The biological mechanisms that direct the generation and accumulation of the vast diversity of metabolites observed in the plant kingdom are not fully understood. An exciting and promising approach to understand these mechanisms is described in the paper by Xie et al. (2009). The authors have coupled state-of-the-art metabolomic analyses with novel bioinformatic techniques to identify apparent ‘metabolic modules’ in turmeric (*Curcuma longa*) rhizomes. A metabolic module is defined as a group of co-regulated metabolites and this approach elegantly represents a basic innovative and practical attempt to understand and predict metabolic pathways using detailed bioinformatics data mining following careful and well-documented GC-MS and LC-MS analyses. More interestingly, the authors describe the use of metabolic modules and their patterns of apparent co-regulation to predict biochemical metabolic pathways. Turmeric was chosen as an excellent model plant to test this approach because it produces a large number of secondary metabolites from diverse metabolic pathways and networks. Among other compounds, turmeric rhizomes accumulate diarylheptanoids derived from the phenylpropanoid pathway, some of which have important bioactivities. Sixteen different growth and development treatments were performed and metabolically profiled, followed by hierarchical cluster analysis in order to detect co-regulated volatile and non-volatile metabolites. Biosynthetic relationship predictions were then made using the metabolite modules that resulted from the processing of the metabolomic information collected (Xie et al., 2009). The dissection of the metabolic information into metabolic modules led to the conclusion that more than one polyketide synthase is involved in diarylheptanoid biosynthesis in turmeric rhizomes, but the potential for using a similar approach in deciphering complex and apparently unrelated metabolic pathways is also very promising.

In recent years it has become evident that biological systems are much more complex than initially thought. The large phytochemical diversity found in nature results from different selection pressures that plants have successfully coped with during evolution or cultivation (Lewinsohn and Gijzen, 2008). However, there are some problematic issues in metabolomics that need to be addressed while attempting to integrate metabolomic data with proteomics and transcriptomics. Such integration is required for the study of the intricate mechanisms that modulate the production and accumulation of plant secondary metabolites. The analysis of a full metabolome is technically problematic, due to the lack of a universal protocol for extraction, identification, and quantification of all the metabolites present in a cell or tissue at a particular time. Different metabolites need different chemical extraction and analysis procedures for their accurate determination. Thus, the precursors often behave differently during extraction and quantification than the final products of a given biochemical pathway and, therefore, either the precursors or the final products may remain undetected or may be erroneously determined when using a more universal chemical extraction and determination procedure. Therefore, in order to minimize such problematic cases, clustering of the data obtained by LC-MS, together with the data obtained by GC-MS, might be needed. Moreover, it could be that several extraction
procedures might need to be jointly analysed and the data generated clustered together to attain a more meaningful representation of the metabolome investigated.

Metabolic modules could result from the action of genes that are concertedly regulated or proteins that share some vicinal localization. In addition, many genes involved in plant secondary metabolism are known to code for enzymes capable of producing multiple products from the same precursor molecule. There are documented cases in which enzymes can accept several precursors to form multiple products, such as the lemon basil (Ocimum basilicum) α-zingiberene synthase enzyme that catalyses the formation of more than 15 different sesquiterpenes from the precursor farnesyl diphosphate and 13 different monoterpenes from geranyl diphosphate (Davidovich-Rikanati et al., 2008). Nevertheless, metabolic modules could result from enzyme promiscuity towards substrates, such as the many acetyl esters that can be derived from the action of one acetyltransferase activity or methyl ethers derived from the action of one O-methyltransferase activity, each displaying broad substrate specificities (Lewinsohn and Gijzen, 2008). Mining metabolic modules instead of individual metabolites might be a more meaningful approach to study plant biochemistry when coupled to other proteomics and transcriptomics methodologies. This, in turn, will help to understand the complex interactions that regulate the formation of plant secondary metabolites. It is expected that metabolic module analyses will turn out to be a very useful tool that will direct researchers to the discovery of major metabolic junctions or key regulatory mechanisms in plant biochemistry.

Chemical analysis of plant metabolomes generates a massive amount of difficult to use data, so novel methods for deciphering and integrating these data into a coherent meaningful output are required. As shown in the featured paper (Xie et al., 2009), the discovery and analysis of metabolic modules can serve as a useful tool for predicting the networking of plant secondary metabolism. This strategy for digitized processing of metabolomic data opens the way for the utilization of metabolomics in plant science as an available tool. Metabolite modules could be useful in promoting the elucidation of unknown structures, particularly of minor, routinely ignored compounds, as well as for biosynthetic investigations and cultivar, tissue or plant comparisons. A strategy based on the combination of mining metabolite modules coupled to the proteomics and transcriptomics methodologies already available could help to understand the complex networking and mechanisms used by plants for the biosynthesis of secondary metabolism and its regulation. It is expected that the elucidation of metabolic modules will be a useful point of reference for more educated comparisons of the transcription levels of key genes and key enzymatic activities as reflected in the composition and content of the metabolites produced in a given biological system (Xie et al., 2008). The applications for clustering coupled to principal component analysis (PCA) extend far beyond biological fields, and is now gaining acceptance in other disciplines as well. Recently, the use of such tools is gaining interest among scholars attempting to rationalize complex non-biological systems such as the stock exchange market (Shapira et al., 2008). The use of such tools is referred to by economists as bio-inspired systems level analysis.

Co-regulated metabolites as well as co-ordinated gene expression, silent and non-silent enzymatic activities, and enzyme promiscuity may be universal features of plant secondary metabolism. The integration of methodologies possessing powerful prediction tools from different disciplines and scientific fields resulting in a broader perspective will be required in order fully to decipher plant metabolism. Thus, our goal of gaining a better understanding of complex biological and non-biological systems consequently pushes forward the development of novel multidisciplinary research methodologies capable of accepting and addressing this complexity.

References