



## A novel platform for automated high-throughput fluxome profiling of metabolic variants

Stephanie Heux, Juliette Poinot, Stéphane Massou, Serguei Sokol,  
Jean-Charles Portais

### ► To cite this version:

Stephanie Heux, Juliette Poinot, Stéphane Massou, Serguei Sokol, Jean-Charles Portais. A novel platform for automated high-throughput fluxome profiling of metabolic variants. *Metabolic Engineering*, 2014, 25, pp.8 - 19. 10.1016/j.ymben.2014.06.001 . hal-01268982

**HAL Id: hal-01268982**

**<https://hal.science/hal-01268982>**

Submitted on 13 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **A novel platform for automated high-throughput fluxome profiling of metabolic variants**

Heux Stéphanie<sup>1,2,3\*</sup>, Poinot Juliette<sup>1,2,3</sup>, Massou Stéphane<sup>1,2,3</sup>, Sokol Serguei<sup>1,2,3</sup>, Portais Jean-Charles<sup>1,2,3</sup>

<sup>1</sup> Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

<sup>2</sup> INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

<sup>3</sup> CNRS, UMR5504, F-31400 Toulouse, France

\* Corresponding author. Mailing address Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INSA Toulouse, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France. Fax: +33 561 55 96 89. E-mail: stephanie.heux@insa-toulouse.fr

## Abstract

Advances in metabolic engineering are enabling the creation of a large number of cell factories. However, high-throughput platforms do not yet exist for rapidly analyzing the metabolic network of the engineered cells. To fill the gap, we developed an integrated solution for fluxome profiling of large sets of biological systems and conditions. This platform combines a robotic system for  $^{13}\text{C}$ -labelling experiments and sampling of labeled material with NMR-based isotopic fingerprinting and automated data interpretation. As a proof-of-concept, this workflow was applied to discriminate between *E. coli* mutants with gradual expression of the glucose-6-phosphate dehydrogenase. Metabolic variants were clearly discriminated while pathways that support metabolic flexibility towards modulation of a single enzyme were elucidating. By directly connecting the data flow between cell cultivation and flux quantification, considerable advances in throughput, robustness, release of resources and screening capacity were achieved. This will undoubtedly facilitate the development of efficient cell factories.

## Abbreviations

MS: mass spectrometry; NMR: nuclear magnetic resonance; HT: high throughput; OD600: optical density at 600 nm; CDW: cell dry weight; RSD: relative standard deviation; G6PDH: glucose-6-phosphate dehydrogenase; PCA: principal component analysis; WT: wild type; R5P:C1: C1 of ribose-5-phosphate; AKG:C4, C4 of alpha-ketoglutarate, OAAA/PYR:C3: C3 of the combined pool of oxaloacetate/pyruvate; PYR:C3C3 of pyruvate; R5P:C3: C3 of ribose-5-phosphate; PPP: pentose phosphate pathway; LiHa: Liquid Handling, HEPA: High Efficiency Particular Air Filter; DO; dissolved oxygen; TSP-d4: 3-(trimethylsilyl)-2,2',3,3'-

tetradeteropropionic acid; HCA: hierarchical clustering analysis; TCA: tricarboxylic acid cycle; ED: Entner-Doudoroff; FTBL: flux table format.

## 5 **Keywords**

High-throughput, fluxomics, automation, NMR, statistical analysis, flux calculation.

## 1. Introduction

In the field of biotechnology, metabolic engineering and synthetic biology, fluxomics has been identified as a key analytical technology (Ellis and Goodacre, 2012; Feng et al., 2010; Sanford et al., 2002; Stephanopoulos, 1999) not only for the rational design of cells but also for comprehensive understanding of the link between genotype and phenotype. Fluxomics, i.e. the cell-wide quantification of metabolic fluxes, reveals the actual operation of metabolic networks in given environmental conditions, resulting from the integrated flow of interactions between all molecular components - genes, mRNAs, proteins and metabolites (Sauer, 2004; Sauer, 2006). Hence, metabolic flux analysis is the most comprehensive description of the metabolic phenotype at the cellular level (Wittmann and Portais, 2013). In practice,  $^{13}\text{C}$ -fluxomics is still a tedious and rather time consuming process. The method combines both  $^{13}\text{C}$ -labeling experiments and mathematical modeling of biochemical networks. Cells are grown on  $^{13}\text{C}$ -labeled substrates to metabolic and isotopic steady state and the labeling patterns of metabolites are monitored by mass spectrometry (MS), nuclear magnetic resonance (NMR), or both. The labeling information can be collected on metabolic end-products such as protein-bound amino acids, which accumulate to larger extents than true metabolic intermediates (Sauer, 2004; Szyperski, 1995; Wiechert, 2001). Combined with quantitative physiological data and a detailed metabolic model of metabolism, the labeling patterns give access to the *in vivo* reaction rates (i.e. fluxes) associated with the cellular network of an organism. When this information is not available, multivariate statistics can be applied to the isotopic data to provide detailed phenotyping of biological systems without any prior knowledge (Raghevedran et al., 2004; Zamboni and Sauer, 2004).

The rapid expansion of fluxomics and its areas of application is driving the need for high-throughput approaches to enable comprehensive metabolic investigations of a growing number

of organisms, engineered mutants and physiological conditions (Ellis and Goodacre, 2012). Significant advances have been made in the field in the last decade. Miniaturized cell cultivation systems have been used to perform  $^{13}\text{C}$ -labelling experiments with less effort and cost in labeled substrate (Balcarcel and Clark, 2003; Betts and Baganz, 2006; Ge et al., 2006; Girard et al., 2001; Huber et al., 2009; Isett et al., 2007; Kocincova et al., 2008; Tang et al., 2009). The sensitivity, speed, and robustness of both NMR-based and MS-based isotopic analysis have been improved (Boisseau et al., 2013; Cahoreau et al., 2012; Fan and Lane, 2008; Fischer et al., 2004; Giraudeau et al., 2012; Giraudeau et al., 2011; Massou et al., 2007a; Peng, 2012). Tools have been developed for large scale isotopic data processing (Millard et al., 2012; Poskar et al., 2012) as along with improved algorithms and software for flux calculation and statistical analysis of isotopic data (Antoniewicz et al., 2007; Quek et al., 2009; Raghevendran et al., 2004; Sokol et al., 2012; Weitzel et al., 2012; Zamboni et al., 2005; Zamboni and Sauer, 2004). These developments have allowed large-scale flux analysis to be applied to microorganisms (Amador-Noguez et al., 2010; Blank et al., 2005; Fischer and Sauer, 2005; Fischer et al., 2004; Haverkorn van Rijsewijk et al., 2011; Wittmann et al., 2004) and mammalian cells (Munger et al., 2008).

However, despite all the above improvements, HT fluxomics is still in early development. In particular, there is no existing HT fluxomics platform which combines all experimental and in silico steps of the complete workflow in a single and fully integrated manner. The design and development of such a platform is hindered by several technical challenges that need to be addressed in parallel. This includes (1) the tight control of cultivation parameters to ensure metabolic and isotopic steady state; (2) the automated monitoring of growth parameters to collect labeled material once steady state is achieved (3) parallel cultivation to increase throughput; (4) miniaturization of working volumes to reduce the cost of labeled substrates and to facilitate the parallel processes; (5) automated, rapid, parallel sampling of labeled material

to avoid degradation of the metabolites; (6) rapid, sensitive measurement of isotopic profiles to cope with both the small amounts of biological material and the large number of samples; (7) automated extraction of labeling information from raw analytical data, which today is still mainly done manually, thereby saving time and effort and improving data robustness (8), data interpretation tools that provide valuable metabolic information in a high throughput manner, i.e. with reduced user supervision.

For the first time, these challenges were met with the development of a fully integrated solution for fluxome analysis that combines a robotic cultivation and sampling workstation for  $^{13}\text{C}$ -labelling experiments with NMR-based isotopic profiling, and tools for processing and interpreting isotopic data. The automation, parallelization, optimization and integration of all steps in the workflow are described in detail. As a proof of concept, the new platform was applied to a set of *Escherichia coli* mutants with varying levels of a single enzyme and grown on two different  $^{13}\text{C}$ -labelled carbon sources. The power of the overall approach to provide discriminating metabolic information for a large number of mutants and conditions as well as its value in generating valuable metabolic knowledge was demonstrated.

## 2. Materials and methods

Figure 1 is a schematic diagram of the complete workflow developed in this study

### 2.1. Bacterial strains

*Escherichia coli* MG1655 strains used in this study are listed in supplementary Table S1. The over-expression and under-expression mutants (Pzwf) were created by replacing the native

promoter of *zwf* by artificial promoters of different strength (Meynial-Salles et al., 2005; Nicolas et al., 2007), resulting in the gradual expression of glucose-6-phosphate dehydrogenase. The deletion mutant ( $\Delta zwf$ ) was obtained with a one-step disruption protocol and the *in vitro* activity of the glucose-6-phosphate dehydrogenase was measured in crude cell extracts (Nicolas et al., 2007).

