IFAC

Structural Sensitivity Analysis of Metabolic Networks[*]

Markus Uhr[*] and Jörg Stelling[*]

[*] *Institute of Computational Science and Swiss Institute of Bioinformatics, ETH Zurich, 8092 Zurich, Switzerland (Tel: +41 44 632 7435; e-mail: joerg.stelling@inf.ethz.ch).*

Abstract: The knowledge about control and dynamics of biological systems is often limited or incomplete but, especially for cell metabolism, network structures are often well-characterized. A major challenge in systems biology, therefore, concerns the reverse-engineering of cellular control structures from the available knowledge on the controlled systems' structures. Here, we propose a method to analyze the sensitivities in metabolic reaction networks that makes use only of the stoichiometry of the metabolic network and the assumption that the biological system is (and remains) in steady state. It is based on least-squares analysis of reaction flux adjustments to disturbances, for which we present an analytic solution. The method can be extended to include multivariate disturbances and, if a reference flux distribution as an operating point is available, to compute relative sensitivities. The resulting sensitivities are instrumental in predicting the regulation of the affected reactions. We demonstrate the utility of the method with the example of a medium-size network model for *E. coli* metabolism. In particular, we focus on the relation of structural network sensitivities to the variance of gene expression data resulting from external perturbations and from the action of cellular control circuits.

Keywords: Modelling and identification.

1. INTRODUCTION

Formal sensitivity analysis is a well-established tool for analyzing the influence of perturbations and of control structures on systems behavior. For the area of systems biology, the former type of analysis is mainly concerned with characterizing the overall robustness of biological circuits. The latter can be employed to investigate design principles underlying complex networks in biology (Stelling et al., 2004). These approaches in general require dynamic, for instance, ODE-based systems models. However, establishing realistic and predictive mathematical models for biological circuits is currently a major bottleneck for the field. For most biological systems of interest, our knowledge of control structures and associated kinetic parameters is insufficient for mechanistic modeling. Correspondingly, identification of nonlinear biochemical systems is a challenge (Feng and Rabitz, 2004).

One possible alternative to analyzing the detailed systems dynamics is to focus on the set of *possible* behaviors that is consistent with structural constraints, in particular, reaction stoichiometries and reversibilities. This type of 'constraints-based' approaches has been particularly successful for analyzing metabolic networks. For one reason, reaction stoichiometries are well-characterized for metabolic networks such that genome-scale models could be established for several organisms (Borodina and Nielsen, 2005). Also, metabolism operates on faster time-scales than cellular processes such as genetic control systems, which allows for focusing the analysis on the steady-state behavior (Price et al., 2004). An approach termed 'flux-balance analysis' (FBA) allows computing flux distributions in metabolic networks that are optimal in terms of a given objective function by solving a linear optimization problem (Varma and Palsson, 1994). This method has, for instance, been used to predict mutant phenotypes, outcomes of evolution, and to analyze structural couplings in large networks (Price et al., 2004). Metabolic pathway analysis, which is based on decomposition of large networks into smaller functional units such as elementary flux modes (EFMs), can be employed for the same purposes. In contrast to FBA, it characterizes the complete solution space of a stoichiometric network model; as a drawback, the determination of pathways is a computationally hard problem and algorithms for genome-scale networks are not available (Schuster et al., 2007).

While methods for the analysis of metabolic networks—the controlled system—exist, it is largely unclear how the available biological knowledge could be employed to understand the corresponding control structures. In particular, the genetic control at slow time–scales that establishes different network operation modes is of interest. For instance, by comparison of computational predictions derived from metabolic network structures alone with experimental data on gene expression, the existence of correlations between metabolic fluxes and genetic control has been demonstrated (Stelling et al., 2002; Bilu et al., 2006). An apparent path to integrated network representations is the design of hybrid models that represent genetic control in an abstracted form such as Boolean logic models. Such

10.3182/20080706-5-KR-1001.1031

models allow for more accurate predictions (Barrett et al., 2005), but they are of limited use in reverse-engineering of control circuits.

