L03. Challenges in handling complex metabolomic data

Michal Daszykowski
Department of Chemometrics, Institute of Chemistry, The University of Silesia, Katowice, Poland

Over the past few years there has been a growing interest in understanding complex biological processes. Owing to the development of powerful analytical instruments, now, this objective seems to be within one's grasp. This is supported by many examples of successfully conducted omics studies described in the literature. In metabolomic studies, one is mainly interested in discovering a unique set of target compounds, often called biomarkers, indicating an on-going process, e.g., development of a disease. Usually non-targeted approaches are firstly used in order to collect as much chemical data as possible about the studied system. At present, hyphenated chromatographic techniques such as liquid chromatography coupled with mass spectrometry (LC-MS) are frequently adopted for metabolomic profiling [1, 2]. They can provide easily a huge amount of chemical information about each sample. The sample, analyzed by the LC-MS technique, can be represented by a two-dimensional chromatographic fingerprint (abundance measured at a given retention time and mass-to-charge ratio). Accordingly, the data collected are multi-dimensional and can contain many irrelevant features and some signals of questionable quality. These issues pose a real problem in data handling (exploration and modeling) and therefore, drawing general conclusions about the experiment and/or testing a hypothesis could be very sophisticated or even impossible. Similar problems related to data explosion and handling were also encountered many years ago, when near infrared (NIR) data appeared and stimulated an extensive chemometric research.

An appropriate use of well-suited chemometric methods can greatly facilitate the analysis of complex proteomic/metabolomic data. Specifically, with chemometric methods an enhancement of signal-to-noise quality, alignment of peak shifts observed among samples due to difficulties in maintaining the same conditions during analysis, can be achieved [3].

References:

1. G. Theodoridis, H.G. Gika, I.D. Wilson, *TrAC*, **27** (2008) 251-260
2. J.C. Lindon, J.K. Nicholson, *TrAC*, **27** (2008) 194-204
3. M. Daszykowski, B. Walczak, *TrAC*, **25** (2006) 1081-1096