Journal of Chromatography A, 1208 (2008) 116–125

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Computer-assisted method development in liquid chromatography–mass spectrometry: New proposals

J. García-Lavandeira, B. Losada, J.A. Martínez-Pontevedra, M. Lores, R. Cela*

Institute of Research and Food Analysis, Analytical Chemistry Laboratory, University of Santiago de Compostela, E-15782 Santiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 23 June 2008
Received in revised form 18 August 2008
Accepted 20 August 2008
Available online 3 September 2008

Keywords:
Computer-assisted method development
LC–MS
Pareto optimality

ABSTRACT

A new approach is proposed for computer-assisted method development in LC–MS. The procedure consists of three stages. Firstly, an accurate retention model is developed for the peaks in the mixture to be separated by use of an iterative approach with isocratic priming data, which is calibrated and validated by means of a few gradient runs. Secondly, a specially developed LC–MS objective function, based on selectivity targets (the selectivity matrix), is calculated and used to evaluate the simulated chromatograms and drive the optimization process. Thirdly, the retention model and the selectivity matrix objective function are used with an evolutionary algorithm in which the concepts of constrained Pareto optimality are applied, to carry out the unattended optimization process. The system was applied to real data for a complex separation and compared with the results provided by a commercial tool for computer-assisted method development.

1. Introduction

Computer-assisted method development (CAMD) in HPLC has been a permanently open research field since the 1970s, with intermittent developments [1–3], and new proposals from both the academic [4–12] and industrial communities [13–18]. In general, most of these proposals have considered some form of resolution as the objective function, alone or frequently combined in time-weighted functions [19–21], often named chromatographic response functions (CRFs). Because the most commonly applied detection system in HPLC has traditionally been the UV–vis detector, in any of its forms, the main goal underlying resolution-based objective functions has been the attainment of a good enough resolution between all the peaks to be separated.

The steadily growing presence of mass spectrometry-based instruments (LC–MS) in research and routine analytical laboratories offers many advantages to the chromatographer but also some added challenges. Because the old separation programs and procedures can now be done more efficiently with LC–MS, the simple transference of elution programs from one instrument to another will waste the enhanced selectivity provided by the mass spectrometer, so is far from being the best solution. Consequently, the old procedures developed for less selected detectors should be adapted specifically to the MS detector and a new, more or less extensive, optimization process carried out. A similar situation appears when developing new separations to be used in routine analysis. The impressive selectivity provided by the MS detector often tempts the analyst not to be too worried by the chromatographic separation. Instead of transferring the separations from the conventional HPLC-UV apparatus in the appropriate way [22,23], some analysts try to inject the samples without even having a chromatographic column in the system, in an attempt to obtain maximum sample throughput. The best procedure however, would be to use the column separation power as far as required to complement the selectivity provided by the detector, while keeping the separation as short as possible [24]. To achieve this aim specific CAMD techniques in LC–MS are required helping the chromatographer achieve rapid access to the optimum separation conditions. Current CAMD commercial tools dealing with LC–MS separations focus on the critical aspect of recognizing the peaks of interest within the bulk of the sample matrix [25,26] and detecting the real number of peaks in the samples [27]. While these aspects are of utmost importance in screening experiments, in routine analysis, the number and nature of analytes is usually known and the goal of the chromatographer is to develop and apply the quickest and most efficient separation procedure. Here, we propose a new tool for LC–MS computer-assisted method development, which can be applied within the common framework of freely available standard computer-assisted techniques. An efficient practical chemometric tool, which runs unattended, has been developed

* Corresponding author.
E-mail address: rafael.cela@usc.es (R. Cela).

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.08.105

for chromatographers carrying out routine analyses with modern liquid chromatography, by considering the unique features of MS detection and by applying Pareto optimality associated with an evolutionary algorithm.

