DTU Library

Integration of metabolome data with metabolic networks reveals reporter reactions

Çakir, Tunahan; Patil, Kiran Raosaheb; Önsan, Zeynep Ilsen; Ülgen, Kutlu Özergin; Kirdar, Betül; Nielsen, Jens

[Link back to DTU Orbit]

molecular
systems
biology

Integration of metabolome data with metabolic networks reveals reporter reactions

Tunahan Çakır[1,3], Kiran Raosaheb Patil[2,3], Zeynep İlsen Önsan[1], Kutlu Özergin Ülgen[1], Betül Kırdar[1] and Jens Nielsen[2,*]

[1] Department of Chemical Engineering, Boğaziçi University, Bebek, Istanbul, Turkey and [2] Center for Microbial Biotechnology, Biocentrum-DTU, Technical University of Denmark, Kgs. Lyngby, Denmark
* Corresponding author. Center for Microbial Biotechnology, Biocentrum-DTU, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
Tel: + 45 45 25 26 96; Fax: + 45 45 88 41 48; E-mail: jn@biocentrum.dtu.dk
[3] These authors contributed equally to this work

Interpreting quantitative metabolome data is a difficult task owing to the high connectivity in metabolic networks and inherent interdependency between enzymatic regulation, metabolite levels and fluxes. Here we present a hypothesis-driven algorithm for the integration of such data with metabolic network topology. The algorithm thus enables identification of reporter reactions, which are reactions where there are significant coordinated changes in the level of surrounding metabolites following environmental/genetic perturbations. Applicability of the algorithm is demonstrated by using data from *Saccharomyces cerevisiae*. The algorithm includes preprocessing of a genome-scale yeast model such that the fraction of measured metabolites within the model is enhanced, and thus it is possible to map significant alterations associated with a perturbation even though a small fraction of the complete metabolome is measured. By combining the results with transcriptome data, we further show that it is possible to infer whether the reactions are hierarchically or metabolically regulated. Hereby, the reported approach represents an attempt to map different layers of regulation within metabolic networks through combination of metabolome and transcriptome data.
Molecular Systems Biology 3 October 2006; doi:10.1038/msb4100085
Subject Categories: bioinformatics; metabolic and regulatory networks
Keywords: data integration; genome-scale stoichiometric model; metabolic regulation; quantitative metabolomics; reporter reactions

Introduction

One of the goals of systems biology is to obtain overall quantitative description of cellular systems. This is currently not achievable as the number of components and interactions involved in these systems is quite large resulting in a very large parameter space. Thus, methods are required to reduce the dimensionality and particularly identify key regulatory points in the many different cellular processes. Metabolism is a good starting point to develop such analysis methods as it is studied in great detail and well annotated. Furthermore, genome-scale metabolic models have been developed for many different cellular systems (Edwards and Palsson, 2000; Förster *et al*, 2003; Sheikh *et al*, 2005), and besides their use for simulation of cellular function (Edwards *et al*, 2001; Famili *et al*, 2003; Price *et al*, 2004), these models can serve as scaffolds for analysis of genome-scale biological data (Covert *et al*, 2004; Borodina and Nielsen, 2005). This has been demonstrated recently for analysis of transcriptome data, where the use of genome-scale metabolic models enabled identification of coregulated sub-networks and reporter metabolites (Patil and Nielsen, 2005). Although transcriptome data provide an overview of the global regulation in the metabolism,

understanding of cellular physiology is incomplete without knowledge of metabolome, owing to the high connectivity in metabolic networks and inherent interdependency between enzymatic regulation, metabolite levels and fluxes (Nielsen, 2003). Metabolites, acting as intermediates of biochemical reactions, play a crucial role within a living cell by connecting many different operating pathways. Metabolite levels are determined by the concentrations and the properties of the surrounding enzymes, making their levels a complex function of many cellular regulatory processes in different dimensions. Thus, the metabolome represents a snapshot of the functioning metabolism of the cell and hence provides valuable information about regulation of several different cellular processes (Villas-Bôas *et al*, 2005c). Consequently, in recent years, there has been increased focus on analysis of the metabolome (Sumner *et al*, 2003; Bino *et al*, 2004; Villas-Bôas *et al*, 2005c). Even though traditional data analysis methods like principal component analysis, clustering analysis and chemometrics have shown to be efficient for analysis of this kind of data (Raamsdonk *et al*, 2001; Allen *et al*, 2003), there are some limitations with these methods for uncovering the underlying biological principles (Weckwerth *et al*, 2004). Furthermore, there are still only few examples of studies on the

use of metabolome data to understand regulatory principles in metabolism.

Functional analysis of cellular metabolism and in particular integration of metabolome data with other omics-data demands (semi-)quantitative measurements of key metabolites. However, a problem with metabolomics is the scarcity of targeted quantitative data, and often metabolome analysis is (at best) semiquantitative even though there is a trend towards more quantitative analysis (Nielsen and Oliver, 2005). Although it is currently not yet possible to quantify all the metabolites in a cellular system (Fernie *et al*, 2004; Goodacre *et al*, 2004), a high-throughput GC–MS method that allows semiquantitative identification of several metabolites in *Saccharomyces cerevisiae* was recently developed (Devantier *et al*, 2005a; Villas-Bôas *et al*, 2005a). In the latter studies, the levels of 52 unique metabolites (out of 584 reported unique metabolites in the genome-scale yeast model; Förster *et al*, 2003) were determined in genetically different yeast strains under different environmental conditions. Specifically, metabolites playing important roles in the central carbon metabolism and amino-acid biosynthesis could be identified.

In order to understand the regulatory principles underlying the changes in metabolite levels, we developed an algorithm that enables integration of such quantitative metabolome data with genome-scale models by using a graph theoretical representation of the metabolism. We demonstrate the application of this algorithm for the metabolome data reported by Villas-Bôas *et al* (2005a) and Devantier *et al* (2005a). We use the significance of changes in the metabolite levels to identify reporter reactions around which the most significant coordinated metabolite changes are observed. Reporter reaction analysis is an attempt to infer the differential reaction significance based on metabolite measurements, and hence provides a basis for understanding the underlying cellular processes responding to the perturbations. We further demonstrate that through combination with transcriptome data, reporter reactions may provide clues on whether regulatory control at a given reaction node is at the metabolic level or at the hierarchical level.

Results and discussion

Model preprocessing

Owing to the large chemical diversity of the metabolome, there is currently no single analytical method that enables analysis of the complete metabolome. Even the best analytical methods reported to date for metabolome analysis therefore only cover a small fraction of the metabolites present in genome-scale metabolic models. The unavailability of data for a large number of metabolites is one of the major problems associated with mapping (and hence integration) of metabolome data on to genome-scale metabolic networks. In order to overcome this fundamental problem, we preprocessed the genome-scale model of Förster *et al* (2003) so as to obtain a reduced model, where the fraction of experimentally measured metabolites was enriched. This processing was carried out by systematically eliminating unmeasured metabolites from the metabolic network. We note that the model preprocessing is dependent on the metabolome data that are available, and the

preprocessing will have to be carried out for each case. However, following the flow-chart depicted in Supplementary Figure 1, this preprocessing is relatively straight forward and can easily be carried out also for other metabolic networks.

