

## Metabolic footprint analysis of recombinant *Escherichia coli* strains during fed-batch fermentations†

Sónia Carneiro,<sup>a</sup> Silas G. Villas-Bôas,<sup>b</sup> Eugénio C. Ferreira<sup>a</sup> and Isabel Rocha<sup>\*a</sup>

Received 3rd August 2010, Accepted 17th November 2010

DOI: 10.1039/c0mb00143k

Metabolic footprinting has become a valuable analytical approach for the characterization of phenotypes and the distinction of specific metabolic states resulting from environmental and/or genetic alterations. The metabolic impact of heterologous protein production in *Escherichia coli* cells is of particular interest, since there are numerous cellular stresses triggered during this process that limit the overall productivity, e.g. the stringent response. Because the knowledge on the metabolic responses in recombinant bioprocesses is still scarce, metabolic footprinting can provide relevant information on the intrinsic metabolic adjustments. Thus, the metabolic footprints generated by *E. coli* W3110 and the  $\Delta relA$  mutant strain during recombinant fed-batch fermentations at different experimental conditions were measured and interpreted. The IPTG-induction of the heterologous protein expression resulted in the rapid accumulation of inhibitors of the glyoxylate shunt in the culture broth, suggesting the clearance of this anaplerotic route to replenish the TCA intermediaries withdrawn for the additional formation of the heterologous protein. Nutritional shifts were also critical in the recombinant cellular metabolism, indicating that cells employ diverse strategies to counteract imbalances in the cellular metabolism, including the secretion of certain metabolites that are, most likely, used as a metabolic relief to survival processes.

### Introduction

The optimization of bioprocesses using recombinant microorganisms is still restrained by the lack of information available on the metabolic responses induced by various stress conditions.<sup>1</sup> Significant knowledge could be gained from a comprehensive analysis of the metabolic footprint (i.e. extracellular metabolite profiling) by inspecting key metabolic changes and understanding their relation with environmental conditions.<sup>2–8</sup>

Some work has been published addressing various metabolic responses during the production of heterologous proteins in *Escherichia coli*.<sup>9</sup> Experimental studies showed that the host cell metabolism undergoes a severe metabolic burden, resulting in rapid exhaustion of essential precursors and cellular energy.<sup>10</sup> Typically, strong expression systems are

employed to assure the production of large amounts of heterologous proteins by the host, which uses a large quantity of metabolic and energy resources in order to maintain and express the foreign DNA.<sup>11</sup>

Heterologous protein production is also believed to diminish flow in the TCA cycle through the withdrawal of intermediates that serve as precursors for amino acid biosynthesis.<sup>12</sup> Moreover, the difference usually observed in amino acid composition of foreign proteins and the average composition of the host proteins contributes to this metabolic imbalance.<sup>13–16</sup>

Stringent response has been associated with the stress phenomenon caused by the depletion of certain metabolic resources, namely amino acids.<sup>17</sup> The increased level of free tRNA molecules, due to the lack of amino acids, triggers this stress response that is characterized by the arrest of the ribosomal translation process and the rapid RelA-mediated accumulation of ppGpp.<sup>18</sup> This nucleotide has been found<sup>19–21</sup> to bind directly to the RNA polymerase, adjusting the transcriptional activity from the expression of genes required for rapid growth, to stress-related genes and amino acid biosynthetic operons.<sup>22–25</sup> The regulatory mechanisms of this ppGpp-induced stress response are known in some detail,<sup>26–28</sup> but the impact of this response on the cellular metabolism has been less studied. Knowledge on how these responses take place and how to dodge them is of great importance, since *E. coli* has become one of the most used microbial systems to produce heterologous proteins.

<sup>a</sup> IBB—Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.  
E-mail: [irocha@deb.uminho.pt](mailto:irocha@deb.uminho.pt); Fax: +351 253 678 986;  
Tel: +351 253 604 408

<sup>b</sup> School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.  
E-mail: [s.villas-boas@auckland.ac.nz](mailto:s.villas-boas@auckland.ac.nz); Fax: +64 9 3737416;  
Tel: +64 9 3737599

† Electronic supplementary information (ESI) available: ESI 1—list of extracellular metabolites detected by GC-MS; ESI 2—table with the peak intensity value corrected for the recovery of the internal standard (D-4-alanine) and normalized with the corresponding biomass concentration. See DOI: 10.1039/c0mb00143k

Here, we aimed at investigating the physiological and metabolic changes in *E. coli* cultures during the production of heterologous proteins by performing a metabolic footprinting analysis. Furthermore, the focus of the study was not only to evaluate the changes of the extracellular metabolite pools during heterologous protein production, but also to assess the effect of removing a gene closely related to the initiation of the stringent response (*relA*) on the cellular behaviour. Thus, the W3110 and the isogenic mutant ( $\Delta relA$ ) *E. coli* strains were grown and induced to produce a heterologous protein (AcGFP1) at different nutritional conditions in a controlled fed-batch mode.

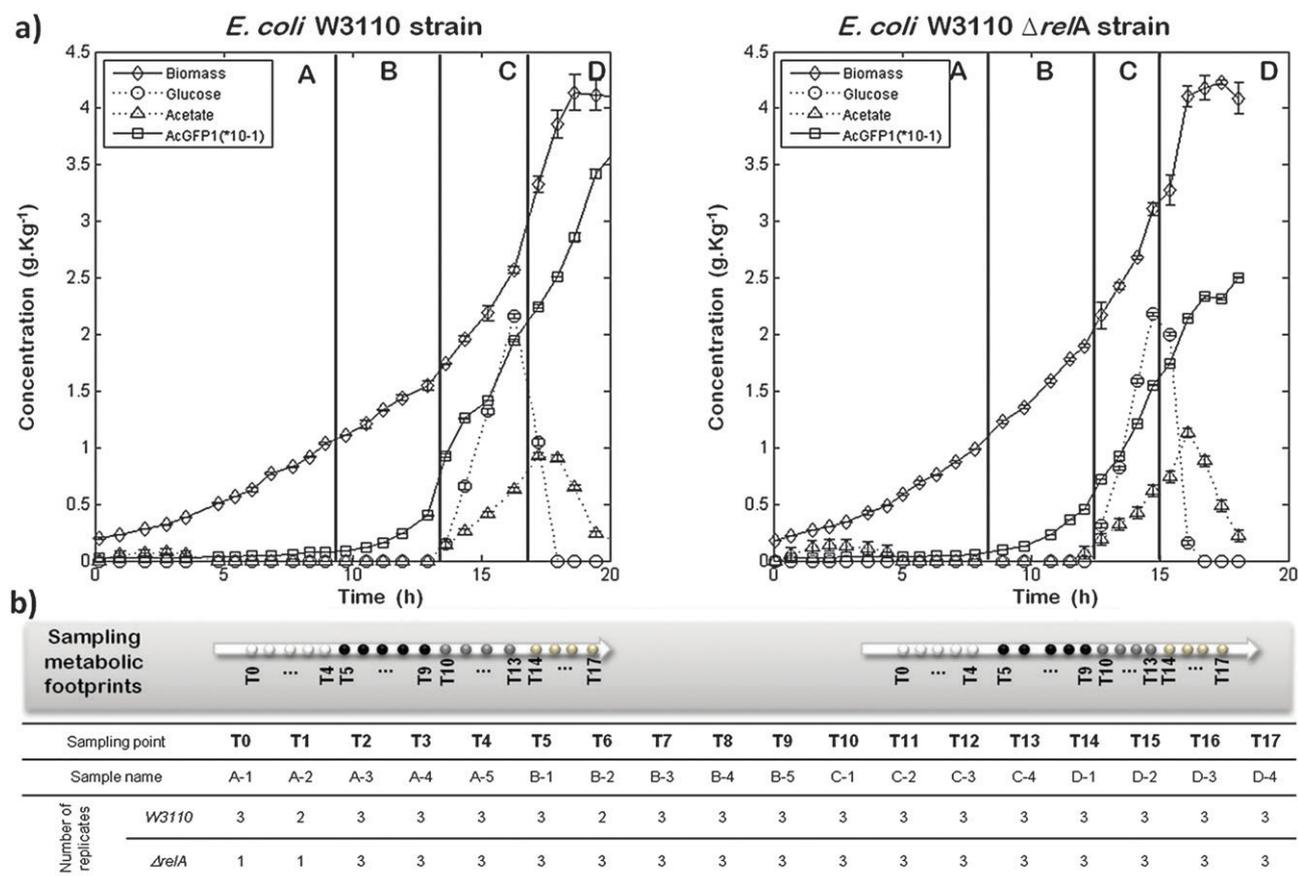
## Results

### Fed-batch fermentations of W3110 and $\Delta relA$ *E. coli* cells

To elucidate the physiological responses of the W3110 and  $\Delta relA$  mutant *E. coli* strains during recombinant fed-batch fermentation, cells were grown aerobically with a closed-loop feeding control to maintain a quasi-steady state growth. Two fed-batch cultures, one with each strain, were started with a low specific growth rate (between 0.09 and 0.16 h<sup>-1</sup>) and growth characteristics were determined prior and after IPTG induction (phases A and B, respectively). Then, the glucose feeding rate

was increased to maintain the specific growth rate around 0.2 h<sup>-1</sup> (phase C) to evaluate the impact of nutritional upshift on the *E. coli* cultures during heterologous protein production. Nutrient limitation by ceasing the glucose feeding (phase D) was finally examined in these *E. coli* cultures until growth is arrested.

