

Minireview

Mathematical models of cell factories: moving towards the core of industrial biotechnology

Marija Cvijovic,[†] Sergio Bordel[†] and Jens Nielsen*
Department of Chemical and Biological Engineering,
Chalmers University of Technology, Göteborg, Sweden.

Summary

Industrial biotechnology involves the utilization of cell factories for the production of fuels and chemicals. Traditionally, the development of highly productive microbial strains has relied on random mutagenesis and screening. The development of predictive mathematical models provides a new paradigm for the rational design of cell factories. Instead of selecting among a set of strains resulting from random mutagenesis, mathematical models allow the researchers to predict *in silico* the outcomes of different genetic manipulations and engineer new strains by performing gene deletions or additions leading to a higher productivity of the desired chemicals. In this review we aim to summarize the main modelling approaches of biological processes and illustrate the particular applications that they have found in the field of industrial microbiology.

Introduction

The rapid progress in molecular biology and the development of tools for directed genetic modifications, high-throughput measurements and genome sequencing have made available important quantities of data that can be used to validate mathematical models and fit their parameters as well as the means to test *in vivo* the validity of the predictions of the models (Price *et al.*, 2003). Combining knowledge of various disciplines (biology, mathematics, physics, biochemistry, molecular biotechnology and computer science) enhances the possibility to elucidate the complexity of biological systems and predict their behaviour.

Received 2 September, 2010; accepted 11 October, 2010. *For correspondence. E-mail nielsenj@chalmers.se; Tel. (+46) 31 772 3804; Fax (+46) 31 772 3801. [†]These authors contributed equally to this work.

In the field of industrial biotechnology there is currently much focus on how systems biology can improve the efficiency of cell factories and in particular speed up the development process (Nielsen and Jewett, 2008), and hereby ensure that new products can be brought to the market faster or there can be a faster improvement of existing bioprocesses. The use of metabolic engineering for improvement of cell factories is not a novel concept (Nielsen, 2001; Stephanopoulos, 2002), but recent synergies with tools developed in systems biology enabled the production of variety of products through biotechnology with significantly reduced time and resources required for commercialization (Otero and Nielsen, 2010; Tyo *et al.*, 2010).

Developing computational tools for data integration and generating *in silico* genome-scale metabolic models (GSMM) enables analysis of the effects of different media and specific mutations on growth and metabolic network adjustments. Numerous valuable predictions have been obtained from GSMM, with the relatively high success rate of 70–90%, depending of the organism and the predictions (Price *et al.*, 2003). *Escherichia coli* and *Saccharomyces cerevisiae* are some of the most exploited organisms in industrial biotechnology. *Escherichia coli* has been used for production of many different recombinant proteins (like human growth hormone) and the yeast *S. cerevisiae* is used for bioethanol production, production of a range of pharmaceutical proteins, fine and bulk chemicals, and nutraceuticals (Table S1).

Biological systems are complex and generally not entirely understood. Mathematical models provide means to better understand processes and unravel some of the complexities. The aim is to construct the model in the simplest possible way, but still retain the most important features of the system. A good model will be able to agree as closely as possible with the real world observations of the phenomenon we are trying to model and at the same time be interrogative.

Depending on the process we want to model, the available data and the goal, biological processes can be modelled using either kinetic or stoichiometric methods.

Dynamic models

Dynamic modelling (Fig. 1) requires knowledge of the kinetics including the parameters of kinetic expressions. Kinetics of the different reactions is used to describe dynamic changes in the state variables, which are typically the concentrations of key compounds. These dynamic models are typically represented as difference equations (*discrete dynamical systems*) or differential equations (*continuous dynamical systems*).

The most common technique for dynamic modelling of biological systems is the ordinary differential equation (ODE) approach. The main characteristic of ODEs is that we can obtain deterministic time series for the variables under investigation. Linear ODE can be solved analytically, while non-linear ODEs are much harder, and in some cases it is impossible to find the solution analytically. In this case the approximate solution is derived using numerical algorithms for solving differential equations.

The rapid development in the field of systems biology led to enormous expansion of computational tools that can be used for system analysis. Most of the tools are freely available for the scientific community and their use will greatly depend on the user's preferences and expertise (Klipp *et al.*, 2007).

Dynamic modelling approaches are used in modelling regulatory processes, like central carbon metabolism (Savageau, 1969a,b; Savageau *et al.*, 1970; Curto *et al.*, 1995; Sorribas *et al.*, 1995; Teusink *et al.*, 1998; 2000; Chassagnole *et al.*, 2002), cell cycle regulation (Novak and Tyson, 1997; Rizzi *et al.*, 1997; Novak *et al.*, 1998; 2001; Chen *et al.*, 2000; 2004; Barberis *et al.*, 2007) and different signalling pathways (Klipp *et al.*, 2005; Papin *et al.*, 2005; Kholodenko, 2006). These models help in understanding complicated dynamic features, such as glycolytic oscillations, regulatory feedback effects or cell cycle oscillations.

A dynamic model describing the central carbon metabolism in *E. coli* (Chassagnole *et al.*, 2002) includes the phosphotransferase system (PTS), glycolysis, pentose-phosphate pathway and storage material and represents the first step towards systematic analysis of metabolism in *E. coli*. The main feature of this model is its application in improvement of microbial production processes. The underlying framework for improvement of production capabilities of desired compounds is built on metabolic control analysis (MCA). Assuming that in a metabolic network, the kinetics of the individual enzymes is known, MCA allows estimating the individual flux control coefficients and hereby identifying targets for overexpression with the objective to increase the flux through the pathway.

Flux control coefficients of glucose uptake by the PTS were calculated from the general model. The highest flux

control coefficient was expectedly observed for the PTS since glucose uptake is irreversible and in the model it is independent of any preceding reactions. The results (Fig. S1) further suggest that there are other enzymes that control glucose uptake apart from the PTS. Phosphofructokinase (PFK) has the second highest value, and equivalent flux control coefficients are observed for pyruvate dehydrogenase (PDH) and glucose-6-phosphate dehydrogenase (G6PDH). The PTS is feedback inhibited by glucose-6-phosphate (G6P) and by its co-product pyruvate, and this explains why G6PDH has a high control on glucose uptake due to its role in G6P-consuming reaction and PDH is exerting flux control on glucose uptake as it is a pyruvate-degrading reaction.

Through combining kinetic modelling and MCA Hoefnagel and colleagues (2002) demonstrated that it was possible to identify targets for improving the flux through biotechnologically relevant pathways in *Lactococcus lactis*. Their approach allowed for simple estimation of flux control coefficients and a further advantage was that the effect of genetic manipulations can be tested directly using the kinetic model. The kinetic model of *L. lactis* comprises of a set of ODEs which describes the time dependence of the metabolite concentrations, while enzymes were modelled using reversible Michaelis–Menten equation.