The yeast genome-scale model includes three compartments (mitochondria, cytosol and external space) with 844 metabolites (559 cytosolic, 164 mitochondrial, 121 external) and 1175 reactions (Förster *et al*, 2003). Within the context of this model, metabolites present in more than one compartment are treated as if they are different entities in each compartment. However, the experimental data used in this analysis (and most of the data sets available to date) can only differentiate between extracellular and intracellular space. As metabolite levels in different cellular compartments are not available, the cytosolic/mitochondrial compartmentation of the model was removed and corresponding metabolites were represented as one, with their corresponding reactions conserved. Also, there are a number of duplicate reactions owing to the presence of isoenzymes in the model, and these reactions were lumped into single reactions, as metabolome data alone do not provide information that enables distinction between the operations of different isoenzymes. As a result, the 'processed' model (uncompartmented model, UNCOMP) consists of 677 metabolites (559 internal, 118 external) with 725 reactions, including transport reactions. With this model, the experimental data used here amount to about 12% of these 677 metabolites (52 internal, 32 external).

Enzyme subsets are enzymes that always operate together in fixed flux proportions at steady state (Pfeiffer *et al*, 1999; Schuster *et al*, 2002), often representing enzymes in linear pathways. Accordingly, the intermediate metabolites in enzyme subsets can be assumed to be similarly affected by the perturbations. UNCOMP was further reduced in size by using METATOOL 4.3 (Pfeiffer *et al*, 1999; Dandekar *et al*, 2003) and thus representing each enzyme subset as a single reaction. The resulting model (enzyme-subset model, ENZSUB-1) consists of 563 metabolites and 590 reactions and it has about 15% of the metabolites measured within the data used. As the removal of the metabolites in linear pathways also led to the omission of six measured metabolites, the reactions containing these metabolites were restored back into the ENZSUB-1 model. To further increase the fraction of the measured metabolites, potentially inactive (or potentially low flux) reactions were removed. This was carried out by using flux balance analysis (FBA) (Varma and Palsson, 1994; Kauffman *et al*, 2003) for simulation of fluxes at specific environmental conditions used in the experiments (aerobic and anaerobic batch cultivation in glucose-limited minimal media). The ENZSUB-1 model was used to simulate the fluxes with the objective of optimum growth. Then, the maximum and the minimum flux for each reaction in the model were obtained by constraining the specific growth rate between its optimum value and 50% of the optimum. Reactions that had zero flux in the FBA analysis (at both optimum values) were considered as potentially invariant between the studied perturbations and thus omitted from the ENZSUB-1 model. The resulting model had 349 reactions involving 267 metabolites. The here-used FBA-based approach for model reduction does not necessarily imply that the eliminated reactions are inactive and that the metabolites involved in
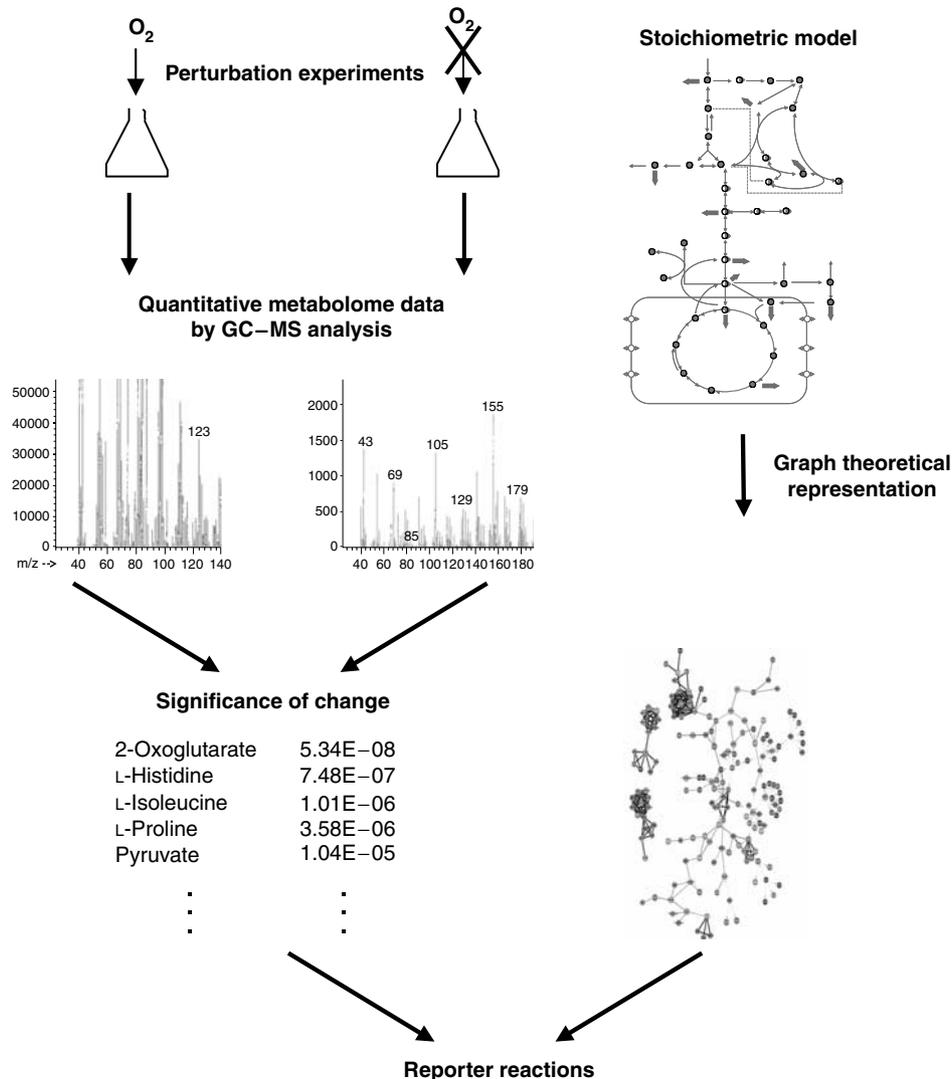
Figure 1 Reporter reaction algorithm to identify differential reaction significance by integrating metabolome data with metabolic networks. Quantitative metabolome data obtained from perturbation experiments are interpreted in terms of significance of change, and mapped onto the stoichiometric network, which is represented as bi-partite undirected graph, to identify reporter reactions.

these reactions not present in the cell. However, it is assumed that as these reactions are likely to carry very low fluxes under the studied conditions, the associated metabolite pools are likely to be weakly affected because of the changes in the fluxes through these reactions. Although this approach is useful, the assumption is not fool-proof as we indeed found certain measured metabolites that were intermediates in pathways with zero fluxes (pimelic acid, PIMExt, myristic acid, C140xt, *trans*-4-hydroxy-L-proline, itaconate, nicotinate, 4-aminobenzoate, THMxt). The first six of these metabolites were detected as 'invariant' by the FBA approach owing to the fact that that these metabolites are not connected to the overall network (Förster *et al*, 2003). However, here we restored back reactions involving these measured metabolites, and the resulting model comprised a total of 285 metabolites participating in 361 reactions (Supplementary Figure 1). Even though certain reactions may be removed from the analysis by using this approach, the algorithm will still correctly identify

reporter reactions, given the metabolome data set. The resulting metabolic network, ENZSUB-2 model, was substantially enriched in terms of the content of measured metabolites (now accounting for about 30%).