In Fig. 1, biomass, glucose, acetate and AcGFP1 concentrations are depicted for each fed-batch culture. Table 1 shows the growth parameters determined at each experimental condition. As demonstrated, prior to IPTG induction the specific growth rates were similar for both strains (around 0.16 h<sup>-1</sup>), but after IPTG induction the growth of the wild-type strain was significantly reduced (0.092 h<sup>-1</sup>) while the  $\Delta relA$  mutant had a specific growth rate of 0.13 h<sup>-1</sup>. In contrast, the AcGFP1 production increased after IPTG induction to a maximum of around  $7 \times 10^{-3}$  g g<sup>-1</sup> DW h<sup>-1</sup> for both cultures. Upon the nutritional upshift (phase C), both strains increased their growth rates and the AcGFP1 production rates. The maximum AcGFP1 production was determined to be close to  $20 \times 10^{-3}$  g g<sup>-1</sup> DW h<sup>-1</sup> in both cultures. At these conditions acetate accumulation was detected in both cultures at rates below  $90 \times 10^{-3}$  g g<sup>-1</sup> DW h<sup>-1</sup>. When the glucose feeding was stopped (phase D) acetate was consumed and growth was arrested. However, AcGFP1 production continued until the carbon sources (*i.e.* glucose and acetate) were completely depleted from the medium.



**Fig. 1** (a) Growth curves for the W3110 [pTRC-His-AcGFP1] and the isogenic derivative  $\Delta relA251::kan$  *E. coli* strains cultured in fed-batch fermentations with closed-loop feeding control. (b) Sampling scheme for GC-MS analysis. Samples were harvested at regular intervals during the four fermentation phases: A—prior to induction; B—after IPTG induction; C—after growth upshift; and D—after nutrient downshift. The first sample (A-1) was taken when the biomass concentration was around 0.6 g kg<sup>-1</sup> (corresponding to sampling point T0) and then samples were collected every 40–50 minutes (corresponding to sampling points T1 to T17). Time series were similarly defined in both fed-batch fermentations.

**Table 1** Growth parameters of the fed-batch cultures of the W3110 and  $\Delta relA$  *E. coli* strains. 95% confidence intervals are indicated after the symbol  $\pm$ 

Strain	Phases	Specific growth rate set point/h <sup>-1</sup> $\mu$	Expected biomass	Specific rates/g g <sup>-1</sup> DW h <sup>-1</sup>				Biomass yield/g <sub>DW</sub> g <sup>-1</sup> substrate
			yield/g <sub>DW</sub> g <sup>-1</sup> substrate $Y_{X/S}$	$\mu$	$q_{AcGFP1} (\times 10^{-3})$	$q_{Gluc}$	$q_{Ac}$	$Y_{X/S}$
W3110	A	0.1	0.35	0.16 $\pm$ 0.020	0.60 $\pm$ 0.12	0.30 $\pm$ 0.067	—	0.59 $\pm$ 0.038
	B	0.1	0.35	0.092 $\pm$ 0.023	6.8 $\pm$ 4.5	0.29 $\pm$ 0.052	—	0.36 $\pm$ 0.065
	C	0.2	0.2	0.17 $\pm$ 0.023	19 $\pm$ 6.0	0.64 $\pm$ 0.11	0.089 $\pm$ 0.0074	0.26 $\pm$ 0.068
	D	—	—	—	5.3 $\pm$ 5.8	0.24 $\pm$ 0.043	-0.49 $\pm$ 0.27	0.31 $\pm$ 0.069 <sup>a</sup>
$\Delta relA$	A	0.1	0.35	0.16 $\pm$ 0.019	0.60 $\pm$ 0.45	0.21 $\pm$ 0.016	—	0.62 $\pm$ 0.059
	B	0.1	0.35	0.13 $\pm$ 0.026	6.9 $\pm$ 3.8	0.25 $\pm$ 0.036	—	0.48 $\pm$ 0.022
	C	0.2	0.2	0.20 $\pm$ 0.030	17 $\pm$ 1.8	0.56 $\pm$ 0.089	0.084 $\pm$ 0.0072	0.36 $\pm$ 0.010
	D	—	—	—	14 $\pm$ 2.1	0.47 $\pm$ 0.14	-0.92 $\pm$ 0.23	0.04 $\pm$ 0.049 <sup>a</sup>

Specific growth rate ( $\mu$ ), AcGFP1 production ( $q_{AcGFP1}$ ), glucose uptake ( $q_{Gluc}$ ) and acetate formation ( $q_{Ac}$ ) specific rates.<sup>a</sup> These parameters were calculated for biomass growth under acetate consumption.

### Metabolic footprint analysis

Gas chromatography–mass spectrometry (GC-MS) has been widely used in the analysis of a large number of compounds such as amino acids, sugars, and organic acids. In this work, a GC-MS-based analytical platform established by Villas-Bôas' group<sup>29</sup> (detailed in the Experimental section) was used to detect amino and nonamino organic acids secreted by *E. coli* cells during fed-batch cultures. Samples taken for metabolic footprinting were harvested at regular time intervals and the amount of samples taken in each growth phase was consistent between fed-batch fermentations (Fig. 1). Time series data were further examined to address any changes in the metabolic footprints resulting from the alteration of culture conditions and to verify if the *relA* gene mutation could actually influence the metabolic behaviour of this recombinant *E. coli* strain.

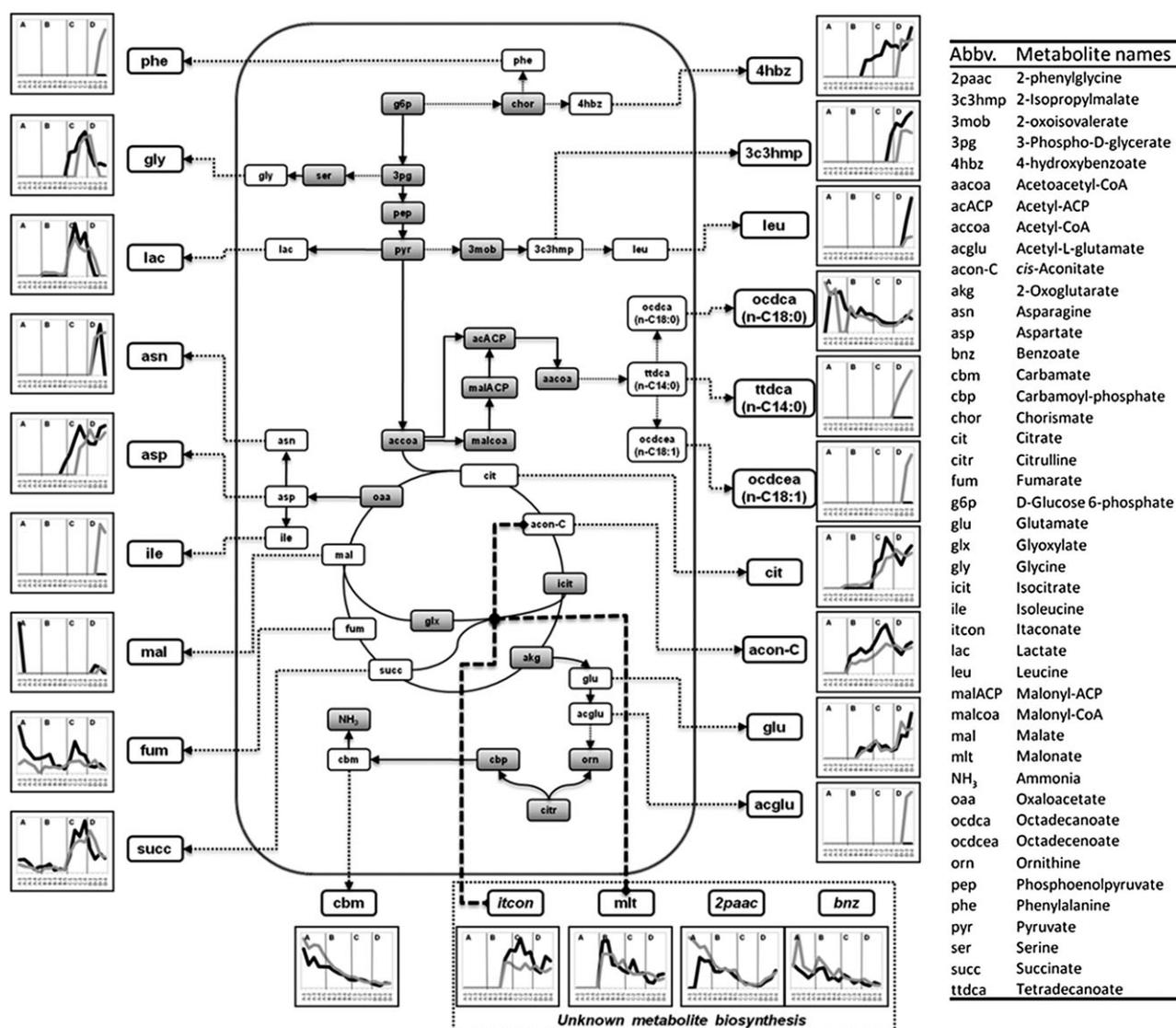
A wide variety of metabolites was detected in the extracellular medium, including fatty, amino and organic acids (see Fig. 2 and ESI 1<sup>†</sup>). These metabolites could either accumulate in the extracellular medium *via* secretion/excretion by cells or by means of the accumulation of components from the feeding media that were not completely exhausted, which is only the case of isoleucine (see media composition in the Materials and Methods section). Some of these metabolites were uniquely detected in the fed-batch fermentation with *E. coli*  $\Delta relA$  (e.g., phenylalanine (phe), acetyl-L-glutamate (acglu), octadecenoate (ocdcea) and tetradecanoate (ttcca)), which indicates that metabolic differences between the two fed-batch experiments might exist. These metabolites were found to be mainly involved in the central carbon metabolism, including the tricarboxylic acid cycle (TCA cycle), the biosynthesis of amino and fatty acids as well as other energy generating metabolic reactions. However, we detected metabolites such as itaconate, malonate, 2-phenylglycine and benzoate that could not be linked directly to any known metabolic pathway of *E. coli*, according to public databases (such as KEGG<sup>30</sup> and EcoCyc)<sup>31</sup> and genome-scale metabolic models (e.g. iAF1260 model of *E. coli* K12).<sup>32</sup> However, these metabolites are known to participate in metabolic reactions of other organisms.<sup>33,34</sup> Therefore, resemblances between enzyme-coding genes from other organisms that produce these metabolites with the *E. coli* genome were investigated, but it was not possible to establish with confidence their participation as

reactants/products in *E. coli* reactions. Yet, itaconate and malonate have been described as *in vitro* inhibitors of a key enzyme of the *E. coli* anaplerotic metabolism, *i.e.*, isocitrate lyase, the first enzyme of the glyoxylate cycle,<sup>35</sup> indicating that at least these metabolites are likely to be produced during *E. coli* growth. Further information is available in the ESI 1.<sup>†</sup>

