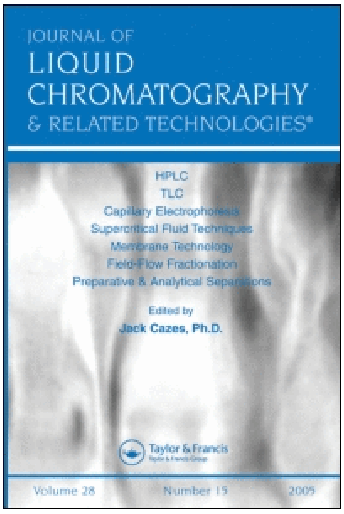

Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Factor Analysis in Normal Phase Liquid Chromatography Of N-Benzylideneanilines

S. Ounnar[a]; M. Righezza[a]; J. R. Chrétien[b]
[a] Institut de Chimie Organique et Analyhque, Universite d'Orleans, Orleans, France [b] Laboratoire de Chimiomdtrie, Universite d'Orleans, Orleans, France

To cite this Article Ounnar, S. , Righezza, M. and Chrétien, J. R.(1998) 'Factor Analysis in Normal Phase Liquid Chromatography Of N-Benzylideneanilines', Journal of Liquid Chromatography & Related Technologies, 21: 13, 2017 — 2037

To link to this Article: DOI: 10.1080/10826079808006603

URL: http://dx.doi.org/10.1080/10826079808006603

J. LIQ. CHROM. & REL. TECHNOL., 21(13), 2017–2037 (1998)

FACTOR ANALYSIS IN NORMAL PHASE LIQUID CHROMATOGRAPHY OF N-BENZYLIDENEANILINES

S. Ounnar[1], M. Righezza[1], J. R. Chrétien[2]

[1] Institut de Chimie Organique et Analytique
Université d'Orléans
B.P. 6759
45067 Orléans Cedex 2, France

[2] Laboratoire de Chimiométrie
Université d'Orléans
B.P. 6759
45067 Orléans Cedex 2, France

ABSTRACT

The retentions of 72 substituted N-benzylideneanilines (NBA) in normal phase liquid chromatography were determined using seven eluents prepared with heptan, tetrahydrofuran, octanol-1, and ethyl acetate. Principal component analysis (PCA) and correspondence factor analysis (CFA) were applied to describe the behavior of the compounds. With PCA, only one factor is necessary to model the retention. CFA offers an analysis of second-order effects and shows how the retention of the relevant compounds is related to the electronic effects. A multilinear regression with the Hammett's constants of substituents is established. The influence of the substituent position on retention is discussed.

INTRODUCTION

Retention in Liquid Chromatography (LC) has been extensively studied with Quantitative Structure Retention Relationships (QSRRs) methods.[1-5] These methods involve multiple regression analysis[6,7] to establish QSRRs between the retention and the structural properties of solutes. In some cases, multiple regression analysis can be improved by the use of factorial analysis (FA).[7-12]

The first step of the hyphenated techniques FA/QSRRs consists in the preprocessing of the selected data set, standardization, normalization or logarithmic transformation, and then the data processing with factor analysis itself. The abstract reproduction of data sets is the second step. This step involves determining the number of necessary abstract factors to reproduce the original data.[13] Finally, the abstract factors which have only a mathematical sense must be interpreted in terms of physico-chemical properties.

In this last step, several approaches are available. The abstract factors can be directly correlated to physico-chemical properties or they can be transformed to obtain new abstract or real factors.

Rotation of the abstract factors,[14,15] involving different rotational methods (varimax, quartimax,...), produces new abstract factors which may be easier to interpret due to the improved discrimination of clusters.

Target factor analysis (TFA) may be used to establish a relation between the abstract and real factors. Target transformation[13] involves a test vector, which represents a physico-chemical property. When target transformation fails, the tested property is rejected, otherwise the next abstract factor is analyzed. Finally, this transformation breaks the data matrix down into two real matrices representing the variations of the tested properties. A good example of target transformation in chromatography is the deconvolution of overlapped peaks.[16] When the behavior of solutes is studied, the test vector must represent the physico-chemical properties of individual solutes.

FA/QSRRs have been successfully applied in chromatography[17-25] with PCA or CFA in conjunction with TFA.[26-30] The physico-chemical properties commonly used in QSRRs are hydrophobicity constants,[31-34] Hammett's constants,[6,35] solvatochromic parameters,[36,37] or solubility parameters.[38] Nevertheless, the correlation between the physico-chemical properties and the transformed factors is not always well established.[30] It can be assumed that transformed factors do not represent pure physico-chemical properties but a combination of several properties. The transformation process applied in TFA might be improved with the more pure abstract factors. This may be achieved by reducing the redundancy and/or the elements which do not contribute to the information contained in the data matrix.

The present paper aims to study the chromatographic retention of substituted NBA with FA/QSRRs. Topological and factor analyses are applied to retention data determined in normal phase liquid chromatography using different eluents. The redundancy of the information and/or the non pertinent solutes are reduced.

The abstract factorial factors, defined by the relevant solutes, will be explained with physico-chemical descriptors such as polar contributions of substituents, dipolar moment, ionization potential, or Hammett's constants.

EXPERIMENTAL

Compounds

The N-benzylideneanilines (NBA) are Schiff bases obtained by the classical reaction of an aromatic amine with an aromatic aldehyde. The general formula of the substituted NBA is: $X\text{-}C_6H_4\text{-}CH{=}N\text{-}C_6H_4\text{-}Y$ where X and Y substituents in position meta or para are listed in Table 1. The substituted NBA will be denoted as X-Y in the text.

Chromatographic Eluents

The eluents consist of solvent mixtures based on heptan plus one modifier. The solvents used include heptan for HPLC (Fisons, Loughborough, Great Britain), tetrahydrofuran (Merck, Darmstadt, Germany), octanol-1 (Merck, Darmstadt, Germany), and ethyl acetate (Prolabo, Paris, France).

