

Turning Metabolomics into Drug Discovery

Asmaa Boufridi and Ronald J. Quinn*

Eskitis Institute for Drug Discovery, Griffith University, QLD 4111 Brisbane, Australia

Metabolomics is the “omics” that studies the whole metabolome. It has a wide range of applications, *inter alia* chemotaxonomy, environmental influences, agriculture, etc. Here we review the application of metabolomics in natural product research. The importance of physicochemical properties to drug delivery are discussed in relation to turning metabolomic studies towards drug discovery. We believe that coupling metabolomic studies with standards of known physicochemical properties in order to calibrate the chromatographic columns can be beneficial in identifying compounds of candidate drug quality.

Keywords: natural products, HPLC, retention time, log P, druggability

1. Introduction

Metabolomics is the characterization and quantification of the large number of metabolites that occur in biological systems. Metabolomics approaches are more and more used in quality control,¹ chemical ecology,² chemotaxonomy³ and identification of novel natural products.⁴ Focusing on problems such as the characterization of primary and secondary metabolites in order to establish a link between the constituents of an organism and its environment (geographical origin, genetic, climates or seasonal changes, etc.) requires complete metabolite fingerprints to detect any differences between the samples and generate hypotheses to explain these differences.

Primary and secondary metabolites refer to small molecules (MW < 1000 Da)⁵ made by living organisms via interaction with biosynthetic enzymes.⁶ Primary metabolites are essential for the survival of the organism, and they include amino acids, carbohydrates and lipids, on the other hand, the secondary metabolites are non-essential for the organism's life but necessary for its survival, and are used as defenses or to communicate with their environment. They include: terpenes, alkaloids, polyketides, hormones and polyphenols.⁵

Metabolomics is also used for dereplication of microorganisms to identify productions of novel compounds.⁷ Multivariate statistical analyses such as principal components analysis (PCA) are used in metabolomics to simplify multi-parametric data into a

correlation pattern. PCA is an unsupervised method that can show a correlation between different samples and highlight either differences or similarities between datasets. In addition to PCA, which compares the variance between different samples, partial least squares (PLS)-discriminant analysis (DA) and other targeted analytical profiling are used in order to establish a degree of “biological similarity” between samples. Multivariate analysis using statistical tools such as unsupervised principal component analysis, supervised discriminant function analysis and Z-score analysis can be used for pattern recognition which may predict that biological activity results from particular features of compounds.⁸

Various chromatographic and spectroscopic techniques, mostly hyphenated techniques (liquid chromatography (LC)-mass spectrometry (MS), LC-nuclear magnetic resonance (NMR), etc.)⁹ are used for metabolomics but high performance liquid chromatography (HPLC) is still the first choice in metabolomics to study the chemical composition of natural crude extracts and the rapid detection of known compounds.¹⁰

In this mini-review article, we discuss the application of metabolomics using chromatographic analysis to study the whole metabolome and how to adapt this information to drug discovery.

2. Applications of Metabolomics

Brazil covers an area of 8,511,996 km² with several ecosystems and important biodiversity¹¹ that has attracted Brazilian studies around natural product metabolomics.

*e-mail: r.quinn@griffith.edu.au

Dereplication was used to identify and quantify tocopherol from Brazil oil nut from different geographic locations in Brazil using HPLC, revealing little changes in the amount of tocopherol (between 144.80 and 234.26 mg g⁻¹). This study showed also the presence of the two main tocopherols (α - and β -tocopherol) in all of the authentic oil samples with differences in their amounts, however, some commercial Brazil nut oils did not show the presence of tocopherols at all.¹² In order to select superior banana genotypes for breeding, dereplication of 29 samples from different genomic groups were studied to determine the phenolic contents and carotenoid profiles. This study aimed to identify samples with high concentrations in carotenoids, which are known as pro-vitamin A. The HPLC analysis showed an appreciable amount of pro-vitamin A carotenoids in the active germosplasm samples comparing with the main cultivars that are currently marketed.¹³ Dereplication of natural products has also been used to study marine organisms; the use of LC-photodiode array detector (PDA)-MS analysis of 14 Brazilian sponge specimens of the genus *Aplysina*, in order to detect bromotyrosine-derived metabolites based on their UV absorption (on the basis of the three most typical chromophores known for Verongida dibromotyrosine-derived metabolites),¹⁴ did not show a significant difference in the chemical profile of the samples. In order to differentiate between the chemical composition of six different *Lippia* species, an ultra HPLC (UHPLC) metabolomics analysis combined with extraction of molecular formulae based on time-of-flight (TOF)-MS data together with the use of filters (including log P) were used.¹⁰ The fast separation capacity of UHPLC and the high quality of the obtained chemical profiles allowed the discrimination of the samples and gave a precise picture of the chemical relationship. In addition, the authors used hierarchical clustering analyses to provide an efficient approach to discover cluster relationships between the samples.¹⁰ Dereplication of 57 leaf extracts from Brazilian Asteraceae species were analyzed by HPLC-MS and subjected to machine learning algorithms in order to determine biomarkers with anti-inflammatory potential. Using a genetic algorithm, 1241 chromatographic peaks out of 6052 were selected according to their anti-inflammatory activities, allowing the determination of 11 biomarkers.¹⁵ In order to distinguish coffee genotypes grown in three different regions in Brazil, gas chromatography-single quad (GC-Q)/MS coupled with statistical analysis showed that some of the 44 metabolites identified can be used as chemomarkers for origin and genotype differentiation.¹⁶ In order to identify analogues of isopimarane diterpenes