Principal components analysis (PCA) was initially performed to investigate whether the samples from different fermentation phases could be distinguished based on their metabolic footprint profiles and to determine the significant metabolic differences between the W3110 and  $\Delta relA$  mutant *E. coli* strains (Fig. 3).

To investigate these differences, metabolic footprint data from the two *E. coli* cultures were compared (Table 2). For each metabolite, the Pearson correlation coefficient ( $r$ ) was calculated for the average  $Z$ -score values in order to evaluate the similarities between metabolite profiles produced in each culture and to find which metabolites are likely to explain the differences depicted in Fig. 3. As shown in Table 2, the majority of metabolites presented Pearson's correlation coefficients that were found significant after performing Student's  $t$  test ( $p$ -values higher than 0.05). However, metabolites like 2-phenylglycine (2paac), fumarate (fum), malate (mal) and octadecanoate (ocdca) showed relatively lower coefficients, indicating that their profiles were distinct between the two *E. coli* cultures. Although there are few metabolites with uncorrelated profiles, differences in sample projections along PCs, in particular samples from phase D, might be associated with the presence of some metabolites that were only detected in one of the fed-batch cultures. For example, the metabolites isoleucine (ile), phenylalanine (phe), acetyl-L-glutamate (acglu), octadecenoate (ocdcea) and tetradecanoate (ttcca) were only detected in the fed-batch culture with the  $\Delta relA$  mutant strain during the last fermentation phase (*i.e.* phase D), indicating that cells were likely at different metabolic states and metabolic activities carried out by the two *E. coli* strains were distinct.

Although the differences found between the  $\Delta relA$  mutant and W3110 strain may result from changes associated with the genetic perturbation, we cannot ignore the intrinsic metabolic variability that is common to most systems. As reported previously,<sup>36</sup> fluctuations may arise because organisms are never in the same exact metabolic stage, even when growing



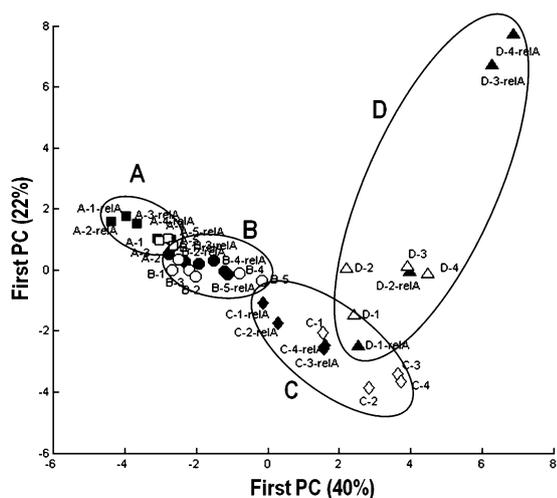
**Fig. 2** Schematic diagram of the *E. coli* metabolic map involving the metabolites secreted into the extracellular medium during recombinant fed-batch cultivations. The accumulation (or assimilation) of metabolites was evaluated along four phases: the pre-IPTG induction phase (A), the post-IPTG induction phase (B), the post-nutrient upshift phase (C) and the glucose limitation phase (D). Graphs represent the time series data relative to the concentration levels of each metabolite in the metabolic footprint along the fermentation process for the W3110 strain (dark line) and the  $\Delta relA$  mutant (light gray line) cultures. Bold dashed lines connecting metabolites and biochemical reactions indicate known inhibitory effects on those reaction-associated enzymes. Grey boxes represented in the metabolic map indicate other metabolites that participate in the metabolism, but were not detected in these experiments. The dashed square lists the detected metabolites with unknown biosynthetic reactions in *E. coli*.

in the same conditions, and small differences in enzyme concentrations may also affect metabolite concentrations. Thus, to better understand these differences, the metabolic footprints from each culture were further examined independently by clustering analyses (Fig. 4 and 5).

Hierarchical clustering analysis (HCA) was used to evaluate the metabolic footprints produced by cells during the fed-batch fermentations. As shown in Fig. 4a, for the W3110 culture, samples from the same fermentation phase, including sample replicates, were clustered together, which indicates that those samples have similar metabolite profiles. Metabolite clusters evidenced the association between metabolite patterns generated along the fermentation process. The starting hypothesis is that metabolites that show a similar variation are related, a relation

that conveys information about their proximity or function within the metabolic map. For example, the metabolite profiles of citrate (cit), aspartate (asp), 4-hydroxybenzoate (4hbz), cis-aconitate (acon-C) and itaconate (itcon) were clustered at these experimental conditions, meaning that in the W3110 *E. coli* culture these metabolites follow a close pattern.

Similarly, samples from the recombinant  $\Delta relA$  mutant were grouped into four major clusters (Fig. 5a). However, in this case, samples were clustered differently. Samples from phase A clustered together, but samples taken immediately after IPTG induction (triplicate samples from B-1 to B-3) clustered together with samples from phase A. The following samples from phase B: i.e. samples B-4 and B-5, including replicates, clustered



**Fig. 3** Principal component analysis (PCA) 2-dimensional projection of samples from fed-batch cultures of *E. coli* grown at different conditions based on mass fragment profiles of extracellular metabolites analysed by GC-MS. The Z-score values corresponding to sample replicates for each metabolite variable were averaged at each sampling point. A total of 36 average profiles were considered in the PCA analysis: 18 from samples taken from the W3110 culture (represented by opened symbols) and the same amount from the  $\Delta relA$  mutant culture (represented by full symbols). For each time series, samples represented by *squares* correspond to samples withdrawn during the pre-induction phase [A], *circles* to the post-induction phase [B], *diamonds* to the phase after growth upshift [C] and *triangles* to the phase after nutrient downshift [D]. Samples are also labelled with letters corresponding to the cultivation phase (A–D), a number that indicates the sampling sequence in each cultivation phase and, in some cases the designation “relA”, indicating that it is a sample from the  $\Delta relA$  mutant culture.

**Table 2** Pearson’s correlation coefficients ( $r$ ) of metabolite profiles from fed-batch cultures of *E. coli* W3110 and the  $\Delta relA$  mutant strain. The corresponding  $p$ -values were calculated by a Student  $t$ -test and metabolite profiles with  $p$ -values lower than 0.05 were considered significantly correlated

Metabolite	$r$	$p$ -value
2paac	0.3370	0.1714
3c3hmp	0.8020	$6.260 \times 10^{-5}$
4hbz	0.5875	$1.030 \times 10^{-2}$
acglu	<i>a</i>	<i>a</i>
acon-C	0.9031	$2.870 \times 10^{-7}$
asn	0.7488	$3.499 \times 10^{-4}$
asp	0.8199	$3.110 \times 10^{-5}$
bnz	0.7222	$7.124 \times 10^{-4}$
cbm	0.9568	$5.320 \times 10^{-10}$
cit	0.9115	$1.430 \times 10^{-7}$
fum	0.3664	0.1347
glu	0.8814	$1.350 \times 10^{-6}$
gly	0.7624	$2.347 \times 10^{-4}$
ile	<i>a</i>	<i>a</i>
itcon	0.8938	$5.820 \times 10^{-7}$
lac	0.9049	$2.490 \times 10^{-7}$
leu	0.9728	$1.380 \times 10^{-11}$
mal	0.01391	0.9563
mlt	0.8640	$3.800 \times 10^{-6}$
ocdca	0.09675	0.7025
ocdcea	<i>a</i>	<i>a</i>
phe	<i>a</i>	<i>a</i>
succ	0.7807	$1.316 \times 10^{-4}$
ttдца	<i>a</i>	<i>a</i>

<sup>a</sup> Undetected in the W3110 culture. See metabolite abbreviations in Fig. 2.

separately. This indicates that changes in the metabolic footprint profiles were not immediate after IPTG induction, as observed in the W3110 culture. Only in the late stage of this fermentation phase those changes in metabolite profiles were detectable. Samples taken immediately after nutrient downshift (triplicate samples D-1 and D-2) clustered with samples from phase C, which indicates that these samples have metabolic profiles more similar to samples from phase C than to those from phase D. This can be due to the fact that, at this sampling time, cells were still consuming the excess of glucose accumulated during phase C (see Fig. 1). The other samples from phase D clustered separately. In this fed-batch culture, metabolites were not clustered in the same groups as observed in the W3110 *E. coli* culture, which reflects the existence of important differences in the metabolic footprints produced by the  $\Delta relA$  mutant cells. For example, the 4-hydroxybenzoate (4hbz) is, at these conditions, clustered with asparagine (asn), 2-isopropylmalate (2paac) and tetradecanoate (ttдца).

