

CHROM. 14,341

NUMERICAL TAXONOMY OF COMMON PHASES FOR GAS-LIQUID CHROMATOGRAPHY, USING CHLOROPHENOXY ALKYL ESTERS AS TEST SUBSTANCES

JACQUES O. DE BEER*

Instituut voor Hygiëne en Epidemiologie, J. Wytmanstraat 14, B-1050 Brussels (Belgium)
and

AUBIN M. HEYNDRICKX

Laboratorium voor Toxicologie, Rijksuniversiteit te Gent, Hospitaalstraat 13, B-9000 Ghent (Belgium)

(Received August 7th, 1981)

SUMMARY

Numerical taxonomy of the retention indices of the methyl and pentafluorobenzyl esters of ten chlorophenoxy alkyl acids has been employed to determine the relationships between phases used for gas-liquid chromatography.

INTRODUCTION

In a previous paper¹ we described and compared the gas-liquid chromatographic (GLC) behaviour of the methyl and pentafluorobenzyl (PFB) ester of ten chlorophenoxy alkyl acids on different liquid phases with increasing McReynolds constants. On liquid phases with similar McReynolds constants, identical separation patterns and analogous retention indices were established. A plot of the retention indices against a specific polarity constant from the McReynolds system gave complete information about the GLC behaviour on the examined phases. On more polar phases the retention indices of the chlorophenoxyacetic acid derivatives showed larger increases in comparison with the other derivatives. This also explained the changes in the elution order with increasing phase polarity.

We now examine whether these selected GLC phases can be classified, using the retention indices of the methyl and PFB esters of the ten structurally related chlorophenoxy alkyl acids as test substances. The method employed is that of numerical taxonomy, previously applied by Massart and co-workers²⁻⁷.

THEORETICAL

Taxonomy can be defined as the classification of individual species into groups with respect to their mutual resemblance. These groups can be linked to form more extensive groups, so generating a hierarchical system. In numerical taxonomy, a part of the "cluster" analysis developed by Sneath and Sokal⁸, the object is to classify

operational taxonomic units (OTUs) according to the values of a set of (taxonomic) characters. For example, solvents used for thin-layer chromatography (TLC) can be classified according to the resulting migration patterns for a mixture of compounds. The TLC solvents are the OTUs and the migration patterns the taxonomic characters, the values of which are given by the respective R_f values in the different solvent systems. The next step is to compare each OTU with the other OTUs and to evaluate their resemblance. The taxonomic similarity between pairs of OTUs is represented by means of one calculated value, the "similarity" coefficient. In chromatography, two important "similarity" coefficients are the distance coefficient and the correlation coefficient.

The distance coefficient

The greater the distance between the taxonomic characters of two OTUs, the greater is the mutual lack of similarity. Many mathematical formulae have been derived to express the distance between OTUs. A convenient formula was elaborated by Sneath and Sokal⁸, which yields the taxonomic or Euclidean distance, D

$$D = \left[\sum_{i=1}^n (x_{ij} - x_{ik})^2 \right]^{1/2}$$

where j and k are the OTUs, in this case the selected GLC phases on which the separations are performed; x_{ij} is the numerical value of the (taxonomic) character i for OTU j , i.e., the retention index of compound i on phase j ; n is the number of (taxonomic) characters or the number of compounds that is injected on each examined GLC phase.

The correlation coefficient

The correlation coefficient must be regarded as a mathematical expression of the relationship and/or proportionality between pairs of OTUs. When two chromatographic systems show a high correlation, the first system yields practically no further information about the second. Moffat and co-workers^{9,10} demonstrated that pairs of chromatographic systems, the retention parameters of which show the lowest correlation, yield the highest "discriminating power".

The dendrogram

After the similarity coefficients have been computed, a "similarity" matrix, $i_{\max} \times i_{\max}$ (i_{\max} = number of OTUs), can be constructed, which is the basis of the actual classification step. The purpose of this is to form "clusters" of OTUs for which within-group distance or variance is minimized and between-group variance is maximized. To achieve this, a variety of linkage or clustering techniques can be applied. The main consideration is to select the smallest distance or the best correlation.

If the most similar GLC phases, e.g., A and B, are assumed to belong to one group A', the distance between this group and the other liquid phases can be calculated:

$$D_{A'C} = \frac{1}{2} (D_{AC} + D_{BC})$$

So the two columns with two series of similarity coefficients for A and B can be eliminated from the matrix and replaced by one column with one series of similarity coefficients for A'. In this way a new matrix is constructed and the procedure is repeated until only one column with one series of coefficients remains. This method of matrix reduction, calculating the arithmetical mean, is called the "weighted pair group" method.

In the next step the dendrogram is constructed which illustrates graphically the resulting classification. From the dendrogram, appropriate chromatographic systems can be selected.

RESULTS

We have examined whether the retention indices (Tables I and II) of the methyl esters or the PFB esters of the ten chlorophenoxy acids are useful as taxonomic characters to demonstrate the similarities or differences between the ten GLC phases. Both the correlation and the distance coefficients were tested.

