Network inference from time-resolved metabolomics data
Hendrickx, D.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Elucidating changes in the distribution of reaction rates in metabolic pathways under different conditions is a central challenge in systems biology. Here we present a method for inferring regulation mechanisms responsible for changes in the distribution of reaction rates across conditions from correlations in time-resolved data. A reversal of correlations between conditions reveals information about regulation mechanisms. With the use of a small in silico hypothetical network, based on only the topology and directionality of a known pathway, several regulation scenarios can be formulated. Confronting these scenarios with experi-

---

3 This chapter is based on Diana M. Hendrickx, Huub C.J. Hoefsloot, Margriet M.W.B. Hendriks, Daniël J. Vis, André B. Canelas, Bas Teusink and Age K. Smilde. Inferring differences in the distribution of reaction rates across conditions. Molecular Biosystems, Volume 8:9 (2012), pages 2415-2423.
Inferring differences in the distribution of reaction rates across conditions

mental data results in a short list of possible pathway regulation mechanisms associated with the reversal of correlations between conditions. This procedure allows for the formulation of regulation scenarios without detailed prior knowledge of kinetics and for the inference of reaction rate changes without rate information. The method was applied to experimental time-resolved metabolomics data from multiple short-term perturbation-response experiments in *S. cerevisiae* across aerobic and anaerobic conditions. The method’s output was validated against a detailed kinetic model of glycolysis in *S. cerevisiae*, which showed that the method can indeed infer the correct regulation scenario.

4.1 Introduction

In living organisms growth, reproduction, maintenance of homeostasis, and adaptation to the environment are enabled by chemical means [185]. The intermediates of metabolic reactions, metabolites, are organized in a large network, consisting of pathways. Elucidating how these pathways function is an important topic in systems biology, because it can contribute to different disciplines [71, 181], such as microbiology [50], plant biology [134, 205] and biomedical sciences [43].

One of the approaches to infer information about the functioning of cellular systems is correlation analysis of metabolomics data [21, 189, 192]. It has been shown that the interpretation of correlations in metabolomics is not straightforward [21, 189], mostly because there is no direct relation between the correlation coefficients and the underlying pathway of reactions [66, 192]. However, interpreting correlations can give more insight in the regulation of biochemical processes [205]. Previous research [21, 189, 192] focused on the origin of correlations in steady state (static) metabolomics data (biological replicates). Table 4.1 gives an overview of how to interpret correlations between metabolites in static data.

Besides correlations, several other approaches have been developed for the reverse engineering of metabolic networks [5, 37]. In a previous study [66], we examined the feasibility of inferring a *de novo* metabolic network from metabolomics data. The conclusion was that the mathe-
Inferring differences in the distribution of reaction rates across conditions

Table 4.1: Interpretation of correlations in steady state metabolomics data.

<table>
<thead>
<tr>
<th>observation</th>
<th>interpretation</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>high positive correlation between two</td>
<td>(1) dominance of one parameter (e.g. an enzyme</td>
<td>[21, 189]</td>
</tr>
<tr>
<td>metabolites in one condition</td>
<td>concentration)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) equilibrium</td>
<td>[21, 189]</td>
</tr>
<tr>
<td>negative correlation between two</td>
<td>(1) metabolites are not in equilibrium</td>
<td>[21]</td>
</tr>
<tr>
<td>metabolites in one condition</td>
<td>(2) metabolites belong to a moiety conservation</td>
<td>[21]</td>
</tr>
<tr>
<td>sustained correlation across multiple</td>
<td>(1) rapid equilibrium</td>
<td>[189]</td>
</tr>
<tr>
<td>conditions</td>
<td>(2) mass conservation</td>
<td>[189]</td>
</tr>
<tr>
<td>reversed correlation between two</td>
<td>(1) change in regulation</td>
<td>[189]</td>
</tr>
<tr>
<td>conditions</td>
<td>(2) existence of multiple steady states</td>
<td>[189]</td>
</tr>
</tbody>
</table>

Mathematical methods require sampling frequencies and noise levels that are not in accordance with contemporary laboratory experiments.

