Assessment of adaptive focused acoustics *versus* manual vortex/freeze-thaw for intracellular metabolite extraction from *Streptomyces lividans* producing recombinant proteins using GC-MS and multi-block principal component analysis

Yankuba Kassama,[*a] Yun Xu,[a] Warwick B. Dunn,[b] Nick Geukens,[c] Jozef Anné[c] and Royston Goodacre[a]

We compared the gas chromatography-mass spectrometry (GC-MS) metabolite profiles of mouse tumour necrosis factor alpha (mTNF-α) secreting *Streptomyces lividans* TK24 to the non-secreting wild type and the wild type harbouring the empty pIJ486 plasmid by multi-block principal component analysis (PCA). The multi-block PCA model successfully identified peaks that were statistically different between the protein secreting and non-secreting strains, and at the same time also uncovered the efficiency of intracellular metabolite extraction by an ultrasonic adaptive focused acoustics (AFA) technique compared to a manual vortex/freeze-thaw method. Fifty-one metabolites were significantly different between the three biological strains and 17 of these were abundant in the mTNF-α secreting strain compared to the non-secreting strains. No significant differences in the number of detected metabolite peaks were observed between the two extraction techniques. However, from the loadings of the multi-block PCA model, as well as univariate statistical analysis, we observed that the relative peak response ratios to the internal standard of 10 metabolites were higher for the AFA extraction, suggesting a more efficient recovery of these metabolites than achieved with the manual vortex/freeze thaw method.

Introduction

Streptomycetes naturally synthesise and secrete a variety of secondary metabolites which are extensively used in the agricultural, environmental and pharmaceutical industries.[1,2] This natural ability has since been exploited to develop expression systems in *Streptomyces lividans* to enhance the secretion of heterologous proteins (so-called biopharmaceuticals).[3–7]

Physiological and genetic factors influencing extracellular protein secretion are already well described in *S. lividans*.[8–10] However, there are very limited metabolome studies that investigate heterologous protein biosynthesis and secretion in actinomycetes.[11] Such studies, as already described in other bacteria,[1,12,13] could identify biosynthetic bottlenecks or develop models for the more efficient expression and subsequent secretion of desirable protein products.

Data acquired from metabolome studies, specifically metabolic profiling where a global and holistic snapshot of the metabolome is obtained, can be influenced by sample collection and treatment methodologies.[14–16] An accurate measurement of intracellular metabolites requires: 1) an effective quenching technique to inhibit metabolic activities and maintain physiological metabolic profiles in analysed samples; 2) an extraction method that allows a high recovery of intracellular metabolites present in a wide range of metabolite classes; and 3) an analytical platform for detecting and quantifying metabolites.[12,17] Whilst the platforms for metabolite detection and data analysis may be relatively similar for a range of sample types, various quenching and extractions techniques are often required for different sample types where cell structure, metabolic activity and cell wall integrity can all influence the efficiency of quenching and extraction procedures.[18]

Adaptive focused acoustics (AFA) involves the localization of high frequency acoustics energy on to sample targets by a transducer. The energy creates rapid cavitation events through the formation and collapse of bubbles leading to cell and tissue disruption. Although the general principle of cell lysis is similar to normal laboratory sonications,[19] the main difference is the operating wavelengths. Generally, sonicators operate at a low frequency (1 KHz) and long wavelength (100 mm) compared to 1 MHz and 1 mm for AFA.[20] This shorter wavelength and higher frequency allows the energy to be focused at a sample target and hence prevents the dissipation of sonication energy leading to heat spots. A controlled shorter wavelength also enables the energy to traverse sample tubes so that the samples are not in direct contact with the probe. In addition, the AFA was coupled to a computerized system that enabled a precise control of the acoustics frequency, temperature and length of sample treatment. The AFA technique is therefore an isothermal and non-contact process that has been employed for cells and tissue disruption to recover target molecules.[21]

[a] *School of Chemistry, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom*
[b] *Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom*
[c] *Laboratory of Bacteriology, Rega Institute, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium*

This study thus compares the efficiency of ultrasonic AFA and manual vortex/freeze-thaw extraction techniques for comparative metabolite profiling of mouse tumour necrosis factor alpha (mTNF-α) expression in *S. lividans* TK24 against the non-secreting wild type and empty pIJ486 plasmid strains; pIJ486 is the vector for mTNF-α and so represents an additional control.

Methods

Bacterial growth

Three strains of *S. lividans* TK24: i) wild type (labelled 'W'), ii) empty pIJ486 plasmid (P) and iii) mTNF-α (T)) were grown as seven biological replicates each in 250 mL Erlenmeyer shake flasks containing 50 mL phage medium (per litre: 10 g glucose, 5 g tryptone, 5 g yeast extract, 5 g Lab Lemco, 0.74 g CaCl$_2$.2H$_2$O, 0.5 g MgSO$_4$.7H$_2$O, pH: 7.2) for 72 h at 28 °C and 300 rpm in a Minitron shaker incubator (INFORS-HT Bottmingen Switzerland).