Equipment

The HPLC equipment includes an L6200 Intelligent Pump, an autosampler model 655A-40, a L3000 Multi channel photo Detector, and the corresponding D-6000 software for detection and acquisition from Merck-Hitachi, Darmstadt, Germany.

Table 1

Description of the 72 Substituted N-Benzylideneanilines

No.	Compounds X - Y	No.	Compounds X - Y
1	H - 4 Br	37	4F - 4Cl
2	4CN - 4F	38	4F - 3Cl
3	H - H	39	4F - 4CN
4	H - 4F	40	4F - $3NO_2$
5	H - 4Cl	41	4F - $4NO_2$
6	H - 3Cl	42	4Cl - $4CH_3$
7	H - 3CN	43	4Cl - H
8	H - 4CN	44	4Cl - 4F
9	H - $3NO_2$	45	4Cl - 4Cl
10	H - $4NO_2$	46	4Cl - 3Cl
11	$4NO_2$ - $4OCH_3$	47	4Cl - $3NO_2$
12	$4NO_2$ - $4CH_3$	48	4Cl - $4NO_2$
13	$4NO_2$ - $3CH_3$	49	3Cl - $4OCH_3$
15	$4NO_2$ - 3F	51	3Cl - H
16	$4NO_2$ - 4F	52	3Cl - 4F
17	$4NO_2$ - 4Cl	53	3Cl - 4Cl
18	$4NO_2$ - 4Br	54	3Cl - 3Cl
19	$4NO_2$ - 3Cl	55	3Cl - 4CN
20	$4OCH_3$ - H	56	$4CF_3$ - $4OCH_3$
21	$4OCH_3$ - 4F	57	$4CF_3$ - $4CH_3$
22	$4OCH_3$ - 4Cl	58	$4CF_3$ - H
23	$4OCH_3$ - 3Cl	59	$4CF_3$ - 4F
24	$4OCH_3$ - $3NO_2$	60	$4CF_3$ - 4Cl
25	$4CH_3$ - H	61	$4CF_3$ - 4CN
26	$4CH_3$ - 4F	62	$4CF_3$ - $3NO_2$
27	$4CH_3$ - 4Cl	63	4CN - $4CH_3$
28	$4CH_3$ - 3Cl	64	4CN - H
29	$4CH_3$ - 4CN	65	$3NO_2$ - $4OCH_3$
30	$4CH_3$ - $3NO_2$	66	$3NO_2$ - $4CH_3$
31	$4CH_3$ - $4NO_2$	67	$3NO_2$ - $3CH_3$
32	$3OCH_3$ - H	68	$3NO_2$ - H
33	$3OCH_3$ - 4F	69	3NO2 - 4F
34	4F - $4CH_3$	70	$3NO_2$ - 4Cl
35	4F - H	71	$3NO_2$ - 3Cl
36	4F - 4F	72	4Br - H

Table 2

Experimental Design - Composition of the Eluents for the Seven Experiments*

Label of Eluents	Tetrahydro-Furan/Heptane 5/95 (v/v)	Ethyl Acetate/ Heptane 6/94 (v/v)	Octanol-1/ Heptane 2/98 (v/v)
	$\overline{\text{THF}}$	$\overline{\text{AC}}$	$\overline{\text{OC}}$
$\overline{\text{THF}}$	1	0	0
$\overline{\text{AC}}$	0	1	0
$\overline{\text{OC}}$	0	0	1
$\overline{\text{THF}}$ $\overline{\text{AC}}$	1/2	1/2	0
$\overline{\text{THF}}$ $\overline{\text{OC}}$	1/2	0	1/2
$\overline{\text{AC}}$ $\overline{\text{OC}}$	0	1/2	1/2
$\overline{\text{THF}}$ $\overline{\text{AC}}$ $\overline{\text{OC}}$	1/3	1/3	1/3

---

* The eluents are prepared from $\overline{\text{THF}}$, $\overline{\text{AC}}$, and $\overline{\text{OC}}$.

The retention of substituted NBA is studied in normal phase liquid chromatography (NP-LC) with an amino bonded stationary phase, 7-µm Lichrospher 100 NH_2 column (125*4 mm ID) used with an oven model 655A-52 (Merck-Hitachi, Darmstadt, Germany) thermostated at 60°C.

Chromatographic Procedures

An experimental design was applied to the chromatographic systems defined by the three main eluents based on binary mixtures of heptan plus one modifier. The eluents were uniformly distributed inside the triangular diagram whose vertices are occupied by the three main binary mixtures. The other eluents occupy the edges and the center of this triangle. The exact eluent composition of the seven chromatographic systems chosen are given in Table 2. The flow rate was 1mL/min and the volume of injection was 10µL.

The solutes were injected simultaneously with two compounds belonging to the series of NBA. Three references were chosen from the series studied. They cover all the retention scale from the less up to the more retained solute.