This chapter presents a study on the inference of changes in reaction rate distribution from correlations in time-resolved data measured under different conditions, applied to the central carbon metabolism of *Saccharomyces cerevisiae*. The results of the correlation analysis are combined with *a priori* information about the reaction scheme to infer possible regulation scenarios for the pathway. This is facilitated by using an in silico hypothetical network model.

Changing conditions are often associated with differences in the distribution of reaction rates [56, 62, 142, 154]. This research shows that correlations in time-resolved metabolomics data, combined with prior knowledge about the pathway, can help to obtain information about changes in the distribution of reaction rates between conditions. This means that correlations in metabolomics time series provide a new computational approach for deriving information about reaction rates, without performing reaction rate analysis. Reaction rates in a metabolic pathway are not straightforward to determine experimentally [39], so
Inferring differences in the distribution of reaction rates across conditions

Table 4.2: Description of the experiments.

<table>
<thead>
<tr>
<th>steady-state condition</th>
<th>perturbation</th>
<th>number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>aerobic</td>
<td>10 mM glucose</td>
<td>4</td>
</tr>
<tr>
<td>aerobic</td>
<td>2.5 mM glucose</td>
<td>3</td>
</tr>
<tr>
<td>aerobic</td>
<td>2.3 mM glucose</td>
<td>1</td>
</tr>
<tr>
<td>anaerobic</td>
<td>1 mM glucose</td>
<td>1</td>
</tr>
<tr>
<td>anaerobic</td>
<td>3 mM glucose</td>
<td>1</td>
</tr>
</tbody>
</table>

the development of computational methods to infer information about reaction rates is very important. Very commonly used methods for the determination of reaction rates are based on measured external fluxes (e.g. metabolic flux analysis [237]). The approach in this chapter makes a link between intracellular metabolite concentrations and intracellular reaction rates by using correlations. This study examines which prior knowledge is needed and how to combine this knowledge with correlations in time-resolved data to detect differences in the distribution of reaction rates across conditions.

4.2 Materials and methods

4.2.1 Data set

The proposed approach was tested on time-resolved metabolomics data from *S. cerevisiae* cultivations. The cells were grown in aerobic (*D* = 0.1 *h*⁻¹) or anaerobic (*D* = 0.05 *h*⁻¹) glucose-limited continuous cultures, as described elsewhere [25, 67, 81, 140] and short-term perturbation-response experiments were performed on the steady state, by introducing a sudden increase in the extracellular glucose concentration (called a glucose pulse). An overview is given in Table 4.2. Quantitative data on absolute levels of intracellular metabolites were obtained by LC-MS/MS, as described elsewhere [25, 81, 140]. The metabolites measured are from glycolysis (glucose-6-phosphate, fructose-6-phosphate, fructose-1,6-bisphosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate) and
Inferring differences in the distribution of reaction rates across conditions

the tricarboxylic acid cycle (pyruvate, citrate, oxoglutarate, succinate, fumarate, malate). Each time series consists of 11 time points. The data are plotted in Supplementary Figures 8-10.

4.2.2 Procedure to infer regulation scenarios

The procedure to infer regulation scenarios is depicted in Figure 4.1. Correlation analysis is performed on time-resolved data measured under different conditions (see section Correlation analysis). A priori information about the studied pathway is used to construct an in silico hypothetical network model (see section In silico hypothetical network model). Regulation scenarios for the pathway are inferred by combining the correlations with the hypothetical model.

![Diagram](image)

Figure 4.1: Procedure to infer regulation scenarios from time-resolved data and a priori information about the pathway.

4.2.2.1 Correlation analysis.

For each experiment, the Pearson product-moment correlation [151] between time profiles of metabolite concentrations, consisting of 11 time points is calculated. In the case of replicates, the correlation coefficients are averaged to obtain one statistic for each condition. Attention must be given to the skewed distribution of Pearson’s correlation coefficient, which makes the mean of the correlations not a good measure of central tendency. To calculate averages, the correlations were converted to scores that have a normal distribution, by applying the Fisher z-transform [54]. The mean z-score over replicates over a condition was
Inferring differences in the distribution of reaction rates across conditions

calculated and converted back to a Pearson correlation coefficient by applying the inverse of the Fisher z-transform.