Metabolic control analysis of the pyruvate branches in *L. lactis* (Fig. S2) indicated that the highest flux control coefficients of the acetolactate branch are not within this branch, as intuitively one would assume, but can be found in the enzymes outside this branch – lactate dehydrogenase (LDH) and NADH oxidase (NOX) (Table S2). Further analysis indicated that 92% of the pyruvate is converted via the acetolactate branch when LDH knockout is combined with NOX overexpression.

Another approach to model complex biological system is to provide detailed representation of smaller modules and then stitch these together to describe a larger system. Fine tuning and wiring of the components in small modules is more effective and controllable than in larger systems. A challenge with this approach is the linking of the different modules, but this can be achieved by defining appropriate input and output signals for each module. Extending this further gives the possibility to link different pathways (modelled as single independent modules) into a larger network.

The high osmolarity glycerol (HOG) pathway has been intensively studied in the literature (Albertyn *et al.*, 1994; Van Wuytswinkel *et al.*, 2000; de Nadal *et al.*, 2002; Hohmann, 2002). The pathway represents the fundamental process by which cells regulate their water balance. It consists of two branches – Sho1 and Sln1, both being transmembrane proteins, being placed upstream of other players in HOG signalling pathway. The fact that HOG

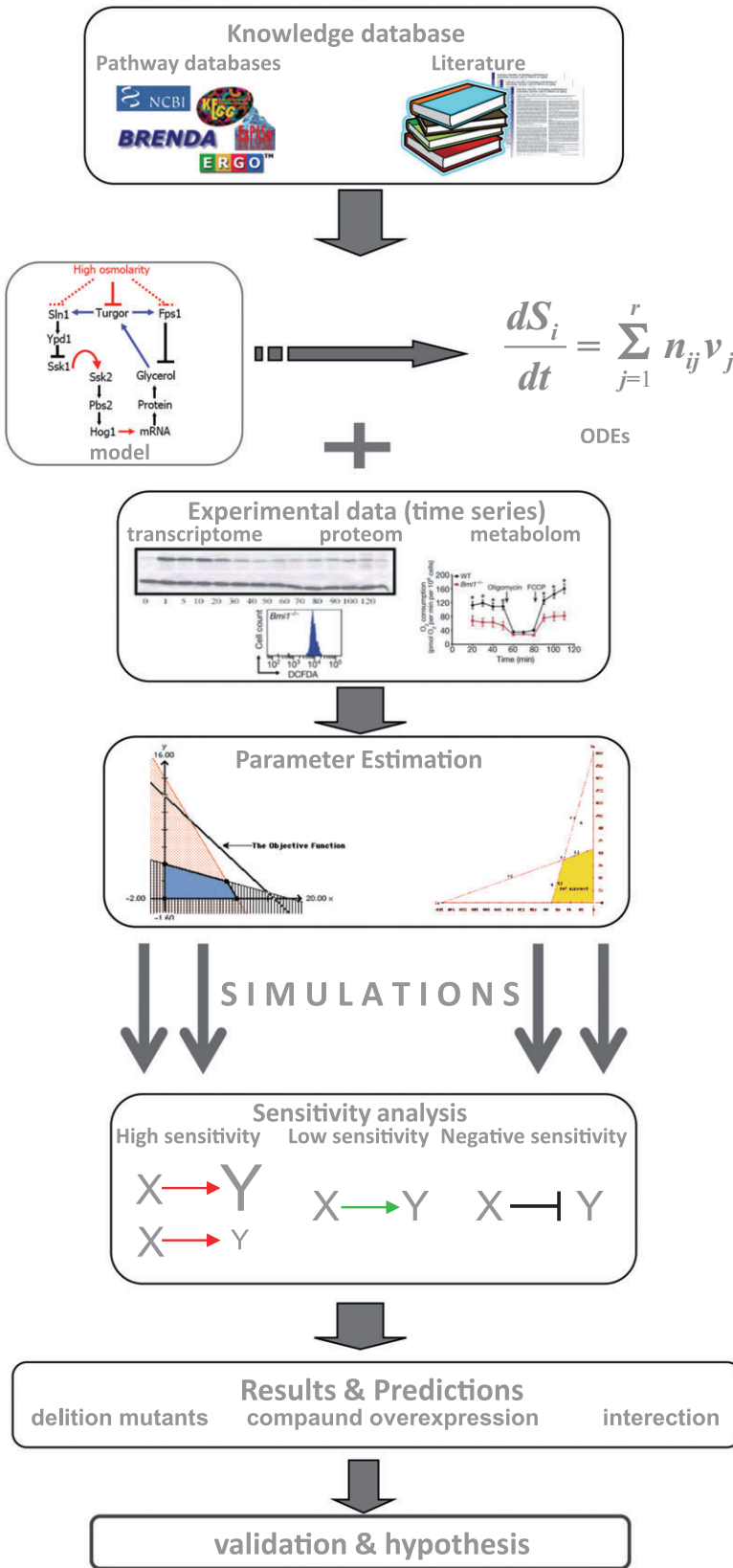


Fig. 1. Dynamic modelling approach – flowchart. The schematic representation of essential steps in the development of dynamic models. Formulation of the set of ODEs based on the given problem and a literature review. The set of ODEs, with the general representation: $\frac{dS_i}{dt} = \sum_{j=1}^r n_{ij} v_j$ gives the rate of change of concentration S_i of the i th compound, while r denotes the number of reactions involved, v_j the rate of j th reaction and n_{ij} represents stoichiometric coefficient of S_i in reaction j . The stoichiometric coefficients give stoichiometric matrix $N = \{n_{ij}\}$, which denotes the network properties. Analysis of matrix N gives information of admissible set of fluxes and conservation relations of the system. When building a model, not all parameters are known or could not be measurable directly, or there is inconsistency between different labs and strains that are used. In those cases, we have to estimate ‘unknown’ parameters by fitting the model to experimental data. Parameter estimation minimizes the error function over the parameters under investigation and using the goodness-of-fit measure the discrepancy between the observed values and the expected values in the model is monitored. Performing sensitivity analysis we get more information about the model. Some of the questions usually asked are: which parameters have the highest influence on the system behaviour or on the other hand which ones do not have any effect on the system, so they do not have to be considered further and can be fixed to some semi-arbitrary value. The main goal of sensitivity analysis is to better understand the dynamic behaviour of the system. Once the model is able to reproduce the input data obtained from experimental studies it can be used for the prediction of untested scenarios.

pathway activity can easily and rapidly be controlled experimentally by extracellular stimuli (Hohmann *et al.*, 2007) makes it a suitable candidate for system level studies via mathematical modelling.

The model developed by Klipp and colleagues (2005) represents a good example of an integrated approach towards a quantitative understanding of the osmotic shock response in yeast, and it also is a good example of linking regulation to a (small-scale) metabolic model. The model has four modules: Phosphorelay, MAP kinase cascade, Gene expression and Metabolism module. Each module was modelled and analysed individually.

