

# Model-independent fluxome profiling from $^2\text{H}$ and $^{13}\text{C}$ experiments for metabolic variant discrimination

Nicola Zamboni and Uwe Sauer

Address: Institute of Biotechnology, ETH Zürich, CH-8093 Zürich, Switzerland.

Correspondence: Uwe Sauer. E-mail: [sauer@biotech.biol.ethz.ch](mailto:sauer@biotech.biol.ethz.ch)

Published: 16 November 2004

Genome **Biology** 2004, **5**:R99The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/12/R99>

Received: 28 August 2004

Revised: 18 October 2004

Accepted: 25 October 2004

© 2004 Zamboni and Sauer; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

We introduce a conceptually novel method for intracellular fluxome profiling from unsupervised statistical analysis of stable isotope labeling. Without *a priori* knowledge on the metabolic system, we identified characteristic flux fingerprints in 10 *Bacillus subtilis* mutants from 132  $^2\text{H}$  and  $^{13}\text{C}$  tracer experiments. Beyond variant discrimination, independent component analysis automatically mapped several fingerprints to their metabolic determinants. The approach is flexible and paves the way to large-scale fluxome profiling of any biological system and condition.

## Background

Genome-wide analyses of cellular mRNA, protein or metabolite complements have become workhorses in biological research that produce unprecedented amounts of data on cellular network composition. In contrast to such compositional information, molecular fluxes through intact metabolic networks link genes and proteins to higher-level functions that result from biochemical and regulatory interactions between the components [1]. As such, quantitative knowledge of *in vivo* molecular fluxes is highly relevant to functional genomics, metabolic engineering and systems biology [2,3]. Intracellular fluxes, or *in vivo* reaction rates, can be assessed by methods of metabolic flux analysis that are based on stable isotopic tracer experiments [4,5], which have successfully unraveled novel biochemical pathways [6,7] and gene functions [8,9]. The presently tedious and limited methodologies, however, hamper broader application to a large range of environmental conditions, isotopic tracers and higher biological systems [4].

We set out to overcome a principal bottleneck in metabolism-wide flux (fluxome [10]) analysis: the requirement for math-

ematical frameworks to interpret the isotopic tracer data from nuclear magnetic resonance (NMR) or mass spectrometric (MS) analyses within a detailed metabolic model [4,5]. Constructing such models requires *a priori* knowledge on possible distributions of the tracer used within the network, and, more importantly, extensive labeling and physiological data to resolve all fluxes within a given model. The lack of such structural knowledge on metabolic pathways and the technical difficulty of acquiring sufficient data hamper studies of metabolism, in particular in higher cells with complex nutrient requirements and for exotic tracer molecules. Hence, fluxome analysis is largely restricted to few  $^{13}\text{C}$ -labeled carbon sources in microbes or plants cultivated in minimal medium [7,11-16].

Here we discriminate mutants/conditions and assess their metabolic impact directly from 'raw' mass-isotope data by unsupervised multivariate statistics without *a priori* knowledge of the biochemical reaction network. To illustrate the applicability of this conceptually novel profiling method, we focused on the reactions of central metabolism in the model

Table 1

*B. subtilis* strains used

Strain	Description of deleted gene
Wild-type	168 <i>trpC2</i>
<i>pgi</i>	P-glucosomerase
<i>yqjI</i>	6-P-gluconate dehydrogenase
<i>sdhC</i>	Succinyl-CoA dehydrogenase component
<i>ytsJ</i>	Malic enzyme
<i>mdh</i>	Malate dehydrogenase
<i>pps</i>	PEP synthetase
<i>ccpA</i>	Main carbon catabolite repressor
<i>cggR</i>	Repressor of the <i>gapA</i> operon
<i>glcP</i>	Hexose/H <sup>+</sup> symporter
<i>glcR</i>	Repressor of PTS system

Strains were provided by S. Aymerich (INRA, CNRS, Thiverval-Grignon, France) and K. Kobayashi (Nara Institute of Science and Technology, Nara, Japan) [17].

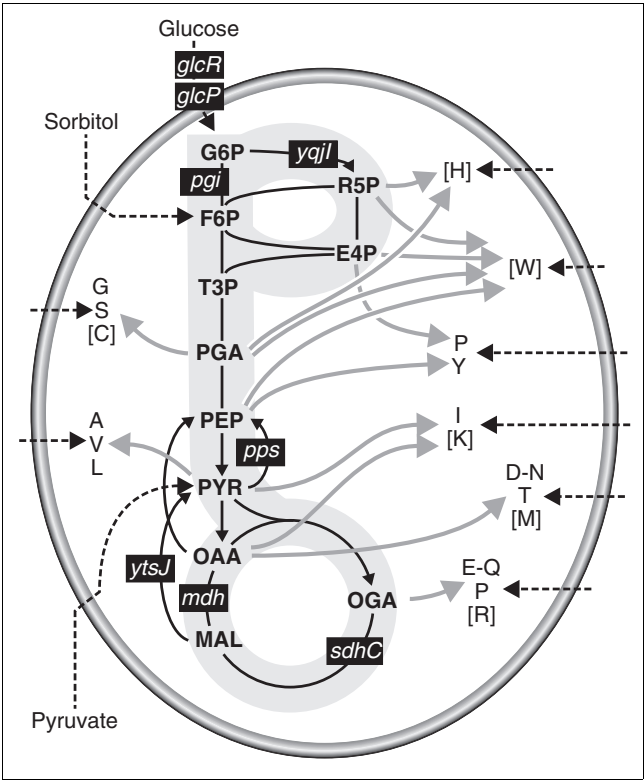
bacterium *Bacillus subtilis*, for which detailed flux data were available to validate the results [9,11,14].

Results

<sup>2</sup>H and <sup>13</sup>C tracer experiments

Environmental and genetic modifications were used to perturb intracellular metabolic activities in *B. subtilis*. In particular, we chose 10 knockout mutants [17] that were affected in metabolic genes or transcriptional regulators linked to central metabolism (Table 1 and Figure 1). These mutants were grown in 1-ml batch cultures [18] with six combinations of the carbon sources [U-<sup>13</sup>C] or [U-<sup>2</sup>H]glucose, [U-<sup>13</sup>C]sorbitol or [3-<sup>13</sup>C]pyruvate and the nitrogen sources ammonium or casein amino acids (CAA). As a proof of concept, we detected the isotopic labeling patterns in proteinogenic amino acids by gas chromatography MS (GC-MS), which provides direct access to several metabolic nodes in the network [6,7,19] (Figure 1). The raw mass isotope data of all mutants under each of the six experimental conditions are given in Additional data file 2.