4.2.2.2 *In silico* hypothetical network model.

Based on only the topology and the directionality of the pathway, we can construct a small hypothetical network as follows. The reaction scheme of the studied pathway is retrieved from MetaCyc (YeastCyc) [26]. No prior knowledge about the kinetic parameters is needed. If reactions are described as rapid equilibrium in the literature, substrate and product can be lumped together. Cofactors and branches of the pathway not occurring in the data set are not included in the hypothetical network. Reaction kinetics are formulated using first order mass-action kinetics with all parameters initially set to 1 for simplicity. Regulation scenarios are inferred with the following forward selection procedure:

- Step 1: Calculate the steady state before the pulse.

- Step 2: Perturb the system from steady state by increasing the inflow. If the qualitative behavior (increase or decrease of metabolite concentrations) after the pulse is in accordance with the experimental data, then it can be concluded that the distribution of the reaction rates does not change after the pulse. If not, then check in the literature if the pathway can have different behavior (e.g. cycle vs. two branches). If different behavior of the pathway is possible, go to Step 3a. If not, go to Step 3b.

- Step 3a: Simulate the following scenarios: 1) change of one rate constant (increase or decrease, for a network with $p$ reactions, there are $2p$ scenarios); 2) change of the behavior of the pathway according to literature. When one or more scenarios are in accordance with the observed correlation in the experimental data, then it can be concluded that these are the possible regulation scenarios after the pulse. If not, then select the scenarios most similar to the data (in qualitative behavior). Go to Step 4.

- Step 3b: Simulate the following scenario: change of one rate constant (increase or decrease, for a network with $p$ reactions, there
Inferring differences in the distribution of reaction rates across conditions

are $2p$ scenarios. When one or more scenarios are in accordance with the observed correlation in the experimental data, then it can be concluded that these are the possible regulation scenarios after the pulse. If not, then select the scenarios most similar to the data (in qualitative behavior). Go to Step 4.

- Step 4: Change one more parameter in the selected scenarios (per selected scenario, the total number of possibilities is twice the number of remaining parameters). When one or more scenarios are in accordance with the observed correlation in the experimental data, then it can be concluded that these are the possible regulation scenarios after the pulse. If not, then select the scenarios most similar to the data (in qualitative behavior). Repeat Step 4 until scenarios are derived that are in accordance with the experimental data.

The method described above infers the simplest possible regulation scenarios that are in accordance with the changes in correlation observed in the experimental data. The inferred regulation scenarios are based on mass balances and therefore independent of the order of the mass-action kinetics used in the hypothetical network model (see paragraph 3 of the Supplementary Material for an example and a general mathematical proof).

Matlab’s ordinary differential equation solver ode15s [127] was used for calculating concentration values over time and determining steady state concentrations.

### 4.3 Results

#### 4.3.1 Data set

Figures 4.2, 4.3, 4.4 and 4.5 summarize the results of the correlation analysis for glycolysis and the TCA cycle. The emphasis is put on a qualitative study of metabolite associations, not on the exact values of the correlation coefficient $r$. Therefore the correlations are divided into categories, based on literature [21, 123]: strong ($|r| \geq 0.8$), moderate ($0.6 \leq |r| < 0.8$), weak ($0.3 \leq |r| < 0.6$) and zero ($|r| < 0.3$). The sign
Inferring differences in the distribution of reaction rates across conditions

Figure 4.2: Correlations between measured glycolytic metabolites under aerobic conditions. Abbreviations: GLUC, glucose; ACALD, acetaldehyde; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate.

of the correlations is also reported.