This model comprises the HOG signalling pathway, gene expression, cellular metabolism of glycerol production and control of cellular volume and osmotic pressure. The entire reaction network consists of 32 ordinary equations and 70 parameters, which were estimated on the basis of steady state and time-course experiments. The model was validated with physiological and genetic perturbations.

Several new aspects of yeast osmoregulation have been revealed using *in silico* approach: (i) the contribution of osmotic and turgor pressure changes to the regulation of biochemical processes, (ii) the role of aquaglyceroporin Fps1p in controlling glycerol accumulation and signalling through the HOG pathway, and (iii) the function of the induced changes of gene expression as long-term contributions to the upregulation of glycerol (Klipp, 2007).

Genome-scale metabolic models (GSMM)

Kinetic models have their limitation in terms of describing large metabolic networks. Here simple stoichiometric models are more appropriate (Fig. 2), and with the appearance of genome sequences it became possible to reconstruct metabolic networks at genome scale. Four years after the first sequences were revealed, the first metabolic model was reconstructed (*Haemophilus influenzae*, Schilling and Palsson, 2000), and today more than 80 reconstructed models exist (Feist *et al.*, 2009; Milne *et al.*, 2009).

In GSMM the metabolic network is represented as a stoichiometric matrix containing the stoichiometric coefficients for all the metabolites in all the cellular reactions. Relevant data on the stoichiometry and occurrence of metabolic reactions are often extracted from annotated genome sequences, pathway databases (KEGG, ExPASy, ERGO), biochemical textbooks and research papers. Based on the stoichiometric coefficients for each metabolite in all reactions it is possible to set up mass balance equations for each metabolite. By assuming that the level of the metabolites is in steady state it is possible to constrain the set of fluxes (or reaction rates) of each of the cellular reactions, resulting in an underdetermined

system of linear equations. The assumption of steady state implies that for each internal metabolite in the network, the sum of the rates of the reactions producing it is equal to the sum of the rates of the reaction consuming it. This assumption has proved to be realistic due to the fact that the relaxation time that the internal metabolites take to reach a steady state after a perturbation is several magnitude orders lower than the doubling time of the cells.

To identify a particular solution to this set of linear equations requires determination of constraints under which the system functions. These constraints can be based on thermodynamics, enzyme activity and '-omics' data (genomics, proteomics, transcriptomics, metabolomics) (Beard *et al.*, 2002; Palsson, 2002; Price *et al.*, 2002), and the integration of high-throughput biological data generally increases the predictive capabilities of genome-scale models.

For exploiting the capabilities of metabolic networks a number of tools are available. Constraint based modelling can be performed using the COBRA TOOLBOX (Becker *et al.*, 2007) and web-based BioOpt application available in BioMet Toolbox (Cvijovic *et al.*, 2010). Structural properties of metabolic networks, like elementary flux modes (EFMs) and null space matrix, can be analysed in METATOOL (von Kamp and Schuster, 2006) or FluxAnalyzer (Klamt *et al.*, 2003). FluxAnalyzer is also able to detect dead-end metabolites in the network, that is, metabolites present only in one reaction, which can therefore not be used. BioMet Toolbox also features tools for transcriptome/proteome/metabolome data analysis (Reporter Features and Reporter Subnetworks applications).

Metabolic networks can be analysed using two different approaches: (i) identification of a unique solution and (ii) pathway analyses (Patil *et al.*, 2004).

Identification of a unique solution

Metabolic flux analysis (MFA) is the solution of the system of linear equations obtained from the mass balances around each internal metabolite. In order to find a unique solution for the fluxes, as many fluxes as degrees of freedom should be determined experimentally. The number of degrees of freedom of the system is equal to the number of reactions minus the rank of the stoichiometric matrix. The number of degrees of freedom is typically big (several hundreds for genome-scale models) and exact solutions using MFA can only be found for simplified metabolic networks.

Flux balance analysis (FBA) is an approach that relies on imposing an objective function to be maximized (or minimized) on the system and then find among the many possible solutions for the system, one that maximizes (or

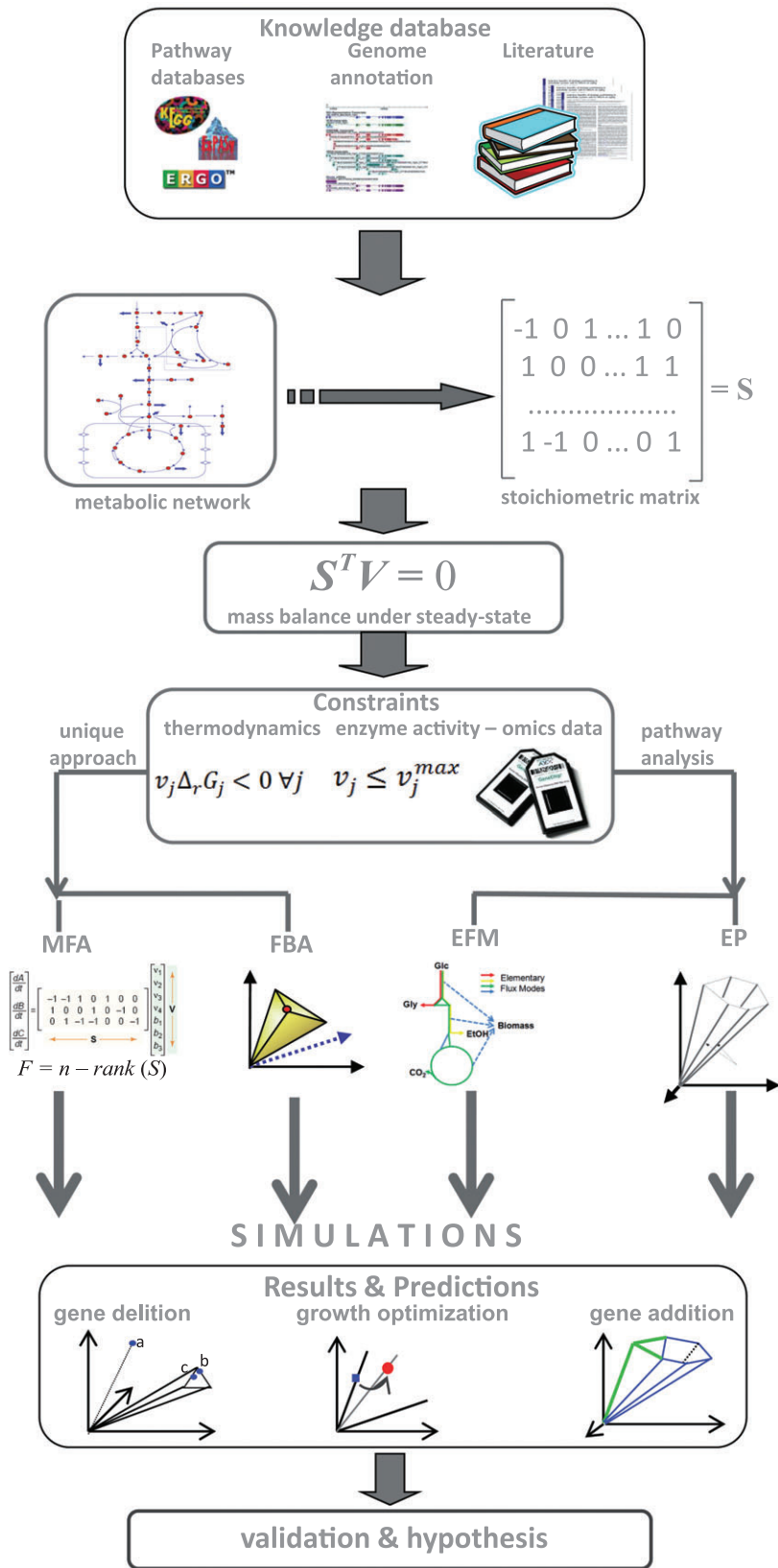


Fig. 2. Genome-scale models – flowchart. This figure illustrates the work-flow from the reconstruction of genome-scale metabolic models to their different applications as predictive tools for metabolic engineering. More detailed information about each step is included in the text.

