Metabolomics: will it stay?

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

Phytochemical Analysis

Received: 19 October 2009; Accepted: 19 October 2009 Published online in Wiley Interscience

(www.interscience.wiley.com) DOI 10.1002/pca.1191

Metabolomics: Will it Stay?

Robert Verpoorte,* Young Hae Choi and Hye Kyong Kim

In the past 20 years, *Phytochemical Analysis* has published numerous papers on the analysis of plant products by means of chromatography, mass spectrometry and NMR-based methods. Most have targeted a specific group of compounds such as alkaloids, terpenoids, flavonoids, coumarins, you name it. Typically, phytochemical analysis deals with the development of a suitable extraction method, yielding a sample that gives good results in the subsequent analytical method in which the compounds are identified and quantified. For almost all known compounds, numerous papers have been published in scientific journals concerning their analysis in different materials, reporting an improved extraction method, or better conditions for the analysis, most often using new ('improved') columns. Also, detection limits are always improving—almost as if there were some sort of world championship for the most sensitive method. If we take the analysis of the pharmaceutically important tropane alkaloids hyoscyamine and scopolamine as an example, we see that, following the first report of their HPLC separation in 1976 (Verpoorte and Baerheim Svendsen, 1976), by the early 1980s the number of publications on the GC and HPLC analysis of these compounds was already more than 80 (Verpoorte and Baerheim Svendsen, 1984). If we now make a literature search, the number of publications on this topic runs into the hundreds, and in the past year alone some 15 papers dealing with the HPLC analysis of these alkaloids have been published. Apparently no standardisation has been achieved in all these years. Official pharmacopoeia methods may have a longer life, but for other research purposes it seems that every research group has its own preferences. One may wonder if this is worth all the efforts of researchers, editors, reviewers and publishers. How much advancement has been made in terms of extraction, sensitivity and selectivity? Would we have advanced more if we had spent the time and money on the real scientific questions for which the methods were developed instead of trying to improve the methods all the time?

These questions have also to be asked when discussing metabolomics. As the latest of the 'omics', it contributes to the total analysis of an organism: genome, transcriptome, proteome and metabolome. The 'omics' thus cover the total chemistry from DNA to organic molecules. Molecular biology is characterised by highly standardised extraction methods as the basis for isolating DNA or RNA for subsequent sequencing and identification. Methodology for sequencing is rapidly improving, both in sensitivity and speed, but the final output is in the form of sequences that can be stored for future data-mining and identification. This is information that is not dependent on the method used for isolation and sequencing. On the level of proteins, however, things become more complex as there are water-soluble proteins and membrane-bound proteins and these cannot be extracted in a single operation. Moreover, levels of proteins differ greatly, the number of building blocks is much larger than for DNA and RNA, and all kinds of chemical modifications occur. One thus sees that a total proteome analysis is not feasible, but rather the analysis has to be divided into different parts, each focusing on a certain subset of the proteome. However, at least the final sequence obtained is again independent of the methods used. For both transcriptomics and proteomics, the quantitative analysis in different samples is still posing major challenges.

When we go to the level of the metabolome, which encompasses small molecules (with an average molecular weight of some 200–500) up to macromolecules (like polysaccharides and lignin) and everything in between, the complexity increases even further. Although the number of elements present is limited (C, H, O, N, P and S being most common) the chemical space that they represent is almost infinite (Verpoorte, 2000, 2009). This is reflected in an enormously broad variety of structures with very different physical and chemical characteristics, such as acidic, basic, oxidant and antioxidant properties and polarity. Also stability varies greatly. Finally some compounds can be present at levels of several per cent of the dry weight of a plant, whereas others are only present in ppm or even lower levels. Whilst the number of compounds in an organism is quite large, no real number is known, although an educated guess would be that it is in the same order as the number of genes in an organism, and this means that in a plant some 30,000 compounds may exist. This all makes the analysis of a metabolome quite a challenging task, and in fact at present there is no method that can give a complete qualitative and quantitative picture of the metabolome. Each method has its own advantages and limitations (Verpoorte *et al.*, 2008).

Obviously from the above discussion it is clear that there is at present no single method that can achieve the analysis of all metabolites, but at least reproducibility seems a more reasonable goal to achieve, though for the methods mentioned there will be clear differences. In particular, chromatographic methods have the problem of dependence on stationary phases, whereas NMR is quite reproducible as it concerns the measurement of physical characteristics of compounds.

In recent years the number of papers with the term metabolomics in the title has exploded; however, in many cases they are in fact not different from the classical targeted approaches, i.e. only a limited selected group of compounds is measured, and no

* Correspondence to: R. Verpoorte, Department of Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, PO Box 9502, 2333CC Leiden, The Netherlands. E-mail: verpoort@chem.leidenuniv.nl

effort is made to arrive at any form of standardisation and building of a public database for long-term data-mining.