Besides the observed differences between the metabolic footprints produced at particular growth conditions, differences between the set of metabolic variables that better characterized each phase (*i.e.* loading profiles represented as vectors) were also explored. Thus, in the next subsections details are exposed on the key metabolites involved in the alteration of the metabolic footprints when cells were: (I) induced to express the heterologous protein; (II) submitted to the nutritional upshift; and (III) submitted to glucose-limited conditions.

### I Impact of IPTG induction on the metabolic footprint.

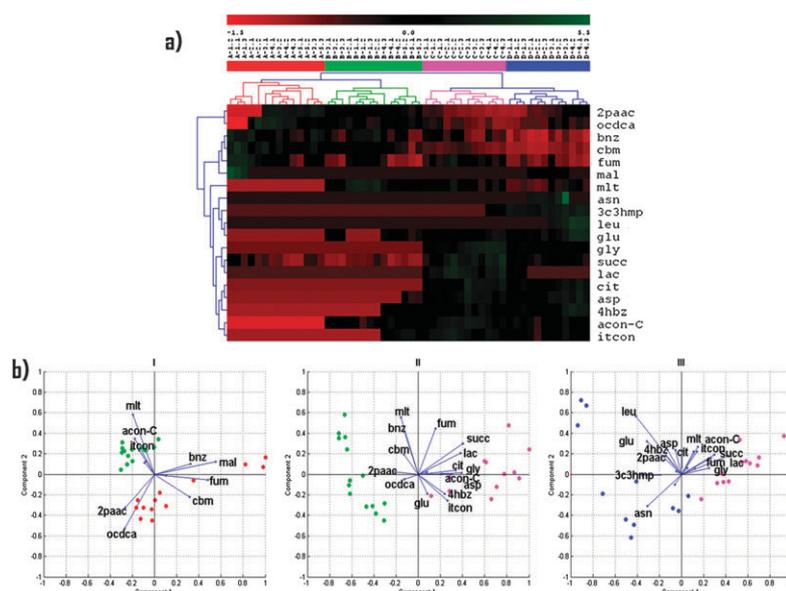
Fig. 4b shows that the metabolites with the highest positive loading profiles in PC2 (malonate (mlt), itaconate (itcon) and *cis*-aconitate (acon-C)) had their levels increased after IPTG induction in the *E. coli* W3110 culture. Similarly, these metabolites also presented high positive loading profiles in PC1 of samples from the  $\Delta relA$  mutant culture (Fig. 5b). In addition, itaconate and *cis*-aconitate were clustered using a Pearson correlation metrics, as represented in the HCA diagrams (Fig. 4a and 5a) presenting a strong association between their levels along both fed-batch fermentations.

### II Impact of nutrient upshift on the metabolic footprint.

Fig. 4b shows that glycine (gly), succinate (succ), lactate (lac), citrate (cit), aspartate (asp), 4-hydroxybenzoate (4hbz), *cis*-aconitate (acon-C) and itaconate (itcon) are the most significant variables that explain the differences projected in PC1 (positive loading profiles), which are the differences between growth phases B and C of the *E. coli* W3110 culture. In fact, as illustrated in Fig. 4a they were grouped in the same metabolite cluster and represent those that were highly accumulated after the nutrient upshift. Similarly, positive loading profiles for PC1 and negative for PC2 of the  $\Delta relA$  mutant culture (Fig. 5b) exposed these metabolites as the most relevant variables characterizing the samples after nutrient upshift, except for the hydroxybenzoate (4hbz) and itaconate (itcon).

### III Impact of glucose limitation on the metabolic footprint.

In the W3110 *E. coli* culture (Fig. 4b), negative loading profiles projected in the PC1, corresponding to leucine (leu), asparagine (asn), 2-isopropylmalate (3c3hmp) and glutamate (glu), showed the largest differences after nutrient downshift.



**Fig. 4** Analysis of the metabolic footprint profiles obtained from the fed-batch aerobic culture of *E. coli* W3110. (a) Hierarchical clustering analysis (HCA) distinguished four data classes (vertical clusters), corresponding to extracellular samples taken during each fermentation phase (including sample replicates), and several metabolite clusters (horizontal clusters) that characterized the metabolic footprint profiles during the recombinant fed-batch cultivation, based on a Pearson correlation metrics. (b) Principal component analysis (PCA) was performed to determine the most significant metabolic changes in the extracellular medium when growth conditions were changed: (I) induction of the heterologous protein expression; (II) nutritional upshift; and (III) glucose downshift. Biplots depict component scores projected in PCs as coloured dots, which represent samples clustered by HCA: samples from the first cluster (red dots), second cluster (green dots), third cluster (magenta dots) and fourth cluster (blue dots); and loading profiles represented as vectors that indicate metabolites that contributed the most to discriminate between sample clusters defined by the HCA. Only metabolites with larger vectors were depicted in the biplots, though the whole metabolic footprint data from the W3110 culture were analysed.

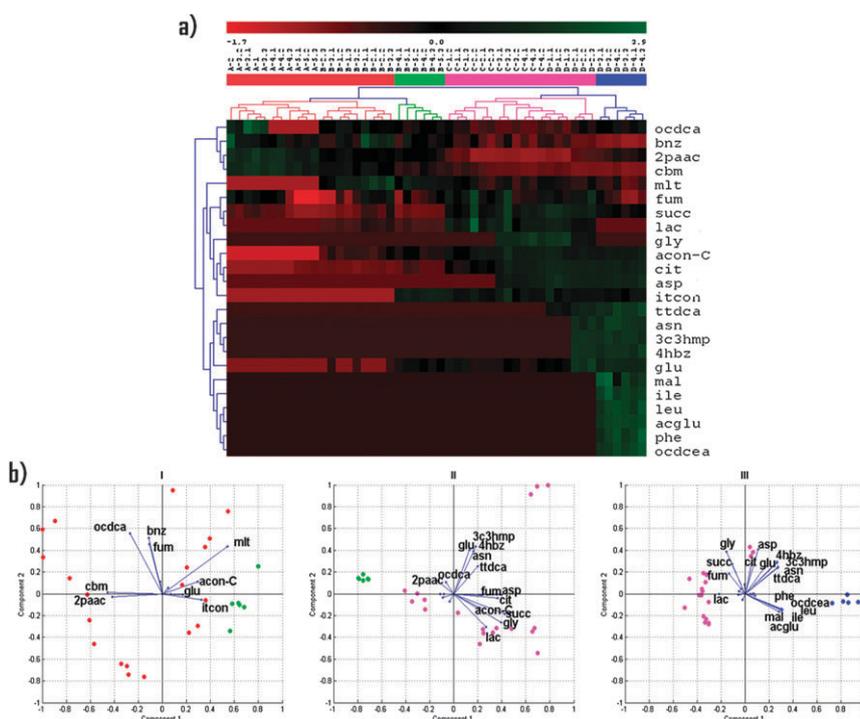
These metabolites were also clustered by HCA (Fig. 4a). The loading profiles of metabolic footprints from the  $\Delta relA$  mutant culture (Fig. 5b) depicted glycine (gly), succinate (succ) and lactate (lac) as adjacent vectors corresponding to metabolites that were immediately assimilated after glucose limitation.

## Discussion

The metabolic footprint analysis is supported on the basis that cells can secrete metabolites to the extracellular medium during growth and/or in response to environmental changes.<sup>37–39</sup> Furthermore, cells may activate a variety of efflux transporters that work like metabolic relief valves or defensive support to survive an antagonistic environment.<sup>40</sup> In the first case, an increase in extracellular metabolites would be associated with an increase in the intracellular concentration of those compounds, while in the second case there would be a gradient that has to be maintained by the cells to achieve specific purposes. In general, the variety and level of the secreted metabolites reflect the metabolic state of the cell and, therefore, may be considered the closest indicator of the phenotype.<sup>4,41–43</sup> Considering this, and the fact that the metabolic impact of recombinant protein expression in the host cells is still not well-understood, the metabolic footprints of the recombinant W3110 and  $\Delta relA$  mutant *E. coli* cells grown at different experimental conditions were analysed.

The metabolic responses of *E. coli* cells were evaluated by measuring some physiological parameters, such as the cellular growth and acetate formation (Fig. 1 and Table 1).