minimizes) the selected objective function. The objective function used for microorganisms is normally the specific growth rate (Edwards *et al.*, 2002), which is consistent with the evolutionary advantage of fastest-growing species. FBA problems are solved using linear programming and the obtained solution is in a corner of the feasible region in the solution space. In order to have growth rate as an objective function, it is crucial to define a biomass stoichiometric equation which should be obtained from the biomass composition of the corresponding microorganism. The biomass equation determines the drainage of the different biomass precursors per unit of biomass produced and therefore the metabolic demands for growth. The processes involved in biomass production, such as protein synthesis or DNA replication, have also a cost in terms of ATP, which is difficult to estimate *a priori*. The right estimation of the energetic costs for biomass production is extremely relevant for the predictive power of the metabolic models.

Pathway analysis

The set of constraints will determine all possible functions of the reconstructed network or all feasible phenotypes. We refer to this set of possible solutions as a solution space. The solution space can be characterized by a set of solutions that span the whole spectrum of steady-state solutions attainable by a metabolic network. The elements of this set are EFMs and extreme pathways (EPs). A recent review on pathway analysis and its applications has been published by Trinh and colleagues (2009).

Elementary flux modes (EFMs) are minimal sets of reactions that can operate in steady state. This means that if any reaction used by an EFM is removed, the remaining reactions will not be able to operate together in steady state (Schuster *et al.*, 1999). The EFMs of a network provide an easy way to understand the effects of gene deletions due to the fact that the deletion of a particular reaction simply makes disappear all the EFMs in which it was involved. This allows the identification of essential reactions and the evaluation of the degree of coupling between different fluxes. However, the number of EFMs grows exponentially with the size of a metabolic network and makes its calculation impossible for genome-scale metabolic networks. Despite the mentioned limitations, the pathway analysis of simplified metabolic models has led to several successful metabolic engineering applications involving the identification of both deletion and overexpression targets as well as the introduction of heterologous genes.

The first successful practical application of pathway analysis for a metabolic engineering application was the optimization of 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) production in *E. coli* (Liao *et al.*, 1996).

A recent application was the development of high-ethanol-producing *E. coli* strains (Trinh *et al.*, 2008). These approaches consist essentially in the deletion of reactions involved in pathways with low yields of the product of interest and the overexpression of reactions involved in the pathways with high yields. However, as Liao and co-workers highlighted in their paper (Liao *et al.*, 1996), there are also examples of unsuccessful overexpression strategies deduced from the EFM approach.

Extreme pathways (EPs) are the elements of the convex basis that characterizes the solution space (Schilling *et al.*, 2000). They are a subset of the EFMs and none of them can be expressed as a positive linear combination of the others. Schilling and colleagues (2000) showed how the projection of the different EPs on a plane defined by the input fluxes of oxygen and the carbon source defines several differentiated regions in the phenotypic phase plane.

Algorithms for *in silico* strain optimization

Traditionally, the improvement of the industrial strains producing valuable compounds was carried out by inducing random mutations and selecting the strains that showed improvements in the production yield. This iterative blind method allowed obtaining high-yield microorganisms; however, the availability of reconstructed metabolic networks allows much more rational and efficient approaches to the problem.

The first rational manipulations of metabolic pathways were focused on genes directly connected with the product-synthesizing pathway (Mingot *et al.*, 1999; Stafford *et al.*, 2002; Koffas *et al.*, 2003; Padilla *et al.*, 2004). All these manipulations part from the identification of a key branch point in a precursor of the desired product. The flux through the path leading to the target product is then increased by deleting genes in the competing pathways or by overexpressing enzymes involved in the desired pathway. In some cases the path leading to the product of interest was not present in the wild-type strain and was introduced by the heterologous expression of enzymes from a different organism.

In order to predict the effects of genetic manipulations on a broader scale, two main methods have been used: FBA and minimization of the metabolic adjustment (MOMA). The FBA method is based on the assumption that the metabolic fluxes in a microorganism are such to maximize the cell growth yield (Edwards and Palsson, 2000). The MOMA approach (Segre *et al.*, 2002) minimizes the distance between the flux distributions of the wild-type strain and the deletion mutant. Another algorithm developed with the same purpose as MOMA is the regulatory On/Off minimization of metabolic flux changes (ROOM). It aims to minimize the number of fluxes that

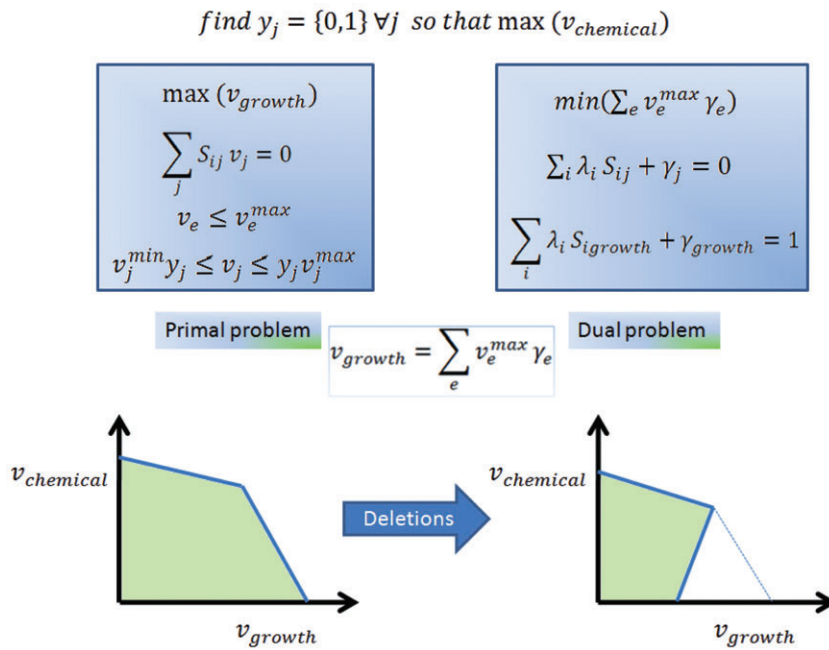


Fig. 3. Formulation of the OptKnock algorithm. The primal and dual problems for the growth optimization are formulated and their solutions are set to be the same. The use of the dual formulation transforms a nested optimization into a single optimization for v_{chemical} . The presence of a gene is expressed by multiplying the maximal and minimal rates of its corresponding reaction by the variable y_j with values 1 or 0 depending on if the gene is present or deleted. The index e represents the reactions with known maximal values (normally exchange fluxes). The rest of the reactions can be constrained to zero when the corresponding genes are deleted or take arbitrary maximal and minimal values when the genes are present. The coefficients λ and γ represent, respectively, the shadow prices for the stoichiometric and maximal rate constraints. The primal and dual problems have the same optimal solution and the shadow prices of the maximal rate constraints become zero when $y_j = 1$. The selected knockouts transform the space of solutions in a way that makes the maximal growth rate attainable correspond to a high rate of production for the target chemical.