In media supplemented with amino acids, cell protein was only partly synthesized from the isotopically labeled substrate. In such cases, current flux-analysis methods such as isotopomer balancing or flux ratio analysis are not applicable [4,5] because they do not account for variations in the labeling patterns due to amino-acid uptake and catabolism. Practically, we tackled here a worst-case scenario: growth in a medium enriched with unlabeled amino acids and profiling of the labeling pattern from tracers in the proteinogenic amino acids, which may potentially originate entirely from the medium. Nevertheless, a sufficiently high fraction of all analyzed amino acids was synthesized *de novo* from the labeled

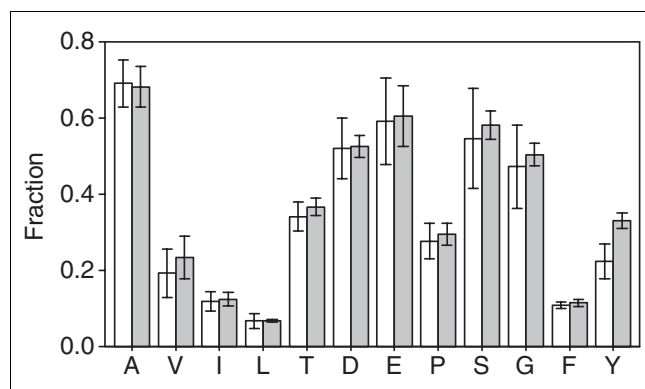


**Figure 1**  
Simplified biochemical reaction network of *Bacillus subtilis* central carbon metabolism. Gray arrows outline the biosynthesis of precursor amino acids that are indicated by their one-letter code. Amino acids in square brackets were not detected. Black dashed arrows illustrate the uptake of substrates. Black boxes highlight pathways or reactions that are affected in the mutants used (see also Table 1). G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; T3P, triose phosphate; PGA, phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; OAA, oxaloacetic acid; MAL, malic acid; OGA, 2-oxoglutarate.

substrates to obtain relevant MS signals, indicating that information on pathway activities was recorded in the labeling patterns (Figure 2). To capture the impact of genetic or environmental modifications, we analyzed the 260-330 raw mass isotope data points for each mutant and condition. This is essentially a table of mass-distribution vectors for all detected amino-acid fragments upon correction for naturally occurring stable isotopes, that is, the list of the relative frequencies of all possible isotope isomers for each detected analyte.

Identification of metabolic determinants for altered flux profiles

For the visualization of metabolic effects, the corrected MS signals of the wild type were subtracted from those of the mutants (Figures 3 and 4). Some mutations, such as *pps*, were silent under the conditions tested and exhibited only noise in the wild-type-normalized data. In other mutants, characteristic profiles of strongly affected amino acids were readily apparent. One example was the almost identical signature of

**Figure 2**

Fraction of amino acids that were synthesized *de novo* from [U-<sup>13</sup>C]glucose (white bars) and sorbitol (gray bars) in batch experiments supplemented with 0.5 g/l casein hydrolysate. Amino acids are given in the one-letter code.

serine (S) fragments in the profiles of the *glcP* and *cggR* mutants during growth on sorbitol with CAA; that is, high fractions of masses  $m_0$  and  $m_3$  and low fractions of  $m_1$  and  $m_2$  (where the subscripts denote the number of <sup>13</sup>C atoms in each amino-acid fragment). While the S signature of the *mdh* mutant on sorbitol with CAA was also distinct, it was different from that in the above two mutants with low  $m_1$ ,  $m_2$ , and  $m_3$  fractions (Figure 3). These characteristic labeling profiles are biochemically very informative and may be linked to precise metabolic causes. For the above examples, the high fraction of uncleaved serine molecules with intact C<sub>3</sub> backbones (that is,  $m_0$  and  $m_3$ ) in *glcP* and *cggR* is evidence of a lower exchange with the glycine pool, when compared with the wild type [19,20]. In the *mdh* mutant, the high fraction of uncleaved but unlabeled S ( $m_0$ ) reveals high incorporation of unlabeled serine from the CAA supplement, and thus low *de novo* bio-synthesis from <sup>13</sup>C-labeled sorbitol.

As well as consistency with the data in the literature, the analysis also revealed new information on pathway activity and regulation that was not previously accessible. One example is the pronounced signatures of the *sdhC* mutant on glucose and sorbitol. Because the *sdhC* mutation disrupts the tricarboxylic acid (TCA) cycle, the wild-type flux through the cycle must be similar on these substrates, both with and without CAA (Figure 3). The *sdhC* signatures of the TCA cycle-derived amino acids aspartate (D) and glutamate (E) were also present in the CAA profiles of the other TCA cycle mutant *mdh*. Their absence on ammonium indicates activity of the malic enzyme-based pyruvate bypass [11] in the *mdh* mutant.

While such a level of detailed biochemical insight is possible, it requires considerable expertise and time to retrieve. Alternatively, metabolic impacts in new mutants can be identified by comparison of the mass fingerprints in mutants with known metabolic lesions. During growth on sorbitol and pyruvate in minimal media but not with CAA, the *CggR*

repressor of the glycolytic *gapA* operon, for example, appears to affect TCA cycle fluxes because the mutant profile matches those of the TCA cycle mutants *sdhC* and *mdh* (Figure 3). In contrast to glucose, sorbitol does not elicit catabolite repression; hence, comparison of sorbitol and glucose profiles can identify repression-dependent effects. Examples are the signatures of the oxaloacetate-derived amino acids isoleucine (I), threonine (T) and aspartate in the *cggR* profile that reveal, by the similarity to the *sdhC* and *mdh* mutants, a TCA cycle flux-promoting effect of *CggR* on sorbitol but not on glucose. This is consistent with the repression of *cggR* on glucose [21], and the TCA cycle effect is probably indirect, through the repression of glycolytic genes [22].

A significant extension beyond the canonical <sup>13</sup>C-tracer methods is the applicability to any isotope, which broadens the observable metabolic processes. Here we used fully deuterated [U-<sup>2</sup>H]glucose that allows us to monitor dehydrogenase activities and water release. The <sup>2</sup>H-label was present exclusively in the variable side chains, because the α-carbon hydrogen was lost in the transaminase reaction. Thus, glycine contains no label and the acidic aspartate and glutamate lose the label proximal to the carboxyl group as a result of exchange with water at the low pH during hydrolysis. The remaining amino acids provided a stable and informative <sup>2</sup>H-pattern (see Additional data file 1). An illustrative example is the *cggR* mutant signatures for the pyruvate-derived amino acids valine (V), leucine (L) and, partially, alanine (A) (Figure 3). In all three cases, reduced  $m_2$  and increased  $m_0$  fractions revealed a double loss of <sup>2</sup>H-label in their common precursor pyruvate at position C-3. This loss of <sup>2</sup>H indicates increased exchange of <sup>2</sup>H with water at the C-3 position of pyruvate (or any upstream triose), which is fully consistent with increased transcription of the glycolytic enolase in the *cggR* mutant on glucose [23] that could catalyze this exchange. As the enolase activity does not affect the carbon backbone, the corresponding patterns cannot be identified in <sup>13</sup>C experiments

### Independent component analysis (ICA)

For large-scale profiling studies, automated mutant classification based on metabolic function without user supervision would be desirable. Initially, we used principal component analysis (PCA), which is often used for graphical representation of multidimensional variables from profiling experiments [24,25], as was recently described for pretreated (summed fractional labels) mass isotope data [26]. From the raw mass isotope data, the first two PCs discriminated, under most conditions, mutants with extreme labeling patterns (see Additional data file 1). The differences become smaller with increasing PCs, and only the initial three to four PCs allowed reliable discrimination. In the present data, PCA tended to discriminate extreme singular labeling patterns in few fragments or, more frequently, combinations of altered patterns in the fragments of many amino acids, as was expected from the variance maximization of PCA. Unfortunately, the