Results corroborated previous works,<sup>14–16</sup> indicating that the decrease on the cellular growth is the major consequence of the metabolic burden on the recombinant host cells due to higher demands of energy and amino acids. The drainage of energy and biosynthetic precursors associated to the expression of foreign proteins imposes severe changes in the metabolic activity of cells and, as a result, reduces the cellular growth. As shown, upon IPTG induction, the specific growth rate of the W3110 and  $\Delta relA$  mutant *E. coli* strains decreased 50% and 32%, respectively. The  $\Delta relA$  mutant strain seems to be less affected, which can be explained by the failure to stimulate the RelA-dependent stringent response, a stress response that has been proposed<sup>17,44,45</sup> to occur when there is a lack of intracellular amino acids associated with the additional requirements for the production of recombinant products. Ultimately, this stress response may induce a decrease of cellular growth and protein production. When increasing the nutrient availability, cells can generate sufficient metabolic and energetic resources for the formation of the heterologous protein, as well as for growth-associated processes, and as a result no significant physiological differences between cultures were observed. However, it was expected that at nutrient deprived conditions (phase D), the physiological responses of the two strains would be, at some extent, distinct. Besides the substrate uptake rates, the estimated physiological parameters did not show significant differences between the W3110 and  $\Delta relA$  mutant *E. coli* cultures. Cellular growth was rapidly arrested and formation of the heterologous protein decreased to similar levels, which did not allow us to deduce any



**Fig. 5** Analysis of the metabolic footprint profiles obtained from the fed-batch aerobic culture of *E. coli*  $\Delta relA$  mutant strain. (a) Hierarchical clustering analysis (HCA) distinguished four sample clusters (vertical clusters) and several metabolite clusters (horizontal clusters). Here, sample clusters did not correspond to each fermentation phases, but illustrate data classes that are characterized by different metabolic states defined by the metabolite footprints detected during the recombinant fed-batch cultivation, based on a Pearson correlation metrics. (b) Principal component analysis (PCA) was performed to determine the most significant metabolic changes in sample clusters. Biplots depict component scores projected in PCs as dots, which represent samples clustered by HCA: samples clustered in the first group (red dots), second cluster (green dots), third cluster (magenta dots) and fourth cluster (blue dots); and loading profiles represented as vectors that indicate metabolites that contributed the most to discriminate between the sample clusters defined by the HCA. Only metabolites with larger vectors were depicted in the biplots, though the whole metabolic footprints from the  $\Delta relA$  culture were analysed.

fundamental alterations in the cellular metabolism caused by the single gene mutation. Therefore, the analysis of extracellular metabolites was invaluable to determine the main consequences on the physiology of *E. coli* cells derived from the *relA* mutation and the experimental conditions.

As previously reported,<sup>37,38</sup> *E. coli* cells secrete metabolites according to the adjustments needed in the cellular metabolism to cope with different physiological demands. A small, but considerable number of metabolites characterized the metabolic footprints produced during the fed-batch processes and their level fluctuations were inspected to estimate their relationship with the intracellular metabolic changes. Metabolic footprints allowed us not only to discriminate between samples withdrawn at different fermentation phases, but also to disclose some metabolic changes that were not evidenced by the physiological characterization of the fed-batch cultures.

According to the PCA analysis (Fig. 3), it can be observed that most samples from the same cultivation phase are close to each other, except for some samples from phases C and D. On the other hand, also from PCA results, it can be concluded that both strains behave similarly during phases A and B, but quite differently in the remaining phases, in particular when comparing PC projections corresponding to samples D-3 and D-4 with D-3-*relA* and D-4-*relA*, respectively.

According to HCA results illustrated in Fig. 4a, samples from the W3110 *E. coli* fed-batch culture were discriminated

according to the fermentation phases, which defined clusters with characteristic metabolic properties. In contrast, samples from the  $\Delta relA$  mutant *E. coli* culture (Fig. 5a) were not equally clustered. When cells were induced to express the heterologous protein (I) or submitted to glucose-limiting conditions (III), the metabolic footprints observed before and immediately after these experimental shifts were equivalent.

In order to justify the different behaviour of both strains regarding the impact of heterologous production and environmental changes in the metabolic profiles, it is hypothesised that the lack of the *relA* gene allowed cells to face these stressful conditions during phases B and D in a more “relaxed” way, therefore making the metabolic profiles closer to the following (case of phase A to B) or previous stages (case of phase C to D). These results are consistent with the higher production rates of the mutant strain in phase D (Table 1).

Although, as seen in Table 2, most of the metabolites exhibit similar patterns between both strains and some of the changes found on the metabolite patterns could be associated with the intrinsic variability of this kind of experiments, the differences found and the accumulation of certain metabolites only detected in the  $\Delta relA$  mutant culture (e.g. acetyl-L-glutamate (acglu)) may result from changes on the cellular metabolism that were not equivalent in the two *E. coli* cultures.

Analysing the metabolic footprints of both cultures independently, and regarding IPTG induction of the heterologous

protein expression in the W3110 *E. coli* fed-batch culture, a set of metabolites were depicted as key elements characterizing the metabolic changes associated with this experimental transition. Three carboxylic acids (malonate, itaconate and *cis*-aconitate) were secreted into the culture broth (Fig. 4b) suggesting that they were released from the cells to prevent any inhibitory effects on the activity of isocitrate lyase. As mentioned before, *cis*-aconitate and itaconate, as well as malonic acid,<sup>35</sup> are inhibitors of this enzyme, controlling the activity of the first reaction of the glyoxylate shunt that was reported<sup>46</sup> as serving an anaplerotic function in the cell during heterologous protein production. Since several TCA intermediates are withdrawn from the TCA cycle as amino acid precursors and need to be replenished, these anaplerotic reactions are central to balance the intracellular levels of TCA metabolites, fulfilling this way the additional biosynthetic requirements associated with the formation of the heterologous protein. The  $\Delta relA$  mutant strain response to IPTG induction was also manifested by the accumulation of the same enzymatic inhibitors: malonate, *cis*-aconitate and itaconate (see Fig. 5b). However, HCA showed that the metabolic footprints produced immediately after IPTG induction did not discriminate these samples from the previous fermentation phase, which suggests that the mutant strain might retard these regulatory responses after induction of the heterologous protein production.

Despite the extensive knowledge on basic aspects such as the changes of growth rates with nutrient concentrations,<sup>47–51</sup> information on the effects of the nutritional upshift during recombinant processes on the metabolic footprint is still scarce. In general, studies are focused in the secretion of acetate, not only because it retards growth and inhibits protein formation, but also because it represents a deviation of carbon that might otherwise be used to generate energy and precursors for biosynthetic purposes.<sup>52–54</sup> In this study, a glucose feeding upshift was applied to increase the specific growth rate during heterologous protein production and the metabolic footprints were analysed. Numerous metabolites were immediately accumulated in the extracellular medium of the W3110 culture after glucose availability increased, such as glycine (gly), 4-hydroxybenzoate (4hbz), lactate (lac), citrate (cit), *cis*-aconitate (acon-C), itaconate (itcon), succinate (succ), fumarate (fum) and aspartate (asp) (Fig. 4b). If we consider that the increase in extracellular concentrations is associated with the accumulation of the corresponding intracellular metabolites, which include TCA intermediaries, amino acids and amino acid precursors, then two hypotheses can be deduced to explain this phenomenon. One is the increase in the overall activity of the metabolic pathways that would lead to a direct increased concentration of all the intermediates due increased enzyme activities. The second hypothesis would be related with an increased inhibitory action that would result in the accumulation of certain metabolites that participate in key reactions that are subjected to regulation. Although not all TCA cycle intermediates, amino acids and amino acid precursors have been measured, there seems to be a clear tendency of increased concentration of all metabolites, indicating that the first hypothesis might be more plausible. Moreover, acetate was also accumulated during this fermentation phase resulted from

the glucose overflow metabolism. It was reported<sup>55</sup> that in the presence of excess glucose, the carbon flux through glycolysis exceeds the capacity of the TCA cycle and acetate is accumulated.

The same set of metabolites, except 4-hydroxybenzoate and itaconate, were accumulated during the  $\Delta relA$  fermentation process at these conditions, which indicate that metabolic adjustments induced by the nutrient upshift resulted in similar alterations in the metabolic footprints from both strains. Although some metabolite patterns revealed differences between strains (Fig. 2), in general, these were not very significant to distinguish the samples (Fig. 3).

The accumulation of lactate at these conditions was found intriguing. Since the conversion of pyruvate to lactate in *E. coli* is usually exclusively induced at anaerobic conditions,<sup>56,57</sup> the presence of this by-product implies that the internal accumulation of pyruvate due to the metabolism overflow overrides any other mechanism known to control the activity of the lactate dehydrogenase (LdhA) enzyme under aerobic conditions<sup>58</sup> or, for some reason, local oxygen deficiencies during the *E. coli* fed-batch process have triggered the ArcAB system and other genes involved in the mixed acid fermentation pathway.<sup>59</sup> The latest assumption seems improbable, as good mixing conditions and dissolved oxygen values above 30% were maintained inside the reactor.

Finally, the metabolic responses to nutritional stress associated with the restriction of glucose feeding during recombinant *E. coli* processes were also evaluated. It is remarkable the amount of metabolites found to be secreted after glucose downshift. Besides the decreasing levels of metabolites that can serve as carbon sources for *E. coli* (e.g. acetic acid), the accumulation of unexpected metabolites, like amino acids (e.g. leucine (leu), asparagine (asn), glutamate (glu) and aspartic (asp)) and amino acid derivatives (e.g. 2-isopropylmalate (3c3hmp)), indicate that the cells accumulated these biosynthetic precursors as a consequence of the delayed protein translation machinery, as estimated by the experimental AcGFP1 formation rates (Table 1). It is acknowledged that under nutrient starvation *E. coli* cells entail complex protective processes that ultimately manage the cellular metabolism to sustain cellular maintenance and viability.<sup>60–62</sup> Apparently the RelA-dependent response is involved in these metabolic readjustments, as the metabolic footprints exhibited by the wild-type and  $\Delta relA$  mutant *E. coli* cultures were somehow divergent (Fig. 3). When comparing the metabolite patterns of both strains after nutrient downshift, we found that acetate utilization by *E. coli* cells has a strong effect on the accumulation of several metabolites, especially in the  $\Delta relA$  mutant strain (Fig. 2). Metabolites such as phenylalanine (phe), isoleucine (ile), leucine (leu), acetyl-L-glutamate (acglu) and octadecenoate (ocdcea) were accumulated in the medium of the  $\Delta relA$  mutant culture shortly after acetate consumption has started (data shown in ESI 1†). Under these conditions, acetate is converted to acetyl coenzyme A (accoA) at the expense of ATP which, in turn, is mainly catabolised *via* glyoxylate cycle that serves as anaplerotic reactions.<sup>63</sup> This shift in the utilization of carbon sources involves the activation of various cellular processes, including the synthesis of new catabolic enzymes and the activation of substrate-specific transport

