Applying a Metabolic Footprinting Approach to Characterize the Impact of the Recombinant Protein Production in *Escherichia coli*

Sónia Carneiro, Silas G. Villas-Bôas, Isabel Rocha, and Eugénio C. Ferreira

Abstract. In this study metabolic footprinting was applied to evaluate the metabolic consequences of protein overproduction at slow growth conditions ($\mu = 0.1$ h^{-1}). The extracellular metabolites detected by gas chromatography-mass spectrometry characterized the metabolic footprints before and after the induction of the recombinant protein production (i.e. pre- and post-induction phases). Metabolic footprinting enabled the discrimination between the two growth phases and exposed significant alterations in the extracellular milieu during the recombinant process.

1 Introduction

Escherichia coli has been exploited for the production of a variety of products, especially recombinant proteins with pharmaceutical applications. Vast efforts have been made to improve the productivity of such bioprocesses, like the optimization of operational conditions, medium composition and the implementation of monitoring and control strategies [11,12,21]. However, the overproduction of these recombinant products often causes cellular stress events that result in slow growth and eventually cessation of growth [2,3,8,18]. The rapid exhaustion of essential metabolic precursors and cellular energy due to the expression of recombinant proteins may result in the imbalance of the metabolic load in the host cell, also called metabolic burden [9]. It is believed that the withdrawal of the intermediates that serve as biochemical precursors explains the decreasing tricarboxylic acid (TCA) cycle activity and consequent acetate production that has been reported in many works [1,5-7,10,22,23,25].

In recent years, various high-throughput experimental techniques have been used to understand the physiological behavior of cells during the operation of these bioprocesses and to develop strategies to overcome some of these limitations, like the design of improved strains that maximize the yield or productivity

Sónia Carneiro · Isabel Rocha · Eugénio C. Ferreira
IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Silas G. Villas-Bôas
School of Biological Sciences, The University of Auckland, 3A Symonds Street, Auckland 1142, New Zealand

of recombinant proteins [4,14]. Transcriptome and proteome analyses have been widely used to investigate the stress response mechanisms associated with over-production of recombinant proteins in *E. coli*, but so far metabolomic approaches were scarcely exploited in the characterization of recombinant cultures. Metabolic footprinting, i.e. the analysis of the entire set of metabolites released from cells into the extracellular medium, can be an effective method to characterize the metabolic state of cells at diverse environmental conditions. The secretion of metabolites during the production of recombinant proteins may reflect the adjustment of the intracellular metabolism in response to the imposed metabolic demands [13], since the intracellular accumulation of certain metabolites due to metabolic imbalances will most probably result in their excretion. Therefore, metabolic footprinting can represent an invaluable tool to generate key information that, together with other experimental data, will help in the optimization of recombinant cultures.

In this study, we investigated the usefulness of metabolic footprinting to assess the impact of the induction of recombinant protein production in *E. coli* cells. To avoid overlapping cellular responses that could be triggered, for example by metabolic overflow, cellular growth was maintained at low rates through the control of the glucose feeding profile.

2 Material and Methods

2.1 Growth Conditions

The *E. coli* strain W3110 (F-, *LAM*-, *IN*[*rrn*D-*rrn*E]1, *rph*-1) was transformed with the cloned pTRC-HisA-AcGFP1 plasmid encoding the production of the re-combinant AcGFP1 protein. The *gfp* gene was amplified from the pAcGFP1 plasmid (from Clontech) that encodes for the green fluorescent protein AcGFP1, a derivative of AcGFP from *Aequorea coerulescens*.

