Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT–IR and direct injection electrospray mass spectrometry

Naheed N. Kaderbhai, David I. Broadhurst, David I. Ellis, Royston Goodacre* and Douglas B. Kell*

Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, Wales, UK

**Abstract**

We sought to test the hypothesis that mutant bacterial strains could be discriminated from each other on the basis of the metabolites they secrete into the medium (their ‘metabolic footprint’), using two methods of ‘global’ metabolite analysis (FT–IR and direct injection electrospray mass spectrometry). The biological system used was based on a published study of *Escherichia coli* tryptophan mutants that had been analysed and discriminated by Yanofsky and colleagues using transcriptome analysis. Wild-type strains supplemented with tryptophan or analogues could be discriminated from controls using FT–IR of 24 h broths, as could each of the mutant strains in both minimal and supplemented media. Direct injection electrospray mass spectrometry with unit mass resolution could also be used to discriminate the strains from each other, and had the advantage that the discrimination required the use of just two or three masses in each case. These were determined via a genetic algorithm. Both methods are rapid, reagentless, reproducible and cheap, and might beneficially be extended to the analysis of gene knockout libraries. Copyright © 2003 John Wiley & Sons, Ltd.

**Keywords:** metabolome; metabolomics; metabolic footprinting; mass spectrometry; FTIR spectroscopy; *Escherichia coli*; tryptophan; genetic algorithm

**Introduction**

The systematic and complete genome sequencing of many organisms, including *Escherichia coli* (Blattner et al., 1997, 1998; Liang et al., 2002; Riley and Serres, 2000), brings the need to establish the cellular functions of all the genes, including the many novel genes, thereby uncovered (Brent, 1999, 2000; Clare and King, 2000; Hieter and Boguski, 1997; Kell and King, 2000; King et al., 2000; Oliver, 1996; Skolnick et al., 2000). Typical strategies include expression profiling at the level of the transcriptome (Burge, 2001; Devaux et al., 2001; Jurgen et al., 2001; Oshima et al., 2002; Tjaden et al., 2002) and the proteome (Akashi and Gojobori, 2002; Auer et al., 1998; Champion et al., 2001; Choe et al., 1998; Dunn, 1998; Futcher et al., 1997, 1999; Han et al., 2001; Joubert-Caron and Caron, 1999; Jurgen et al., 2001; Kabir and Shimizu, 2001; Loo et al., 2001; Thomas, 1999) and, given the importance of metabolism to cellular physiology (Cornish-Bowden and Cardenas, 2000, 1999), at the level of metabolism. With the emergence of many useful genomics-derived databases of metabolism (e.g. Goryanin et al., 1999; Kanehisa and Goto, 2000; Karp et al., 1996, 2002, 2000b; Mendes, 2002; Ouzounis and Karp, 2000; Schilling and Palsson, 2000; Tomita et al., 1999; Wixon and Kell, 2000), experimental
Metabolic footprinting in *E. coli* 377

analysis of the metabolome (Kell and Mendes, 2000; Oliver, 1996, 1997; Oliver et al., 1998; Raamsdonk et al., 2001; ter Kuile and Westerhoff, 2001), the complement of low molecular weight metabolites in a cell (Covert et al., 2001; Fiehn, 2001, 2002; Fiehn et al., 2000, 2001; Hall et al., 2002; Kose et al., 2001; Roessner et al., 2001; Taylor et al., 2002) is now being seen as another essential step for improving our understanding of the living cell. In order to study these metabolites a number of factors have to be considered, as the changes that take place will be multiple, will affect molecules with very different chemistries, and there will be both intracellular and extracellular changes (e.g. Kramer, 1994). Their quantification, detection and identification using new technologies and methodologies provide both hurdles and opportunities.

