ORIGINAL PAPER

VOC-based metabolic profiling for food spoilage detection with the application to detecting *Salmonella typhimurium*-contaminated pork

Yun Xu · William Cheung · Catherine L. Winder · Royston Goodacre

Abstract In this study, we investigated the feasibility of using a novel volatile organic compound (VOC)-based metabolic profiling approach with a newly devised chemometrics methodology which combined rapid multivariate analysis on total ion currents with in-depth peak deconvolution on selected regions to characterise the spoilage progress of pork. We also tested if such approach possessed enough discriminatory information to differentiate natural spoiled pork from pork contaminated with *Salmonella typhimurium*, a food poisoning pathogen commonly recovered from pork products. Spoilage was monitored in this study over a 72-h period at 0-, 24-, 48- and 72-h time points after the artificial contamination with the salmonellae. At each time point, the VOCs from six individual pork chops were collected for spoiled vs. contaminated meat. Analysis of the VOCs was performed by gas chromatography/mass spectrometry (GC/MS). The data generated by GC/MS analysis were initially subjected to multivariate analysis using principal component analysis (PCA) and multi-block PCA. The loading plots were then used to identify regions in the chromatograms which appeared important to the separation shown in the PCA/multi-block PCA scores plot.

Peak deconvolution was then performed only on those regions using a modified hierarchical multivariate curve resolution procedure for curve resolution to generate a concentration profiles matrix **C** and the corresponding pure spectra matrix **S**. Following this, the pure mass spectra (**S**) of the peaks in those region were exported to NIST 02 mass library for chemical identification. A clear separation between the two types of samples was observed from the PCA models, and after deconvolution and univariate analysis using *N*-way ANOVA, a total of 16 significant metabolites were identified which showed difference between natural spoiled pork and those contaminated with *S. typhimurium*.

Keywords VOC analysis · Pork · *Salmonella typhimurium* · Peak deconvolution · Metabolic profiling

Introduction

In recent years, non-targeted metabolite profiling-based approaches have been widely applied to many types of sample matrices like plant extractions, biofluids from animals or humans, etc. to address various problems like gene function analysis, disease diagnostic and more [1–3]. However, until now, there were few reports of using metabolite profiling on volatile organic compounds (VOCs). Compared to other types of analysis, VOC analysis has its unique advantage which is that the sampling process is normally rapid, noninvasive and can be done in situ and the collected VOCs can be readily analysed using gas chromatography with little to no sample pretreatment.

The objective of this study was to investigate the possibility of using VOC-based metabolite profiling for food spoilage detection. Conventional methods of microbial

Y. Xu (✉) · W. Cheung · C. L. Winder · R. Goodacre
School of Chemistry, University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK
e-mail: yun.xu-2@manchester.ac.uk

C. L. Winder · R. Goodacre
Manchester Centre for Integrative Systems Biology,
Manchester Interdisciplinary Biocentre, University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK

spoilage detection are normally carried out by determining the total viable bacterial counts taken from the surface of the suspected meat; this typically requires an incubation period of 24 to 72 h for bacteria colony formation. Furthermore, this does not include additional time for selective enrichment of the sample and subsequent analysis; therefore, the total time required for the analysis of meat spoilage could be in the orders of weeks rather than days. Since food spoilage is the result of decomposition and the formation of metabolites caused by the growth and enzymatic activity of microorganisms, this is normally accompanied by the emission of unpleasant odour [4, 5]. Therefore, as the odour intensifies with the progression of the spoilage process, the analysis of VOCs potentially provides a rapid and quantifiable estimation of the level of microbial contamination. Previous studies have demonstrated a correlation between the intensity of the odour emitted from spoiled pork and viable counts. It has been reported that [6–8] at levels of 10^7 colony forming units per square centimetre (c.f.u. cm^{-2}), off-odours may become evident in the form of a faint 'dairy'-type aroma, and once the surface population of bacteria has reached 10^8 c.f.u. cm^{-2}, the supply of simple carbohydrates has been exhausted and recognisable off-odours develop, leading to what is known as 'sensory' spoilage. The development of off-odours is dependent upon the extent to which free amino acid utilisation has occurred, and these odours have been variously described as dairy/buttery/fatty/cheesy at 10^7 c.f.u. cm^{-2} through to a sickly sweet/fruity aroma at 10^8 c.f.u. cm^{-2} and finally putrid odour at 10^9 c.f.u. cm^{-2}. VOCs are readily analysed using headspace collection, and this process is noninvasive and so attractive for food sampling. Since VOCs from the spoilage process are mostly caused by the action of microorganisms, it would also be advantageous to establish a method using the analysis of VOCs which allows the discrimination between pathogenic and natural spoilage microorganisms.

In this study, we investigated such a possibility using pork as a test food matrix and *Salmonella typhimurium* as a relevant model pathogen. *S. typhimurium* is commonly associated with microbial spoilage of meat products including pork [9] and is pathogenic to humans; symptoms are gastrointestinal and infection is typified by bloody diarrhoea with mucus, fever, vomiting and abdominal cramps. The sampling of VOCs can be achieved using a variety of methods. The most popular sampling method for VOC analysis is head space sampling, and this includes dynamic head space sampling and static head space sampling [10]. Dynamic head space sampling generally offers better sensitivity, albeit requires more sophisticated equipment which makes it more suitable for the detection of analytes at low levels. Static head space sampling is more often used because of its better adaptivity and lower cost.

Head space sampling can be achieved by sampling the air above the samples directly by using an airtight syringe or combined with a pre-concentration device such as solid phase micro-extraction. In this study, we employed a polydimethylsilicone (PDMS) patch-based approach reported by Riazanskaia et al. [11] as the PDMS patch can be easily fit into a glass Petri dish (see below).

Fresh pork chops were purchased from a local supermarket to avoid variation caused by different hygiene status between the pork chops. Each pork chop was butterfly cut into two pieces to provide a sterile surface for the study. The two 'matched' pieces from the same pork chop were either subjected to a natural spoilage process or contaminated with *S. typhimurium*. VOCs from each piece were sampled via head space using a PDMS patch and subsequently analysed using gas chromatography/mass spectrometry (GC/MS). Four time points over a period of 72 h were monitored, and the VOC profiles were analysed using a newly devised chemometrics methodology which combined rapid multivariate analysis on total ion currents (TICs) with in-depth peak deconvolution on selected regions and univariate statistics testing to reveal the natural trend of the whole data set and also identify the potentially interesting metabolites.