## 2.2. Media and cultivation conditions

All *E. coli* cells were grown on minimal synthetic medium containing (per liter) 2.5 g of either glucose or xylose, 17.4 Na<sub>2</sub>HPO<sub>4</sub>.12H<sub>2</sub>O, 3.03 g of KH<sub>2</sub>PO<sub>4</sub>, 0.51 g of NaCl, 2.04 g of NH<sub>4</sub>Cl, 0.49 g of MgSO<sub>4</sub>, 4.38 mg of CaCl<sub>2</sub>, 15 mg of Na<sub>2</sub>EDTA.2H<sub>2</sub>O, 4.5 mg of ZnSO<sub>4</sub>.7H<sub>2</sub>O, 0.3 mg of CoCl<sub>2</sub>.6H<sub>2</sub>O, 1 mg of MnCl<sub>2</sub>.4H<sub>2</sub>O, 1 mg of H<sub>3</sub>BO<sub>3</sub>, 0.4 mg of Na<sub>2</sub>MoO<sub>4</sub>.2 H<sub>2</sub>O, 3 mg of FeSO<sub>4</sub>.7H<sub>2</sub>O and 0.3 mg of CuSO<sub>4</sub>.5H<sub>2</sub>O, 0.1 g of thiamine. All the cultivations were performed at 37 °C. *E. coli* cells were freshly inoculated from a glycerol stock on a LB (10 g/l of tryptone, 5 g/l of yeast extract and 10 g/l of NaCl) liquid medium for six hours. Liquid pre-cultures containing minimal synthetic medium were inoculated from the LB cultures. Cultures were inoculated with 0.57 mg CDW.L<sup>-1</sup>. The minimal synthetic labeled medium for <sup>13</sup>C-labeling experiments contained 13.9 mM of a mixture containing either 20% (mol/mol) [U-<sup>13</sup>C]-glucose and 80% (mol/mol) [1-<sup>13</sup>C]-glucose or [U-<sup>13</sup>C]-xylose (99% of <sup>13</sup>C atom, Eurisotop, France) and 80% (mol/mol) [1-<sup>13</sup>C]- xylose (99% of <sup>13</sup>C atom, Eurisotop, France). Cells were grown in mini stirred tank reactors.

## 2.3. Automated cultivation and sampling platform

The system presented here incorporates a bioreaction block developed by Weuster-Botz et al. (Kusterer et al., 2008; Puskeiler et al., 2005) in an automatic workstation (Freedom EVO 200,



TECAN, Switzerland) (Fig. 1). The bioreaction block has a capacity of 48 stirred tank bioreactors equipped with solid state sensors for oxygen and pH and a working volume ranging from 8 to 15 ml. A lid with magnetic impellers on the central hollow axles enabled stirring, circulation and distribution of sterile inlet air into the head space. A series of 48 holes in the lid allowed the Liquid Handling (LiHa) robotic arm to access each bioreactor. The temperature of the bioreaction block was regulated using a water circulation bath. The bioreactor block was also designed to minimize evaporation and contamination of the medium in each individual bioreactor. For a detailed description of this bioreactor block, see Weuster-Botz et al. (Kusterer et al., 2008; Puskeiler et al., 2005). The robotic workstation was equipped with three robotic arms to handle liquids, transport tubes, microtiter plates, centrifuge buckets, filters, height modules (i.e. a barcode reader, a cooling module, a heating module, a plate washer, a plate reader, a HCL dispenser, a centrifuge and a filter station) and a high efficiency particulate air filter hood to provide sterile conditions. The robot was activated via control software (EVOware Pipetting Software and Pegasus Event Control Software, TECAN). A database (SQL Manager 2008) made it possible to collect the barcodes of each destination tube, values of pH, dissolved oxygen (DO) and optical density ( $OD_{600}$ ) with a time stamp for each bioreactor, a volume and time stamp for the liquids dispensed into each bioreactor, a volume and a time stamp, and the origin (i.e. name of the bioreactor) of the liquid dispensed into each destination tube and the name of the protocol used. The automated protocol used in this study is described in Supplementary Fig. S1 and Supplementary Note.

## 2.4. Processing of physiological data

The maximum specific growth rate was calculated from the time profiles of biomass as described previously (Heux et al., 2011). The specific rates for substrate consumption and product formation were derived from the NMR measurement (see section below) of glucose,

xylose and acetate in supernatants. The specific consumption rate of oxygen was derived from online measurement. All specific rates were determined during the exponential growth phase using a previously validated correlation factor of  $0.38 \text{ g CDW.L}^{-1} \text{ OD}_{600} \text{ Unit}^{-1}$  (Nicolas et al., 2007) .

5

## 2.5. Sample preparation for NMR analysis

Samples of cultivation broth were automatically collected by the pipetting arm of the workstation, and were centrifuged to separate cells from supernatants. The cell pellets were automatically washed twice with physiological buffer (0.9% of NaCl). To release protein-bound amino acids from cellular proteins, cell pellets collected with the workstation were hydrolyzed 12 hours with 6N HCl at 110°C. Because strong acids can be hazardous and deleterious for all metal parts of the robot, the decision was made to not automatize this step. Hence, addition of HCl to cell pellets was done manually in a safety hood. HCl was evaporated under low pressure (20 mbar, 50 °C). Biomass hydrolysates were washed twice in D<sub>2</sub>O using the same evaporation method. The dried hydrolysates were resuspended in 600 µL D<sub>2</sub>O and centrifuged; 50 µL of 3-(trimethylsilyl)-2,2',3,3'-tetradeuteriopropionic acid (TSP-d<sub>4</sub>) at 0.74 g/L and 50 µL of DCl at 1.3 M were added to 500 µl of the mixture. Samples containing supernatants were prepared for the determination of glucose and acetate concentrations, by adding 100 µL of TSP-d<sub>4</sub> at 4,3 mM to 500 µL of supernatant.

## 2.6. NMR experiments

All samples were analyzed on a Bruker Ascend 800 MHz magnet using a 5 mm QCI cryoprobe. The magnet was equipped with a sample changer (Bruker SampleJet) with a capacity of 500 NMR tubes. Using this device, samples can be stored at 6 °C and pre-heated to 25 °C just prior

the NMR, thereby ensuring automation and appropriate storage conditions. The 1D  $^1\text{H}$ -NMR spectra of both biomass hydrolysates and supernatants were collected. The NMR acquisition parameters were 256 transients per spectrum, 128K data points, 6.83 s acquisition time, 5 s recycle delay,  $30^\circ$  pulse angle, and presaturation of the water signal during the relaxation delay.

5 The FIDs were zero-filled, Fourier transformed with 0.5-Hz exponential line broadening, manually phase corrected, automatically baseline corrected, and aligned to the TSP-d4 signal.

To facilitate the automated interpretation of NMR data, the  $^1\text{H}$ -NMR spectra were further processed as follows. The signals from the solvent and from TSP-d4 were replaced by averaged signal noise regions. Spectra were then normalized to the total spectrum area. The data size was

10 reduced 21-fold. To this end, data were first smoothed by convolving with a Gaussian filter constructed in such a manner that its  $\pm 2\sigma$ -support was composed of 21 points. Then every 21<sup>st</sup> point was used to produce a dimensionally reduced spectrum containing 11039 data points instead of the original 256 k data points. A program developed in R (R Development Core Team, 2009) was generated to automatically apply the above processing to all spectra.

## 2.7. Assignment of $^1\text{H}$ -NMR signals and assignment database

A separate labeled biomass hydrolysate sample was produced specifically to assign and characterize all signals in the NMR spectra. This sample was obtained from the hydrolysis of

20 100 mg dry biomass collected from *E. coli* (strain Nissle 1917) cells grown on a mixture of 50% (mol/mol) [ $\text{U-}^{13}\text{C}$ ]-glucose and 50% (mol/mol) unlabeled glucose in shake flask culture.

A 2D ZQF-TOCSY spectrum (Massou et al., 2007b) was collected from this sample from which the NMR signals were assigned. In total, 70 NMR resonances showing  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear couplings were detected. Among the 70 protonated carbon position detected, 59 were

25 unambiguously assigned to the proton resonances of 16 amino acids and of ribose. The

remaining 11 protonated carbon positions could not be assigned but showed  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear couplings, and hence corresponded to protonated carbon positions of unidentified compounds. The cross-peak sections of the 2D ZQF-TOCSY spectrum were used to determine the spectral characteristics of each resonance. For each one, the following information was extracted and compiled in a database: the lower and upper frequency boundaries; the nature of the signal (i.e. center or satellites); the name of the compound and the corresponding carbon position of the compound; the number of protons attached to the carbon. In addition, based on data from Szyperski et al. (Szyperski, 1995), the name and the corresponding carbon position of the metabolic precursor of the proteinogenic amino acids were specified. The final assignment database contained 11 unidentified and 59 identified protonated carbon positions, 176 satellite signals and 88 center signals, corresponding to a total of 28 compounds (Supplementary Table S2).