In order to further focus the analysis only on reactions involving measured metabolites, ENZSUB-3 was constructed by keeping only reactions that involved at least one measured metabolite. Additionally, only one member of the NADH/ NAD^+, NADPH/$NADP^+$, $FADH_2$/FAD^+ cofactor pairs, when available, was retained in the remaining reactions as the levels of members of each pair were assumed to be interdependent. The resulting metabolic network, ENZSUB-3, included a total of 178 metabolites participating in 139 reactions, which corresponds to more than 47% of the available quantitative metabolome data (Supplementary Figure 1). The 139 reactions included in the model are given in Supplementary Table 1.

The significance of the change in the levels of metabolites between any two conditions was calculated by applying a

statistical test (see Materials and methods). However, it is difficult to deduce which reactions in the cell are affected most by only judging the significance of the change in metabolite levels, as the number of the metabolic reactions in the cell is high and one metabolite usually appears in more than one reaction. Thus, we calculated a normalized Z-score for each reaction based on the z-values of its neighboring metabolites (P-values of individual metabolites were converted to Z-scores by using inverse normal cumulative distribution function; see Materials and methods). Here we assume that the calculated reaction Z-scores can be regarded as an indicator of the significance of how the reactions respond to the studied perturbation at metabolic level. This assumption is based on the fact that metabolite levels are governed by changes in fluxes and enzyme activities (Nielsen, 2003). Reactions exhibiting significant changes (typically $z > 1.28$, corresponding to $P < 0.10$) for the perturbations analyzed were identified by using the graph representation of the derived metabolic model, ENZSUB-3, and are listed in Tables I and II. A loose cutoff was deliberately chosen, as we did not want to be too biased in the light of the fact that measurements were not available for all of the metabolites in the model, and thus the resultant P-values are in fact, in general, shifted to high values owing to randomly selected P-values for those unmeasured metabolites.

Effect of an altered redox metabolism and oxygen availability

As a first demonstration of our approach, we considered data from metabolome analysis of two different *S. cerevisiae* strains, a wild-type laboratory strain (CEN.PK.113-7D) and a redox-engineered strain, which was carried out in batch cultures under two different environmental conditions (aerobic and anaerobic) in standard mineral media with glucose as the sole carbon source (Villas-Bôas *et al*, 2005a). The redox-engineered strain carrying a deletion of the NADPH-dependent glutamate dehydrogenase encoded by *GDH1* and an overexpression of the NADH-dependent glutamate dehydrogenase encoded by *GDH2* was constructed by dos Santos *et al* (2003). Three different perturbations were analyzed here: genetic change under both aerobic and anaerobic conditions (wild-type versus redox-engineered strain), and environmental change for the wild-type strain (aerobic versus anaerobic). As it was reported that sample-to-sample variability exceeds flask-to-flask variability, replicate samples from different shake flasks were treated equivalently (Villas-Bôas *et al*, 2005a). Accordingly, the metabolome data set includes around 15 intracellular and nine extracellular replicates for each experimental condition. The data set used in this study is available in Supplementary information as normalized abundances of GC–MS peaks.

Comparison of the wild-type and mutant strains revealed that the genetic changes do not alter the basic growth characteristics in aerobic (dos Santos *et al*, 2003) and anaerobic (Nissen *et al*, 2000) batch cultivations. Our approach, however, captures the associated changes in different cellular pathways, by identifying a number of significantly affected reactions due to these perturbations. The detected reactions (Table I) belong to many different amino-acid pathways, indicating a widespread effect of the mutation on the cellular metabolism. The present integrated approach also differentiates between the genetic perturbation under aerobic and anaerobic conditions, as there are reactions that are specific to each condition.

Genetic perturbations (wild type versus redox engineered) used in the present study are directly related to a changed redox metabolism. Environmental perturbation (aerobic

Table I Reactions with significant Z-scores ($P < 0.10$, $z > 1.28$) in response to genetic perturbations by altered redox metabolism and environmental perturbation by oxygen availability[a,b,c]

Genetic perturbation (aerobic)			Genetic perturbation (anaerobic)			Environmental perturbation (wild-type strain)		
VALsyn	(4/4)	2.90	AGX1	(4/4)	2.67	UGA[ES]	(5/5)	2.41
ALT	(4/4)	2.83	ALT	(4/4)	2.35	ALT	(4/4)	2.34
LEUsyn[ES]	(5/6)	2.66	**PROsc**	(2/3)	2.08	AGX1	(4/4)	2.34
TYRsyn	(3/4)	2.54	LEUsyn[ES]	(5/6)	1.80	CAR2	(3/4)	1.95
CAR2	(3/4)	2.50	ASP3-1	(2/3)	1.78	LEUsyn[ES]	(5/6)	1.95
PHEsyn[ES]	(3/5)	2.25	**U46_**	(3/4)	1.64	TYRsyn	(3/4)	1.92
AGX1	(4/4)	2.01	**CHA1p**	(2/3)	1.58	VALsyn	(4/4)	1.87
AAT	(4/4)	1.86	PHEsyn[ES]	(3/5)	1.57	PHEsyn[ES]	(3/5)	1.74
ILEsyn[ES]	(6/7)	1.77	**PUT1**	(2/3)	1.55	SERsyn[ES]	(4/6)	1.67
SUCsc	(2/3)	1.66	VALsyn	(4/4)	1.54	**GAD1**	(2/3)	1.47
SDH	(2/3)	1.63	**GLY1**	(2/3)	1.50	**GDH13**	(3/4))	1.44
HISsyn[ES]	(4/10)	1.58	SERsyn[ES]	(4/6)	1.41	ASP3-1	(2/3)	1.39
ASP3-1	(2/3)	1.57				GDH2	(3/4)	1.38
GDH2	(3/4)	1.55				**MYRsc**	(2/2)	1.36
DLD	(2/4)	1.51				ILEsyn[ES]	(6/7)	1.36
UGA[ES]	(5/5)	1.48				HISsyn[ES]	(4/10)	1.34
SERsyn[ES]	(4/6)	1.46				**GLYsyn**	(2/4)	1.30
LEU4	(2/4)	1.36				**U155_**	(4/4)	1.29
FUM	(2/2)	1.28						

The number of measured metabolites and the total number of metabolites for each reaction are also given in parentheses. The explicit form of the reactions can be followed from Supplementary Table 1.
[a]Reactions specific to each perturbation are given in bold letters.
[b]ES means that the corresponding reaction is an enzyme subset consisting of a combination of more than one reaction.
[c]sc in some of the reaction names stands for 'secretion', indicating that they are secretion reactions.

