Assessment of column selection systems using Partial Least Squares☆

Petar Žuvela [a], J. Jay Liu [a,*], Alina Plenis [b], Tomasz Bączek [b]

[a] Department of Chemical Engineering, Pukyong National University, 365 Sinseon-ro, 608-739 Busan, South Korea
[b] Department of Pharmaceutical Chemistry, Medical University of Gdańsk, Hallera 107, 80-416 Gdańsk, Poland

ARTICLE INFO

ABSTRACT

Column selection systems based on calculation of a scalar measure based on Euclidean distance between chromatographic columns, suffer from the same issue. For diverse values of their parameters, identical or near-identical values can be calculated. Proper use of chemometric methods can not only provide a remedy, but also reveal underlying correlation between them. In this work, parameters of a well-established column selection system (CSS) developed at Katholieke Universiteit Leuven (KUL CSS) have been directly correlated to parameters of selectivity (retention time, resolution, and peak/valley ratio) toward pharmaceuticals, by employing Partial Least Squares (PLS). Two case studies were evaluated, separation of alfuzosin, lamotrigine, and their impurities, respectively. Within them, comprehensive correlation structure was revealed, which was thoroughly interpreted, confirming a causal relationship between KUL parameters and parameters of column performance. Furthermore, it was shown that the developed methodology can be applied to any distance-based column selection system.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

When it comes to separation in pharmaceutical analysis, RP-LC has become a de-facto standard [1]. Availability of more than 1000 commercial stationary phases [2,3], makes column selection a crucial process. Therefore, it is not surprising that numerous papers [4–22] describing various attempts to achieve this are being continuously published.

One of the earliest developed systems for column selection was the linear solvation energy relationship (LSER) model introduced by Abraham et al. [23,24]. In this approach chromatographic retention was considered an additive function of molecular interactions (i.e. solvatochromic parameters) allowing in turn: column selection. It was followed by the Tanaka system [25] based on six selectivity parameters for characterization of commercial packing column materials. After that, systems were based on ranking or sorting the columns according to a one-dimensional measure calculated from Euclidean distance between various selectivity parameters (Table 1).

In the Euerby system [26–28] that was the Column Difference Factor (CDF) [28], constructed out of six parameters of the Tanaka

system. In application of Snyder and Dolan's [29–33] hydrophobic-subtraction model, the measure for column ranking was the column selectivity function (F_s) [34] constructed out of five selectivity parameters [29], while in the KUL column selection system (CSS) introduced by Hoogmartens et al. [2,35–41] the measure was the F metric [38] based on four selectivity parameters.

In order to test the applicability of the KUL CSS, several studies were published. Its parameters were related to other CSSs, as well as selectivity studies of pharmaceuticals (further in text: column performance) according to the system suitability test (SST) as recommended by the European Pharmacopeia (Ph. Eur.) [42].

Thereby, Dragovic et al. [39] have related its parameters to parameters of Snyder and Dolan's hydrophobic-subtraction model, Haghedooren et al. [43] to parameters of the Euerby system, while there are several studies relating them to relating them to column performance parameters [37,44–46]. Separate multivariate analyses such as factor analysis (FA), principal component analysis (PCA), or hierarchical cluster analysis (HCA) were employed for that purpose.

Conclusions on existence [43–46] or non-existence [39] of correlation were based on visual observations of the score and loading plots, e.g. similar orientation of columns and parameters in distinct score, and loading plots, respectively. In fact, results presented in the study by Dragovic et al. [39] are contradictory. Based on comparison of results originating from separate PCA analyses, the authors concluded there is no correlation between the two systems. Their subsequent results show that both of the studied column

Table 1
Parameters of various column selection systems and their corresponding chromatographic properties.

Parameter	Description	Property
Euerby system [28]		
k_{PB}	Relative retention time of pentylbenzene	Amount of alkyl chains
α_{CH2}	Selectivity factor between amylbenzene and butylbenzene	Hydrophobicity
$\alpha_{C/P}$	Selectivity factor between caffeine and phenol	Hydrogen bonding capacity
$\alpha_{T/O}$	Selectivity factor between triphenylene and o-terphenyl	Steric selectivity
$\alpha_{B/P \text{ pH } 2.7}$	Selectivity factor between benzylamine and phenol at pH 2.7	Ion exchange capacity at pH < 3
$\alpha_{B/P \text{ pH } 7.6}$	Selectivity factor between benzylamine and phenol at pH 7.6	Ion exchange capacity at pH > 7
Snyder and Dolan's system [33]		
H	Column parameter H	Hydrophobicity
S^*	Column parameter S^*	Steric resistance
A	Column parameter A	Hydrogen-bond acidity
B	Column parameter B	Hydrogen-bond basicity
C	Column parameter C	Cation-exchange activity
KUL column selection system [37]		
k'_{amb}	Retention factor of amylbenzene	Hydrophobicity
$rk'_{ba/ph \text{ pH } 2.7}$	Relative retention factor of benzylamine/phenol at pH 2.7	Silanol activity
$rk'_{tri/o\text{-}ter}$	Relative retention factor of triphenylene/o-terphenyl	Steric selectivity
$k'_{2,2'\text{-}d}$	The retention factor of 2,2'-dipyridyl	Silanol activity and metal impurities

selection systems exhibit similar probability of selecting an identical column in respect to the same reference.