Nuclear magnetic resonance (NMR) spectroscopy has been used for determination of *in vivo* metabolite levels in intact cells and cell extracts (Nicholson et al., 1999; Raamsdonk et al., 2001; Warne et al., 2000) but the low sensitivity of the method in most laboratories restricts its use (Hartbrich et al., 1996). More recently, a novel NMR spectroscopic approach to the direct biochemical characterization of bacterial culture broths was presented (Abel et al., 1999). Low molecular weight organic components of broth supernatants from cultures of *Streptomyces citricolor* were analysed using one- (1D) and two-dimensional (2D) $^1$H-NMR spectroscopic methods; it was possible to identify and monitor simultaneously a range of media substrates and excreted metabolites which included 2-phenylethylamine, trehalose, succinate, acetate, uridine and aristeromycin. Signals were extensively overlapped in the $^1$H-NMR spectra of the whole broth mixtures, so directly coupled HPLC–NMR spectroscopy was also applied to the analysis of broth supernatants to aid spectral assignments. Multiple bond correlation for structural elucidation and peak assignments of individual components was also conducted using 2D NMR methods based on $^1$H–$^1$H and $^1$H–$^{13}$C correlations. This work showed that high-resolution NMR spectroscopic methods could provide a rapid and efficient means of investigating microbial metabolism directly without invasive or destructive sample pretreatment. van Eijk developed a high-throughput screening method using LC–MS (Caceres et al., 2000; van Eijk et al., 1999) to study the amino acids, in which the simultaneous application (and thus measurement) of multiple amino acid tracers was used coupled to liquid chromatography and mass spectrometry, resulting in the measurement of both the concentration and isotope enrichment of *O*-phthaldialdehyde (OPA)-derivatized plasma amino acids in one run. Considering the easier and cheaper derivatization procedure and instrumentation, the simultaneous collection of isomeric distribution spectra (enabling the application of multiple labelled components) and concentration data, the method presents an attractive alternative to traditional GC–MS applications for amino acids. Therefore, combining liquid chromatography with electrospray mass spectrometry (LC–ESI–MS) methods is attractive (Buchholz et al., 2001; Cole, 1997; Gaskell, 1997; Magera et al., 2000) and recent innovations with extended capabilities allow both small metabolite and large biomolecular analysis via LC–ESI–MS (Krishnamurthy et al., 1999) and direct infusion and flow-injection ESI–MS (Goodacre et al., 1999, 2000, 2002, 2003; Vaidyanathan et al., 2001, 2002). In particular, LC–DAD–MS, MS–MS and MALDI–TOF may be highly automated, opening up high-throughput screening (HTS) and easier and simpler data acquisition and analysis.

Ferenci and co-workers have adopted a metabolomic approach to the analysis of *E. coli* (Liu et al., 2000; Tweeddale et al., 1998, 1999). They used radioisotopic labelling and extracted the cells, and apparently the medium, with a final concentration of 67% boiling ethanol for 30 min. One may assume that only the most stable metabolites survived this treatment, and indeed only about a dozen were identified via 2D thin-layer chromatography. We have chosen to use mass spectrometry for our analyses as it is a sensitive and rapid technique, does not require radio-isotopes, and is potentially capable of discriminating many more metabolites (than TLC) from their mass/charge ratio alone (and if such is available to identify many more via tandem analyses; Rashed et al., 1997; Vaidyanathan et al., 2002). Additionally, FT–IR, a method which measures the overall composition of a sample by detecting the molecular vibrations and other motions of chemical bonds, is another excellent method for rapid screening of microbial samples (Goodacre et al., 1998a, 1998b; Naumann et al., 1995; Oliver et al., 1998; Timmins et al., 1998).
Table 1. Genotypes of strains provided by Professor Yanofsky and used in the present work

<table>
<thead>
<tr>
<th>W3110</th>
<th>Wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CY15682</td>
<td>TrpR2 (repressor minus)</td>
</tr>
<tr>
<td>CY15500</td>
<td>TnaA2 (tryptophanase minus)</td>
</tr>
<tr>
<td>CY15001</td>
<td>TrpR2 tnaA2</td>
</tr>
<tr>
<td>CY15602</td>
<td>ΔtrpEA2 (trp operon deleted)</td>
</tr>
<tr>
<td>CY15680</td>
<td>TrpR2 ΔtrpEA2 tnaA2 trpA46PR9</td>
</tr>
<tr>
<td>CY15681</td>
<td>TnaA2 tnaA46PR9</td>
</tr>
</tbody>
</table>

