Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1096 (2005) 40–49
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

The chemometric resolution and quantification of overlapped peaks form comprehensive two-dimensional liquid chromatography

Carlos G. Fraga *, Cynthia A. Corley

Department of Chemistry, 2355 Fairchild Dr., Suite 2N225, United States Air Force Academy, CO 80840-6230, USA

Available online 21 April 2005

Abstract

The chemometric resolution and quantification of overlapped peaks from comprehensive two-dimensional (2D) liquid chromatography (LC × LC) data are demonstrated. The LC × LC data is produced from an in-house LC × LC analyzer that couples an anion-exchange column via a multi-port valve with a reversed-phase column connected to a UV absorbance detector. Three test mixtures, each containing a target analyte, are subjected to partial LC × LC separations to simulate likely cases of signal overlap. The resulting unresolved target-analyte signals are then analyzed by the standard-addition method and two chemometric methods. The LC × LC analyses of a test mixture and its corresponding standard-addition mixture results in two data matrices, one for each mixture. The stacking of these two data matrices produces a data structure that can then be analyzed by trilinear chemometric methods. One method, the generalized rank annihilation method (GRAM), uses a non-iterative eigenvalue-based approach to mathematically resolve overlapped trilinear signals. The other method, parallel factor analysis (PARAFAC), uses an iterative approach to resolve trilinear signals by the optimization of initial estimates using alternating least squares and signal constraints. In this paper, GRAM followed by PARAFAC analysis is shown to produce better qualitative and quantitative results than using each method separately. For instance, for all three test mixtures, the GRAM-PARAFAC approach improved quantitative accuracy by at least a factor of 4 and quantitative precision by more than 2 when compared to GRAM alone. This paper also introduces a new means of correcting run-to-run retention time shifts in comprehensive 2D chromatographic data.
Published by Elsevier B.V.

Keywords: Liquid chromatography, two-dimensional; Complex mixtures; GRAM; PARAFAC; Three-way data; Trilinear

1. Introduction

Comprehensive two-dimensional (2D) liquid chromatography (LC × LC) is well suited for the separation and analysis of semi and non-volatile compounds in complex mixtures. Like comprehensive 2D gas chromatography (GC × GC), LC × LC's enhanced peak capacity provides a greater separation space to resolve chemical components. In the 1990s, several papers used the enhanced separation power of LC × LC to successfully analyze complex mixtures that were primarily biological [1–9]. Recent LC × LC papers have expanded the use of LC × LC to other sources of complex mixtures such as food products [10–17]. However, unlike GC × GC the chemometric analysis of LC × LC data has not been heavily pursued, even though it would benefit LC × LC as it has GC × GC. For instance, several papers have successfully applied chemometric methods to GC × GC data to reveal hidden chemical information [18–25]. The goal of the chemometric methods discussed in this paper is to mathematically reveal and quantify overlapped signals, which inevitably occur in very complex mixtures. The generalized rank annihilation method (GRAM) is one chemometric method discussed in this paper. It has been successful at resolving and quantifying severely overlapped GC × GC peaks [18–20]. It has also been applied to overlapped signals from other hyphenated chromatographic methods producing structured 2D data [26–32]. Another chemometric applied to GC × GC data is known as PARAFAC. It has been used to successfully resolve overlapped signals in data from GC × GC—time-of-flight mass spectrometry (GC × GC—TOFMS) [24,25]. Reference [25] demonstrated that coupling PARAFAC with trilinear decomposition (TLD), which is a method similar to GRAM, gave better results than TLD alone. Better signal resolution

* Corresponding author. Tel.: +1 719 333 6044; fax: +1 719 333 2947.
E-mail address: carlos.fraga@usafa.af.mil (C.G. Fraga).

0021-9673/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.chroma.2005.03.118

with PARAFAC agrees with the findings of other authors [33,34].

In this paper, chemometric resolution and quantification of unresolved LC × LC data is applied for the first time. The LC × LC data was obtained using an in-house LC × LC analyzer that couples an ion-exchange (IC) column and a reversed phase (RP) column with a single-wavelength UV absorbance detector. The work described in this paper is similar to the first application of GRAM to GC × GC flame-ionization data [18]. However, PARAFAC is used to improve upon the GRAM results. Better signal resolution and quantitative results are gained when GRAM and PARAFAC are coupled as opposed to individually. In addition, a new method for correcting run-to-run retention time shifts in 2D data is introduced.

