Identification of Genome-Scale Metabolic Network Models Using Experimentally Measured Flux Profiles

Paula Jouhten

011735018

Helsinki 22.02.2007

1 Introduction

The capability of an organism to perform biochemical conversions is encoded in the genome. The genome contains the genes that encode the enzymes that perform the biochemical reactions. The complement of the biochemical reactions encoded in the genome is modelled as a genome-scale metabolic network preserving the connections between the enzymes performing the reactions and the metabolites that are either substrates or products of the reactions. The reactions are thus connected through shared metabolites and form a network. A metabolic flux is the rate at which a metabolite is converted to another in a reaction. The metabolic fluxes in a cell are the ultimate physiological response to the gene function and to the extracellular conditions.

Cell function is regulated in different functional levels. Gene expression is under regulation, enzyme synthesis can be controlled and enzyme activities can be adjusted for examples by inhibitors and activators. The regulation is sensitive to both external and intracellular conditions through sensor mechanisms. The regulation leads to adjusting the biochemical conversion capabilities according to the particular requirements. Thus, all the reactions are not necessarily active at the same time and the activities of the reactions can vary extensively depending on the conditions.

Genome-scale metabolic network reconstructions for micro-organisms have been published (Duarte *et al.*, 2004; Förster *et al.*, 2003). *In silico* predictions of the function of an organism's metabolism: the optimal growth rate, the maximal production rates and corresponding metabolic flux distributions, the viability after gene knockout, can be made using the metabolic network (Famili *et al.*, 2003). The reconstruction process of a genome-scale metabolic network requires gene annotation information for coupling the genes with correct biochemical reactions to be included in the metabolic network. However, the gene annotation information for many organisms is incomplete. Thus, the reconstruction of the ill-characterised organisms leads in formation of gaps in the metabolic network (Herrgård *et al.*, 2006).

The optimal metabolic network identification (OMNI) method aims to determine a set of active reactions that best agree between the *in silico* predicted and experimentally

measured metabolic flux distributions in a cell. The OMNI method can be used to fill gaps in the metabolic networks of ill-characterised organisms, to identify reactions that are inactive or extensively down-regulated and forming bottlenecks in evolved and engineered strains and even to evaluate functions of poorly annotated enzymes and to identify correct alternative reaction mechanisms (Herrgård *et al.*, 2006).

2 Constraint-based modelling

Constraint-based modelling is used for studying phenotypical behaviour of an organism. The existing large metabolic models are stoichiometric models describing the metabolic capabilities of the organisms. The models describe the metabolite-enzyme connectivities and the stoichiometries of the possible reactions. The reaction stoichiometries define the reactants and their molar ratios in the reactions. The constraint-based modelling approach complements the stoichiometric model with additional physicochemical constraints such as thermodynamic and reaction rate capacity constraints (Price *et al.*, 2003). The stoichiometric and additional constraints define the possible metabolic flux distribution outcome of a metabolic network, the feasible flux distribution solution space, Figure1. A metabolic flux can be thought as the rate at which material is processed through a reaction or in other words a reaction rate in moles per unit of time. A metabolic flux distribution is a vector of values for each of the metabolic fluxes in a cell in defined conditions and at a specific time point. A complement of all the metabolic fluxes in a cell is a fluxome. All the points in the defined feasible flux distribution solution space, that is a polyhedral cone, are possible physiological states of a cell. With only few experimentally measured parameters, e.g. the extracellular consumption and production rates and the growth rate, the constraint-based models can be used to determine an optimal metabolic flux distribution in a cell given a particular objective function using flux balance analysis (Famili *et al.*, 2003)
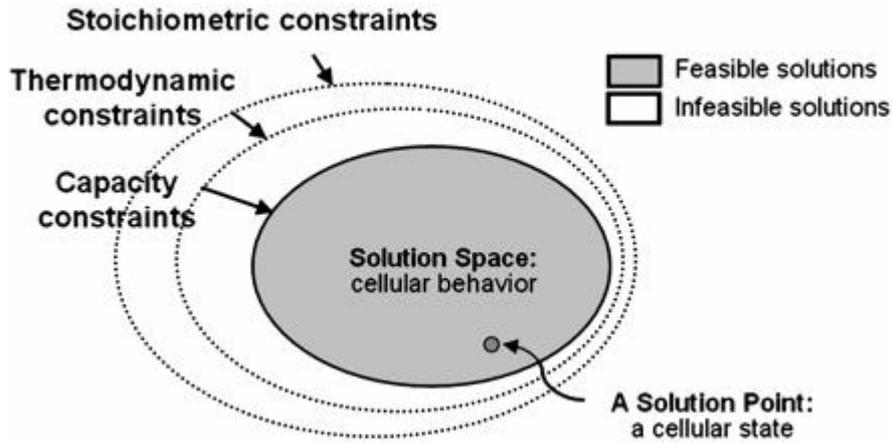
Figure 1. Constraint-based modelling approach integrates stoichiometric, thermodynamic and reaction rate capacity constraints to define a space of feasible metabolic flux distribution solutions where the cellular state solution points lie (Famili *et al.*, 2003).