Materials and methods

Culture and chemicals

A culture of *S. typhimurium* strain 4/74, whose genome sequence is known, was kindly provided by Professor Tim Brocklehurst (Institute of Food Research, Norwich, UK). This strain was sub-cultured on Lab M LAB028 blood agar plates three times after cold storage to establish phenotypic stability. A single colony was harvested using a sterile plastic loop and transferred to 250 mL of sterile nutrient broth and incubated at 37 °C for 16 h, resulting in a suspension of 5×10^7 c.f.u./mL according to plate counts. An aliquot (5 mL, $\sim 2.5 \times 10^8$ cells) of the suspension solution was transferred to a Falcon tube and centrifuged at $4,810 \times g$ for 10 min. The supernatant was removed and the pellet resuspended in 50 mL of sterile physiological saline solution (0.9% NaCl) and centrifuged at $4,810 \times g$ for 10 min. The supernatant was removed and the pellet was resuspended in 50 mL of sterile saline solution. This was repeated three times in total to remove the remaining media. The washed pellet was resuspended in 50 mL of sterile saline solution and used for the artificial contamination of the pork.

Sampling and extraction

A total of 24 boneless pork chops (weight 200–300 g) were purchased from a local supermarket. Each pork chop was

cut to a unified size of 12–14 cm^2 surface area and butterfly cut into two pieces with equivalent surface area. This provided a sterile surface for the study. A digital photo (2,048×1,536 resolution, ~3,000,000 pixels) was taken for each piece of pork alongside a ruler to enable the estimation of pixels per square centimetre and thus the surface areas in the section of pork. The butterfly piece was then placed in a large sterile glass Petri dish (Fisher Scientific, PDS-100-04L) lined with sterilised Whatman grade 40 filter paper (cat no. 90-7501-06) to which 2 mL of sterilised physiological saline solution was added to act as a moisture source to prevent the surface of the meat from drying out. The two 'matched' pieces from the same pork chop were used to produce a control (natural spoilage) and salmonellae-contaminated sample. An aliquot (1 mL) of sterile physiological saline was added to the surface of the meat and spread across the surface using a sterile plastic loop. The sample was incubated as detailed below to allow for the colonisation by the natural spoilage microorganisms. The second section of pork was contaminated with the saline suspension of *S. typhimurium* (1 mL, see above) and also spread across the surface of the meat using a sterile plastic loop. The Petri dishes were wrapped in two autoclave bags individually and sealed to minimise VOC cross-contamination and incubated at 25 °C. The headspace was sampled at a total of four time points: 0, 24, 48 and 72 h after the contamination. At each time point, six pork chops, i.e. 12 pieces, six of each type (naturally spoiled and *Salmonella*-contaminated) were sampled using PDMS patches.

PDMS patches were cut from a single silicone elastomer sheet (cat no. 751-624-16; Goodfellow Cambridge Limited) to the following dimensions, 20×15×50 mm. The patches were washed in a 5% Decon 90 solution with ultrapure water (18 MΩ) three times. The patches were then rinsed with a copious amount of ultrapure water and conditioned in a vacuum oven for 15 h at 180 °C (pressure was maintained at −1,000 mbar). Once conditioned, the patches were transferred directly into clean thermal desorption (TD) tubes (cat no. C-TBE10: Markes International Ltd.) and capped. The TD tubes were stored at room temperature within an airtight glass container filled with a layer of molecular sieves (cat no. M2635,8-12 mesh beads, Sigma). All of the tubes were used within 48 h of cleaning and conditioning, and a random selection of 20% of the TD tubes was analysed after cleaning to confirm that the batch was clean and free from contaminants before use. The patches were stuck onto the underside of the front cover of the Petri dish due to its natural adhesiveness. The samples were then resealed and placed back in the incubator for 1 h. The patches were removed and placed back into the TD tubes and analysed by GC/MS. All samples were analysed within <24 h after collection.

GC/MS analysis

A Markes International Unity 1 thermal desorption unit was connected directly through the front injector assembly of a Varian CP 3800 gas chromatograph coupled to a 2200 quadrupole ion trap mass analyzer. The Unity 1 transfer line was connected directly to the analytical column within the GC oven through the use of a Valco zero dead volume micro union (Restek, cat no. 20148). The cold trap packing material was Tenax-TA carbograph 1 TD. The transfer line to the GC was kept at 150 °C isothermally.

The analytical column was an Agilent HP-5 60 m× 0.25 mm (I.D) with a film thickness of 0.25 μm and with a stationary phase composition of 5% phenyl/95% methyl capillary column. To ensure splitless injection efficient desorption within the cold trap within the Mark 1 Unity unit, a minimum flow rate of 1.5 mL/min is required. The on-column pressure was adjusted to 85.5 kPa.

The thermal desorption profile is as follows: sample desorption 180 °C for 3 min, the cold trap kept at −10 °C. After the initial first stage sample desorption, the cold trap was then heated to 300 °C for 3 min.

The GC profile used a 70 °C initial temperature hold for 10 min; this was then ramped to 250 °C at the rate of 3 °C/min and held for 10 min. The transfer line to the MS was maintained at 270 °C isothermally. The MS was maintained at 200 °C with electron impact ionisation source at 70 eV. The mass range used was from 40 to 400 amu with a scan rate of 1.03 scan per second. Cold trap blanks and column blanks were ran after each sample to ensure that the system was free of artefacts before the next sample was analysed.

Data analysis

All the data were recorded as Varian sms files, the files were converted to netCDF files using Palisade mass transit programme (Scientific Instrument Services, Ringoes, USA), and the netCDF files were then loaded into Matlab (Mathworks, MA, USA) using the mexnc toolbox.

The procedure used for data analysis was performed as given below, and a detailed discussion about the reason of using such a procedure is explained fully in "Results and discussion". The multivariate pattern recognition was mainly conducted on the TIC. The TICs were firstly baseline-corrected using alternative least square as described by Paul et al. [12]. Alignment of the data was performed by correlation optimised warping (COW) [13]. The two parameters of COW, the number of segments and the slacking size, were optimised with a simplex optimisation procedure as described by Skov et al. [14]. For each aligned TIC, the results of the warping, i.e. the position of every segment before and after the warping, were recorded and applied to each mass channel of that chromatogram so

that not only the 1D TICs but also the 2D chromatograms were both aligned using the same alignment settings. The baseline-corrected and aligned TICs were further normalised to the surface area of each sample (see above) and $\log_{10}$-scaled. The data were $\log_{10}$-scaled because of the large difference in the peak intensities between different time points, and such scaling is necessary to prevent the multivariate analysis from being dominated by large peaks.

