Journal of Chromatography A, 1409 (2015) 241–250

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Sum of ranking differences to rank stationary phases used in packed column supercritical fluid chromatography

Caroline West[a,*], Maria A. Khalikova[b], Eric Lesellier[a], Károly Héberger[c]

[a] Univ Orleans, CNRS, Institut de Chimie Organique et Analytique (ICOA), UMR 7311, B.P. 6759, rue de Chartres, F-45067 cedex 2 Orléans , France
[b] Charles University in Prague, Faculty of Pharmacy in Hradec Králové, Department of Analytical Chemistry, Heyrovského 1203, 500 05 Hradec Králové, Czech Republic
[c] Plasma Chemistry Group, Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok krt. 2, H-1117 Budapest XI., Hungary

ARTICLE INFO

Article history:
Received 5 June 2015
Received in revised form 16 July 2015
Accepted 16 July 2015
Available online 20 July 2015

Keywords:
Column classification
Method development
Stationary phases
Selectivity
Supercritical fluid chromatography

ABSTRACT

The identification of a suitable stationary phase in supercritical fluid chromatography (SFC) is a major source of difficulty for those with little experience in this technique. Several protocols have been suggested for column classification in high-performance liquid chromatography (HPLC), gas chromatography (GC), and SFC. However, none of the proposed classification schemes received general acceptance. A fair way to compare columns was proposed with the sum of ranking differences (SRD). In this project, we used the retention data obtained for 86 test compounds with varied polarity and structure, analyzed on 71 different stationary phases encompassing the full range in polarity of commercial packed columns currently available to the SFC chromatographer, with a single set of mobile phase and operating conditions (carbon dioxide–methanol mobile phase, 25 °C, 150 bar outlet pressure, 3 ml/min). First, a reference column was selected and the 70 remaining columns were ranked based on this reference column and the retention data obtained on the 86 analytes. As these analytes previously served for the calculation of linear solvation energy relationships (LSER) on the 71 columns, SRD ranks were compared to LSER methodology. Finally, an external comparison based on the analysis of 10 other analytes (UV filters) related the observed selectivity to SRD ranking. Comparison of elution orders of the UV filters to the SRD rankings is highly supportive of the adequacy of SRD methodology to select similar and dissimilar columns.

1. Introduction

The use of chromatographic mobile phases containing carbon dioxide as a major component, as in supercritical fluid chromatography (SFC), allows to obtain rapid separations with high efficiency on packed columns, which could favor the replacement of numerous HPLC methods by SFC ones. In SFC, the same carbon dioxide-based mobile phase can be used with both polar and nonpolar stationary phases. This is an advantage, but also a source of difficulty for the inexperienced chromatographer willing to develop a new SFC method. Indeed, the selection of an appropriate column from the continuously expanding group of potentially suitable ones is becoming a complex task. Consequently, the need for a classification of stationary phases in SFC appeared.

In the past years, characterization of stationary phases employed in SFC was achieved through the use of linear solvation energy relationships (LSER) using Abraham descriptors, also known as the solvation parameter model [1–4]. LSER methodology is based on the analysis of a large group of analytes with structural diversity and known features to calculate quantitative structure-retention relationships relating retention data (log k) to molecular descriptors and results in "solvation vectors" defined by five coordinates related to five types of interactions (van der Waals interactions and hydrogen bonding). Using this method, more than 70 stationary phases encompassing the full range of polarities available to SFC chromatographers were previously characterized with constant mobile phase and operating conditions [5–11]. LSER results were also used to draw a classification map (spider diagram), designed as a projection of the solvation vectors onto a two-dimensional plane. While this map is useful in a general observation of the selectivity space available to SFC, some stationary phases appear to be highly clustered (especially polar stationary phases). As a complement, it would be useful to have a tool to measure the selectivity

* Corresponding author. Tel.: +33 238 494 778; fax: +33 238 417 281.
E-mail address: caroline.west@univ-orleans.fr (C. West).

http://dx.doi.org/10.1016/j.chroma.2015.07.071
0021-9673/

similarity between two columns, especially when one column needs to be replaced by another. This case may occur in two different instances: (i) when the stationary phase currently in use is not satisfying for reasons of retention or selectivity or (ii) when the stationary phase currently in use is providing adequate retention and selectivity but poor peak shapes or efficiency. In the first case, a stationary phase with different (or orthogonal) selectivity must be identified while in the second case, a stationary phase with close selectivity but possible improved efficiency must be identified.

Several methods were previously proposed to quantify the selectivity similarity between stationary phases employed in HPLC, GC, or SFC.

A simple quantification of the differences between two chromatographic systems is available with the Pearson's correlation coefficient r, calculated between the retention factors measured on the two systems. To assess selectivity differences, Neue et al. [12] derived the s^2 coefficient, a distance measure defined as:

$$s^2 = 1 - r^2 \quad (1)$$

With this criterion, small s^2 values indicate close selectivity, and vice versa. However, neither the r nor the s^2 values were ever implemented as a ranking tool.