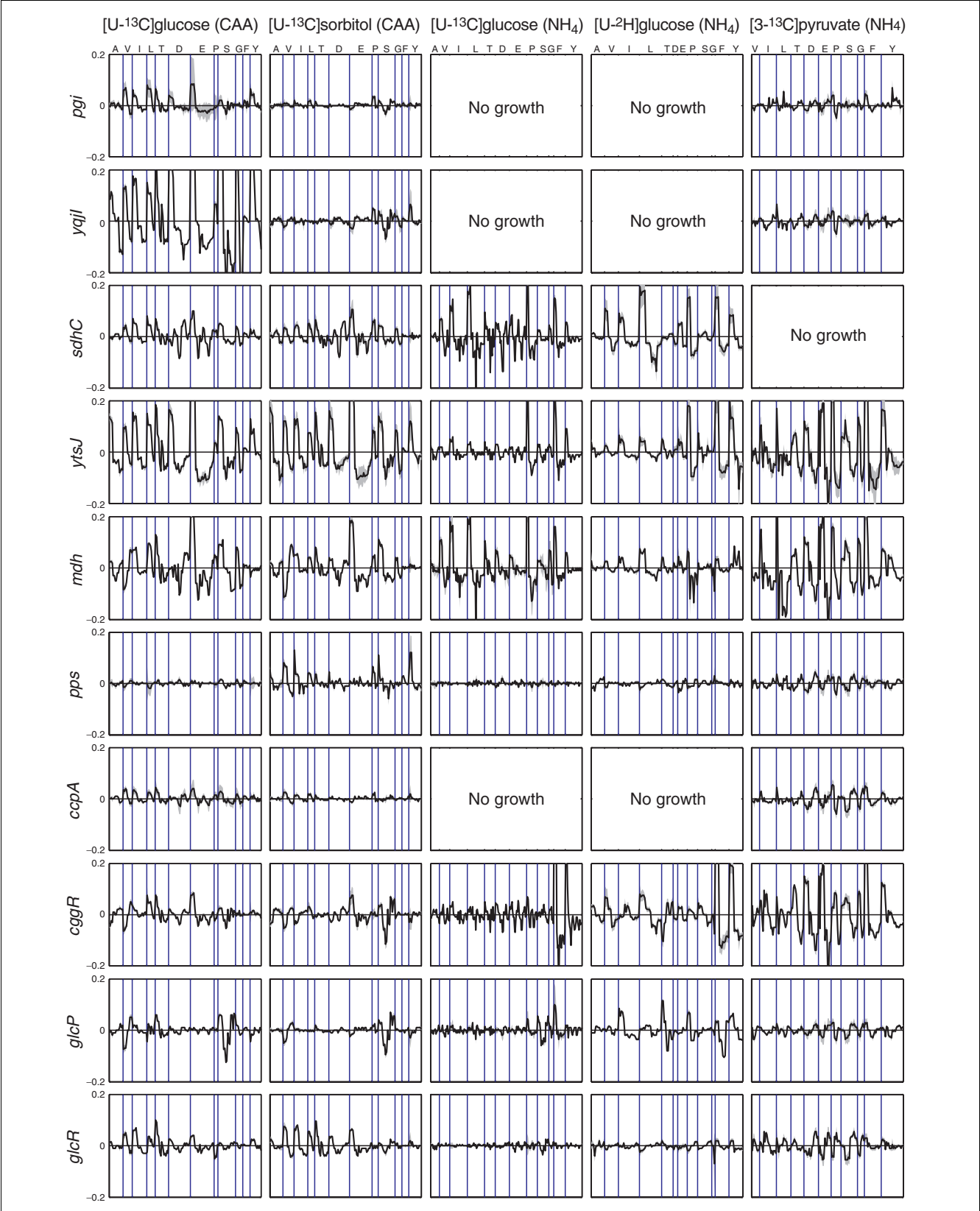


Figure 3 (see legend on next page)

**Figure 3** (see previous page)

Comparison of labeling profiles in amino acids of *B. subtilis* mutants that were normalized by subtraction with the wild-type values obtained under the same condition, as obtained from five different medium compositions. The line deviates above (or below) the null line when an amino acid (represented by their one letter code at the top of the first panel) mass is more (or less) abundant in the mutant than in the parent. For each amino acid, the available data points are in the order of their total mass fragment. Gray areas represent the deviation of the normalized values, based on duplicate analyses of mutant and wild type. To reduce the dimension of the data for visual comparison, we excluded those values that, on average, accounted for less than 5% of the fragment pool in all mutants under a given condition.

resulting complex PCs are difficult to interpret metabolically, and thus are of limited biochemical relevance.

Consequently we used independent component analysis (ICA) for unsupervised, automatic recognition of conserved labeling patterns that are biochemically relevant. The underlying assumption is that these patterns result from the superposition of independent metabolic activities. Each activity causes a specific shift in the mass distributions of one or more intermediates. ICA seeks to separate the observed variables into non-gaussian components that are statistically as independent as possible [27]. Generally, ICA clearly discriminated mutants and conditions from the corrected (non-normalized) MS data (see Additional data file 1). While the weights in PCs were more broadly distributed among the input variables, ICs were dominated by fewer, sharper peaks (Figure 4).