The correlation coefficients

The PFB esters. In Table III the correlation coefficients between the retention indices of the ten PFB esters on each GLC phase are compiled in a 10×10 matrix. It is clear that the GLC phases XE-60 (6) and OV-225 (7) exhibit the best relationship. In contrast, the correlation between phases DC-200 and OV-275 is very poor. So phases 6 and 7 may be considered as belonging to the same group, 6', which is related to the other phases by the arithmetical mean values:

$$\begin{array}{l}
 (6) + (7) \rightarrow (6') \\
 r = 0.9999
 \end{array}
 \begin{array}{l}
 \left\{ \begin{array}{l} 1 \text{ versus } 6: 0.9879 \\ 1 \text{ versus } 7: 0.9885 \end{array} \right\} \rightarrow 1 \text{ versus } 6' = 0.9882 \\
 \left\{ \begin{array}{l} 2 \text{ versus } 6: 0.9795 \\ 2 \text{ versus } 7: 0.9802 \end{array} \right\} \rightarrow 2 \text{ versus } 6' = 0.9799 \\
 \left\{ \begin{array}{l} 3 \text{ versus } 6: 0.9824 \\ 3 \text{ versus } 7: 0.9829 \end{array} \right\} \rightarrow 3 \text{ versus } 6' = 0.9827 \\
 \left\{ \begin{array}{l} 4 \text{ versus } 6: 0.9950 \\ 4 \text{ versus } 7: 0.9955 \end{array} \right\} \rightarrow 4 \text{ versus } 6' = 0.9953 \\
 \left\{ \begin{array}{l} 5 \text{ versus } 6: 0.9956 \\ 5 \text{ versus } 7: 0.9962 \end{array} \right\} \rightarrow 5 \text{ versus } 6' = 0.9959 \\
 \left\{ \begin{array}{l} 8 \text{ versus } 6: 0.9981 \\ 8 \text{ versus } 7: 0.9981 \end{array} \right\} \rightarrow 8 \text{ versus } 6' = 0.9981 \\
 \left\{ \begin{array}{l} 9 \text{ versus } 6: 0.9928 \\ 9 \text{ versus } 7: 0.9928 \end{array} \right\} \rightarrow 9 \text{ versus } 6' = 0.9928 \\
 \left\{ \begin{array}{l} 10 \text{ versus } 6: 0.9832 \\ 10 \text{ versus } 7: 0.9832 \end{array} \right\} \rightarrow 10 \text{ versus } 6' = 0.9832
 \end{array}$$

A new reduced 9×9 matrix can be obtained (Table IV). In this matrix the best relationship is shown for phases 4 (OV-17) and 5 (QF-1), belonging to a common group 4'. This matrix is also further reduced as follows:

TABLE I

RETENTION INDICES OF METHYL ESTERS

Abbreviations: CPIP = *p*-chlorophenoxyisobutyric acid; MCPP = methylchlorophenoxy-2-propionic acid; MCPA = methylchlorophenoxyacetic acid; 2,4-DP = 2,4-dichlorophenoxy-2-propionic acid; 2,4-D = 2,4-dichlorophenoxyacetic acid; 2,4,5-TP = 2,4,5-trichlorophenoxy-2-propionic acid; 2,4,5-T = 2,4,5-trichlorophenoxyacetic acid; MCPB = methylchlorophenoxybutyric acid; 2,4-DB = 2,4-dichlorophenoxybutyric acid; 2,4,5-TB = 2,4,5-trichlorophenoxybutyric acid (for formulae cf. ref. 1).

Ester	<i>Apollane</i> (C_{87})	DC-200	DC-11	OV-17	QF-1	XE-60	OV-225	NPGA	FFAP	OV-275
(1) CPIB-CH ₃	1444	1475	1498	1703	1891	1907	1955	1950	2025	2156
(2) MCPP-CH ₃	1497	1519	1546	1769	1947	1985	2035	2038	2115	2227
(3) MCPA-CH ₃	1522	1524	1549	1812	1994	2059	2113	2144	2242	2362
(4) 2,4-DP-CH ₃	1567	1576	1597	1847	2034	2089	2142	2144	2242	2347
(5) 2,4-D-CH ₃	1594	1594	1615	1900	2099	2189	2241	2271	2395	2521
(6) 2,4,5-TP-CH ₃	1705	1716	1738	1988	2170	2235	2288	2300	2371	2446
(7) 2,4,5-T-CH ₃	1734	1736	1753	2049	2241	2349	2403	2453	2551	2644
(8) MCPB-CH ₃	1734	1740	1756	2024	2236	2288	2337	2352	2420	2510
(9) 2,4-DB-CH ₃	1796	1800	1813	2106	2316	2393	2444	2461	2551	2628
(10) 2,4,5-TB-CH ₃	1968	1955	1962	2266	2479	2570	2619	2651	2719	2769