in complex crude extracts of Velloziaceae, electrospray ionization tandem MS (ESI-MS/MS) was applied to a series of isolated diterpenes from the same family. The compounds have different number of hydroxyls that were observed in the mass spectrum by the multiple loss of water. The authors suggested that the intensity of protonated and cationized compounds could be used to differentiate these compounds in complex mixtures. The MS/MS fragmentation did not show a difference between the isopimarane diterpenes, but could be used to distinguish them from their isomers.¹⁷ Because of the challenges encountered in the identification and characterization of all metabolites during metabolomic studies, a rational and sequential method was investigated, 14 different solvents were evaluated for their extraction efficiency on *Jatropha gossypifolia*. Design of experiments (DoE) and partial least squares (PLS) were used to correlate the physicochemical properties of the different solvents with the chromatographic profiles. It has been shown that the physicochemical properties of the solvent can influence significantly the extraction capacity.¹⁸ Targeted and untargeted metabolomics were used in order to discover biomarkers and to search for novel active compounds against neglected diseases in Brazil (Chagas disease, dengue, leishmaniasis, leprosy, malaria, schistosomiasis and tuberculosis).¹⁹ It was shown that most of the metabolites identified were amino acids, carbohydrates, lipids, organic acids, nucleosides and fatty acids.¹⁹ Green and brown Brazilian propolis were collected from different regions in Brazil, analyzed by GC-MS and analyzed by multivariate analysis in order to evaluate their chemical profiles and identify active compounds. Green and brown propolis showed different chemical compositions: brown propolis was rich in triterpenoids, however, green propolis was rich in sesquiterpenes and steroids.²⁰

3. Chemometric Analysis: PCA

Because of the huge amount of data acquired during metabolomic studies, the use of chemometric and statistical analysis is indispensable for data mining and visual interpretation.²¹ PCA is one of the most widely used multivariate statistical analysis tools used in metabolomics analysis.^{16,22} It is a reliable and easy tool to compare metabolomics profiling. It has been shown that the use of ¹H NMR metabolomics with PCA analysis can be very effective in the discrimination of samples: (i) this technique allowed the differentiation of 12 *Cannabis sativa* cultivars and was shown to be very promising for the authentication and quality control of *C. sativa*;²³ and (ii) the same method

allowed the discrimination of 11 *Ilex* species, and allowed the identification of arbutin, which has not been reported before as a constituent of *Ilex* species and was found to be a biomarker of 8 of the studied species.²⁴

PCA combined with molecular networking and dereplication have been shown to be valuable tools for researchers focusing on microbial extract; molecular networking allows rapid comparison of metabolites profiles from complex crude fermentation extracts, for effective chemical dereplication and discovery of novel compounds.^{25,26} The influence of bacterial isolation location (from Scottish and Antarctic sediments) on bioactive secondary metabolite production within the marine environment was studied using molecular networks for the first time. This study showed a high degree of biogeographic influence upon secondary metabolite production and comparative metabolomics aided in the metabolite dereplication of over 3500 parent ions from these marine sources and has provided targets for further purification.²⁵

