Tanaka-parameter based approach for chromatographic column selection

How to locate 'k' columns with the most dissimilar chromatographic selectivity from a selection of 'n' columns in a database

K. Kassam,[1*] D. Tsarev,[1] and M. Euerby[2]

[1] Advanced Chemistry Development, Inc. (ACD/Labs), 8 King Street East, Suite 107, Toronto, ON., M5C 1B5. Canada
[2] University of Strathclyde, Strathclyde Institute of Pharmacy and Biomedical Sciences, 161 Cathedral Street, Glasgow, G4 0RE, UK

Introduction

A common starting point in chromatographic method development is to screen different combinations of columns and mobile phases that are expected to give the desired retention and good peak shape. From a Quality by Design (QbD) perspective, selection of a set of columns with a wide range of chromatographic selectivity is imperative to investigate maximum design space.

However, how can the subset of columns be selected for screening to achieve maximum specificity with a minimum number of experiments?

With the plethora of columns available, each with different stationary phase ligands and different marketed characteristics, the selection of columns to be included in a screen can be an iterative trial and error process.

Principal Component Analysis (PCA) has typically been used to identify the most different columns in a group of columns as shown in Figure 1, even though it is not commonly used in practice.

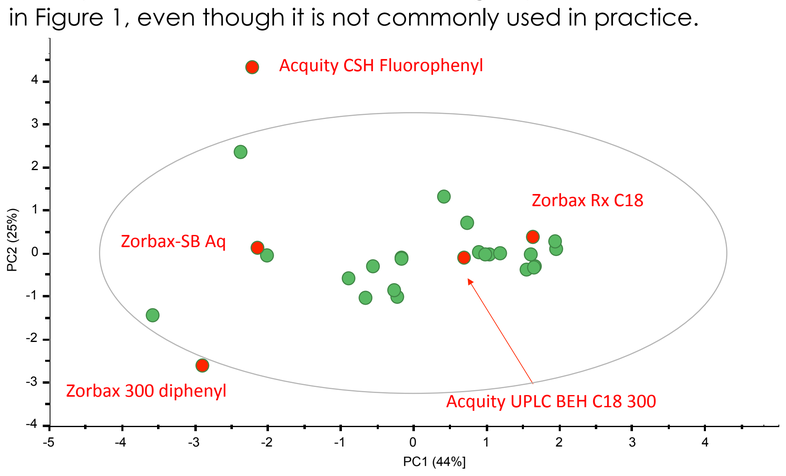


Figure 1 PCA score plot of the 27 columns investigated in this study as expressed by their Tanaka chromatographic parameters. (Red points in the plot represent dissimilar columns as selected by ACD/Labs' new tool for column comparison.)

The problem with the PCA approach is that it requires expensive chemometric software and each time a column is added or subtracted from the database a new PCA must be performed.

Tanaka chromatographic selectivity parameters[1] combined with weighed and scaled Euclidean distances between two multi-dimensional points, referred to as Column Distance Factor (CDF), have been successfully used to identify the degree of similarity between two columns.[2] This is useful for the identification of replacement/backup columns. It is less useful for the identification of the most diverse columns for screening.

In this poster, we will describe an approach to quickly and confidently select the most dissimilar columns from an inventory of available ones, based on Tanaka parameters. This approach has led to the development of a tool that is included in ACD/Labs' chromatography software expanding the use of the existing database of 340 columns.[3] The new tool allows selection of *n* columns of interest from the database and, subsequently, the identification of *k* most dissimilar columns among these with the most different chromatographic selectivity by maximizing the CDF for any pair of columns.

n = subset of columns selected by the user from within the database for comparison

k = desired number of dissimilar columns

The theory bit!

In order to calculate the distance, d_j, between a column *j* and the average column in the selection, the following approach has been used:

$$d_j = \sqrt{\sum_{i=1}^{6} z_{ij}^2}$$

Where:

i = variable 1 to 6 in the Tanaka test
j = column 1 to *k* to investigate
x_{ij} = variable *i* for column *j*
x_{avgi} = average value for variable *i*
s_i = standard deviation for variable *i*
$x_{refi} = x_{avgi}$
w_i = weight for variable
y_{ij} = normalised variable = $(x_{ij} - x_{refi})/s_i$
z_{ij} = normalised and weighted variable = $y_{ij} \cdot w_i$

Recommended Approach

The recommended approach is to select a group of modern and available columns (i.e., *n* columns) that are expected to provide good stability, adequate retention, and good peak shape for the analytes in question. Thereafter, a subset of columns should be selected (i.e., *k* columns, number depends on column oven size) that provide the largest differences in chromatographic selectivity for screening. This is usually done by selecting stationary phases with different ligands based on information from column manufacturers, but this is not always reliable. In fact a C18 column can be more similar to a phenyl column than another C18 column.