Table 3

Retention Indices of the N-Benzylideneanilines Studied with the Seven Eluents

Solute	$\overline{\mathrm{THF}}$	$\overline{\mathrm{AC}}$	$\overline{\mathrm{OC}}$	$\overline{\mathrm{THF}}\ \overline{\mathrm{AC}}$	$\overline{\mathrm{THF}}\ \overline{\mathrm{OC}}$	$\overline{\mathrm{AC}}\ \overline{\mathrm{OC}}$	$\overline{\mathrm{THF}}\ \overline{\mathrm{AC}}\ \overline{\mathrm{OC}}$
1	121.08	122.00	138.99	122.09	130.52	131.20	131.05
2	269.59	266.38	267.74	268.36	269.14	268.25	268.58
3	116.27	121.23	145.87	116.18	130.52	133.48	129.87
4	119.34	125.73	141.56	120.13	129.26	133.48	128.68
5	117.30	119.34	136.69	118.12	126.66	128.85	126.23
6	117.30	117.41	135.75	118.12	126.66	126.42	126.31
7	226.37	231.98	252.63	229.45	235.79	240.37	236.41
8	241.46	247.31	266.08	243.52	248.02	255.81	250.44
9	200.00	204.32	204.60	202.05	202.30	203.50	200.00
10	200.00	206.65	209.67	202.05	204.63	208.89	200.00
11	287.22	285.47	279.49	286.04	284.78	282.99	285.93
12	191.71	189.64	187.47	190.61	189.03	189.19	190.73
13	184.30	184.04	185.02	184.15	185.70	184.27	183.73
14	100.00	100.00	100.00	100.00	100.00	100.00	100.00
15	219.31	214.13	200.00	217.68	211.36	207.57	214.10
16	230.31	222.30	208.34	227.03	222.18	215.91	222.24
17	229.32	216.98	200.00	224.97	219.57	210.20	219.31
18	233.59	219.50	203.07	228.65	225.71	212.77	223.39
19	228.44	216.64	203.12	224.52	221.00	210.20	219.30
20	182.38	189.34	200.00	184.54	191.12	196.73	190.73
21	186.35	200.00	200.00	189.78	191.10	194.79	191.19
22	186.53	189.34	190.18	187.23	188.29	189.72	188.97
23	181.45	183.86	186.00	181.42	184.33	183.70	183.73
24	285.49	287.19	285.74	285.55	284.95	285.72	285.33
25	112.29	115.44	143.66	115.96	126.66	131.20	126.23
26	114.33	117.54	135.28	117.05	122.59	125.18	122.42
27	110.19	111.00	126.77	107.22	119.75	121.31	117.03
28	112.29	111.98	128.60	112.78	118.29	119.98	118.41
29	230.58	237.10	265.29	232.79	236.37	142.35	237.43
30	192.92	195.69	195.77	193.86	194.20	195.28	196.84
31	192.17	200.00	200.00	193.06	192.66	195.28	193.36
32	185.71	189.96	191.44	186.80	188.14	189.19	187.84
33	103.54	111.07	128.21	109.52	116.80	121.31	117.03
34	104.68	100.00	130.07	106.06	115.28	117.23	112.74
35	112.29	117.41	135.58	116.99	125.33	125.18	124.98
36	121.08	123.07	133.42	121.64	123.92	126.42	124.98
37	119.21	115.41	119.17	119.98	122.59	121.31	121.11
38	120.94	118.39	128.21	121.76	125.33	122.62	119.77
39	271.92	272.19	291.51	271.28	273.70	275.49	274.68
40	228.16	224.60	213.92	227.66	223.40	218.36	223.39
41	227.95	224.22	214.79	225.99	223.55	220.17	225.66
42	102.36	103.63	117.80	104.92	108.87	111.42	101.68
43	112.29	112.37	130.92	112.73	121.18	121.31	115.56
44	120.12	112.37	122.95	117.83	121.18	119.98	115.56
45	115.33	110.19	116.05	115.82	118.29	109.89	100.00
46	119.08	110.19	115.98	116.91	121.18	112.92	115.63

Table 3 (Continued)

Solute	$\overline{\text{THF}}$	$\overline{\text{AC}}$	$\overline{\text{OC}}$	$\overline{\text{THF}}$ $\overline{\text{AC}}$	$\overline{\text{THF}}$ $\overline{\text{OC}}$	$\overline{\text{AC}}$ $\overline{\text{OC}}$	$\overline{\text{THF}}$ $\overline{\text{AC}}$ $\overline{\text{OC}}$
47	231.20	221.54	204.56	226.80	224.28	213.41	222.24
48	225.74	217.15	201.55	222.35	219.57	210.20	219.30
49	178.01	180.44	182.04	178.11	180.80	180.79	181.29
50	115.33	114.43	127.53	114.82	122.48	119.98	121.11
51	121.08	119.34	131.55	118.74	129.26	127.65	120.98
52	130.72	126.63	132.90	128.93	131.77	127.65	131.05
53	130.44	124.95	126.45	127.17	131.77	123.91	122.33
54	128.99	124.70	126.64	128.93	131.77	126.42	129.85
55	280.80	272.50	272.78	278.06	280.87	276.59	279.42
56	172.01	174.67	162.89	171.28	166.17	165.73	167.06
57	200.00	200.00	200.00	200.00	200.00	200.00	200.00
58	108.02	106.99	108.58	108.33	100.00	106.74	100.00
59	115.33	112.37	106.58	114.74	107.18	106.74	109.74
60	114.33	102.39	108.62	108.30	100.00	105.11	100.00
61	286.78	277.48	152.43	281.61	273.42	268.25	266.61
62	248.02	234.81	200.00	243.89	229.15	215.91	230.62
63	225.79	228.16	233.90	227.09	228.54	230.98	228.44
64	236.18	239.85	252.95	236.49	240.56	244.79	240.94
65	300.00	300.00	300.00	300.00	300.00	300.00	300.00
66	212.84	212.84	206.77	212.70	212.69	210.20	212.08
67	206.17	200.00	195.99	200.00	203.88	202.11	200.00
68	223.41	223.76	227.27	224.24	224.52	223.69	225.66
69	255.74	249.87	239.68	253.09	248.22	244.31	249.98
70	260.82	248.97	236.60	256.20	255.43	242.35	253.14
71	254.43	242.42	232.89	249.49	245.02	236.81	245.31
72	115.33	112.46	132.04	114.77	123.97	122.62	121.11

The three reference solutes selected with the lowest to the highest retention are: $4CF_3$-4Me, $4NO_2$-H and $3NO_2$-4OMe. From the measured capacity factors, retention indices (I_X) are calculated as follows:

$$I_x = 100 * n + 100 * \frac{\log(k'_x) - \log(k'_n)}{\log(k'_{n+1}) - \log(k'_n)}$$

where x represents the studied compound and n, n+1 are the reference compounds.