However, metabolomic fingerprinting of microbial strains has two major difficulties: (a) the turnover time of the intracellular metabolites can be very fast (De Koning and van Dam, 1992), necessitating the very rapid quenching of metabolism; and (b) the intracellular volume of a typical microbial suspension at a concentration of 1 mg wet weight/ml occupies only some 0.1% of the total volume, and removing the small intracellular volume from the much larger extracellular volume can (indeed, is likely to) lead to contamination of the former by the latter. However, given that microbes are rarely if ever enjoying balanced growth, and must (and do) secrete any number of substances into the medium, it occurred to us that we might make a virtue of necessity by exploiting the fact that what is secreted must reflect the exact genetic make-up of the strain in question (Allen et al., 2003) and might therefore be used, according to the principles of ‘guilt by association’ (Oliver, 2000) or supervised learning (Kell and King, 2000), for the purposes of functional genomics in gene knockout strains. What was not known, however, was whether the metabolic footprints would be either sufficiently reproducible or discriminating as a different kind of ‘fingerprinting’ technique (Fiehn, 2001) to allow such discrimination in E. coli. We therefore decided to test this hypothesis explicitly.

We investigated tryptophan metabolism mutant strains selected by Yanofsky and co-workers (Khodursky et al., 2000a; Yanofsky and Horn, 1994), who studied changes in gene expression profiles of the strains in response to tryptophan-supplemented and partially or totally tryptophan-starved conditions in defined media (Tao et al., 1999). These strains had deletions in tryptophan repressor gene (trpR) or tryptophanase gene (tnaA2) (Kamath and Yanofsky, 1992), tryptophan operon deletion (trpEA2) and a leaky auxotroph, trpA bradytroph (trpA46PR9) (see Table 1). Hierarchical clustering of the profiles revealed changes in expression of a total of 691 genes with identification and functional roles assigned to 169 of the genes. As the transcriptome profiling showed specific functional alterations (Featherstone and Broadie, 2002; ter Kuile and Westerhoff, 2001) we predicted that there might be metabolite changes as well and these changes were monitored in filtered culture supernatants and sample analysis [after cellular metabolite (‘fingerprinting’) extraction using both perchloric and hot ethanol methods after LC separation of a range of metabolites and direct culture supernatant (‘footprinting’), using FT–IR and ESI–MS. The FT–IR and MS spectral profiles were processed chemometrically (Beavis et al., 2000; Goodacre and Kell, 2003; Raamsdonk et al., 2001; Shaw et al., 1999a; Smith, 1998).

Materials and methods

Strains, kindly provided by Professor Yanofsky, with genotypes as shown in Table 1, were grown in defined media composed of 0.2 g glucose, 0.2 g MgSO4·7H2O, 2 g citric acid, 10 g anhydrous K2HPO4 and 3.5 g NaNH4HPO4·4H2O per litre (Heatwole and Somerville, 1992), with added indole acrylate or tryptophan where stated. The inoculum was grown under the same conditions as the experimental flasks in 2 ml volumes in sterile tubes for 24 h at 30°C and 200 rpm. Experimental 500 ml flasks with 100 ml medium were grown for 24 h, as above. The final OD600 of the cultures was approx 2, equivalent to final cell densities of some 109/ml.

Intracellular metabolite (fingerprinting) extraction methods

Duplicate samples for metabolite extraction were withdrawn 24 h after growth.
Perchloric acid extraction

A 5 ml volume of a culture was squirted into 15 ml −40 °C pre-chilled 60% methanol buffered with 70 mM HEPES/KOH, pH 7.5, mixed rapidly and centrifuged at −20 °C for 5 min at 5000 × g. The supernatant was discarded and the pellet was resuspended in 2 ml 35% (v/v) perchloric acid and stored at −80 °C for 1 h, thawed on ice and centrifuged as before. A 2 ml volume was withdrawn from the supernatant and neutralized with 4 × 200 µl 5 M K2CO3 (pH checked and adjusted if necessary). The sample was frozen at −80 °C, thawed and pH-monitored again. It was re-centrifuged and the supernatant was removed, aliquoted and stored frozen at −40 °C. For the control, a 5 ml volume was treated similarly to the metabolite extraction with perchloric acid (Meyer et al., 1999). Perchloric acid was preferred here to ethanol extraction as many more metabolites were extracted (unpublished; but see also Buchholz et al., 2001).