Reverse-engineering of the associated controllers appears feasible because the effective dimension of the control problem may be much smaller than suggested by the complexity of metabolic networks (Barrett et al., 2006). Several approaches to identify closed–loop reactions to perturbations, or to pinpoint control mechanisms from the structure of the controlled network have been proposed. They rely on assumptions on optimal rejection of perturbations in the sense of a minimal deviation from a given operating point, using continuous ('minimization of metabolic adjustment', MOMA; Segre et al. (2002)) or discrete ('regulatory on/off minimization', ROOM; Shlomi et al. (2005)) distance metrics to quantify the deviation between original and perturbed state. Alternatively, 'structural kinetic modeling' makes use of randomly parametrized Jacobians for a given, generic systems model for the same purpose (Steuer et al., 2006).

Despite these advances, an equivalent to formal sensitivity analysis for dynamic systems is lacking. Here, we propose an approach to sensitivity analysis that uses the network structure only and is based on an analytical solution of a least-squares optimization problem (section 2). We apply our method to predict control points in the metabolic network of the bacterium *Escherichia coli* as a model system and validate the predictions with published experimental data on metabolic gene expression for a wide range of experimental conditions and perturbations (section 3).

2. NETWORK SENSITIVITY

2.1 Metabolic Network Fundamentals

Nutrient molecules are taken up by the cell and converted by chemical reactions to intermediary metabolites and finally to biomass to enable life. One possibility to model the biochemical processes in a bacterium is to look at the stoichiometry of the chemical reactions occurring in that organism. The interconversion of the different molecules can be modeled as a reaction network. The nodes of this network represent the metabolites that are part of cell metabolism and the edges represent the chemical reactions.

Mathematically, the network structure can be represented as a matrix $\mathbf{N}$, called the *stoichiometric matrix*. The metabolites in a reaction network form the rows of $\mathbf{N}$ and the reactions build the columns of the matrix. In a *column* of the stoichiometric matrix, a negative entry denotes a substrate of the reaction and a positive entry denotes a reaction product. Zero entries indicate that a metabolite is not affected by a reaction.

The stoichiometric matrix is a systems invariant that relates the rates $\mathbf{v}$ of the reactions (*fluxes*) to the concentrations $\mathbf{c}$ of the metabolites. If a metabolic reaction network has m metabolites and n reactions (usually, $m < n$), then $\mathbf{N} \in \mathbb{R}^{m \times n}$ and

$$\frac{d\mathbf{c}}{dt} = \mathbf{N}\mathbf{v}. \tag{1}$$

Equation 1 is an ordinary differential equation and states that the changes of metabolite concentrations are a linear function of the reaction rates.

The assumption that the system is in steady state can be represented as

$$\mathbf{N}\mathbf{v} = 0, \tag{2}$$

meaning that the fluxes in the metabolic network are such that metabolite concentrations do not change.

2.2 Structural Network Sensitivity

Based on the steady-state assumption, we aim at identifying chemical reactions that are highly sensitive to disturbances in reaction fluxes. The only information that is used to compute the sensitivities is the stoichiometry of the metabolic network.

Assume that a reaction k is subjected to a disturbance δ_k. If none of the undisturbed fluxes changes, the resulting flux is (in general) not in steady state any more. It is thus reasonable to assume that the undisturbed fluxes will adjust, such that, given the disturbed flux $v_k + \delta_k$, the overall flux distribution is in steady state again (Segre et al., 2002; Shlomi et al., 2005).

The sensitivity of a reaction i to a disturbance δ_k in reaction k is defined as the minimal adjustment d_i^* needed to bring the overall network flux in steady state again. Let $\mathbf{v}$ denote a flux distribution in steady state. If the disturbance δ_k and the adjustments d_i are written as a vector $\mathbf{d}$, the condition for the new, disturbed flux is

$$\mathbf{N}(\mathbf{v} + \mathbf{d}) = 0 \tag{3}$$

There exist infinitely many possible adjustments d_i to bring the overall flux $\mathbf{v} + \mathbf{d}$ in steady state. We look for minimal adjustments d_i^* in the sense that their sum of squares is minimized.

The same ideas hold if multiple simultaneous disturbances in several reactions are taken into account. In the following, subscripts $[k]$ denote the set of reactions that are disturbed simultaneously and subscripts $[i]$ refer to undisturbed reactions that can freely adjust (*independent* reactions). For example, the vectorial disturbances are written as $\boldsymbol{\delta}_{[k]}$.

