Metabolomics: what is new?

Verpoorte, R.; Choi, Y.H.; Kim, H.K.

Plenary Lecture

Flavour and Fragrance Journal

Received: 14 January 2010; Accepted: 15 January 2010; Published online in Wiley InterScience: 18 February 2010

(www.interscience.wiley.com) DOI 10.1002/ffj.1982

Metabolomics: what's new?†

R. Verpoorte,* Y. H. Choi and H. K. Kim

Keywords: metabolomics; reproducibility and chemical structures; high-throughput analysis; identification and quantitation; public metabolomic database

In 1966, Baerheim Svendsen and Karlsen reported the analysis of three essential oils by means of gas chromatography (GC);[1] in 2009, Clara Grosso reported the metabolomics of *Tornabenea* by means of gas chromatography–mass spectroscopy (GC–MS).[2] What is the difference? Metabolomics is the latest of the '-omics'. It is a major tool in functional genomics, quality control of botanicals, studies on the activity of medicinal plants, and systems-biology types of studies of microorganisms, cells, animals, plants, etc.[3,4] Metabolomics has the ambitious objective to identify and quantify all metabolites in an organism – ambitious because the number of compounds in an organism is probably in the same order as the number of genes, e.g. in a single plant some 30 000 compounds might be present, with very different structures, polarity, at very different levels and many of these compounds might be instable. That means a great challenge for any analytical chemical tool to deal with such a complexity, but in fact, essential oil analysis faced the same challenge more than 40 years ago. Since then, however, the technology of separation methods, MS and nuclear magnetic resonance spectroscopy (NMR) have made great advances, but probably most importantly is the information technology that now allows us to work with huge datasets and to use chemometric methods to extract information out of all the data.

Listing the requirements of any method for meeting the above-mentioned analytical challenges gives the following required features:

- Highly reproducible.
- Detection of broad variety of chemical structures.
- High resolution to deal with large number of compounds.
- High-throughput analysis.
- Identification of compounds.
- Ease of quantitative analysis.
- Suitable for a public metabolomics database.

At present, the major technologies applied in metabolomics are chromatography [liquid (LC) and gas (GC)] coupled with mass spectroscopy, mass spectrometry (MSn) and nuclear magnetic resonance spectroscopy (NMR). Each of these methods has its advantages and disadvantages considering the requirements mentioned, and numerous reviews deal with this.[5,6] Here we briefly summarize the most important points.

Reproducibility

Reproducibility is a must for long-lasting public databases. For this aspect, NMR scores best. We all know how difficult it is to standardize chromatographic separations. Analyses performed on Friday afternoon and Monday morning showing different chromatograms is a well-known problem. The dream of all natural product people in the past 50 years has been to develop reproducible analytical methods for compounds in plants. However, we know from the literature that instead of a dream we have a nightmare. Some 25 years ago we[7,8] reviewed the analysis of the tropane alkaloids hyoscyamine and scopolamine – 69 references on HPLC and 32 on GC analysis. But every year many new studies are again published on this separation, always with the same aim. Apparently, standardization has failed. A major reason is the improving quality of equipment and chromatographic materials. But even a seemingly trivial aspect, the method of sample preparation, has not yet been standardized. Harvesting, storage, grinding, extraction – steps that carry the risk of artifact formation or loss of compounds – need to be standardized to achieve the goal of high reproducibility.[9]

We should learn from molecular biologists how they have been able to standardize their methods. They use standardized commercial kits, which can be used by anyone but will give highly reproducible results. These methods are the basis of the public databases for genes and proteins, thus enabling data-mining in all data produced in the past years. In natural products analysis, we at best have our own database with chromatograms of essential oils or alkaloids, or whatever restricted class of compounds, that can be used as an in-house reference for our own work, but one public database with all data on the analysis of any kind of natural products is still missing. Only some databases restricted to certain classes of natural products exist.