Culture supernatant sample preparation (footprinting)

Extracellular secretion of intracellular metabolites was monitored in the culture medium after 24 h growth of the bacterial strains under varying growth conditions at 30 °C and 200 rpm, with cells removed using 0.22 µm filter units. Controls were fresh media at 0 h and wild-type grown on freshly prepared medium on three separate days with samples taken for analysis.

In this type of strategy (Hastie et al., 2001; Kell and King, 2000), the accepted norm is to ‘train’ using a subset of samples and project in the data from a different set of replicates to ensure (i.e. demonstrate) that one is not overfitting the data. Thus we used a number of different replicates from different days for this purpose (see legends to Figures).

FT–IR analysis

Analysis by FT–IR with automated HTS was carried out with each sample run as six replicates of 10 µl volume/well on 100-well aluminium plate (Goodacre et al., 1998a). The plate was oven-dried at 50 °C for 30 min prior to analysis and loaded onto the motorized stage of a reflectance TLC accessory of a Bruker IFS28 FT–IR spectrometer (Bruker Spectrospin, Coventry, UK) equipped with a mercury–cadmium–telluride (MCT) detector cooled with liquid N2. The spectral range was 4000–600 cm−1 and 256 co-adds were used (Winson et al., 1997).

ESI–MS analysis for footprinting

A Waters Alliance 2690 HPLC linked to a Photodiode array detector 996 (DAD) and Micromass LCT electrospray mass spectrometer were used for analysis of the metabolites. A 10 µl sample volume was first separated on a 200 × 4 mm chiral Nucleodex β-OH column using 12 mM ammonium acetate : methanol (99 : 1) as eluent at a flow rate of 500 µl/min and an isotropic gradient with a 10 min metabolite separation and 30 min column wash. A 40 µl/min stream was directed after splitting the volume into the MS for further analysis in the range 65–815 m/z. The MS was optimized with capillary voltage at 2000 V, source temperature 80 °C, desolvation temperature at 150 °C, nebulizer and desolvation gas flow at 90 and 540 l/h, and sample cone and extraction cone voltage at 40 V and 11 V, respectively.

ESI–MS for footprinting

The automated analysis was performed using the same instruments as above. Samples were diluted 10-fold in 30% HPLC grade methanol and 0.1% formic acid made up to volume with HPLC grade water. The samples were de-gassed and large particles were removed by microcentrifugation (Eppendorf microfuge) at full speed for 3–5 min. Volumes of 100 µl were dispensed into pre-labelled glass inserts and placed in tubes in the LC carousel. A 20 µl volume sample was loaded into the sample loop using LC solvents (70% 10 mM formic acid/30% HPLC grade methanol) and pumped at 0.5 ml/min. The total scan cycle was 1 s (0.9 s scan and 0.1 s interscan delay) and the complete run time was 2 min. The MS was optimized, leading to the following final conditions: capillary voltage at 2000 V, source temperature at 80 °C, desolvation temperature at 150 °C, nebulizer and desolvation gas flow at 90 and 540 l/h, and sample cone and extraction cone voltage at 40 V and 11 V, respectively (Vaidyanathan et al., 2001).
Chemometric data processing methods

The FT–IR and mass spectrometric methods described above produce vast amounts of potentially useful data (Benton, 1996), e.g. LC–MS produces a spectrochromatogram (an array of the MS vs. time) for each sample analysed plus a diode array detector trace. Each spectrochromatogram can typically hold 10⁶ values (depending upon the MS range and sampling rates). In their native form, such data are extremely difficult to interpret. To turn such data into information of chemical or biological interest, some sort of multivariate statistical analysis must be employed.

Data processing for FT–IR spectral analysis

Raw data were exported to MATLAB as a matrix object using the Opus software (Bruker Spectrospin). Data preprocessing was carried out on autoscaling by normalization to unit variance (Winston et al., 1997).

Cluster analysis of FT–IR spectra

Principal components analysis

PCA (Causton, 1987; Jolliffe, 1986; and see below) was performed on the original data set to give a new reduced set of orthogonal variables called principal components (PCs), the first few of which typically account for >95% of the variance.

Discriminant function analysis

DFA is a supervised projection method (Manly, 1994); a priori information about sample grouping in the data set is used to produce measures of within-group variance and between-group variance. This information is then used to define discriminant functions that optimally separate the a priori groups (in this case the groups are defined as replicates). In this implementation, the first n PC scores are used as the data source for DFA, where n is chosen using cross-validation (Radovic et al., 2001).