3 Flux balance analysis

The metabolic flux analysis (MFA) aims to determine the flux distribution in a cell. The extracellular fluxes such as the consumption and production rates of compounds by cells and the growth rate can be experimentally determined but the intracellular metabolic fluxes cannot be measured directly. The stoichiometric coefficients, the molar ratios of participating reactions, from a stoichiometric metabolic model can be collected into a stoichiometric matrix N, the metabolites in rows and fluxes in columns. The stoichiometric matrix describes the metabolic system and is independent of time, the metabolite concentrations and kinetics of the enzymes catalysing the reactions. If the metabolic flux distribution is the vector v, the changes of the metabolite concentrations over time can be described by a system of differential equations, Equation 1 (Klamt and Stelling, 2006).

$$\frac{dc(t)}{dt} = N \cdot v(t) \tag{1}$$

Where $c(t)$ is a vector of the metabolite concentrations. The elements of the metabolic flux distribution vector, the metabolic fluxes, are functions of time, the time-dependent metabolite concentrations and often not exactly known parameters p describing the kinetic properties of the enzymes catalysing the reactions, Equation 2 (Klamt and Stelling, 2006).

$$v(t) = f(c(t), p, t) \tag{2}$$

A steady-state of metabolism can be assumed when cells are growing at constant rate since the biochemical conversions are fast compared to regulatory events. Constant growth experiments are usually continuous cultivations of a population of micro-organisms where there is a constant supply of substrates and a constant withdrawal of culture broth and mixing of the culture broth can be assumed ideal. A limiting substrate holds the growth rate constant. Under a steady-state assumption the reaction rates and the intracellular metabolite pool sizes can be assumed to remain constant over time (Stephanopoulos *et al.*, 1998). Thus, the changes of the metabolite concentrations in time can be set to zero and a homogenous system of linear mass balance equations for the metabolites is obtained, Equation 3, (Bonarius *et al.*, 1997).

$$0 = N \cdot v \tag{3}$$

The dilution effect of the macromolecule pools caused by the growth of cells is usually modelled as an artificial reaction consuming macromolecular cell biomass constituents and producing an artificial metabolite biomass (Stephanopoulos *et al.*, 1998).

