Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Phytochemistry 62 (2003) 859–863
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry

Royston Goodacre, Emma V. York, James K. Heald, Ian M. Scott*

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

Received 12 July 2002; received in revised form 2 September 2002

Abstract

Metabolic fingerprints were obtained from unfractionated *Pharbitis nil* leaf sap samples by direct infusion into an electrospray ionization mass spectrometer. Analyses took less than 30 s per sample and yielded complex mass spectra. Various chemometric methods, including discriminant function analysis and the machine-learning methods of artificial neural networks and genetic programming, could discriminate the metabolic fingerprints of plants subjected to different photoperiod treatments. This rapid automated analytical procedure could find use in a variety of phytochemical applications requiring high sample throughput.

Keywords: *Pharbitis nil*; Convolvulaceae; Japanese Morning Glory; Electrospray ionization mass spectrometry; Neural networks; Genetic programming; Metabolic fingerprinting

1. Introduction

Advances in analytical technology promise ever-expanding possibilities for metabolic characterization of: functions of new genes revealed by genome-sequencing programmes (Kell et al., 2001; Fiehn, 2002); physiological changes due to environmental stress (Fiehn et al., 2000b; Johnson et al., 2000); novel natural products (Wolfender et al., 1998); modes of action of bioactive compounds (Aranibar et al., 2001); or metabolic changes in transgenic crops (Noteborn et al., 2000). If plant metabolomics is to fulfil such aims, however, it needs to develop methodologies that encompass the diversity of phytochemicals, of which 200,000 have been estimated to occur in the plant kingdom (Fiehn, 2002), while achieving a sufficiently high analytical throughput to make the required experimentation feasible.

Hyphenated techniques, which couple GC or LC separations to MS, NMR or photodiode-array analysis, cope impressively with the complexity of minimally fractionated plant extracts (Wolfender et al., 1998; Fiehn et al., 2000a), but for high throughput the chromatographic step is a disadvantage, which is compounded when derivatization is required. In consequence, there is increasing interest in the analysis of plant extracts without chromatographic separation, using spectroscopic techniques such as NMR (Aranibar et al., 2001), FT-IR (Johnson et al., 2000) or pyrolysis MS (Goodacre et al., 1992). Such approaches may be termed 'metabolic fingerprinting' as they emphasize the pattern-recognition of metabolic phenotypes, rather than the cataloguing of specific compounds. Metabolic fingerprinting generally requires chemometric interpretation of the complexity resulting from simultaneous acquisition of analytical data on hundreds of metabolites (Goodacre et al., 1992; Beavis et al., 2000; Johnson et al., 2000; Aranibar et al., 2001).

Electrospray ionization (ESI)-MS, a soft-ionization technique that generates 'molecular' ions (Cole, 1997), has relatively unexplored potential for metabolic fingerprinting in plants. Schröder (1996) suggested that the 'molecular' ions in a complex sample may be sufficiently distinguished by their m/z values alone for the conventional LC column to be omitted and the unfractionated sample to be introduced directly into the ESI-MS, using flow-injection (Vaidyanathan et al., 2002) or direct-infusion (Zahn et al., 2001). ESI-MS has found recent application for the rapid characterization of micro-

* Corresponding author. Tel.: +44-1970-622341; fax: +44-1970-622350.
E-mail address: ias@aber.ac.uk (I. M. Scott).

0031-9422/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0031-9422(02)00718-5

organisms (Goodacre et al., 1999; Vaidyanathan et al., 2001), for rapid estimation of secondary metabolite expression in actinomycetes (Zahn et al., 2001), and for semi-quantitative determination of specific plant metabolites (Favretto et al., 2001).

The present study evaluates ESI-MS, in combination with a variety of post-analysis chemometric methods, as a possible new tool for rapid metabolic fingerprinting of complex phytochemical samples. *Pharbitis nil* was chosen as an experimental model for which advances in plant metabolomics might contribute to progress on a long-standing issue, the physiological nature of floral induction (Takeba and Takimoto, 1966; Durdan et al., 2000). We show that it is possible to discriminate metabolic fingerprints rapidly acquired by ESI-MS for *P. nil* leaves in different physiological states.