The results show metabolite pairs that are positively correlated under all studied experimental conditions: G6P-F6P, 3PG-PEP and FUM-MAL. There were two pairs where no or only weak correlation was observed for both aerobic and anaerobic conditions: 3PG-PYR and PEP-PYR. For the metabolite pairs FBP-3PG, FBP-PEP, CIT-OGL, CIT-FUM and CIT-MAL, the response to a glucose pulse shows negative correlation under aerobic conditions and positive correlation under anaerobic conditions. Under aerobic conditions, positive correlation for the 10 mM glucose pulses and zero correlation for the 2.3-2.5 mM glucose
Inferring differences in the distribution of reaction rates across conditions

1 - 3 mM glucose anaerobic

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6P-F6P</td>
<td>strong positive (r &gt;= 0.8)</td>
</tr>
<tr>
<td>F6P-FBP</td>
<td>moderate positive (0.6 &lt;= r &lt; 0.8)</td>
</tr>
<tr>
<td>3PG-PEP</td>
<td>low positive (0.3 &lt;= r &lt; 0.6)</td>
</tr>
<tr>
<td>no connection</td>
<td>very low (zero) (-0.3 &lt; r &lt; 0.3)</td>
</tr>
<tr>
<td>MAL-FUM</td>
<td>low negative (-0.6 &lt; r &lt;= -0.3)</td>
</tr>
<tr>
<td>3PG-PEP</td>
<td>moderate negative (-0.8 &lt; r &lt;= -0.6)</td>
</tr>
<tr>
<td>strong negative (r &lt;= -0.8)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.3: Correlations between measured glycolytic metabolites under anaerobic conditions. Abbreviations: see Figure 4.2.

pulses is observed for SUC-MAL and SUC-OGL. Under anaerobic conditions, responses to glucose pulses are negatively correlated for these two metabolite pairs. The sustained correlations between the metabolite pairs G6P-F6P and MAL-FUM in the time-resolved data correspond to results from previous studies of correlations in steady-state data and point to rapid equilibrium reactions [62, 189]. From the preserved correlation of 3PG-PEP we can infer that both reactions from 3PG to 2PG and from 2PG to PEP, catalyzed by phosphoglyceromutase and enolase respectively, have to be rapid equilibrium reactions, which is in accordance with the literature [35]. Furthermore, in both steady state and time-resolved data, there is a relation between the displacement from equilibrium of a reaction and the strength of the correlation between substrate and product. The closer the reaction is to its equilibrium, the stronger the correlation between substrate and product [21].

The reversed correlation between aerobic and anaerobic conditions for some of the metabolite pairs in glycolysis and the TCA cycle were unexpected. A small hypothetical network model was used to determine the regulation mechanisms that could explain this behavior.
Inferring differences in the distribution of reaction rates across conditions

Figure 4.4: Correlations between measured metabolites of the TCA cycle under aerobic conditions. Abbreviations: PYR, pyruvate; CIT, citrate; OGL, oxoglutarate; SUC, succinate; FUM, fumarate; MAL, malate.
Inferring differences in the distribution of reaction rates across conditions

1-3 mM glucose anaerobic

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PYR</td>
<td>strong positive (r &gt;= 0.8)</td>
</tr>
<tr>
<td>CIT</td>
<td>moderate positive (0.6 &lt;= r &lt; 0.8)</td>
</tr>
<tr>
<td>SUC</td>
<td>low positive (0.3 &lt;= r &lt; 0.6)</td>
</tr>
<tr>
<td>FUM</td>
<td>no connection</td>
</tr>
<tr>
<td>MAL</td>
<td>very low (zero) (-0.3 &lt; r &lt; 0.3)</td>
</tr>
<tr>
<td>OGL</td>
<td>low negative (-0.6 &lt; r &lt;= -0.3)</td>
</tr>
<tr>
<td>1-3 mM glucose anaerobic</td>
<td>moderate negative (-0.8 &lt; r &lt;= -0.6)</td>
</tr>
<tr>
<td></td>
<td>strong negative (r &lt;= -0.8)</td>
</tr>
</tbody>
</table>

Figure 4.5: Correlations between measured metabolites of the TCA cycle under anaerobic conditions. Abbreviations: see Figure 4.4.