pass zero to a non-zero flux (or the opposite) between the two metabolic states (Shlomi *et al.*, 2005).

The mentioned methods provided prediction tools for the effects of gene deletions; however, the combinatorial number of possible gene deletions to be tested required the introduction of computer algorithms to find the optimal combinations of deletions necessary to optimize the desired target.

The first algorithm introduced for microbial strain optimization was OptKnock (Burgard *et al.*, 2003). The function to be optimized by this algorithm is the production rate of the target product and the variables are the genes to be deleted. The solution involves two nested optimizations, which makes it difficult to solve. In order to overcome this problem, the dual optimization problem for growth is also formulated (Fig. 3). In this way the nested optimization can be transformed into a simple optimization of the production rate of the desired chemical with the constraints corresponding to the primal and dual problems plus the condition of equality between the two solutions.

The OptKnock algorithm was used to construct lactic acid-producing strains of *E. coli* with very positive results (Fong *et al.*, 2005).

The OptKnock algorithm found three optimal solutions for the optimization of lactic acid production. The first one involved the deletion of the *adhE* and *pta* genes (Fig. S3). This solution seems trivial, as those genes are involved in the production of ethanol and acetate, which are the two main products secreted under anaerobic conditions. The second solution involved *pta* and *pfk* deletions. The *pfk* deletion is a less trivial solution. It deviates flux through the Entner–Doudoroff pathway and

increases the production of NADH and pyruvate, the two necessary metabolites for lactic acid production. The third solution involves the already three mentioned targets plus the *glk* gene. The deletion of the glucokinase couples the phosphorylation of glucose to the transformation of PEP into pyruvate.

Three new strains were built with the *adhE-pta*, the *pta-pfk* and the *adhE-pta-pfk-glk* deletions. The designed deletion strains did not show optimal growth rates and lactic acid yields in the first generation; however, after a process of adaptive evolution, the growth rate and the lactic acid production (which is coupled to the growth rate in these strains) showed up to threefold increases. This finding clearly shows that the growth rate is a suitable objective function for microorganisms.

OptKnock has found a number of industrial applications, which is reflected in an increasing number of patents. Some examples are: 'Methods and organisms for the growth-coupled production of succinate' (Burgard and Van Dien, 2007); 'Methods and Organisms for Growth-Coupled Production of 3-Hydroxypropionic Acid' (Burgard and Van Dien, 2008a,b); 'Methods and organisms for the growth-coupled production of 1,4-butanediol' (Burgard *et al.*, 2009).

The OptKnock algorithm, when it is applied to genome-scale models, is computationally demanding, and normally only a relatively small set of deletion candidates can be tested, such as the reactions in the central carbon metabolism (Burgard *et al.*, 2003) or amino acid metabolism (Pharkya *et al.*, 2003).

Other authors (Alper *et al.*, 2005) have used an iterative approach that consists in simulating the effects of single

deletions, selecting the deletion that results in the highest increase in the product yield and performing new deletions on the selected mutant. This approach was successfully applied to the production of lycopene in *E. coli*.

The parallelism existing with the process of biological evolution and their lower computational cost make genetic algorithms an attractive tool for the design of new strains. OptGene (Fig. S4) (Patil *et al.*, 2005) is an algorithm that relates the production rate of the desired compound to the evolutionary fitness of the microorganism.

The OptGene algorithm is more computationally efficient than the OptKnock algorithm and allows using non-linear objective functions (such as the amount of product per unit of time). However OptGene does not guarantee finding the absolute optimal solution. The optimal convergence rates for OptGene were found for a population of 125 individuals and a mutation probability of 1 per genome size and generation. The method was used to optimize the production of vanillin, glycerol and succinate in *S. cerevisiae* and it converged to the final solution in less than 1000 generations.

OptGene has been used to improve the production of sesquiterpenes in *S. cerevisiae* (Asadollahi *et al.*, 2009). It has also inspired the apparition of new genetic algorithms such as CIED (Fowler *et al.*, 2009), which differs in the mutant selection method and has been used to increase the production of flavanone in *E. coli*.

In order to illustrate OptGene's applications it is interesting to mention its recent utilization for the design of high-succinic-acid-producing strains of *S. cerevisiae* (J.M. Otero, D. Cimini, K.R. Patil, S.G. Poulsen, L. Olsson and J. Nielsen, submitted). The use of yeasts for the production of succinic acid has some advantages in relation to the traditional producing microorganisms (all of them prokaryotic). *Saccharomyces cerevisiae* can grow under pH conditions between 3 and 6, which allows the production of succinic acid and not succinate salts, as it is the case under the neutral pH conditions necessary for bacteria. The acid growth medium also protects the reactor against bacterial contamination.

The OptGene algorithm proposed a combination of three gene deletions for the overproduction of succinic acid: *sdh3*, *ser3* and *ser33*. The *sdh3* enzyme catalyses the transformation of succinate into fumarate in the Krebs cycle; this is a straightforward solution. The deletions of *ser3* and *ser33* act in a more complex way. The deletion of these genes cuts the pathway from 3-P-Glycerate to serine. After this deletion, the only way left to the cell to produce serine (which is necessary for growth) is to synthesize it from glycine (Fig. S5). Glycine is obtained from glyoxylate which is itself synthesized from isocitrate-producing succinate as a by-product that must be secreted. In this way the production of succinate is coupled to the cell growth.

The mutant *sdh3Δser3Δser33Δ* was constructed. The new strain was cultivated in a series of six flasks with decreasing concentrations of glycine in order to create a selective pressure favouring the production of glycine from isocitrate and therefore releasing succinate. An increase of almost eightfold in the succinate yield was observed in the evolved strain. Three more cultures in absence of glycine were performed in order to select the strain for faster growth.