Variety of Chemical Structures

Concerning the problem of broad chemical characteristics, each method has different constraints. In GC it is obviously volatility

* Correspondence to: R. Verpoorte, Department of Pharmacognosy, Section of Metabolomics, Institute of Biology Leiden, PO Box 9502, 2300RA Leiden, The Netherlands. E-mail: VERPOORT@LACDR.LeidenUniv.NL

Department of Pharmacognosy, Section of Metabolomics, Institute of Biology Leiden, The Netherlands

† This paper was published in *Flavour and Fragrance Journal* as part of the special issue based on the lectures given at the 40th International Symposium on Essential Oils, held at Savigliano, Italy, 6–9 September 2009, organized by Carlo Bicchi and Patrizia Rubiolo

and stability that hamper the analysis of most plant secondary metabolites by GC. As long as analytes are volatile enough as such or as TMS derivatives, GC is the preferable method for metabolomic analysis. In particular, headspace trapping techniques combined with GC–MS are quite suitable analytical approaches for metabolomic analysis of volatiles.[10] Headspace solid-phase micro-extraction (SPME) fibres are commonly used to collect and concentrate the volatiles released into the air space above the plant material. SPME applications have been found to have high sensitivity, reproducibility and robustness.[11–14]

In most GC–MS analyses, electron-impact (EI) analysis is being used as a suitable ionization method, which has been proved to provide very reproducible MS-fragmentation that can be used for the identification of metabolites with the help of a database. However, despite the high reproducibility of the fragmentation pattern, the limitation of the method is that information on the molecular weights (MWs) of some compounds is not always obtained because of over-strong fragmentation. To solve the problem of EI ionization atmospheric pressure, a chemical ionization (APCI) method has been coupled to GC to improve the detection of the MWs. For the identification of volatiles in cucumber and tomato, the APCI mode was compared to EI for GC analysis. The combination of EI and APCI by simultaneous detection made the identification of volatiles much easier.[15]

In LC and NMR, it is the polarity of the compounds that is the restriction. These methods are in fact most suitable for detecting middle polar/polar compounds. In NMR-based metabolomic analysis, mostly ^{1}H-NMR has been used for polar plant secondary metabolites. In the case of vegetable oils, ^{13}C-NMR has been used for analysis of the oil components.[16,34] In MS, the question is about whether a compound is easy to ionize; in particular, some phenolic compounds are known to be poorly ionized.

Consequently, in all cases one has to live with a compromise, a combination of methods is required to have the broadest possible coverage of the metabolome.

Number of Compounds

There is no doubt that the combination of chromatography and MS has the highest possible resolution, with the two dimensions of the two hyphenated methods. Using two-dimensional (2D) chromatographic methods in combination with MS, the resolution is undoubtedly the highest that that can be achieved. However, in the case of GC, volatility and stability interfere with this. High resolution MS–MS might in fact be preferable to the hyphenated methods, as it overcomes the reproducibility problems of chromatographic methods, as mentioned above.

High-throughput Analysis

The chromatographic methods are the most time consuming, with a run time of 30–60 min needed to deal with the broad variety of compounds. UPLC may reduce this to 10–15 min, but MS and NMR are without doubt much faster. Dedicated NMR equipment can run a ^{1}H-NMR spectrum of a crude extract in less than 1 min. But for all methods, one should also consider the time needed for extraction and sample preparation, which in the case of GC is far more complicated, as a chemical derivatization step is required, whereas in the case of NMR the extraction can be done with the deuterated NMR solvent.[24–33]

Identification of Compounds

For this aspect, one should consider two groups of compounds, known and unknown. Known compounds can in all cases be identified by their specific characteristics (retention time, molecular weight to several decimal places, specific NMR signals). However, in the case of an unknown compound, a retention time has no meaning and the molecular weight may give at best the molecular formula in terms of carbons, hydrogens, oxygens, nitrogen, etc., but for all the known natural products there is already a large number of isomers, not to mention novel structures. Only by NMR, through correlation between the signals, is information obtained about the chemical structure; by applying 2D NMR many times, structural elucidation of unknowns can even be done in crude extracts.[17] NMR is thus most suited for the direct identification of compounds in mixtures. In addition, linking NMR to HPLC can help in further identification of minor compounds.[18] It should be noted that NMR is the only non-destructive method, allowing the same sample to be analysed by other methods afterwards.