This experimental study yields a 69x7 retention data matrix which represents 69 indices calculated for the seven eluents (Table 3). The indices of the three reference compounds were not introduced into the matrix since their indices are constant.

THEORETICAL

Topological Analysis

In accordance with the DARC (description, acquisition, retrieval, and computer-aided design)[39] topological system and the DARC/PELCO[40] (perturbation of an environment which is limited, concentric, and ordered) procedures, the environment of each compound is described starting from an origin taken as the focus. Each substituent is localized in an environment which is limited, concentric, and ordered (ELCO). The influence of each substituent on retention is calculated in the form of a perturbation term corresponding to the formal replacement of a hydrogen atom by this substituent. Evaluation of the perturbation term is done on statistical bases and is an average, taking into account all the compounds in which it is found.

Principal Component Analysis

Principal Component Analysis (PCA)[8-9] is a linear algebra technique which attempts to describe the formulation of the observed data on the basis of a small number of underlying factors. The direct result of PCA is a set of abstract factors which do not have direct physical meaning.

The starting point in PCA is a data matrix denoted by [D]. The rows in the table are termed objects, i.e., solutes. The columns are termed variables and comprise the measurements made on the objects, i.e., the retention times for every solute studied (i) with the (j) eluent. The matrix of retention [D] is submitted to PCA to calculate the row and column vectors, R_k and C_k, respectively, in order to recalculate [D] as:

$$[D] = \sum_{i}^{k} R_k C_k$$

Where k is the number of abstract factors in the new and reduced space.

Correspondence Factor Analysis

Correspondence Factor Analysis (CFA)[10-11] is mathematically related to PCA, differing in the preprocessing and scaling of the data.

The experimental data matrix studied with CFA is a table of M(IxJ) of non-negative numbers k_{ij} in which the row and column sums are non-zero. This data matrix is transformed in the matrix F of relative frequencies by dividing each element k_{ij} by the sum n of elements. Elements f_{ij} of F are calculated with the following equation :

$$f_{ij} = \frac{k_{ij}}{n} \text{ with } n = \sum_i \sum_j k_{ij}$$

A symmetric study is applied to row and column elements. For the row study, each element is converted into a row profile by dividing each term of the row i by $f_{i.}$.

$$f_{i.} = \sum_j f_{ij} \text{ (for column elements } f_{.j} = \sum_i f_{ij} \text{)}$$

The row and the column elements define a cloud NI and NJ, in the column space R^J and in the row space R^I, respectively, in which each element has a coordinate. The CFA of the cloud NI, or NJ, searches for a series of eigenvectors for which the inertia of the cloud projection is maximal.

From the set of eigenvectors in the R^J or R^I space, the matrix of relative frequencies F can be reconstructed. The relative frequencies $\overline{f_{ij}}$ are calculated with the expression:

$$\overline{f_{ij}} = f_{i.} \times f_{.j} \times \left(1 + \sum_{q>1} \left[\sqrt{\lambda_q} \times \phi_{jq} \times \psi_{iq}\right]\right)$$

where:

- q is the number of extracted eigenvectors,
- λ_q is the q^{ith} eigenvalue,
- ϕ_{jq} is the eigenvector matrix in the column space R^J and
- ψ_{iq} is the eigenvector matrix in the row space R^I.

On the factorial axis defined by the corresponding eigenvector, the coordinate of each data point is calculated. Each element can be projected on a factorial plane, i.e. the first factorial plane defined by the two first factorial axes. On this plane, the direction given by the point projection and one of the factorial axes, i.e., the first one, forms an angle θ.

Table 4

Perturbation Terms for the NBA Series Studied with the Seven Eluents*

Focus with Eluent	Contribution	Substituent	Perturbation Term
$\overline{\text{THF}}$	118.09	Br	4.73
$\overline{\text{AC}}$	116.47	Cl	4.51
$\overline{\text{OC}}$	120.23	F	3.19
		CN	138.67
		NO_2	100.44
		OMe	57.94
		Me	-14.12
		CF_3	-3.84

---

* The contribution of the focus, unsubstituted NBA, is given for the three main eluents $\overline{\text{THF}}$, AC , and $\overline{\text{OC}}$. The square multiple regression coefficient R^2 equals 0.934 and F test is 603.4.

The quantity $cos^2\theta$ is called the contribution to the inertia of the cloud explained by the considered axis. This quantity can be expressed as a percentage. If the contribution is high, θ is low and the profile vector of the element is correlated with the axis.

RESULTS AND DISCUSSION

Principal Component Analysis

To obtain an overall view of the main trends governing retention, PCA is applied to the data matrix. The PCA of the retention indices gives rise to a first factorial plane representing 100% of the information content. The first axis accounts for 99% of the information content.

Considering the first factorial axis, the solutes are projected in two groups. The first group includes the solutes which have the lowest retention. These compounds are substituted in X or Y position with alkyl or halogen substituents. In the second group, the solutes are substituted with an acceptor

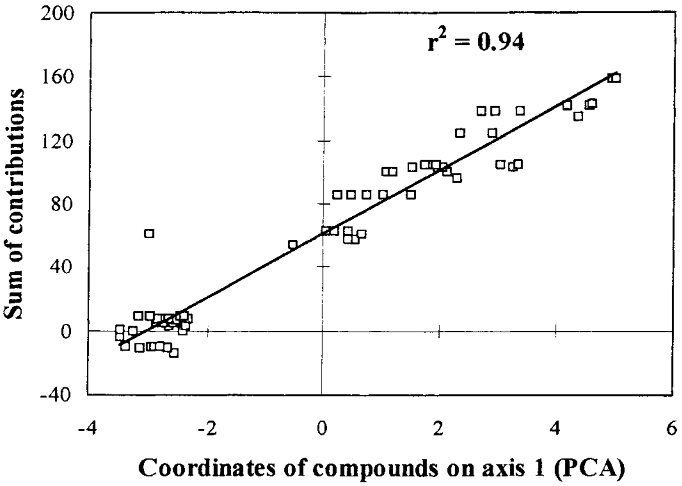

Figure 1. Correlation between the coordinates on axis 1 of PCA and the sum of contributions to the retention of each compound. The square regression coefficient r^2 equals 0.94.