2. Results and discussion

The mean number of floral buds produced per *P. nil* cv Violet plant by 3 months after a single short day (SD) was 7.8±S.E. 2.1 ($n=15$), compared to only 0.4±0.24 on controls kept in long days (LDs). This confirmed that a single short day (SD) is sufficient to condition the transition from vegetative growth to flowering in this photoperiodically sensitive plant (Takeba and Takimoto, 1966; Durdan et al., 2000). Leaf sap, which is potentially appropriate material for flowering studies in view of the classical evidence for a mobile flowering signal (Takeba and Takimoto, 1966), was expressed at intervals after the SD treatment and compared with LD controls.

Unfractionated sap samples from 5 plants were each analyzed ×3 by ESI-MS. Each ESI-MS analysis took a maximum of only 30 s and yielded complex spectra (Fig. 1) with peaks at m/z values consistent with $[M–H]^-$ ions of organic acids and sugars known to occur in leaf apoplast, such as malate (m/z 133), hexose (m/z 179), citrate (m/z 191), sucrose (m/z 341), shikimate (m/z 173), 2-oxoglutarate (m/z 145) and fumarate (m/z 115) (Lopez-Millan et al., 2001). Spectra from SD- or LD-treated plants were not amenable to discrimination by visual comparison (Fig. 1).

Two basic approaches were used to test if spectra from the different groups of plant saps could be discriminated by chemometrics. *Unsupervised* chemometric methods seek to separate the data into clusters without reference to known classes (Everitt, 1993), and have been used to define metabolic phenotypes in GC–MS profiles of *Arabidopsis thaliana* extracts (Fiehn et al., 2000a). In contrast, *supervised learning* methods develop models by refining the accuracy of predictions for a set of examples with a known class structure, and can prove more successful than cluster analyses in discriminating plant materials (Goodacre et al., 1992; Johnson et al., 2000).

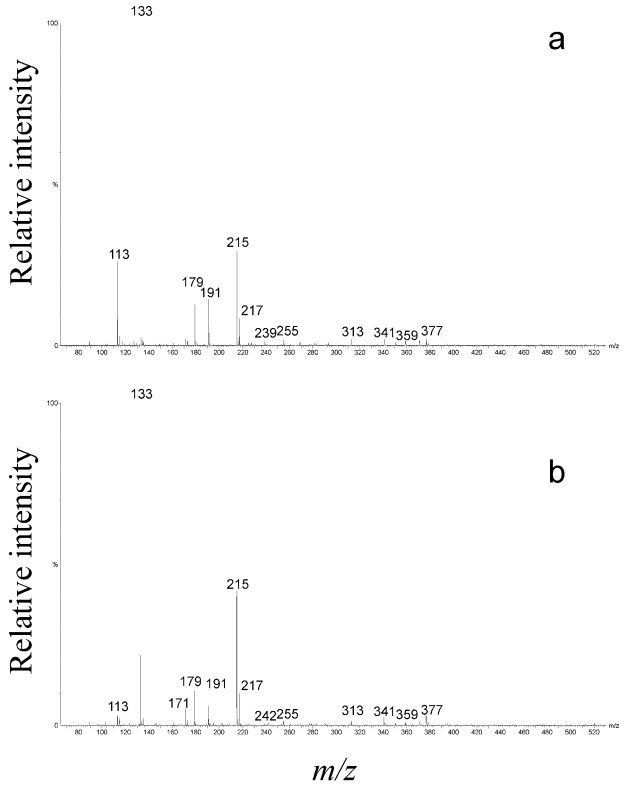

Fig. 1. Negative ion mass spectra obtained by ESI-MS analysis of unfractionated leaf sap from (a) LD control plant, and (b) SD-treated plant, each sampled 1 week after the SD.

Principal components analysis (PCA) was initially used to reduce the dimensionality of the ESI-MS data. However, no clustering according to the 6 treatments (i.e. plants sampled 24 h, 48 h or 1 week after SD treatment, with respective LD controls) was observed in PC scores plots (data not shown). Discriminant function analysis (DFA) was therefore performed on the PCs with the a priori knowledge of which spectra were from each treatment. In an iterative optimization process, it proved necessary to use the first 20 PCs as inputs into the DFA algorithm for the best class separation in DFA space. PC-DFA was performed on 4 of the 5 plants from each treatment and to test the validity of this cluster analysis the mass spectra from the 5th plant from each treatment were projected into this PC-DFA space as detailed elsewhere (Radovic et al., 2001). The projected mass spectra for each treatment fell close to the corresponding spectra used to construct this cluster space (Fig. 2), suggesting that the reproducibility of the ESI-MS approach is high. The clusters for each of the 6 treatments were wholly or partially separated, indicating that these plants were physiologically different.