Table II Effect of media change (standard medium versus VHG medium) on each strain analyzed by the developed approach; reactions with significant Z-scores ($P<0.10$, $z>1.28$) are shown[a]

Media change for laboratory strain (CEN.PK113-7D)		z_{RE}	z_{GE}	Media change for industrial strain (Red Star)		z_{RE}	z_{GE}
ALT	(4/4)	2.50	2.48	ALT	(4/4)	2.48	1.80
AGX1	(4/4)	2.45	0.86	AGX1	(4/4)	2.31	2.21
UGA[ES]	(5/5)	2.18	1.69	UGA[ES]	(5/5)	2.23	0.38
ECM40	(3/4)	1.85	2.39	**U155_**	(4/4)	2.01	—
GLUsc	(2/3)	1.85	1.17	ASN	(4/7)	1.85	0.54
ASN	(4/7)	1.84	2.30	**TYRsyn**	(3/4)	1.84	3.41
CAR2	(3/4)	1.74	0.57	GLUsc	(2/3)	1.81	0.79
LYSsyn[ES]	(7/8)	1.67	2.46	PHEsyn[ES]	(3/5)	1.65	0.98
TRP23	(3/5)	1.67	1.31	TRP23	(3/5)	1.65	0.78
ASP3-1	(2/3)	1.47	1.45	PROsc	(2/3)	1.45	0.90
CHA1p	(2/3)	1.47	0.93	ALAsc	(2/3)	1.45	0.75
U42_-43_	(2/3)	1.47	—	GLYsc	(2/3)	1.45	0.80
ASPsc	(2/3)	1.43	2.09	LACsc	(2/3)	1.45	0.86
PROsc	(2/3)	1.43	1.40	PYRsc	(2/3)	1.45	0.86
ALAsc	(2/3)	1.43	1.90	SUCsc	(2/3)	1.45	—
GLYsc	(2/3)	1.43	1.66	**CITsc**	(2/3)	1.45	—
LACsc	(2/3)	1.43	0.81	**AKGsc**	(2/3)	1.45	—
PYRsc	(2/3)	1.43	0.81	**U88_**	(2/3)	1.43	—
SUCsc	(2/3)	1.43	—	GAD1	(2/3)	1.43	1.21
GLY1	(2/3)	1.41	0.41	ILEsyn[ES]	(6/7)	1.42	1.90
VALsc	(2/3)	1.38	1.55	ASP3-1	(2/3)	1.41	1.34
PHEsyn[ES]	(3/5)	1.36	1.66	U42_-43	(2/3)	1.41	—
GAD1	(2/3)	1.29	1.46	**LEUsyn[ES]**	(5/6)	1.38	0.91
				ASPsc	(2/3)	1.29	0.86
				ARG5,6-8[ES]	(4/8)	1.28	0.88

Z-scores of the gene expression changes are also given for comparison. z_{RE}: Z-scores of reactions calculated by the developed approach, z_{GE}: Z-scores of genes/gene groups calculated from associated P-values from transcriptome data. The number of measured metabolites and the total number of metabolites for each reaction are also given in parentheses.
[a]Reactions specific to each perturbation are given in bold letters.

versus anaerobic) is, however, also associated with a changed redox metabolism owing to the direct effect of oxygen availability on the operation of the TCA cycle and the pentose phosphate pathway, and hence on the redox state of the cell. This is also reflected in the identified reporter reactions, as a number of common significantly changed reactions are observed for the two different types of perturbations (Table I and Supplementary Table 2).

The glutamate decarboxylase reaction (GAD1) appears as a significantly changed reaction specific to the environmental perturbation of the wild-type cells, which implies a major role of this reaction during respiratory growth (Table I). Indeed, it was reported (McCammon *et al*, 2003) that the defects in any of the 15 TCA cycle genes, associated with the slowing down of the respiratory metabolism, result in a substantial decrease in the mRNA levels of *GAD1*, which is in agreement with our findings. GAD1 constitutes the first step of the glutamate catabolic pathway towards succinate (Coleman *et al*, 2001). The downstream steps of the pathway are catalyzed by Uga1p and Uga2p (UGA[ES]), which are affected most by the environmental perturbation (Table I). Detection of all reactions of this pathway (GAD1, UGA[ES]) as responsive to the oxygen availability (Figure 2A) indicates that they have a key role in succinate production via glutamate under anaerobic conditions where the yeast is secreting succinate. In fact, this pathway was found to be activated during oxidative (Coleman *et al*, 2001) or osmotic (sugar) (Erasmus *et al*, 2003) stress to control the redox balance of the cell.

Although the glyoxylate cycle is generally believed to be repressed during growth on glucose, Villas-Bôas *et al* (2005b) found that an alternative pathway for glyoxylate biosynthesis is active in *S. cerevisiae*. Examination of the Z-scores of reactions involving glyoxylate for all the analyzed perturbations revealed that AGX1 (reaction of enzyme encoded by *YFL030w*), which enables synthesis of glyoxylate from glycine, has much higher scores for all the perturbations compared to the reactions of the glyoxylate pathway (ICL and MLS) (Figure 2B). Thus, our analysis supports the presence of an alternative pathway catalyzed by AGX1 leading to the biosynthesis of glyoxylate from glycine.

Reporter reaction analysis also identifies that the genetic perturbation results in metabolic changes around the genes that are perturbed (Figure 2C). Thus, the reaction responsible for the overexpressed gene in the redox-engineered strain, GDH2, has a significant Z-score for the genetic perturbation under aerobic condition. It should be mentioned that a genetic perturbation of a gene should not necessarily result in that the corresponding reaction comes out as a reporter reaction, as certain genetic perturbations may lead to only small changes in metabolite levels. However, in this case, there are two genetic modifications around α-ketoglutarate and glutamate (deletion of *GDH1* and overexpression of *GDH2*) that lead to identification of GDH2 as reporter. For the genetic change under anaerobic conditions, the detected significance of GDH2 is comparably lower. However, an indirect effect of the genetic modification in the glutamate biosynthesis can be observed
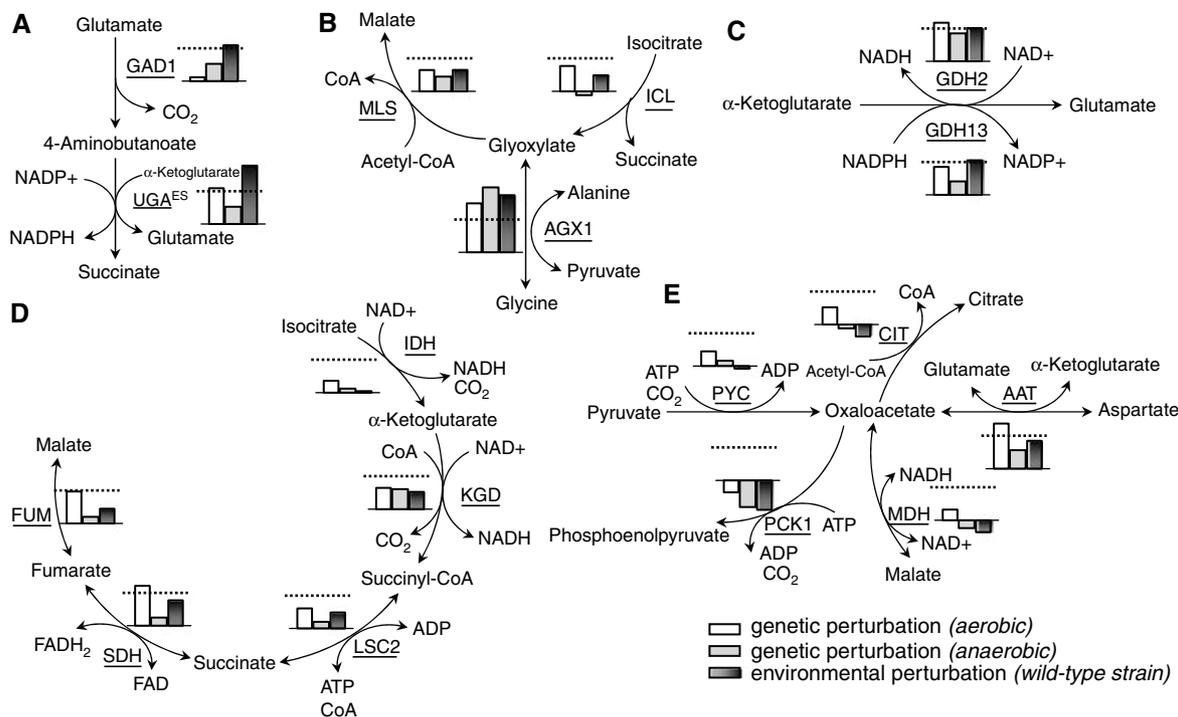
Figure 2 Example pathway structures based on *Z*-scores of reactions, which demonstrate the metabolomic response of the selected reactions in the reported case studies, namely, the effect of an altered redox metabolism and aerobic/anaerobic growth. The dashed lines correspond to the cutoff of 1.28 (*P*=0.10). See text for detailed discussion. (**A**) Glutamate catabolic pathway. (**B**) Glyoxylate metabolism. (**C**) Glutamate dehydrogenation reactions. (**D**) TCA cycle. (**E**) Oxaloacetate metabolism.