Cells were first grown in a shake flask pre-culture using minimal medium consisting of 5 g·kg^{-1} of glucose, 6 g·kg^{-1} of Na$_2$HPO$_4$, 3 g·kg^{-1} of KH$_2$PO$_4$, 0.5 g·kg^{-1} of NaCl, 1 g·kg^{-1} of NH$_4$Cl, 0.015 g·kg^{-1} of CaCl$_2$, 0.12 g·kg^{-1} of MgSO$_4$.7H$_2$O, 0.34 g·kg^{-1} of thiamine, 2 mL·kg^{-1} of trace-element solution (described elsewhere [16], 2 mL·kg^{-1} of vitamins solution (described elsewhere [16]), 20 mg·kg^{-1}of L-isoleucine and 100 mg·kg^{-1} of ampicillin. For fed-batch cultures, cells were thereafter transferred to a fermenter with the same minimal medium, except glucose. The feeding medium consisted of 50 g·kg^{-1} of glucose, 10 g·kg^{-1} of NH$_4$Cl, 4 g·kg^{-1} of MgSO$_4$.7H$_2$O, 20 mg·kg^{-1}of L-isoleucine and 100 mg·kg^{-1} of ampicillin.

Fed-batch fermentation was conducted in a 5L fermenter (Biostat MD) with a working volume of 2 L at 37°C, pH 7 and dissolved oxygen (DO) above 30%. The induction of AcGFP1 protein production was performed with 1.5 mM IPTG (isopropyl β-D-thiogalactoside) when the culture reached an optical density (OD600nm) of 2.3. Fermentation conditions were monitored and controlled via a computer supervisory system. A closed-loop feeding control algorithm was employed to maintain the growth rate (μ) constant in the fed-batch culture [17]. The

algorithm is based on a Monod kinetic model using glucose as the only growth-limiting substrate and can be represented by the following equation:

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

To maintain the specific growth rate (μ) at $0.1h^{-1}$, the feeding profile was computed based on the growth yield in glucose ($Y_{X/S}$) that was set to 0.35 and the concentration of glucose in the feed (S_f), kept at 50 g·kg^{-1}. The culture medium weight (W_R) was measured online, while the biomass concentration (X) was estimated based on the initial concentration.

2.2   Analytical Methods

Samples were taken from the fermenter approximately every 30 minutes for the determination of OD600nm, AcGFP1 fluorescence, glucose and acetate concentrations and the GC-MS analysis of extracellular amino and nonamino organic acids. In order to determine the cell dry weight, 10 mL of broth were centrifuged at 10000 g for 20 min at 4ºC, washed twice with deionized water and dried at 105 °C to constant weight. The production 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 seconds. His-Tag purification of the AcGFP1 was performed with HiTrap columns (GE Healthcare Bio-Sciences AB) and the concentration was determined by the Bradford method using BSA as standard. The metabolic footprints were analyzed by GC-MS. After the lyophilization of 1 mL of each sample in triplicates, chemical derivatization was performed using the methyl chloroformate (MCF) method described elsewhere [26]. Samples were thereafter analyzed with a GC-MS system – a GC7890 coupled to an MSD5975 - (Agilent Technologies) equipped with a ZB-1701 GC capillary column, 30 m x 250 mm id x 0.15 mm (film thickness) with 5 m guard column (from Phenomenex), at a constant flow rate of 1.0 mL/min of helium. Samples (1 µL) were injected onto the column under pulsed splitless mode (1.8 bars until 1 min, 20 mL/min split flow after 1.01 min) and the detector was set with a scan interval of 1.47 seconds and m/z range of 38-650.

2.3   Data Processing and Statistical Analysis

The mass fragmentation spectrum was analyzed with the Automated Mass Spectral Deconvolution and Identification System (AMDIS) [24] to identify the metabolites matching the analytical chemical standards. The peak intensity values of the identified metabolites in the spectrum were normalized by the peak intensity of the internal standard (D-4-alanine) and the corresponding biomass concentration of the sample. Further data processing and statistical analysis were performed with

MultiExperiment Viewer (MeV) v4.5 [20]. The normalized peak intensity values were $\log_2$ transformed and further computed using K-means clustering (KMC), hierarchical clustering (HCL), and principal component analysis (PCA). K-means method was used to group the metabolic profiles into k clusters, while hierarchical clustering distributed samples and metabolites into branched groups represented by a two dimensional tree. Euclidean distance metrics were used in both clustering methods. PCA was further used to visualize whether the samples could be differentiated based on their metabolic profiles.