ESI–MS preprocessing

In order to simplify any subsequent statistical analysis, two simple pre-processing algorithms were applied to the ESI–MS spectrochromatograms. First, each ESI–MS array was reduced into a single ‘aggregate’ MS vector by summing the ion counts of a given m/z ratio over the total scan cycle. Each MS vector was then ‘binned’ to unit m/z ratio (i.e. ion counts of fractional m/z ratios were added to the nearest integer m/z). Thus, after this initial data reduction an ESI–MS spectrochromatogram with MS range 65–815 m/z will be reduced to a single vector having 750 values. This is a highly efficient strategy since, depending on the scan rate, the file sizes are reduced from tens of megabytes to a few kilobytes.

Multivariate analysis

Before employing any multivariate analysis each MS vector is normalized to the total ion count (which is given a value of 10⁶). This is done so that different spectra can be compared quantitatively. Once a set of N spectra (with mass range p) is concatenated into a single matrix (N objects × p variables) each column of the data set can be optionally normalized to unit variance. This is done to eliminate bias, in subsequent analysis, toward any column that contains either large absolute values or large variances (Martens and Næs, 1989). However, we note that normalization can sometimes be more detrimental than helpful. If there are a large number of redundant variables in the data, the noise on such variables is amplified to the same importance as relevant variables. This can easily cloud any underlying statistical trends.

In order to cluster the spectral data, principal components analysis (PCA) was used (Causton, 1987; Jolliffe, 1986). PCA involves projecting the original X-matrix (N × p) onto a d-dimensional subspace using a projection (or loading) matrix, thus creating object coordinates (a score matrix) in a new coordinate system. This is achieved by the method known as singular value decomposition (SVD) of X:

\[ X_{N \times p} = U_{N \times d} \Lambda_{d \times d} L_{p \times d}^T = T_{N \times d} L_{p \times d}^T \]

where \( U \) is the unweighted (normalized) score matrix and \( T \) is the weighted (or biased) score matrix. \( L \) is the loading matrix, where the columns of \( L \) are known as eigenvectors or loading-PCs. \( \Lambda \) is a diagonal matrix (i.e. all of the off-diagonal elements are equal to zero) containing the square roots of the first d eigenvalues of the co-variance matrix (\( X^T X \)) where, \( d < N \) and \( d < p \).
The principal components (PCs) can be considered as a basis set used to project the original data matrix, $X$, onto the scores, $T$. In other words, the new coordinates are linear combinations of the original variables, e.g. the elements of the first principal component can be represented as:

$$t_1 = x_{11}l_{11} + x_{12}l_{21} + \ldots + x_{1p}l_{p1}$$

$$t_2 = x_{21}l_{11} + x_{22}l_{21} + \ldots + x_{2p}l_{p1}$$

$$\vdots$$

$$t_n = x_{n1}l_{11} + x_{n2}l_{n1} + \ldots + x_{np}l_{p1}$$

The influence of each of the original variables on the new PCs (i.e. the contents of the loading matrix) is determined on the basis of the maximum variance criterion. The first PC is considered to lie in the direction describing maximum variance in the original data. Each subsequent PC lies in an orthogonal direction of maximum variance that has not been considered by the former components. The number of PCs computed for a given data set is up to the analyst; however, usually as many PCs are calculated as are needed to explain a pre-set percentage of the total variance in the original data (the total number of PCs possible is equal to the number of original variables). It is also possible to use PCA analysis on a subset of the variables chosen via a genetic algorithm (GA) (Broadhurst et al., 1997) and we have exploited such GA–PCA analysis of the ESI–MS data here. In particularly favourable cases the discrimination can be made on the basis of just two or three variables, which allows the display of data in a 2D or 3D plot of the actual variables themselves (as opposed to the PCs: Taylor et al., 1998). Finding these variables is a combinatorial optimization problem (Cook et al., 1998), as the number of pairs and triplets which can be formed from 750 (i.e. the mass spectral) variables is respectively 280 875 and 70 031 500; hence the need for the GA.