## 2.8. AEDE: Automated Enrichment Data Extraction tool

A tool named AEDE programmed in R (R Development Core Team, 2009) was developed to model the NMR spectra as linear combinations of single resonances corresponding to each protonated carbon position measured in the NMR experiment. The resonance of a proton attached to a carbon in a given compound is comprised of a “central” signal ( $^{12}\text{C}$ -bound protons) and two “satellite” signals ( $^{13}\text{C}$ -bound protons) positioned symmetrically to the central signal. For a given carbon,  $^{13}\text{C}$  enrichment corresponds to the ratio of the area of the satellite signals to the total area of the resonance (i.e. sum of the area of the satellite and central signals). Each signal was modeled as a triangle (Fig. 1C). The width (i.e. lower and upper frequency boundaries) of the triangles was taken from the assignment database (see previous section). Their heights were related to the parameters to be fitted in an optimization process and were

proportional to 1) the number of attached protons (determined from the molecule structure); 2) the relative proportions of  $^{12}\text{C}$  and  $^{13}\text{C}$  nuclei (for the central and satellite signals respectively) in the considered carbon position (to be fitted); and 3) a weight that could be assimilated to a concentration of the considered compounds (to be fitted). Molecules that have incorporated multiple  $^{13}\text{C}$  atoms express both short-range and long-range  $^1\text{H}$ - $^{13}\text{C}$  couplings, which can influence the shape (i.e. height and width) of the  $^1\text{H}$ -NMR signals. In addition, in certain regions,  $^1\text{H}$ -NMR spectra obtained on isotopically labeled biomass hydrolysate present a high degree of overlap. For that reason, the assignment database was built using a biomass hydrolysate selected to ensure both maximum signal width and a high rate of overlapping signals (see previous section). Firstly, this situation ensures that each triangle will be wide enough to cover the entire signal whatever its shape, such that each signal will be assigned to a specific triangle (and not divided across two or more triangles). Secondly, multiple triangles can be introduced to account for the overlapping signals known to occur in dense regions of the spectra. Therefore, by construction, the influence of both  $^1\text{H}$ - $^{13}\text{C}$  couplings and overlapping are taken into account in the calculation of  $^{13}\text{C}$  enrichment.

This is summed up in the following formula: 
$$S = T_c ((1 - {}^{13}\text{C}).M) + T_s ({}^{13}\text{C}.M) + b$$

where  $S$  is a simulated spectrum,  $T_c$  and  $T_s$  are matrices whose columns are basis spectra composed of central ( $T_c$ ) and satellite ( $T_s$ ) signals per carbon atom position; row number (in  $T_c$  and  $T_s$ ) is equal to the number of points in sub-sampled spectra; column number is equal to the number of modeled carbon enrichments;  ${}^{13}\text{C}$  is a vector of  $^{13}\text{C}$ -enrichments,  $\mathbf{1}$  is a vector of the same length as  ${}^{13}\text{C}$  composed of 1's so that  $(\mathbf{1} - {}^{13}\text{C})$  is equal to  ${}^{12}\text{C}$ , a vector of  $^{12}\text{C}$  proportions; a scalar value  $b$  is a baseline level intensity; finally,  $M$  is a vector of compound concentrations where each concentration is repeated as many times as there are modeled carbons in the

molecule. In this way,  $M$  has the same length as  $^{13}C$  and a low dot operator  $Z=X.Y$  expresses a component-wise vector multiplication  $Z_i=X_i.Y_i$  for all vector components  $i$ .

A least-square problem was formulated and solved to fit a simulated spectrum  $S$  to a measured spectrum  $D$  by minimizing the following cost function  $Q$ :  $Q(^{13}C, M, b) = \|S - D\|^2$  where  $Q$  is a

5 function of unknown parameters to be fitted: enrichments  $^{13}C$ , concentrations  $M$  and baseline  $b$ . The minimization of  $Q$  is subject to inequalities  $0 \leq ^{13}C \leq 1$  (by definition of  $^{13}C$ ) and  $M \geq 10^{-10}$  as concentration must be positive to avoid degeneration of the Jacobian matrix; it is also subject to equality constraints. These equalities are derived from the fact that many carbon positions in the measured compounds come from the same metabolic precursor (Supplementary

10 Table S3) and, consequently, their  $^{13}C$  enrichments must be identical. Equalities are automatically generated from the information entered in the assignment database in “Name of metabolic precursor” (Supplementary Table S2). For a total of 59 protonated carbon positions identified, we formulated 35 equalities (Supplementary Table S3). The formulated least-square problem with inequality and equality constraints was solved by NLISC algorithm (Sokol et al.,

15 2012) resulting in the estimation of  $^{13}C$  enrichments of 70 protonated carbon positions, 28 metabolite concentrations and 1 constant baseline level for each spectrum. Confidence intervals at 95% level for protonated carbon positions were also provided. The confidence intervals were computed from the standard deviation of residual vector and Jacobian matrix. They tell how precisely the parameters have been determined. If a confidence interval is very wide, the

20 parameter is not defined with a good precision. Different output file formats are generated to facilitate the interpretation of labeling data. For multivariate statistical analysis, a matrix containing the  $^{13}C$  enrichment values across the experimental conditions and replicates is provided in CSV format. The  $^{13}C$  enrichment data averaged over the four biological replicates and their standard deviations are in a data file that is automatically generated to perform flux

25 calculations using influx\_s software (Sokol et al., 2012). To validate the modeling approach,

standard solutions of glucose containing various proportions of [1- $^{13}\text{C}$ ]-glucose (from 5 to 90% of  $^{13}\text{C}$  atom in C1 position) and of lysine containing 50% (mol/mol) [U- $^{13}\text{C}$ ]-lysine and 50% (mol/mol) unlabeled lysine were prepared and analyzed by  $^1\text{H}$ -NMR. All spectra were integrated either manually or automatically (i.e. using AEDE). The accuracy was evaluated by calculating the mean of the differences between measured - either manual or modeled - data with theoretical data. The precision was estimated as the square root of the mean squared error issued from the linear regression of measured vs theoretical values for glucose. For lysine, the precision was estimated as the standard deviation of the difference between manual or modeled data with theoretical data. Standard solutions were prepared as described in (Cahoreau et al., 2012).

## 2.9. Statistical data analysis

Estimates of the  $^{13}\text{C}$  enrichments of the protonated carbon positions across the 80 spectra and obtained as solutions to the least squares problem (described in the previous section), were used as input variables for principal component analysis (PCA) and hierarchical clustering analysis (HCA) (Varmuza and Filzmoser, 2009). The  $^{13}\text{C}$  enrichment values with a RSD higher than 20% (i.e. 5) were discarded from the analysis, resulting in a total of 17 protonated carbon positions for each condition. The computation was performed with R (R Development Core Team, 2009) using the classPCA and hcaSpectra functions of the ChemoSpec Packages, an R package for the chemometric analysis of spectra (Hanson, 2012). Before applying PCA, the data were centered and scaled using Pareto scaling (van den Berg et al., 2006). For the HCA, Euclidean distance was used as the metric and the clustering method used complete linkage. To identify differences between conditions at the variable level (i.e.  $^{13}\text{C}$  enrichments), a two-way ANOVA was applied. This enabled the main effects of each factor (i.e. genetic and

environmental conditions) to be determined but also checked for significant cross-effects between the factors.

## **2.10. Metabolic flux analysis**

5

Metabolic fluxes were calculated from the estimates of  $^{13}\text{C}$  enrichments of 23 protonated carbon positions across the 80 spectra and obtained as the solution of the least squares problem described in the previous section.  $^{13}\text{C}$  enrichment values for each protonated carbon position were averaged across the biological replicates. The metabolic network considered for flux  
10 calculations contained the main pathways of *E. coli* central metabolism: glycolysis, the pentose phosphate pathway (PPP), the tricarboxylic acid cycle (TCA), the glyoxylate shunt and anaplerotic reactions. It also included the Entner-Doudoroff (ED) pathway and amino acid biosynthesis reactions (Nicolas et al., 2007; Revelles et al., 2013). For growth on xylose, the metabolic network was adapted to the metabolism of the sugar (Nakahigashi et al., 2009). Flux  
15 calculations were performed using the software influx\_s (Sokol et al., 2012), in which both mass balances and carbon atom transitions describing the bioreaction network were used. Measurable extracellular fluxes (glucose uptake, acetate production and specific precursor requirements for biomass) were used as additional measurements.