The algorithms mentioned so far are used to improve microbial phenotypes by performing gene deletions. In many cases, metabolic engineering involves expressing heterologous genes in host organisms that lack reactions involved in the production of the desired compound. The creators of the OptKnock algorithm developed the OptStrain algorithm (Pharkya *et al.*, 2004), which uses a database of known biological transformations to find the minimal set of non-native genes to be expressed in a host organism in order to obtain the desired product with an optimal yield.

OptStrain is limited by the reactions contained in the available database. A more general algorithm named BNICE (Hatzimanikatis *et al.*, 2005) has been proposed to generate *de novo* pathways using generalized enzyme reactions. This method can involve metabolic intermediates that are not included in the available reaction databases.

As we have seen, the existing algorithms for strain optimization rely on stoichiometric considerations and do not include regulatory information. The obtained outputs are sets of reactions to be removed or added to a metabolic network in order to modify its topology and couple the growth rate with the production of the desired compounds. The effects of overexpression of metabolic genes are still poorly understood and a systematic algorithm to find overexpression targets with efficiency comparable to OptKnock or OptGene is still missing.

Integration of thermodynamics in the analysis of metabolic networks

According to the second law of thermodynamics, a chemical transformation at constant pressure and temperature occurs in the direction of negative Gibbs free energy.

The thermodynamic feasibility condition must be satisfied by all the reactions in the metabolic network simultaneously. This condition allows defining additional constraints to the mass balances. It has been used to determine the feasibility of biosynthetic pathways under physiological conditions (Mavrovouniotis, 1993; Pissarra and Nielsen, 1997) and is a useful tool for the design of *de novo* biosynthetic pathways (Hatzimanikatis *et al.*, 2005).

The analysis of thermodynamic feasibility in metabolic networks has been denominated Network-embedded

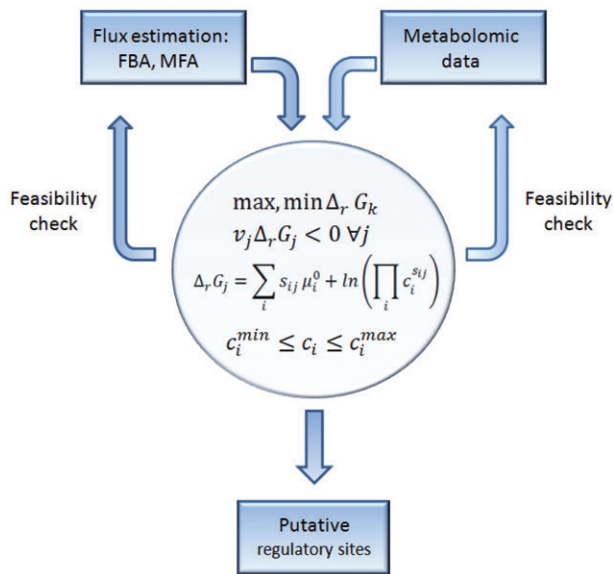


Fig. 4. Schema of NET analysis. Each reaction in the network has to be thermodynamically feasible. This limits the interval of variation of the metabolite concentrations.

thermodynamic analysis (NET analysis, Fig. 4) and is based on the integration of metabolomics data (Kümmel *et al.*, 2006). NET analysis consists in the application of an optimization procedure to find the interval of variation of reaction Gibbs free energy for each metabolic reaction. The concentrations of unmeasured metabolites are allowed to vary between 0.001 and 10 mM (Fraenkel, 1992).

Network-embedded thermodynamic analysis can be used to check the thermodynamic consistency of metabolomic data and the assumed flux directions. It also allows specifying concentration ranges for unmeasured metabolites and determining whether each reaction operates far or close to the equilibrium. The reactions operating far from the equilibrium are more likely to be flux controlling (Wang *et al.*, 2004).

There are several important limitations for the use of NET analysis. The metabolomics data are often incomplete and lacking enough precision to predict accurately the *in vivo* chemical potentials. For example, from seven metabolomics data sets analysed (Kümmel *et al.*, 2006), only four were thermodynamically consistent. The effects of ionic strength and pH have been often overlooked and in many cases change completely the results of the analysis (Maskow and von Stockar, 2005). The standard chemical potentials of most of the metabolites in genome-scale models are unknown and the errors in their estimations play also a role in decreasing the accuracy of the thermodynamic analysis. An improved group contribution method for the determination of standard chemical potentials in complex metabolic networks has been recently published (Jankowski *et al.*, 2008).

The first effort to estimate the reaction Gibbs free energies for all the reactions in a genome-scale model (Henry *et al.*, 2006) led to the iHJ873 model for *E. coli*. This model included standard Gibbs free energies for each of its reactions. It contains fewer reactions than the iJR904 model because those reactions involving metabolites with unknown chemical potentials were lumped together. The iHJ873 model was used to identify the less thermodynamically favourable reactions in the network and analyse the biological implications of removing them from the model.

The same authors moved a step forward (Henry *et al.*, 2007) and included in their analysis the effects of ionic strength and pH. In the same paper, an improved analysis framework with respect to NET analysis was proposed. The new framework was named thermodynamics-based metabolic flux analysis (TMFA) and differs from NET analysis in the fact that the reaction rates are not obtained independently of the thermodynamic analysis. In TMFA, the thermodynamic directionality constraints are added to the mass balance constraints during the calculation of the flux distribution.

The flux distributions obtained using TMFA do not include loops, which are by definition thermodynamically infeasible. The infeasibility of loops has been mentioned in the literature (Beard *et al.*, 2004); however, a reliable method to remove loops from the flux distributions has not yet been implemented in the common calculus packages. TMFA would be the ideal framework.

Other important finding of Henry and co-workers was the fact that the reactions that operate far from the equilibrium are likely to be situated at the beginning of linear pathways or in branching points. This is consistent with their expected role as flux-controlling steps.

The range of concentration variability allowed for the metabolites was between 20 and 0.001 mM. Only one essential reaction appeared to be infeasible under the used concentration rank, it was the dihydroorotase. The reaction became feasible by allowing the estimated reference Gibbs free energy to vary in an interval of two standard deviations (obtained from the group contribution method) or increasing the concentration rank of its metabolites by 20%. The thermodynamic bottleneck nature of dihydroorotase is consistent with the fact that in mammals, this enzyme belongs to a multi-enzyme complex, which could be a way to use substrate channelling to overcome unfavourable Gibbs free energies.

The main weakness of the existing thermodynamic analysis approaches is the fact that they assume a concentration rank for the metabolites *a priori*. Once the reaction directions have been determined using TMFA, a new concentration rank can be defined for each metabolite; however, this rank is still very broad (Henry *et al.*,

2007) and is not really a useful predictor of the cellular metabolic pools.

Conclusions and perspectives

Mathematical models have shown to be promising tools for the design of new highly productive microbial strains. These models provide a predictive framework that allows testing *in silico* the outcomes of genetic manipulations. In this way some manipulations can be discarded *a priori* and others can be selected for further validation. This allows saving both time and resources in the strain development process.