from the presence of transaminase activity associated with some of the identified reporter reactions for this perturbation (conversion of glutamate to α-ketoglutarate by ALT, LEUsyn[ES], PHEsyn[ES], VALsyn, SERsyn[ES]; Table I and Supplementary Table 2). On the other hand, the aerobic–anaerobic shift for the wild type gives rise to nearly the same *Z*-score for GDH2 reaction as the genetic perturbation under aerobic conditions. One explanation for this similarity in behavior would be that oxygen availability may have a direct effect on glutamate dehydrogenase genes; that is, cessation of oxygen uptake or manipulation of redox metabolism may result in similar effects on this node in the metabolism. In fact, in chemostat cultures, *GDH2* is associated with a significant transcription change when subjected to the same environmental perturbation (Piper *et al*, 2002). On the other hand, it is not possible to make a definite interpretation about the effect of the mutation on the deleted gene, *GDH1*, by looking at the *Z*-score of GDH13 reaction, as the reaction catalyzed by Gdh1p is identical to that catalyzed by Gdh3p. Consequently, what is reflected by this *Z*-score is the 'combined' response of these two enzymes. The reason that the GDH13 reaction is not identified as a reporter reaction whereas the GDH2 reaction is identified can only be explained by either a different response in the cofactor level as a consequence of the perturbations, that is, the NADPH/NADP$^+$ levels do not change as much as the NADH/NAD$^+$ levels, or due to measurement errors of these cofactors (these cofactors are inherently difficult to measure).

As the TCA cycle activity is known to be low under anaerobic conditions, the associated effect of genetic mutation under this condition is expected to be weaker than the other two perturbations analyzed. The *Z*-scores for the SDH and FUM reactions (both being part of the TCA cycle) are clearly in agreement with this expectation (Figure 2D). These two reactions are also members of the electron transport system, and this further explains why the metabolites surrounding these reactions exhibit remarkably weaker coordinated change in the genetic perturbation under anaerobic condition than in the other perturbations.

Similarly, the *Z*-scores of key reactions involving oxaloacetate suggest that these reactions are mainly affected in the redox-engineered strain under aerobic conditions (Figure 2E), and AAT, a transamination reaction leading to the conversion of oxaloacetate to aspartate, appears to be the key reaction where oxaloacetate is involved. There are no literature data available about the effect of the genetic perturbation on this metabolic reaction, but as the genetic perturbation results in a changed ratio of glutamate to 2-oxoglutarate (Villas-Bôas *et al*, 2005a), it may have affected this important transamination reaction.

Effect of very high-gravity fermentation

As a second demonstration of our approach, we used metabolome data from two different *S. cerevisiae* strains, a laboratory strain (CEN.PK.113-7D) and an industrial strain used for fuel ethanol production (hereafter termed as 'Red Star'). For both strains, the data were obtained from anaerobic batch cultures under two different cultivation conditions: exponential growth in glucose-containing standard mineral media and the stationary phase in maltodextrin-containing

very high-gravity (VHG) mineral media (Devantier *et al*, 2005a). Environmental perturbations obtained through variation in the media were analyzed here for each strain. The intracellular metabolome data set includes four replicates for the standard medium and eight replicates for the VHG medium. The extracellular metabolome data set has six replicates for each condition. The complete data set is available in Supplementary information.

As for the first case study discussed above, the two media perturbations analyzed revealed the same trend for the glyoxylate reactions, pointing to substantial regulation of the AGX1 reaction node in both perturbations (data not shown). In case of the glutamate metabolism, all the reactions have noticeably higher *Z*-scores, except GDH2, implying that this pathway is highly affected by VHG-associated media changes. All of the TCA cycle reactions shown in Figure 2D have very low *Z*-scores, in accordance with the fact that the cycle is barely operational under any of the experimental conditions studied (anaerobic fermentations). For reactions involving oxaloacetate, AAT again appears to play the major role as observed in the first data set, in parallel with the graph shown in Figure 2E.

The reaction governed by Gad1p, which catalyzes decarboxylation of glutamate—a reaction that is generally considered to be associated with stress—is found to be significantly changed in both strains when the media were changed (Table II). A noticeably lower score was obtained for comparison of the two strains grown on the standard medium (results not shown), which shows that the standard medium imposes less stress compared with the VHG medium where sugar and ethanol stresses are predominant. The appearance of all reactions (GAD1, UGA^{ES}) involved in the glutamate catabolic pathway as reporter reactions when the media are perturbed (Table II) points to the fact that this perturbation has a major effect on the amino-acid metabolism, and probably also on the redox balance in the cell. The results of transcriptome analysis for the same strains in standard and VHG media (Devantier *et al*, 2005b) indicate that the strains have differences in their redox balancing, confirming our finding.

A large number of transport reactions were found to have significant *Z*-scores (Table II). GC–MS analysis of extracellular metabolites in the VHG medium revealed many more metabolites compared to what is found in the standard medium, explaining the appearance of transport reactions as significant. The here-reported algorithm allowed us to identify and quantify the secretion reactions which are mostly affected from the media change, by integrating both intracellular and extracellular measurements to the reaction network. Secretion of a number of amino acids (glutamate, aspartate, proline, alanine and glycine), and succinate, pyruvate and lactate is commonly and significantly regulated in response to media perturbation for both the laboratory and the red star strain. On the other hand, detection of strain-specific secretion patterns (valine, citrate and α-ketoglutarate; Table II) points to differences in operation of the metabolic network in the two strains, possibly arising from the difference in the redox metabolism of the two strains.