The system in Equation 3 is usually underdetermined because of a high number of degrees of freedom. The trivial solution to Equation 3, $v = 0$, would represent a thermodynamic equilibrium instead of a living organism. The stoichiometric matrix N and the metabolic flux distribution vector v can be rearranged into measured and unknown parts so that the measured parts contain the measured extracellular fluxes and the rest of the fluxes are considered in the unknown part, Equation 4 (Klamt *et al.*, 2002).

$$0 = N \cdot v = N_m \cdot v_m + N_c \cdot v_c \tag{4}$$

Where v_m contains the measured fluxes and N_m contains the stoichiometric coefficients of the measured fluxes and v_c contains the rest of the fluxes and N_c

contains the rest of the stoichiometric coefficients (Stephanopoulos *et al.*, 1998). If rank(N_c) < number of unknown fluxes, the system is underdetermined. If rank(N_c) = number of unknown fluxes, the system is determined. If rank(N_c) < number of metabolites in the system, the system is redundant and thus there are rows in N_c that can be expressed as linear combinations of other rows. If rank(N_c) = number of metabolites in the system, the system is not redundant (Klamt *et al.*, 2002). A unique intracellular flux distribution can be determined if the system is determined and not redundant. Thus the non-singular square matrix N_c can be inverted and the v_c can be solved, Equation 5, (Stephanopoulos *et al.*, 1998).

$$v_c = -\left(N_c\right)^{-1} N_m \cdot v_m \tag{5}$$

However, due to the complex structures of metabolic networks, a MFA system usually is underdetermined. Isotopic-tracer experiments, in particular carbon-13 labelling experiments, are the means to provide additional experimental input data. The label is introduced into cells by feeding them with labelled substrate. The fate of the label depends on the active metabolic pathways in the cell and the fate of the isotopic label can be determined either from the metabolites or from the biomass constituents as amino acids either by mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectroscopy (Stephanopoulos *et al.*, 1998). The drawback is that the isotopically labelled compounds are expensive, the experiments are laborious and they do not always provide enough independent data to enable reliable computational estimation of the intracellular flux distribution.

The flux balance analysis (FBA) is a particular case of the metabolic flux analysis. The FBA is based on solving a system of metabolite mass balance equations under a steady state assumption with external measured consumption and production fluxes and growth rates as the only experimental input data (Bonarius *et al.*, 1997). If the metabolite mass balance equation system remains underdetermined a linear objective function is set for linear programming and additional constraints from a constraint-based model are included and thus the optimal flux distribution is obtained as a solution of (Herrgård *et al.*, 2006)

$$\max a^T v$$

subject to $N \cdot v = 0$

$$0 \le v_i \le v_i^{\max} .$$ (6)

where the vector a contains the objective coefficients.

However, setting up a biologically meaningful objective function is not a simple task. Maximising growth is a suitable objective function for many micro-organisms in many growth conditions and has even been shown to be valid for knock-out strains after evolving in the specific growth conditions (Edwards *et al.*, 2001). On the other hand there is experimental evidence of even bacterial strains that do not optimise for growth but compromise between the growth requirements and robustness functions (Fischer and Sauer, 2005). Eukaryotic cells like yeasts may optimise for growth in some conditions (Famili *et al.*, 2003) but for example *Saccharomyces cerevisiae* shows respirofermentative metabolism in glucose excess conditions due to carbon source repression that limits the respiration and optimal biomass yield. Mammalian cells in tissues are extreme examples of cells that cannot be thought to optimise growth but other functions instead.

The biosynthetic fluxes are coupled to the growth rate of an organism. They are either put into the objective function when the growth is optimised or they are considered as additional measured fluxes when an estimate of the biomass composition is available and an artificial biomass synthesis flux can be set up (Herrgård *et al.*, 2006). In an objective function the stoichiometry of the cell biomass synthesising reactions has to be known exactly. The macromolecular composition of cell biomass can be experimentally measured but some of the methods provide results that are prone to large experimental errors (Lange and Heijnen, 2001). Furthermore, the biosynthetic routes of many organisms are not completely known and may vary depending on the growth conditions as the macromolecular composition of the biomass. Also the compositions of macromolecules may vary growth condition dependently.

The constraint-based modelling approach obviously provides erroneous predictions on the metabolic flux distributions if the structure of the metabolic network model is

incorrect. Discrepancies between the model predictions and experimental data are observed if the model stoichiometry includes errors. The Optimal Metabolic Network Identification (OMNI) method can be used to define the correct active set of reactions in the stoichiometric model to improve the model predictions (Herrgård *et al.*, 2006).