4. Metabolomics and Drug Discovery: What Makes a Drug? (Importance of Physicochemical Properties)

Metabolomic studies aim to study the whole metabolome, to identify known compounds or to characterize a class of compounds, sometimes without isolating the metabolites. On the other hand, a drug is a single compound that has to be delivered to the patient and reach a suitable site to treat disease. Progressing from metabolomics analysis to drug discovery requires knowledge of physicochemical properties. A drug should have certain physicochemical properties, defined by Lipinski *et al.*²⁷ as a rule-of-five; this rule is a set of four simple physicochemical properties for oral bioavailability: MW \leq 500 Da, log P \leq 5, H-bond donors \leq 5, and H-bond acceptors \leq 10. Recently, data concerning the cause of attrition, toxicology and safety of 812 oral small-molecule drug candidates from AstraZeneca, Eli Lilly, GlaxoSmithKline (GSK) and Pfizer between 2000-2010 were pooled. A set of physicochemical descriptors was also calculated for those drug candidates in order to establish links between the physicochemical properties and compound attrition. It has been shown that there is a greater attrition for more lipophilic compounds. In addition, the majority of compounds had shown desirable ranges, with 75% of compounds having a MW < 499 Da and 75% having log P < 4.4.²⁸

Turning metabolomics to drug discovery might be achieved in a very simple way, our hypothesis is to use a mixture of compounds with known physicochemical

properties to calibrate the HPLC column and limit the area of the metabolome to be considered in order to progress down a drug discovery pathway.

5. Retention Time (t_R) and log P

HPLC as used for metabolomics can be used as a surrogate for log P upon suitable calibration. Prediction of retention parameters have been obtained²⁹ using chemoinformatic, principally quantitative structure-retention relationship (QSRR) studies, to obtain statistical correlation between chromatographic retention time and theoretical properties of the metabolites.²⁹ t_R is correlated to the physicochemical parameters of a given compound, specifically to log P. Retention time estimation based on the calculated log P is increasingly used, for example, allowing the differentiation between isomeric flavonoid aglycones that eluted at different retention times and possessed different log P values without the need for isolation.¹⁰ However, for more hydrophobic molecules, GSK showed that measuring the chromatographic hydrophobicity index (CHI) might be more reliable than measuring log P as the standard model in drug discovery, and also, CHI has been shown to be linear, non-solubility dependent, predictable and relevant.³⁰

6. Conclusions

Figure 1 shows the relationship between metabolome space, drug space and the two most important physicochemical properties (log P and MW), and illustrates how using these properties with retention time can help to limit metabolome space to that portion relevant to drug space.

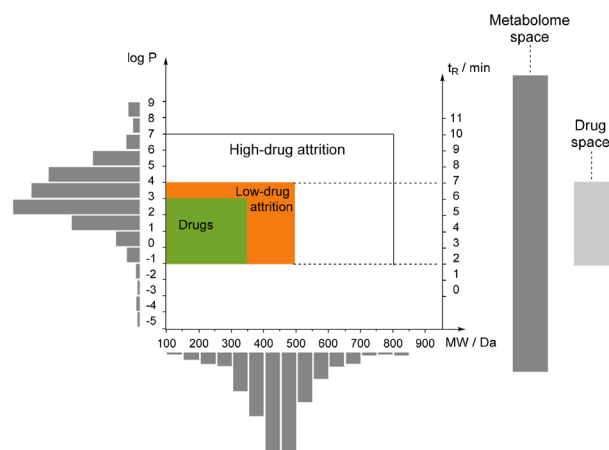


Figure 1. Scheme representing the distribution of drugs (green), low-drug attrition (orange),³¹ molecular mass (Da) (grey columns),²⁸ calculated log P (grey bars)²⁸ and retention time (t_R)³² in metabolome space.

The physicochemical properties of a given compound are critical to drug delivery and metabolism. Metabolomics characterizes a wider range of metabolites; using standards with known physicochemical properties (log P or CHI) in order to calibrate the chromatographic columns can be beneficial in identifying compounds of candidate drug quality (Figure 1).

Acknowledgments

This research was supported by an Australian Research Council *Discovery Projects* funding (project No. DP130102400).



Asmaa Boufridi is a natural product chemist. She obtained her Master's degree from the University of Blida in Algeria in 2011, and her PhD in marine natural products from the University Paris-Sud in France in 2014. Asmaa joined the Eskitis Institute, Griffith University, in 2015 as a research fellow with Prof Ron Quinn. Her research focuses on the isolation and structural elucidation of novel and active compounds from plants and marine organisms using NMR metabolomics.