Sample preparation

Two replicates from each of the seven fermentation cultures (5 mL) were quenched in 25 mL 60% aqueous methanol containing 10 mM HEPES (−40 °C). The samples were then immediately centrifuged at 4000 g for 5 min at −9 °C, supernatant removed and the cell pellets snap-frozen in liquid nitrogen and stored at −80 °C.

Intracellular metabolites from one set of the quenched replicate samples were extracted in 1 mL 100% methanol (−40 °C). The manual extraction involved three cycles of 1 min vigorous manual vortex followed by freeze-thaw in liquid nitrogen and thawing on ice. The second replicate samples were suspended in 1 mL 100% methanol and then transferred into 5 ml glass screw-cap vials for the ultrasonic adaptive focus acoustics treatment in a Covaris™ S1 single tube system (Covaris Inc., Woburn, MA, US). The treatment setting was 20% duty cycle, 10 mV intensity, 200 cycles per burst for 2 min per sample and the instrument chiller was set to 4 °C. All the extracts were centrifuged at 4000 g for 10 min at −9 °C and the supernatant collected and stored at −80 °C for GC-MS analysis.

GC-MS

GC-MS analysis was performed as originally described by Pope *et al.*[22] In short, 900 μL of extract was spiked with 100 μL of an internal standard solution (0.16 mg mL^{-1} succinic d_4 acid), vortex mixed for 15 s and lyophilised overnight (HETO VR MAXI vacuum centrifuge attached to a HETO CT/DW 60E cooling trap; Thermo Life Sciences, Basingstoke, UK). The lyophilised samples were chemically derivatised with 20 mg mL^{-1} *O*-methylhydroxyamine in pyridine (50 μL) and the mixture heated at 40 °C for 90 min. *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (50 μL) was added and the mixture further heated for 90 min at 40 °C. The derivatised samples were analysed by GC (Agilent 6890 gas chromatograph; Agilent Technologies, Wokingham, UK) coupled to a LECO Pegasus III time-of-flight mass spectrometer, (TOF/MS; LECO, Stockport, UK). Raw data baseline correction, deconvolution, and metabolite identification by searching against an in-house prepared

library and the NIST/EPA/NIH02 (http://www.nist.gov/srd/nist1.htm) library, were performed using ChromaTof version 2.12. Processed data were exported in ASCII format as a list of metabolite peaks and associated peak areas. All data were normalized to the internal standard (response ratio = peak area metabolite/peak area internal standard). All data analysis was performed in Matlab (Mathworks, MA, US).

Data analysis

This study compared two different extraction methods (analytical factor) and three different strains (biological factor) both of which will have significant influences on the GC-MS metabolic profiles. It is also possible that there might be some delicate interactions between these two factors, *e.g.*, some extraction methods might be more effective on one strain than the others. Therefore it is unrealistic to assume that each class forms a homogeneous distribution in metabolite hyperspace. As suggested by Cattell, when classical factor analysis models like PCA are applied to heterogeneous samples, the latent factors obtained will be "neither clear species differentiators nor optimal individual differentiators".[23] In univariate analysis, such a problem is normally solved by rearranging data into "blocks" so that in each block there is only one influential factor to be tested and use 2-way (or N-way if there were more than 2 influential factors) ANOVA to assess if the influence of each factor is statistically significant or not. In this study, we extend such a methodology to multivariate factor analysis, where samples are rearranged into blocks and subjected to multi-block PCA.

There have been no less than four multi-block PCA models reported[24] and each of them has their own unique properties.[25] Nevertheless, all of these models are designed to find the common trend between different blocks and display it in their latent factors which are commonly refered to as the 'super scores'. In addition, multi-block PCA also provides' 'block scores' and 'block loadings' which are similar to the scores and loadings in classical PCA except they represent the relative position of samples and contribution of variables within each block. Such a characteristic can be very useful to deconvolute two interacting factors like the problem within this study. Given two factors, *A* and *B*, which have *a* and *b* levels respectively; to examine whether factor *A* has significant influence on the data, we can rearrange data into *b* blocks so that each block contains all samples at the same level as factor *B*. By doing so, the influence of factor *B* no longer exists in the row space while the influence of *A* becomes a common trend between different blocks which should be modelled by the multi-block PCA (if it does exist) and can be seen in the super scores plot. Similarly, if we rearrange data in to *a* blocks, the influence of factor *B* should be revealed in the super scores plot as well. In this study, a multi-block PCA model called consensus PCA-W (CPCA-W), a CPCA[26] model modified by addition of a normalising block loadings step, was employed. Before building the model, the data were first arranged into blocks and then each block was mean centred and standardised so that each variable has unit standard deviation.

First, the data were rearranged into two blocks; each block contained all the samples using the same extraction method (either manual vortex or AFA). Another way was to rearrange

data into three blocks; each block contained all the samples of the same strain (T, W or P). Then we applied CPCA-W on these rearranged data sets to determine the significant differences between the extraction methods (from the 3-blocks model) and biological variability (from the 2-blocks model) within the metabolic profiles. In addition, like classical PCA, the variables which contribute to the separation shown in the scores can be found in the block loadings plot.