In the coming years, metabolomics will require a major input from natural products chemists to identify all unknowns observed in many metabolomic studies, e.g. in GC–MS studies of plant metabolomes, less than half of all peaks observed have been identified.

Ease of Quantitation

In the chromatography- and MS-based methods, each compound has a different detector response, which means that, for quantitation of each individual compound, a calibration curve has to be made, and with new columns these curves have to be validated again. In MS, the matrix may play an important role in the efficiency of the ionization of a compound and thus in the detector response. In fact, this means that for these methods at best a relative quantitation for an individual compound can be made, e.g. percentage-wise, the level of a compound under different experimental conditions of a plant can be given in comparison with a control. It is in this aspect that NMR is superior to all the other metabolomic tools. In NMR the signal intensity of a proton is only dependent on the molar concentration, which means that all compounds can be directly compared and, by using a single internal standard, the absolute quantities for all compounds can be calculated.

Public Metabolomic Database

From the discussion above, it becomes clear that a public database for metabolomics, which means a database in which the total datasets of all measurements are collected for future use, will be difficult to achieve, first of all because of the problems of reproducibility. Certainly alignment programmes exist for chromatography, but these are of little use if a certain type of column no longer exists. Moreover, by aligning unknown compounds, errors can not be excluded. Mass spectrometry suffers from matrix effects that may influence the quantitation. Thus, direct comparison of the MS spectra of samples of different origin, and particularly of different species, may pose problems. NMR is probably the most promising technology for a public database, as when measured in the same solvent and under controlled pH, for most compounds high reproducibility of the spectra is

achieved. The success of NMR-based metabonomics in urine analysis is probably the best evidence for NMR as a reliable and reproducible method to generate data for long-term data-mining.[4,19] One of the reasons for success is, of course, that urine samples are measured as such; there are no discussions about how to harvest, grind and extract the material. In the field of essential oil analysis, many different methods are already being used, and thinking about the analysis of the total metabolome, the number of possibilities for extraction is almost infinite. Without coming to clear decisions about the methods to be applied, it will be difficult to achieve a public database. At least we have been able, in our work, to define standard protocols for NMR-based metabolomics, which allows us to do data-mining in a database with more than 20 000 spectra of crude plant extracts, and to use our NMR database with 1D and 2D spectra of more than 500 common plant metabolites, all measured under standard conditions to enable compound identification.[20] These protocols are particularly suitable for secondary metabolites of intermediate polarity. Other protocols for NMR of primary metabolites have been reported by Kruger *et al*.[21] De Vos *et al*.[22] published a protocol for LC–MS of a wide spectrum of compounds, and Lisec *et al*.[23] reported a protocol for GC–MS, particularly of primary metabolites.

These tools can be applied in a broad variety of applications: quality control of medicinal plants (*Ephedra* spp.,[24] *Ginkgo biloba*[25]), plant resistance against pests and diseases (*Senecio* spp.,[26] *Brassica rapa*,[27] *Catharanthus roseus*,[28] tobacco[29]), identification of active compounds in medicinal plants (*Galphimia glauca*[30]), and chemotaxonomy (*Ilex* spp.,[31] cannabis,[32] *Strychnos* spp.[33]). In particular, the use of metabolomics, together with proteomics and transcriptomics, will be a major area of basic functional genomics studies. In systems biology approaches to study, e.g. plant resistance against pests and diseases, metabolomics will be crucial, as it describes the phenotype by means of chemistry. Many definitions have been given for systems biology, and we will not discuss these here but just mention the essence, which is the unbiased observation of an organism under different conditions using any available tool, e.g. visual, physiological, molecular biological, chemical and physical observations. 'Unbiased' means that there is no hypothesis to start with, just observations that, by means of chemometrics, are analysed for any correlations, groupings, biomarkers, etc., after which one may formulate a hypothesis – a model to explain the findings. Chemometrics is thus essential to be able to deal with the large number of variables. In particular, various multivariate analysis methods, such as unsupervised principal component analysis (PCA), and the supervised methods, such as partial least square discriminant analysis (PLS-DA) and orthogonal projection-to-latent structures discriminant analysis (O-PLS-DA), are important tools for finding correlations in large datasets.[19]