4 Optimal Metabolic Network Identification (OMNI)

The OMNI method aims to identify the correct active set of reactions from a larger set of potentially active enzymes for example from a genome-scale constraint-based metabolic model. The method is applicable to different type of cases where the active network structure is unclear. If the metabolic network reconstruction contains gaps, the OMNI method can be used to identify the gap-filling reactions that make the model predictions optimally agree with the experimental data from a set of potentially active reactions. If an evolved strain does not show optimal growth, the OMNI method can be used to identify the possible bottleneck reactions, which are either inactive or severely down-regulated (Herrgård *et al.*, 2006).

4.1 Bilevel optimisation problem

The OMNI is based on measured rates, the growth rate and the substrate uptake and product secretion rates, and on intracellular flux data from MFA. The problem of finding an optimal active reaction set to match the model predictions with experimental data is formulated as a bilevel mixed-integer optimisation problem, Figure 2. The inner problem solves a flux balance analysis problem for the particular metabolic network structure. Thus the reaction set included in the model is fixed for the inner problem. The feasible flux distribution solution space, the polyhedral cone, is different for the different reaction sets. In the inner FBA problem an optimal flux distribution is solved for the particular feasible solution space. Whereas the outer problem searches for an optimal network structure, in other words a best set of active reactions, to obtain predictions that match optimally with the experimental data (Herrgård *et al.*, 2006).

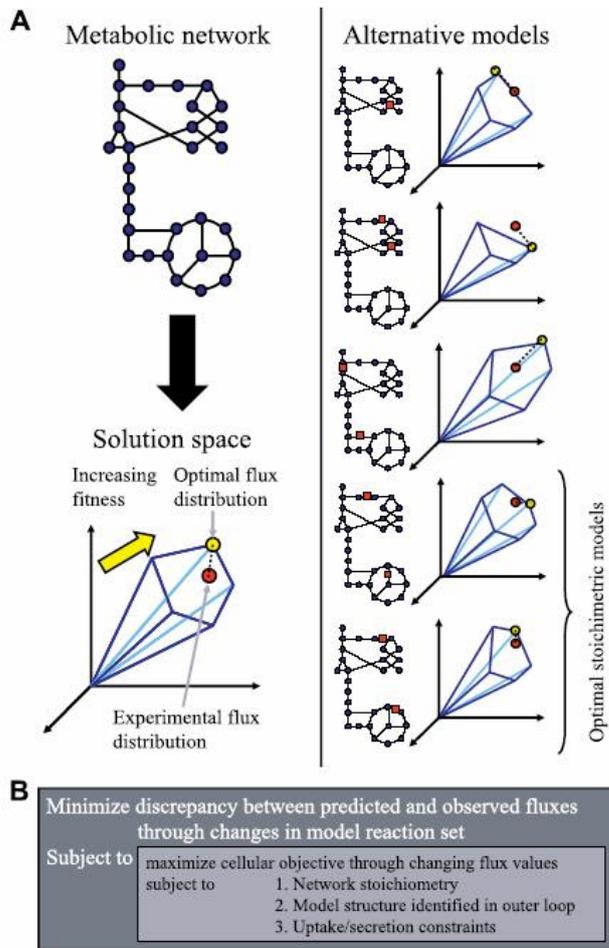**Figure 2.** An illustration of the bilevel optimisation approach of the OMNI method. (A) The altered model structure results in changes in the FBA-predicted optimal flux distribution. (B) The optimal network structure identification is formulated as a bilevel optimisation problem (Herrgård *et al.*, 2006).