As the change in the fermentation medium led to ethanol and osmotic stress for both strains (Devantier *et al*, 2005a), it is not surprising that many of the reactions are shared in the identified lists for the two strains in the media comparison (Table II). Transcriptome analysis of this data set revealed that a substantial part of the significantly changed genes were involved in protein synthesis and amino-acid metabolism (Devantier *et al*, 2005b). Thus, amino-acid pathway reactions detected by our analysis (Table II) are in accordance with the transcriptome data. Absence of amino-acid synthesis in VHG media owing to the cessation of growth in the stationary phase can be a possible cause of the observed differences.

Integration of metabolome data with transcriptome data for understanding regulation

For the latter case study, where the effect of a VHG medium was analyzed on the metabolome of laboratory and industrial strains, a genome-wide expression analysis was also performed (Devantier *et al*, 2005b). This basically enables further evaluation of mode of regulation for the different reactions in the reduced metabolic network. ter Kuile and Westerhoff (2001) introduced the concept of metabolic regulation and hierarchical regulation, where the first indicates that regulation of flux is at the level of enzyme kinetics, that is, through changes of the metabolite levels, and the second indicates that regulation of flux is at the level of enzyme production/activity (transcription/translation/post-translational modification). As both metabolite data and transcription data are available for this case study, we looked into whether it was possible to identify the type of regulation at the individual reaction level. A major obstacle for this kind of analysis is, however, that we do not have information about changes in fluxes for the analyzed conditions, and such data would also be difficult to obtain. Although there are efficient methods for obtaining data on the metabolic fluxes in the central carbon metabolism (Nielsen, 2003), it is difficult to get good estimates for the fluxes in all pathways of the metabolic network analyzed here, and even though the fluxes can be calculated by using FBA, this method is not well suited to give precise estimates for the actual fluxes in networks where there are redundant pathways. In order to proceed with analysis, we therefore assumed that whenever there was a coordinated significant change in metabolite levels around a reaction, then it is very likely that the flux through this reaction is also changing. However, there is no guarantee that the flux through this reaction is also changed as there could also be a change in the enzyme concentration, or there could even be altered allosteric regulation of the enzyme, thus keeping the flux unchanged. Thus, our assumption may result in identification of some false positives, but still the analysis would clearly lead to identification of reactions around which there is at least one level of regulation (and possibly several levels of regulation), and we will therefore refer to these reactions as being metabolically regulated. For all the reactions that are not identified as reporter reactions, we cannot infer anything about whether the flux has changed, but still we can deduce from the transcription data whether there has occurred regulation at the hierarchical level, and even though this does not necessarily mean hierarchical regulation of the flux, we will refer to these reactions as being hierarchically regulated.

This deduction can still be informative as indicator of the logic of transcriptional regulatory machinery governing gene expression. For cases where there was a significant change at the transcriptional level for an identified reporter reaction, we considered this to be a situation where there was mixed regulation.

The metabolic network includes several enzymes (hence reactions) governed by multiple genes. Thus, in order to infer about the significance of change in expression levels for the reactions, we summed the transcript levels for all genes coding for the same reaction before applying the statistical test. The P-values of transcripts were then calculated by using a t-test with unequal variance, and further converted into Z-scores to enable a comparison with the Z-scores of reactions based on metabolome data.

Using this approach, we grouped all the reactions of the metabolic network into whether they were metabolically or hierarchically regulated (or a combination or not regulated at all) for the VHG data set. To score the magnitude of the regulation at the hierarchical and metabolic levels, we used the corresponding Z-scores. The qualitative evaluation of Z-scores emerging from the transcriptome and the metabolome data enabled us to get an indication of regulation within the metabolic network (see Supplementary Figure 2 and Supplementary Table 3). The cases where only the transcript Z-score is significantly changed can be scored as points with possible hierarchical regulation, whereas the opposite case where only the metabolite-based Z-score has significantly changed implies metabolic regulation of the corresponding reaction (Rossel et al, 2005). When both Z-scores are significant, there is regulation shared at both levels, and when none of the Z-scores are significant, it is not possible to infer about at which level there is regulation.

Of the 121 reactions in the model having corresponding genes associated with them, the number of reactions predicted to be regulated hierarchically, metabolically and at both levels were 56, 7 and 14, respectively for the media perturbation with the laboratory strain, and 31, 14 and 5 for the same perturbation with the industrial strain (Figure 3 and Supplementary Table 3). For the laboratory strain, 44 reactions were found to be relatively irresponsive to the perturbation. On the other hand, the number of potentially unregulated reactions was much higher (71) for the industrial strain. One explanation for the observed predominance of transcriptional regulation could be the fact that the strains protect themselves against the applied perturbation by mainly changing their gene expression to minimize the changes in the metabolome, an observation also encountered in plants (Hirai et al, 2004). Figure 3 and Supplementary Table 3 suggest that metabolic regulation is
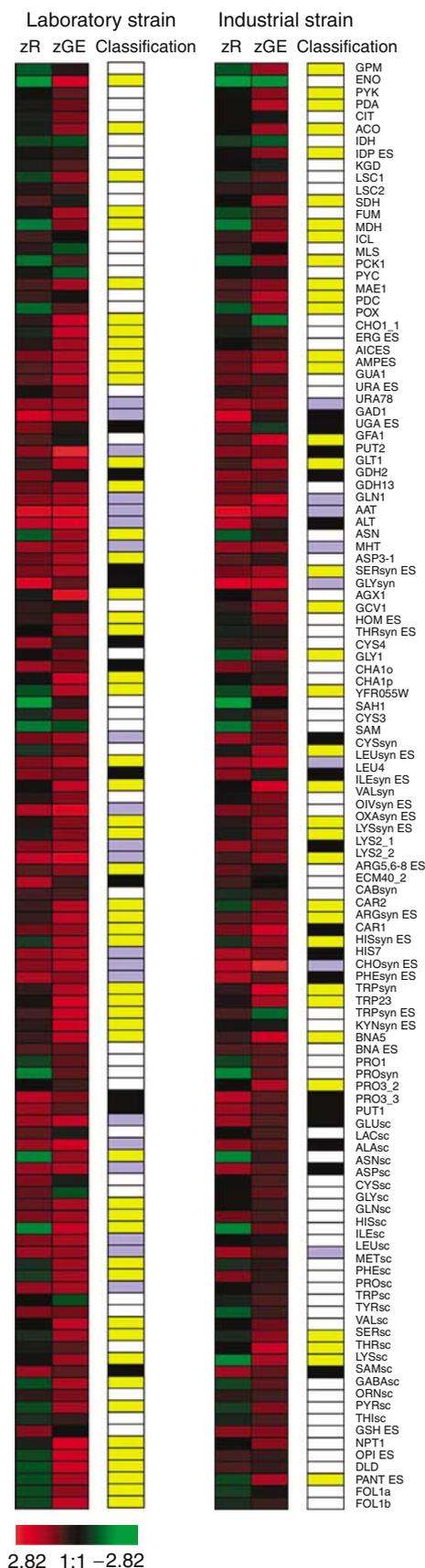

Figure 3 Magnitude of the regulation for the reactions of the metabolic network, ENZSUB3, at the hierarchical and metabolic levels for the effect of VHG fermentation media on laboratory (CEN.PK113-7D) and industrial (RS) strains. Z-scores calculated based on gene expression changes (zGE) and based on changes in the surrounding metabolites (zRE) are shown. Red means a positive Z-score and green means a negative Z-score indicating that the regulation is insignificant. Reactions were color-coded with respect to their Z-scores using $z=1.28$ ($P=0.10$) as the cutoff value to decide on the corresponding regulation type. Yellow denotes hierarchical regulation, black metabolical regulation, violet mixed regulation and white statistically insignificant score for both types.