group (CN, NO_2, CF_3) or strong donor (OCH_3). To give a physico-chemical meaning to the first factorial axis, let us consider the contribution of each subtituent to retention. The contribution is calculated according to topological analysis. The focus is the unsubtituted N-benzylideneaniline. Three values are calculated for the focus corresponding to the three main eluents. The perturbation term represents the contribution to the retention of the different subtituents in position *X* or Y, meta or para. The values of the contribution to retention are listed in Table 4.

In the NP-LC systems studied, the perturbation terms of substituent reflect the contribution to the polarity. For a substituted compound X-Y, its polarity is proportional to the sum contributed by each substituent X and Y. The physico-chemical meaning of the first factorial axis can be explained with the sum of the substituents contributions.

The coordinate of the compounds on the first factorial axis and the sum of contributions are correlated (Figure 1). The square regression coefficient r^2 equals 0.94.

Table 5

Relative Contribution (%) of the Studied Solutes to the Inertia Explained by Axes 1 and 2 of CFA

Compound	Axis 1	Axis 2	Compound	Axes 1	Axes 2
H - 4Br	1.80	1.20	4F - 4Cl	0.01	2.00
4CN - 4F	0.10	0.01	4F - 3Cl	0.20	0.70
H - H*	5.50	0.50	4F - 4CN	0.70	0.60
H - 4F*	2.80	0.40	4F - 3NO$_2$	1.20	0.10
H - 4Cl*	2.30	0.40	4F - 4NO$_2$	0.90	0.01
H - 3Cl*	2.00	1.70	4Cl - 4CH$_3$	1.50	0.10
H - 3CN	1.70	0.50	4Cl - H*	2.20	0.60
H - 4CN	1.40	1.50	4Cl - 4F	0.10	2.40
H - 3NO$_2$	0.01	1.80	4Cl - 4Cl	0.01	0.50
H - 4NO$_2$	0.10	3.00	4Cl - 3Cl	0.10	7.30
4NO$_2$ - 4OCH$_3$	0.40	0.20	4Cl - 3NO$_2$*	3.10	1.00
4NO$_2$ - 4CH$_3$	0.20	0.20	4Cl - 4NO$_2$*	2.60	0.90
4NO$_2$ - 3CH$_3$	0.01	0.10	3Cl - 4OCH$_3$	0.01	0.30
4NO$_2$ - 3F	1.90	0.10	3Cl - 4CH$_3$	0.90	1.70
4NO$_2$ - 4F*	2.20	0.10	3Cl - H	0.80	1.50
4NO$_2$ - 4Cl*	3.60	1.30	3Cl - 4F	0.01	1.00
4NO$_2$ - 4Br*	3.80	3.50	3Cl - 4Cl	0.10	1.40
4NO$_2$ - 3Cl*	2.80	1.80	3Cl - 3Cl	0.10	2.40
4OCH$_3$ - H	1.00	1.00	3Cl - 4CN	0.30	1.30
4OCH$_3$ - 4F	0.20	9.30	4CF$_3$ - 4OCH$_3$	0.80	4.20
4OCH$_3$ - 4Cl	0.01	0.80	4CF$_3$ - H	0.01	6.90
4OCH$_3$ - 3Cl	0.01	0.30	4CF$_3$ - 4F	1.00	1.40
4OCH$_3$ - 3NO$_2$	0.10	1.10	4CF$_3$ - 4Cl	0.20	0.50
4CH$_3$ - H*	6.00	1.10	4CF$_3$ - 4CN*	4.00	0.40
4CH$_3$ - 4F*	2.30	0.01	4CF$_3$ - 3NO$_2$*	9.70	0.10
4CH$_3$ - 4Cl*	2.30	1.10	4CN - 4CH$_3$	0.10	0.70
4CH$_3$ - 3Cl	1.60	0.40	4CN - H	0.60	1.00
4CH$_3$ - 4CN*	2.80	3.00	3NO$_2$ - 4CH$_3$	0.30	0.20
4CH$_3$ - 3NO$_2$	0.01	0.70	3NO$_2$ - 3CH$_3$	0.40	0.20
4CH$_3$ - 4NO$_2$	0.01	5.70	3NO$_2$ - H	0.01	0.20
3OHC$_3$ - H	0.01	1.40	3NO$_2$ - 4F	1.20	0.01
3OHC$_3$ - 4F*	3.40	0.01	3NO$_2$ - 4Cl*	2.30	2.50
4F - 4CH$_3$*	4.90	6.70	3NO$_2$ - 3Cl	1.90	0.70

Table 5 (Continued)

Compound	Axis 1	Axis 2	Compound	Axes 1	Axes 2
4F - H*	2.70	0.40	4Br - H	1.90	3.60
4F - 4F	0.70	0.40			

---

* Solutes noted with an asterisk have a contribution to the inertia explained by Axes 1 higher than 2%.

The first factorial axis reflects the influence of the polarity of substituents on retention. In a first approximation, retention is based principally on the polarity of substituent. A more in-depth analysis must be applied to study the second-order effects on the retention mechanism with CFA.

Correspondence Factor Analysis

CFA of the complete data matrix gives rise to the first factorial plane representing 91% of the total information with 85% on axis 1 and 6% on axis 2. The solutes contributions to the information explained by the first two factorial axes are given in Table 5. The plot of the contributions of each solute to the information explained by the first and second factorial axes (Figure 2), with log-log scale, indicates that only a few solutes make a significant contribution. It may be assumed that each axis is determined by thirty-three significant solutes which have a contribution between 1% and 10%. This assumption does not mean that solutes which do not have a significant contribution have no significant chromatographic retention. It means that these solutes have a behavior similar to that calculated from the marginal distributions. Only solutes which have a significant difference between the experimental and the calculated retention contribute to the inertia explained by the factorial axes.