### 3. Results and discussion

#### 3.1. Development of a robotic cultivation and sampling system

5 A workstation enabling i) the automated parallel cultivation of bacteria with  $^{13}\text{C}$ -labelled substrates and ii) the automated sampling of labeled metabolites under steady state conditions was designed and constructed. Cell cultivation was automated, parallelized and miniaturized by incorporating a block of 48 micro-scale bioreactors (Kusterer et al., 2008; Puskeiler et al., 2005) in the automatic workstation. This system allows fully automated and reproducible  
10 monitoring of growth, pH and oxygen (Fig. 1A) for each individual bioreactor, but also automated control of pH, temperature and stirrer speed. Tight control of growth parameters is vital to ensure stable physiological conditions throughout cultivation. The working volume of a single bioreactor can be set between 8 and 15 mL, i.e., a valuable trade-off between cultivation and throughput constraints while requiring limited amount of labeled substrate for each  
15 experimental condition. Each bioreactor is equipped with a sampling port that allows the cultivation broth to be collected automatically for various purposes. For stationary flux determination, cells have to be collected at metabolic and isotopic steady-states (Wiechert and Nöh, 2005), which is achieved during the exponential growth phase in batch culture (Sauer, 2004). Growth is monitored by collecting broth samples and reading optical densities ( $\text{OD}_{600}$ ).  
20 When the  $\text{OD}_{600}$  corresponding to steady state is reached, the labeled material is automatically sampled (Supplementary Fig. S1 and Supplementary Note 1). The target ODs are set by the user, and are defined in preliminary investigations where metabolic steady state was validated from the stability of input (substrate) and output (products, biomass) fluxes or of intracellular concentrations. The isotopic steady-state is usually reached after the 5<sup>th</sup> or 6<sup>th</sup> generation time  
25 (Toya et al., 2010), as verified from the stability of labelling patterns in independent experiments. Hence labeled samples were collected after 6 generation times in the experiments

reported here. Cultivation broths are collected by the pipetting arm, and cells are separated from the medium by centrifugation. Cell pellets and supernatants are stored at -20 °C until analysis. The monitoring of growth and sampling of labeled material can be performed independently for each individual bioreactor, so that strains or mutants, or conditions, with different growth rates can be investigated in parallel. Since all the cultivation and sampling processes – including centrifugation steps - are automated, no human intervention is needed. The user is automatically informed – by email – of all key operations so that he/she can monitor the entire process. The time needed to collect one cultivation broth from one bioreactor is one minute (including pipetting and the movement of the tube to the centrifuge). Because the pipetting arm can take eight samples simultaneously in one minute, sample collection for the 48 bioreactors can be completed in six minutes. The total time needed to recover the cell pellets is longer because three successive centrifugation steps are needed to remove all media components. Here again, pipetting and centrifugation can be parallelized. In the automation procedure, priority is given to the collection of cultivation broths over later steps (i.e Step 6 to 9, Supplementary Fig. S1) to ensure the labeled material is collected under steady-state conditions and to avoid degradation of labeled material during sampling. In addition, the three robotic arms and other robot modules can work independently, providing further parallelization capability, and increasing process flexibility. These different processes are controlled by a Tecan Event control software, which was customized to fit the specific requirements of the workstation and to manage all events and priorities. This software also included a database for storing and sharing information from which the following parameters were extracted for further data interpretation: pipetting volume, sampling time, and growth, pH and oxygen parameters. Cultivation systems designed for flux analysis have already been described (Fischer et al., 2004; Nanchen et al., 2006; Wittmann et al., 2004) but manual intervention was required for monitoring growth, or for sampling labeled material, or both. In comparison, the robotic

platform described here enables full automation of all cultivation and sampling steps, thereby saving human effort and time. The machine can operate 24 hours a day; and up to 96 samples can be collected and processed in parallel, continuously and with little human supervision. Both automation and the use of bar-coding to track samples and sub-samples make it possible to avoid the experimental errors that are inherent to manual operations. Most importantly, cultivations are performed under fully controlled conditions and can be applied to a broad variety of organisms that can be grown and sampled in stirred-tank bioreactors. This includes bacteria (Knorr et al., 2007; Kusterer et al., 2008), yeast (Kusterer et al., 2008), filamentous fungi (Hortsch et al., 2011) and mammalian cells (Ge et al., 2006; Legmann et al., 2009)

### **3.2. 1D-<sup>1</sup>H-NMR based isotopic profiling**

Isotopic profiles were collected by NMR since this technique has many attractive features for HT fluxomics. Because of its high resolution, chromatographic separation is not required hence sample preparation is significantly decreased. In addition, NMR is nondestructive enabling samples to be re-used for further analysis (e.g. using different NMR pulse sequences or MS-based isotopomer profiling). Most importantly, NMR provides direct information on the position of the label in the molecule (Schmidt et al., 1997; Tang et al., 2009), ensuring a maximum amount of information is available to solve metabolic fluxes. Lastly, the capability of NMR for its high speed, robustness and reproducibility has been demonstrated through the widespread metabolic profiling based applications (Beckonert et al., 2007; Lenz and Wilson, 2007). In this work, labeling information was extracted from protein-bound amino acids since proteins are very abundant and chemically stable. This is an advantage for the NMR sensitivity, the miniaturization of sample sizes and the stability of labeled material during the sampling process. In addition the chemical composition of across samples is quite similar thus greatly

facilitating setting up automated processing tools for isotopic data extraction, as shown further in this work. Finally this is the most widespread approach for steady-state flux analysis thus readily amenable for automation.

In this study, the isotopic profiles of protein-bound amino-acids were obtained from the 1D  $^1\text{H}$ -NMR analysis of biomass hydrolysates. To ensure throughput, samples were placed in an NMR auto-sampler and profiled by 1D-  $^1\text{H}$ -NMR sequence. In the first trials, the  $^1\text{H}$ -NMR spectra were collected in one hour per sample, which was reduced to 15 minutes with further optimization of the entire process. As illustrated with alanine (Fig. 1B), the  $^{13}\text{C}$  content of a protonated carbon position of a molecule can be measured by  $^1\text{H}$ -NMR. This is due to heteronuclear  $^1\text{H}$ - $^{13}\text{C}$  ( $J_{\text{CH}}$ ) couplings, by which the signal of protons bound to a  $^{13}\text{C}$  atom are split into two “satellite” signals that appear symmetrically to the “central” signal of protons bound to a  $^{12}\text{C}$  atom. The area of the satellite signals relative to the total resonance area (i.e. satellite + central signals) is a direct measurement of the  $^{13}\text{C}$  enrichment in the corresponding carbon position of the molecule. In addition,  $^1\text{H}$ -NMR is much more sensitive than  $^{13}\text{C}$ -NMR, which is critical given the low amount of labeled material that can be obtained in mini-bioreactors. Indeed, instead of the 50 to 200 milligrams of biomass (cell dry weight, CDW) that are classically collected to measure the labeling patterns of protein-bound amino acids by NMR (Massou et al., 2007a), only 1.5 mg CDW can be collected per sample using the mini bioreactors described above. Recent progress in NMR technology means fully exploitable spectra can be obtained from such a small amount of material. This is illustrated in this in this work where 1D  $^1\text{H}$ -NMR spectra with very good signal-to-noise ratios (Fig. 1B) were obtained on a very small amount of biological material using a 800 MHz NMR spectrometer equipped with a cryocooled probe head. The high resolution of the NMR spectra also helped separate overlapping signals. Finally, a high degree of reproducibility was obtained with an average relative standard deviation (RSD) of 7.8% between biological replicates (Supplementary Table S4).

### 3.3. Automated extraction of $^{13}\text{C}$ enrichment data using AEDE

Whether automated or not, the extraction of isotopic enrichments from 1D  $^1\text{H}$ -NMR is a challenging task. The samples are chemically complex and give rise to a high number of  $^1\text{H}$  signals. Because of the narrow spectral width of  $^1\text{H}$  NMR spectra, there is significant overlap between signals. With labeled material, the spectral complexity is further amplified because each signal of a  $^1\text{H}$  nucleus attached to a carbon (i.e. protonated carbon position) is split into three signals (one central signal + two satellite signals) due to the  $J_{\text{CH}}$  couplings, as illustrated in Figure 1B. To resolve the spectral complexity in a HT-compatible manner, we developed a tool, named AEDE (for Automated Enrichment Data Extraction), that automatically extracts the  $^{13}\text{C}$  enrichments from the 1D  $^1\text{H}$ -NMR spectra. AEDE relies on an assignment database containing 59 protonated carbon positions assigned to the amino acids and sugars released from biomass hydrolysis (Supplementary Table S2). Details on the construction of the assignment database are given in Material and Methods below. This database was used to model each detected signal (i.e. central and satellite) as a triangle whose width is equal to the interval between the lower and upper frequency boundaries, and whose height is related to the parameters to be fitted in the optimization process. At the end of the optimization process, the  $^{13}\text{C}$  enrichment of all protonated carbon was provided. The accuracy and precision of this method was evaluated with two types of standard solutions. First, it was applied to a series of solutions containing various proportions of  $[1\text{-}^{13}\text{C}]$  glucose. The enrichment of the C1 position of glucose could be measured with an accuracy of -0.0062 and a precision of 0.0064 (Supplementary Table S5). This is slightly higher than obtained by manual integration of the same peaks (0.0008 and 0.0034, respectively), but still very good. Then the method was applied to a solution of 50%  $[\text{U-}^{13}\text{C}]$ -Lysine, which provides a situation where peaks overlap in the  $^1\text{H}$ -