2.3 2-norm Minimization

To find the minimal adjustments d_i^* of the independent reactions, the following minimization problem must be solved

$$\begin{aligned} \min \quad & \|\mathbf{d}\|_2 \\ \text{s.t.} \quad & \mathbf{N}\mathbf{d} = 0 \\ & \mathbf{d}_{[k]} = \boldsymbol{\delta}_{[k]} \end{aligned}. \tag{4}$$

Problem (4) has an analytical solution. To derive this solution, consider the following partition of the stoichiometric matrix

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_{[i]} & \mathbf{N}_{[k]} \end{bmatrix}$$

into columns of independent reactions (indexed $[i]$) and disturbed reactions (indexed $[k]$). The partition of the flux vectors $\mathbf{v}$ and $\mathbf{d}$ is done analogously into independent and disturbed fluxes

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_{[i]} \\ \mathbf{v}_{[k]} \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} \mathbf{d}_{[i]} \\ \mathbf{d}_{[k]} \end{bmatrix}.$$

The solution to problem (4) can be found by reformulating the problem as a least squares problem

$$\min_{\mathbf{d}_{[i]}} \quad \left\| \mathbf{N}_{[i]}\mathbf{d}_{[i]} = -\mathbf{N}_{[k]}\boldsymbol{\delta}_{[k]} \right\|_2 \tag{5}$$

and by solving it through Singular Value Decomposition (SVD) of the matrix $\mathbf{N}_{[i]}$

$$\mathbf{N}_{[i]} \triangleq \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^{\mathsf{T}} = \begin{bmatrix} \mathbf{U}_{[r]} & \mathbf{U}_{[n]} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_{[r]}^{\mathsf{T}} \\ \mathbf{V}_{[n]}^{\mathsf{T}} \end{bmatrix}.$$

The matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and the matrix $\boldsymbol{\Sigma}$ is diagonal with the so-called singular values on its diagonal. As in general the stoichiometric matrix $\mathbf{N}_{[i]}$ does not have full rank, some singular values are zero. The diagonal matrix $\boldsymbol{\Sigma}$ can thus be written as an $r \times r$ matrix $\boldsymbol{\Sigma}_{11}$ which has a non-zero diagonal (r being the rank of $\mathbf{N}_{[i]}$) and a zero rest. The splitting of $\mathbf{U}$ and $\mathbf{V}$ is also according to the rank r. Thus, the columns of $\mathbf{U}_{[r]}$ form an orthogonal basis of the range of $\mathbf{N}_{[i]}$ and the columns of $\mathbf{V}_{[n]}$ form an orthogonal basis of the null space of $\mathbf{N}_{[i]}$.

Then, the least squares adjustments $\mathbf{d}_{[i]}^*$ are given as

$$\mathbf{d}_{[i]}^* = -\mathbf{V}_{[r]}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{U}_{[r]}^{\mathsf{T}}\mathbf{N}_{[k]}\boldsymbol{\delta}_{[k]}. \tag{6}$$

The adjustments $\mathbf{d}_{[i]}^*$ are a linear function of the disturbances $\boldsymbol{\delta}_{[k]}$. It is thus reasonable to define the matrix of this linear function as the sensitivity to disturbances in reactions $[k]$, i.e.

$$\mathbf{S}_{[i],[k]} \triangleq -\mathbf{V}_{[r]}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{U}_{[r]}^{\mathsf{T}}\mathbf{N}_{[k]}. \tag{7}$$

An element (i,k) of the matrix $\mathbf{S}_{[i],[k]}$ can be interpreted as the ratio of the i'th adjustment to the k'th disturbance:

$$s_{i,k} = \frac{d_i^*}{\delta_k}$$

It is important to see that the computation of the sensitivity matrix $\mathbf{S}_{[i],[k]}$ is based on the solution of problem (5) which is a least squares problem that always yields a solution. However, it is always possible to choose a set of disturbed reactions $[k]$ such that for certain disturbances $\boldsymbol{\delta}_{[k]}$ there exists no adjustment vector $\mathbf{d}_{[i]}$ that fulfills $\mathbf{N}\mathbf{d} = 0$. To check whether the system can indeed adjust to arbitrary disturbances $\boldsymbol{\delta}_{[k]}$, one computes the residual

$$\rho = \|\mathbf{N}_{[i]}\mathbf{d}_{[i]} + \mathbf{N}_{[k]}\boldsymbol{\delta}_{[k]}\|.$$

Based on the previously computed SVD of $\mathbf{N}_{[i]}$, the residual is

$$\rho = \| -\mathbf{U}_{[n]}^{\mathsf{T}}\mathbf{N}_{[k]}\boldsymbol{\delta}_{[k]}\|. \tag{8}$$