Principal component analysis (PCA) [15, 16] was performed on the pre-processed TICs to reveal the natural trend of the data. As previously demonstrated [17–19], the multi-block PCA models, e.g. consensus PCA (CPCA) [17, 18], are easier to interpret compared to classical PCA when there are two or more influential factors. Hence, two further CPCA models were built on two rearranged multi-block matrices to elucidate the changes caused by the two major factors of interest: (1) natural spoilage vs. salmonella-contaminated spoilage and (2) the changes over time during spoilage. A diagrammatic representation of this rearrangement of the data is shown in Fig. 1. The block loadings plots were used to identify potential interesting time windows which may contain metabolites responsible for the separation shown in the scores plot. These time windows were then selected for peak deconvolution; each time window contains 100–150 data points with 8–15 potential peaks.

On each time window, peak deconvolution was performed on each sampling time point separately due to the large difference in the peak intensities between different time points. Within the same time window, all the chromatograms at the same sampling time point were concatenated to form a matrix $\mathbf{X}$. An orthogonal projection approach [20] coupled with Dubin–Watson statistics [21, 22] were performed to provide an initial estimation of the pure spectra matrix $\mathbf{S_{est}}$ and the number of components in the $\mathbf{X}$. An initial curve resolution was performed on the $\mathbf{X}$ using alternating regression (AR) [23, 24] based on $\mathbf{S_{est}}$. Non-negativity and unimodality constraints were applied to concentration profiles, whilst non-negativity constraint was applied to spectra. The resolved concentration profiles matrix $\mathbf{C}$ and the pure spectra matrix $\mathbf{S}$ were visually inspected and obvious artefacts, e.g. irregular peak shape, highly variable retention time, etc., were removed. The remaining components in $\mathbf{C}$ and $\mathbf{S}$ were multiplied together to produce a reconstructed data matrix, $\mathbf{X_{est}}$, which was superimposed on the original $\mathbf{X}$ and subjected to visual inspection. The unfitted peaks were identified and the spectra at the apex of those unfitted peaks were selected and appended to $\mathbf{S}$ to generate the new estimated spectra matrix $\mathbf{S_{est}}$. Another AR curve resolution was then performed using the new S. This was repeated until most significant peaks were fitted and no obvious artefacts were noted in the deconvolved $\mathbf{C}$ and $\mathbf{S}$ matrices. This procedure is illustrated as a flowchart in Fig. 2. The peak area of each peak was calculated by integrating the concentration profile of each peak.
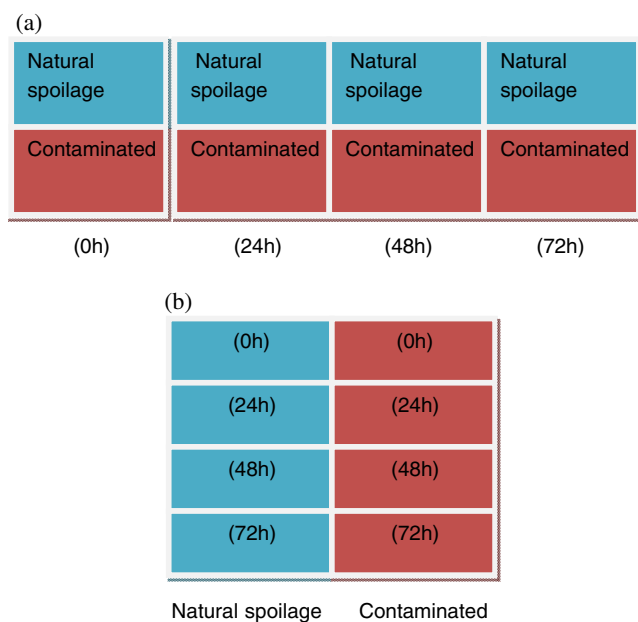

Fig. 1 Graphic illustration of the CPCA models. **a** CPCA model 1, aimed at highlighting the difference between two types of samples. **b** CPCA model 2, aimed at highlighting the changing-over-time effect whilst the spoilage progresses
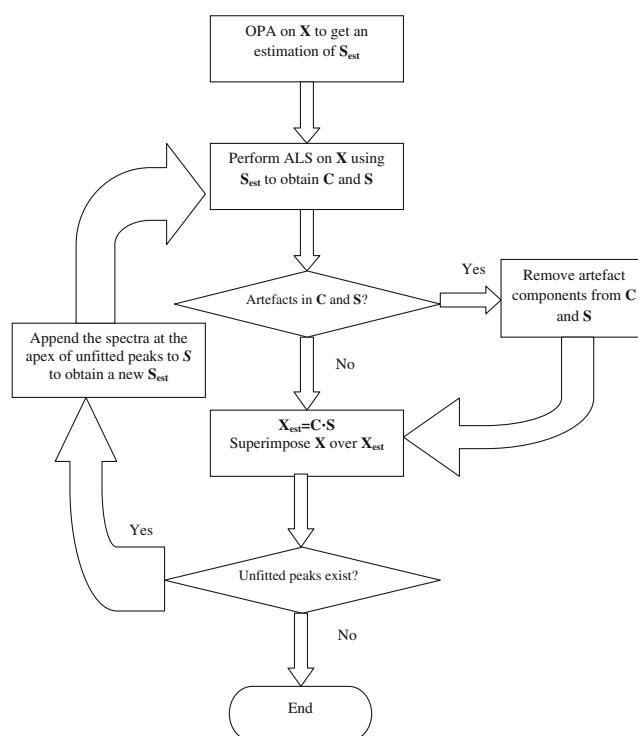