Furthermore, we applied the Friedman test (a non-parametric version of balanced design two-way ANOVA) on each variable to identify statistically significant variables and examine the consistency between the variables discovered by CPCA-W block loadings and those discovered by the Friedman test.

It is important to note that although CPCA-W is still an unsupervised method, by rearranging data into blocks, we did make use of some *a priori* knowledge. Hence it is necessary to validate the results revealed by the CPCA-W model. Therefore we tested the robustness of the trend revealed in the CPCA-W model *via* a cross-validation coupled with a simple distance classification according to the following procedure:

(1) Remove one sample (all profiles from a biological replicate) from each block and build a CPCA-W model on the remaining samples and keep the first 3 PCs (no less than 80% of overall variations were captured).

(2) Calculate the centre and covariance matrix of each class (*e.g.* each strain for two blocks CPCA-W model or each extraction method for three blocks CPCA-W model) within each block.

(3) Project the samples being left out in step (1), *i.e.* the 'test samples', into each of the blocks *via* the block loadings and calculate the lack-of-fit errors using eqn (1) for each block and assign the testing sample to the block with the lowest error.

$$e_i = \|\boldsymbol{x} - \boldsymbol{x} \cdot \boldsymbol{P}_i \cdot \boldsymbol{P}'_i\| \tag{1}$$

where e_i is the lack-of-fit error for block i and $\boldsymbol{P}_i$ is the block loadings for block i, $i = 1$, 2 and 3 for three blocks model or $i = 1$ and 2 for two blocks model; $\boldsymbol{x}$ is the testing sample.

(4) Using eqn (2), calculate the Mahalanobis distance, d, from the block scores of the test sample to each class centre within the block that the test sample was originally assigned to. Next, predict the class membership of the test sample as the one with the shortest distance.

$$d_j = \sqrt{(\mathbf{x} - \bar{\mathbf{x}}_j) \cdot \boldsymbol{C}_j^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}}_j)'} \tag{2}$$

where $\bar{x}_j$ and C_j^{-1} are the centre and the inverse of the covariance matrix of class j, $j = 1$, 2 and 3 for the three blocks model or $j = 1$ and 2 for the two blocks model.

(5) Repeat steps (1) to (4) and remove a different set of samples each time for testing until all the samples have been left out once. Next, calculate the correctly classified rates (CCR) for block membership and class membership classification.

(6) Compare the CCRs obtained *via* cross-validation as described above against a null distribution of CCRs obtained from a permutation test. The permutation test was performed by randomly shuffling the order of data, then rearranging the data into 2 or 3 arbitrary blocks with the same number of samples for each block as was applied to the original model. The operations described from step (1) to (5) were applied to the permuted data, the CCRs of block and class membership predictions were then recorded. This test was repeated 1,000 times, each time with different randomly shuffled data. The CCRs obtained from these tests formed the null distributions and were used to assess the significance level of the CCRs obtained from the original data. If the block and/or class structure is authentic, it can be expected that the CCRs from the original data will be much better than the majority of those from the null distribution.

Another potential caveat is that in other applications, like processing monitoring where CPCA-W has been employed, the arrangement of the samples are normally in the way that each row are the same sample analysed by different analytical platforms (blocks). This is not the case in this study when it comes to dividing the data into 3 blocks. When the data has been divided into 3 blocks, each row contains 3 different biological samples coming from 3 different bacterial strains, although subjected to different extraction processes. However, if the common trend (*e.g.*, in this case, the separation between classes) between different blocks is genuine, it is not important to consider whether the samples within each row come from the same biological origin or not. Even if they do in fact originate from different sources, the superscores will still be a linear combination of different samples coming from that same class. If the separation between classes is genuine, the separation should still be revealed no matter which subset of samples from within the class have been chosen. To prove this, we performed another permutation test as described below:

(1) A *within-class permuted* data set was generated by shuffling the order of samples within each block independently, the shuffling was restricted within each class only, *i.e.* after the shuffling, the class structure of each block remained the same. In our case, the first seven samples were subjected to vortex/freeze-thaw extraction whilst the next seven were subjected to AFA extraction.

(2) Another *globally permuted* data set was generated, also by shuffling the order of samples within each block, except that this time the shuffling was performed on the full block, regardless of the class membership of the samples. Hence the class structure differs from that previously applied. Nevertheless we still assume the first seven samples are from one "artificial class" and the next seven are from a different one.

(3) CPCA-W models of these two permuted data sets respectively were built and the first 3 PCs for each model were retained. The separation between two classes was then measured by using two-samples Hotelling's t^2 statistics[27] which essentially is the Mahalanobis distance (see eqn (2)) between two class centroids. The larger the statistic, the better the two classes are separated.

(4) Repeat step (1) to (3) 100 times and collect the Hotelling's t^2 statistics obtained from the two permuted data sets and compare the distributions of the two sets of statistics.