**Results and discussion**

The synthesis, utilization and degradation of tryptophan in *E. coli* has been studied extensively, with regulation of its operon being effected by both repression and attenuation (transcription termination) (Yanofsky, 2000; Yanofsky and Horn, 1994; Yanofsky et al., 1993, 1996). More recently a selected set of mutants and wild-type W3110 (control) strains were used to study the changes in expression profiles in response to altered tryptophan availability during early growth phase, and 15 genes organized in nine operons exhibited changes. The set of experiments conducted here for our study were different from those of Khodursky et al. (2000b) as the cultures were grown for 24 h into stationary phase. The wild-type W3110 and mutant strains with *trpR2* (repressor minus), *tnaA2* (tryptophanase minus), *trpEA2* (tryptophan operon minus) and *trpA bradytroph* were grown under three different growth conditions: in minimal medium; in the presence of excess tryptophan; and tryptophan starvation induced by indole acrylate (a tryptophan analogue). Indole acrylate prevents tryptophan repression (Ilic et al., 1994) by inhibiting the charging of tRNAtrp by tryptophanyl-tRNA synthetase, which in turn effects both repression and attenuation (transcription termination) of the tryptophan operon. As arginine biosynthetic genes are sensitive to tryptophan starvation, very mild starvation conditions were imposed and ideally a study would involve the use of near-isogenic strains (Khodursky et al., 2000b). Intracellular metabolite data are not displayed in this report as no meaningful clustering of replicates was observed. However, preliminary experimental results had indicated that differences between strains could be detected using the filter-sterilized media samples from cultures grown for 24 h, and therefore these samples were analysed using FT–IR and mass spectrometry.

**FT–IR analysis of *E. coli* tryptophan metabolism mutants**

FT–IR spectral analysis is used routinely in our laboratory for high-throughput screening of a wide range of microbes (Goodacre et al., 1998b; Oliver et al., 1998) and their products (McGovern et al., 1999, 2002; Shaw et al., 1999b; Winson et al., 1997, 1998). The data produced by FT–IR spectroscopy are multidimensional and thus chemometric data analysis is required. Additionally, characteristic vibrations can lead to the identification of specific metabolites (e.g. Goodacre et al., 2000; Johnson et al., 2000; McGovern et al., 2002).

The DFA biplots of FT–IR data from replicate samples are shown in Figures 1, 2 and 3, with representative FT–IR spectra being illustrated in Figure 1A. Figure 1 shows *E. coli* W3110 wild-type strain grown for 24 h in normal growth...
Footprinting set 3/2

Absorbance (arbitrary units)

Wave number (cm\(^{-1}\))

DFA model built and cross validated using the first 10 PCA scores

Copyright © 2003 John Wiley & Sons, Ltd.
Metabolic footprinting in *E. coli*

medium without any additions (1), in medium supplemented with 50 µg/ml tryptophan (2) or indole acrylate at 10 µg/ml (3) and 15 µg/ml (4). Clustering of the five replicates for each of the samples is well defined. That the projected spectra are recovered in the correct group clearly demonstrates the high reproducibility of FT–IR here. The samples in the presence of added tryptophan (2) and/or indole acrylate (3 and 4) are clearly separating away from the normal minimal growth medium (1) and tryptophan samples cluster away from the indole acrylate samples.

Figure 2 shows the DFA biplot of selected tryptophan mutant strains of *E. coli* grown in minimal medium only. The replicates and projected data cluster together for each sample and W3110 (1) clusters away from the other strains. The tryptophanase-negative strains, tnaA2 (2) and tnaA2 bradytroph (3) cluster together, whereas trpR2 mutant (4) harbouring a repressor deletion clusters away from the other strains. The DFA of selected tryptophan mutant strains grown in the presence of added tryptophan in Figure 3 show that the tryptophan repressor, trpR2, deletion strain and...
DFA model built and cross validated using the first 18 PCA scores

Figure 3. DFA biplot of FT–IR data showing the relationship between culture media from wild-type W3110 (1), 15682 trpR2 (2), 15000 tnaA2 (3), 15001 trpR2 tnaA2 (4), 15602 ΔtrpEA2 (5) and 15680 trpR2 ΔtrpEA2(6) strains grown in minimal media supplemented with 50 µg/ml tryptophan (which causes tryptophan repression). Cross-validation of the DFA model was performed, whereby the original data set was divided into two subsets, one of which was used to train the model (closed circles) and the other subsequently used to validate it (open circles). This process serves to ensure that the optimal number of principal components (PCs) is used to build the DFA model; in this case, 18 PCs were needed.

total tryptophan operon deletion (6) cluster away from the other strains and the W3110 wild-type but are closer to the strain with the operon deletion only (5). Similarly, the strain with only trpR2 deletion (2) clusters away from all the other strains but is closer to tnaA2 and trpA2 deletions (4) and tnaA2 deletion (3) strains.