Calculation of Euclidean distance was employed in several works related to column classification, under varied names (F, F_s, or CDF) and all based on the same principle [13–15]. Basically, it is related to the initial calculation of coordinates for each column in a selectivity space, issued from chromatographic measurements (such as methylene selectivity, shape selectivity, silanol activity, etc.). The distance between two columns defined by these coordinates is then calculated, sometimes with weighting coefficients, based on the relative influence of each type of interaction on the global selectivity. A limit value may be provided indicating when two columns should be considered as equivalent or different. For instance, Németh et al. indicate maximum and bottom values to identify similar or orthogonal columns respectively [16].

In the past, we have also used the calculation of the angle between the solvation vectors [17–19], which is related to the distance criterion described above in a sinusoidal fashion.

The methodology of sum of ranking differences (SRD) [20–22] detailed in the following section is a possible alternative to these methods. SRD was previously applied to solve different problems:

(i) selection of the best polarity measure (indicator) for small organic molecules [23];
(ii) comparison of PLS models in near-infrared spectroscopy [24];
(iii) testing panel consistency in examination of organoleptic properties of foods [25];
(iv) comparing performance indicators for model goodness [26];
(v) ranking of QSAR models and splits [27];
(vi) testing the performance of pure spectrum resolution (Raman mapping) [28];
(vii) selection of best quantitative structure–activity relationship (QSAR) models [29];
(viii) classification of chromatograms based on separation capacity and peak symmetry [30]; and
(ix) comparison of HPLC columns [20], just to name a few.

In the present paper, we thus investigated the applicability of SRD methodology for the measurement of selectivity similarity between stationary phases employed in SFC. To demonstrate the potential of this method for real cases, a sample application (analysis of UV filters in cosmetic creams) is first developed on a column selected as reference one, then the method is transferred to other columns designated by SRD ranking as providing close or distant selectivity to the reference one.

2. Sum of ranking differences: procedure

The procedure of SRD is fairly simple to understand and will be explained here with relation to the present topic: the comparison of chromatographic systems.

For all columns in the set, a common set of retention data must be available. In our case, we had retention factors (k) for 86 analytes that could be eluted from 71 columns in identical operating conditions. The data are thus arranged in a matrix form with the analytes enumerated in the rows and the stationary phases arranged in the columns.

Then a column must be selected as a reference. It can be a real column: for instance, when one has started developing a chromatographic method with one particular column that needs to be replaced by another one. But it could also be a "virtual" column, as would be a hypothetical average column calculated from the set of all available (and previously characterized) columns. This would naturally depend on the initial selection of columns, and would change every time a new column would be introduced in the column set.

For each column that must be ranked, including the reference column, the retention data are then organized in order of increasing retention. The analytes are then ranked: the analyte with the lowest retention on column C_i ranks 1_{Ci}, the second eluted ranks 2_{Ci}, etc. The analyte last eluted ranks 86_{Ci}. When two analytes have the same retention (co-eluted species), the program "without ties" calculates ranks using the top down (or left to right) principle, meaning that the analytes are ranked in order of appearance. This is an erroneous ranking but causes no issues when the proportion of such cases remains low. When too many co-elution cases occur, it is preferable to use the program "with ties" to give equal ranks to equal retention (further details can be found in [22]).

The data table of analyte ranks is then organized according to increasing analyte rank (elution order) on the reference column. For each analyte and each column, the analyte ranks are compared. For instance, the analyte that eluted first on the reference column (analyte rank 1_{Cref}) may have been the third eluted on column C_i (analyte rank 3_{Ci}). The absolute value of the difference in their analyte rank is thus 2. The difference (absolute value) is calculated for each analyte on column C_i, and in the end these differences are summed up to obtain the sum of ranking differences (SRD score). The column with the smallest SRD score is thus the closest to the reference column in terms of selectivity, because elution orders are most similar on these columns. On the opposite, the column with the largest SRD score value is the most dissimilar to the reference column and should provide significantly different selectivity.

Finally, an SRD rank can be attributed to each column, depending on how close it is to the reference column. In our case, we had 71 columns in total, thus the SRD ranks range from 1 (closest to the reference column selected) to 70 (farthest to the reference column).

Validation of the SRD method can be carried out with simulated random vectors for comparison, which is a kind of permutation test: if the number of columns is smaller than 14, the theoretical SRD distributions of random numbers are applied, otherwise the distribution is approximated by a Gaussian fitted curve. This procedure is called "comparison of ranks by random numbers" (CRRN) [21]. Naturally, leave many out (e.g. sevenfold cross-validation) can also be carried out: a portion has repeatedly been left out and the ranking is done on the remaining number of analytes. In this way, uncertainty of the SRD values can be determined.

It can be noted that normalization of the initial retention data was not required in our case, because only elution order is taken into account, and this does not depend on column dimensions that were not always identical to all columns in the set. On the other hand, the primer SRD data was scaled between 0 and 100 to be comparable. This is done to facilitate comparisons with other studies.