If CPCA-W is not sensitive to the order of samples providing that they belong to the same class, the Hotelling t^2 statistics obtained from the within-class permuted data set should be significantly higher than those coming from the globally permuted data set. In addition, we also want to measure the stability of the block scores upon permutations of samples since it will influence the interpretability of the model. Hence for each CPCA-W model generated for the within class permuted data,

we performed a Procrustes transformation[28] between the block scores of each block from the permuted data set and those from the corresponding block of the original data set without the permutation. The standardised Procrustes distance was then calculated to measure the dissimilarity between these two sets of block scores. The standardised Procrustes distance varies from 0 to 1, a perfect match of the two matrices will result in the value of 0 while if there is nothing in common between them the value will be 1. For each comparison, 3 distance values were obtained, one for each block (*i.e.* P, T and W) respectively. Low Procrustes distances indicate that the relative position of the samples within the block scores are relatively unchanged whilst large Procrustes distances indicate that the pattern revealed in the block scores are constantly changing upon permutation.

Results and discussion

This study compares the efficiency of AFA and manual vortex/freeze-thaw processes for the extraction of intracellular metabolites from *S. lividans*. The efficiency of extraction here refers to the number and intensity of detected metabolite peaks. A separate objective was to see if it was possible to discriminate between the three strains of *S. lividans*.

AFA has not been used for metabolite extraction before. Therefore as an exploratory technique applied to global metabolite profiling analysis the AFA technique was optimized by varying the duty cycle, voltage intensity, cycles per burst and length of sample treatment. The best extraction setting was determined by how quickly the bacterial mycelia were completely disrupted without any residual cellular suspension and with minimal changes in sample temperature. This was obtained at 20% duty cycle, 10 mV intensity, 200 cycles per burst for 2 min. On the other hand, the extent of cellular disruption by manual freeze-thaw extraction was not assessed in this study. However, several studies have already demonstrated that freeze-thaw in 100% methanol to facilitate cell permeability is the best metabolite extraction method in a range of microorganisms.[13,16,29] This is based upon qualitative observations of cells in light microscopy, the species of metabolites recovered and the reproducibility of GC-MS data. Other methods including boiling ethanol, acid or alkalinic extraction for instance, were reported to be less appropriate for global metabolite analysis since some metabolites are heat labile or they may not be stable at extreme pH.[29,30] Therefore after quenching of metabolism in methanol (which is also the preferred approach for microbial analyses[13,16,29]) we opted to compare manual vortex freeze-thaw extraction in 100% methanol with the new AFA approach for metabolite extraction.

Averages of 120 metabolite peaks were detected for each strain and using visible inspection alone there were no noticeable differences between AFA and manual vortex in terms of the total numbers of metabolite peaks detected. This suggested that there were no differences in the types of chemical species that were recovered using AFA *versus* manual extraction. The data were normalized to the peak area of the internal standard and subsequently standardised (each variable was mean centred and divided by its standard deviation) prior to multivariate analysis to ensure that any observed clustering was not biased towards peaks with high intensities. This is because metabolomics data are typically characterized by a small number of metabolites at high concentrations and a larger number present at lower concentrations.

Comparison of extraction techniques

A plot of the scores of the first two principal components (Fig. 1) obtained from classical PCA showed no obvious separation between the two extraction methods. Instead the samples appeared to cluster according to biological variations, *viz.* the three strain types. Particularly, strains expressing mTNF-α (CT or MT for AFA and manual extraction respectively) were discriminated from the non-secreting wildtype (CW and MW)
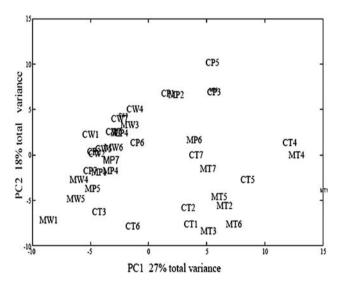

Fig. 1 Principal component plot of all samples showing separation between the mTNF-α secreting strain and the non-secreting wild type and empty pIJ486 strains. CP = AFA extraction and pIJ486; CW = AFA extraction and wild type; CT = AFA extraction and mTNF-α; MP = manual vortex and pIJ486 plasmid; MW = manual vortex and wild type; MT = manual vortex and mTNF-α. The numbers represent the biological replicates.
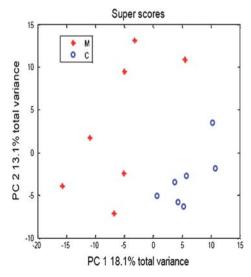

Fig. 2 Super scores plot of 3-blocks CPCA-W model. *M* = manual vortex; *C* = AFA extraction.

and empty pIJ486 strains (CP and MP). By contrast, the super scores obtained from the 3 blocks CPCA-W model (with the influence of strains suppressed) showed that there was a clear separation between the two different extraction methods, AFA and manual vortex/freeze-thaw, in the first PC as shown in Fig. 2. Furthermore, within group variations were smaller in AFA extraction than manual vortex, suggesting that AFA was a more reproducible extraction technique.