For the residual to be zero for disturbances $\boldsymbol{\delta}_{[k]}$ in all directions, $\mathbf{N}_{[k]}$ must be orthogonal to all columns of $\mathbf{U}_{[n]}^{\mathsf{T}}$

$$\mathbf{U}_{[n]}^{\mathsf{T}}\mathbf{N}_{[k]} = \mathbf{0}. \tag{9}$$

The testing equation (9) can be interpreted as follows. All vectors that can be formed by $\mathbf{N}_{[k]}\boldsymbol{\delta}_{[k]}$ (denoted as range($\mathbf{N}_{[k]}$)) must also be formed by the adjustments $\mathbf{N}_{[i]}\mathbf{d}_{[i]}$ (denoted as range($\mathbf{N}_{[i]}$)), see Eq. (5). What we want to know is whether range($\mathbf{N}_{[k]}$) $\subseteq$ range($\mathbf{N}_{[i]}$). This is equivalent to testing whether the orthogonal complement (denoted as $^{\perp}$) of range($\mathbf{N}_{[i]}$) intersected with range($\mathbf{N}_{[k]}$) is the null element

$$\text{range}(\mathbf{N}_{[i]})^{\perp} \cap \text{range}(\mathbf{N}_{[k]}) = \mathbf{0}.$$

A possible way to circumvent these problems would be to only allow vectors $\boldsymbol{\delta}_{[k]}$ that can be balanced by adjust-

ments $\mathbf{d}_{[i]}$ even if range($\mathbf{N}_{[k]}$) $\nsubseteq$ range($\mathbf{N}_{[i]}$). Disturbances with this property can be written as

$$\boldsymbol{\delta}_{[k]} = \mathbf{V}_{[r]}\boldsymbol{\beta}$$

for any $\boldsymbol{\beta}$.

This approach does not take into consideration the reversibilities of the reactions. Thus, if a flux $\mathbf{v}$ is close to the boundary of feasible fluxes, the least squares adjustments might lead to fluxes that are chemically impossible for a finite disturbance δ_k. If, however, the sensitivity is considered as adjustments to infinitesimal disturbances, the definition of the sensitivity in Eq. (7) holds except for those points lying *exactly* on the boundary.

2.4 Relative Sensitivities

The sensitivity matrix (7) defines absolute sensitivities in that it gives a measure for the magnitude of the absolute adjustments $\mathbf{d}_{[i]}$. If a reference flux $\mathbf{v}^{\text{ref}}$ (e. g. an operating point) is known, one can compute relative adjustments $\mathbf{r}_{[i]}$:

$$r_i \triangleq \frac{d_i}{v_i^{\text{ref}}}. \tag{10}$$

Using matrix notation, we can write the relative adjustments as

$$\mathbf{r}_{[i]} = \text{diag}(\mathbf{v}_{[i]}^{\text{ref}})^{-1}\mathbf{d}_{[i]} \tag{11}$$

and a relation between relative adjustments $\mathbf{r}_{[i]}$ and relative disturbances $\boldsymbol{\rho}_{[k]}$ (which are defined analogously to Eq. (10)) can be derived

$$\begin{aligned} \mathbf{r}_{[i]} &= \text{diag}(\mathbf{v}_{[i]}^{\text{ref}})^{-1}\mathbf{d}_{[i]} \\ &= \text{diag}(\mathbf{v}_{[i]}^{\text{ref}})^{-1}\mathbf{S}_{[i],[k]}\boldsymbol{\delta}_{[k]} \\ &= \text{diag}(\mathbf{v}_{[i]}^{\text{ref}})^{-1}\mathbf{S}_{[i],[k]}\text{diag}(\mathbf{v}_{[k]}^{\text{ref}})\boldsymbol{\rho}_{[k]}. \end{aligned} \tag{12}$$

With this definition, it is straightforward to compute the relative sensitivities $\mathbf{S}_{[i],[k]}^{\text{rel}}$:

$$\mathbf{S}_{[i],[k]}^{\text{rel}} \triangleq \text{diag}(\mathbf{v}_{[i]}^{\text{ref}})^{-1}\mathbf{S}_{[i],[k]}\text{diag}(\mathbf{v}_{[k]}^{\text{ref}}). \tag{13}$$

Note that in order to compute the relative sensitivities, a complete reference flux distribution $\mathbf{v}^{\text{ref}}$ must be given. In practice, it is often difficult to get experimental data of a complete flux distribution. As a possible resort, flux distributions obtained from *flux balance analysis* (FBA) (Varma and Palsson, 1994) might be used. These are approximations of the *in vivo* fluxes and computational predictions depend on the objective functions employed in FBA (Schuetz et al., 2007). Despite these limitations, the structural sensitivity framework allows for a direct relation between controlled network structure and possible features of the corresponding controller.