Mechanistic models based on differential equations have been applied to study relatively small systems such as metabolic or signalling pathways but their scale-up at the genome-scale level has proved a complex task that would require collecting kinetic information from all the biochemical processes in the cell. Genome-scale stoichiometric metabolic models have together with various simulation algorithms shown to be very efficient for identification of complex metabolic engineering strategies. A major advantage of these models is that they do not rely on any information about the kinetics of the individual reactions within the metabolic network.

In this review some interesting examples of the applications of mathematical modelling to the design of efficient cell factories have been discussed. Biological systems are still too complex to be analysed in the deterministic way typically applied in other branches of science, but systems biology is providing an increasing number of success stories in the field of industrial biotechnology. These examples are a very encouraging trend that could be the beginning of a shift from the traditionally descriptive approach of biology to a predictive one.

Acknowledgements

The authors would like to thank Dina Petranovic for helpful discussions and ideas. We acknowledge the Chalmers Foundation, Knut and Alice Wallenberg Foundation and the EU-funded project SYSINBIO (contract No. 212766) for funding.

References

Albertyn, J., Hohmann, S., and Prior, B.A. (1994) Characterization of the osmotic-stress response in *Saccharomyces cerevisiae*: osmotic stress and glucose repression regulate glycerol-3-phosphate dehydrogenase independently. *Curr Genet* **25**: 12–18.

Alper, H., Jin, Y.S., Moxley, J.F., and Stephanopoulos, G. (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* **7**: 155–164.

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

Barberis, M., Klipp, E., Vanoni, M., and Alberghina, L. (2007) Cell size at S phase initiation: an emergent property of the G1/S network. *PLoS Comput Biol* **3**: 649–666.

Beard, D.A., Liang, S.D., and Qian, H. (2002) Energy balance for analysis of complex metabolic networks. *Biophys J* **83**: 79–86.

Beard, D.A., Babson, E., Curtis, E., and Qian, H. (2004) Thermodynamic constraints for biochemical networks. *J Theor Biol* **228**: 327–333.

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.O., and Herrgard, M.J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* **2**: 727–738.

Burgard, A.P., and Van Dien, S.J. (2007) Methods and organisms for the growth-coupled production of succinate. Patent: 20070111294 (USA).

Burgard, A.P., and Van Dien, S.J. (2008a) Methods and organisms for the growth-coupled production of 3-hydroxypropionic acid. Patent: 20080199926 (USA).

Burgard, A.P., and Van Dien, S.J. (2008b) Methods and organisms for the growth-coupled production of 1,4-butanediol. Patent: 20090047719 (USA).

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

Burgard, A.P., Van Dien, S.J., and Burk, M.J. (2009) Methods and organisms for the growth-coupled production of 1,4-butanediol. Patent: WO/2009/023493.

Chassagnole, C., Rizzi, N.N., Schmid, J.W., Mauch, K., and Reuss, M. (2002) Dynamic modeling of central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng* **79**: 53–73.

Chen, K.C., Csikasz-Nagy, A., Gyorfy, B., Val, J., Novak, B., and Tyson, J.J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol Biol Cell* **11**: 369–391.

Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., and Tyson, J.J. (2004) Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell* **15**: 3841–3862.

Curto, R., Sorribas, A., and Cascante, M. (1995) Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model definition and nomenclature. *Math Biosci* **130**: 25–50.

Cvijovic, M., Olivares-Hernandez, R., Agren, R., Dahr, N., Vongsangnak, W., Nookaew, I., *et al.* (2010) BioMet Toolbox: genome-wide analysis of metabolism. *Nucleic Acids Res* **38**: 144–149.

Edwards, J.S., and Palsson, B.O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97**: 5528–5533.

Edwards, J.S., Covert, M., and Palsson, B. (2002) Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol* **4**: 133–140.

- Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.Ø. (2009) Reconstruction of biochemical networks in microbial organisms. *Nat Rev Microbiol* **7**: 129–143.
- Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., and Palsson, B.O. (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* **91**: 643–648.
- Fowler, Z.L., Gikandi, W.W., and Koffas, M.A. (2009) Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl Environ Microbiol* **75**: 5831–5839.
- Fraenkel, D.C. (1992) Genetics and intermediary metabolism. *Annu Rev Genet* **26**: 159–177.
- Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., and Broadbelt, L.J. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* **21**: 1603–1609.
- Henry, C.S., Jankowski, M.D., Broadbelt, L.J., and Hatzimanikatis, V. (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys J* **90**: 1453–1461.
- Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2007) Thermodynamics-based metabolic flux analysis. *Biophys J* **92**: 1792–1805.
- Hoefnagel, M.H., Starrenburg, M.J., Martens, D.E., Hugenholz, J., Kleerebezem, M., Van, S., II, *et al.* (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology* **148**: 1003–1013.
- Hohmann, S. (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* **66**: 300–372.
- Hohmann, S., Krantz, M., and Nordlander, B. (2007) Yeast osmoregulation. *Methods Enzymol* **428**: 29–45.
- Jankowski, M.D., Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys J* **95**: 1487–1499.
- von Kamp, A., and Schuster, S. (2006) Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* **22**: 1930–1931.
- Kholodenko, B.N. (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* **7**: 165–176.
- Klamt, S., Stelling, J., Ginkel, M., and Gilles, E.D. (2003) FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* **19**: 261–269.
- Klipp, E. (2007) Modelling dynamic processes in yeast. *Yeast* **24**: 943–959.
- Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., and Hohmann, S. (2005) Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* **23**: 975–982.
- Klipp, E., Liebermeister, W., Helbig, A., Kowald, A., and Schaber, J. (2007) Systems biology standards – the community speaks. *Nat Biotechnol* **25**: 390–391.
- Koffas, M.A., Jung, G.Y., and Stephanopoulos, G. (2003) Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab Eng* **5**: 32–41.
- Kümmel, A., Panke, S., and Heinemann, M. (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* **2**: 2006–0034.
- Liao, J.C., Hou, S.Y., and Chao, Y.P. (1996) Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol Bioeng* **52**: 129–140.
- Maskow, T., and von Stockar, U. (2005) How reliable are thermodynamic feasibility statements of biochemical pathways? *Biotechnol Bioeng* **92**: 223–230.
- Mavrovouniotis, M.L. (1993) Identification of localized and distributed bottlenecks in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol* **1**: 275–283.
- Milne, C.B., Kim, P.J., Eddy, J.A., and Price, N.D. (2009) Accomplishments in genome-scale *in silico* modeling for industrial and medical biotechnology. *Biotechnol J* **4**: 1653–1670.
- Mingot, J.M., Penalva, M.A., and Fernandez-Canon, J.M. (1999) Disruption of *phacA*, an *Aspergillus nidulans* gene encoding a novel cytochrome P450 monooxygenase catalyzing phenylacetate 2-hydroxylation, results in penicillin overproduction. *J Biol Chem* **274**: 14545–14550.
- de Nadal, E., Alepuz, P.M., and Posas, F. (2002) Dealing with osmostress through MAP kinase activation. *EMBO Rep* **3**: 735–740.
- Nielsen, J. (2001) Metabolic engineering. *Appl Microbiol Biotechnol* **55**: 263–283.
- Nielsen, J., and Jewett, M.C. (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res* **8**: 122–131.
- Novak, B., and Tyson, J.J. (1997) Modeling the control of DNA replication in fission yeast. *Proc Natl Acad Sci USA* **94**: 9147–9152.
- Novak, B., Csikasz-Nagy, A., Gyorfy, B., Chen, K., and Tyson, J.J. (1998) Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys Chem* **72**: 185–200.
- Novak, B., Pataki, Z., Ciliberto, A., and Tyson, J.J. (2001) Mathematical model of the cell division cycle of fission yeast. *Chaos* **11**: 277–286.
- Otero, J.M., and Nielsen, J. (2010) Industrial systems biology. *Biotechnol Bioeng* **105**: 439–460.
- Padilla, L., Kramer, R., Stephanopoulos, G., and Agosin, E. (2004) Overproduction of trehalose: heterologous expression of *Escherichia coli* trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in *Corynebacterium glutamicum*. *Appl Environ Microbiol* **70**: 370–376.
- Palsson, B. (2002) *In silico* biology through ‘omics’. *Nat Biotechnol* **20**: 649–650.
- Papin, J.A., Hunter, T., Palsson, B.O., and Subramaniam, S. (2005) Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* **6**: 99–111.
- Patil, K.R., Akesson, M., and Nielsen, J. (2004) Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* **15**: 64–69.
- Patil, K.R., Rocha, I., Forster, J., and Nielsen, J. (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* **6**: 308.