Ronald Quinn obtained his PhD in Chemistry from the University of New South Wales, Australia in 1970. He joined Griffith University in 1982 and is Professor and Director of the Eskitis Institute for Drug Discovery. His research concentrates on the use of molecules as probes to understand interactions in biological systems, conducting research on bioactive natural products and bioaffinity mass spectrometry to identify protein-ligand complexes. He is a Member of the Order of Australia (AM) for service to scientific research, in the field of chemistry and a Fellow of the Australian Academy of Technological Sciences and Engineering.

References

- Batista, A. N. L.; Colombo, R.; de Pascoli, I. C.; Teles, H. L.; Silva, G. H.; Bomfim, G. C. C.; Burgos, R. C. R.; Cavalheiro, A. J.; Bolzani, V. S.; Silva, D. H. S.; Reimberg, M. C. H.; *Rev. Bras. Farmacogn.* **2011**, *21*, 402.
- Paul, C.; Mausz, M. A.; Pohnert, G.; *Metabolomics* **2012**, *9*, 349; Kuhlisch, C.; Pohnert, G.; *Nat. Prod. Rep.* **2015**, *32*, 937.
- Kim, H. K.; Saifullah; Khan, S.; Wilson, E. G.; Kricun, S. D. P.; Meissner, A.; Goralder, S.; Deelder, A. M.; Choi, Y. H.; Verpoorte, R.; *Phytochemistry* **2010**, *71*, 773; Ernst, M.; Silva, D. B.; Silva, R.; Monge, M.; Semir, J.; Vêncio, R. Z. N.; Lopes, N. P.; *Anal. Chim. Acta* **2015**, *859*, 46.
- Li, Z.; Wang, Y.; Ouyang, H.; Lu, Y.; Qiu, Y.; Feng, Y.; Jiang, H.; Zhou, X.; Yang, S.; *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2015**, *988*, 45; Nielsen, K. F.; Smedsgaard, J.; *J. Chromatogr. A* **2003**, *1002*, 111.
- Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S.; *J. Chromatogr. A* **2015**, *1382*, 136.
- Sturm, N.; Quinn, R. J.; Kellenberger, E.; *Planta Med.* **2015**, *81*, 467.
- Nielsen, K. F.; Mansson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O.; *J. Nat. Prod.* **2011**, *74*, 2338; Eshell, M.; Harvey, L.; Edrada-Ebel, R.; McNeil, B.; *Toxins* **2015**, *7*, 439; Cheng, C.; MacIntyre, L.; Abdelmohsen, U. R.; Horn, H.; Polymenakou, P. N.; Edrada-Ebel, R.; Hentschel, U.; *PLoS One* **2015**, *10*, e0138528.
- Tawfike, A. F.; Viegmann, C.; Edrada-Ebel, R.; *Methods Mol. Biol.* **2013**, *1055*, 227; Yuliana, N. D.; Khatib, A.; Choi, Y. H.; Verpoorte, R.; *Phytother. Res.* **2011**, *25*, 157; Yuliana, N. D.; Khatib, A.; Verpoorte, R.; Choi, Y. H.; *Anal. Chem.* **2011**, *83*, 6902; Barding, G. A.; Orr, D. J.; Sathnur, S. M.; Larive, C. K.; *Anal. Bioanal. Chem.* **2013**, *405*, 8409.
- Wolfender, J.-L.; Queiroz, E. F.; Hostettmann, K.; *Expert Opin. Drug Discovery* **2006**, *1*, 237; Morais, R. P.; Lira, S. P.; Selegim, M. H. R.; Berlinck, R. G. S.; *J. Braz. Chem. Soc.* **2010**, *21*, 2262; Carnevale Neto, F.; Siquitelli, C. D.; Pilon, A. C.; Silva, D. H. S.; Bolzani, V. S.; Castro-Gamboa, I.; *J. Braz. Chem. Soc.* **2013**, *24*, 758.
- Funari, C. S.; Eugster, P. J.; Martel, S.; Carrupt, P.-A.; Wolfender, J.-L.; Silva, D. H. S.; *J. Chromatogr. A* **2012**, *1259*, 167.
- Valli, M.; Pivatto, M.; Danuello, A.; Castro-Gamboa, I.; Silva, D. H. S.; Cavalheiro, A. J.; Araújo, A. R.; Furlan, M.; Lopes, M. N.; Bolzani, V. S.; *Quim. Nova* **2012**, *35*, 2278.
- Funasaki, M.; Menezes, I. S.; Barroso, H. S.; Zanotto, S. P.; Carioca, C. R. F.; *Acta Amazonica* **2013**, *43*, 505.
- Borges, C. V.; Amorim, V. B. O.; Ramlov, F.; Ledo, C. A. S.; Donato, M.; Maraschin, M.; Amorim, E. P.; *Food Chem.* **2014**, *145*, 496.
- Silva, M. M.; Bergamasco, J.; Lira, S. P.; Lopes, N. P.; Hajdu, E.; Peixinho, S.; Berlinck, R. G. S.; *Aust. J. Chem.* **2010**, *63*, 886.
- Chagas-Paula, D. A.; Oliveira, T. B.; Zhang, T.; Edrada-Ebel, R.; Da Costa, F. B.; *Planta Med.* **2015**, *81*, 450.
- da Silva Taveira, J. H.; Borém, F. M.; Figueiredo, L. P.; Reis,