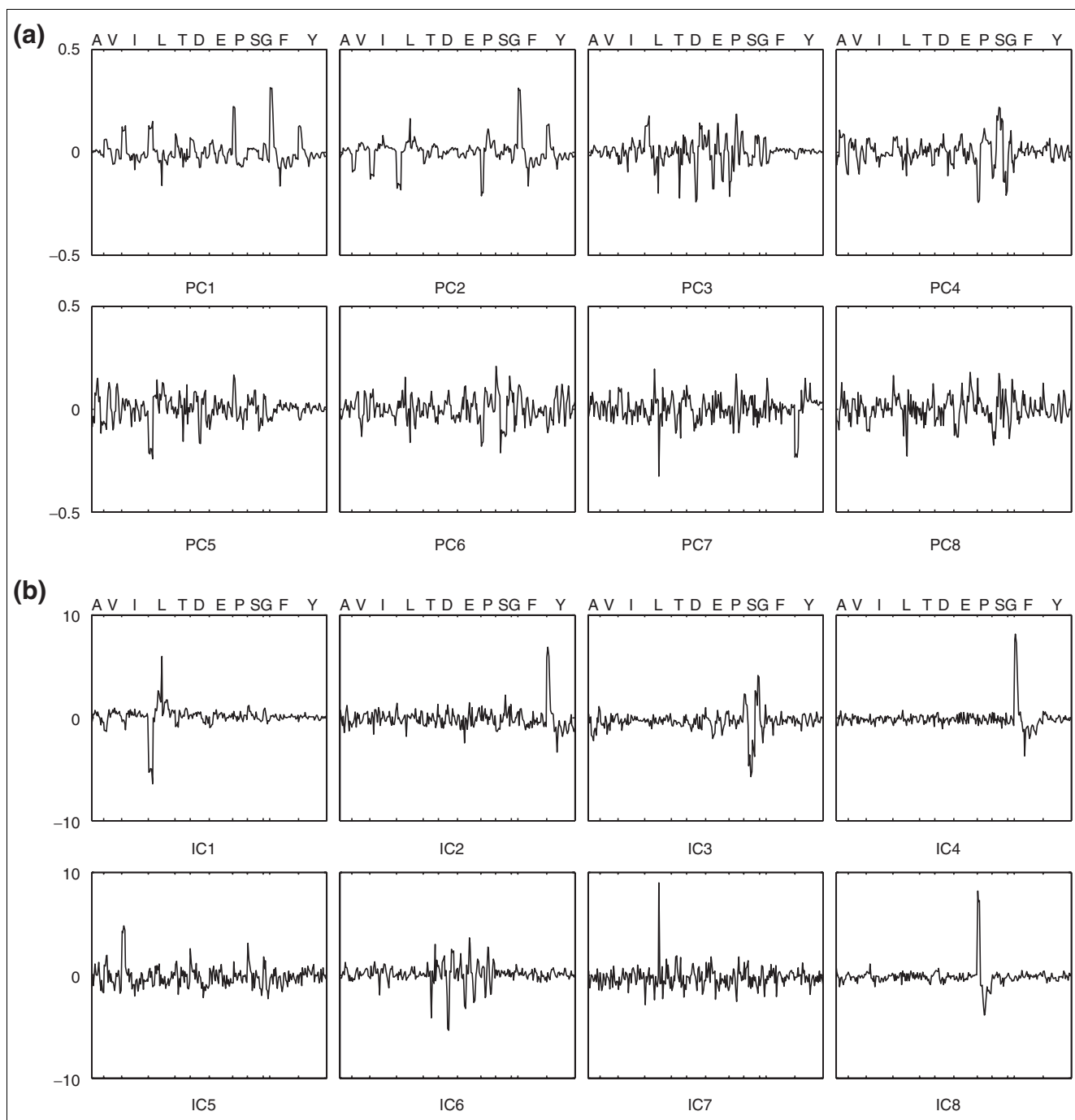
For the particular example of the [U-<sup>13</sup>C]sorbitol with ammonium experiment, we explored the ICA results in more detail (Figure 5). The first, striking, observation was that the second IC contains the biochemically redundant signals of  $m_2$  T,  $m_2$  D, and  $m_1$  and  $m_3$  E (highlighted in red in Figure 5a) that arise from acetyl-CoA units in the TCA cycle [19]. This shows that ICA automatically provides insights into the biosynthetic linkage between amino acids with a resolution that eclipses visual comparison of the normalized signatures. For amino acids, this information was of course previously available, but statistical identification of biochemical relations could potentially also be obtained for less well-characterized compounds. Second, ICA often clustered biosynthetically related signals in the same component (Figure 5): IC7 grouped the similar signatures of phenylalanine (F) and tyrosine (Y) together; IC1 reports labeling shifts in glycine (G) and partially serine; and IC4 concentrated high weights in signals of the pyruvate derivatives alanine, valine and leucine (highlighted in blue in Figure 5). While isoleucine is also synthesized from pyruvate, it had only a marginal weight in IC4 because of interference from its second precursor oxaloacetate. Third, specific signatures of proline (P), leucine and serine are clearly recognized in IC3, IC8 (highlighted in green in Figure 5a), and IC10, respectively. These signatures reflect those previously identified in the normalized profiles (Figures 3 and 5c). Among the remaining components, IC5 and IC6 emphasize outliers in the *cggR* and *lysJ* MS data, respectively, whereas the noisy IC9 profile indicates that the identified ICs in our small data-set approach a limit.

Akin to PCA, ICA allowed us to discriminate mutants from the corrected MS data (Figure 5b and Additional data file 1). On sorbitol, mutants such as *pgi*, *yqiI*, *pps*, *glcP* and *glcR* were mostly silent, and typically projected in proximity to the parent strain. In contrast to PCA, ICs classified the mutants on the basis of specific metabolic effects. In some cases (IC2 or IC4 in Figure 5b), the IC defined well-separated clusters of mutants, usually two groups, reflecting a binary (on-off) effect. In the majority of the components, however, the even distribution between the extremes reveals progressive metabolic responses (for example, IC3, IC7 or IC10). Overall, the ICs correlated favorably with the signatures of wild-type-normalized profiles (Figure 5 and Additional data file 1). Thus, ICA clearly outperformed PCA by its capacity for unsupervised recognition of metabolic responses and its ability to correlate biochemically redundant information in the data.

### Comparison of PCA and ICA with analytically determined flux ratios

For most experimental conditions tested, mathematical frameworks for numerical flux analysis such as isotopomer balancing or flux-ratio analysis [4,5] were not available. Only the [U-<sup>13</sup>C]glucose minimal medium experiments allowed a direct comparison of fluxome profiles with flux ratios. Therefore, we examined whether any of the statistically identified PCs and ICs was linearly correlated with eight analytically determined flux ratios [9,19] that were obtained from the same MS data (Figure 6). For PCs, the correlation coefficients decreased with increasing component number, and singular correlations could not be detected between individual PC-flux ratio pairs. Generally, the ICs were much better correlated with the flux ratios, for particular pairs with coefficients close to 0.90. This indicates that the identified ICs define signatures in the mass distribution of the analytes that bear high metabolic relevance, similarly to analytically derived flux ratios.

Notably, IC6 was almost perfectly correlated with the flux ratio of oxaloacetate derived through the TCA cycle (Figure 6). This IC contained high weights in TCA-cycle-derived amino acids signals that are linked to the incorporation of C<sub>2</sub> units from acetyl-CoA (Figure 4). As shown above, the projection of a data point on the axis defined by a component reflects the presence of the fluxome signature in its labeling patterns, and hence directly quantifies the occurrence of a particular metabolic activity. When plotting the projection

**Figure 4**

Weights of input variables. Weights of input variables in the first eight components obtained by **(a)** PCA and **(b)** ICA from the corrected MS data of the [U-<sup>13</sup>C]glucose experiment with ammonium.

versus the numerical values, the IC6-derived data exhibited a highly linear correlation, while the correlation coefficient was almost halved for PC3, the closest relative to IC6 (Figure 7). This confirms numerically the enhanced capacity of ICA to capture essential and independent information for a complex metabolic trait such as the TCA cycle activity. The extraordi-

narily high correlation coefficient of 0.99 demonstrates that IC6 represents very closely the analytically deduced TCA-cycle flux ratio. This is surprising because IC6 was statistically identified from 265 masses, whereas the flux ratio was calculated on the basis of a large body of biochemical background information [19,20].