mainly predominant for secretion reactions and amino-acid pathways with or without simultaneous hierarchical regulation, the sole exceptions being proline and methionine/cysteine pathways. It is logical to identify the latter as subjected to different regulation, as they are involved in pathways with sulfur assimilation and there were no direct perturbation on sulfur utilization in the experimental study. The type of regulation for a number of reactions differs between the two strains, which supports the finding that gene expression pattern can vary within different *S. cerevisiae* strains (Ferea *et al*, 1999; Brem *et al*, 2002; Townsend *et al*, 2003; Jansen *et al*, 2005). Ferea *et al* (1999) have reported altered expression levels of genes involved in metabolite transport for strains obtained by adaptive evolution in glucose-limited cultures. This observation presents an interesting analogy to our analysis, as the industrial strain is also likely to be a result of adaptive evolution. Similarly, different wild-type strains were found to have widespread variations in expression of genes involved in amino-acid metabolism (Townsend *et al*, 2003). In order to further validate that the metabolism is different in the industrial and laboratory strains, we performed principal component analysis of the metabolome data for the VHG medium data set (Supplementary Figure 3). This shows a clear distinction of the strains, indicating that the strains behave remarkably different at the level of metabolome. Our analysis systematically combines the transcriptome and metabolome and deduces the underlying regulation causing these differences in metabolism. Notably, following a change to a high-gravity fermentation medium, transcriptional regulation of metabolism is much more predominant in the laboratory strain as compared to the industrial strain, whereas the number of reporter reactions between two strains is around the same with a 70% overlap (Table II). This strongly suggests that although the industrial strain has a better adaptation of its transcriptional program for high-gravity media, there is still a metabolic regulation pattern that is similar to that of the laboratory strain. The difference in strains in terms of their response to the same perturbation is, again, very visible in the secretion reactions where laboratory strain attempts to regulate them also at the transcriptional level, whereas industrial strain relies predominantly on metabolic control (Figure 3 and Supplementary Table 3). The lesser degree of transcriptional regulation in the industrial strain could benefit the cells by reducing the investment of resources in transcriptional regulatory machinery.

Conclusions

In the present study, an integrative algorithm based on metabolome data was introduced for the identification of reporter reactions, defined as the reactions that are responding to a genetic or environmental perturbation through a coordinated variation in the levels of surrounding metabolites. We demonstrate that the algorithm functions, even with a small number of measured metabolites (84), which is a typical situation for several currently used technologies. Moreover, the method developed is suitable for mapping the entire alterations associated with a specific perturbation, depending on the advances in analytical detection techniques enabling the measurement of a larger number of metabolites.

Furthermore, when integrated with transcriptome data, our approach can be used to infer information about whether a reaction is metabolically or hierarchically regulated. Our analysis can therefore be regarded as a genome-scale approach towards the integration of different types of omics data by using metabolic networks as a scaffold in order to understand the architecture of metabolic regulatory circuits. Furthermore, our model-driven analysis is flexible and will further allow integration of other types of omics data, such as proteomics, and this will further refine the method presented herein to account for the genome-scale alterations in response to genetic as well as environmental perturbations, and hence allow genome-scale identification of regulation in the metabolism.

Materials and methods

Graph representation

In the present study, the metabolic network ENZSUB-3 was represented as a bipartite undirected graph to identify reporter reactions. Reactions and metabolites were both taken as nodes, and the edges denoted the interactions between them (Patil and Nielsen, 2005). Hence, the graph consisted of 317 nodes.

Different genetic and environmental perturbations associated with the two data sets (Devantier *et al*, 2005a; Villas-Bôas *et al*, 2005a) were analyzed. The graph representation was used to identify 'reporter reactions' for these perturbations. The algorithm used in the simulations is a modification of the algorithm recently developed by Patil and Nielsen (2005), which was based on the analysis of transcriptome data to identify so-called reporter metabolites, the spots in the metabolism with substantial transcriptional regulation. The modified algorithm herein has the capability of identifying reporter reactions, the putative key points in the metabolism in terms of metabolic regulation (Figure 1).

Significance test

The significance of change for the experimental metabolite levels between any two conditions was determined by comparing the levels with the aid of a statistical test, thereby quantifying the effect of the associated perturbation. For each of the perturbations, the statistical test was applied to the experimental data following the normalization process described by Villas-Bôas *et al* (2005a). Briefly, the normalization process is such that the within-group variances among replicates are reduced and between-group variances are maximized. The Mann–Whitney rank-sum U-test is a nonparametric statistical test, which has no *a priori* assumption about the distribution type of the data. It was preferred over the standard t-test, as the distribution of levels of some of the metabolites among the replicates, especially $NAD+$ and NADPH, was found to be skewed rather than normal distributed. The Student's t-test assumes normal distribution of the data and compares the mean values, whereas the U-test compares medians rather than means. Furthermore, median is a better measure for skewed distributions as it is less sensitive to the extreme scores that can be encountered in the replicates.

Strategy for the lack of data

As the utilized reporter reaction algorithm depends on the scoring of reactions based on the P-values of involved metabolites, the lack of P-values for the 94 metabolites that remain unmeasured in the final ENZSUB-3 model must be handled. Random assignment from GC–MS peaks was used to overcome the problem of the unavailable data. GC–MS spectra contain a large number of unknown peaks due to unmeasured metabolites. All the peaks in GC–MS spectra were deconvoluted for each replicate. The output was normalized by using a Python code, which minimizes the sample variability within the classes (Villas-Bôas *et al*, 2005a). Afterwards, the peaks in the spectra

within a selected time interval (0.15 min) were binned to account for the fluctuations in the retention times using a MATLAB algorithm. This has resulted in the overall detection of 236 unknown peaks for the first data set (Villas-Bôas *et al*, 2005a), with 116, 178 and 201 non-zero peak comparisons for genetic perturbations under aerobic and anaerobic conditions and environmental perturbations respectively, and 240 unknown peaks for the second data set (Devantier *et al*, 2005a) with 129 and 174 non-zero peak comparisons for the environmental perturbation of laboratory and industrial strains, respectively. The significance of change for these unknown peaks was quantified for each perturbation by means of *P*-values using the *U*-test. These *P*-values were randomly assigned to the unmeasured metabolites.