systems. It seems that the metabolic imbalance caused by these metabolic activities coupled with the additional formation of the heterologous protein is the basis for the accumulation of several metabolites, including amino acids, that was more critical in the  $\Delta relA$  strain (where protein production rates were also higher in this phase). During glucose starvation the translational apparatus, as well as cellular growth, are limited *via* transcriptional control of several growth-associated genes, like ribosomal operons. The ppGpp-stringent control has been related to this cellular response to nutritional deprivation that redirects the RNA polymerase transcriptional activity from stable RNA (ribosomal and transfer RNA) synthesis to stress-related genes, in particular genes that have protective functions. In the absence of the RelA activity, the ppGpp accumulation is limited and the translational apparatus stays unaffected. This seems to be the main cause for the only slight reduction of heterologous protein production rate in the  $\Delta relA$  mutant. However, while protein production seems to be unaffected by nutrient starvation, bacterial growth was arrested in this strain, probably by action of responses mediated by other stress proteins. These facts apparently generated an imbalance in the mutant strain's metabolism (probably due to differences in amino acid composition between heterologous and average *E. coli* proteins) that was evidenced by the accumulation of some amino and fatty acids. For example, the synthesis of fatty acids is known to be inhibited during glucose starvation, which did not seem to happen in the mutant strain. However, since these compounds had not been used for biomass production, there was an accumulation of octadecenoate (ocdcea) and tetradecanoate (ttcca) in the  $\Delta relA$  mutant strain. Moreover, isoleucine (ile), acetyl-L-glutamate (acglu) and phenylalanine (phe) were also accumulated at these conditions by the  $\Delta relA$  mutant strain. It seems evident that the failure to accumulate ppGpp at these conditions allowed a continuous production of the heterologous protein (although slightly reduced) resulting in an unbalanced drainage of precursors. Some metabolites were gradually replenished while others (less required for the production of AcGFP1) were over-accumulated. The fact that most metabolic resources were probably redirected to the formation of AcGFP1 may explain the higher protein synthesis rate and lower biomass yield observed for the  $\Delta relA$  mutant strain when compared to the W3110 strain. It is clear that the *relA* mutation influences the natural cellular responses to nutritional downshifts, which can delay, or even suppress, *E. coli* survival and resistance.

## Conclusions

Although the characterization of metabolites in *E. coli* culture broths has been performed using various detection methods, such analyses are mostly confined to specific metabolites and have not been done in a global scale. For example, the secretion of acetate during aerobic *E. coli* fermentations is regularly measured, because it is considered a major obstacle to enhanced heterologous protein production. However, the present work reveals that the complexity of the generated metabolic footprints at different culture conditions is much higher than what has been admitted. As demonstrated, *E. coli*

secretes a vast array of metabolites that participate in a wide range of metabolic pathways. Although the metabolism in *E. coli* has been studied more intensively than in any other bacterium, only recently it has become clear that targeted studies do not provide an accurate picture of the cellular metabolism. A typical metabolomic approach is expected to generate new knowledge from the comprehensive analysis of the metabolome and the distinctive metabolic patterns produced at different environmental and genetic conditions.

Metabolic footprints resulted from the IPTG-induction of the heterologous protein expression have shown that there is a rapid accumulation of unexpected metabolites in the culture broth. The secretion of the isocitrate lyase inhibitors suggests that the anaplerotic glyoxylate shunt was activated to replenish the TCA intermediaries engaged in the additional formation of the heterologous protein. Moreover, the detection of compounds unknown to participate in *E. coli* metabolism reinforces the importance of unbiased analytical approaches in research.

When cells are exposed to conditions of nutrient-excess, the uncoupling of maximum glucose uptake rates and the TCA fluxes results in a metabolic overflow with consequent accumulation of overflow metabolites such as acetate and lactate. Assuming that the demand for metabolic resources for cellular growth and heterologous protein production is exceeded, the secretion of these metabolites can be understood as a metabolic relief required to avoid adverse effects from the imbalanced cellular metabolism.<sup>4</sup> Also, TCA intermediaries (*e.g.* fumarate, succinate, citrate and *cis*-aconitate) and some amino acids (*e.g.* aspartate and glycine) were secreted at these conditions, probably as a result of the increased activities of these pathways.

At nutrient-limited conditions (*i.e.* when glucose feeding was discontinued), some of these metabolites (*e.g.* acetate) were assimilated by the cells as carbon and energy sources. These metabolic activities, coupled with the formation of the heterologous protein may have resulted into severe rearrangements in the cellular metabolism that led to the secretion of amino acids like: phenylalanine, asparagine, and leucine; and the acetyl-L-glutamate. Once again, these metabolic imbalances were more pronounced in the  $\Delta relA$  strain, which fails to trigger RelA-dependent processes to respond to nutrient deprivation.

The metabolic flexibility exposed by the alterations in the metabolic footprints, evidenced that cells entail diverse cellular mechanisms to sense and rapidly counteract the adverse environmental conditions. This stringent behaviour was prevalent in the W3110 strain, while the  $\Delta relA$  strain showed some difficulties to cope immediately with the metabolic imbalances caused by the formation of heterologous protein. It is evident that, although some disadvantages might have been indicated concerning the metabolic behaviour of the  $\Delta relA$  strain (*e.g.* failure to manage metabolic imbalances), the enhanced production rate of the heterologous protein represents a major benefit.

Metabolic footprinting, more than most other analytical strategies, is a rapid and non-invasive analysis, representing a powerful approach for the characterization of phenotypes and the distinction of specific metabolic states due to

environmental or genetic alterations. Nevertheless, metabolic footprints are just a shallow representation of the metabolic state of cells and the full understanding of the underlying mechanisms controlling these metabolic imbalances caused by the heterologous protein production, require further inspection of key metabolites (*e.g.* metabolites that are important nodes in the metabolic network) or the combination with other experimental strategies (*e.g.* gene expression and proteomics).

## Experimental

### Microbial strains

*E. coli* strains W3110 (F<sup>-</sup>, LAM<sup>-</sup>, IN[*rrnD-rrnE*]1, *rph-1*) and the isogenic mutant containing the  $\Delta relA251::kan$  allele (obtained from M. Cashel, National Institute of Health, USA) were transformed with the cloned pTRC-HisA-AcGFP1 plasmid encoding the expression of the recombinant AcGFP1 protein. The *gfp* gene was amplified from the pAcGFP1 plasmid (Clontech, Takara Bio Company, USA) that encodes for the green fluorescent protein AcGFP1, a derivative of AcGFP from *Aequorea coerulescens*. The PCR product was then cloned into the pTRC-HisA vector (Invitrogen Corporation and Applied Biosystems Inc, USA) that contains a *trc* promoter for high-level expression of the fusion protein and an ampicillin resistance gene for propagation and selection in *E. coli*.

### Growth conditions

Precultures were prepared in 500 mL shaking flasks filled with 300 mL of minimal medium consisting of 5 g kg<sup>-1</sup> of glucose, 6 g kg<sup>-1</sup> of Na<sub>2</sub>HPO<sub>4</sub>, 3 g kg<sup>-1</sup> of KH<sub>2</sub>PO<sub>4</sub>, 0.5 g kg<sup>-1</sup> of NaCl, 1 g kg<sup>-1</sup> of NH<sub>4</sub>Cl, 0.015 g kg<sup>-1</sup> of CaCl<sub>2</sub>, 0.12 g kg<sup>-1</sup> of MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.34 g kg<sup>-1</sup> of thiamine, 2 mL kg<sup>-1</sup> of trace-element solution (described elsewhere)<sup>64</sup> and 2 mL kg<sup>-1</sup> of vitamins solution (described elsewhere).<sup>64</sup> The minimal medium containing additional 20 mg kg<sup>-1</sup> of L-isoleucine and 100 mg kg<sup>-1</sup> of ampicillin was used to grow the recombinant wild-type strain, while this same medium with further addition of 20 mg kg<sup>-1</sup> L-valine and 25 mg kg<sup>-1</sup> kanamycin was used to grow the  $\Delta relA$  mutant strain. Cells were thereafter washed and transferred to a 5 L fermenter (Biostat MD, Sartorius) with a working volume of 2 L containing the same minimal medium, except glucose. The fed-batch operation was started immediately after inoculation at 37 °C, pH 7 and dissolved oxygen (DO) above 30%. The feed media used contained 50 g kg<sup>-1</sup> of glucose, 10 g kg<sup>-1</sup> of NH<sub>4</sub>Cl, 4 g kg<sup>-1</sup> of MgSO<sub>4</sub>·7H<sub>2</sub>O and the additional requirements for amino acids and antibiotics as described before. The induction of AcGFP1 production was performed with 1.5 mM IPTG (isopropyl  $\beta$ -D-thiogalactoside) when the microbial culture reached an OD<sub>600 nm</sub> of 2.3. Fermentation conditions were monitored and controlled *via* a computer control system. A closed-loop feeding control algorithm was employed to maintain a constant specific growth rate ( $\mu$ ) in the fed-batch culture.<sup>65</sup> The algorithm is based on a Monod kinetic model using glucose as the only growth-limiting substrate. The model combines terms for cell growth ( $X\mu$ ), glucose consumption

( $Y_{X/S}S_f$ ) and the online measurement of culture medium weight ( $W_R$ ) to control the feeding profile, represented by:

$$F = \frac{X\mu W_R}{Y_{X/S}S_f}$$

The biomass concentration ( $X$ ) was initially measured by optical density and estimated at each acquisition time (every 3 min). The predicted growth yields on glucose ( $Y_{X/S}$ ) were set to 0.35 and 0.2, when the specific growth rates were set to 0.1 and 0.2 h<sup>-1</sup>, respectively. The fed-batch experiments were at first conducted at a setpoint of  $\mu = 0.1$  h<sup>-1</sup>, corresponding to the pre-induction (A) and the post-induction (B) phases. Afterwards the setpoint was changed to 0.2 h<sup>-1</sup>, which was kept during almost 4 h, corresponding to a nutritional upshift phase (phase C). When the feeding was ceased (glucose limitation phase or phase D), growth was followed until the OD<sub>600nm</sub> dropped.