Mathematically the bilevel-optimisation problem is formulated as (Herrgård *et al.*, 2006):

$$y^{opt} = \arg\min_{y} \sum_{i \in M} \omega_i \left| v_i^{opt} - v_i^{\exp} \right|$$

$$\text{subject to} \quad v^{opt} = \arg\max_{v} a^T v$$

$$\text{subject to} \quad N \cdot v = 0$$

$$0 \leq v_j \leq v_j^{\max} \quad j \in F$$

$$0 \leq v_k \leq v_k^{\max} y_k \quad k \in D \tag{7}$$

$$v_l = v_l^{\exp} \quad l \in E$$

$$v_{biomass}^{opt} \geq v_{biomass}^{\min}$$

$$y_k = \{0,1\} \forall k \in D$$

$$\sum_{k \in D} (1 - y_k) = K$$

The vector a contains the objective coefficients. The vector v^{max} contains the maximum fluxes from the model constraints. The vector v^{opt} is the optimal flux distribution for a given set of reactions and v^{exp} is the experimentally measured flux distribution. The vector w contains weights for the measured fluxes. The vector y is a binary vector. The elements of the vector y indicate if the reaction is included in the model or not. The corresponding reactions to the elements of y that get the value one are included in the model. F is a set of reactions that are essential and cannot be excluded from the model and D is a set of reactions that can be removed from the model to improve the model predictions. The reactions in the set F have high evidence of being active whereas the set D contains the potential reactions to be inferred. The set D can, for example, contain all the reactions that are not included in the set F in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2004) for filling gaps in the reconstructed metabolic networks of poorly characterised organisms. K is the number of allowed reaction deletions. M is the set of reactions with measured fluxes. The set E contains the reactions with constrained fluxes. The set E can include measured extracellular fluxes. In general it is possible to divide reactions to sets M and E in different ways. The fluxes in the constraint set E are assumed to be measured more accurately than the fluxes in the set M that are used in the objective (Herrgård *et al.*, 2006).

9

The inner problem is a FBA problem that is solved for a particular set of reactions defined by the binary vector y, the elements of the vector y are parameters, whereas the elements of y are variables in the outer problem that is solved for minimised discrepancy between the optimal predicted flux distribution and the experimental flux distribution. The outer problem searches through the space of possible binary vectors y, that is dependent on the reaction sets F and D, for the best possible metabolic model structure given the experimental flux distribution (Herrgård *et al.*, 2006). The OptKnock computational strain design method utilises a similar kind of bilevel optimisation strategy, where the outer problem optimises for a single specific production flux of an organism. However, the OptKnock method does not include any experimental flux data in the input data (Burgard *et al.*, 2003).

4.2 Transformation into a single-level MILP

Solving the bilevel optimisation problem directly is time consuming due to the numerous possible reaction combinations to be included or excluded from the model (Herrgård *et al.*, 2006). Instead the overall bilevel optimisation problem can be formulated as a single mixed-integer linear program (MILP) as in the OptKnock method (Burgard *et al.*, 2003). According to the linear programming duality theory there exist a unique dual optimisation problem for each primal linear programming problem and the optimal objective values of the primal and dual problems are equal. Thus, the objectives of the primal and dual problems can be set equal to one another, the respective constraints of the problems accumulated and the bilevel optimisation problem formulation becomes transformed into a single MILP (Burgard *et al.*, 2003). In the transformation of the OMNI optimisation problem into a single MILP the inner problem is converted into a set constraint and thus, the OMNI method will always evaluate the distance between the experimental flux distribution and the one of the predicted alternative flux distributions that is the closest to the experimental flux distribution (Herrgård *et al.*, 2006).

To convert the OMNI optimisation problem in Equation 5 into a single large MILP the objective function in Equation 7 can be expressed in terms of only linear terms and constraints as (Herrgård *et al.*, 2006):

$$\sum_{i \in M} \omega_i \left(\Delta_i^+ + \Delta_i^- \right)$$

$$\text{where} \quad \Delta_i^+ \geq v_i^{opt} - v_i^{\exp} \qquad\qquad (8)$$

$$\Delta_i^- \geq v_i^{\exp} - v_i^{opt}$$

$$\Delta_i^+, \Delta_i^- \geq 0$$

A constraint for searching unique model structures that give equally good predictions, the equal value of the objective function, can be formulated as (Herrgård *et al.*, 2006):

$$\sum_{k \in R_n} y_k > 0 \ \ n = 1...N \qquad\qquad (9)$$

N is the number of previously obtained active reaction set solutions.