Although the success of PC-DFA proved that unsupervised methods could discriminate the treatments, the failure of PCA to yield clusters prevented access to the discriminatory m/z values via loadings plots, as used for

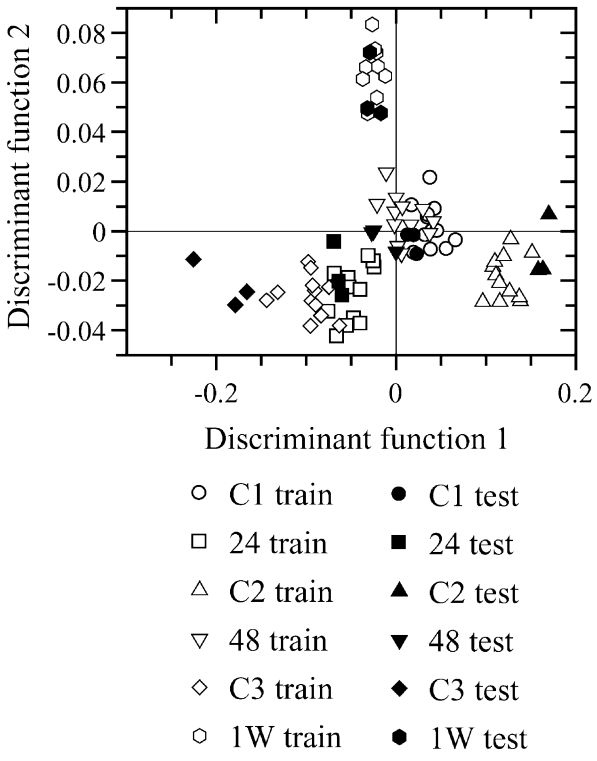

Fig. 2. PC-DFA biplot based on ESI-MS data showing the relationships between sap samples. PC-DFA was performed on triplicate mass spectra from 4 of the 5 plants from each treatment (open symbols) and test sets of the triplicate mass spectra from the 5th plant from each treatment (closed symbols) were projected into this PC-DFA space. Key: 24, 48, 1w = induced plants 24 h, 48 h, 1 week after SD; C1, C2, C3 = LD-control plants sampled at 24 h, 48 h, 1 week.

example by Fiehn et al. (2000a). Supervised learning methods were therefore also tested.

Supervised learning routines were first employed to ascertain if the mass spectra were sufficiently reproducible to be used to recognize each treatment. Arguably the most significant of the supervised methods to date have been artificial neural networks (ANNs), which have been used to authenticate virgin olive oils from adulterated oils from their pyrolysis mass spectra (Goodacre et al., 1992), and to discriminate FT-IR data on tomato fruit grown under saline or non-saline conditions (Johnson et al., 2000) or NMR data on maize treated with different herbicides (Aranibar et al., 2001). The ESI-MS data for 4 of the 5 replicate plants were used as training sets for ANNs, while the remaining replicates were used as the unseen test set. ANNs were trained over 10^3 calculations with the knowledge of which treatment had been performed on the plants. As 6 different treatments were to be assessed, the output was binary encoded in 6 nodes (see Table 1 for details). To avoid possible over-fitting of the data and to develop a parsimonious model (Seasholtz and Kowalski, 1993), rather than use each entire mass spectrum (comprising 933 m/z values) the first 100 most characteristic m/z values were used. For this purpose, m/z values were

Table 1
Identity of sap samples in the test set as judged by ANN analysis of their FI-ESI-MS data (identified samples have the largest values)