3   Results

In this study the impact of the production of recombinant proteins in the *E. coli* metabolism was investigated. The resulting growth pattern (Figure 1) shows that the biomass formation was affected by the AcGFP1 production, since before the IPTG-induction the experimental growth rate (0.16 h^{-1}) was higher than the growth rate imposed by the feeding profile. In turn, after IPTG induction the estimated biomass concentration was closer to the experimentally determined, corresponding to a growth rate of approximately 0.09 h^{-1}. This suggests that the biomass yields from glucose ($Y_{X/S}$) in the two growth phases were considerably different. In the first phase (pre-induction) glucose entering the cell was used for growth and maintenance, while in the second phase (post-induction) glucose was also allocated to AcGFP1 formation.
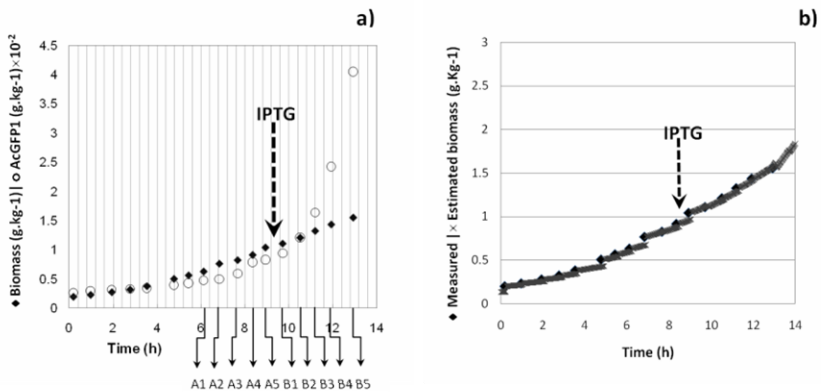

Fig. 1 Growth profile of the recombinant *E. coli* during the controlled fed-batch fermentation. a) The biomass formation and the AcGFP1 production were monitored in the pre- and post-induction phases. The dashed arrow indicates IPTG addition to the culture and the other arrows point to the samples analyzed in the GC-MS. Samples were identified by the letters A and B, corresponding to the pre- and post induction phases, respectively. b) The feeding profile generated by the model was periodically updated for the estimated biomass concentration
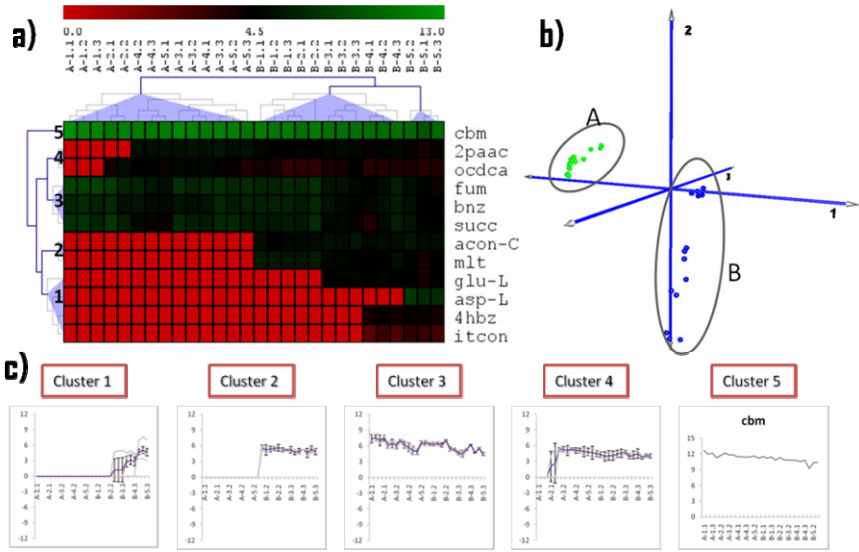
Fig. 2 Analysis of the metabolic footprints. a) Hierarchical clustering (HCL) defined clusters in the sample and metabolite dimensions. b) Principal component analysis (PCA) grouped samples from the pre-induction phase (group A) isolated from samples taken after IPTG-induction (group B). c) K-Means Clustering (KMC) gives the resulting clusters based on the KMC input parameters (50 interactions upper limit). Metabolite abbreviations: *cbm* - carbamic acid; *2paac* - 2-phenylaminoacetic acid; *ocdca* – stearic acid; *fum* – fumaric acid; *bnz* – benzoic acid; *succ* – succinic acid; *acon-C* – *cis*-aconitic acid; *mlt* – malonic acid; *glu-L* – L-glutamic acid; *asp-L* – L-aspartic acid; *4hbz* - hydroxybenzoic acid; *itcon* – itaconic acid