Detailed insight into the differences within each strain resulting from the different extraction methods can be seen by examining the block scores plots (Fig. 3a–3c). It can be seen that in the block scores plot of the T block, that the manual vortex was particularly irreproducible, which resulted in very large within group variation. On the other hand, AFA extraction showed much smaller within group variability across all 3 different strains suggesting that the extraction performance of AFA is more consistent than that of manual vortex/freeze-thaw. This is probably because the AFA is automatically controlled and therefore sample treatments (temperature and cycle time) were much more effectively controlled.

Similar to classic PCA, variables (metabolites) that contribute to the trend shown in the scores plot are revealed in the respective block loadings plot. However, the drawback of loadings plots is that they do not provide any statistical significance assessment. Hence we also applied the Friedman test to each variable. Since it is a large multiple comparisons problem (a total number of 120 comparisons) the threshold for significance of p values is down-adjusted by setting the false discovery rate (FDR)[31] to $q < 0.05$. Twelve variables were identified as being significantly different between the two extraction methods. These variables are highlighted by circles in the block loadings plot (Fig. 3d–3f). It can be seen that nearly all of these variables appear on the edge of the block loading plot which indicates that they contributed significantly to the separation shown in the block scores plot. Only succinic acid was consistently extracted with more efficiency by manual vortex in all 3 blocks and leucine was favoured by manual vortex at P and T blocks. The remaining significant metabolite variables were more efficiently recovered by AFA. The relative peak intensities of all these metabolites along with their tentative identifications are given in Fig. 4. Overall the AFA technique achieved up to five times greater peak intensities compared to that of manual vortex extraction.

Biological variability

In order to identify the differences between the three strains, data were rearranged into two blocks: AFA block (C block) and manual vortex block (M black). The same type of analysis was performed on this two-block data. The super scores plot is shown in Fig. 5 and the blocks scores along with the block loadings plots are shown in Fig. 6. Similar to those of the three blocks model, the separation of strains appeared to be improved in the CPCA-W super scores plot as well as in the respective block scores plot compared to the classical PCA. Clear separation between strains observed in both M and C block scores indicates that even though AFA has some advantages over manual vortex/freeze-thaw in terms of extraction efficiency, manual vortex/freeze-thaw
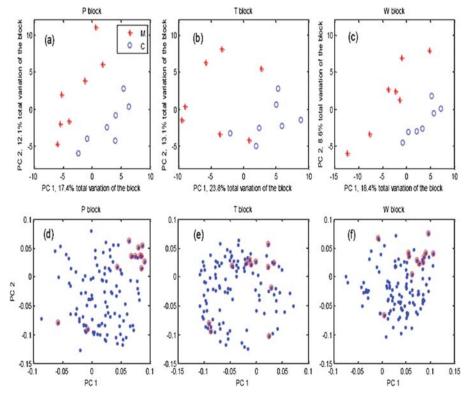

Fig. 3 Block scores and loadings plots of the three blocks CPCA-W model: T = mTNF-α secreting; P = empty pIJ486; W = wild type; M = manual vortex and C = AFA. (a) – (c) are block scores plots of P, T and W blocks respectively and (d) – (f) are their respective block loadings plots. Variables identified as statistically significant by the Friedman test are highlighted by circles.
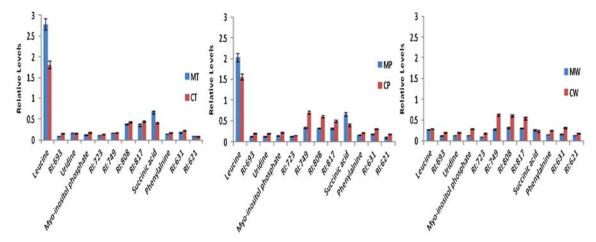
Fig. 4 Relative peak intensities of the significant metabolites between the AFA and manual vortex extraction techniques for all three biological strains. GC peaks that were not positively identified are represented by their retention time (RI followed by time in seconds). CP = AFA extraction and pIJ486; CW = AFA extraction and wild type; CT = AFA extraction and mTNF-α; MP = Manual vortex and pIJ486 plasmid; MW = manual vortex and wild type; MT = manual vortex and mTNF-α. Relative levels are square roots of actual metabolite peak intensities. Bar heights show the mean of all seven biological replicates and error bars show standard deviations.
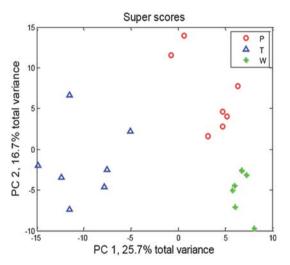

Fig. 5 Super scores plot of 2 blocks CPCA-W model. P = pIJ486, T = mTNF-α and W = wild type.

Fig. 6 Block scores plots and block loadings plot of the two block principal component analysis showing the significant metabolites between the biological strains: P = pIJ486, T = mTNF-α and W = wild type. M and C block represent manual vortex and AFA respectively.

is still efficient enough for the purpose of strain discrimination in this study.