	Identity of samples judged by trained ANNs					
Correct identity:	C1	24	C2	48	C3	1w
C1	**1.0**	0.0	0.0	0.1	−0.1	−0.1
24	0.1	**0.7**	−0.1	0.2	0.2	−0.1
C2	0.1	−0.1	**1.1**	0.0	0.0	0.0
48	−0.1	0.0	0.1	**0.9**	0.1	−0.1
C3	0.0	0.1	−0.1	−0.1	**1.1**	−0.1
1w	0.1	−0.1	0.0	0.0	0.0	**0.9**

Training sets consisted of sap samples from plants ($n = 4$) subjected to same treatments as test set. Key: 24, 48, 1w = sap from induced plants 24 h, 48 h, 1 week after SD; C1, C2, C3 = sap from LD-control plants sampled at 24 h, 48 h, 1 week. Bold font shows the winning node.

ranked according to their discriminatory power by calculating the outer variance between groups to inner group variance ratio as described by Eshuis et al. (1977). Five ANNs, with 100 input nodes (one per m/z value) and 6 output nodes (one per plant treatment) connected by a single hidden layer containing 8 nodes, were trained for 10^3 calculations (epochs), which yielded a very low (ca. 0.06) average root mean squared error between the ANN output and that expected (Table 1). The training set were therefore all identified correctly and more importantly the test set were also all assessed accurately.

Although ANNs are a powerful modelling technique, the information they use to construct the calibration model is largely inaccessible. By contrast, certain evolutionary computation methods such as genetic programming (GP) yield descriptive models which can allow the interpretation of metabolite profiles in chemical terms (Goodacre et al., 2000; Johnson et al., 2000; Kell et al., 2001). Therefore GP was used to determine m/z values that discriminated the spectra for the training sets described above. Thus, a GP that was evolved to discriminate plants 1 week after SD from the rest generated rules involving m/z 520, 229 and 143. Although these m/z values represented very minor peaks in the sap spectra, when they were plotted as a pseudo-3D plot, 1-week plants could indeed be separated from the others (Fig. 3).

These results add ESI-MS to the array of techniques suitable for metabolic fingerprinting in phytochemistry. For this purpose, ESI-MS has relative merits and demerits. As a metabolic fingerprinting technique, ESI-MS was competitive in terms of speed with NMR (Aranibar et al., 2001), FT-IR (Johnson et al., 2000) or pyrolysis MS (Goodacre et al., 1992). Metabolic fingerprints produced by ESI-MS should prove amenable to identification of specific metabolites, especially in conjunction with MS/MS analysis. On the debit side, ESI-MS fingerprints will be less than global metabolic profiles since positive and negative ESI modes will emphasize

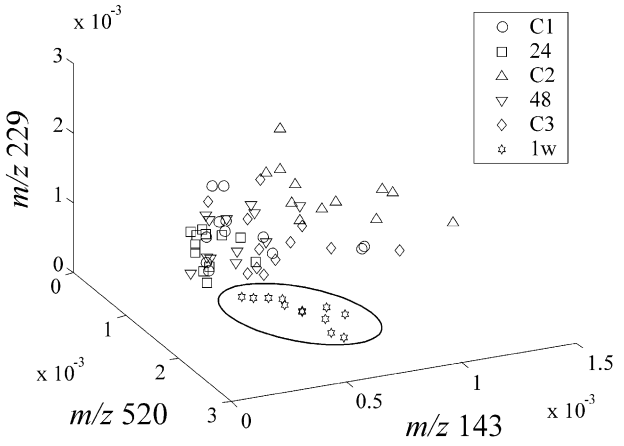

Fig. 3. Pseudo-3D plot of intensities of m/z 143, 229 and 520 ions in ESI-MS spectra of sap samples, with clustered data for plants sampled 1 week after SD (1w) highlighted. Key as Fig. 2.

different arrays of metabolites, while metabolite representation is likely to be semi-quantitative due to ion-suppression effects (Law and Temesi, 2000) that may be maximized without chromatographic separation of analytes. While ESI-MS therefore represents a potentially valuable compromise between speed and information, it needs to be used with appropriate caution.

2.1. Concluding remarks

In combination with appropriate chemometrics, ESI-MS analysis of unfractionated plant extracts offers a metabolic fingerprinting technique with considerable potential for applications where high-throughput screening is desired.