Tanaka Column Parameters

Tanaka parameters have been successfully used to develop a tool that characterizes equivalent and complementary columns. The tool is available free online or as part of ACD/Labs' commercial chromatography software.

Hydrophobicity—depends on carbon load and surface area. It reflects the hydrophobic retention and surface area of the column.

$k'_{PB} = k'_{\text{n-pentylbenzene}}$

Hydrophobic selectivity or methylene group selectivity of the column.

$\alpha_{CH_2} = k'_{\text{n-pentylbenzene}} / k'_{\text{n-butylbenzene}}$

Steric selectivity (or Shape Selectivity)—accounts for the capacity of a stationary phase to discriminate compounds of identical elemental composition but different three dimensional structure. The parameter depends on the length of the bonded chain, bonding density, and the type of bonding (i.e., monomeric or polymeric).

$\alpha_{(T/O)} = k'_{\text{triphenylene}} / k'_{\text{o-terphenyl}}$

Hydrogen bonding capacity (or Silanol capacity)—depends on the amount of available silanol groups present on the phase and reflects the H-bonding capacity of a column.

$\alpha_{(C/P)} = k'_{\text{caffeine}} / k'_{\text{phenol}}$

Ion exchange capacity at pH 7.6—at this pH most silanols possess a negative charge while benzylamine is positively charged, hence electrostatic attraction. This parameter reflects the total number of free silanol groups on the column.

$\alpha_{(B/P)}$ at pH 7.6 = $k'_{\text{benzylamine}} / k'_{\text{phenol}}$ at pH 7.6

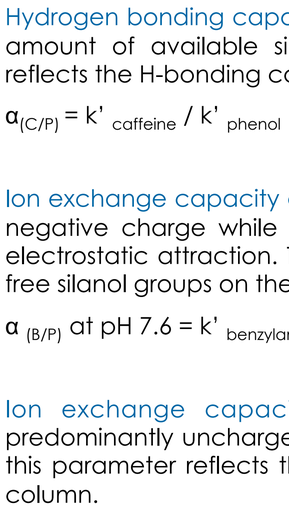
Ion exchange capacity at pH 2.7—at this pH silanols are predominantly uncharged while benzylamine is protonated, hence this parameter reflects the number of acidic silanol groups on the column.

$\alpha_{(B/P)}$ at pH 2.7 = $k'_{\text{benzylamine}} / k'_{\text{phenol}}$ at pH 2.7

Results & Discussion

The columns of interest can be selected from ACD/Labs' database of 340 columns—characterized with Tanaka parameters (provided by M.R. Euerby and P. Petersson). For this study, a subset of 27 columns (from Waters and Agilent) were chosen and the 5 most dissimilar were identified (Figure 2).

In order to get a result similar to what is obtained by PCA (Figure 1) it is necessary to apply a weighting factor of 2 for hydrophobicity parameters (Figure 2, k'PB and αCH_2). Thereby the hydrophobic parameters get a similar importance as the other Tanaka parameters.


List of Candidate Columns (List B):

#	Sel	Column Name
1	☐	Acquity BEH C18 1.7 um
2	☐	Acquity BEH C8
3	☐	Acquity BEH Phenyl 1.7um
4	☐	Acquity BEH Shield RP18 1.7 um
5	☐	Acquity CSH C18 1.7um
6	☐	Acquity CSH Fluorophenyl 1.7um
7	☐	Acquity CSH Phenyl hexyl 1.7um
8	☐	Acquity HSS C18 1.8 um
9	☐	Acquity HSS C18 SB 1.8 um
10	☐	Acquity HSS T3 1.8 um
11	☐	Acquity UPLC BEH C18 130 1.7 um
12	☐	Acquity UPLC BEH C18 300 1.7 um
13	☐	Zorbax 300 diphenyl 1.8um
14	☐	Zorbax 300 SB-C18 1.8um
15	☐	Zorbax 300 SB-C3 1.8um
16	☐	Zorbax Bonus-RP
17	☐	Zorbax Eclipse plus C18 1.8 um
18	☐	Zorbax Eclipse plus C18 3.5 um
19	☐	Zorbax Eclipse XB-phenyl
20	☐	Zorbax Eclipse XDB-C18
21	☐	Zorbax Extend C18
22	☐	Zorbax Rx C18
23	☐	Zorbax SB-C18
24	☐	Zorbax SB-C18 300
25	☐	Zorbax SB-C3
26	☐	Zorbax SB-C8
27	☐	Zorbax-SB Aq