According to Malinowski considerations[13] of the physico-chemical meaning of the successive and perpendicular abstract factorial axes, particular attention is paid to the physico-chemical properties described by the first factorial axis.

To give physico-chemical meaning to the first abstract axis, the solutes which have a significant contribution are identified. From the original data matrix, the solutes projected on the first factorial plane are suppressed successively from the center of the plane to the extremities of axis 1.

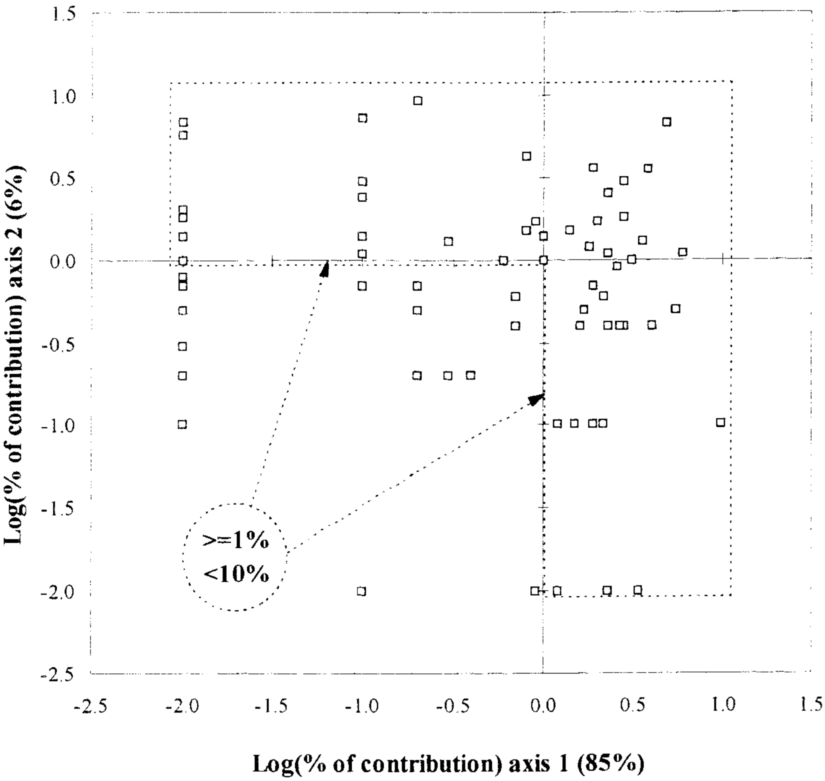

Figure 2. The log-log plot of the solute contributions to the information comprised by the first and the second factorial axes of CFA.

The step by step reduction of the data matrix gives rise to five successive data matrices which have 33, 28, 21, 14, and 9 principal solutes. The five data sub-matrices are analyzed with CFA. The solutes which have not a significant contribution to the first factorial axis are considered as complementary data: their coordinates on the first and second factorial axes are calculated on each factorial plane defined by the five sub-matrices. The five CFA lead to five factorial sub-planes in which each principal or complementary solute is projected simultaneously. The five factorial sub-planes are compared to the previous factorial planes calculated with 69 principal solutes which is considered as a reference.

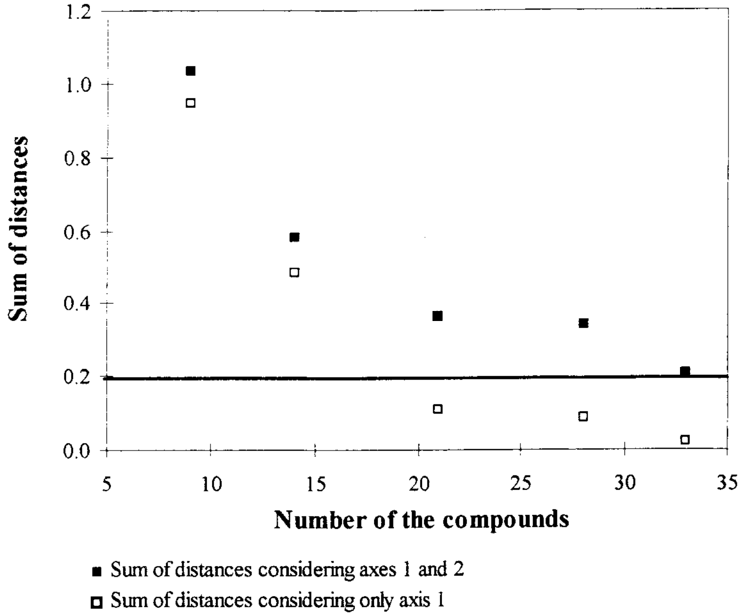

Figure 3. The plot of the sum of distances between the solutes projections on the reference plane and the five sub-planes, considering the two factorial axes or only axis 1 of CFA versus the number of principal solutes. The horizontal straight line gives the limit of the sum of projection distances. Above this limit, the sub-plane and the reference plane are not similar.

For each sub-plane, the sum of distances between each solute projection on the reference plane and its projection on the sub-plane, is calculated considering either the two factorial axes or only axis 1. The five sums are plotted versus the number of principal solutes and reported in Figure 3.

The experimental error measurement indicates that the standard deviation of a solute projection is around 5.5% on the first factorial axis. This deviation, applied to the 69 solutes gives rise to the limit of the sum of projection distances which equals 0.2. This value is reported in Figure 3 as a horizontal straight line. Above this limit, the difference between the solutes projections indicates that the two factorial planes, the sub-plane, and the reference plane, are not similar. It means that significant solutes are missing in the data sub-matrix.