In most cases the OMNI MILP can be solved to optimality using standard solvers run on a single workstation within few hours. If the upper and lower bounds for the fluxes can be set tighter the computational time is reduced. For many reactions the bounds can be set to zero depending on the growth conditions without any loss of information (Herrgård *et al.*, 2006).

5 Identification of flux bottlenecks in E. coli knock-out strains

For a fraction of knock-out strains FBA significantly over predicted the growth rates at the end point of evolution for 45-50 d. The *E. coli* iJR904 genome-scale metabolic model overpredicted the growth rates of two independently evolved endpoint strains of each triose phosphate isomerase (*tpi*), phosphoenolpyruvate carboxylase (*ppc*) and phosphoglucose isomerase (*pgi*) deletions strains by an average of 22 % compared to the data from evolution experiments. The deletion strains had been characterised with metabolic flux and gene expression profiling experiments. The OMNI method was applied to identify the causes of the discrepancies between the model predictions and the experimental data. Since the discrepancies were over predictions of the growth rates, except for *pgi*E2 that was thus not included in the analysis, it was assumed that for some reasons some of the reactions in the model could not operate optimally in the experiments and thus the OMNI method could improve the model predictions. The corresponding *in silico* reactions were deleted from the genome-scale model. The set of measured target fluxes in the objective function of the OMNI was the growth rate

and 23 intracellular fluxes that had been determined from the strains. One to four reaction deletions were allowed for the OMNI. The model predictions did not improve significantly with over four reaction deletions (Herrgård *et al.*, 2006).

The results of the study are shown in Table 1 from Herrgård *et al.*, 2006. For all the strains included in the analysis the OMNI was able to identify a model structure that significantly improved the model predictions. Increasing the number of deleted reaction from the model up to four further improved the model predictions. Significant improvement of the model predictions required deletion of at least two reactions and thus the bilevel-optimisation approach of the OMNI method is highly efficient compared to full enumeration of all possible reaction deletions (Herrgård *et al.*, 2006).

Table 1. The reaction bottlenecks identified by the OMNI method for evolved *E. coli* knockout strains. The value of the OMNI objective function for the corresponding number of reaction deletions is shown. Also the corresponding errors in the growth rate prediction are given (Herrgård *et al.*, 2006).