The distinct pattern of clustering of media using FT–IR data analysis derived from a single strain cultured under diverse growth conditions clearly suggested that there are obvious changes in the extracellular metabolite composition. This could be induced by growth media supplemented with tryptophan or indole acrylate, because tryptophan metabolism is tightly regulated by the presence of tryptophan and indole acrylate in the medium. These changes could also be attributed to the uptake of nutrients from the growth medium or secretion of intracellular metabolites into the medium during growth. Using a subset of strains carrying defined gene mutations, DFA analysis shows distinct clusters, which are wholly reproducible at the mutant level, as confirmed by the projection of ‘unknown’ biological replicates into PC-DFA space. Strains with single gene deletions for tryptophanase and/or tryptophan repressor proteins show DFA clusters displaced from those of the strains carrying the polycistronic deletion of tryptophan operon. Thus a marked effect on tryptophan metabolism is generated by single or multiple gene deletions. In conclusion, clustering using FT–IR analysis can easily separate strains according to their genotype and thus metabolomics can provide a rapid high-content screen for genetic lesions.

ESI–MS analysis of E. coli tryptophan metabolism mutants

Additionally, these samples were also analysed using direct injection mass spectrometry, which has been used successfully to identify bacteria
from crude cell-free extract preparations via a complex milieu of large and small chemicals (Magera et al., 2000; Morris and Cooper, 2000; Tiller et al., 2000; Vaidyanathan et al., 2001, 2002; van Eijk et al., 1999). Samples of 24 h culture media for this study were also analysed using ESI–MS in the positive ion mode, as we were focusing on changes in tryptophan metabolites. Mass spectrometric data are also high-dimensional and must be preprocessed before chemometric analysis. Representative ESI–MS spectra are given in Figure 4A.

Figure 4. (A) Illustrative mass spectral data from stationary phase supernatants of wild-type strains of E. coli prepared as described in the text, (a) three separate experiments to show reproducibility and (b) in different media. (B) GA–PCA-derived plot of ESI–MS data of culture media from wild-type W3110 grown in the presence of 50 µg/ml tryptophan (1), 10 µg/ml (2) and 15 µg/ml (3) indole acrylate for 24 h. The MS data of wild-type W3110 grown in minimal medium alone were subtracted as background masses. The axes represent the normalized ion counts of the stated m/z variables.

Copyright © 2003 John Wiley & Sons, Ltd.
The GA–PCA-derived plots of ESI–MS showing selected $m/z$ ions for filter-sterilized culture media of wild-type W3110 and the other strains grown for 24 h are shown in Figures 4, 5 and 6. All plots show very close clustering of the six replicate samples. In order to remove the effect of the wild-type’s growth on the metabolic footprints, the MS data of the wild-type, E. coli W3110, grown only in minimal medium, were subtracted as a ‘background’ from all the sample data shown in these figures.

The analysis showed that the samples of wild-type grown in different media in Figure 4 could be clearly discriminated using just two analyte ions. The $m/z$ 190 and 260, alone or together, allowed clear separation of the wild-type grown in supplemented medium with indole acrylate at 15 $\mu$g/ml (2), showing the greatest variance with $m/z$ 260 at around 1150 normalized ion counts (NIC) and with $m/z$ 190 at around 170 NIC. In the presence of indole acrylate at the lower concentration of 10 $\mu$g/ml (1), there were around 360 NIC of $m/z$ 260 and around 75 NIC for $m/z$ of 190. By contrast, medium supplemented with tryptophan (0) showed no significant discrimination using either of these masses and clustered around zero at the origin.

Figure 5 shows a 2D $m/z$ plot of 260 vs. 381 derived from GA–PCA of the strains grown in minimal media only. The trpR2(7) strain clearly separated from the others with $m/z$ of 381 and about 250 NIC and tnaA2 trpA46PR9 bradytroph (6) separated with $m/z$ 260 only with an NIC of around 450. The tnaA2 (5) strain clustered around zero, suggesting that it was similar (at least in these two analytes) to the wild-type.