## Discussion

For the example of central and amino-acid metabolism in *B. subtilis*, we show that fluxome profiling by multivariate statistics from mass isotopomer distribution analysis is meaningful for the discrimination of mutants or conditions on the basis of their metabolic behavior, and applicable to conditions that are inaccessible to previous flux analysis. In sharp contrast to metabolome concentration data [24,25], fluxome profiles contain functional information on the operation of fully assembled networks [1,4]. As shown here by ICA, this approach enables us to distill the essential signatures of independent metabolic activities, and supports the identification of the underlying biochemical causality. Because no model or *a priori* knowledge on the investigated system is required, the metabolic imprints of any tracer atom and molecule can be followed in virtually any biological system, including multicellular organisms in complex multisubstrate media.

Similarly, *a priori* knowledge of the number of ICs to be computed is not a prerequisite. As a matter of fact, the optimal number depends primarily on the labeling patterns and can hardly be estimated from the dataset dimensions. An underestimate will generally leave some relevant signatures unrecognized, whereas an overestimate will lead to an increased fraction of components reflecting measurement or biological noise. Although statistical significance can be assessed with duplicates, this becomes prohibitive with large datasets (that is, hundreds of mutants or analytes) or reduced availability of replicas. The bottleneck resides in the stochastic approach of most ICA algorithms, for which independent runs result in different ICs or ordering thereof. Instead, algorithmic and statistical reliability of the ICs can be evaluated by repeating the estimation several times either with randomly chosen initial guesses or by slightly varying the dataset (bootstrapping [28]), respectively, and then clustering all results to identify robust ICs [29].

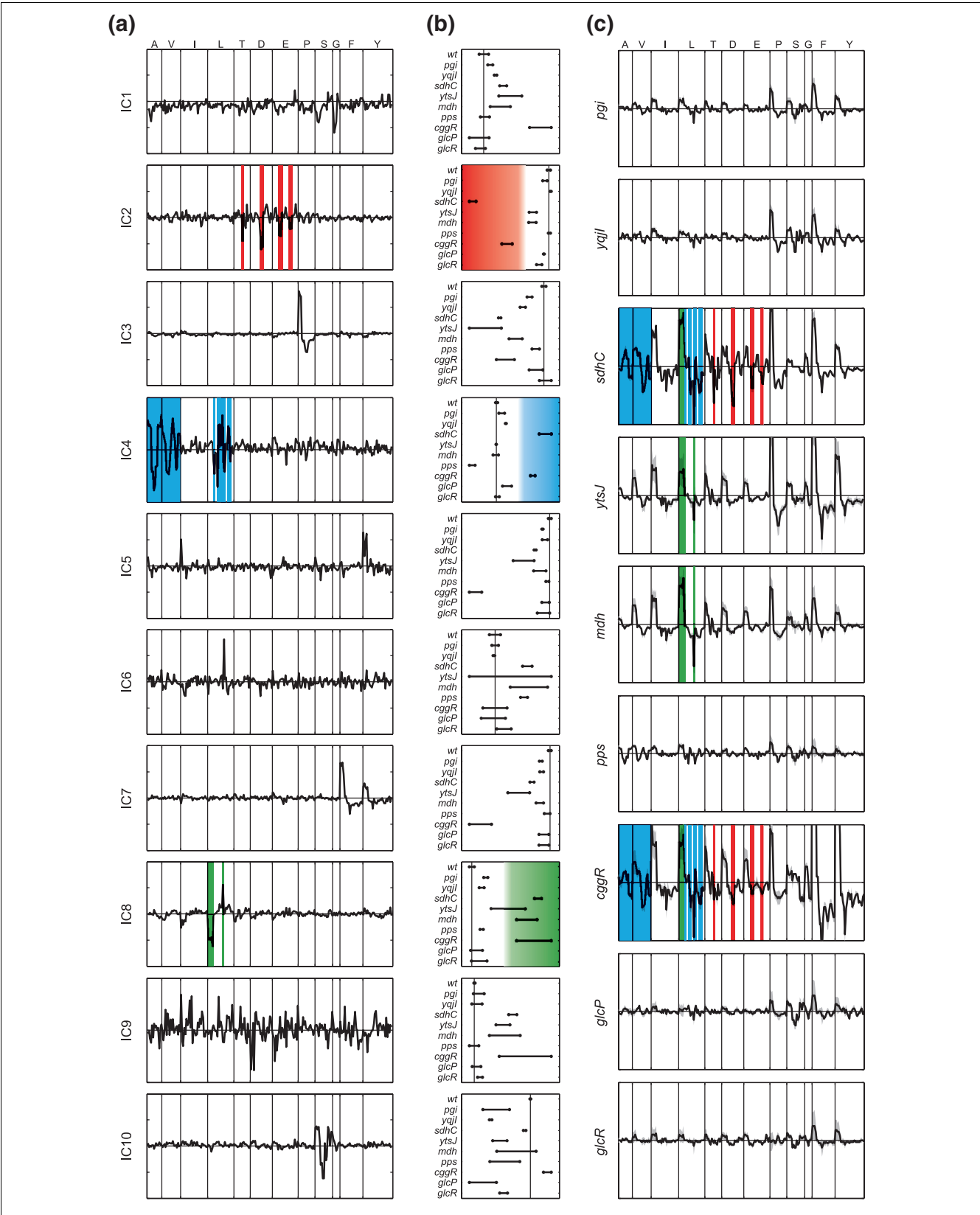
Two factors directly affect the results that can be obtained by comparative fluxome profiling: the detected analytes and the choice of isotopic tracer. As well as polymer-based analytes such as the proteinogenic amino acids monitored here, fluxome profiles can be detected in any set of intra- or extracellu-

lar metabolites, thereby widening the observable metabolic processes. The choice of tracer depends, to some extent, on the metabolic subsystem of interest. Uniformly labeled substrates provide a more global perspective because they allow assessment of the scrambling of any carbon backbone and, in the case of experiments performed in rich media, also allow quantification of the fraction of *de novo* biosynthesis from the tracer relative to the uptake of a medium component. Similarly, uniformly deuterated substrates or  $^2\text{H}_2\text{O}$  are valuable for simultaneously capturing a wide number of ICs that are affected by the release, binding and exchange of water or protons. Substrates that are labeled at specific positions, in contrast, enable deeper interrogation of particular sub-networks, for example,  $[1-^{13}\text{C}]$ hexoses for the initial catabolic reactions [8,19] or  $[1-^{13}\text{C}]$ aspartate to assess urea cycle activity.