2. Theory

2.1. The retention model

The retention model provides a mathematical description of the chromatographic behavior of analytes departing from a small set of experimental runs, thus allowing simulation of chromatograms under different scenarios. It is an extremely critical step in CAMD because the reliability and accuracy of the derived simulations cannot be better than the quality of the retention model developed. Most commercial CAMD tools develop the retention model for reversed phase separations, based on the linear strength theory and other well-known chromatographic relationships [14–16,20,22,28,29] using data from a few scouting gradient runs, which limit the initial experimental effort. Alternatively, isocratic measurements can be used to feed the mathematical retention model [4,6–9,11,12,30], which demands a little more initial experimental effort, but provides somewhat better priming data precision, which is less affected by the instrument configuration and enables modeling of compounds that are not linearly retained. One important point is the definition of the interval of modifier percentages covered by the retention model. Both the scouting gradients and the isocratic priming data approaches have serious problems when this interval becomes large. In the first case, because the retention linearity cannot be granted. In the second, because frequent missing data arising from too retained peaks limits the data available. To avoid this, the approach presented here applies a three-stage recursive strategy that allows the extension of the modifier intervals from 5 to 95% by default and from 0 to 100% if required (e.g. ternary solvent systems). The strategy departs from the available isocratic priming data by adjusting exponential regression models to peak retention data. These initial peak retention models are generally somewhat less accurate than the models provided by the gradient scouting technique. In the second stage these models are calibrated against at least two gradients of different shape (linear, multilinear, curved or stepwise), starting and ending conditions. In this process, the retention of all the peaks in the calibration gradients are simulated using the initial retention models and compared with the experimental values provided. The differences in retention for each individual peak are used to adjust the parameters of the retention model to obtain the better possible adjustment to both the experimental isocratic and gradient data. For linearly retained peaks, this procedure provides usually accurate enough retention models, mostly equivalent to those generated from scouting gradients. However, for non-linearly retained peaks further adjustments are needed. The third calibration stage involves releasing the mathematical model restrictions (e.g. using Gaussian, hyperbolic or spline fittings for selected peaks) allowing even abnormal retention behavior modeling (e.g. u-shaped peaks in some polar embedded phases [31]). This third stage is carried out only for the selected peaks thus retaining the models achieved in the second stage for the remaining ones. Peaks to be submitted to this third calibration stage are selected by the chromatographer based on graphical tools showing the retention model against the experimental data both in isocratic and gradient runs. In this way, highly adaptive modeling is developed, which ensures the accurate prediction of retention under all elution conditions for the different analytes to be separated in the mixture.

2.2. The objective function

Any auto-optimizable system requires an objective function to be able to proceed to the optimum without the intervention of the chromatographer. Although the objective functions applied in CAMD are traditionally based on critical-pair resolution or on weighted sums or products of resolution derived functions and runtime, in LC–MS, good resolution between all pairs of peaks forming the mixture to be analyzed is generally not necessary and this has to be accounted for. In the present approach, a new objective function has been developed which has been named the "selectivity matrix". The selectivity matrix is calculated by departing from the mass spectra of the peaks to be separated under the following hypothesis, i.e. that to be able to qualify and quantify accurately a peak in the chromatogram we need a "clean" quantification ion plus two or three "clean" qualification fragments, where "clean" means that there is no interference from the spectra of other peaks or from the sample matrix background spectrum. Consequently, two or more peaks can overlap extensively provided the quantification and qualification peaks are sufficiently different (mutually not interfering) and intense enough to be measured accurately after the background is subtracted. In other words, to be able to quantify a peak it is necessary that signals for co-eluting peaks in the peak's quantification and qualification slots are null or extremely small.

The selectivity matrix is the arrangement of resolution constraints for any peak in the mixture in terms of the allowable spectral interference from other peaks. If two peaks exhibit different "clean" quantification and qualification fragments, those peaks can overlap fully in the final optimized separation so the associated value in the selectivity matrix for those peaks is zero. On the contrary, for peaks showing more or less critical spectral selectivity conflicts, the resolution values imposed by the selectivity matrix must be large (e.g. 1.5–2.0). Thus, this process is equivalent for weighing each peak in terms of the resolution needed for that peak being the weighting factors defined by the spectral interferences.

The practical method of calculating the selectivity matrix consists of comparing the mass spectrum of each peak with the sum of all mass spectra (except that being compared) in the mixture. The graph in Fig. 1 corresponding to a simple mixture of four peaks exemplifies the calculation process. Here, for peak number 1, the four most intense fragments were selected as the quantification and qualification fragments. In comparison with the spectral sum of the remaining peaks and background, there is only slight interference from other signals in the quantification fragment and qualification fragments 1 and 3. On the contrary, there is strong interference in qualification fragment 2. There are now two possibilities. The first is re-selection of the qualification fragments to avoid the interference in Q2 if there are some alternative fragments of sufficient intensity. The second possibility is to force separation of peaks that may be responsible for the interference in Q2 (in that case, peak number 3 in the mixture). If the first option is feasible, the row corresponding to peak 1 in the selectivity matrix will be formed by zeros, meaning that this peak may overlap with any other in the chromatogram. On the contrary, the situation illustrated in Fig. 1 will provide the following line in the selectivity matrix for peak 1:

	Peak 2	Peak 3	Peak 4
Peak 1	0.0	1.5	0.0
.....	.....	.....	.....

and thus peaks 1 and 3 should be separated to allow accurate identification and quantification. It is important to appreciate that selectivity requirements for a pair of peaks are not necessarily sym-