The GA–PCA-derived plot of the organisms grown in the presence of 50 $\mu$g/ml tryptophan is shown in Figure 6. This pseudo-3D plot of $m/z$ 115243 and 288 clearly distinguished the trpR2 (10), tnaA2 (11), trpR2 tnaA2 (12), $\Delta$trpEA2 (13) and trpR2 $\Delta$trpEA2 (14) into five tight clusters. The degradation of tryptophan leads to indole (Goodacre and Kell, 1993; Prinsen et al., 1997), pyruvate and ammonia, and the MS analysis of the supernatant medium shows that a mass of 288 clearly discriminates trpR2 tnaA2 and $\Delta$trpEA2.
Figure 5. GA–PCA-derived plot of ESI–MS data of culture media from 15000 tnaA2 (1), 15681 tnaA2 trpA46PR9 bradytroph (2) and 15682 trpR2(3) strains grown in minimal media only for 24 h. The MS data of wild-type W3110 were subtracted as background masses from the MS data of each of the mutant strains. The axes represent the normalized ion counts of the stated m/z variables.

Figure 6. Pseudo-3D GA–PCA-derived plot of ESI–MS data of culture media from 15682 trpR2 (1), 15000 tnaA2 (2), 15001 trpR2 tnaA2 (3), 15602 ΔtrpEA2 (4) and 15680 trpR2 ΔtrpEA2(5) strains grown in minimal media supplemented with 50 µg/ml tryptophan for 24 h. The MS data of wild-type W3110 were subtracted as background masses from the MS data of each of the mutant strains. The axes represent the normalized ion counts of the stated m/z variables.
strains. Indole-3-glycerol phosphate is the penultimate intermediate of tryptophan synthesis, and has a mass of 287, which, when protonated in positive-ion ESI–MS, gives it an m/z of 288 (Mohammed et al., 1999). It is thus highly likely that the m/z 288 analyte is therefore indole-3-glycerol phosphate, and a functional genomics strategy with access to a tandem instrument would establish this.

In conclusion, these rapid spectroscopic methods allowed us to discriminate these closely related single-gene knockout strains from their metabolic footprints alone. Thus, they can be used to detect small phenotypic differences that other conventional phenotyping and global profiling approaches would miss, opening up the possibility of gaining useful information from knockouts with subtle phenotypes, especially in functional genomics studies with large libraries of such gene knockouts. The footprinting approach, which does not rely upon the identification of any peaks, can be used without prior knowledge of the likely function of the genes of interest, and can supply data that could indicate potential functions for genes. The FT–IR is rapid but is chemically unselective, and is better for a very rapid ‘fingerprinting’ type of study in which it is not of great interest to identify the metabolites of interest (Fiehn, 2001). By contrast, the ESI–MS is slightly slower but can give an indication of the metabolites contributing to the differences between the strains. More definitive identification would require other methods, such as tandem mass spectrometry. Nevertheless, as rapid and reagentless approaches, FT–IR and ESI–MS of metabolic footprints are both much quicker and cheaper than are transcriptomics and proteomics.

Acknowledgements

We thank the BBSRC for financial support, and Simon Andrews, Arne Buchholz, Simon Doig and Sue Stovell for useful discussions.

References


Caceres A, Cardenas S, Gallego M, Rodriguez A, Valcarcel M. 1996. Simultaneous applications to pyrolysis mass spectrometry of the differences between the strains. More definitive identification would require other methods, such as tandem mass spectrometry. Nevertheless, as rapid and reagentless approaches, FT–IR and ESI–MS of metabolic footprints are both much quicker and cheaper than are transcriptomics and proteomics.

Acknowledgements

We thank the BBSRC for financial support, and Simon Andrews, Arne Buchholz, Simon Doig and Sue Stovell for useful discussions.

References


Copyright © 2003 John Wiley & Sons, Ltd.
Metabolic footprinting in E. coli


Dunn M. 1998. Proteome analysis: chemists are analysing the total protein content of organisms to gain insights into complex cellular processes and accelerate the drug discovery process. Chem Br 34: 54–58.