NMR spectra. In this case, the accuracy was 0.10 by manual integration and 0.02 by automatic fitting, while the precision was respectively 0.064 and 0.054. The strong differences between both approaches were mainly due to the overlapping signals which were much better integrated by the automatic fitting than with manual integration. Together these data demonstrate the benefit of using AEDE tool to provide  $^{13}\text{C}$  enrichments in even complex situation with enough accuracy and precision. AEDE tool was further challenged with four  $^1\text{H}$ -NMR spectra of biomass hydrolysates obtained from four replicate experiments in which wild type (WT) *E. coli* cells were grown on labeled glucose (Fig. 1B). Each spectrum was efficiently simulated using triangle modeling despite the broad diversity of signal shapes and significant overlaps between the detected signals (Fig. 1C), confirming the robustness of the method. The processing of spectra was fast; it took only a few seconds to generate a raw vector containing  $^{13}\text{C}$  enrichment values for all the protonated carbon positions for each spectrum. The mean and standard deviation for the replicates were also provided. Of the 236 enrichment values, 85% (201 measurements) were measured with an acceptable precision. The RSD over replicates was below 20% for 63% of enrichment data. The precision of the model could be further improved by introducing metabolic constraints. The carbon positions of different compounds that originate from the same carbon position of the same metabolic precursor necessarily have the same  $^{13}\text{C}$ -enrichment at steady state. Hence equality constraints can be introduced in the model for all carbon positions of the same metabolic origin (Supplementary Table S2 and S3). The metabolic origin of the different carbons depends on the biosynthetic routes and is organism-dependent. For *E. coli*, the 59 identified carbon positions of amino-acids and ribose detected in the  $^1\text{H}$ -NMR spectra of biomass hydrolysates derive from 24 independent carbon positions in nine metabolic precursors (3-phosphoglycerate, acetyl-CoA,  $\alpha$ -ketoglutarate, erythrose-4-phosphate, oxaloacetate, phosphoenol-pyruvate, pyruvate, ribose-5-phosphate and 5-methyl-tetrahydrofolate). One precursor carbon, ribose C5, was discarded from this analysis because

the only signal from which the  $^{13}\text{C}$ -enrichment could be obtained was badly resolved. A final total of 23 metabolic constraints were included in the model (with a total of 35 equality equations) and applied during the automated extraction of  $^{13}\text{C}$ -enrichments (Supplementary Table S3). When applied to the four replicate biological samples, these equality constraints improved both the precision and the robustness of the extracted  $^{13}\text{C}$ -enrichment data. The confidence intervals for all 92 measured enrichment data (23 carbon positions in metabolic precursors x 4 replicates) were significantly reduced and the RSD over replicates was below 20% for 76% of the enrichment data. Finally, the average of the  $^{13}\text{C}$  enrichments of each redundant carbon position obtained without equality constraints corresponded to the value obtained with equalities (data not shown).

The above data demonstrate the power of the proposed AEDE tool to approximate  $^{13}\text{C}$  enrichments in a fast, robust and sufficiently accurate way. To facilitate the interpretation of labeling data, various output formats can be generated so that statistical and flux calculations can be directly launched in parallel (Fig. 1D). Most importantly, the AEDE tool is not restricted to *E. coli* but widely applicable to biomass hydrolysates derived from any biological source. The equality constraints provided for *E. coli* can be directly exploited for a broad range of microorganisms and conditions. For organisms or conditions for which they cannot be applied directly, the metabolic origin of the various carbon positions of the amino acids should be specified in the assignment database. Alternatively, if the metabolic origin cannot be specified,  $^{13}\text{C}$  enrichments can be extracted without applying equality constraints. Lastly, the AEDE tool described here can be used on  $^1\text{H}$ -NMR spectra of biomass hydrolysates produced independently of the robotic cultivation and sampling platform, using any standard cultivation devices.

### 3.4. Application to variants gradually expressing glucose-6-phosphate dehydrogenase

To experimentally assess the platform for large scale analysis, but also to demonstrate its value for providing detailed biological knowledge, it was applied to ten *E. coli* strains with gradual expression of the *zwf* gene encoding the glucose-6-phosphate dehydrogenase (G6PDH) (Supplementary Table S1). G6PDH is a key enzyme in central metabolism, involved in the partition of carbon flows between glycolysis and the pentose phosphate pathway (PPP). While consequences of the absence and of the increase in G6PDH activity for the metabolic network have been reported (Fischer and Sauer, 2003; Hua et al., 2003; Nicolas et al., 2007; Zhao et al., 2004a; Zhao et al., 2004b), no information is available on the effects of modulating enzyme activity at the flux level. The regulation of metabolism relies on the fine tuning of enzyme activities rather than on switching enzymes on or off. Consequently, the comprehensive understanding of how the quantitative modulation of a single enzyme impacts the entire metabolic network can provide insight into metabolic regulation at the system level.

To address this question, the ten strains were grown in minimum medium containing 13.9 mM of a mixture of 20% (mol/mol) [U-<sup>13</sup>C]-glucose + 80 % (mol/mol) [1-<sup>13</sup>C]-glucose. In addition to glucose, the strains were also grown on a mixture of 20% (mol/mol) [U-<sup>13</sup>C]-xylose + 80 % (mol/mol) [1-<sup>13</sup>C]-xylose. Xylose enters the central metabolism downstream from the mutation, and is consequently a valuable way to check for potential indirect effects. In total, 80 <sup>13</sup>C-labelling experiments were performed, including 10 strains, two labeled inputs, and four replicates per condition. The results are detailed and discussed one by one in the following sections.

### 3.5. Growing and sampling the G6PDH variants with the robotic system



The physiological data obtained for the 80 cultivations are summarized in Figure 2. For each strain and condition, the growth rates were highly reproducible (i.e. with an averaged relative standard deviation of 7.2%) among the four biological replicates (Fig. 2A). The rates of cells grown on xylose were slightly lower than those grown on glucose (Fig. 2A-D). The modulation of G6PDH activity had no significant effect on the growth rate and substrate uptake rate on either sugar (Fig. 2A-B). This result is in agreement with reports showing that the deletion of *zwf* or its over-expression has no impact on growth (Fischer and Sauer, 2003; Hua et al., 2003; Nicolas et al., 2007; Zhao et al., 2004a; Zhao et al., 2004b). However, significant changes in acetate production and in oxygen consumption were observed between the different strains when these were grown on glucose (Fig. 2C-D). Compared to strains with reduced G6PDH activity, strains over-expressing *zwf* showed higher specific oxygen uptake rates and lower acetate production rates. These results indicated higher respiration activity in the over-expressing mutants.

Cell pellets and supernatants were automatically collected in all 80 individual experiments, i.e. 160 samples in total. The target OD<sub>600</sub> for sampling of labeled material was defined in preliminary experiments. In our cultivation conditions, metabolic and isotopic steady-states were achieved at OD<sub>600</sub> values of between 1.0 and 1.5, i.e. at mid-exponential phase. During this sampling period, glucose, xylose, and biomass showed constant specific rates indicating that metabolic steady state was achieved. Isotopic steady state was assumed to be reached since cells were harvested after six generations. Moreover, at this sampling point, the amount of unlabeled biomass coming from the inoculum was negligible (below 2%). In practice, OD<sub>600</sub> measurements are made every 15 minutes, meaning the target OD may be reached and exceeded between two measurements. To get round this problem, the target OD was defined at a low value (1.0) on the workstation. Given that the growth rates in all the study conditions were similar (Fig. 2A), parallel sampling of all cultivations was extremely challenging because all

48 reactors reached the target OD at almost the same time. The harvesting times were highly reproducible among the four biological replicates for each condition and very close between all conditions (Fig. 2E). Thanks to the parallelization of sampling tasks, cells could be harvested with high timing precision despite the close sampling times. The actual ODs at the actual time of harvest were all between 1.054 and 1.282 (Fig. 2E), meaning that all samples were collected at metabolic and isotopic steady state.

### 3.6. Discrimination of G6PDH variants based on isotopic profiles

Cell pellets from the 80 cultivations were hydrolyzed and biomass hydrolysates were analyzed by 1D  $^1\text{H}$ -NMR. The  $^{13}\text{C}$ -enrichments were extracted from the spectra using the AEDE tool. To classify the G6PDH variants and conditions and possibly deduce qualitative metabolic information without prior knowledge, unsupervised multivariate analysis - principal component analysis (PCA) - was applied to  $^{13}\text{C}$  enrichments data. The first two principal components (PC1 and PC2) explained 76% of the total variance (Fig. 3A). Cells grown on labeled glucose were clearly separated from cells grown on labeled xylose by the first component (PC1) (Fig. 3A). When grown on labeled xylose, the different strains could not be distinguished, whereas on labeled glucose, three groups were distinguished. The first cluster included strains that over-expressed *zwf* (Pzwf1 to Pzwf4) and the WT strain. The second cluster was composed of the strain with slightly reduced expression of the *zwf* strain (Pzwf1.4) and the third cluster, which was clearly separated from the first one, included strains with strongly reduced expression of *zwf* (Pzwf 1.1 to 1.3) and the *zwf* deletion mutant ( $\Delta\text{zwf}$ ). These data were confirmed by hierarchical cluster analysis (Supplementary Fig. S2). Because the isotopic profiles were collected by NMR, direct information on the position of metabolites that receive the label is provided. This allows direct identification of the carbon positions with discriminating  $^{13}\text{C}$