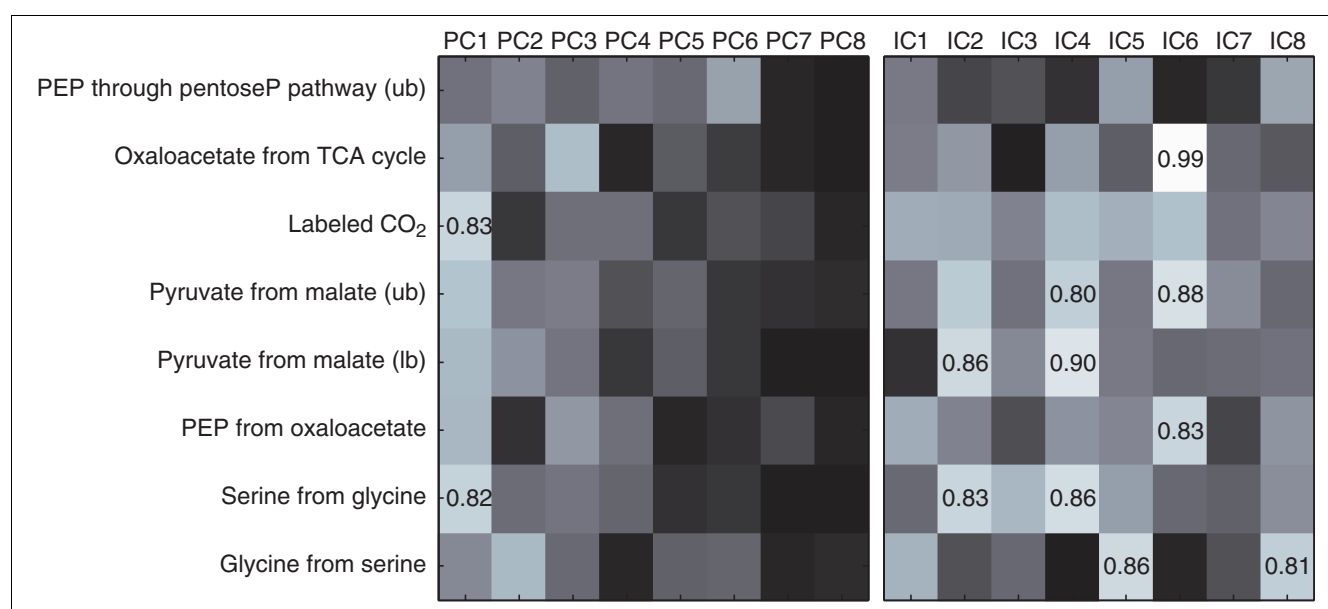
The results also revealed new biological information on pathway activity, function or regulation. First, both glycolysis and the pentose phosphate pathway actively catabolized glucose in the presence of CAA, because the *pgi* and *yqiI* mutant signatures were different from the wild type and from each other. On sorbitol, in contrast, the same mutants were very similar to the wild type, suggesting that both reactions are only marginally involved in catabolism of this sugar. Second, the Krebs cycle flux was similar on glucose and sorbitol (with and without CAA), as deduced from the similarly pronounced signatures of the *sdhC* mutant. Third, absence of the *sdhC* signatures in the Krebs cycle-derived amino acids aspartate and glutamate of the *mdh* mutant when grown with ammonium (but not CAA) indicates activity of the malic enzyme-based pyruvate bypass [30]. Fourth, activity of the NADP-dependent malic enzyme appears to be independent of catabolite repression because pronounced signatures of the *ytsJ* mutant were seen on all substrates. The gluconeogenic phosphoenolpyruvate synthetase Pps, in contrast, was inactive in the presence of the repressing glucose but active on pyruvate or sorbitol. Fifth, as discussed above the data reveal a Krebs cycle-promoting effect of the repressor CggR on sorbitol but not on glucose, most likely through the repression of glycolytic genes [22].

**Figure 5** (see following page)

Fluxome profiling by independent component analysis of *B. subtilis* mutants grown on a 50:50 mixture of  $[\text{U}-^{13}\text{C}]$ - and naturally labeled sorbitol with ammonium. **(a)** Weights of input variables (amino-acid mass-distribution vectors) in the mixing matrix of 10 ICs. **(b)** Projections (on x-axis) of samples on the IC shown in (a). The vertical line is drawn to intersect the average of the wild-type values. **(c)** Wild-type-normalized labeling profiles. Colors are used to highlight those aspects of the amino-acid profiles that were identified by ICA as relevant for the discrimination of the samples (b) along selected components.



**Figure 5** (see legend on previous page)

**Figure 6**

Correlation between analytically derived metabolic flux ratios (on the y-axis) [19] and the projections of the data on the first eight components obtained by PCA and ICA for the [U-<sup>13</sup>C]glucose experiment with ammonium. The brightness reflects the correlation coefficient, with black and white corresponding to values of 0 and 1, respectively. For coefficients higher than 0.8, the numerical value is reported. ub, upper bound; lb, lower bound.

The comparative fluxome profiling presented here complements traditional flux analysis because it enables potentially rapid and automated identification of relevant mutants or conditions from large-scale datasets, for example from entire mutant libraries. The approach is quantitative in terms of the relative difference between variants, but qualitative with respect to the *in vivo* flux. Interesting variants are then subjected to deeper interrogation of the specific metabolic phenomenon identified. Besides mere data mining, fluxome profiling also has the potential to identify complex functional traits in higher cells where current flux methods fail, and possibly even identify the underlying biochemical mechanism of discriminant mass isotope signatures.