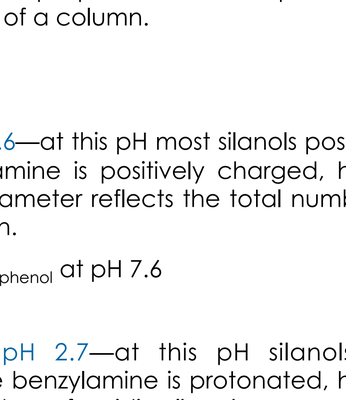


Figure 2 The list of 27 stationary phases evaluated in the study with Tanaka weightings and number of desired dissimilar columns.

The software generates a Hierarchical Clustering Dendrogram (Figure 3) which allows identification of dissimilar/similar columns and helps the user visualise how they are grouped together.

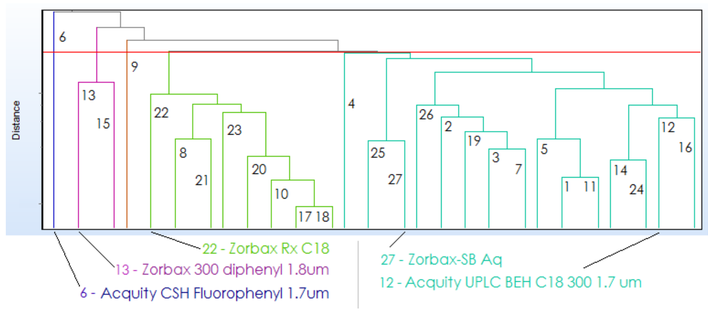


Figure 3 Hierarchical Clustering Dendrogram for the 27 column subset under investigation.

In addition, a radar plot is generated (Figure 4) which highlights the differences between the five selected columns as shown by their Tanaka parameters.

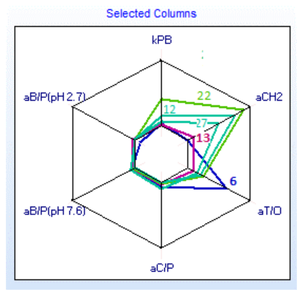


Figure 4 Radar plots of the Tanaka parameters for the five dissimilar columns selected by the software.

A summary report can easily be generated to rationalize and justify column selection in accordance with QbD principles (Figure 5).

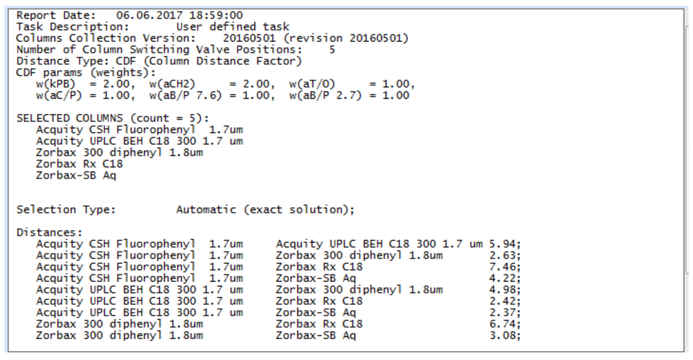


Figure 5 Report generated by ACD/Labs software providing justification for the five most dissimilar columns selected for screening.

Conclusion

Separation scientists can use software to help investigate chromatographic space with minimal chromatographic runs for greater confidence in the specificity of their methods. Since the first parameter under investigation is often stationary phase, the ability to quickly and confidently select a subset of the most dissimilar columns for screening can help jump start a method development project.

In this study we have shown that ACD/Labs' new tool for selection of the most dissimilar columns from within a database of characterized columns, provides comparable results to a PCA approach with greater simplicity and efficiency.

References

1.N. Tanaka et al. (**1989**). Chromatographic Characterization of Silica C18 Packing Materials. Correlation between a Preparation Method and Retention Behavior of Stationary Phase. *J. Chromatogr. Sci.* 27: 721-728.
2.M.R. Euerby and P. Petersson. (**2003**). Chromatographic classification and comparison of commercially available reversed-phase liquid chromatographic columns using principal component analysis. *J. Chromatogr. A.* 994: 13-36.
3.www.acdlabs.com/columnselector