The Friedman test identified 51 significant variables at a level of $q<0.05$ and they are also highlighted by circles in the loadings plot (Fig. 6). Among these, 12 metabolites showed significant difference between the two extraction methods and of these only 4 (of which leucine and succinic acid were positively identified) were significantly different between the three strains. This suggests that the majority of metabolites important for differentiating the different strains were largely extraction method independent. In addition, a much higher number of significant variables also suggest that the dominating factor in this study is the difference between strains rather than that between the different extraction methods. Perhaps, this is the reason that no clear separation between extraction methods can be seen in the classical PCA scores plot. The relative peak intensities of these variables (metabolites) are shown in Fig. 7. The levels of many of
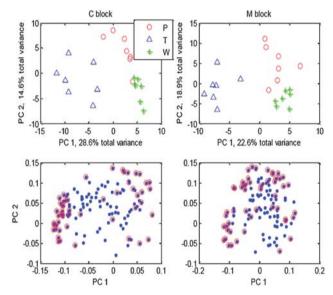
the metabolites were at least ten fold higher in strains that were either expressing mTNF-α or contained the empty pIJ486 plasmid compared to the wildtype, reflecting the influence of mTNF-α biosynthesis and the metabolic burden of harbouring the empty plasmid.

The highest levels of 17 metabolites (including alanine, leucine, glycerol, aspartic acid, gamma-aminobutyric acid, glycerol-3-phosphate, tetradecanoic acid, myo-inositol and sugars) were detected in mTNF-α expressing strains relative to the non-secreting strains of the wildtype and pIJ486 in both AFA and manual vortex extractions. Valine, isoleucine, sucrose, succinic acid, methionine and proline were higher in the mTNF-α secreting strain but only by manual vortex extraction while threonine was favoured by AFA (Fig. 7).
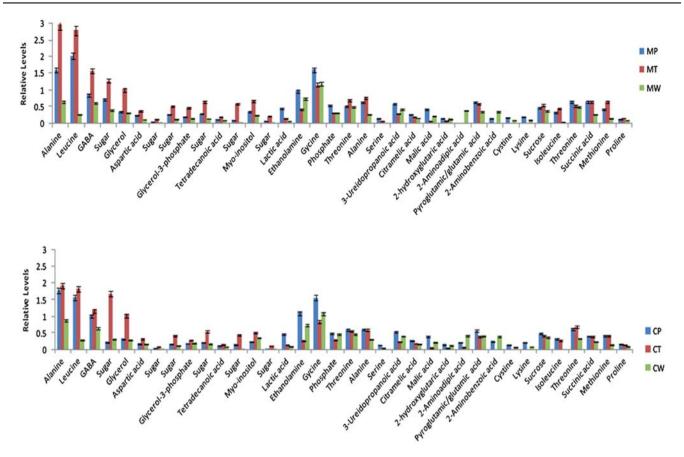
Fig. 7 Relative peak intensities of only the positively identified significant metabolites between the three different *S. lividans* strains. Fourteen metabolites show consistent higher levels in mTNF-α strains for both AFA (a) and manual vortex (b) extraction while 10 were favoured in pIJ486 strains. CP = AFA extraction and pIJ486; CW = AFA extraction and wild type; CT = AFA extraction and mTNF-α; MP = Manual vortex and pIJ486 plasmid; MW = manual vortex and wild type; MT = manual vortex and mTNF-α. Relative levels are square roots of actual metabolite peak intensities. Bar heights show the mean of all seven biological replicates and error bars show standard deviations.

Although the observed elevated levels of these metabolites in mTNF-α secretion strains have not so far been directly linked to exact pathways of protein expression and secretion in *S. lividans*, some of the metabolites have already been reported to perform essential biosynthetic and regulatory functions in both eukaryotes and bacteria. Leucine, isoleucine and valine for instance constitute the branch chain amino acids (BCAA). The metabolism of these amino acids has been show to correlate to protein biosynthesis.[32–34] Leucine up regulates protein synthesis by stimulating the initiation of mRNA translation.[35,36] Similarly, myo-inositol has been reported to be an essential precursor molecule for the biosynthesis of streptomycin in *Streptomyces griseus*.[37]

The translocation and secretion of mTNF-α in *S. lividans* is *via* the Sec pathway.[10] Sec-dependent secretion is energy driven mainly from the proton motive force that is generated by the electrochemical energy potential across the cell membrane and ATP hydrolysis by SecA.[38,39] In this study sugars, glycerol and glycerol-3-phosphate were up to 50-fold higher in mTNF-α strains (Fig. 7). These metabolites could therefore be potential energy sources and at the same time contribute to maintaining a chemical gradient across the membrane in order to drive protein secretion. In addition, the concentration of many acidic metabolites (lactic acid, 3-ureidopropanoic acid, citramalic acid, malic

acid, 2-aminoadipic acid, 2-hydroxyglutamic acid, 2-aminobenzoic acid, glutamic acid and pyruglutamic acid), that could affect the electrical potential across the membrane especially when exported into the extracellular growth media, were significantly lower in mTNF-α strains compared to the non-secreting strains (Fig. 7).