Reaction	Number of Reactions	OMNI Objective	Growth Rate Error (%)	Expression Score 1	Expression Score 2	Reaction	Description	Subsystem
pgiE1	0	12.1	32.6	—	—	—	—	—
	2	11.5	27.5	0.2	99.0	MTHFC	Methenyltetrahydrofolate cyclohydrolase	Folate metabolism
						NADH6	NADH dehydrogenase	Oxidative phosphorylation
	3	10.4	20.7	0.4	1.3	DRPA	Deoxyribose-phosphate aldolase	Alternate carbon metabolism
						NADH6	NADH dehydrogenase	Oxidative phosphorylation
						FRD3	Fumarate reductase	Citrate cycle (TCA)
ppcE1	0	60.1	24.4	—	—	—	—	—
	1	51.7	23.0	14.0	13.7	TKT2	Transketolase	Pentose phosphate cycle
	2	46.2	22.5	99.0	99.0	AKGDH	2-oxogluterate dehydrogenase	Citrate cycle (TCA)
						TKT2	Transketolase	Pentose phosphate cycle
	3	34.7	13.6	3.0	4.7	AKGDH	2-oxogluterate dehydrogenase	Citrate cycle (TCA)
						NADH6	NADH dehydrogenase	Oxidative phosphorylation
						FRD3	Fumarate reductase	Citrate cycle (TCA)
ppcE2	0	49.1	17.0	—	—	—	—	—
	1	40.3	15.8	99.0	13.4	TKT2	Transketolase	Pentose phosphate cycle
	2	35.2	15.2	1.9	99.0	AKGDH	2-oxogluterate dehydrogenase	Citrate cycle (TCA)
						TKT2	Transketolase	Pentose phosphate cycle
	3	28.8	7.2	9.8	13.5	NADH6	NADH dehydrogenase	Oxidative phosphorylation
						NADH8	NADH dehydrogenase	Oxidative phosphorylation
						SUCOAS	Succinyl-CoA synthetase (ADP-forming)	Citrate cycle (TCA)
	4	26.7	6.9	4.7	2.8	FDH2	Formate dehydrogenase	Oxidative phosphorylation
						AKGDH	2-oxogluterate dehydrogenase	Citrate cycle (TCA)
						NADH6	NADH dehydrogenase	Oxidative phosphorylation
						NADH8	NADH dehydrogenase	Oxidative phosphorylation
tpiE1	0	33.6	29.0	—	—	—	—	—
	1	25.5	21.9	1.2	13.0	PGL	6-phosphogluconolactonase	Pentose phosphate cycle
	2	9.8	20.2	1.0	13.2	EDD	6-phosphogluconate dehydratase	Pentose phosphate cycle
						PGL	6-phosphogluconolactonase	Pentose phosphate cycle
tpiE2	0	33.1	25.3	—	—	—	—	—
	1	24.3	18.5	13.8	13.0	PGL	6-phosphogluconolactonase	Pentose phosphate cycle
	2	10.7	16.8	99.0	13.4	EDD	6-phosphogluconate dehydratase	Pentose phosphate cycle
						PGL	6-phosphogluconolactonase	Pentose phosphate cycle
	3	10.4	12.7	1.9	13.0	CYTBO3	Cytochrome oxidase bo3	Oxidative phosphorylation
						EDA	2-dehydro-3-deoxy-phosphogluconate aldolase	Pentose phosphate cycle
						PGL	6-phosphogluconolactonase	Pentose phosphate cycle
	4	9.9	16.2	99.0	99.0	EDD	6-phosphogluconate dehydratase	Pentose phosphate cycle
						PGL	6-phosphogluconolactonase	Pentose phosphate cycle
						GALU	UTP-glucose-1-phosphate uridylyltransferase	Alternate carbon metabolism
						PPM2	Phosphopentomutase 2 (deoxyribose)	Alternate carbon metabolism

The correspondence of the bottleneck reactions identified by the OMNI method and the gene expression data from the strains compared to the *E. coli* wild-type strain was also studied. In many cases the genes corresponding to the bottleneck reactions were downregulated in the evolved strain compared to the wild-type. However, the gene expression changes were quite similar in all the strains even though the OMNI method identified distinctive bottleneck reaction sets for the strains (Herrgård *et al.*, 2006).

Based on the bottleneck reaction identification by the OMNI method, regulatory constraints limiting the optimal performance could be estimated. The bottleneck reactions in the pentose-phosphate pathway were identified for *tpi*E1 and *tpi*E2 strains and downregulation of the corresponding genes compared to the wild-type strain was observed. The reason for bottlenecks in the pentose-phosphate pathway is not clear but it could be argued that the reason is the aim to reduce the production of NADPH by the pentose-phosphate pathway because the reduced consumption of NADPH by the reactions linked to the deleted *tpi* gene reaction (Herrgård *et al.*, 2006).