enrichments in the PCA, from which metabolic information can be derived. The most discriminating variables were the C1 of ribose-5-phosphate (R5P:C1), the C4 of alpha-ketoglutarate (AKG:C4), the C3 of oxaloacetate or pyruvate (OAAA/PYR:C3) (Szyperski, 1995) and the C3 of pyruvate (PYR:C3) (Fig. 3A). Significant changes (pvalue <0.001, two-way ANOVA) in the  $^{13}\text{C}$  enrichment of these carbon positions were observed between strains and carbon sources (Fig. 3B). This is consistent with the different fate of the C1 of glucose in these strains. In the  $\Delta zwf$  strain, the C1 of glucose will end up in the C3 of pyruvate via glycolysis and then in the C4 alpha-ketoglutarate via the TCA cycle. Label will be also recovered in the C1 of ribose-5-phosphate via the non-oxidative branch of the pentose phosphate pathway (PPP) (Hua et al., 2003; Nakahigashi et al., 2009; Nicolas et al., 2007; Zhao et al., 2004a; Zhao et al., 2004b). For strains with active G6PDH (Supplementary Table S1), part of the C1 of glucose is lost as  $\text{CO}_2$  in the oxidative branch of the PPP, leading to decreased  $^{13}\text{C}$  enrichments for PYR:C3; AKG:C4 and R5P:C1. In contrast, only minor changes were observed between *zwf* over-expressing strains and the WT strain despite an up to 15-fold increase in G6PDH activity (Supplementary Table S1), which shows that higher G6PDH activity does not result in significantly higher PPP flux (Nicolas et al., 2007). Strains grown on labeled xylose were not distinguished on the basis of the C1 of R5P and showed similar  $^{13}\text{C}$  content to the  $\Delta zwf$  strain grown on labeled glucose. This is in agreement with a previous report that, when strains are grown on xylose, there is little activity of G6PDH (Nakahigashi et al., 2009), meaning that pentoses-phosphates are poorly recycled via the oxidative PPP in the experimental conditions concerned.

The above data demonstrate the relevance of multivariate analysis of  $^{13}\text{C}$  enrichment data for screening purpose and model-independent fluxome profiling (Zamboni and Sauer, 2004). They also demonstrate that NMR-based isotopic profiling can provide direct information on changes in pathway activity. This approach can be applied in a high throughput manner,

thereby enabling the large-scale investigation of biochemical pathways and networks for a broad range of organisms.

### 3.7. Estimation of *in vivo* fluxes in G6PDH variants

5

To capture quantitative changes in metabolism caused by the modulation of G6PDH activity, <sup>13</sup>C enrichment data and physiological data were used as measurement inputs for flux estimation (Supplementary file 3). Flux maps were automatically generated for all 20 conditions tested in five minutes using influx\_s software (Sokol et al., 2012). The results are summarized in Figure 4 and the flux values are listed in Supplementary Table S6. Given the type (positional <sup>13</sup>C enrichments) and limited number of isotopic data available, the estimated fluxes were of good quality and in good agreement with data in the literature. For example, for the WT and the 15-fold *zwf* over-expressing strains, the G6PDH fluxes were 3.6 and 3.8 mmol/gDW/h, respectively (Supplementary Table S6), i.e., close to the values previously reported for these strains (Fischer et al., 2004; Nicolas et al., 2007; Sauer et al., 2004). The precision (Supplementary Note 2) of the estimated fluxes varied depending on the biochemical reaction, and was higher for glycolysis and PPP than for anaplerotic reactions. It should be kept in mind that this is the result of a HT process in which all flux maps were generated automatically and in parallel, and thus represents a valuable trade-off between the quality of flux information and number of conditions investigated. When more precise flux values are required, more detailed isotopomer analysis - by either 2D NMR or MS, or both – can be performed on the same labeled samples.

No differences in flux distributions were observed between the 10 strains grown on xylose. There was no flux in the oxidative PPP whatever the strain (Fig. 4A), demonstrating the absence of pentoses-phosphate recycling in the conditions tested here. Interesting features were

observed in the case of strains grown on glucose. In *zwf* under-expressing strains, the absolute G6PDH flux gradually increased with increased enzyme activity (Fig. 4B and Supplementary Table S6) but not in a linear fashion. For instance, the strain Pzwf1.3 retained only 6% G6PDH activity compared to the WT strain but was still able to carry 22% of the G6PDH flux measured in the WT strain. This means that a low G6PDH activity can constrain the metabolic network enough to direct a significant part of carbon flow towards the PPP. In contrast, the G6PDH flux remained constant in all strains that over-expressed *zwf* at a flux value close to that in the WT strain (Fig. 4B and Supplementary Table S6) despite an up to 15-fold increase in G6PDH *in vitro* activity and thus in the amount of active enzyme in the cells (Hüsemann and Papoutsakis, 1989). Taken together, these data show that G6PDH controls the PPP flux – as well as the partitioning of carbon between glycolysis and the PPP- when its concentration is low but not when its concentration is higher than the WT level. The WT level represents the lowest amount needed to obtain the highest flux, suggesting that the concentration of G6PDH is optimal in the WT strain. This is in accordance with previous study showing that important cellular process such as PPP or glycolysis are expressed above the demand to enable a fast switching toward higher NADPH or ATP demand (Valgepea et al., 2013). However, such a conclusion should be interpreted with care since optimality, here, is defined as a function of the PPP flux, which has no physiological sense *per se*. Nevertheless, the flux data point to the lack of correlation between G6PDH concentration and G6PDH flux, indicating that the actual PPP flux is not determined by genetic control but rather by metabolic (Becker et al., 2007; Moritz et al., 2000) or post-transcriptional control (Revelles et al., 2013). In the  $\Delta zwf$  mutant, fluxes through the non-oxidative PPP were reversed in comparison to the WT strain (Supplementary Table S6), as reported in previous studies (Hua et al., 2003; Nicolas et al., 2007; Zhao et al., 2004a; Zhao et al., 2004b). This was also observed in strains Pzwf 1.1 to Pzwf 1.3. This indicates that when the flux through oxidative PPP decreases, the synthesis of PPP-derived compounds is

compensated for by the reverse activity of the non-oxidative PPP. Together these data strongly suggests that *E. coli* uses parallel strategies (i.e. upkeep of a high G6PDH flux and activation of the non-oxidative PPP in that particular case) to maintain growth upon a single genetic perturbation. This is in agreement with previous studies showing the robustness of *E. coli* metabolic network against changes in its structure (Behre et al., 2008; Ishii et al., 2007; Kohlstedt et al., 2010).

Overall, the flux data are consistent with the multivariate analysis of  $^{13}\text{C}$  enrichment data, but go one step further by elucidating the pathways that support metabolic flexibility in *E. coli* towards modulation of a single enzyme.

## 5. Conclusion

The integrated fluxome platform presented here is a first but critical step towards robotic screening of quantitative metabolic phenotypes. Due to the parallelization and automation of the workflow, significant improvements in throughput, robustness and release of resources were achieved. The 80 flux data collected on *E. coli* G6PDH-modulation mutants were generated in four days, excluding preparation time but including all automated and manual steps from cultivation to calculated fluxes. This means that 20 flux data can be generated per day, which, to the best of our knowledge, is a major advance in the field. Moreover, the data are obtained from cultivations performed in tightly controlled physiological conditions, which is of critical importance for the biological meaning of the data. However, although the approach used here to deduce metabolic flux can be used for a broad range of microbes, it is not sufficient to resolve fluxes in mammalian cells because of their metabolic complexity (i.e. co-substrate consumption, slow metabolite turnover, metabolite channeling between compartments) (Mueller and Heinzle, 2013). However, obtaining labeling information on free metabolites

rather than on amino acids is possible using this platform. The robot includes an automatic process, not described here, for metabolite harvesting and extraction that is sufficiently rapid to quench metabolic activity, so that isotope patterns can be measured by MS on free intermediary metabolites. In addition to pave the way for solving fluxes in mammalian cells in a high throughput manner, this MS labelling data can also be used to complement the NMR labelling data. The opportunity to directly connect the data flow between cell cultivation and quantification of fluxes while bringing it significantly closer to being a high-throughput technology adds a new dimension to fluxomics. This will definitely be a valuable framework for the large-scale analysis of mutants, the development and metabolic optimization of production cells, the large-scale analysis of the effects of biological effectors, of drugs, among others. Such an advance would benefit a wide range of fields including not only metabolic engineering, synthetic biology and biotechnology but also toxicology and pharmacology, biomarker and drug development.

## Acknowledgements

MetaToul (Metabolomics & Fluxomics Facilities, Toulouse, France, [www.metatoul.fr](http://www.metatoul.fr)) and its staff members Edern Cahoreau and Lindsay Peyriga are gratefully acknowledged for technical support and access to NMR. MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, [www.metabohub.fr](http://www.metabohub.fr)). MetaToul is supported by Grants from the Région Midi-Pyrénées, the European Regional Development Fund, the SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). We thank Cecile Nicolas for providing the *E. coli* mutants. This work was supported by INRA.

5

10

15

## References

- 20 Amador-Noguez, D., Feng, X. J., Fan, J., Roquet, N., Rabitz, H., Rabinowitz, J. D., 2010.  
Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic  
acid cycle in *Clostridium acetobutylicum*. *J. Bacteriol.* 192, 4452-61.