4.3.2 Hypothetical network model

4.3.2.1 Glycolysis.

To examine which information about the network can be inferred from the reversed correlation between FBP and 3PG, the hypothetical network model of Figure 4.6 can be used. This hypothetical network is obtained as follows. G6P and F6P are in rapid equilibrium (see results correlation analysis) and are lumped together. The equilibrium pool of GAP and DHAP [201] is presented by one node. The pair of reactions GAP → BPG and BPG → 3PG is lumped together because BPG is not measured. The part of glycolysis after 3PG is left out because this part of glycolysis is not relevant to derive regulation scenarios from the negative correlation between FBP and 3PG. FBP and 3PG correspond to B and D in the hypothetical network model. Under aerobic conditions, a negative correlation between FBP and 3PG in the first few minutes after the perturbation is observed in the data because FBP increases and 3PG decreases (See Supplementary Figures 9(a-d) and 10(a-d)). Based on mass balances, an increase in B and a decrease in D as an initial response to the pulse is only possible in the following situations:

- a smaller rate of the reaction converting B to C (a lower value for $k_2$);
Inferring differences in the distribution of reaction rates across conditions

- a larger rate of the reaction converting $C$ to $E$ in the branch (a higher value for $k_3$);
- a smaller rate of the reaction converting $C$ to $D$ (a lower value for $k_4$);
- a larger rate of the reaction that further metabolizes $D$ (a higher value for $k_6$);

Thus by reasoning, regulation scenarios corresponding to the reversal of correlations between conditions can be inferred. In the case of a more complex network, there is a need for a more automated way to infer regulation scenarios. This can be done by using first order mass-action kinetics for the hypothetical model and performing simulations. The mass balances for the metabolites are

\[
\begin{align*}
\frac{d[A]}{dt} &= \text{in} - k_1 \cdot [A] \\
\frac{d[B]}{dt} &= k_1 \cdot [A] - k_2 \cdot [B] \\
\frac{d[C]}{dt} &= k_2 \cdot [B] - (k_3 + k_4) \cdot [C] \\
\frac{d[D]}{dt} &= k_4 \cdot [C] - k_6 \cdot [D] \\
\frac{d[E]}{dt} &= k_3 \cdot [C] - k_5 \cdot [E]
\end{align*}
\]

with parameter values $\text{in} = 1$ and $k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = 1$. The steady state of this system is $[A] = [B] = 1$ and $[C] = [D] = [E] = 0.5$. A pulse, a sudden increase in the inflow, is simulated by changing the inflow to $\text{in} = 2$. When the pulse does not cause relative changes in reaction rates, all metabolite concentrations increase (see Figure 4.7). This is not in accordance with the behavior of the experimental data under aerobic conditions. Supplementary Table 1 shows the results of the simulations when one parameter is changed. There are four regulation scenarios leading to an increase in $B$ and a decrease in $D$ as an initial
Inferring differences in the distribution of reaction rates across conditions

response to the pulse (see Figure 4.8), which are the same as those found by reasoning.

Correlation analysis of time-resolved data under different conditions is a valuable method to infer regulation scenarios. The simplest possible scenarios leading to the increase or decrease of metabolite concentrations are considered. The ones that fit the qualitative behavior of the experimental data are scenarios that could be true regulation. This means that a simple first order network model can be used if the pathway is too complex to derive regulation by reasoning. This is also shown for the TCA cycle in paragraph 4.3.2.2.

This example illustrated how the possible regulation scenarios can be inferred from the reversal of correlations under aerobic conditions. These regulation scenarios do not depend on the kinetics (e.g. first or second order kinetics) used in the hypothetical network model. Paragraph 3 of the Supplementary Material shows that the initial qualitative behavior of the metabolite concentration profiles after the pulse is independent of the order of the mass-action kinetics.

If the pulse does not cause relative changes in reaction rates, the response shows positive correlation between B and D (see Figure 4.7). This is what happens in glycolysis after a glucose pulse under anaerobic conditions.