Considering both axes 1 and 2, Figure 3 shows a high deviation, over the error line, starting with 35 solutes. This is not surprising since the process of matrix reduction regards only the contribution of the solutes to the inertia of the first axis. Some solutes, which have a significant contribution to the inertia defined by axis 2, are excluded. This set of solutes can be used to determine the physico-chemical meaning of axis 1 only.

Considering only axis 1, the progressive reduction of solute in the data matrix (Figure 3) causes a regular increase of distances between the solute projections. The sum of distances crosses the error line when the number of principal solutes becomes lower than 21. These 21 solutes (noted with an asterisk in Table 5) have a contribution to the information content of axis 1 that is higher than 2.0%. This indicates that the first factorial axis can be defined considering only 21 solutes instead of 69. The first axis is linked to the same physico-chemical properties when it is defined with the total matrix (69x7) or the sub-matrix (21x7).

The difference of projections between the 21 solutes in the total matrix and the sub-matrix is clearly shown in Figure 4. From the total matrix, only 21 selected solutes are retained and framed in bold. On the same graph, the 21 solutes of the sub-matrix are projected with the error indicator on axis 1. When the experimental error is taken into account, the solutes projection of the total and sub-matrices can be considered as identical.

To give a physico-chemical meaning to axis 1, four physico-chemical descriptors have been evaluated: the total dipolar moment (μ), the ionization potential (IP) which represent the global molecular properties and the Hammett's constants of substituents X and Y (σ_X, σ_Y) in positions meta or para. The results of regression analysis between the coordinate of solutes on axis 1 and each descriptor are given in Table 6.

In normal phase liquid chromatography, the retention mechanism is principally governed by dipolar and electronic interactions. From a chromatographic point of view, it appears that the main trend represented by axis 1 is not correlated to the dipolar moment; the square of regression coefficient r^2 is equal to 0.544. The significant correlation between solutes coordinate on the first axis and ionization potential, r^2 equals 0.92, reflects the charge transfer interactions. This electronic descriptor IP takes into account the global molecular property. It can be useful to use local electronic descriptors as the Hammett constants evaluated for X and Y substituents in position meta or para.

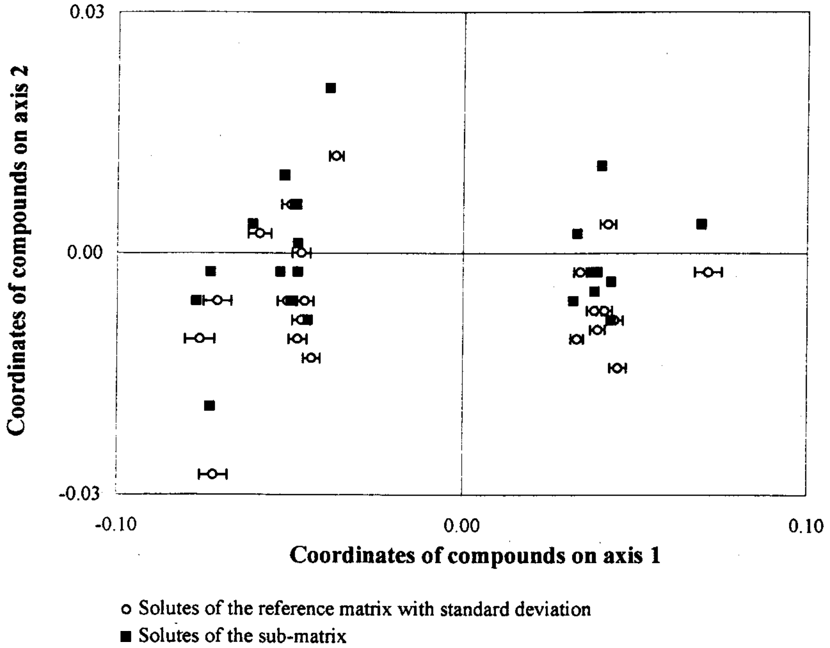

Figure 4. The solutes projection of the reference and the sub-matrix on the first factorial plane of CFA. The standard deviation is drawn for the solutes of the reference matrix.

Table 6

Correlation Coefficients Between the 21 Solutes Coordinates on First Factorial Axis of CFA and the Three Structural Descriptors; Dipolar Moment, Ionization Potential, and Hammett's Constants

	CFA Coordinates on Axis 1 Versus		
	Dipolar Moment (μ)	Ionization Potential (IP)	Hammett's Constants ($\sigma_x\sigma_y$)
R^2	0.54	0.92	0.94
Fisher	F(1,19) = 22.70	F(1,19) = 211.96	F(2,18) = 151.13

The multilinear correlation coefficient, R^2, of the solute (i) coordinate on axis 1 (C_{1i}) versus combination of σ_X and σ_Y equals 0.94. The correlation is represented by the following expression:

$$C_{1i} = (-0.059 \pm 0.008) + (0.100 \pm 0.016)\sigma_X + (0.085 \pm 0.021)\sigma_Y$$

This correlation is principally obtained with the relevant solutes substituted in position 4X or 4Y; 15 solutes are substituted in position 4X, 13 in position 4Y, 2 in position 3X and 4 in position 3Y. This indicates that the position para has a higher contribution to the information content than the meta position. Considering only the position para, the numbers of relevant solutes which are substituted in position 4X with NO_2, CH_3 or halogen are 4, 4, and 5, respectively. In position 4Y, these numbers are 1, 1, and 9, respectively. Substituent in position X has a large contribution to the information content. These results are in accordance with spectroscopic studies (NMR, IR)[41-42] of the physico-chemical properties of the NBA which have found that the same substituent has a stronger contribution to the electronic effects in the position X than in the position Y.