- Pharkya, P., Burgard, A.P., and Maranas, C.D. (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* **84**: 887–899.
- Pharkya, P., Burgard, A.P., and Maranas, C.D. (2004) Opt-Strain: a computational framework for redesign of microbial production systems. *Genome Res* **14**: 2367–2376.
- Pissarra, P.D., and Nielsen, J. (1997) Thermodynamics of metabolic pathways for penicillin production: analysis of thermodynamic feasibility and free energy changes during fed-batch cultivation. *Biotechnol Prog* **13**: 156–165.
- Price, N.D., Famili, I., Beard, D.A., and Palsson, B.O. (2002) Extreme pathways and Kirchhoff's second law. *Biophys J* **83**: 2879–2882.
- Price, N.D., Papin, J.A., Schilling, C.H., and Palsson, B.O. (2003) Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol* **21**: 162–169.
- Rizzi, M., Baltés, M., Theobald, U., and Reuss, M. (1997) *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnol Bioeng* **55**: 592–608.
- Savageau, M.A. (1969a) Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* **25**: 365–369.
- Savageau, M.A. (1969b) Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* **25**: 370–379.
- Savageau, M.A. (1970) Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol* **26**: 215–226.
- Schilling, C.H., and Palsson, B.O. (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J Theor Biol* **203**: 249–283.
- Schilling, C.H., Letscher, D., and Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic functions from a pathway-oriented perspective. *J Theor Biol* **203**: 229–284.
- Schuster, S., Dandekar, T., and Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biothechnol* **17**: 53–60.
- Segre, D., Vitkup, D., and Church, G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99**: 15112–15117.
- Schilling, C.H., Edwards, J.S., Letscher, D., and Palsson, B.O. (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng* **71**: 286–306.
- Shlomi, T., Berkman, O., and Ruppin, E. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* **102**: 7695–7700.
- Sorribas, A., Curto, R., and Cascante, M. (1995) Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model validation and dynamic behavior. *Math Biosci* **130**: 71–84.
- Stafford, D.E., Yanagimachi, K.S., Lessard, P.A., Rijhwani, S.K., Sinskey, A.J., and Stephanopoulos, G. (2002) Optimizing bioconversion pathways through systems analysis and metabolic engineering. *Proc Natl Acad Sci USA* **99**: 1801–1806.
- Stephanopoulos, G. (2002) Metabolic engineering by genome shuffling. *Nat Biotechnol* **20**: 666–668.
- Teusink, B., Walsh, M.C., van Dam, K., and Westerhoff, H.V. (1998) The danger of metabolic pathways with turbo design. *Trends Biochem Sci* **23**: 162–169.
- Teusink, B., Passarge, J., Reijenga, C.A., Esgalhado, E., van der Weijden, C.C., Schepper, M., *et al.* (2000) Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* **267**: 5313–5329.
- Trinh, C.T., Unrean, P., and Sreenc, F. (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol* **74**: 3634–3643.
- Trinh, C.T., Wlaschin, A., and Sreenc, F. (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* **81**: 813–826.
- Tyo, K.E., Kocharin, K., and Nielsen, J. (2010) Toward design-based engineering of industrial microbes. *Curr Opin Microbiol* **13**: 255–262.
- Van Wuytswinkel, O., Reiser, V., Siderius, M., Kelders, M.C., Ammerer, G., Ruis, H., and Mager, W.H. (2000) Response of *Saccharomyces cerevisiae* to severe osmotic stress: evidence for a novel activation mechanism of the HOG MAP kinase pathway. *Mol Microbiol* **37**: 382–397.
- Wang, L., Birol, I., and Hatzimanikatis, V. (2004) Metabolic control analysis under uncertainty: framework development and case studies. *Biophys J* **87**: 3750–3763.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Flux control coefficients of the reactions in the system on the glucose uptake by phosphotransferase system.

Fig. S2. Schematic representation of pyruvate branch in *L. lactis*.

Fig. S3. The OptKnock algorithm found four deletions to improve the lactic acid production in *E. coli*.

Fig. S4. Flow chart of an evolutionary algorithm. OptGene defines a population with a fixed number of individuals. Each individual is characterized by a 'chromosome', which is a list of genes labelled with ones if they are present or zeros if they are absent. The algorithm can be initialized by assigning present status to all genes or by assigning present or absent status randomly. Each individual in the population is assigned a fitness score that determines whether it will be propagated to the next generation. The fitness score is calculated using FBA, MOMA or other optimization criteria. The individuals to be propagated to the next generation are selected with a probability determined from the fitness score. The 'chromosomes' of the selected individuals are crossed over to generate the next population and the next iteration starts. The iterations are repeated until an individual with the desired phenotype appears.

Fig. S5. Deletions coupling succinate production to growth. After deleting the genes *ser3* and *ser33*, the cell is forced to synthesize the serine necessary for growth from glyoxylate. Glyoxylate itself is obtained from isocitrate, generating succinate as a by-product. The deletion of *sdh3* avoids the degradation of succinate, which has to be secreted from the cell. The first-generation mutant grows very slowly and needs external serine supply, but after evolution in a medium with decreasing serine concentration the cells become adapted to produce serine via glyoxylate-secreting succinate as a by-product.

Table S1. Overview of available genome-scale metabolic models of some of the industrially most exploited organisms.

Table S2. Flux control coefficients predicted by the model through acetolactate synthesis (ALS) in *Lactococcus lactis*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.