Fig. 2 Flowchart of the GC/MS peak deconvolution process

The deconvolved peaks at different sampling time points were then combined to build an overall peak table. The peaks with high similarity in the deconvolved spectrum (correlation coefficient >0.9) and also similar retention time (±1 data points, ~2 s) were considered the same peak and merged as the same variable. If in one sample no peak can be found to match the above two criteria, this peak is considered as "missing" in this sample. It is noteworthy that during the peak deconvolution, it is not possible to fit all of the peaks in all the chromatograms, and hence, it is safer to treat "not exist" as "missing" rather than an arbitrary "0" because "0" values can have very a significant impact on the statistical testing. Once the peak table was compiled, the values were log-scaled and N-way ANOVA (although this is a balanced two-way ANOVA experiment design, the two-way ANOVA cannot be applied directly due to the missing values) was applied to each variable to identify statistically important peaks which showed significant difference between different time points, different types of samples (naturally spoiled or *Salmonella*-contaminated) or both. The deconvolved mass spectra of the significant peaks were exported as NIST text files and imported into NIST 02 mass library for metabolite identification. A tentative chemical identification is made if a match factor (either match or reverse match) is >750 within the NIST library matching software. In our experience, a match factor >750 can generally be considered as a good match, and such a threshold has been used previously [19, 25]. In order to achieve definitive identification, it would be necessary to analyse the pure reference chemical and incorporate at least one additional orthogonal parameter (e.g. retention index) [26]. As most of these chemicals are not readily available, this was not conducted in this study.

Results and discussion

Two categories of methods are routinely employed to analyse the data generated by hyphened chromatography such as GC/MS, LC/MS or CE/MS. One is to use the TIC directly and treat it as a format of spectrum and apply multivariate methods such as PCA and PLS-DA to the TICs directly [27]. The advantage of this method is that less effort on the data pre-processing is needed and the risk of introducing artefacts caused by the data pre-processing into the data is relatively low; therefore, researchers can quickly gain insights into the data. However, the main drawback of using this approach is that the extra dimension brought by using hyphened chromatography, e.g. the mass spectrometric dimension, is completely ignored and the results provide little to no chemical information (i.e. the identification of the chromatographic peaks), which is against the very purpose of using MS compared to FID. By contrast, in

another category of methods focused on performing a form of peak deconvolution on the 2D chromatograms using the information provided by the second dimension (i.e. spectrum at each time point), one can deconvolve all (or at least the most significant) the chromatographic peaks and their "pure" spectra to form a peak table with relative concentrations and then apply multivariate analysis on the peak table [24, 28]. This way, in theory, provides the most comprehensive information of the data set. Various methods have been developed during the last few decades; the latest method is called hierarchical multivariate curve resolution (H-MCR), reported by Jonsson et al. [24]. The main advantage of H-MCR is that by stacking multiple chromatograms together and assuming a trilinear data structure, the curve resolution results can be significantly improved compared to previously reported methods which apply curve resolution to each chromatogram separately. Furthermore, it is even possible to resolve co-eluting metabolites which results in fully overlapped chromatographic peaks using H-MCR because in multiple samples, the relative concentrations will be different; this is impossible for methods which deconvolve chromatograms individually. However, this type of method has a much higher requirement on the quality of the data compared to the methods which use TICs directly; else, the deconvolved results will contain artefacts and can be misleading.

In this study, we found that the results obtained by applying H-MCR were far from satisfactory (results not shown). By adopting the end criterion proposed by Jonsson et al., a large number of artefacts were found from the deconvolved peaks, and these manifested themselves as irregular-shaped peaks, whilst a large amount of peaks in the chromatograms appeared left unfitted. There were many cases found wherein when the peak requirement can no longer be met by introducing more components into the deconvolution process, there were still one or more significant peaks in the chromatograms that remained unfitted, and the extra introduced components were modelled as artefacts (e.g. irregular-shaped peaks) in the end results. We believe this is because our data employed a quadrupole mass filter which has a much lower sampling rate and also a mass resolution comparing to the more expensive time-of-flight MS used by Jonsson and colleagues.

This suggested a problem with the data structure rather than the H-MCR algorithms per se, and since many researchers to do not have access to ToF-MS and use quadrupoles as an alternative, devising a new strategy for handling such data is needed. We concluded that H-MCR was still a valid approach to peak deconvolution, but that we could not follow the semi-automatic procedure described by the Jonsson et al. Extensive manual inspection had to be performed on the deconvolved results and some adjustments had to be made on the curve resolution (e.g.

removal of artefacts, insertion of better initial estimations and re-computing the curve resolution) accordingly, and this would be a very time-consuming process given the amount of data we have collected. In addition, it is likely that it is in fact not necessary to perform full deconvolution on all the peaks from all the chromatograms as not every peak in the chromatogram will have biological significance. Therefore, we decided to combine the two categories of the methods described above:

- We firstly performed the multivariate analyses of PCA and multi-block PCA on the TICs directly to reveal the trend of data set. The potentially interesting chromatographic regions were identified through the inspections of the loadings plots.
- Next these selected regions were then subjected to a modified H-MCR procedure as described in "Data analysis" for curve resolution, and the resolved peaks were assessed using statistical test. The resolved mass

(a)

(b)

Fig. 4 Block scores plots from the two CPCA models. **a** Block score plots of the first CPCA model; *N block* naturally spoiled samples, *S block Salmonella*-contaminated samples. **b** Block scores of the second CPCA model. *N* natural spoilage, *C* contaminated with *S. typhimurium*

spectra of those peaks being identified as significant were exported to NIST 02 mass spectra library (National Institute of Standards and Technology, MD, USA) for structure identification.