Reporter reaction analysis

Resultant *P*-values were converted to *Z*-scores using an inverse normal cumulative distribution function for further analysis. Each reaction in the constructed graph was scored by calculating the score of the sub-network formed by its k neighboring metabolites, and *z*-values of the metabolites were used in the scoring:

$$Z_{\text{reaction}} = \frac{1}{\sqrt{k}} \sum Z_{\text{metabolite}, k}$$

Z_{reaction} score was then corrected for background distribution using the mean (μ_k) and standard deviation (σ_k) of *Z*-scores of metabolite groups of the same size, obtained by random sampling from the same metabolic network

$$Z_{\text{corrected reaction}} = \frac{Z_{\text{reaction}} - \mu_k}{\sigma_k}$$

In order to minimize the sensitivity of reporter reactions to the randomly selected *P*-values for the non-measured metabolites as mentioned above, the reporter reaction algorithm was executed 1000 times by repeating the random assignment in each case. This repetition eliminated the effect of the *P*-values of the assigned peaks on results. For each reaction, the *Z*-scores in each repetition were averaged to get a final *Z*-score. Those reactions with the highest *Z*-scores (typically $Z > 1.28$, corresponding to $P < 0.10$) can be defined as reporter reactions for a system with complete metabolome data. As available experimental data were not complete, the calculated *Z*-scores were used for deducing the relative significance of the reactions in the analyzed perturbations. In other words, we mainly focus on comparative analysis of reactions among the studied perturbations, as revealed by Figure 2, rather than comparing a reaction to another based on its *Z*-score. The underlying reason is to avoid potentially incorrect conclusions due to the unmeasured metabolites, which have randomly assigned *P*-values. Additionally, the analyzed reactions have a high percentage of measured metabolite content, as indicated in Tables I and II. In the case of low coverage of measured metabolite content, this method should be followed with caution as the resultant *Z*-scores of reactions will become insignificant, and such reactions will not be picked up as reporters. However, in future, when analytical methods have been further improved, it is likely that more metabolites can be measured, and one will overcome this shortcoming and our approach may then be used to infer more solidly about the level of regulation at different parts of large metabolic networks. Based on the features of our algorithm, we suggest certain guidelines for the metabolome measurements in order to effectively exploit our approach: (i) measurement of metabolites that participate in many reactions (hubs in the metabolic network) will increase the performance of the algorithm and (ii) measurement of metabolites that participate in certain closely related pathways (metabolites that are closely placed in the network) will increase the confidence in the obtained *Z*-scores for reactions in those pathways (see Supplementary note 1 for further discussion).

Computational tools

METATOOL 4.3 (Pfeiffer *et al*, 1999) was used for the identification of enzyme subsets in the UNCOMP model. The codes written in MATLAB 7.0 (MathWorks Inc.) were utilized for the model preprocessing summarized above and to call the algorithm written in C++ for

reporter reaction identification. FBA was performed using in-house software BioOpt employing LINDO API for linear optimization. Deconvolution of peaks in GC–MS spectra for the identification of metabolites based on a metabolite library and for the random peak assignment was achieved using AMDIS software (Stein, 1999), and the peak normalization software was kindly provided by JF Moxley.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

References

Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21:** 692–696

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* **9:** 418–425

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

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296:** 752–755

Coleman ST, Fang TK, Rovinsky SA, Turano FJ, Moye-Rowley WS (2001) Expression of a glutamate dehydrogenase homologue is required for normal oxidative stress tolerance in *Saccharomyces verevisiae*. *J Biol Chem* **276:** 244–250

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429:** 92–96

Dandekar T, Moldenhauer F, Bulik S, Bertram H, Schuster S (2003) A method for classifying metabolites in topological pathway analyses based on minimization of pathway number. *BioSystems* **70:** 255–270

Devantier R, Scheithauer B, Villas-Bôas SG, Pedersen S, Olsson L (2005a) Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. *Biotechnol Bioeng* **90:** 703–714

Devantier R, Pedersen S, Olsson L (2005b) Transcription analysis of *S. cerevisiae* in VHG fermentation. *Ind Biotechnol* **1:** 51–63

dos Santos MM, Thygesen G, Kötter P, Olsson L, Nielsen J (2003) Aerobic physiology of redox-engineered *Saccharomyces cerevisiae* strains modified in the ammonium assimilation for increased NADPH availability. *FEMS Yeast Res* **4:** 59–68

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

Edwards JS, Ibarra RU, Palsson BO (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19:** 125–130

Erasmus DJ, van der Merwe GK, van Vuuren HJJ (2003) Genome-wide expression analyses: metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress. *FEMS Yeast Res* **4:** 375–399

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

Ferea TL, Botstein D, Brown PO, Rosenzweig RF (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Nat Acad Sci USA* **96:** 9721–9726

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* **5:** 1–6

Förster J, Famili I, Fu P, Palson BO, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* **13:** 244–253

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **22:** 245–252

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **101:** 10205–10210

Jansen MLA, Diderich JA, Mashego M, Hassane A, de Winde JH, Daran-Lapujade P, Pronk JT (2005) Prolonged selection in aerobic, glucose-limited chemostat cultures of *S. cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology* **151:** 1657–1669

Kauffman KJ, Prakash P, Edwards JE (2003) Advances in flux balance analysis. *Curr Opin Biotechnol* **14:** 491–496

McCammon MT, Epstein CB., Przybyla-Zawislak B, McAlister-Henn L, Butow RA (2003) Global transcription analysis of Krebs tricarboxylic acid cycle mutants reveals an alternating pattern of gene expression and effects on hypoxic and oxidative genes. *Mol Biol Cell* **14:** 958–972

Nielsen J (2003) It is all about metabolic fluxes. *J Bacteriol* **185:** 7031–7035

Nielsen J, Oliver S (2005) The next wave in metabolome analysis. *Trends Biotechnol* **23:** 544–546

Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J (2000) Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. *Metab Eng* **2:** 69–77

Patil KR, Nielsen J (2005) Uncovering transcriptional regulation of metabolism using metabolic network topology. *Proc Nat Acad Sci USA* **102:** 2685–2689

Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S (1999) Metatool: for studying metabolic networks. *Bioinformatics* **15:** 251–257

Piper MDW, Daran-Lapujade P, Bro C, Regenberg B, Knudsen S, Nielsen J, Pronk JT (2002) Reproducibility of oligonucleotide microarray transcriptome analyses. *J Biol Chem* **277:** 37001–37008

Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2:** 886–897

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* **19:** 45–50

Rossel S, van der Weijden CC, Kruckeberg AL, Bakker BM, Westerhoff HV (2005) Hierarchical and metabolic regulation of glucose influx in starved *Saccharomyces cerevisiae*. *FEMS Yeast Res* **5:** 611–619

Schuster S, Klamt S, Weckwerth W, Moldenhauer F, Pfeiffer T (2002) Use of network analysis of metabolic systems in bioengineering. *Bioproc Biosyst Eng* **24:** 363–372

Sheikh K, Forster J, Nielsen LK (2005) Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*. *Biotechnol Prog* **21:** 112–121

Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J Am Soc Mass Spectrom* **10:** 770–781

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62:** 817–836

ter Kuile BH, Westerhoff HV (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* **500:** 169–171

Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. *Mol Biol Evol* **20:** 955–963

Varma A, Palsson BO (1994) Metabolic flux balancing: basic concepts, scientific and pratctical use. *Bio/technology* **12:** 994–998

Villas-Bôas SG, Moxley JF, Åkesson M, Stephanopoulos G, Nielsen J (2005a) High-throughput metabolic state analysis: the missing link in integrated functional genomics. *Biochemical J* **388:** 669–677

Villas-Bôas SG, Åkesson M, Nielsen J (2005b) Biosynthesis of glyoxylate from glycine in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **5:** 703–709

Villas-Bôas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J (2005c) Mass spectrometry in metabolome analysis. *Mass Spectrom Rev* **24:** 613–646

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Nat Acad Sci USA* **101:** 7809–7814