4.3.2.2 TCA cycle.

The hypothetical network model in Figure 4.9 was used to infer possible regulation scenarios from the reversed correlations between metabolites in the TCA cycle. To obtain the hypothetical network, the following pairs of reactions were lumped together because their common intermediate is not measured: PYR → ACoA and ACoA → CIT ($k_1$); CIT → ISOCIT and ISOCIT → OGL ($k_2$); OGL → SUCCoA and SUCCoA → SUC ($k_3$); PYR → OAA and OAA → MAL ($k_7$). The TCA cycle can show two types of behavior. It plays a crucial role in ATP production under respiratory conditions (for example glucose-limited aerobic conditions) and behaves as a cycle. Under fermentative (for example anaerobic) conditions, TCA cycle activity has only a role in the formation of precursors of amino acids and does not act as a cycle, but as two
Inferring differences in the distribution of reaction rates across conditions

Figure 4.6: *In silico* hypothetical network model used to study reversal of correlations in glycolysis. Abbreviations: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; GAP, glyceraldehyde-3-phosphate; DHAP, dihydroxyacetone phosphate; BPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; GLYC, glycerol.

Figure 4.7: Response of the metabolites in the *in silico* hypothetical network model of Figure 4.6 to an increase in the inflow. The numbers on the axes are in arbitrary units.
Inferring differences in the distribution of reaction rates across conditions

Figure 4.8: Changes in the parameters that lead to the emergence of a negative correlation between B and D: a) a lower value for $k_2$; b) a higher value for $k_3$; c) a lower value for $k_4$; d) a higher value for $k_6$. The numbers on the axes are in arbitrary units.
Inferring differences in the distribution of reaction rates across conditions

branches, leading to oxoglutarate and succinate [22, 161]. Both types of TCA cycle behavior have to be present in the model and prior knowledge about the way the pathway acts before the perturbation is necessary to apply the procedure depicted in Figure 4.1. The mass balances for this system are (see Figure 4.9 for an explanation of the symbols):

\[
\begin{align*}
\frac{d[A]}{dt} &= in - (k_1 + k_7 + k_8) * [A] \\
\frac{d[B]}{dt} &= k_8 * [A] - k_9 * [B] \\
\frac{d[C]}{dt} &= k_7 * [A] - (k_{5b} + k_6) * [C] + k_{5a} * [E] \\
\frac{d[D]}{dt} &= k_1 * [A] + k_6 * [C] + k_2 * [D] \\
\frac{d[E]}{dt} &= k_{5b} * [C] - (k_{4b} + k_{5a}) * [E] + k_{4a} * [G] \\
\frac{d[F]}{dt} &= k_2 * [D] - (k_3 + k_{11}) * [F] \\
\frac{d[G]}{dt} &= k_{4b} * [E] - (k_{4a} + k_{10}) * [G] + k_3 * [F]
\end{align*}
\]

Parameter values are: \( k_{4b} = k_{5b} = k_7 = 0 \) and all other parameters are 1 when the pathway acts as a cycle (for example aerobic glucose-limited conditions); \( k_3 = k_{4a} = k_{5a} = k_6 = 0 \) and all other parameters are 1 when the pathway acts as two branches (for example anaerobic conditions).

When the pathway acts as a cycle, the steady state of this system is \([A] = [B] = 1/2; [C] = [E] = [G] = 1/6; [D] = 2/3; [F] = 1/3\). A pulse is simulated by changing the inflow to \( in = 2 \). If the pulse did not cause relative changes in reaction rates, all of the metabolite concentrations would increase (see Supplementary Figure 4). In the experimental data, citrate decreases and the other metabolites increase (see Supplementary Figures 9(e-h) and 10(e-h)), which results in the emergence of negative correlations between metabolites in the TCA cycle (CIT-OGL, CIT-MAL and CIT-FUM). A decrease in D and an increase of A, C, E, F and
Inferring differences in the distribution of reaction rates across conditions

G are observed for the model in the following cases (see Supplementary Table 2):

- After the pulse, the pathway acts as two branches (switch from respiration to fermentation)(see Supplementary Figure 5a).
- After the pulse, there is a larger rate of the reaction converting citrate($D$) to oxoglutarate($F$)(higher value for $k_2$, see Supplementary Figure 5b).

