, Daran JM. When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Mol Syst Biol*. 2006;2:49.
128. Mashego MR, van Gulik WM, Heijnen JJ. Metabolome dynamic responses of *Saccharomyces cerevisiae* to simultaneous rapid perturbations in external electron acceptor and electron donor. *FEMS Yeast Res*. 2007;7:48–66.
129. Voit EO, Alvarez-Vasquez F, Sims KJ. Analysis of dynamic labeling data. *Math Biosci*. 2004;191:83–99.
130. Maranas CD. Computational analysis and redesign of biological networks. In *IECA 2006*, Jeju Island; 2006.
131. Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst*. 2005;130:606–625.
132. Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ. Microbial metabolomics: past, present and future methodologies. *Biotechnol Lett*. 2007;29:1–11.
133. Oldiges M, Lutz S, Pflug S, Schroer K, Stein N, Wiendahl C. Metabolomics: current state and evolving methodologies and tools. *Appl Microbiol Biotechnol*. 2007;76:495–511.
134. Villas-Boas SG, Bruheim P. Cold glycerol-saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Anal Biochem*. 2007;370:87–97.
135. Benevolensky SV, Clifton D, Fraenkel DG. The effect of increased PGI on glucose metabolism in *Saccharomyces cerevisiae*. *In situ* study of the glycolytic pathway in *Saccharomyces cerevisiae*. *J Biol Chem*. 1994;269:4878–4882.
136. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem*. 2000;267:5313–5329.
137. Bucholz A, Takors R, Wandrey C. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatography-electrospray ionization tandem mass spectrometric techniques. *Anal Biochem*. 2001;295:129–137.
138. Schaefer U, Boos W, Takors R, Weuster-Botz D. Automated sampling device for monitoring intracellular metabolite dynamics. *Anal Biochem*. 1999;270:88–96.
139. Abdel-Hamid AM, Attwood MM, Guest JR. Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*. *Microbiology*. 2001;147:1483–1498.
140. Li M, Yao S, Shimizu K. Effect of *poxB* gene knockout on metabolism in *Escherichia coli* based on growth characteristics and enzyme activities. *World J Microbiol Biotechnol*. 2007;23:573–580.
141. Canonaco F, Hess TA, Heri S, Wang T, Szyperski T, Sauer U. Metabolic flux response to phosphoglucose isomerase knockout in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiol Lett*. 2001;104:247–252.
142. Lerondel G, Doan T, Zamboni N, Sauer U, Ayermich S. YtsJ has the major physiological role of the four paralogous malic enzyme isoforms in *Bacillus subtilis*. *J Bacteriol*. 2006;188: 4727–4736.
143. Sauer U, Eikmanns BJ. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev*. 2005;29:765–794.
144. Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol*. 1973;27:65–104.
145. Beven K. A manifesto for the equifinality thesis. *J Hydrol*. 2006;320:18–36.
146. Mashego MR, Jansen MLA, Vinke JL, van Gulik WM, Heijnen JJ. Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS Yeast Res*. 2005;5:419–430.
147. Patnaik PR. Are microbes intelligent beings? An assessment of cybernetic modeling. *Biotechnol Adv*. 2000;18:267–288.
148. Varner J, Ramkrishna D. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. *Biotechnol Progr*. 1999;15:407–425.
149. Varner J, Ramkrishna D. Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. *Metab Eng*. 1999;1:88–116.
150. Baloo S, Ramkrishna D. Metabolic regulation in bacterial cultures: I. *Biotechnol Bioeng*. 1991;38:1337–1352.
151. Baloo S, Ramkrishna D. Metabolic regulation in bacterial cultures: II. *Biotechnol Bioeng*. 1991;38:1353–1362.
152. Turner BG, Ramkrishna D, Jansen NB. Cybernetic modeling of bacterial cultures at low gowth rates: single-substrate systems. *Biotechnol Bioeng*. 1989;1989:252–161.
153. Varner J, Ramkrishna D. Application of cybernetic models to metabolic engineering: investigation of storage pathways. *Biotechnol Bioeng*. 1998;58:282–291.
154. Ramakrishna R, Ramkrishna D, Konopka AE. Cybernetic modeling of growth in mixed, substitutable substrate environments: preferential and simultaneous utilization. *Biotechnol Bioeng*. 1996;52:141–151.
155. Straight JV, Ramkrishna D. Cybernetic modeling and regulation of metabolic pathways. growth on complementary nutrients. *Biotechnol Prog*. 1994;10:574–587.
156. Zak SH. *Systems and Control*. New York: Oxford University Press; 2003.
157. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. *Nature*. 2002;420:190–193.
158. Namjoshi A, Kienle A, Ramkrishna D. Steady-state multiplicity in bioreactors: bifurcation analysis of cybernetic models. *Chem Eng Sci*. 2003;58:793–800.
159. Narang A, Konopka A, Ramkrishna D. New patterns of mixed-substrate utilization during batch growth of *Escherichia coli* K12. *Biotechnol Bioeng*. 1997;55:747–757.
160. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool for chemometrics. *Chemom Intell Lab*. 2001;58:109–130.
161. Wold S, Trygg J, Berglund A, Antti H. Some recent developments in PLS modeling. *Chemom Intell Lab*. 2001;58:131–150.
162. Li R, Meng G, Gao N, Xie H. Combined use of partial least-squares regression and neural network for residual life estimation of large generator stator insulation. *Meas Sci Technol*. 2007;18:2074–2082.
163. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006;7:142.
164. Teppola S, Mujunen SP, Minkkinen P. Partial least squares modeling of an activated sludge plant: a case study. *Chemom Intell Lab*. 1997;38:197–208.
165. Martens H, Martens M. Modified jack-knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression (PLSR). *Food Qual Prefer*. 2000;11:5–16.
166. Lopes JA, Costa PF, Alves TP, Menezes JC. Chemometrics in bioprocess engineering: process analytical technology (PAT) applications. *Chemometr Intell Lab*. 2004;72:269–275.
167. Lee DS, Vanrolleghem PA, Park JM. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *J Biotechnol*. 2005;115:317–328.
168. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003;19:524–531.

Manuscript received Nov. 3, 2009.