6 Conclusions

The most prominent application of the OMNI method is the identification of the correct active set of reactions in evolved strains showing slower than optimal growth rate based on the model predictions. An example of this application was given in the article by Herrgård *et al.*, (2006) where evolved *E. coli* strains were studied. However, the suggested gap-filling application (Herrgård *et al.*, 2006) can be expected to be computationally extremely time consuming if the set of potential reactions to be included in the model cannot be held small. The other suggested application in evaluating functions of poorly annotated enzymes and identifying the correct alternative reaction mechanisms (Herrgård *et al.*, 2006) are not straightforward applications and for solving these problems there are possibly more efficient methods. Alternative reaction mechanisms could only be studied in the level of reaction stoichiometry with the OMNI method. The enzyme function annotation problem is hard to solve by searching for optimal biochemical conversions, reactions, for the model since even though an optimal set of biochemical conversions is found it does not necessarily couple enzymes to the biochemical conversions.

The OMNI method uses metabolic flux distribution data as input data (Herrgård *et al.*, 2006). If only extracellular rates as the growth rate and uptake and production rates are included in comparisons of the experimental flux data and the model predictions, the distribution of intracellular fluxes may have many alternative solutions that are equally good and thus even different model structures could give show equally good predictions. On the other hand obtaining reliable information on the intracellular flux distribution to be used as input data would require isotopic labelling experiments and that is laborious and expensive to be carried out for a larger set of for example knock-out strains.

The objective function for the FBA problem is in a central role in the OMNI method as in all FBA applications if biologically meaningful results are desired. Maximising growth is more suitable objective for simple bacteria than for eukaryotic micro-organisms. The stoichiometry of an artificial flux for biomass synthesis should be set up carefully.

In the formulation of the bilevel-optimisation problem the number of allowed deletions was given exactly without further explanation (Herrgård *et al.*, 2006). If it was replaced with an inequality, would the computation become too time-consuming to be practical?

7 References

Bonarius, H., Schmid, G. and Tramper, J., Flux analysis of underdetermined metabolic networks: the quest for the missing constraints, *TIBTECH*, **15** (1997) 308-314.

Burgard, A., Pharkya, P. and Maranas, C., OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization, *Biotech. Bioeng.*, **84** (2003) 647-657.

Duarte, N., Herrgård, M. and Palsson B., Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model, *Genome Res.*, **14** (2004) 1298-1309.

Edwards, J., Ibarra, R. and Palsson, B., *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, *Nat. Biotechnol.*, **19** (2001) 125-130.

Famili, I., Förster, J., Nielsen, J. and Palsson, B., *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network, *Proc. Natl. Acad. Sci.*, **100** (2003) 13134-13139.

Fischer, E. and Sauer, U., Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism, *Nat. Genetics*, **37** (2005) 636-640.

Förster, J., Famili, I., Fu, P., Palsson, B. and Nielsen, J., Genome-scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network, *Genome Res.*, **13** (2003) 244-253.

Herrgård, M., Fong, S., and Palsson, B., Identification of Genome-Scale Metabolic Network Models Using Experimentally Measured Flux Profiles, *Comp. Biol.*, **2** (2006) 676-686.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M., The KEGG resource for depichering the genome, *Nucleic Acid res.*, **32** (2004) D277-D280.

Klamt, S., Schuster, S., and Gilles, E., Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria, *Biotechnol. Bioeng.*, **77** (2002) 734-751.

Klamt, S. and Stelling, J., Stoichiometric and Constraint-based Modeling, In *System Modeling in Cellular Biology*, Edited by Szallasi, Z., Stelling, J. and Periwal, V., The MIT Press, Cambridge Massachusetts 2006, p. 73- 96.

Lange, H. and Heijnen, J., Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*, *Biotechnol. Bioeng.*, **75** (2001) 334-344.

Price, N., Papin, J., Schilling, C. and Palsson, B., Genome-scale microbial in silico models: the constraints-based approach, *Trends Biotechnol.*, **21** (2003) 162-169.

Stephanopoulos, G., Aristidou, A. and Nielsen, J., *Metabolic Engineering: Principles and Methodologies*, Academic Press, London 1998, 725p.