3. ANALYSIS OF *E. COLI* METABOLISM

We start from the hypothesis that structural network sensitivities are related to variability in gene expression data because if a reaction is very sensitive it requires regulation. Thus, if there are disturbances in the network, sensitive reactions need stronger regulation than insensitive reactions, which should be reflected by high variability of gene expression in the sensitive reactions.

To analyze the metabolism of *E. coli* we use a medium scale model of central carbon metabolism (Stelling et al., 2002). Such models capture the basic properties of cell metabolism but have only reduced demands for computational power as compared to full genome-scale models. The model has 97 metabolites and 118 reactions (i.e. the stoichiometric matrix $\mathbf{N} \in \mathbb{R}^{97 \times 118}$). An interesting property of this network is that the stoichiometric matrix has full row rank, which means that there is no redundancy in the compact stoichiometric model.

3.1 Computation of Network Sensitivities

To be able to analyze the relative sensitivities, we first generate reference fluxes. To this end, we use the elementary flux modes (EFMs) of the *E. coli* metabolic network. Elementary flux modes represent independent minimal subsets of reactions that permit steady-state fluxes. More mathematically, they are the extreme rays of the so called "flux cone", which is the set of all feasible fluxes under the steady state assumption and irreversibility constraints. EFMs are unique up to a scaling factor and their linear combination allows for computing all feasible flux distributions, that is,

$$\mathbf{v} = \sum_j \xi_j \mathbf{e}_j \qquad (14)$$

where ξ_j is a non-negative scaling factor for the j-th EFM $\mathbf{e}_j$ (Schuster et al., 2007).

For the computation of reference fluxes, we consider all EFMs of the *E. coli* model with glucose uptake and biomass production. This simulates growth of *E. coli* on glucose as their only nutrient source. For the 97×118 model, there are roughly $27\,000$ EFMs with these properties (Stelling et al., 2002).

EFMs for generating a reference flux are selected according to their biomass yield. This is the ratio of biomass production flux to glucose uptake flux, which is a linear fractional function. Fluxes with optimal yield $y^{\max}$ are nonnegative linear combinations of the EFMs with highest yield value. The selection of EFMs employs a parameter α such that only EFMs e_j with yield $y_j \geq \alpha \cdot y^{\max}$ are used to generate a reference flux. The values for α and corresponding numbers of EFMs used here are compiled in Table 1. To find nonnegative linear combinations of the EFMs, we draw the coefficients randomly from a uniform distribution in $(0, 1)$.

Table 1. EFM selection for reconstruction of a flux operating point.

α	Number of EFMs
0.01	21 592
0.1	20 996
0.5	8841
0.9	1425
0.95	708
0.97	400
0.99	90
0.995	35
1	2

Given the reference fluxes, we are now able to compute the relative sensitivities. We compute the relative sensitivities for a univariate disturbance in every reaction of the
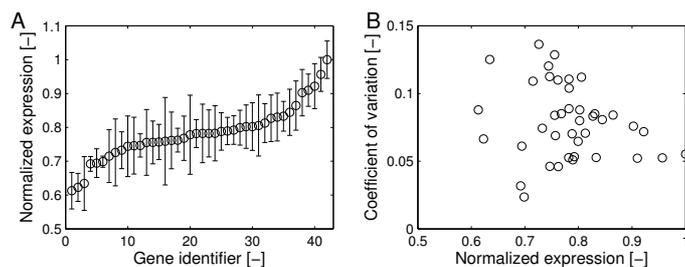
Fig. 1. Experimentally determined variability in metabolic gene expression for 42 genes in *E. coli* that are included in the stoichiometric model. (A) Average gene expression and standard deviation of expression. (B) Coefficient of variation as a function of the normalized expression level.

network and then take the average over all disturbances. Because the generation of reference fluxes depends on computational methods, they might have errors compared to wild-type fluxes. Therefore, we sample 20 different reference fluxes for each disturbance and take the average.