By adopting this approach, the excessive amount of time performing just curve resolution was avoided, and as reported below, we were still able to gain useful chemical insight into the data, even using a relatively cheap, routinely used GC/MS. More importantly, even though there were some imperfections in the deconvolved data, which were unavoidable, this did not have a negative influence on the results obtained from the multivariate analysis.
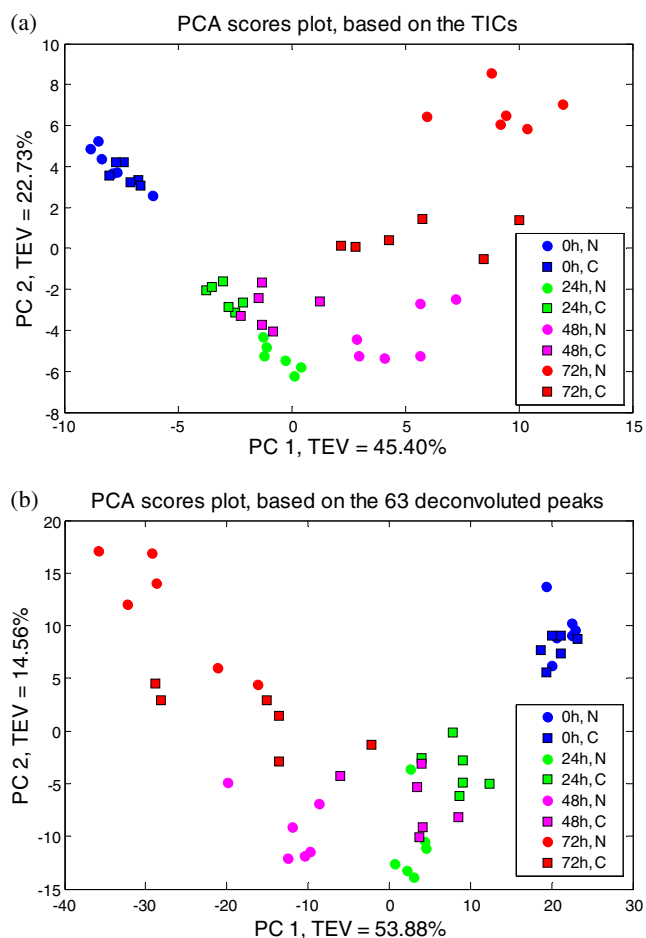

(a)

(b)

Fig. 3 PCA score plots. **a** Score plots from the PCA on TICs; **b** Score plots from the PCA on 63 deconvoluted unique peaks (VOCs). N natural spoilage, C contaminated with *S. typhimurium*, TEV = total explained variance
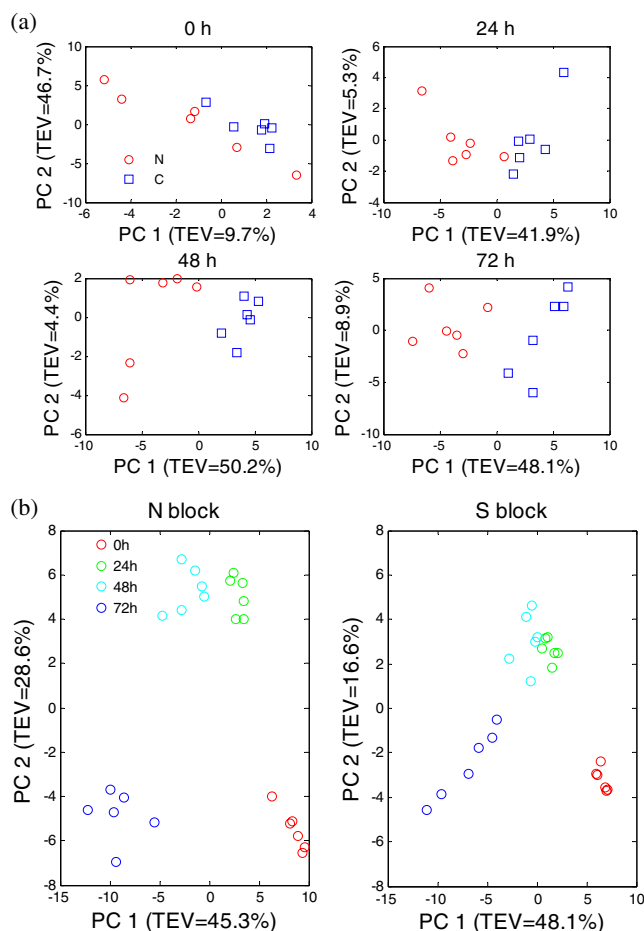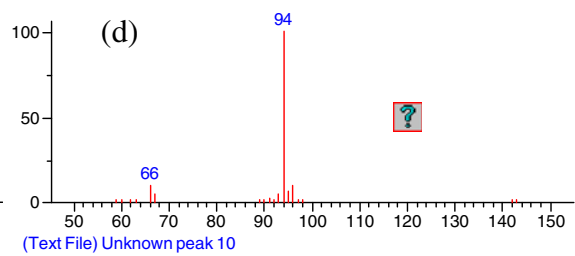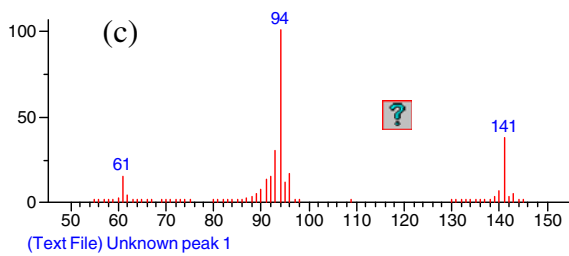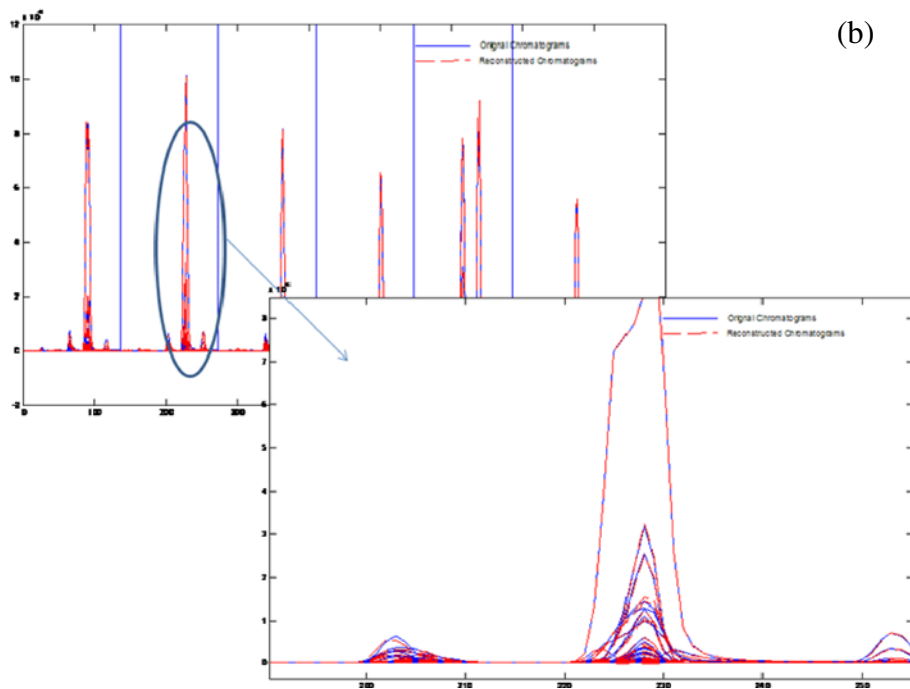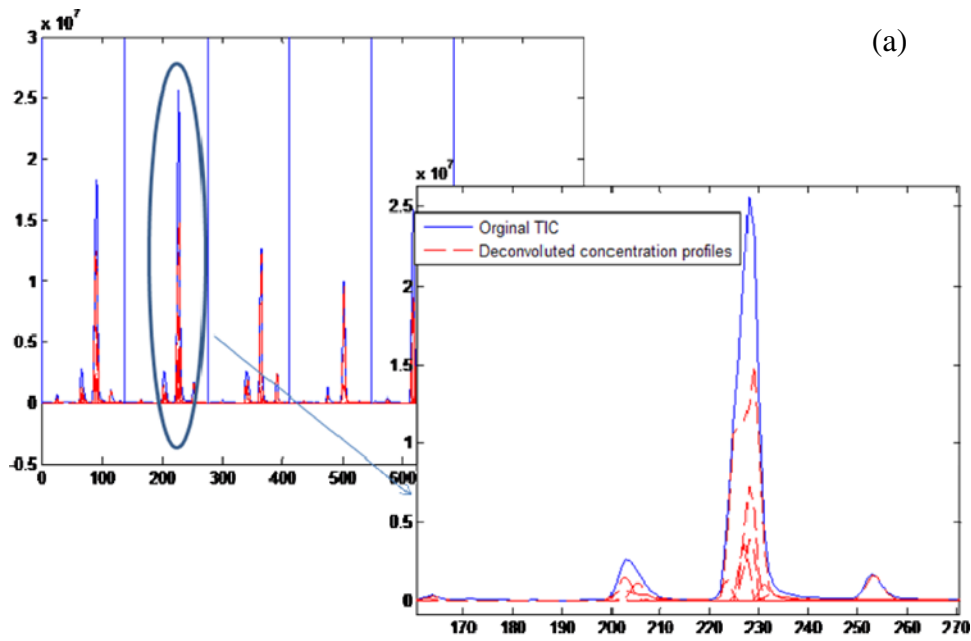
Fig. 5 Example of the results from peak deconvolution. **a** The deconvoluted concentration profiles superimposed on the TIC. **b** Reconstructed 2D chromatogram obtained by **C** × **S** superimposed on the original 2D chromatogram. **c** Mass spectrum of the peak before the deconvolution. **d** Mass spectrum of the peak after the deconvolution