CPCA-W model validation

The results of cross-validation on the original data and randomly shuffled data are shown in Table 1. The prediction accuracy from the original data was much better than that of the randomly shuffled data for every type of prediction. Among those 1,000

Table 1 Mean prediction results of cross-validation on the original and randomly shuffled data. The number of correctly classified samples ± standard deviation is shown in the bracket

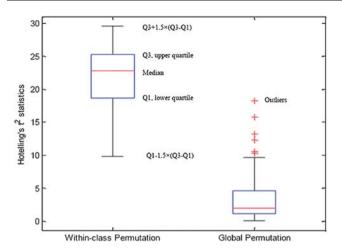| Original Data	Block	Class
3 – blocks model	83.3% (35/42)	76.2% (32/42)
2 – blocks model	76.2% (32/42)	90.5% (38/42)
Randomly Shuffled Data		
3 – blocks model	33.3% (14.0 ± 3.7)	46.4% (19.5 ± 3.5)
2 – blocks model	49.1% (20.6 ± 3.9)	31.7% (13.3 ± 3.7)

Fig. 8 The distribution of the Hotelling t^2 statistics from the second permutation test.

tests, there was not a single case that achieved such high prediction accuracy as was achieved with the original dataset. In addition, the average CCRs of shuffled data are close to the expected performance of a pure random classifier (33.3% for 3-class classification and 50% for 2-class classification), suggesting that no significant bias was introduced by rearranging the data into blocks for CPCA-W.

The results of the second permutation test are presented in Fig. 8 and 9. From Fig. 8 it is easy to see that the Hotelling's t^2 statistics obtained from the within class permutations (a median of 23.8) are much higher than those from the global permutation datasets (a median of 2.2). There are some rare cases in the global permutation data sets which obtained high Hotelling's t^2 statistics as well. By examining their class labels we found that those for the permutations were in fact largely correct (10–13 labels out of 14 were what they were originally suppose to be) hence they could also be considered as within class permutations. This proved that the CPCA-W model does not require that samples within the same row are of the same origin (*e.g.* same biological
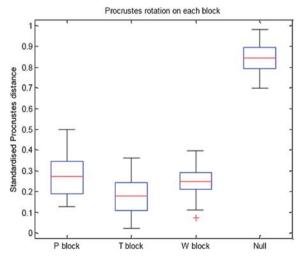
sample). Providing the samples are from within the same class and the separation between classes is genuine, the order of samples within each block is not important for the CPCA-W model to recognise such a separation. In addition, the Procrustes distances obtained by comparing the within-class permuted data set with the original data set without any permutation are also low (Fig. 9). The median distances are 0.28, 0.19 and 0.27 for the 3 blocks respectively. For comparison, the distances obtained by comparing 2 randomly generated matrices with exactly the same sized block scores as the original CPCA-W model have the median of 0.85 (the Null column in Fig. 9). This suggests that the relative position of the samples block scores is also not sensitive to the order of samples. Upon appropriate rotation, similar results can be reached regardless of the order of samples within each class.

Conclusion

The AFA metabolite extraction is a non-contact technique and compared to the manual vortex/freeze-thaw method both the temperature and cycle times are better controlled thus resulting in more repeatable metabolite extractions and hence more reproducible data. Although there were no significant differences between the numbers and species of metabolites detected by the two techniques, this study demonstrates that the recovery of intracellular metabolites was more efficient using the AFA technique for some specific metabolites. Although AFA technology offers no significant advantage for global metabolite analysis of *S. lividans* in this study, it would nevertheless be particularly useful for metabolite analyses where high throughput automation is needed. We also identified a list of specific metabolites that could have potential significance in protein synthesis and secretion in *S. lividans*, and these will be explored further.

Finally, we also demonstrated the ability of the multiblock PCA methodology to deconvolute two interacting factors and thus provide a more interpretable model that can clearly identify the contribution of each factor, even when one of them is not so obvious in the dataset. This methodology has been rigorously validated by using two types of permutation tests, thus we believe that such an approach can be very useful for complicated metabolic studies.

References

1 I. Borodina, P. Krabben and J. Nielsen, *Genome Res.*, 2005, **15**, 820–829.
2 A. L. Demain and A. Fang, *Adv Biochem Eng Biotechnol*, 2000, **69**, 1–39.

Fig. 9 The distributions of the standardized Procrustes distance taken from the second permutation test.

3 C. Binnie, J. Douglas Cossar and D. I. H. Stewart, *Trends Biotechnol.*, 1997, **15**, 315–320.

4 G. F. Payne, N. DelaCruz and S. J. Coppella, *Appl. Microbiol. Biotechnol.*, 1990, **33**, 395–400.

5 C. Pozidis, E. Lammertyn, A. S. Politou, J. Anne, A. S. Tsiftsoglou, G. Sianidis and A. Economou, *Biotechnol. Bioeng.*, 2001, **72**, 611–619.