## Materials and methods

### Strains and growth conditions

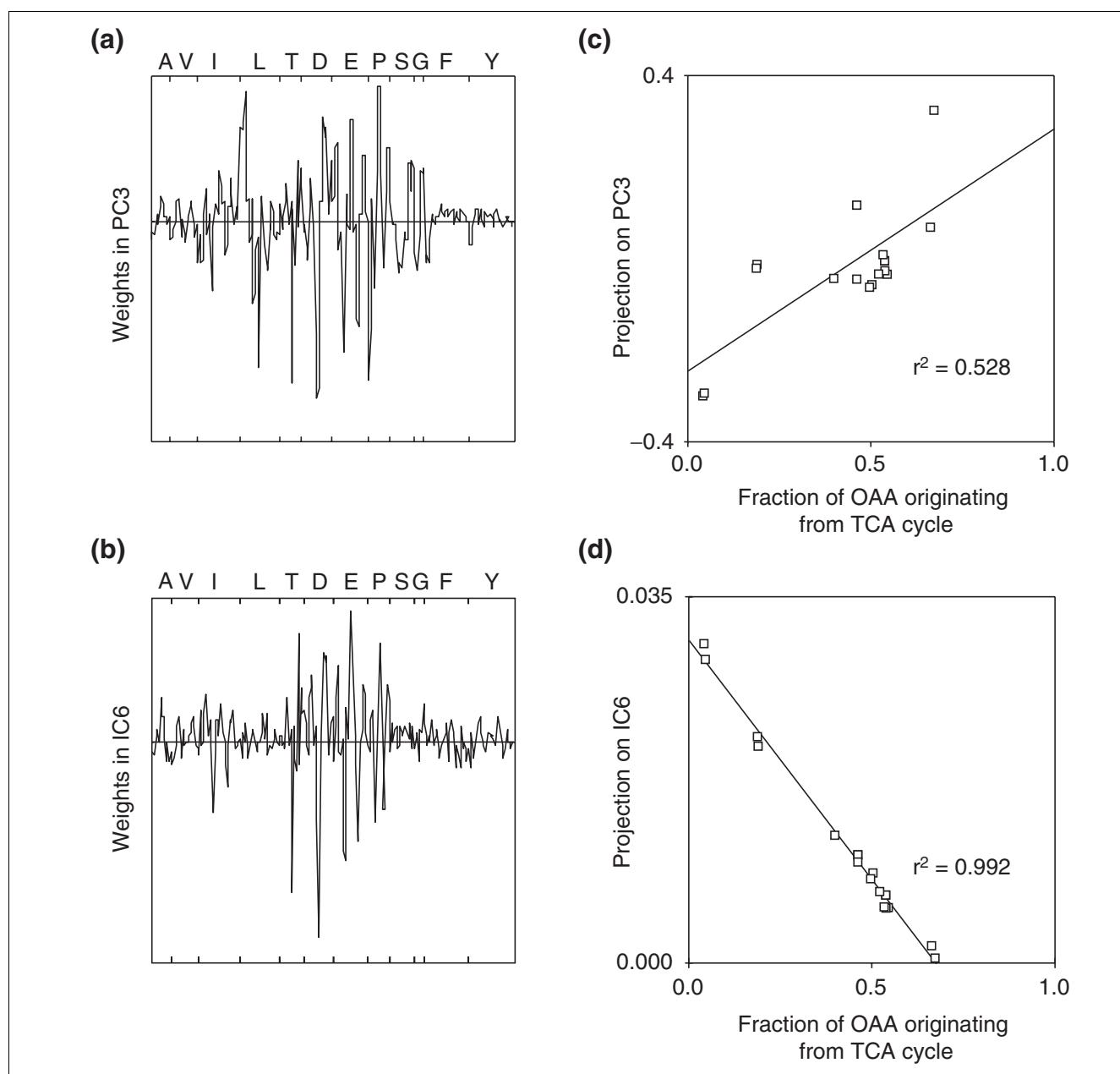
Wild-type *B. subtilis* 168 (*trpC2*) [31] and knockout mutants containing an antibiotic marker in single genes [17] were grown in M9 minimal medium [9] at pH 7.0 with 50 mg tryptophan. Six different combinations of <sup>2</sup>H- or <sup>13</sup>C-labeled isotopic tracers (3 g/l) and nitrogen sources were used: (i + ii) uniformly <sup>13</sup>C-labeled [U-<sup>13</sup>C]glucose with either 0.5 g/l CAA (Sigma) or 1 g/l NH<sub>4</sub>Cl; (iii + iv) [U-<sup>13</sup>C]sorbitol with either 0.5 g/l CAA or 1 g/l NH<sub>4</sub>Cl; (v) [U-<sup>2</sup>H]glucose ([1,2,3,4,5,6,6-<sup>2</sup>H]glucose) with 1 g/l NH<sub>4</sub>Cl; and (vi) [3-<sup>13</sup>C]pyruvate with 1 g/l NH<sub>4</sub>Cl and twofold higher concentrations of phosphate to ensure pH buffering. [U-<sup>13</sup>C]glucose (Martek Biosciences), [U-<sup>13</sup>C]sorbitol (Omicron Biochemicals), and [1,2,3,4,5,6,6-<sup>2</sup>H]glucose (Euriso-Top) were supplemented as 50:50 mix-

tures of labeled and unlabeled isotopomers. Pyruvate was supplied entirely as the [3-<sup>13</sup>C] isotopomer (Euriso-Top).

Aerobic batch cultures were grown in silicone-covered, deep-well microtiter plates at 37°C and 300 rpm in a 5-cm orbital shaker [18]. Frozen stocks were used to inoculate 1 ml LB medium with selective antibiotics. After 10 h of incubation, 10 μl were used to inoculate 1 ml M9 medium with 5 g/l glucose and selective antibiotics, incubated for 12 h, and 10 μl of these precultures were used to inoculate 1.2 ml of M9 medium with isotopic tracers. Cultures were harvested upon entry into stationary phase (assessed by visual evaluation). Because the length of batch growth varied, cultures with CAA, with NH<sub>4</sub>Cl, and with pyruvate were harvested after 10, 14 and 24 h, respectively. Labeling patterns in the analyzed proteinogenic amino acids are rather stable [10,19]; hence differences of a few hours in growth phase at harvest were irrelevant. This was also confirmed in separate (data not shown) and duplicate experiments for each combination of strain and medium that was independently started from culture stocks.

### GC-MS analysis and data preprocessing

Cell harvest, protein hydrolysis and GC-MS analysis of amino acids were done exactly as described before [19,32]. Amino acid mass distributions were derived from the spectra after correction for the natural abundance of stable isotopes [19]. Since amino acids are fragmented during electron impact ionization in the MS, we obtained three to five fragments with partially redundant information for each amino acid. For each fragment, a normalized vector  $m_0, m_1, \dots, m_n$ , expresses

**Figure 7**

Weights of input variables in the component that is linked to TCA cycle activity, identified by either (a) PCA or (b) ICA from the [U- $^{13}\text{C}$ ]glucose experiment with ammonium. In (c) and (d), the projections of the mutant data on the component shown in (a) and (b), respectively, were plotted versus the analytically derived fraction of oxaloacetate (OAA) originating from TCA cycle [19]. The correlation coefficients are for linear fits.

the fraction of molecules that are labeled at 0,1, ...,  $n$  positions, depending on the total number  $n$  of carbon or hydrogen atoms present. Considering all corrected fragment vectors obtained per sample, a complete dataset typically consisted of about 260 and 330 single mass values from  $^{13}\text{C}$  and  $^2\text{H}$  experiments, respectively, depending on the quality of the MS measurement.

### Multivariate data analysis

To obtain a new representation of the multivariate MS data and to make their essential structure accessible, we applied PCA to the corrected fragment vectors. This approach projects the input variables in an orthogonal space that is spanned by the PCs. Among the infinite number of possibilities, each successive PC is selected to maximize the variance of the projected data and to be orthonormal to the previous ones [33]. Consequently, PCA concentrates the maximum

and nonredundant information of the entire dataset in the minimal number of dimensions, and thus is best suited for data compression [27]. The computation was performed with Matlab (The Mathworks) using the *princomp* function of the Statistics toolbox 4.0. No input vectors were eliminated from the dataset to filter outliers in PCA, because this operation affected only PCs with higher order but only marginally PC1 and PC2.

To reveal hidden information in the labeling patterns, the corrected MS vectors were subjected to ICA [27], which is frequently used in the neurosciences [34,35] and in gene-expression studies [36,37]. For ICA, we assume that independent metabolic processes such as reactions or pathways produce characteristic fingerprints in the labeling pattern. These metabolic fingerprints are defined by  $m$  fundamental components  $S = (s_1, \dots, s_m)^T$ , each of which is represented by a vector of  $p$  MS-signals. We assumed that the experimental data  $X = (x_1, \dots, x_n)^T$ , with  $n$  vectors of  $p$  corrected MS signals for each mutant/condition, result from a linear combination of the  $m$  fundamental processes, given by  $x_i = a_{i1}s_1 + \dots + a_{im}s_m$ . In matrix notation, this leads to  $X_{p \times n} = A_{p \times m} S_{m \times n}$ , with  $A$  as the mixing or loading matrix. ICA seeks to estimate the unknown terms  $A$  and  $S$  from the observed values  $X$  but has different objectives from PCA. Briefly, ICA identifies statistically ICs by selecting those with maximum non-gaussianity [27]. Hence, ICs are nonlinearly decorrelated and assumed to have non-gaussian distributions. Because of the central limit theorem, which states that the sum of non-gaussian random variables is closer to gaussianity than the original ones, ICs are identified by selecting the linear combinations of the observed variables that have maximum non-gaussianity [27]. In particular, we used the publicly available FastICA 2.1 algorithm [38] to estimate the number of components that were equal to the number of strains in the dataset, excluding duplicates. The data dimension was not reduced (by PCA) before IC computation.

