Journal of Chromatography B, 910 (2012) 31–45

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

Trends in data processing of comprehensive two-dimensional chromatography: State of the art☆

João T.V. Matos, Regina M.B.O. Duarte, Armando C. Duarte*

CESAM & Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal

ARTICLE INFO

Article history:
Received 26 October 2011
Accepted 29 June 2012
Available online 7 July 2012

Keywords:
Comprehensive two-dimensional chromatography
Multidimensional separations
Chemometrics
Peak detection
Resolution
Data treatment

ABSTRACT

The operation of advanced chromatographic systems, namely comprehensive two-dimensional (2D) chromatography coupled to multidimensional detectors, allows achieving a great deal of data that need special care to be processed in order to characterize and quantify as much as possible the analytes under study. The aim of this review is to identify the main trends, research needs and gaps on the techniques for data processing of multidimensional data sets obtained from comprehensive 2D chromatography. The following topics have been identified as the most promising for new developments in the near future: data acquisition and handling, peak detection and quantification, measurement of overlapping of 2D peaks, and data analysis software for 2D chromatography. The rational supporting most of the data processing techniques is based on the generalization of one-dimensional (1D) chromatography although algorithms, such as the inverted watershed algorithm, use the 2D chromatographic data as such. However, for processing more complex N-way data there is a need for using more sophisticated techniques. Apart from using other concepts from 1D chromatography, which have not been tested for 2D chromatography, there is still room for new improvements and developments in algorithms and software for dealing with 2D comprehensive chromatographic data.

Contents

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, Edited by Pedro Araujo and Bjørn Grung.

* Corresponding author. Tel.: +351 234 370200; fax: +351 234 370084.
E-mail address: aduarte@ua.pt (A.C. Duarte).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.06.039

1. Introduction

The development of several one-dimensional (1D) separation techniques, such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE), led to the generalized idea by the end of the twentieth century, that these techniques could be just finely tuned in order to solve all the practical problems in Analytical Chemistry [1]. However, the need for analysis of increasingly complex samples with a large number of compounds, highlighted the limitations of such techniques, and prompted the development of technologies with a much higher separation capacity in order they could take full advantage of coupling them to advanced detection systems, such as mass spectrometry and nuclear magnetic resonance spectroscopy.

The need for improving the analytical figures of merit associated to the research explosion in proteomics and metabolomics, and the ever increasing requirements for adequate identification and quantification of proteins, glycoproteins and metabolite products has prompted a need to push separation techniques to their limits. Furthermore, even when 1D chromatography could produce acceptable results, they do not have the separation power to deal with complex samples, and their use in such cases would mean spending a lot of time for analysis [1,2]. The obvious response to this lack of separation power of 1D techniques is the development of multidimensional chromatographic systems using two or more independent separation mechanisms.

Multidimensional separation can be understood as a separation system capable of discriminating the components from a mixture, using different separation mechanisms which are connected but do not interact among themselves, that is, they should be completely independent from each other. There are two modes of operation of multidimensional chromatography: heart-cutting and comprehensive. In the heart-cutting mode, only some selected fractions are transferred from the first into the second separation system, and the results become two separate 1D data sets. On the other hand, a separation is comprehensive when the whole sample is subjected to two different separation mechanisms, the separation (resolution) obtained in the first dimension is essentially maintained, and the chromatogram obtained is representative of the entire sample (after pre-treatment), which requires that either no sample goes to waste (everything passes through the detector) or a sufficient number of second-dimension chromatograms are recorded very frequently across the width of a first-dimension peak [3].

As reviewed by Phillips and Beens [4], the comprehensive multidimensional chromatography became more relevant after the development of comprehensive 2D gas chromatography (GC × GC), more than a decade after the development of a first comprehensive 2D liquid chromatography (LC × LC) by Erni and Frei [5]. Although the data sets resulting from GC × GC have received more attention than those resulting from any other comprehensive 2D chromatographic technique, they are formally equivalent and most of the work developed for GC × GC can be applied with small modifications to other chromatographic combinations, such as LC × LC, LC × GC, and LC × CE.

The data collected from advanced chromatographic systems designed for the analysis of complex samples contain huge amounts of information that need complex processing algorithms in order to take advantage of such powerful analytical systems. For instance, analysis of a sample with n replicates in a 2D chromatographic system coupled to a multichannel detector, such as a diode array detector (DAD) or a mass spectrometer (MS) can produce a so-called four-way data set. This terminology can be better understood through a schematic representation of different data sets derived from different types of analyses and orders of instruments, as shown in Fig. 1. The interpretation of these data sets is based on the order of the analytical signal, which was thoroughly discussed in 1994 by Booksh and Kowalski [6]. Fig. 1A represents both a first-order tensor (i.e., a vector) data that changes over the time of the first-dimension and a two-order tensor (i.e., a matrix) of data. While the former can be obtained by a first-order instrument, such as 1D chromatography system coupled to a single channel detector, the two-order tensor of data is derived from a second-order instrument, which is defined as an instrument capable of generating a data set that also changes over time. A 2D chromatographic system coupled to a single channel detector or 1D hyphenated chromatographic techniques (e.g. GC/MS or MS/MS) are good examples of such second-order instruments. It should be mentioned that the first-order tensor of data can also be produced by discarding information from a second-order tensor acquired in a second-order instrument. This is usually performed to build the first-order profile of the data set. The order of the data produced can be further increased if one combines an additional first-order instrument to the second-order instrument. This is depicted in Fig. 1B, which represents a third-order tensor data acquired in a 2D chromatography system coupled to a multichannel detector (third-order instrument). Finally, a four-way data, represented in Fig. 1C, is classified as a hypercube of data obtained by stacking the data from n replicates acquired in a third-order instrument.

The aim of this review is to discuss the state of the art in data processing for multidimensional data sets obtained in different types of 2D chromatography, from the pre-treatment until the quantification of the identified chromatographic peaks. The discussion will lead to the identification of the main trends in data processing of comprehensive 2D chromatography and it will pinpoint the gaps and research needs that should be tackled in this field.

2. Data acquisition and handling in comprehensive 2D chromatography

The massive amount of data generated from the current high-resolution analytical instrumentation requires the use of computerized assistance for data processing and transformation. The 2D chromatography is no exception, and the use of informatics tools has become essential for transforming the raw analytical data into fit for purpose information. The 2D chromatography produces a considerable amount of data in a relatively short time when applied to the separation of complex mixtures. Such an enhancement in performance provides an order-of-magnitude increase in peak capacity, when compared to 1D chromatography.

Handling of 2D chromatography data is a challenging task in Analytical Chemistry. The acquisition of data in real time from 2D chromatography coupled to detectors, such as DAD or MS, generates huge data files, that can reach more than 10 million data points which may lead to considerable problems in storage and processing [2,7]. The greatest challenges lies in producing automatic tools capable of processing and converting the data matrix under useful forms without losing control on the analysis of samples for obtaining raw data, and transformation of data into useful chemical