CONCLUSION

In the NP-LC systems studied, the chromatographic behavior of substituted NBA is governed principally by polar interactions and, to some extent, by electronic interactions. The principal abstract factors of PCA and CFA are correlated to real factors represented by physico-chemical properties. The first factorial axis of PCA is correlated to the contribution of substituents to the polarity. That of CFA is correlated to the combination of Hammett's constants σ_X and σ_Y for the 21 relevant solutes. The influence of substituent position is established for the set of relevant solutes. The preponderance of 4X position on electronic effects is in accordance with molecular spectroscopy studies of NBA.

REFERENCES

1. R. Kalizsan, J. Chromatogr A., **656**, 417-435 (1993).

2. L. C. Tan, P. W. Carr, J. Chromatogr A., **656**, 521-535 (1993).

3 . J. G. Dorsey, M. G. Khaledi, J. Chromatogr A., **656**, 485-499 (1993).

4. W. J. Lambert, J. Chromatogr A., **656**, 469-484 (1993).

5. L. C. Sander, S. A. Wise, J. Chromatogr A., **656**, 335-351 (1993).

6. O. Exner, **Correlation Analysis of Chemical Data**, Plenum Press, New York, 1988.

7. M. Mellinger, Chemometrics and Intelligent Laboratory Systems, **2**, 29-36 (1987).

8. S. Wold, K. Esbensen, P. Geladi, Chemometrics and Intelligent Laboratory Systems, **2**, 37-52 (1987).

9. J. Joliffe, **Principal Component Analysis**, Springer, Berlin, 1986.

10. M. J. Greenacre, **Theory and Application of Correspondence Analysis,** Academic Press, London, 1984.

11. M. Mellinger, Chemometrics and Intelligent Laboratory Systems, **2**, 61-77 (1987).

12. R. Kalizsan, **Quantitative Structure Chromatographic Retention Relationships,** Wiley Ed., New York, 1987.

13. E. R. Malinowski, **Factor Analysis in Chemistry**, Second Edition, Wiley, New York, 1991.

14. R. Kalizsan, K. Osmialowski, B. J. Bassler, R. A. Hartwick, J. Chromatogr., **499**, 333-344 (1990).

15. R. Kalizsan, K. Osmialowski, J. Chromatogr., **506**, 3-16 (1990).

16. P. Gemperline, J. Chem. Inf. Comput. Sci., **24**, 206-212 (1984).

17. P. T. Funke, E. D. Malinowski, D. E. Martire, L. Z. Pollara, Sep. Sci., **1**, 661-676 (1966).

18. M. Chastrette, J. Chromatogr. Sci., **14**, 357-359 (1976).

19. S. Wold, K. Anderson, J. Chromatogr., **80**, 43-59 (1973).

20. D. H. Mac Closkey, S. J. Hawkes, J. Chromatogr. Sci., **13**, 1-5 (1975).

21. S. R. Lowry, G. L. Ritter, H. S Woodruff, T. L. Isenhour, J. Chromatogr. Sci., **14**, 126-129 (1976).

22. P. H. Wiener, C. J. Dack, D. G. Howery, J. Chromatogr., **69**, 249-260 (1972).

23. R. B. Selzer, D. G. Howery, J. Chromatogr., **115**, 139-151 (1975).

24. D. G. Howery, Anal. Chem., **46**, 829-833 (1974).

25. P. H. Weiner, J. F. Parcher, Anal. Chem., **45**, 302-307 (1973).

26. J. F. K. Huber, C. A. M. Meijers, J. A. R. J. Hulsman, Anal. Chem., **44**, 111-116 (1972).

27. J. F. K. Huber, E. T. Alderlieste, H. Harren, H. Poppe, Anal. Chem., **45**, 1337-1343 (1973).

28. J. H. Kindswater, P. H. Weiner, T. J. Klingen, Anal. Chem., **46**, 982-988 (1974).

29. C. H. Lochmüller, C. Reese, A. J. Aschman, S. J. Breiner, J. Chromatogr. A., **656**, 3-18 (1993).

30. B. Walczak, L. Morin-Allory, M. Lafosse, M. Dreux, J. R. Chrétien, J. Chromatogr., **395**, 183-202 (1987).

31. R. E. Koopmans, R. F. Rekker, J. Chromatogr., **285**, 267-279 (1984).

32. P. Jandera, J. Chromatogr., **A 656**, 437-467 (1993).

33. T. Braumann, G. Weber, L. H. Grimme, J. Chromatogr., **261**, 329-343 (1983).

34. B. Walczak, M. Dreux , J. R. Chrétien, K.Szymoniak, M. Lafosse, L. Morin-Allory, J. P. Doucet, J. Chromatogr., **353**, 109-121 (1986).

35. B. Walczak, L. Morin-Allory, J. R. Chrétien, M. Lafosse, M. Dreux, Chemometrics and Intelligent Laboratory Systems, **1**, 79-90 (1986).

36. K. Jinno, A. Ishigaki, J. H. R. C. C., **5**, 668-673 (1982).

37. P. C. Sadek, P. W. Carr, R. M. Doherty, M. J. Kamlet, R. W. Taft, M. H. Abraham, Anal. Chem., **57**, 2971-2978 (1985).

38. T. Hanai, J. Chromatogr., **550**, 313-324 (1991).

39. J. E. Dubois, **Computer Representation and Manipulation of Chemical Information,** W. T. Wipke, S. Heller, R. Feldmann, E. Hyde (Editors), Wiley, New York, 1974.

40. J. R. Chrétien, J. E. Dubois, Anal. Chem., **49**, 747-756 (1977).

41. A. Echevarria, J. Miller, M. G. Nascimento, Magn. Res. Chem., **23**, 809-813 (1985).

42. Y. M. Issa, A. M. Hindamey, Y. A. Marghalani, R. M. Issa, Proc. Pakistan Acad. Sci., **16**, 19-24 (1979).

Received September 24, 1997
Accepted November 18, 1997
Manuscript 4632