In what way is metabolomics different from the classical methods? There are many definitions for metabolomics and metabonomics, but perhaps definitions are of limited value and it may be better to describe the ultimate goal, which is the highly reproducible qualitative and quantitative analysis of all metabolites in an organism under certain conditions. Reproducibility is very important as metabolomic results should be available for data-mining in a public library, similar to the other 'omics'. The strength of such an approach is best seen in NMR-based metabonomics of urine (Lindon *et al.*, 2007). Because of the high reproducibility, results of 20 years of work are available for data-mining, resulting in continuous new discoveries by combining old data with those of novel experiments (Lindon *et al.*, 2007; Nicholson and Lindon, 2008).

In fact metabolomics is thus first of all about standardisation and validation to obtain highly reproducible results that over many years to come can be used for data-mining. Some methods that seem to fulfill these demands have been reported in the past years for NMR (Ludwig and Viant, 2010; Verpoorte *et al.*, 2007, 2008), GC-MS (Lisec *et al.*, 2006) and LC-MS (De Vos *et al.*, 2007). However, still a real large scale database with metabolomic data from many plant species has not been achieved. Even in a field such as essential oil analysis, after more than 40 years of GC-work since the 1960s (Svendsen and Karlsen, 1967), there is not yet a public database in which everybody can deposit results, thus making them available for data-mining. The major reason is probably that each individual researcher has reasons to choose optimal conditions for a specific experiment. Certainly conditions may be selected that separate as many compounds as possible for a specific material; however, the problems of identifying each single peak are huge if no comparison is possible with other results.

The question is whether the quality of a metabolomic study is in the quantity or the quality of the data, i.e. see more peaks or signals with many unknowns and uncertainties, or have fewer peaks or signals but all with a complete identification and including the possibility to compare results directly with other studies using chemometric methods. In our view, it is the latter which is needed since that forms the basis from which, in a subsequent step, one may go into more detail by using alternative methods. The first step should be involve globally accepted methods for which databases exist for depositing results and with which data-mining is possible. NMR has been shown to be such a method in the analysis of body fluids, in metabonomics (Nicholson and Lindon, 2008; Lindon and Nicholson, 2008). Other methods have not yet reached this level of standardisation. As mentioned above, we see more and more that the term 'metabolomics' is used for methods that have been used already in the past decades, without any consideration about developing a method that yields sustainable results. If the field is not able to come to rigid standardisation of the methods employed (Sumner *et al.*, 2007), we fear that support for metabolomics will gradually diminish, leaving metabonomics as the example of how metabolomics should have been developed.

This special issue of *Phytochemical Analysis* offers a number of exciting examples of how metabolomics can contribute to solving all kinds of biological questions. Initiatives such as the MIAMI protocols for transcriptomics (Brazma *et al.*, 2001) are also required for metabolomics. The first step on the road has been made, but considering the complexity of the metabolome, and the diversity of analytical tools, it will not be an easy task to come to general standardised and validated protocols.

References

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. 2001. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* **29**: 365–371.

De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD. 2007. Untargeted large scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2**: 778–791.

Lindon JC, Nicholson JK. 2008. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem* **1**: 45–69.

Lindon JC, Holmes E, Nicholson JK. 2007. *The Handbook of Metabonomics and Metabolomics*. Elsevier: Amsterdam.

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. 2006. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* **1**: 387–396.

Ludwig C, Viant MR. 2009. Two-dimensional *J*-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem Anal* **21**: 22–32.

Nicholson JK, Lindon JC. 2008. Systems biology—metabonomics. *Nature* **455**: 1054–1056.

Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**: 211–221.

Svendsen AB, Karlsen J. 1967. Gas chromatography of monoterpene hydrocarbons from etheric oils on packed separation columns with low content of fluid stationary phase. Terpenes and related compounds. V. *Planta Med* **15**: 1–5.

Verpoorte R. 2000. Plant secondary metabolism. In *Metabolic Engineering of Plant Secondary Metabolism*, Verpoorte R, Alfermann AW (eds). Kluwer Academic: Dordrecht; 1–29.

Verpoorte R. 2010. *Comprehensive Natural Products Chemistry II*, Vol. 3, Elsvier: Amsterdam. In press.

Verpoorte R, Baerheim Svendsen A. 1976. High-performance liquid chromatography of some tropane alkaloids. *J Chromatogr* **120**: 203–205.

Verpoorte R, Baerheim Svendsen A. 1984. *Chromatography of alkaloids. Part II. GLC and HPLC*. Chromatography Library Vol. 23B. Elsevier: Amsterdam.

Verpoorte R, Choi YH, Kim HK. 2007. NMR-based metabolomics at work in phytochemistry. *Phytochem Rev* **6**: 3–14.

Verpoorte R, Choi YH, Mustafa RN, Kim HK. 2008. Metabolomics: back to basics. *Phytochem Rev* **7**: 525–538.