About the Editor

Serge Rudaz

Serge Rudaz studied pharmacy in Switzerland, where he obtained his PhD in 1997. Later, he joined the National Research Center in Rome (Italy) for a postdoctoral position concerning the application of capillary electrophoresis hyphenated to MS for chiral separation in biological fluids. From 1998 to 2011, he was Master-Assistant in Phytochemistry at the University of Lausanne (Switzerland), then Maître d'Enseignement et de Recherche at the University of Geneva (Switzerland), where he developed new strategies for untargeted metabolomic analyses. Since 2010, he has been Director of the Group of Pharmaceutical Analytical Chemistry in the School of Pharmaceutical Sciences. He was promoted to Associate Professor at the University of Geneva in 2012. He has so far contributed to the field of pharmaceutical analysis with diverse activities including invited lectures and invited professorships in various universities (Lyon, Pavia). He is a member and reviewer of numerous journals and societies. He is also member of several congress scientific boards, such as the 'Chimiométrie' series. He has published more than 220 scientific papers and 10 book chapters in analytics, mainly focused on separation sciences, bioanalysis data treatment and metabolomics. Currently, he is interested in (Ultra High-Pressure) LC and capillary electrophoresis coupled to MS, advances in sample preparation analysis of pharmaceuticals, biological matrices, clinical and preclinical studies, including metabolism and toxicological analysis, validation and regulation (ISO 17025) and various chemometric approaches including experimental design and data mining for metabolomics.

For reprint orders, please contact reprints@future-science.com

Foreword

Identification and data-processing methods in metabolomics

Serge Rudaz

Numerous domains of life sciences have recently benefited from new technological developments of analytical platforms. Today, assessing relationships between genes, proteins and metabolites has become a mature field of research. System biology, a concept proposed some years ago, appears now to be a reality thanks to new capabilities for recording signals related to the complexity of living organisms. In this context, metabolomics, as the ultimate point of the related 'omics technologies, can be considered as the most functional view of a given system and the most appropriate tool among the 'omics sciences for assessing biochemical processes related to complex phenotypes.

The choice of the data acquisition and analysis strategy significantly depends on the questions that are asked. While a very small number of metabolites can be sufficient for diagnostic purposes (as in the case of disease evaluation), an extended set of compounds may be desired when a biochemical network is under examination. Metabolite fingerprinting applications have rapidly increased in the last few years, but this untargeted strategy generates massive datasets of various structures that remain challenging to handle and exploit.

As discussed in several chapters of this book, valuable metabolic results can be obtained in a very efficient and high-throughput manner, but to benefit to the largest extent from this data deluge, numerous issues have to be solved by adapted data-mining strategies, including dimensionality reduction. The choice of the most appropriate model for a given dataset

therefore constitutes an important challenge, and extracting valuable information has therefore become crucial. New chemometric approaches still need to be implemented to push forward metabolomics as an important paradigm shift of the knowledge discovery process (i.e., the data-driven strategy).

A key aspect of metabolomic data analysis, specifically discussed in this book, concerns the metabolite identification process – one of the main bottlenecks in the metabolomic workflow. This step is particularly challenging when no isolation of putative compounds can be envisaged, as in plant metabolomics, for example. Biomarker identification plays a central role, as it is mandatory to evaluate the quality of the information obtained from mathematical models and also to link results with biological knowledge, when available. Precious information can be obtained from MS detection in various metabolomic studies and the MW of the biomarkers constitutes the starting point of most identification strategies. However, identifying a given biomarker based solely on the MS information is not a straightforward task. Additional structural information could be obtained using fragmentation, high-resolution measurements and/or advanced MS acquisition mode such as ion mobility-based approaches.

Accessing metabolic databases should considerably increase the speed of the biomarker identification process. Unfortunately, most of the available repositories remain incomplete and suffer from the fact that MS/MS fragmentation patterns are often specific to the acquisition platforms. For now, comparisons of MS/MS spectra using in-house libraries can be performed but new algorithms are mandatory to assist researchers in the interpretation of the fragmentation, in the perspective of a structural identification. As presented here, the other major tool for metabolomics studies, NMR, will certainly contribute to the constant improvement of metabolomics thanks to its reproducibility and high-throughput capacity, which does not require specific sample preparation.

The next challenge for metabolomics will probably be related to the capacity for handling multiple data tables generated by several analytical platforms. For numerous laboratories, this consists of processing an independent data analysis of the different datasets, manually comparing the results using high-level data integration and attempting to establish links between discriminant features. However, to discover the possible interrelations between features or signals that are measured independently, multiblock models should be considered. The latter will provide new insights for the simultaneous analysis of multiple data tables, assessing the relevance of the selected analytical approaches and leading to better understanding of some of the numerous relationships that could exist between different datasets of possibly related variables. Furthermore, the possibility to combine the information afforded

by MS/MS and NMR spectra will efficiently help scientists to accelerate the pace of biomarker identification.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.