- N.; Franca, A. S.; Harding, S. A.; Tsai, C.-J.; *Food Res. Int.* **2014**, *61*, 75.
17. Pinto, A. C.; Vessecchi, R.; da Silva, C. G.; Amorim, A. C. L.; dos Santos Jr., H. M.; Rezende, M. J. C.; Gates, P. J.; Rezende, C. M.; Lopes, N. P.; *Rapid Commun. Mass Spectrom.* **2016**, *30*, 61.
18. Pilon, A. C.; Carnevale Neto, F.; Freire, R. T.; Cardoso, P.; Carneiro, R. L.; Bolzani, V. S.; Castro-Gamboa, I.; *J. Sep. Sci.*, *in press*, DOI: 10.1002/jssc.201500892.
19. Castro, C. B.; Pires, D. O.; *Bull. Mar. Sci.* **2001**, *69*, 357.
20. Bittencourt, M. L. F.; Ribeiro, P. R.; Franco, R. L. P.; Hilhorst, H. W. M.; de Castro, R. D.; Fernandez, L. G.; *Food Res. Int.* **2015**, *76*, 449.
21. Marhuenda-Egea, F. C.; Gonsálvez-Álvarez, R. D.; Lledó-Bosch, B.; Ten, J.; Bernabeu, R.; *Anal. Chem.* **2013**, *85*, 3053; Kuhn, S.; Bernardi Ogliaari, J.; Dias, P. F.; da Silva Santos, M.; Ferreira, A. G.; Bonham, C. C.; Wood, K. V.; Maraschin, M.; *J. Agric. Food Chem.* **2010**, *58*, 2194.
22. Nguyen, H. T.; Lee, D.-K.; Lee, W. J.; Lee, G.-J.; Yoon, S. J.; Shin, B.-k.; Nguyen, M. D.; Park, J. H.; Lee, J.; Kwon, S. W.; *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2016**, *1012-1013*, 61.
23. Choi, Y. H.; Kim, H. K.; Hazekamp, A.; Erkelens, C.; Lefeber, A. W. M.; Verpoorte, R.; *J. Nat. Prod.* **2004**, *67*, 953.
24. Choi, Y. H.; Sertic, S.; Kim, H. K.; Wilson, E. G.; Michopoulos, F.; Lefeber, A. W. M.; Erkelens, C.; Prat Kricun, S. D.; Verpoorte, R.; *J. Agric. Food Chem.* **2005**, *53*, 1237.
25. Purves, K.; Macintyre, L.; Brennan, D.; Hreggviðsson, G.; Kuttner, E.; Ásgeirsdóttir, M.; Young, L.; Green, D.; Edrada-Ebel, R.; Duncan, K.; *Metabolites* **2016**, *6*, 2.
26. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C.; *J. Nat. Prod.* **2013**, *76*, 1686.
27. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.; *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
28. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A.; *Nat. Rev. Drug Discovery* **2015**, *14*, 475.
29. Oliveira, T. B.; Gobbo-Neto, L.; Schmidt, T. J.; Da Costa, F. B.; *J. Chem. Inf. Model.* **2015**, *55*, 26.
30. Young, R. J.; Green, D. V. S.; Luscombe, C. N.; Hill, A. P.; *Drug Discovery Today* **2011**, *16*, 822.
31. Pascolutti, M.; Quinn, R. J.; *Drug Discovery Today* **2014**, *19*, 215.
32. Camp, D.; Davis, R. A.; Campitelli, M.; Ebdon, J.; Quinn, R. J.; *J. Nat. Prod.* **2012**, *75*, 72.

Submitted: February 4, 2016

Published online: March 18, 2016