The first scenario, the TCA cycle has different behavior before and after the pulse, is in accordance with results of experiments from Kresnowati et al. [104].

When the pathway acts as two branches (for example in the anaerobic case), the following steady state is found: $[A] = [B] = [C] = [D] = [E] = [F] = [G] = 1/3$. A pulse is simulated by changing the inflow to $in = 2$. All concentration profiles would increase if the pulse did not cause relative changes in reaction rates (see Supplementary Figure 6). In the experimental data, pyruvate increases and citrate, malate, fumarate and oxoglutarate decrease (see Supplementary Figure 8(c-d)). Succinate increases and is negatively correlated with citrate, malate, fumarate and oxoglutarate. Changing only one reaction rate did not result in a scenario that is in accordance with the data (see Supplementary Table 3). Decreasing $k_1$ agreed best with the experimental data. Therefore all scenarios where $k_1$ is decreased and a second parameter changed (increase or decrease) were examined. Supplementary Table 4 shows that none of these scenarios is fully consistent with the experimental data. The two scenarios that show the most similarities with the experimental data are: 1) a decrease in $k_1$ and an increase in $k_{4b}$; 2) a decrease in $k_1$ and an increase in $k_{5b}$. All scenarios that are combinations of the former two and a change of a third parameter were simulated. The results are shown in Supplementary Table 5. An increase of A and G and a decrease of C, D, E and F is observed in the following situations:

- a smaller rate from pyruvate(A) to citrate(D), a larger rate from fumarate(E) to succinate(G), a larger rate from malate(C) to fumarate(E) (smaller value for $k_1$, higher values for $k_{4b}$ and $k_{5b}$, see Supplementary Figure 7a);
Inferring differences in the distribution of reaction rates across conditions

Figure 4.9: *In silico* hypothetical network model used to study reversal of correlations in the TCA cycle. Blue arrows: reaction included only in the TCA cycle under respiration (cycle); red arrows: reaction included only in the TCA cycle under fermentation (two branches). Reaction scheme based on Camarasa *et al.* [22]. Abbreviations: PYR, pyruvate; ETH, ethanol; OAA, oxaloacetate; MAL, malate; CIT, citrate; ISOCIT, isocitrate; FUM, fumarate; OGL, oxoglutarate; SUC, succinate.

- a smaller rate from pyruvate(A) to citrate(D), a larger rate from fumarate(E) to succinate(G), a smaller rate from pyruvate(A) to malate(C)(smaller values for $k_1$ and $k_7$, higher value for $k_{4b}$, see Supplementary Figure 7b);

### 4.4 Validation

The proposed method was validated for glycolysis by examining whether one of the regulation scenarios inferred from the correlation analysis is in accordance with the way glycolysis is regulated after a glucose pulse under aerobic conditions. For this purpose we used the latest detailed kinetic model for aerobic glycolysis in *S. cerevisiae*, developed by van Eu-
Inferring differences in the distribution of reaction rates across conditions

nen [209](see Supplementary Material for details on the model). Figure 4.10 shows the rate of the reactions from FBP to GAP/DHAP ($v_{ALD}$), from DHAP to glycerol ($v_{GLY}$), from BPG to 3PG ($v_{PGK}$) and from 3PG to 2PG ($v_{PGM}$) under aerobic conditions, simulated with the model. The rates of the reactions from FBP to GAP/DHAP and from BPG to 3PG are increasing (see Figure 4.10a and c), and rate of reactions from DHAP to glycerol is constant (due to the model’s assumption of constant rates to the branches, see Supplementary Material and Figure 4.10b). This disproves the first three scenarios proposed in paragraph 4.3.2.1. Figure 4.10d confirms that the fourth glycolysis-case regulation scenario, a larger rate of the reaction from 3PG to 2PG, is the correct one. Larger reaction rates in lower glycolysis after a glucose pulse under aerobic conditions can be explained by the feed-forward activation of pyruvate kinase by FBP, which increases rapidly after the pulse [79] (see Figure 4.11). Because of the feed-forward activation, the reactions in the lower part of glycolysis become faster. As a consequence, the concentration of 3PG decreases and the correlation between FBP and 3PG is negative. More details are given in paragraph 2 of the Supplementary Material.