As described above, the chromatograms, using TIC alone, were first analysed using classic PCA and the resulting scores plot is shown in Fig. 3a. Inspection of this plot shows that the four sampling time points were clearly separated. In addition, the separation between two types of samples (natural spoilage and salmonella-contaminated) can be observed from 24 h onwards, and seemingly, the discrimination between the sample types improves as the spoilage progresses. Such a trend is even easier to observe from the block scores in the CPCA models once one of the two major influential factors had been suppressed (in the first CPCA model, the influence of the progression of spoilage is suppressed (Fig. 4a), whilst in the second CPCA model, the influence of different type of samples is suppressed (Fig. 4b)) by rearranging the data into a multi-block matrix as illustrated in Fig. 1. In the score plots (Fig. 4a) from the first CPCA model, which highlights the difference between two types of samples, natural spoilage vs. salmonella-contaminated, the sample types were separated across the first two PCs, particularly the first PC, from 24 h which accounted for 41.9% total explained variance (TEV) and increased with the following time points to approximately 50% TEV. The time-dependent trends of the two types of samples are clearly shown in the score plots (Fig. 4b) from the second CPCA model, and it can be seen that the naturally spoiled samples appeared to have a "longer" time trajectory compared to that of salmonellae-contaminated samples; this can also be seen from the classical PCA score plots.

On examination of the loadings plots of the two CPCA models (shown in Electronic supplementary material (ESM) Figs. S-1 and S-2), it appeared that the most important features are within the first 25 min of the chromatogram, with the exception of a predominating peak at ~32 min which was more abundant in the *S. typhimurium*-contaminated samples at 72 h. Hence, we performed peak deconvolution on the first 25 min plus the region between 30–33 min of each chromatogram as described before. A typical example of the results from the peak deconvolution is given in Fig. 5. It can be seen that by superimposing the original and the reconstructed chromatograms, the most significant *m/z* were well fitted, and this gives a certain confidence on the quality of the deconvolved mass spectra. The mass spectra before and after the peak deconvolution of one significant peak are also given in this figure. No satisfactory match can be found using the mass spectrum before the deconvolution. By contrast, using the mass spectrum after the deconvolution, it is possible to identify this feature as 5-methylpyrimidine with a high matching factor of 814. A total number of 63 unique peaks (metabolites) were extracted by this deconvolution processing from the first 25 min of the chromatograms, and 48 of them can be identified through mass spectra matching in the NIST 2.0 library search engine with confidence (match factor and reverse match factor >750). The predominating peak at 32 min appeared to be a siloxane-type peak which is very unlikely to be of a biological origin and more likely to be some type of column bleeding material. Its occurrence in one type of sample at only one

Table 1 List of the identified significant metabolites

Identified chemical	Match scores	Reverse match scores	CAS no.	Change[a]
Phenylethyl alcohol	875	914	60-12-8	↑
1-Butanol, 3-methyl acetate	846	847	123-92-2	↑
Dimethyl disulfide	787	887	624-92-0	↑
2-Heptanone	854	862	110-43-0	↑
2,5-Dimethylpyrazine	891	899	123-32-0	↓
Methoxybenzene	919	932	696-62-8	↓
Phenol	925	936	108-95-2	↓
Dimethyltrisulfide	922	929	3658-80-8	↓
2-Octanone	815	861	111-13-7	↓
Butanoic acid, 3-methyl, 2-methylpropylester	887	906	589-59-3	↓
5-Methylpyrimidine	814	885	2036-41-1	↓
4-Methyl-2-oxovaleric acid	778	844	816-66-0	↓
2-Octenal	811	817	2548-87-0	↓
L-5-Propylthiomethylhydantoin	734	753	71100-43-1	↑
Propanoic acid, 2-methyl, 3-methylbutylester	766	855	2050-01-3	↓
Toluene	889	901	108-88-3	↓

[a] An upward arrow means an increase in *Salmonella*-contaminated relative to natural spoilage and vice versa

time point could be purely by chance. Hence, it was excluded from the statistical analysis. In addition, there were 52 missing values (i.e. a peak that could not be found a match in a certain chromatogram) in total, accounting for 1.72% of the total number of the elements in the final peak table, and the correlation coefficients of the matched peaks varied from 0.92 to 0.99, with the majority of the matches having a correlation coefficient higher than 0.96. This suggested that even the peak deconvolution procedure was conducted independently four times on different set of samples (samples from four different time points); the results are highly consistent and can be seen as an evidence showing that the peak deconvolution is robust and reproducible.