1.1. Chemometric methods

GRAM is a non-iterative eigenvalue-based method used to resolve and quantify the bilinear signals of compounds that vary in concentration between two data matrices. LC × LC signals for the most part are bilinear. That is, an LC × LC signal for an analyte can be mathematically represented by the product of two vectors, each representing that analyte's signal from one HPLC column. One of the two data matrices subjected to GRAM analysis is called the sample data matrix. It contains the signal for the target analyte from the LC × LC analysis of a test mixture. The other data matrix is called the standard data matrix. It contains the signal for the target analyte from the LC × LC analysis of a standard-addition mixture. That mixture is made by spiking a known amount of target analyte into a portion of the test mixture. Different versions of the GRAM algorithm exist [35–38]. The GRAM algorithm used is based on the one from Wilson et al. [38]. The only input required for GRAM analysis is an estimate of the number of different component signals present in the data matrices. Several methods exist for estimating the number of component signals [39–43]. The data requirements for the GRAM analysis of LC × LC data are identical to those listed for GC × GC [18]. The key requirement for GRAM analysis is that the two data matrices (sample and standard) must be trilinear. In other words, when the data matrices are stacked to make three-way data (i.e., a cube of data), the bilinear signals in common between the data matrices must match perfectly excluding signal intensity. For two stacked LC × LC data matrices, three vectors represent the trilinear signal of an analyte. One vector is the normalized analyte signal for one column and another is the normalized analyte signal for the other column. The third vector represents the relative amount of the analyte in the two data matrices.

PARAFAC is an iterative three-way method that resolves overlapped signals through the optimization of initial estimates using alternating least squares (ALS) and signal constraints. Non-negative and uni-modality are the standard signal constrains. PARAFAC as a chemometric method is well documented in literature [34,44]. The initial estimates used by PARAFAC can come from an eigenvalue-based method (e.g., GRAM), random values, random orthogonalized values, and singular values. PARAFAC, like GRAM, is a trilinear-based method that calculates the three vectors that represent the trilinear signal of an analyte in two stacked LC × LC data matrices. Once each vector is obtained, the resolved signal for that analyte can be reconstructed using the three vectors. In this paper, the target analyte's concentration in a given test mixture is calculated from the vector representing the relative amount of the analyte in the test mixture and its standard-addition mixture. The target analyte's actual concentration in the test mixture is calculated using the known concentration of the spiked analyte in the standard-addition mixture.

1.2. Retention time alignment

Run-to-run retention time shifts are a main cause of non-trilinearity in three-way chromatographic data [30,31,45]. Hence, retention-time shifts need to be corrected prior to chemometric analysis by GRAM or PARAFAC. Rank alignment has been successful at correcting retention-time shifts in three-way data [19,28,31,46]. It is an iterative 2D technique that shifts the bilinear signals in one matrix relative to the other until a minimum in the percent residual variance is reached [47]. At that point, the bilinear signals in common between the two stacked data matrices are aligned. Unfortunately in some cases rank alignment did not correct the run-to-run retention time shifts of the LC × LC data. Therefore, an alternative alignment method was developed. This new retention time alignment method, however, is beyond the scope of this paper and will not be discussed in depth. It involves incrementally applying a time-shift correction to the LC × LC data followed by GRAM and then PARAFAC analysis. The right shift provides the smallest sum of squares or best data fit between the PARAFAC data and the raw data. Simulations have shown that for trilinear methods the fit between raw data and processed data improves as the degree of retention-time shift decreases [45].

2. Experimental

2.1. Test mixtures

Three aqueous test mixtures, A–C, were prepared. The solutes were *p*-chlorobenzoic acid (99% Aldrich Chemical Co., Milwaukee, WI, USA), benzoic acid (99.5% Aldrich), uracil (99+% Acros Organics, Morris Plains, NJ, USA), pyruvic acid (99+% Acros), maleic acid (99% Acros), fumaric acid (99% Aldrich), and phenyl phosphoric acid (98% Aldrich). The water used was purified by a Milli-Q system (Millipore Corp., Milford, MA). Mixture A contained *p*-chlorobenzoic acid (50.0 mg/mL or ppm) and benzoic acid (50.0 ppm). Mixture B contained uracil (5.00 ppm) and pyruvic acid (200.0 ppm). Mixture C contained fumaric acid