3. Experimental

3.1. Plant material

Pharbitis nil Chois cv Violet seed were obtained from Dr. R.J. Herbert, University College, Worcester, U.K. (Durdan et al., 2000). Plants were raised in John Innes No. 2 compost with weekly Chempak 3 feeds (Chempak, Hoddesdon, U.K), in a glasshouse (minimum 18 °C) under LDs of 16 h daylight supplemented when necessary with 400 W lights. After ca. 1 month, plants were either subjected to a single SD (8 h light/16 h dark) and then returned to LDs, or kept permanently in LDs as controls; thus, 6 treatments were made in all.

3.2. Mass spectral analyses of leaf sap

Plants ($n = 5$) were sampled 24 h, 48 h or 1 week after the end of the SD, with controls kept concurrently in LDs. Sap was expelled from the cut petioles of detached leaves placed in a Model 600 pressure chamber (PMS Instrument Co., Corvallis, OR), and 10 μl were collected per plant. Sap samples were diluted to 1 ml in 50% *iso*-PrOH-H_2O, loaded into a gastight syringe and infused for up to 30 s, via a Harvard Apparatus Pump 11 at a flow rate of 5 μl min^{-1}, into an LCT ESI-MS (Micromass, Manchester, UK) operating in negative ion mode at an optimized sample cone voltage of 25 V. Capillary voltage was 2500 V, extraction cone voltage 11 V, source temp. 80 °C, desolvation temp. 50 °C, desolvation and nebulizer gas flow rates were 550 and 95 l h^{-1} respectively.

3.3. Chemometric analyses

Mass spectral data in the range m/z 65-1000 (normalized to per cent total ion count at each nominal m/z) were exported from the MassLynx 3.2 software used to control the ESI-MS into the mathematical package Matlab 5.0 (The MathWorks, Natick, MA) running under Microsoft Windows NT. Matlab was employed to perform PCA (Jolliffe, 1986) and DFA (Manly, 1994; Goodacre et al., 1998). Back-propagation ANN analyses (Bishop, 1995) were carried out with NeuFrame 3.0 (Neusciences, Southampton, UK) as detailed by Goodacre et al. (1998). Gmax-bio (Aber Genomic Computing, Aberystwyth, UK), running under Windows NT, was used to perform GP analyses as detailed by (Ellis et al., 2002).

Acknowledgements

We are grateful to Dr. M. Holland for technical assistance, Dr. R.J. Herbert for the plant material, and to the BBSRC Engineering and Biological Systems and Plant and Microbial Sciences Committees for financial support.

References

Aranibar, N., Singh, B.K., Stockton, G.W., Ott, K.-H., 2001. Automated mode-of-action detection by metabolic profiling. Biochem. Biophys. Res. Comm. 286, 150–155.

Beavis, R.C., Colby, S.M., Goodacre, R., Harrington, P.B., Reilly, J.P., Sokolow, S., Wilkerson, C.W., 2000. Artificial intelligence and expert systems in mass spectrometry. In: Meyers, R.A. (Ed.), Encyclopedia of Analytical Chemistry. John Wiley, Chichester, UK, pp. 11558–11597.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Cole, R.B. (Ed.), 1997. Electrospray Ionization Mass Spectrometry. Wiley, New York.

Durdan, S.F., Herbert, R.J., Rogers, H.J., Francis, D., 2000. The determination time of the carpel whorl is differentially sensitive to carbohydrate supply in *Pharbitis nil*. Plant Physiol. 123, 189–200.

Ellis, D.I., Broadhurst, D., Kell, D.B., Rowland, J.J., Goodacre, R., 2002. Rapid and quantitative detection of the microbial spoilage of meat using FT-IR spectroscopy and machine learning. Appl. Env. Microbiol. 68, 2822–2828.

Eshuis, W., Kistemaker, P.G., Meuzelaar, H.L.C., 1977. Some numerical aspects of reproducibility and specificity. In: Jones, C.E.R., Cramers, C.A. (Eds.), Analytical Pyrolysis. Elsevier, Amsterdam, pp. 151–156.

Everitt, B.S., 1993. Cluster Analysis. Edward Arnold, London.