Since the experiments were performed at slow growth, no acetate was accumulated. However, unexpected metabolites were detected in the extracellular medium during the fed-batch experiment. According to the GC-MS results, important differences were identified in the metabolic footprints in both growth phases: pre-induction (phase A) and post-induction (phase B).

Results from HCL and PCA analyses (Figure 2a and 2b) show that samples taken during phases A and B presented distinct metabolic footprints. Although HCL shows that samples B5 cluster in a small sub-branch, it is clear that the metabolic footprints sampled before IPTG-induction (phase A) are distinct from those sampled after IPTG-induction (phase B). This was also confirmed by the PCA graph that assigned samples into two major groups corresponding to the growth phases. Further examination of the clusters generated by the K-means method (Fig. 2c) reveals that some metabolites presented similar profiles along the fermentation process. For example, cluster 2 was characterized by the extracellular accumulation of *cis*-aconitic acid (*acon-C*) and malonic acid (*mlt*) immediately after the IPTG induction, while metabolites within cluster 1 were secreted approximately two hours later, when the AcGFP1 production was most evident. Carbamic

acid (*cbm*), involved in the production of a source of nitrogen for *E. coli* growth, was the only metabolite that did not cluster with any other metabolite.

4 Discussion and Conclusions

Both PCA and HCL analyses indicate that the recombinant protein production in *E. coli* has a high impact in the cellular metabolism. Several metabolites were secreted into the medium after IPTG induction, in particular metabolites that were collected within clusters 1 and 2. For the other metabolites, it is not possible to discern if their profiles were influenced by the production of recombinant protein or by the changing fed-batch conditions (biomass concentration increases during the experiment). However, metabolites collected in clusters 1 and 2 are more likely to be modulated by the recombinant process since the accumulation of these metabolites was only observed in phase B.

The accumulation of *cis*-aconitic acid, that participates in the TCA cycle and two other amino acids, which precursors are TCA intermediates, such as the L-glutamic and L-aspartic acids that are synthesized via the enzymatic conversion of the α-ketoglutaric and oxaloacetic acids, respectively, suggest that this pathway is notably affected by the production of the recombinant protein. As reported [15,19], the metabolic effects imposed by the plasmid maintenance and formation of the recombinant product were found to be associated with the metabolic burden caused by the withdrawal of some metabolic intermediates that serve as biochemical precursors. As a consequence, the activity of many biochemical pathways is affected and it is frequently observed a decrease in the TCA cycle activity and the production of by-products that, at high concentrations, might be toxic to the cells. Therefore, by unbalancing the activity of certain reactions, it is expected that some metabolites are accumulated and subsequently secreted into the extracellular medium. This supports the idea that the metabolic adjustments required to compensate the additional production of a recombinant product are not as efficient as desired from a bioprocess optimization perspective.

Although the metabolic footprint measurements are not entirely informative, since they do not provide a comprehensive analysis of the intracellular metabolic changes, they can be used as variables in a multivariate statistical process control, like dynamic principal component analysis, for the on-line bioprocess monitoring. Ultimately, this information can help to characterize the physiological state and culturing performance during recombinant processes. From an engineering point of view, it is crucial to operate these processes in order to achieve consistent and reproducible qualities by developing modeling strategies using real-time observation of process variables, and furthermore to detect in advance abnormal production conditions.