It can thus be concluded that when no comprehensive model about the studied pathway is available, the possible regulation scenarios for a given pathway can be inferred from correlations in time-resolved data with the help of a simple hypothetical model.

4.5 Discussion

In this study, a correlation-based method for inferring regulation scenarios from time-resolved metabolomics data is illustrated on the central carbon metabolism of *S. cerevisiae*. The scenarios are inferred by changing one or more parameters in a hypothetical network that is a simplification of the studied pathway. An important advantage of this method is that information about regulation can be obtained without detailed knowledge of kinetic parameters. Furthermore, the method uses straightforward mathematics and is easy to implement. A major limitation of the method is that the number of scenarios that need to be tested increases with the number of reactions in the hypothe-
Inferring differences in the distribution of reaction rates across conditions

tical network model. In the worst case, $3^n$ scenarios need to be tested for a network with $n$ reactions (every parameter can increase, decrease or remain unchanged). This can be avoided by using the forward selection procedure described in paragraph 4.2.2.2. For example, for the TCA cycle under anaerobic conditions, the reaction scheme consists of nine reactions, so in principle $3^9 = 19683$ scenarios have to be tested. By using the forward selection procedure, only 62 scenarios had to be checked (see Supplementary Tables 3-5).

A drawback of the forward selection method is that added parameters can not be removed later, which can lead to local solutions [122].

Applicability of methods based on time-resolved data is currently restricted by the measured pathways, which is usually only a small part of the metabolism. Often not all metabolites of a pathway are measured. When the common intermediates of reactions are not measured, the reactions can be lumped together.

Because analytical methods continuously improve, it is expected that future time-resolved metabolomics data will cover larger parts of metabolism.
Inferring differences in the distribution of reaction rates across conditions

Figure 4.10: Reaction rates before and after a 2.3 mM glucose pulse under aerobic conditions, simulated with the model of van Eunen [209]. a) rate of the reaction from FBP to GAP/DHAP ($v_{ALD}$) b) rate of reactions from DHAP to glycerol ($v_{GLY}$) c) rate of the reaction from BPG to 3PG ($v_{PGK}$) d) rate of the reaction from 3PG to 2PG ($v_{PGM}$). Abbreviations: ALD, aldolase; GLY, glycerol; PGK, 3-phosphoglycerate kinase; PGM, phosphoglycerate mutase. From the scenarios described in paragraph 4.3.2.1, only the fourth one is in accordance with these figures. Figures a-c disprove the first three scenarios. We can conclude that an increase in the reaction rate that further metabolizes 3PG (to 2PG) is the correct one.
Inferring differences in the distribution of reaction rates across conditions

Figure 4.11: Feed-forward activation of pyruvate kinase by FBP. Abbreviations: GLUC, glucose; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; PYR, pyruvate; PYK, pyruvate kinase.

4.6 Conclusion

In this study, a priori information about the topology and directionality of a pathway is combined with time-resolved data of metabolites to infer possible regulation mechanisms of the network. By combining correlations in time-resolved metabolomics data with a simple hypothetical network model, possible regulation scenarios could be formulated without a priori knowledge of detailed kinetics. Method validation with a detailed kinetic model as benchmark confirmed that the proposed method was indeed able to infer the correct regulation scenario in aerobic glycolysis of S. cerevisiae after a glucose pulse from the correlations in time-resolved metabolomics data.
Inferring differences in the distribution of reaction rates across conditions

Acknowledgments

This project was financed by the Netherlands Metabolomics Centre (NMC) which is part of the Netherlands Genomics Initiative /Netherlands Organization for Scientific Research. Karen van Eunen and Barbara Bakker (University Medical Center Groningen) are gratefully acknowledged for providing us with the simulation models.

Supplementary Data

The supplementary data is too extensive to be integrally included in this thesis. It can be accessed online at http://www.rsc.org/suppdata/mb/c2/c2mb25015b/c2mb25015b.pdf