N-way ANOVA was applied on the 63 metabolites. It appeared that the majority of these VOCs (52 out of 63) show significant difference with false discovery rate [29] $Q \leq 0.05$ between different time points, and these metabolites generally increased as spoilage progresses. This matches our organoleptic observations that the odour from the spoiled pork intensifies as spoilage progresses. In addition, 19 metabolites were identified as being significant in that they showed differences ($Q \leq 0.05$) between the sample types. Among those 19 significant peaks, 16 of them can be putatively identified through mass spectra matching. These significant peaks with their chemical identifications are listed in Table 1, and the box whisker plots along with the corresponding deconvoluted mass spectra of one metabolite, 5-methylpyrimidine which is also the exemplary peak given in Fig. 5, is shown in Fig. 6. Such plots of other significant metabolites are listed in the supplementary information (ESM Figs. S-3 to S-20).

Once deconvolution was complete, classical PCA was performed on the peak areas of the 63 deconvolved peaks (without the siloxan peak at 32 min), and the scores plot is shown in Fig. 3b. It is clear that the result is almost identical to the one obtained from TICs (Fig. 3a), except that the separation on the last time point (72 h) was not as striking as before; this is likely to be because the artefact siloxane peak was excluded from the analysis. This proved that in using TICs alone, it was still possible to extract valuable information from the data set without the need to go through laborious peak deconvolution process. On the other hand, the predominating siloxane peak found at 32 min in this study also suggested that the results obtained from TICs analysis alone can be misleading because no chemical information was obtained from the TICs, and hence, the findings are prone to false positive errors. We believe that a good approach is to combine the two methodology using TICs to get an initial view of the data set and then perform targeted peak deconvolution and structure identification to confirm the findings from the
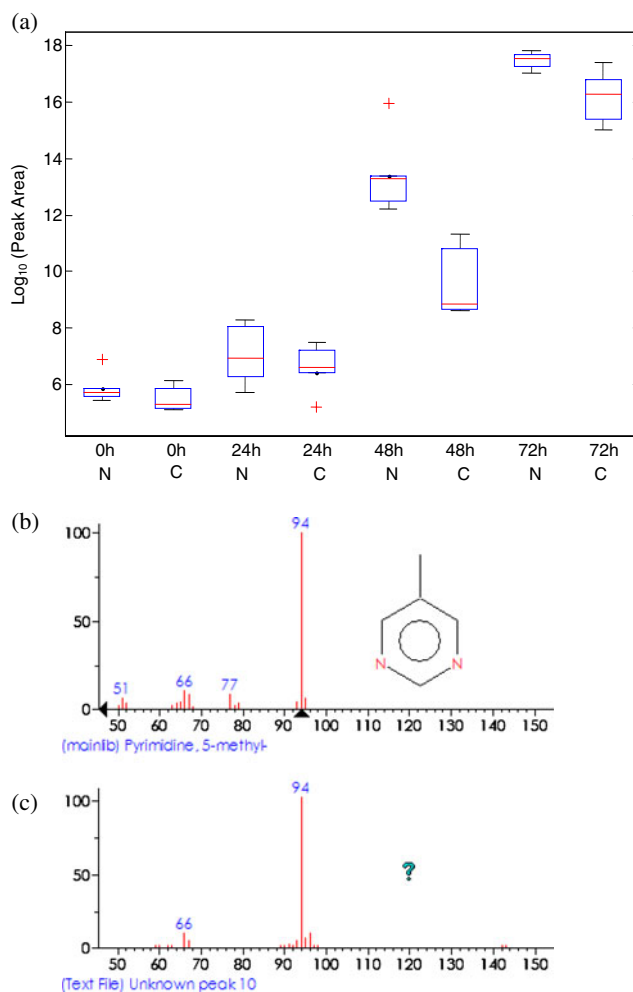

Fig. 6 **a** Box-whisker plot of 5-methylpyrimidine. **b** Standard mass spectrum. **c** Extracted mass spectrum of the target peak

analysis based on TICs. Employing such a strategy enables researchers to achieve results within a reasonable timescale whilst maximising the knowledge obtained from the data.

From Table 1, it is clear that five VOCs are significantly increased in *S. typhimurium* relative to pork that has been allowed to spoil naturally. These are phenylethyl alcohol ($C_8H_{10}O$), 1-butanol, 3-methyl acetate ($C_7H_{14}O_2$), dimethyl disulfide ($C_2H_6S_2$), 2-heptanone ($C_7H_{14}O$) and L-5-propylthio methylhydantoin ($C_7H_{12}N_2O_2$).

2-Heptanone is naturally found in certain foodstuffs including beer, bread and various cheeses [30]. As these are all microbially produced food (albeit from fungi), it is perhaps not surprising that we identified this metabolite. Although 2-heptanone has been observed in the headspace of *Enterobacteriaceae* [31], the concentration of the metabolite was reported at a higher concentration in *Escherichia coli* compared to *S. typhimurium*; however, the growth was performed on artificial culture rather than on the surface of meat in which the metabolic source of 2-heptanone has not yet been reported.

It is known that phenylethyl alcohol is a metabolite that is part of the phenylalanine pathway (KEGG: reaction R02611), so this suggests that from the meat substrate, *Salmonella* is utilising the amino acid phenylalanine preferentially for growth compared with other organisms found on the meat surface naturally. In a similar manner, dimethyl disulfide may be present due to degradation of sulphur-containing amino acids such as methionine and cysteine, as has been reported previously for *S. typhimurium* contamination in vegetables rather than meat [32].