- Antoniewicz, M. R., Kelleher, J. K., Stephanopoulos, G., 2007. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68-86.
- Balcarcel, R. R., Clark, L. M., 2003. Metabolic screening of mammalian cell cultures using well-plates. *Biotechnology progress.* 19, 98-108.
- 5 Becker, J., Klopprogge, C., Herold, A., Zelder, O., Bolten, C. J., Wittmann, C., 2007. Metabolic flux engineering of L-lysine production in *Corynebacterium glutamicum*--over expression and modification of G6P dehydrogenase. *Journal of biotechnology.* 132, 99-109.
- 10 Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C., Nicholson, J. K., 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols.* 2, 2692-703.
- Behre, J., Wilhelm, T., von Kamp, A., Rupp, E., Schuster, S., 2008. Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of theoretical biology.* 15 252, 433-41.
- Betts, J. I., Baganz, F., 2006. Miniature bioreactors: current practices and future opportunities. *Microb. Cell. Fact.* 5, 21.
- Blank, L. M., Kuepfer, L., Sauer, U., 2005. Large-scale <sup>13</sup>C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* 6, 20 R49.
- Boisseau, R., Charrier, B., Massou, S., Portais, J. C., Akoka, S., Giraudeau, P., 2013. Fast Spatially Encoded 3D NMR Strategies for C-Based Metabolic Flux Analysis. *Anal. Chem.*
- 25 Cahoreau, E., Peyriga, L., Hubert, J., Bringaud, F., Massou, S., Portais, J. C., 2012. Isotopic profiling of <sup>13</sup>C-labeled biological samples by two-dimensional heteronuclear J-resolved nuclear magnetic resonance spectroscopy. *Anal. Biochem.* 427, 158-63.
- Ellis, D. I., Goodacre, R., 2012. Metabolomics-assisted synthetic biology. *Curr. Opin. Biotechnol.* 23, 22-8.
- 30 Fan, T. W. M., Lane, A. N., 2008. Structure-based profiling of metabolites and isotopomers by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 52, 69.
- Feng, X., Page, L., Rubens, J., Chircus, L., Colletti, P., Pakrasi, H. B., Tang, Y. J., 2010. Bridging the gap between fluxomics and industrial biotechnology. *J. Biomed. Biotechnol.* 2010, 460717.
- 35 Fischer, E., Sauer, U., 2003. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur. J. Biochem.* 270, 880-91.
- Fischer, E., Sauer, U., 2005. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics.* 37, 636-40.
- Fischer, E., Zamboni, N., Sauer, U., 2004. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived <sup>13</sup>C constraints. *Anal. Biochem.* 325, 40 308-16.
- Ge, X., Hanson, M., Shen, H., Kostov, Y., Brorson, K. A., Frey, D. D., Moreira, A. R., Rao, G., 2006. Validation of an optical sensor-based high-throughput bioreactor system for mammalian cell culture. *Journal of biotechnology.* 122, 293-306.
- 45 Girard, P., Jordan, M., Tsao, M., Wurm, F. M., 2001. Small-scale bioreactor system for process development and optimization. *Biochem Eng J.* 7, 117-119.
- Giraudeau, P., Cahoreau, E., Massou, S., Pathan, M., Portais, J. C., Akoka, S., 2012. UFJCSY: a fast 3D NMR method for measuring isotopic enrichments in complex samples. *Chemphyschem : a European journal of chemical physics and physical chemistry.* 13, 3098-101.

- Giraudeau, P., Massou, S., Robin, Y., Cahoreau, E., Portais, J. C., Akoka, S., 2011. Ultrafast quantitative 2D NMR: an efficient tool for the measurement of specific isotopic enrichments in complex biological mixtures. *Anal. Chem.* 83, 3112-9.
- Hanson, B. A., ChemoSpec: Exploratory Chemometrics for Spectroscopy. R package version 1.51-2. 2012.
- Haverkorn van Rijsewijk, B. R., Nanchen, A., Nallet, S., Kleijn, R. J., Sauer, U., 2011. Large-scale <sup>13</sup>C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol. Syst. Biol.* 7, 477.
- Heux, S., Philippe, B., Portais, J. C., 2011. High-throughput workflow for monitoring and mining bioprocess data and its application to inferring the physiological response of *Escherichia coli* to perturbations. *Appl. Environ. Microbiol.* 77, 7040-9.
- Hortsch, R., Krispin, H., Weuster-Botz, D., 2011. Process performance of parallel bioreactors for batch cultivation of *Streptomyces tendae*. *Bioprocess. Biosyst. Eng.* 34, 297-304.
- Hua, Q., Yang, C., Baba, T., Mori, H., Shimizu, K., 2003. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacteriol.* 185, 7053-67.
- Huber, R., Ritter, D., Hering, T., Hillmer, A. K., Kensy, F., Muller, C., Wang, L., Buchs, J., 2009. Robo-Lector - a novel platform for automated high-throughput cultivations in microtiter plates with high information content. *Microb. Cell. Fact.* 8, 42.
- Hüsemann, M. W., Papoutsakis, E., 1989. Comparison between in vivo and in vitro enzyme activities in continuous and batch fermentations of *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* 30, 585-595.
- Isett, K., George, H., Herber, W., Amanullah, A., 2007. Twenty-four-well plate miniature bioreactor high-throughput system: assessment for microbial cultivations. *Biotechnol. Bioeng.* 98, 1017-28.
- Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., Tomita, M., 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*. 316, 593-7.
- Knorr, B., Schlieker, H., Hohmann, H.-P., Weuster-Botz, D., 2007. Scale-down and parallel operation of the riboflavin production process with *Bacillus subtilis*. *Biochem. Eng. J.* 33, 263-274.
- Kocincova, A. S., Nagl, S., Arain, S., Krause, C., Borisov, S. M., Arnold, M., Wolfbeis, O. S., 2008. Multiplex bacterial growth monitoring in 24-well microplates using a dual optical sensor for dissolved oxygen and pH. *Biotechnol. Bioeng.* 100, 430-8.
- Kohlstedt, M., Becker, J., Wittmann, C., 2010. Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism. *Appl. Microbiol. Biotechnol.* 88, 1065-75.
- Kusterer, A., Krause, C., Kaufmann, K., Arnold, M., Weuster-Botz, D., 2008. Fully automated single-use stirred-tank bioreactors for parallel microbial cultivations. *Bioprocess. Biosyst. Eng.* 31, 207-15.
- Legmann, R., Schreyer, H. B., Combs, R. G., McCormick, E. L., Russo, A. P., Rodgers, S. T., 2009. A predictive high-throughput scale-down model of monoclonal antibody production in CHO cells. *Biotechnol. Bioeng.* 104, 1107-20.
- Lenz, E. M., Wilson, I. D., 2007. Analytical strategies in metabolomics. *Journal of proteome research*. 6, 443-58.

- Massou, S., Nicolas, C., Letisse, F., Portais, J. C., 2007a. Application of 2D-TOCSY NMR to the measurement of specific  $^{13}\text{C}$ -enrichments in complex mixtures of  $^{13}\text{C}$ -labeled metabolites. *Metab. Eng.* 9, 252-7.
- Massou, S., Nicolas, C., Letisse, F., Portais, J. C., 2007b. NMR-based fluxomics: quantitative 2D NMR methods for isotopomers analysis. *Phytochemistry*. 68, 2330-40.
- Meynial-Salles, I., Cervin, M. A., Soucaille, P., 2005. New tool for metabolic pathway engineering in *Escherichia coli*: one-step method to modulate expression of chromosomal genes. *Appl. Environ. Microbiol.* 71, 2140-4.
- Millard, P., Letisse, F., Sokol, S., Portais, J. C., 2012. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics*. 28, 1294-6.
- Moritz, B., Striegel, K., De Graaf, A. A., Sahm, H., 2000. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux in vivo. *Eur. J. Biochem.* 267, 3442-52.
- Mueller, D., Heinzle, E., 2013. Stable isotope-assisted metabolomics to detect metabolic flux changes in mammalian cell cultures. *Curr. Opin. Biotechnol.* 24, 54-9.
- Munger, J., Bennett, B. D., Parikh, A., Feng, X. J., McArdle, J., Rabitz, H. A., Shenk, T., Rabinowitz, J. D., 2008. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nature biotechnology*. 26, 1179-86.
- Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H., Tomita, M., 2009. Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol. Syst. Biol.* 5, 306.
- Nanchen, A., Schicker, A., Sauer, U., 2006. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl. Environ. Microbiol.* 72, 1164-72.
- Nicolas, C., Kiefer, P., Letisse, F., Krömer, J., Massou, S., Soucaille, P., Wittmann, C., Lindley, N. D., Portais, J. C., 2007. Response of the central metabolism of *Escherichia coli* to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase. *FEBS Lett.* 581, 3771-6.
- Peng, L., 2012. Tandem Mass Spectrometry: A New Platform for Fluxomics. *J Proteomics Bioinform.* 5.
- Poskar, C. H., Huege, J., Krach, C., Franke, M., Shachar-Hill, Y., Junker, B. H., 2012. iMS2Flux--a high-throughput processing tool for stable isotope labeled mass spectrometric data used for metabolic flux analysis. *BMC Bioinf.* 13, 295.
- Puskeiler, R., Kaufmann, K., Weuster-Botz, D., 2005. Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD). *Biotechnol. Bioeng.* 89, 512-23.
- Quek, L. E., Wittmann, C., Nielsen, L. K., Krömer, J. O., 2009. OpenFLUX: efficient modelling software for  $^{13}\text{C}$ -based metabolic flux analysis. *Microb. Cell. Fact.* 8, 25.
- R Development Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2009.
- Raghevendran, V., Gombert, A. K., Christensen, B., Kotter, P., Nielsen, J., 2004. Phenotypic characterization of glucose repression mutants of *Saccharomyces cerevisiae* using experiments with  $^{13}\text{C}$ -labelled glucose. *Yeast*. 21, 769-79.
- Revelles, O., Millard, P., Nougayrede, J. P., Dobrindt, U., Oswald, E., Letisse, F., Portais, J. C., 2013. The Carbon Storage Regulator (Csr) System Exerts a Nutrient-Specific Control over Central Metabolism in *Escherichia coli* Strain Nissle 1917. *PloS one*. 8, e66386.