Where do we stand now? Do we have public databases for metabolomic data, similar to gene and protein sequence databases? Can we expect these in the near future? What are the hurdles, and can we overcome them? Many questions, few answers, which is maybe not so surprising after all. For a total analysis of all metabolites, we must conclude that we still have no suitable standard method. However, the basic idea of metabolomics, an unbiased analysis of the metabolome instead of a targeted analysis of certain compounds, has already shown to be a very powerful tool in life sciences; it is a novel gateway to new discoveries. It is also about sustainable methods and results. Observations, which are results of our experiments, are the basis of all science; making these available in an unbiased format will be of great value.

But coming back to the first question, what is the difference between metabolomics and essential oil analysis? In fact there is no real difference – one may consider essential oil analysis as a metabolomic '*avant la lettre*'. Like many things, science is cyclic; after the excitement of the successful total analysis of an essential oil by GC(–MS), it became more or less routine and thus of less interest. The application of GC–MS as a method to measure metabolites in an organism in connection with molecular biological studies brought it to back into the spotlight; the name 'metabolomics' did the rest. However, analysing the publications that now appear in large numbers with the word 'metabolomics' in the title shows that most studies are just targeted analyses of the type that has already been made for several decades. If anything is to be targeted, it is cell type, as metabolomics now uses complete plant parts in which numerous different types of cells are present, each with its own specific (micro-) metabolome, and inside the cell the different cellular compartments each have their (nano-) metabolomes. To understand the regulation and role of the production of plant secondary metabolites, such as essential oils, cell targeted analysis would be very useful. Again, essential oil research has already shown many years ago that this is feasible, e.g. for glandular hairs. To be able to better understand a living organism, we need to look at all four dimensions of life – three of space and one of time.

Thus, it seems that some interesting challenges lie ahead for the analysis of essential oils as part of the analysis of the metabolome.

References

1. A. Baerheim Svendsen, J. Karlsen. *Planta Med*. **1966**, *14*, 376.
2. C. Grosso, G. Teixeira, I. Gomes, E. S. Martins, J. G. Barroso, L. G. Pedro, A. C. Figueiredo. *Biochem. Syst. Ecol*. **2009**, *94*, 153.
3. J. K. Nicholson, J. C. Lindon, E. Holmes. *Xenobiotica* **1999**, *21*, 1181.
4. J. K. Nicholson, J. C. Lindon. *Nature* **2008**, *455*, 1054.
5. R. Verpoorte, Y. H. Choi, H. K. Kim. *Phytochem. Rev*. **2007**, *6*, 3.
6. R. Verpoorte, Y. H. Choi, R. N. Mustafa, H. K. Kim. *Phytochem. Rev*. **2008**, *7*, 525.
7. A. B. Svendsen, R. Verpoorte. In *Chromatography of Alkaloids, Part I, Thin-layer Chromatography*. Chromatography Library, Vol. 23A. Elsevier: Amsterdam, **1983**; 534.
8. R. Verpoorte, A. B. Svendsen. In *Chromatography of Alkaloids, Part II, GLC and HPLC*. Chromatography Library, Vol. 23B. Elsevier: Amsterdam, **1984**; 467.
9. F. Maltese, F. van der Kooy, R. Verpoorte. *Nat. Prod. Commun*. **2009**, *4*, 447.
10. Y. M. Tikunov, F. W. A. Verstappen, R. D. Hall. In *Metabolomics: Methods and Protocols*, W. Weckwerth (ed.). Humana: Totowa, NJ, **2007**; 39–53.
11. F. Augusto, A. L. P. Valente, E. S. Tada, S. R. Rivellino. *J. Chromatogr. A* **2000**, *873*, 117.
12. X. Yang, T. Preppard. *J. Agric. Food Chem.* **1994**, *42*, 1925.
13. A. J. Matich, D. D. Rowan, N. H. Banks. *Anal. Chem.* **1996**, *68*, 4114.
14. J. Song, B. D. Gardner, J. F. Holland, R. M. Beaudry. *J. Agric. Food Chem*. **1997**, *45*, 1801.
15. S. Surawang, N. Rattanapanone, R. Linforth, A. J. Taylor. *Nat. Sci*. **2003**, *37*, 468.
16. G. Vlahov. *Anal. Chim. Acta* **2006**, *577*, 281.
17. Y. S. Liang, H. K. Kim, A. W. M. Lefeber, C. Erkelens, Y. H. Choi, R. Verpoorte. *J. Chromatogr. A* **2006**, *1112*, 148.
18. G. Glauser, D. Guillarme, E. Grata, J. Boccard, A. Thiocone, P. A. Carrupt. *J. Chromatogr. A* **2008**, *1180*, 90.
19. J. C. Lindon, J. K. Nicholson. *Annu. Rev. Anal. Chem.* **2008**, *1*, 45.
20. H. K. Kim, Y. H. Choi, R. Verpoorte. *Nat. Protoc*. (in press).
21. N. J. Kruger, M. A. Troncoso-Ponce, R. G. Ratcliffe. *Nat. Protoc.* **2008**, *3*, 1001.