With respect to the other volatile species, these have not been extensively reported in the literature, and so it is difficult to be very specific. However, it is known that VOCs are used by microorganisms to give the VOC producer an advantage in terms of colonisation by suppressing the growth of other bacteria [33], so these may be defence chemicals elicited by *S. typhimurium*. It is already known that phenylethyl alcohol has antimicrobial property; hence, this metabolite could also be a part of the defence mechanism of *S. typhimurium* to inhibit the growth of other bacteria.

Conclusions

In this study, we have demonstrated the feasibility of developing a VOC metabolite profiling method based on headspace analysis using PDMS capture followed by thermal desorption into GC/MS. This analytical approach generated information-rich metabolite data that we have shown were able to discriminate successfully *S. typhimurium*-contaminated and naturally spoiled pork during spoilage at 25 °C over 72 h. This suggests that VOC analysis has the potential to become a valuable tool for noninvasive, rapid detection of pathogenic microorganisms.

A set of potentially interesting metabolites have been putatively identified through this study, and this will aid the development of specific and sensitive analytical tools for the detection of pathogens in food. Further work is needed to improve the confidence of the metabolite identification; that is to say to go from putative to definitive identification [26]. Currently, the metabolite databases containing VOCs is still at a very preliminary stage and more international effort is needed to increase the metabolite coverage in these databases to the current level offered by derivative-based GC/M metabolite databases.

In addition, we have also developed a chemometric methodology that combines fast multivariate analysis on TICs with more time-consuming peak deconvolution on selected regions of the chromatograms that contain discriminatory information. Satisfactory results were obtained rapidly from the data generated by a routinely used MS instrument employing a quadrupole mass filter. This makes it particularly suitable for low-cost, rapid pilot studies. If further investigations are desired, more labour-intensive and expensive instrumentation can be used (viz. GC-ToF-MS or GCxGC-ToF-MS) followed by more in-depth deconvolution. In this investigation, we have employed this methodology to investigate the differences in the production of VOCs in the natural and artificial spoilage of pork. We have highlighted several metabolites which are characteristic of spoilage by the pathogenic bacteria and with future work may be used as indicators for noninvasive biomarkers of pathogenic spoilage of meat.

References

1. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Nat Biotechnol 18:1157–1161
2. Harrigan GG, Goodacre R (2003) Metabolic profiling: its role in biomarker discovery and gene function analysis. Kluwer, Boston
3. O'Hagan S, Dunn WB, Brown M, Knowles M, Knowles JD, Kell DB (2005) Anal Chem 77:290–303
4. Ellis DI, Goodacre R (2001) Trends Food Sci Tech 12:414–424
5. Ellis DI, Broadhurst D, Kell DB, Rowland JJ, Goodacre R (2002) Appl Environ Microbiol 68:2822–2828
6. Jackson TC, Acuff GR, Dickson JS (1997) In: Doyle MP, Beuchat LR, Montville TJ (eds) Food microbiology: fundamentals and frontiers. ASM, Washington
7. Jay JM (1996) Modern food microbiology. Chapman & Hall, London
8. Stanbridge LH, Davies AR (1998) In: Davies A, Board R (eds) The microbiology of meat and poultry. Blackie Academic & Professional, London
9. Oliveiraa CJB, Carvalhoc LFOS, Fernandesd SA, Tavechiod AT, Domingues FJ Jr (2005) Int J Food Microbiol 2:267–271
10. Mitra S (2003) Sample preparation technique in analytical chemistry. Wiley, New Jersey
11. Riazanskaia S, Blackburn G, Harker M, Taylor D, Thomas CLP (2008) Analyst 133:1020–1027
12. Eilers PHC (2004) Anal Chem 76:404–411
13. vest Nielsen NP, Carstensen JM, Smedegarrad J (1998) J Chromatogr A 805:17–35
14. Skov T, van den Berg F, Tomasi G, Bro R (2006) J Chemometr 20:484–497
15. Wold S (1987) Chemom Intell Lab Syst 2:37–52
16. Brereton RG (2003) Data analysis for the laboratory and chemical plant. Wiley, Chichester
17. Westerhuis JA, Kourti T, MacGregor JF (1998) J Chemometr 12:301–321
18. Smilde AK, Westerhuis JA, de Jong S (2003) J Chemometr 17:323–337
19. Biais B, Allwood JW, Deborde C, Xu Y, Maucourt M, Beauvoit B, Dunn WB, Jacob D, Goodacre R, Rolin D, Moing A (2009) Anal Chem 81:2884–2894
20. Cuesta Sanchez F, Toft J, van den Bogaert B, Massart DL (1996) Anal Chem 68:79–85

21. Durbin J, Watson GS (1950) Biometrika 37:409–428
22. Durbin J, Watson GS (1951) Biometrika 38:159–179
23. Tauler R (1995) Chemom Intell Lab Syst 30:133–146
24. Jonsson P, Johansson AI, Gullberg J, Trygg J, Jiye A, Grung B, Marklund S, Sjöström M, Antti H, Moritz T (2005) Anal Chem 77:5635–5642
25. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis RM, Stephens GM, Goodacre R (2008) Anal Chem 80:2939–2948
26. Sumner LW, Amberg A, Barrett D, Beger R, Beale MH, Daykin C, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, Hardy N, Higashi R, Kopka J, Lindon JC, Lane AN, Marriott P, Nicholls AW, Reily MD, Viant M (2007) Metabolomics 3:211–221
27. vest Nielsen NP, Smedsgaard J, Frisvad JC (1999) Anal Chem 71:727–735
28. Dixon SJ, Xu Y, Brereton RG, Soini HA, Novotny MV, Oberzaucher E, Grammer K, Penn DJ (2007) Chemom Intell Lab Syst 87:161–172
29. Benjamini Y, Hochberg Y (1995) J Roy Statist Soc Ser B 57:289–300
30. Maeda T, Kim JH, Ubukata Y, Morita N (2009) Eur Food Res Tech 228:1438–2377
31. Siripatrawan U (2008) Sensor Actuator B Chem 133:414–419
32. Siripatrawan U, Harte BR (2007) Anal Chim Acta 581:63–70
33. Kai M, Haustein M, Molina F, Petri A, Scholz B, Piechulla B (2009) Appl Microbiol Biotechnol 81:1001–1012