- Sanford, K., Soucaille, P., Whited, G., Chotani, G., 2002. Genomics to fluxomics and physiomics - pathway engineering. *Curr. Opin. Microbiol.* 5, 318-22.
- Sauer, U., 2004. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.* 15, 58-63.
- 5 Sauer, U., 2006. Metabolic networks in motion:  $^{13}\text{C}$ -based flux analysis. *Mol. Syst. Biol.* 2, 62.
- Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., Fischer, E., 2004. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J. Biol. Chem.* 279, 6613-9.
- 10 Schmidt, K., Carlsen, M., Nielsen, J., Villadsen, J., 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55, 831-40.
- Sokol, S., Millard, P., Portais, J. C., 2012. influx\_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics.* 15 28, 687-93.
- Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. *Metab. Eng.* 1, 1-11.
- Szyperski, T., 1995. Biosynthetically directed fractional  $^{13}\text{C}$ -labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. *Eur. J. Biochem.* 232, 433-48.
- 20 Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. E., Keasling, J. D., 2009. Advances in analysis of microbial metabolic fluxes via  $^{13}\text{C}$  isotopic labeling. *Mass spectrometry reviews.* 28, 362-75.
- Toya, Y., Ishii, N., Nakahigashi, K., Hirasawa, T., Soga, T., Tomita, M., Shimizu, K., 2010.  $^{13}\text{C}$ -metabolic flux analysis for batch culture of *Escherichia coli* and its Pyk and Pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology progress.* 26, 975-92.
- 25 Valgepea, K., Adamberg, K., Seiman, A., Vilu, R., 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Molecular bioSystems.* 9, 2344-58.
- 30 van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., van der Werf, M. J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics.* 7, 142.
- Varmuza, K., Filzmoser, P., 2009. Introduction to multivariate statistical analysis in chemometrics.
- 35 Weitzel, M., Nöh, K., Dalman, T., Niedenfuhr, S., Stute, B., Wiechert, W., 2012. 13CFLUX2 - High-Performance Software Suite for  $^{13}\text{C}$ -Metabolic Flux Analysis. *Bioinformatics.*
- Wiechert, W., 2001.  $^{13}\text{C}$  metabolic flux analysis. *Metab. Eng.* 3, 195-206.
- Wiechert, W., Nöh, K., 2005. From stationary to instationary metabolic flux analysis. *Adv. Biochem. Eng. Biotechnol.* 92, 145-72.
- 40 Wittmann, C., Kim, H. M., Heinzle, E., 2004. Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. *Biotechnol. Bioeng.* 87, 1-6.
- Wittmann, C., Portais, J.-C., 2013. Metabolic Flux Analysis. *Metabolomics in Practice.* Wiley-VCH Verlag GmbH & Co. KGaA, pp. 285-312.
- Zamboni, N., Fischer, E., Sauer, U., 2005. FiatFlux--a software for metabolic flux analysis from  $^{13}\text{C}$ -glucose experiments. *BMC Bioinf.* 6, 209.
- 45 Zamboni, N., Sauer, U., 2004. Model-independent fluxome profiling from  $^2\text{H}$  and  $^{13}\text{C}$  experiments for metabolic variant discrimination. *Genome Biol.* 5, R99.
- Zhao, J., Baba, T., Mori, H., Shimizu, K., 2004a. Effect of zwf gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metab. Eng.* 6, 164-74.

Zhao, J., Baba, T., Mori, H., Shimizu, K., 2004b. Global metabolic response of *Escherichia coli* to *gnd* or *zwf* gene-knockout, based on <sup>13</sup>C-labeling experiments and the measurement of enzyme activities. *Appl. Microbiol. Biotechnol.* 64, 91-8.

## Supplementary material captions

As noted in the text, supplementary materials are available in the online version of this paper.

Supplementary file 1 contains supplementary Figures S1 to S2 and Notes. Figure S1 shows the

5 workflow of the automated cultivation with  $^{13}\text{C}$ -labeled substrate and sampling of labeled biomass using the robotic platform. Figure S2 shows the hierarchical cluster analysis of the  $^{13}\text{C}$  enrichment data. Supplementary Note 1 describes the protocol used for automated cultivation with  $^{13}\text{C}$ -labeled substrate and sampling of labeled biomass using the robotic platform. Supplementary Note 2 describes the calculation of confidence interval for the flux estimation.

10 Supplementary file 2 contains supplementary Tables S1-S6. Table S1 is the list of strains used in this study. Table S2 is the NMR assignment database. Table S3 is the list of equality constraints used with the AEDE tool. Table S4 contains the reproducibility of the NMR spectra between the biological replicates. Table S5 is the comparison of the  $^{13}\text{C}$  enrichment between manual and AEDE integration. Table S6 contains the relative and absolute flux data and their  
15 standard deviations.

Supplementary file 3 contains copies of the FTBLs used for flux calculations.

## Figure captions

Figure 1: Integrated platform for fluxome profiling. **A:** *E. coli* cells were grown and harvested on a robotic platform. This platform allowed the preparation, running, monitoring (growth, pH and oxygen) and control (pH, temperature, stirrer speed) of 48 parallel micro-scale (10 ml) cultivations with  $^{13}\text{C}$ -labelled substrates. An automated process allowed harvesting of biomass at targeted optical density ( $\text{OD}_{600}$ ) and preparation of samples (i.e. supernatant and cell pellets) for further analysis. Examples of offline measurement of growth (OD at 600 nm) and online measurement of pH and oxygen during batch cultivation of a wild type (WT) *E. coli* strain on  $^{13}\text{C}$ -labelled glucose in the bioreaction block ( $V = 15 \text{ ml}$ ,  $T = 37^\circ\text{C}$ , stirred at 2200 rpm) are presented. Mean and standard deviations (red area) of 4 replicates are shown. **B:** Labeling patterns were measured on a high resolution (800 MHz) NMR spectrometer equipped with a cryoprobe and an autosampler. A representative 1D- $^1\text{H}$ -NMR spectrum of labeled biomass hydrolysates (1.5 mg of dry biomass) obtained from WT *E. coli* cells grown on labeled glucose is shown. **C:** Spectral data were automatically processed using a program developed in R. This allows automatic quantification of  $^{13}\text{C}$  enrichments from the spectra by incorporating  $^1\text{H}$ - $^{13}\text{C}$  and  $^1\text{H}$ - $^{12}\text{C}$  signals using triangle modeling. A subset of 4 protonated carbon positions is shown as an example. Triangles for  $^1\text{H}$ - $^{13}\text{C}$  and  $^1\text{H}$ - $^{12}\text{C}$  signals of the corresponding carbon position are dark gray and triangles for others signals are pale gray. **D:** Processed data are delivered in formats suitable for statistical and flux analysis.

Figure 2: Physiological parameters on glucose and xylose of the metabolic variants of *E. coli*.

The 10 *E. coli* strains with gradual expression of glucose-6-phosphate dehydrogenase (Supplementary Table S1) and the wild type strain were grown on glucose and xylose in the bioreaction block ( $V = 15 \text{ ml}$ ,  $T = 37^\circ\text{C}$ , stirring at 2200 rpm). **A:** Growth rate in  $\text{h}^{-1}$ . **B:** Specific substrate uptake rate in  $\text{mmol/gdw/h}$ . **C:** Specific acetate secretion rate in  $\text{mmol/gdw/h}$ .

**D:** Specific oxygen uptake rate in mmol/gdw/h. **E:** OD600 at harvesting point. Mean and standard deviation of four replicates per condition are shown.

Figure 3: Statistical analysis of labeling patterns of proteinogenic amino acids from the metabolic variants of *E. coli*. **A:** Biplot of the  $^{13}\text{C}$  enrichments for 23 carbon positions of amino acids and metabolic precursors across the 80 conditions using principal component analysis. **B:**  $^{13}\text{C}$  enrichment of protonated carbon positions showing significant changes (pvalue <0.001, Two way ANOVA) between strains and carbon sources. See abbreviations in Supplementary Table S2.

Figure 4: Fluxes on glucose and xylose of the metabolic variants of *E. coli*. **A:** Distribution of relative fluxes of key metabolic pathways in 20 (10 strains grown on two different substrates) conditions. Fluxes were normalized by substrate uptake. **B:** Absolute values for selected metabolic fluxes plotted as a function of glucose-6-phosphate dehydrogenase (G6PDH) activity relative to the wild type strain (WT). Abbreviations and the standard deviation of each flux are listed in Supplementary Table S4. Dashed line indicates the WT level