6 G. Sianidis, C. Pozidis, F. Becker, K. Vrancken, C. Sjoeholm, S. Karamanou, M. Takamiya-Wik, L. van Mellaert, T. Schaefer and J. Anné, *J. Biotechnol.*, 2006, **121**, 498–507.

7 L. Van Mellaert, C. Dillen, P. Proost, E. Sablon, R. DeLeys, A. Van Broekhoven, H. Heremans, J. Van Damme, H. Eyssen and J. Anne, *Gene*, 1994, **150**, 153–158.

8 R. Morosoli, S. Ostiguy and C. Dupont, *Can. J. Microbiol.*, 1999, **45**, 1043–1049.

9 A. Pandey, A. Shukla and S. K. Majumdar, *African Journal of Biotechnology*, 2005, **4**, 909–910.

10 K. Schaerlaekens, E. Lammertyn, N. Geukens, S. De Keersmaeker, J. Anné and L. Van Mellaert, *J. Biotechnol.*, 2004, **112**, 279–288.

11 B. Kammerer, R. Kahlich, S. Laufer, S. M. Li, L. Heide and C. H. Gleiter, *Anal. Biochem.*, 2004, **335**, 17–29.

12 D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic and S. G. Oliver, *Nat. Rev. Microbiol.*, 2005, **3**, 557–565.

13 S. G. Villas-Boas, J. Hojer-Pedersen, M. Akesson, J. Smedsgaard and J. Nielsen, *Yeast*, 2005, **22**, 1155–1169.

14 W. B. Dunn, *Phys. Biol.*, 2008, **5**, 011001.

15 S. G. Villas-Boas, S. Mas, M. Akesson, J. Smedsgaard and J. Nielsen, *Mass Spectrom. Rev.*, 2005, **24**, 613–646.

16 C. L. Winder, W. B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G. M. Stephens and R. Goodacre, *Anal. Chem.*, 2008, **80**, 2939–2948.

17 R. Goodacre, *J. Exp. Bot.*, 2005, **56**, 245–254.

18 S. G. Villas-Bôas and P. Bruheim, *Anal. Biochem.*, 2007, **370**, 87–97.

19 M. S. Doulah, *Biotechnol. Bioeng.*, 1977, **19**, 649–660.

20 J. A. Laugharn, Jr and B. S. Garrison, EP Patent 1 125 121, 2007.

21 M. D. Wenger, P. DePhillips and D. G. Bracewell, *Biotechnol. Prog.*, 2008, **24**, 606–614.

22 G. A. Pope, D. A. MacKenzie, M. Defernez, M. A. Aroso, L. J. Fuller, F. A. Mellon, W. B. Dunn, M. Brown, R. Goodacre and D. B. Kell, *Yeast*, 2007, **24**, 667–679.

23 R. B. Cattell, *Biometrics*, 1965, **21**, 405–435.

24 A. K. Smilde, J. A. Westerhuis and S. De Jong, *J. Chemom.*, 2003, **17**, 323–337.

25 J. A. Westerhuis, T. Kourti and J. F. MacGregor, *J. Chemom.*, 1998, **12**, 301–321.

26 S. Wold, S. Hellberg, T. Lundstedt, M. Sjostrom and H. Wold, *Proceedings on PLS Model Building: Theory an Application Frankfurt*, Germany, 1987.

27 H. Hotelling, *Ann. Math. Stat.*, 1931, **2**, 360–378.

28 J. C. Gower and G. B. Dijksterhuis, *Procrustes problems*, Oxford University Press, 2004.

29 R. P. Maharjan and T. Ferenci, *Anal. Biochem.*, 2003, **313**, 145–154.

30 H. Hajjaj, P. J. Blanc, G. Goma and J. Francois, *FEMS Microbiol. Lett.*, 1998, **164**, 195–200.

31 Y. Benjamini and Y. Hochberg, *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, **57**, 289–300.

32 J. C. Anthony, T. G. Anthony, S. R. Kimball and L. S. Jefferson, *Journal of Nutrition*, 2001, **131**, 856–860.

33 E. De Rossi, R. Leva, L. Gusberti, P. L. Manachini and G. Riccardi, *Gene*, 1995, **166**, 127–132.

34 D. Petranovic and I. Mijakovic, *J. Microbiol. Methods*, 2004, **56**, 133–136.

35 T. G. Anthony, J. C. Anthony, F. Yoshizawa, S. R. Kimball and L. S. Jefferson, *Journal of Nutrition*, 2001, **131**, 1171–1176.

36 Z. Liu, L. A. Jahn, L. Wei, W. Long and E. J. Barrett, *J. Clin. Endocrinol. Metab.*, 2002, **87**, 5553–5558.

37 F. Pittner, Tovarova II, E. Y. Kornitskaya, A. S. Khokhlov and O. Hoffmann-Ostenhof, *Mol. Cell. Biochem.*, 1979, **25**, 43–46.

38 A. Economou, *Trends Microbiol.*, 1999, **7**, 315–320.

39 A. P. Pugsley, *Microbiology and Molecular Biology Reviews*, 1993, **57**, 50–108.