Favretto, D., Piovan, A., Filippini, R., Caniato, R., 2001. Monitoring the production yields of vincristine and vinblastine in *Catharanthus roseus* from somatic embryogenesis. Semiquantitative determination by flow-injection electrospray ionization mass spectrometry. Rapid Comm. Mass Spec. 15, 364–369.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Tretheway, R.N., Willmitzer, L., 2000a. Metabolite profiling for plant functional genomics. Nature Biotech. 18, 1157–1161.

Fiehn, O., Kopka, J., Tretheway, R.N., Willmitzer, L., 2000b. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal. Chem. 72, 3573–3580.

Fiehn, O., 2002. Metabolomics—the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

Goodacre, R., Kell, D.B., Bianchi, G., 1992. Neural networks and olive oil. Nature 359, 594.

Goodacre, R., Timmins, E:M., Burton, R., Kaderbhai, N., Woodward, A.M., Kell, D.B., Rooney, P.J., 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiol. 144, 1157–1170.

Goodacre, R., Heald, J.K., Kell, D.B., 1999. Characterization of intact microorganisms using electrospray ionisation mass spectrometry. FEMS Microbiol. Lett. 176, 17–24.

Goodacre, R., Shann, B., Gilbert, R.J., Timmins, E:M., McGovern, A.C., Alsberg, B.K., Kell, D.B., Logan, N.A., 2000. Detection of the dipicolinic acid biomarker in *Bacillus* spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. Anal. Chem. 72, 119–127.

Johnson, H.E., Gilbert, R.J., Winson, M.K., Goodacre, R., Smith, A.R., Rowland, J.J., Hall, M.A., Kell, D.B., 2000. Explanatory analysis of the metabolome using genetic programming of simple, interpretable rules. Genet. Prog. Evol. Mach. 1, 243–258.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer-Verlag, New York.

Kell, D.B., Darby, R.M., Draper, J., 2001. Genomic computing. Explanatory analysis of plant expression profiling data using machine learning. Plant Physiol. 126, 943–951.

Law, B., Temesi, D., 2000. Factors to consider in the development of generic bioanalytical high-performance liquid chromatographic-mass spectrometric methods to support drug discovery. J. Chromat. B 748, 21–30.

Lopez-Millan, A.F., Morales, F., Abadia, A., Abadia, J., 2001. Iron deficiency-associated changes in the composition of the leaf apoplastic fluid from field-grown pear (*Pyrus communis*) trees. J. Exp. Bot. 52, 1489–1498.

Manly, B.F.J., 1994. Multivariate Statistical Methods: A Primer. Chapman and Hall, London.

Noteborn, H.P.J.M., Lommen, A., van der Jagt, R.C., Weseman, J.M., 2000. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. J. Biotech. 77, 103–114.

Radovic, B.S., Goodacre, R., Anklam, E., 2000. Contribution of pyrolysis mass spectrometry (Py-MS) to authenticity testing of honey. J. Anal. Appl. Pyrol. 60, 79–87.

Schröder, H.F., 1996. Polar organic pollutants from textile industries in the wastewater treatment process—biochemical and physico-chemical elimination and degradation monitoring by LC-MS, FIA-MS and MS-MS. Trends Anal. Chem. 15, 349–362.

Seasholtz, M.B., Kowalski, B., 1993. The parsimony principle applied to multivariate calibration. Anal. Chim. Acta 277, 165–177.

Takeba, G., Takimoto, A., 1966. Translocation of the floral stimulus in *Pharbitis nil*. Bot. Mag. Tokyo 79, 811–814.

Vaidyanathan, S., Rowland, J.J., Kell, D.B., Goodacre, R., 2001. Discrimination of aerobic endospore-forming bacteria via electrospray-ionization mass spectrometry of whole cell suspensions. Anal. Chem. 73, 4134–4144.

Vaidyanathan, S., Kell, D.B., Goodacre, R., 2002. Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. J. Am. Soc. Mass Spec. 13, 118–128.

Wolfender, J.-L., Rodriguez, S., Hostettmann, K., 1998. Liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy for the screening of plant constituents. J. Chromat. A 794, 299–316.

Zahn, J.A., Higgs, R.E., Hilton, M.D., 2001. Use of direct-infusion electrospray mass spectrometry to guide empirical development of improved conditions for expression of secondary metabolites from actinomycetes. Appl. Environ. Microbiol. 67, 377–386.