Although PCA was applied in development of several column selection systems [25–28,35,40], it remains insufficient for correlating them between themselves, or with column performance. This is due to the fact that in PCA latent variables are constructed in the direction of maximum variance [47] of a single set of analyzed data (i.e. variables of **X**-space). Moreover, distance-based column selection systems themselves all suffer from the same obvious drawback. For different values of source parameters, identical or near-identical values of a scalar measure can be calculated. Recently city-block distance [48,49] was shown to be an interesting alternative. However, it does not entirely resolve this issue.

Therefore, an entirely different approach was devised for tackling this problem, by employing Partial Least Squares (PLS). In PLS, latent variables are constructed not only in the direction of maximum variance of variables of **X**-space, but in the direction of maximum correlation with latent variables of **Y**-space [50]. PLS also provides useful tools such as score plots, loading plots, contribution plots, and **w*****c** plots for analyzing **X**- and **Y**-space.

In turn, with such an approach, PLS was used for exploratory analysis of causal relationships between the two sets of data, instead of prediction. It not only revealed important drawbacks of a well-established CSS, such as ambiguity of source parameters and inconsistencies in their physic-chemical interpretation, but also provided them remedy.

In order to support the methodology, two case studies were evaluated. Namely, correlation of KUL CSS parameters to retention parameters for separation of (1) alfuzosin hydrochloride and its two impurities on 36 columns [45], as well as separation of (2) lamotrigine and its seven impurities on 28 columns [46], respectively. KUL CSS was selected for both, since the principle of its calculation was identical to the other mentioned systems, while it required measurement of only four chromatographic parameters, making it more convenient for column selection.

2. Materials and methods

2.1. KUL column selection system procedure

Procedure of the KUL column selection system (CSS) begins by selecting a reference chromatographic column. In the two case studies, Inertsil and Hypersil BDS C18 columns were selected as references, as recommended by European Pharmacopeia (Ph. Eur.) [42], respectively. Then, KUL parameter values are measured for a selected reference, followed by the observed columns. Upon *autoscaling* the experimentally obtained values, the *F*-values, which represent similarity of column *i* to the reference, are calculated according to Eq. (1).

$$F = (k'_{amb,ref} - k'_{amb,i})^2 + (rk'_{ba/ph \text{ pH } 2.7,ref} - rk'_{ba/ph \text{ pH } 2.7,ref})^2$$
$$+ (rk'_{tri/ter,ref} - rk'_{tri/tre,i})^2 + (k'_{2,2'\text{-}d,ref} - k'_{2,2'\text{-}d,i}) \quad (1)$$

Four KUL parameters were obtained for analytes of both case studies using three isocratic chromatographic analyses in a defined order [A–B–C], as described in [35]. The KUL CSS datasets (Tables S1 and S2) consist out of four retention parameters as provided in Szulfer et al. [45,46].

2.2. Column performance based on the separation of Alfuzosin from its related compounds

Alfuzosin hydrochloride is a quinazoline derivative, an α_1-adrenoreceptor blocker used in treatment of urinary obstruction due to benign prostatic hyperplasia [51,52], as well as trial treatment of hypertension [53]. According to the Ph. Eur. [42], in a solution of alfuzosin hydrochloride five impurities can be found, namely, impurity A (N-[3-[(4-amino-6,7-dimethoxyquinazolin-2-yl)(methyl)amino]propyl]furan-2-carboxamide), impurity B (2-chloro-6,7-dimethoxyquinazolin-4-amine), impurity C ((RS)-N-[3-[(4-amino-6,7-dimethoxyquinazolin-2-yl)amino]propyl]-N-methyltetrahydrofuran-2-carboxamide), impurity D (N-(4-amino-6,7-dimethoxyquinazolin-2-yl)-N-methylpropane-1,3-diamine), and impurity E (N-[3-[(4-amino-6,7-dimethoxyquinazolin-2-yl)(methyl)amino]propyl]formamide), out of which A and D are the most important, and should be clearly separated by RP-LC. Therefore, column performance evaluated based on the separation of alfuzosin from its impurities was carried out as detailed in Ph. Eur. [42] on the RP-LC column studied.

Comprehensive description of the conditions and procedure of the first case study analysis can be found in [45]. The column performance dataset for case study 1 (Table S3) consists out of six chromatographic parameters: retention time of impurity D ($t_{R, imp. D}$), retention time ($t_{R, Alf.}$) and resolution ($R_{s, Alf.}$) of alfuzosin hydrochloride, retention time ($t_{R, imp. A}$), and resolution ($R_{s, imp. A}$) of impurity A, as well as the system suitability test (SST) parameter which represents peak-to-valley (p/v) ratio [54] between height