3.2 Comparison with Gene Expression Data

For the experimental data, we compiled a gene expression data set from the Many Microbe Microarrays Database (http://m3d.bu.edu) (Faith et al., 2007) that includes 507 experimental conditions corresponding to a representative set of random perturbations. We focus on those 42 metabolic genes that are included in the stoichiometric model. The experimental data show only small (< 2-fold) variability in average gene expression levels (Fig. 1A), but distinctions in the variances. However, as shown in Fig. 1B, variability is not related to average expression levels. We find no statistically significant correlation between the coefficient of variation (CV) and the average expression ($r = 0.03$ and $p = 0.85$ for Spearman's rank-order correlation), indicating that prediction of a flux distribution alone will not yield insight into control of the network.

Average structural network sensitivities, in contrast, can show a high correlation with the experimentally determined CV of gene expression, depending on the choice of α (Fig. 2). With increasing α and corresponding decrease of the share of EFMs used in generating the reference flux distributions, the Spearman's rank-ordered correlations r increase. The two scenarios with $r \approx 0.5$ (Fig. 2A) correspond to $\alpha > 0.99$, that is, reference flux distributions with close-to-optimal yield.

To assess the significance of these correlations, we computed the corresponding p-values that give the probability that no correlation exists. As shown in Fig. 2B, only the predictions for high α are statistically significant in contrast to, for instance, negative correlations for low α that would contradict our starting hypothesis. This is in accordance with previous findings (e.g., from FBA) that optimal biomass yield is an appropriate objective function to predict fluxes in metabolic networks of microbes (Price et al., 2004; Schuetz et al., 2007). Note also that the correlation coefficients for gene expression variability are higher than those obtained with a simple flux variability score (as defined by the difference between maximal and
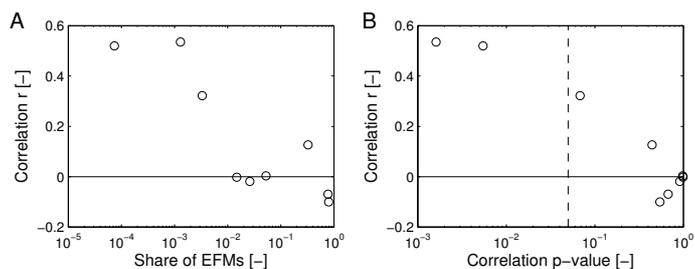
Fig. 2. Correlation between structural network sensitivities and experimentally observed variation of gene expression for metabolic genes. (A) Spearman's rank correlation coefficient r depending on the share of EFMs with highest biomass yield (number of α-optimal EFMs divided by the total number of EFMs) used for calculating the flux operating point. (B) Statistical significance of the correlations evaluated by the associated p-values; the dashed line indicates a significance level of $p = 0.05$.

minimal flux through each reaction) in a previous analysis of a yeast network ($r = 0.17$) (Bilu et al., 2006); higher p-values compared to that study result from a smaller sample size (number of metabolic genes) considered here.

For practical applications of the approach, a ranking of most sensitive—and, hence, most likely regulated—genes could be useful because this would allow a more targeted, detailed experimental analysis of potential key control points in metabolic networks. Next, we therefore investigated how well structural network sensitivities could identify those genes with most variable expression. More specifically, we rank-ordered the experimental data according to their CV, generated corresponding lists of rank-ordered sensitivities with different lengths, and evaluated the overlap between both.

The null model for statistical evaluation of prediction results follows a hypergeometric distribution. It captures the number of successes in L samples from a finite population of size N without replacement. Here, N is the total number of genes and L is the number of most variable genes to be predicted. The expected value for the ratio of correct to total predictions for the null model, φ_0, and the corresponding variance, σ_0^2, are given by

$$\varphi_0 = \frac{L}{N} \text{ and } \sigma_0^2 = \frac{(N-L)^2}{N^2(N-1)}. \qquad (15)$$

With this, we can calculate the 95% confidence intervals for the null model as $\varphi_0 \pm 1.96 \cdot \sigma_0$.