## Additional data files

The following additional data is available with the online version of this paper. Additional data file 1 contains three figures (Additional Figure 1 shows the mass distribution in the  $^2\text{H}$  experiment; Additional Figure 2 shows mutant discrimination by PCA (less relevant than by ICA); Additional Figure 3 is a complete representation of the 660 ICs (10 ICs in 6 experiments for 11 strains). All the raw data is contained in six Excel tables in Additional data file 2.

## References

- Hellerstein MK: **In vivo measurement of fluxes through metabolic pathways: the missing link in functional genomics and pharmaceutical research.** *Annu Rev Nutr* 2003, **23**:379-402.
- Bailey JE: **Lessons from metabolic engineering for functional genomics and drug discovery.** *Nat Biotechnol* 1999, **17**:616-618.
- Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Metabolic pathways in the post-genome era.** *Trends Biochem Sci* 2003, **28**:250-258.
- Sauer U: **High-throughput phenomics: experimental methods for mapping fluxomes.** *Curr Opin Biotechnol* 2004, **15**:58-63.
- Wiechert W:  **$^{13}\text{C}$  metabolic flux analysis.** *Metab Eng* 2001, **3**:195-206.
- Fischer E, Sauer U: **A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*.** *J Biol Chem* 2003, **278**:46446-46451.
- Gunnarsson N, Mortensen UH, Sosio M, Nielsen J: **Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species.** *Mol Microbiol* 2004, **52**:895-902.
- Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E: **The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*.** *J Biol Chem* 2004, **279**:6613-6619.
- Zamboni N, Fischer E, Laudert D, Aymerich S, Hohmann HP, Sauer U: **The *Bacillus subtilis* *yqjI* gene is the major 6-P gluconate dehydrogenase in the pentose phosphate pathway.** *J Bacteriol* 2004, **186**:4528-4534.
- Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, Wüthrich K, Bailey JE: **Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism.** *J Bacteriol* 1999, **181**:6679-6688.
- Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wüthrich K: **Metabolic fluxes in riboflavin-producing *Bacillus subtilis*.** *Nat Biotechnol* 1997, **15**:448-452.
- Klappa M, Aon JC, Stephanopoulos G: **Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry.** *Eur J Biochem* 2003, **270**:3525-3542.
- Petersen S, de Graaf AA, Eggeling L, Möllney M, Wiechert W, Sahm H: **In vivo quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*.** *J Biol Chem* 2000, **275**:35932-35941.
- Dauner M, Storni T, Sauer U: ***Bacillus subtilis* metabolism and energetics in carbon-limited and carbon-excess chemostat culture.** *J Bacteriol* 2001, **183**:7308-7317.
- Schwender J, Ohlrogge JB, Shachar-Hill Y: **A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos.** *J Biol Chem* 2003, **278**:29442-29453.
- Roessner-Tunali U, Liu J, Leisse A, Balbo I, Perez-Melis A, Willmitzer L, Fernie AR: **Kinetics of labelling of organic and amino acids in potato tubers by gas chromatography-mass spectrometry following incubation in  $^{13}\text{C}$  labelled isotopes.** *Plant J* 2004, **39**:668-679.
- Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, et al.: **Essential *Bacillus subtilis* genes.** *Proc Natl Acad Sci USA* 2003, **100**:4678-4683.
- Duetz WA, Ruedi L, Hermann R, O'Connor K, Buchs J, Witholt B: **Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates.** *Appl Environ Microbiol* 2000, **66**:2641-2646.
- Fischer E, Sauer U: **Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS.** *Eur J Biochem* 2003, **270**:880-891.
- Szyperski T: **Biosynthetically directed fractional  $^{13}\text{C}$ -labeling of proteinogenic amino acids: an efficient analytical tool to investigate intermediary metabolism.** *Eur J Biochem* 1995, **232**:433-448.
- Fillinger S, Boschi-Muller S, Azza S, Dervyn E, Branlant G, Aymerich S: **Two glyceraldehyde 3-phosphate dehydrogenases with opposite physiological roles in a non-photosynthetic bacterium.** *J Biol Chem* 2000, **275**:14031-14037.
- Doan T, Aymerich S: **Regulation of the central glycolytic genes in *Bacillus subtilis*: binding of the repressor CggR to its single DNA target sequence is modulated by fructose-1, 6-bisphosphate.** *Mol Microbiol* 2003, **47**:1709-1721.
- Ludwig H, Homuth G, Schmalisch M, Dyka FM, Hecker M, Stülke J: **Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon.** *Mol Microbiol* 2001, **41**:409-422.
- Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB: **High-throughput classification of yeast mutants for functional genomics using metabolic footprinting.** *Nat Biotechnol* 2003, **21**:692-696.
- Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: **Metabolite profiling for plant functional genomics.** *Nat Biotechnol* 2000, **18**:1157-1161.
- Raghevedran V, Gombert AK, Nielsen J: **Phenotypic characteri-**

- zation of glucose repression mutants in *Saccharomyces cerevisiae* using experiments with  $^{13}\text{C}$ -labelled glucose. *Yeast* 2004, **21**:769-779.
27. Hyvärinen A, Karhunen J, Oja E: *Independent Component Analysis* New York: John Wiley & Sons; 2001.
  28. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning* New York: Springer-Verlag; 2001.
  29. Himberg J, Hyvärinen A, Esposito F: **Validating the independent components of neuroimaging time series via clustering and visualization.** *Neuroimage* 2004, **22**:1214-1222.
  30. Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wüthrich K: **Metabolic fluxes in riboflavin-producing *Bacillus subtilis*.** *Nat Biotechnol* 1997, **15**:448-452.
  31. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, et al.: **The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.** *Nature* 1997, **390**:249-256.
  32. Fischer E, Zamboni N, Sauer U: **High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived  $^{13}\text{C}$  constraints.** *Anal Biochem* 2004, **325**:308-316.
  33. Jolliffe IT: *Principal Component Analysis* 2nd edition. New York: Springer Verlag; 2002.
  34. Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmeilin R: **Dynamic imaging of coherent sources: studying neural interactions in the human brain.** *Proc Natl Acad Sci USA* 2001, **98**:694-699.
  35. Brown GD, Yamada S, Sejnowski TJ: **Independent component analysis at the neural cocktail party.** *Trends Neurosci* 2001, **24**:54-63.
  36. Lee SI, Batzoglou S: **Application of independent component analysis to microarrays.** *Genome Biol* 2003, **4**:R76.
  37. Liebermeister W: **Linear modes of gene expression determined by independent component analysis.** *Bioinformatics* 2002, **18**:51-60.
  38. **HUT-CIS: the FastICA package for MATLAB** [<http://www.cis.hut.fi/projects/ica/fastica>]