### Sampling and analytical procedures

Samples were collected from the fermentation broth at regular time intervals (40–50 min) along the four cultivation phases from each fed-batch fermentation, one with each *E. coli* strain. Cell growth was monitored by measuring optical density (OD<sub>600nm</sub>) and cell dry weight. In order to determine cell dry weight, 10 mL of broth were centrifuged at 10 000 g for 20 min at 4 °C, washed twice with deionised water and dried at 105 °C to constant weight. The expression level of AcGFP1 was determined by fluorescence measurements at a Jasco FP-6200 spectrofluorometer with excitation and emission wavelengths of 475 and 505 nm, respectively, a bandwidth of 10 nm and a high sensitivity response in 0.1 s. His-Tag purification of the AcGFP1 was performed with HiTrap columns (GE Healthcare Bio-Sciences AB, Sweden) and the concentration was determined by the Bradford method using BSA as standard. For further analysis, culture samples were centrifuged (15 min, 3000 rpm, 4 °C) and the resulting supernatants were immediately filtered and collected. Afterwards the samples were stored at –20 °C for subsequent analysis and lyophilisation. Glucose and acetate were analysed by HPLC with a refractive index detector (Jasco, Canada) and a Chrompack organic acids column (Varian, USA) at 35 °C. The mobile phase consisted in a 0.01 N solution of H<sub>2</sub>SO<sub>4</sub> at a flow rate of 0.6 mL min<sup>-1</sup>.

### Derivatization and GC-MS analysis

For the GC-MS analysis, only the supernatants collected from samples for which biomass concentration was above 0.6 g kg<sup>-1</sup> were used to determine metabolic footprints (see Fig. 1b). One millilitre of the supernatants was lyophilized in triplicate. The lyophilized samples were then derivatized using the methyl chloroformate (MCF) method<sup>66</sup> and analyzed with a GC-MS system-GC7890 coupled to an MSD5975—(Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with a ZB-1701 GC capillary column, 30 m × 250  $\mu$ m id × 0.15 mm (film thickness) with 5 m guard column (Phenomenex, Inc., Torrance, CA, USA), at a constant flow rate of 1.0 mL min<sup>-1</sup> of helium. Samples (1  $\mu$ L) were injected onto the column under a pulsed splitless mode (1.8 bars until 1 min, 20 mL min<sup>-1</sup> split flow after 1.01 min) and the detector was set with a scan

interval of 1.47 s and  $m/z$  range of 38–650. The oven temperature was initially held at 45 °C for 2 min. Thereafter, the temperature was raised with a gradient of 9 °C min<sup>-1</sup> until 180 °C and held at this value for 5 min. The temperature was raised again at a gradient of 40 °C min<sup>-1</sup> in three steps: until 220 °C (held for 5 min), 240 °C (held for 11.5 min) and finally 280 °C (held for 2 min). The temperature of the inlet was 290 °C, the interface temperature 250 °C, and the quadrupole temperature 200 °C.

### Data processing and statistical analysis

Time series data consisted in a sequence of samples lyophilized, derivatized and analyzed in triplicate (in most cases), for the four cultivation phases for each fed-batch culture. Each time series consisted in 18 sampling points, where the first five sampling points (T0 to T4) corresponded to samples collected during phase A (prior to IPTG induction), the next five (T5 to T9) to samples from phase B (after IPTG induction), other four (T10 to T13) to samples from the growth upshift phase (C) and, the last four (T14 to T17) from the nutrient downshift phase (D). In the end, about 50–52 samples (including replicates) were analysed by GC-MS to characterize the metabolic footprints from each *E. coli* culture.

The mass fragmentation spectrum was analysed using the Automated Mass Spectral Deconvolution and Identification System (AMDIS)<sup>67</sup> to identify the compounds through matching with a library constructed by using analytical chemical standards. The peak intensity values from the AMDIS analysis were corrected for the recovery of the internal standard (D-4-alanine), *i.e.* divided by the intensity peak of the internal standard in each sample, and normalized with the corresponding biomass concentration, *i.e.* further divided by the biomass concentration measured in each sampling point. Data are provided in ESI 2.† These peak intensity values were thereafter transformed into Z-scores, where each metabolite  $K$  has mean peak intensity values equal to zero and standard deviation of one. Z-score values were calculated by subtracting the average peak intensity values for a metabolite  $K$  among all the  $n$  samples (including replicates) from the peak intensity value ( $I_{K,i}$ ) for that metabolite in sample  $i$ , and dividing that result by the standard deviation of all measured peak intensities corresponding to that metabolite  $K$ , according to:

$$Z - \text{score}_{K,i} = \frac{(I_{K,i} - \text{mean } I_{K,1..k,n})}{\text{SD}_{K,1..k,n}}$$

Further data processing and statistical analysis were performed with MATLAB (version 2009b, The Mathworks, Inc) and Multi-Experiment Viewer (MeV).<sup>68</sup> A first principal component analysis (PCA) was performed using the *princomp* MATLAB function on a matrix of the Z-score values of each measured metabolite (with averages for sample replicates) that included samples from all cultivation phases and from both *E. coli* cultures, *i.e.* a total of 36 average profiles were used in this analysis.

The degree of association between the metabolite profiles produced by the W3110 and  $\Delta\text{relA}$  *E. coli* cultures was determined using Pearson's correlation coefficients ( $r$ ) estimated by the MATLAB *corr* function. P-values associated with each Pearson correlation coefficient were calculated using a Student

$t$  distribution to test the null hypothesis ( $H_0$ ) of no significant correlation between the metabolite profiles from the two cultures, against the alternative hypothesis ( $H_1$ ) that establishes a significant correlation between the profiles. The  $p$ -value for Pearson's correlation coefficient is based on the test statistic,  $t$ , with  $n - 2$  degrees of freedom:

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$$

To further examine metabolic changes following alterations in growth conditions, two clustering analyses were performed on GC-MS data acquired from independent fed-batch cultures: hierarchical clustering analysis (HCA) and PCA. Samples used were not averaged among replicates in these clustering analyses. Hierarchical clustering analysis (HCA) was used to cluster samples and metabolites based on the Pearson correlation metrics (MeV software) and was performed on matrices (corresponding to each *E. coli* cultures) of Z-score values determined for each detected metabolite in samples (including replicates) collected from four different cultivation phases (A to D). Two-dimensional score plots and loading profiles of the principal components (PC) were used to visualize the relative contribution of individual metabolites to discriminate between sample clusters previously determined by HCA. In other words, each biplot was constructed based on component scores and loading profiles estimated by the *princomp* MATLAB function with samples from two clusters defined previously by HCA, in order to identify the metabolic shifts that best characterize the discrimination of the two sample clusters. Since samples were clustered in a sequential time course, samples included in the first cluster were compared with samples from the second cluster, samples from the second cluster with samples from the third cluster, and so on. Three biplots were used to compare the four sample clusters determined in each HCA, which were associated with specific growth shifts performed during the fed-batch cultures: (I) IPTG induction of the heterologous protein expression, (II) growth upshift and (III) glucose downshift. Although these shifts between sample clusters and cultivation phases were in agreement only for the W3310 culture, it was assumed that metabolic shifts were fundamentally caused by these growth alterations. Metabolic shifts are characterized in each biplot where sample scores are represented by colored dots according to the cluster enclosing that sample and loading profiles are represented as vectors that indicate which metabolites presented the highest alterations among samples (larger vectors).

### Acknowledgements

The authors thank to Raphael Aggio for assisting in the automatic refinement and correction of the GC-MS data. This work was supported in part by the research project Bridging Systems and Synthetic Biology for the development of Improved Microbial Cell Factories (MIT-Pt/BS-BB/0082/2008) and HeliSysBio-Molecular Systems Biology *Helicobacter pylori* (FCT PTDC/EBB-EBI/104235/2008), both financed by the Portuguese Fundação para a Ciência e Tecnologia. Sónia Carneiro was also supported by a PhD grant from the same institution (ref. SFRH/BD/22863/2005).