Simulation results for different values of α are shown in Fig. 3. The share of sensitive genes equals L/N and the prediction accuracy refers to the relative overlap between top–variable genes from the experimental data, and predicted genes with highest structural sensitivity. In agreement with the correlation studies above, we find that the flux operating point has a large influence if the predictions are significantly different from the null model, that is, a random selection of genes. In particular, when only EFMs with highest biomass yield are used for constructing the reference flux distributions, we find an $\approx$4-fold enrichment of sensitive genes in the predicted lists compared to the null model for the top 10-60% genes (Fig. 3D).
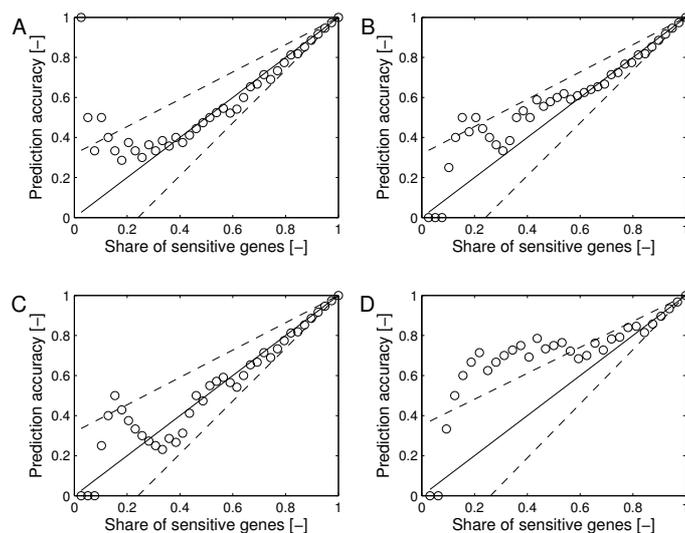
Fig. 3. Prediction of gene expression variability. Prediction accuracy is determined as the number of genes that show highest experimental variability (CV) *and* are identified as most sensitive by structural sensitivity analysis divided by the number of most variable genes considered for different shares of top-sensitive genes. The panels refer to different scenarios for constructing (random) flux operating points: (A) $\alpha = 0.01$, (B) $\alpha = 0.50$, (C) $\alpha = 0.90$, and (D) $\alpha = 0.995$. Circles refer to structural sensitivity predictions, while the solid line and the dashed lines denote predictions and 95% confidence intervals for the random model, respectively.

Note that these predictions are not perfect; for instance, only in the scenario with $\alpha = 0.01$ (Fig. 3A) the genes with highest expression variability are identified correctly. A possible reason for these mismatches between predictions and experimental observations is that the stoichiometric model used here covers only a small core of the metabolic network currently known in *E. coli*.

Figure 4 visualizes the correspondence between the reactions' sensitivities with the expression variability of their encoding genes in most of central carbon metabolism. The reactions are divided into four categories according to their sensitivity and gene variability, showing matches and mismatches of our method. Note that not all reactions could be uniquely assigend to encoding genes. These reactions were not considered in our analysis.

4. CONCLUSION

We have shown a procedure to identify highly sensitive reactions in metabolic networks. It makes use of network structure alone, which is data that is often available early in the investigation of biological organisms. The kinetics of the biochemical reactions are neglected (for the reasons mentioned in the Introduction). This implies that we had to make the assumption of least adjustments to disturbances (Segre et al., 2002) and the interpretation of the results was possible only in a statistical sense.