22. R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino, R. D. Hall. *Nat. Protoc.* **2007**, *2*, 778.
23. J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A. R. Fernie. *Nat. Protoc.* **2006**, *1*, 387.
24. H. K. Kim, Y. H. Choi, C. Erkelens, A. W. M. Lefeber, R. Verpoorte. *Chem. Pharm. Bull*. **2005**, *53*, 105.
25. S. Y. Yang, H. K. Kim, A. W. M. Lefeber, C. Erkelens, N. Angelova, Y. H. Choi, R. Verpoorte. *Planta Med*. **2006**, *72*, 364.
26. K. A. Leiss, Y. H. Choi, I. B. Abdel-Farid, R. Verpoorte, P. G. L. Klinkhamer. *J. Chem Ecol.* **2009**, *35*, 219.
27. H. T. Widarto, E. Van der Meijden, A. W. M. Lefeber, C. Erkelens, H. K. Kim, Y. H. Choi, R. Verpoorte. *J. Chem. Ecol*. **2006**, *32*, 2417.
28. Y. H. Choi, E. C. Tapias, H. K. Kim, A. W. M. Lefeber, C. Erkelens, J. Th. J. Verhoeven, R. Verpoorte. *Plant Physiol.* **2004**, *135*, 2398.
29. Y. H. Choi, H. K. Kim, H. J. M. Linthorst, J. G. Hollander, A. W. M. Lefeber, C. Erkelens, J.-M. Nuzillard, R. Verpoorte. *J. Nat. Prod*. **2006**, *69*, 742.
30. A. T. Cardoso-Taketa, R. Pereda-Miranda, Y. H. Choi, R. Verpoorte, M. L. Villarreal. *Planta Med.* **2008**, *74*, 1295.
31. Y. H. Choi, S. Sertic, H. K. Kim, E. G. Wilson, F. Michopoulosa, A. M. W. Lefeber, C. Erkelens, S. D. P. Kricun, R. Verpoorte. *J. Agric. Food Chem*. **2005**, *53*, 1237.
32. Y. H. Choi, H. K. Kim, A. Hazekamp, C. Erkelens, A. W. M. Lefeber, R. Verpoorte. *J. Nat. Prod.* **2004**, *67*, 953.
33. M. Frédérich, Y. H. Choi, L. Angenot, G. Harnischfeger, A. M. W. Lefeber, R. Verpoorte. *Phytochemistry* **2004**, *65*, 1993.
34. V. Formacek, K. H. Kubeczka. *Essential Oils Analysis by Capillary Gas Chromatography and Carbon-13 NMR Spectroscopy*. Wiley: Chichester, **1983**.