## References

- 1 S. Gnoth, M. Jenzsch, R. Simutis and A. Lubbert, *Bioprocess Biosyst. Eng.*, 2008, **31**(1), 21–39.
- 2 J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver and D. B. Kell, *Nat. Biotechnol.*, 2003, **21**(6), 692–696.
- 3 J. Allen, H. M. Davey, D. Broadhurst, J. J. Rowland, S. G. Oliver and D. B. Kell, *Appl. Environ. Microbiol.*, 2004, **70**(10), 6157–6165.
- 4 D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic and S. G. Oliver, *Nat. Rev. Microbiol.*, 2005, **3**(7), 557–565.
- 5 D. B. Kell, J. K. Allen, H. M. Davey, D. I. Broadhurst, J. J. Rowland and S. G. Oliver, *Yeast*, 2003, **20**, S335.
- 6 S. G. Villas-Bôas, C. D. Moon, S. Noel, H. Hussein, W. J. Kelly, M. Cao, G. A. Lane, A. L. Cookson and G. T. Attwood, *J. Biotechnol.*, 2008, **134**(1–2), 55–63.
- 7 S. G. Villas-Bôas, S. Noel, G. A. Lane, G. Attwood and A. Cookson, *Anal. Biochem.*, 2006, **349**(2), 297–305.
- 8 G. A. Pope, D. A. MacKenzie, M. Defernez, M. A. Aroso, L. J. Fuller, F. A. Mellon, W. B. Dunn, M. Brown, R. Goodacre, D. B. Kell, M. E. Marvin, E. J. Louis and I. N. Roberts, *Yeast*, 2007, **24**(8), 667–679.
- 9 C. P. Chou, *Appl. Microbiol. Biotechnol.*, 2007, **76**(3), 521–532.
- 10 I. S. Aldor, D. C. Krawitz, W. Forrest, C. Chen, J. C. Nishihara, J. C. Joly and K. M. Champion, *Appl. Environ. Microbiol.*, 2005, **71**(4), 1717–1728.
- 11 J. H. Seo, D. G. Kang and H. J. Cha, *Biotechnol. Appl. Biochem.*, 2003, **37**(Pt 2), 103–107.
- 12 B. Jurgen, H. Y. Lin, S. Riemschneider, C. Scharf, P. Neubauer, R. Schmid, M. Hecker and T. Schweder, *Biotechnol. Bioeng.*, 2000, **70**(2), 217–224.
- 13 B. R. Glick, *Biotechnol. Adv.*, 1995, **13**(2), 247–261.
- 14 J. Bonomo and R. T. Gill, *Biotechnol. Bioeng.*, 2005, **90**(1), 116–126.
- 15 W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis and D. S. Kompala, *Biotechnol. Bioeng.*, 1990, **35**(7), 668–681.
- 16 S. W. Harcum, *J. Biotechnol.*, 2002, **93**(3), 189–202.
- 17 S. W. Harcum and W. E. Bentley, *Appl. Biochem. Biotechnol.*, 1999, **80**(1), 23–37.
- 18 V. Jain, M. Kumar and D. Chatterji, *J. Microbiol.*, 2006, **44**(1), 1–10.
- 19 J. Wu and J. Xie, *J. Cell. Physiol.*, 2009, **220**(2), 297–302.
- 20 D. Chatterji, N. Fujita and A. Ishihama, *Genes Cells*, 1998, **3**(5), 279–287.
- 21 I. I. Touloukhou, I. Shulgina and V. J. Hernandez, *J. Biol. Chem.*, 2001, **276**(2), 1220–1225.
- 22 L. U. Magnusson, A. Farewell and T. Nystrom, *Trends Microbiol.*, 2005, **13**(5), 236–242.
- 23 J. W. Roberts, *Genes Dev.*, 2009, **23**(2), 143–146.
- 24 T. Durfee, A. M. Hansen, H. Zhi, F. R. Blattner and D. J. Jin, *J. Bacteriol.*, 2008, **190**(3), 1084–1096.
- 25 A. Srivatsan and J. D. Wang, *Curr. Opin. Microbiol.*, 2008, **11**(2), 100–105.
- 26 I. Artsimovitch, V. Patlan, S. I. Sekine, M. N. Vassilyeva, T. Hosaka, K. Ochi, S. Yokoyama and D. G. Vassilyev, *Cell*, 2004, **117**(3), 299–310.
- 27 D. Chatterji and A. K. Ojha, *Curr. Opin. Microbiol.*, 2001, **4**(2), 160–165.
- 28 M. Jishage, K. Kvint, V. Shingler and T. Nystrom, *Genes Dev.*, 2002, **16**(10), 1260–1270.
- 29 K. F. Smart, J. R. van Houtte, R. B. M. Aggio and S. G. Villas-Bôas, *Nat. Protocols*, 2010, **5**(10), 1709–1729.
- 30 M. Kanehisa and S. Goto, *Nucleic Acids Res.*, 2000, **28**, 27–30.
- 31 I. M. Keseler, C. Bonavides-Martinez, J. Collado-Vides, S. Gama-Castro, R. P. Gunsalus, D. A. Johnson, M. Krummenacker, L. M. Nolan, S. Paley, I. T. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, A. G. Shearer and P. D. Karp, *Nucleic Acids Res.*, 2009, **37**(Database issue), D464–D470.
- 32 A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. O. Palsson, *Mol. Syst. Biol.*, 2007, **3**, 121.
- 33 T. Willke and K. D. Vorlop, *Appl. Microbiol. Biotechnol.*, 2001, **56**(3–4), 289–295.
- 34 Y. S. Kim, *J. Biochem. Mol. Biol.*, 2002, **35**(5), 443–451.
- 35 J. C. Hoyt, E. F. Robertson, K. A. Berlyn and H. C. Reeves, *Biochim. Biophys. Acta*, 1988, **966**(1), 30–35.
- 36 R. Steuer, J. Kurths, O. Fiehn and W. Weckwerth, *Bioinformatics*, 2003, **19**(8), 1019–1026.
- 37 I. Arana, C. Seco, K. Epelde, A. Muela, A. Fernandez-Astorga and I. Barcina, *Antonie Van Leeuwenhoek*, 2004, **86**(2), 189–199.
- 38 H. Y. Lin, F. Hoffmann, A. Rozkov, S. O. Enfors, U. Rinas and P. Neubauer, *Biotechnol. Bioeng.*, 2004, **87**(5), 602–613.
- 39 S. Srinivasan, J. Ostling, T. Charlton, N. R. de, K. Takayama and S. Kjelleberg, *J. Bacteriol.*, 1998, **180**(2), 201–209.
- 40 T. K. Van Dyk, L. J. Templeton, K. A. Cantera, P. L. Sharpe and F. S. Sariaslani, *J. Bacteriol.*, 2004, **186**(21), 7196–7204.
- 41 O. Fiehn, *Plant Mol. Biol.*, 2002, **48**(1–2), 155–171.
- 42 L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, D. K. van and S. G. Oliver, *Nat. Biotechnol.*, 2001, **19**(1), 45–50.
- 43 S. G. Villas-Bôas, J. F. Moxley, M. Akesson, G. Stephanopoulos and S. Nielsen, *Biochem. J.*, 2005, **388**, 669–677.
- 44 F. T. Haddadin and S. W. Harcum, *Biotechnol. Bioeng.*, 2005, **90**(2), 127–153.
- 45 L. Andersson, S. Yang, P. Neubauer and S.-O. Enfors, *J. Biotechnol.*, 1996, **46**(3), 255–263.
- 46 C. Wittmann, J. Weber, E. Betiku, J. Kromer, D. Bohm and U. Rinas, *J. Biotechnol.*, 2007, **132**(4), 375–384.
- 47 Q. Hua, C. Yang, T. Oshima, H. Mori and K. Shimizu, *Appl. Environ. Microbiol.*, 2004, **70**(4), 2354–2366.
- 48 T. Ferenci, *Curr. Opin. Microbiol.*, 1999, **2**, 208–213.
- 49 H. Tweeddale, L. Notley-McRobb and T. Ferenci, *J. Bacteriol.*, 1998, **180**(19), 5109–5116.
- 50 A. G. Marr, *Microbiol. Rev.*, 1991, **55**(2), 316–333.
- 51 U. Lendenmann and T. Egli, *Biotechnol. Bioeng.*, 1998, **59**(1), 99–107.
- 52 M. Van de Walle and J. Shiloach, *Biotechnol. Bioeng.*, 1998, **57**(1), 71–78.
- 53 D. C. Suarez and B. V. Kilikian, *Process Biochem.*, 2000, **35**(9), 1051–1055.
- 54 M. A. Eiteman and E. Altman, *Trends Biotechnol.*, 2006, **24**(11), 530–536.
- 55 R. A. Majewski and M. M. Domach, *Biotechnol. Bioeng.*, 1990, **35**, 732–738.
- 56 D. P. Clark, *FEMS Microbiol. Rev.*, 1989, **63**(3), 223–234.
- 57 E. M. Tarmy and N. O. Kaplan, *J. Biol. Chem.*, 1968, **243**(10), 2587–2596.
- 58 G. R. Jiang, S. Nikolova and D. P. Clark, *Microbiology*, 2001, **147**(Pt 9), 2437–2446.
- 59 B. Xu, M. Jahic, G. Blomsten and S. O. Enfors, *Appl. Microbiol. Biotechnol.*, 1999, **51**(5), 564–571.
- 60 T. Nystrom, *Curr. Opin. Microbiol.*, 1999, **2**(2), 214–219.
- 61 P. Neubauer, M. Ahman, M. Tornkvist, G. Larsson and S. O. Enfors, *J. Biotechnol.*, 1995, **43**(3), 195–204.
- 62 A. Matin, *Mol. Microbiol.*, 1991, **5**(1), 3–10.
- 63 T. D. Brown, M. C. Jones-Mortimer and H. L. Kornberg, *J. Gen. Microbiol.*, 1977, **102**(2), 327–336.
- 64 I. Rocha and E. C. Ferreira, *Anal. Chim. Acta*, 2002, **462**(2), 293–304.
- 65 I. Rocha, A. C. A. Veloso, S. Carneiro, R. Costa and E. C. Ferreira, 17th World Congress The International Federation of Automatic Control, 2008.
- 66 S. G. Villas-Bôas, D. G. Delicado, M. Akesson and J. Nielsen, *Anal. Biochem.*, 2003, **322**(1), 134–138.
- 67 S. E. Stein, *J. Am. Soc. Mass Spectrom.*, 1999, **10**(8), 770–781.
- 68 A. I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush and J. Quackenbush, *BioTechniques*, 2003, **34**(2), 374–378.