Because all computations rely on standard procedures from numerical linear algebra, the method is easy to
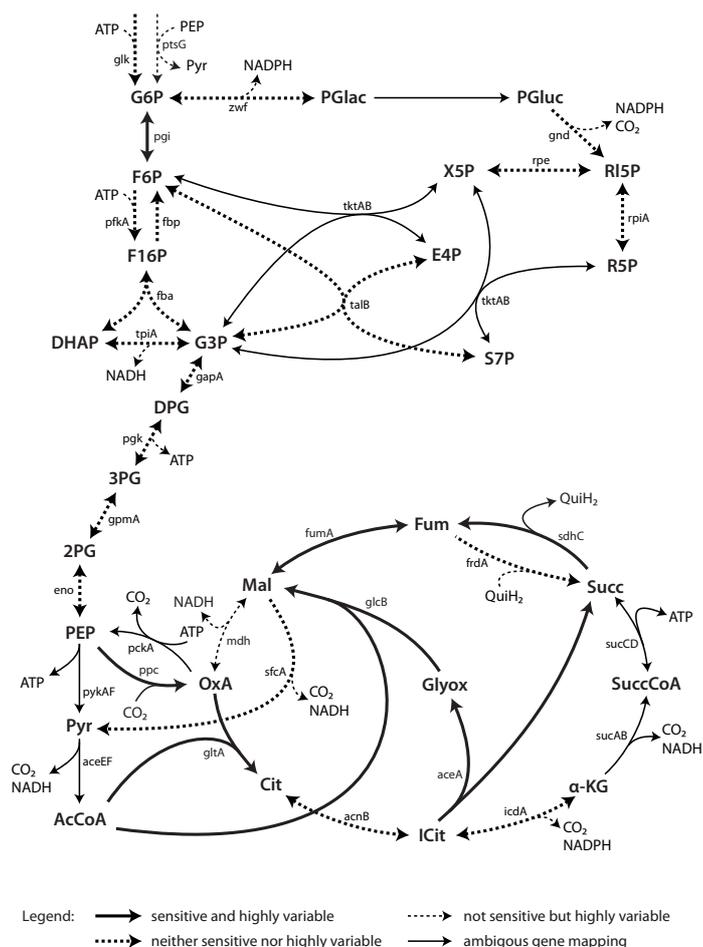
Fig. 4. Correspondence between structural sensitivity of reactions and gene variability of encoding genes for part of the CCM network of *E. coli* used in this analysis. The reactions are classified as highly sensitive and, correspondingly, gene expression is classified as highly variable if they are among the 14 most sensitive reactions or most variable genes. Three out of 14 reactions are found sensitive but their encoding genes are not highly variable (none shown) and another 3 of 14 are not sensitive but their gene expression is higly variable (2 shown).

implement and applicable to larger metabolic networks. Genome-scale models have been established for many model organisms such as *E. coli* (Reed et al., 2003) and could be employed in future work. In particular, calculating structural network sensitivities scales well with network size because the main computational burden lies in the computation of the SVD, which is an $O(n^3)$ process. However, for these larger networks, EFMs cannot be computed yet; for specifying the flux operating point, sampling methods such as those described in (Barrett et al., 2006) could be employed instead.

We demonstrate the usefulness of the method with a model of central carbon metabolism of *E. coli* and experimental data of metabolic gene expression. We were able to show significant correlation between the predicted sensitivities and the variation of gene expression even with such a limited model. Altogether, thus, validation of the structural sensitivity approach with a simple model as above, in combination with the method's scalability, could make this a useful tool for experimental design to uncover control structures that impinge on metabolic networks.

REFERENCES

C. L. Barrett, C. D. Herring, J. L. Reed, and B. O. Palsson. The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states. *Proc Natl Acad Sci U S A*, 102:19103–19108, 2005.

C.L. Barrett, N.D. Price, and B.O. Palsson. Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition. *BMC Bioinformatics*, 7:132, 2006.

Y. Bilu, T. Shlomi, N. Barkai, and E. Ruppin. Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. *PLoS Comput Biol*, 2:e106, 2006.

I. Borodina and J. Nielsen. From genomes to in silico cells via metabolic networks. *Curr Opin Biotechnol*, 16:350–355, 2005.

J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol*, 5:e8, 2007.

X. Feng and H. Rabitz. Optimal identification of biochemical reaction networks. *Biophys. J.*, 86:1270–81, 2004.

N. D. Price, J. L. Reed, and B. O. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2:886–897, 2004.

J.L. Reed, T.D. Vo, C.H. Schilling, and B.O. Palsson. An expanded genome-scale model of *Escherichia* coli K-12 (iJR904 GSM/GPR). *Genome Biol.*, 4:R54, 2003.

R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol Syst Biol*, 3:119, 2007.

S. Schuster, A. von Kamp, and M. Pachkov. Understanding the roadmap of metabolism by pathway analysis. *Methods Mol Biol*, 358:199–226, 2007.

D. Segre, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA*, 99:15112–15117, 2002.

T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*, 102:7695–7700, 2005.

J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420: 190–93, 2002.

J. Stelling, E.D. Gilles, and F.J. Doyle III. Robustness properties of circadian clock architectures. *Proc. Natl. Acad. Sci. U.S.A.*, 101:13210–15, 2004.

R. Steuer, T. Gross, J. Selbig, and B. Blasius. Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.*, 103:11868–73, 2006.

A. Varma and B. O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nat Biotech*, 12:994–998, 1994.