References

1. Akesson, M., Hagander, P., Axelsson, J.P.: Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. Biotechnol. Bioeng. 73, 223–230 (2001)
2. Bentley, W.E., Mirjalili, N., Andersen, D.C., et al.: Plasmid-encoded protein - the principal factor in the metabolic burden associated with recombinant bacteria. Biotechnol. Bioeng. 35, 668–681 (1990)
3. Bonomo, J., Gill, R.T.: Amino acid content of recombinant proteins influences the metabolic burden response. Biotechnol. Bioeng. 90, 116–126 (2005)
4. Bulter, T., Bernstein, J.R., Liao, J.C.: A perspective of metabolic engineering strategies: Moving up the systems hierarchy. Biotechnol. Bioeng. 84, 815–821 (2003)
5. Chou, C.P.: Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Appl. Microbiol. Biotechnol. 76, 521–532 (2007)
6. Dittrich, C.R., Bennett, G.N., San, K.Y.: Characterization of the acetate-producing pathways in *Escherichia coli*. Biotechnol. Progr. 21, 1062–1067 (2005)
7. Eiteman, M.A., Altman, E.: Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotechnol. 24, 530–536 (2006)
8. Ganusov, V.V., Brilkov, A.V.: Estimating the instability parameters of plasmid-bearing cells. I. Chemostat culture. J Theor. Biol. 219, 193–205 (2002)
9. Glick, B.R.: Metabolic load and heterologous gene-expression. Biotechnol. Adv. 13, 247–261 (1995)
10. Jana, S., Deb, J.K.: Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl. Microbiol. Biotechnol. 67, 289–298 (2005)
11. Levisauskas, D., Galvanauskas, V., Henrich, S., et al.: Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. Bioprocess Biosyst. Eng. 25, 255–262 (2003)
12. Mahadevan, R., Doyle, F.J.: On-line optimization of recombinant product in a fed-batch bioreactor. Biotechnol. Progr. 19, 639–646 (2003)
13. Mapelli, V., Olsson, L., Nielsen, J.: Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. Trends Biotechnol. 26, 490–497 (2008)
14. Menzella, H.G., Ceccarelli, E.A., Gramajo, H.C.: Novel *Escherichia coli* strain allows efficient recombinant protein production using lactose as inducer. Biotechnol. Bioeng. 82, 809–817 (2003)
15. Ozkan, P., Sariyar, B., Utkur, F.O., et al.: Metabolic flux analysis of recombinant protein overproduction in *Escherichia coli*. Biochem. Eng. J. 22, 167–195 (2005)
16. Rocha, I., Ferreira, E.C.: On-line simultaneous monitoring of glucose and acetate with FIA during high cell density fermentation of recombinant *E. coli*. Anal. Chim. Acta. 462, 293–304 (2002)
17. Rocha, I., Veloso, A.C.A., Carneiro, S., et al.: Implementation of a specific rate controller in a fed-batch *E. coli* fermentation. In: Proceedings of the 17th World Congress The International Federation of Automatic Control (2008)
18. Rozkov, A., vignone-Rossa, C.A., Ertl, P.F., et al.: Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. Biotechnol. Bioeng. 88, 909–915 (2004)
19. Saeed, A.I., Sharov, V., White, J., et al.: TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34, 374–378 (2003)

20. Saucedo, V.M., Karim, M.N.: Analysis and comparison of input-output models in a recombinant fed-batch fermentation. J. Ferment. Bioeng. 83, 70–78 (1997)
21. Shiloach, J., Kaufman, J., Guillard, A.S., et al.: Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by *Escherichia coli* BL21 ((*λ*DE3) and *Escherichia coli* JM109. Biotechnol. Bioeng. 49, 421–428 (1996)
22. Shimizu, N., Fukuzono, S., Fujimori, K., et al.: Fed-batch cultures of recombinant *Escherichia coli* with inhibitory substance concentration monitoring. J. Ferment. Technol. 66, 187–191 (1988)
23. Stein, S.E.: An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J. Am. Soc. Mass Spectrom. 10, 770–781 (1999)
24. Suarez, D.C., Kilikian, B.V.: Acetic acid accumulation in aerobic growth of recombinant *Escherichia coli*. Process Biochem. 35, 1051–1055 (2000)
25. Villas-Boas, S.G., Delicado, D.G., Akesson, M., et al.: Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. Anal. Biochem. 322, 134–138 (2003)