TABLE II

RETENTION INDICES OF PFB ESTERS

Ester	<i>Apollane</i> (C_{87})	DC-200	DC-11	OV-17	QF-1	XE-60	OV-225	NPGA	FFAP	OV-275
(1) CPIB-PFB	1848	1935	1953	2162	2574	2548	2563	2557	2521	2746
(2) MCPP-PFB	1902	1978	1998	2229	2635	2627	2647	2652	2620	2819
(3) MCPA-PFB	1964	2021	2037	2316	2733	2760	2771	2821	2820	3006
(4) 2,4-DP-PFB	1980	2044	2060	2316	2733	2760	2771	2778	2758	2952
(5) 2,4-D-PFB	2041	2090	2118	2408	2849	2912	2919	2966	2982	3182
(6) 2,4,5-TP-PFB	2118	2179	2191	2460	2882	2912	2919	2943	2904	3062
(7) 2,4,5-T-PFB	2187	2228	2240	2561	3012	3080	3085	3150	3147	3304
(8) MCPB-PFB	2205	2263	2272	2561	3012	3023	3034	3067	3047	3182
(9) 2,4-DB-PFB	2278	2329	2332	2646	3106	3135	3144	3184	3178	3304
(10) 2,4,5-TB-PFB	2444	2480	2492	2815	3291	3347	3343	3392	3363	3460

TABLE IV
REDUCED 9 × 9 MATRIX

	1	2	3	4	5	6'	8	9	10
1	1.0000	0.9988	0.9991	0.9981	0.9969	0.9822	0.9791	0.9648	0.9443
2		1.0000	0.9995	0.9942	0.9927	0.9799	0.9683	0.9515	0.9281
3			1.0000	0.9951	0.9940	0.9827	0.9714	0.9553	0.9333
4				1.0000	0.9995	0.9953	0.9894	0.9790	0.9622
5					1.0000	0.9959	0.9904	0.9811	0.9656
6'						1.0000	0.9981	0.9928	0.9832
8							1.0000	0.9972	0.9908
9								1.0000	0.9971
10									1.0000

$$\begin{aligned}
 & \left. \begin{array}{l} \{ 1 \text{ versus } 4: 0.9981 \} \\ \{ 1 \text{ versus } 5: 0.9969 \} \end{array} \right\} \rightarrow 1 \text{ versus } 4' = 0.9975 \\
 & \left. \begin{array}{l} \{ 2 \text{ versus } 4: 0.9942 \} \\ \{ 2 \text{ versus } 5: 0.9927 \} \end{array} \right\} \rightarrow 2 \text{ versus } 4' = 0.9935 \\
 & \left. \begin{array}{l} \{ 3 \text{ versus } 4: 0.9951 \} \\ \{ 3 \text{ versus } 5: 0.9940 \} \end{array} \right\} \rightarrow 3 \text{ versus } 4' = 0.9946 \\
 & \left. \begin{array}{l} \{ 6' \text{ versus } 4: 0.9953 \} \\ \{ 6' \text{ versus } 5: 0.9959 \} \end{array} \right\} \rightarrow 6' \text{ versus } 4' = 0.9956 \\
 & \left. \begin{array}{l} \{ 8 \text{ versus } 5: 0.9894 \} \\ \{ 8 \text{ versus } 5: 0.9904 \} \end{array} \right\} \rightarrow 8 \text{ versus } 4' = 0.9899 \\
 & \left. \begin{array}{l} \{ 9 \text{ versus } 4: 0.9790 \} \\ \{ 9 \text{ versus } 5: 0.9811 \} \end{array} \right\} \rightarrow 9 \text{ versus } 4' = 0.9801 \\
 & \left. \begin{array}{l} \{ 10 \text{ versus } 4: 0.9622 \} \\ \{ 10 \text{ versus } 5: 0.9656 \} \end{array} \right\} \rightarrow 10 \text{ versus } 4' = 0.9639
 \end{aligned}$$

In the same way, new matrixes (8×8 , 7×7 , ...) can be constructed which are reduced step by step.

The best correlation coefficients in each reduced matrix are:

$$\begin{array}{llll}
 6 + 7 \rightarrow 6' & r = 0.9999 & \text{in } 10 \times 10 \text{ matrix} \\
 4 + 5 \rightarrow 4' & r = 0.9995 & 9 \times 9 \\
 2 + 3 \rightarrow 2' & r = 0.9995 & 8 \times 8 \\
 1 + 2' \rightarrow 1' & r = 0.9990 & 7 \times 7 \\
 6' + 8 \rightarrow 8' & r = 0.9981 & 6 \times 6 \\
 9 + 10 \rightarrow 9' & r = 0.9971 & 5 \times 5 \\
 1' + 4' \rightarrow 1'' & r = 0.9958 & 4 \times 4 \\
 8' + 9' \rightarrow 8'' & r = 0.9937 & 3 \times 3 \\
 8'' + 1'' \rightarrow 1''' & r = 0.9730 & 2 \times 2
 \end{array}$$

Hence the dendrogram can be plotted as shown in Fig. 1. This illustrates the relationship of one individual phase with other grouped phases. So it seems that OV-17 and QF-1 are more similar to the apolar phases and the NPGA is more similar to XE-60 and OV-225 than to FFAP. On this dendrogram we can also distinguish two separate groups of liquid phases with different chromatographic properties.

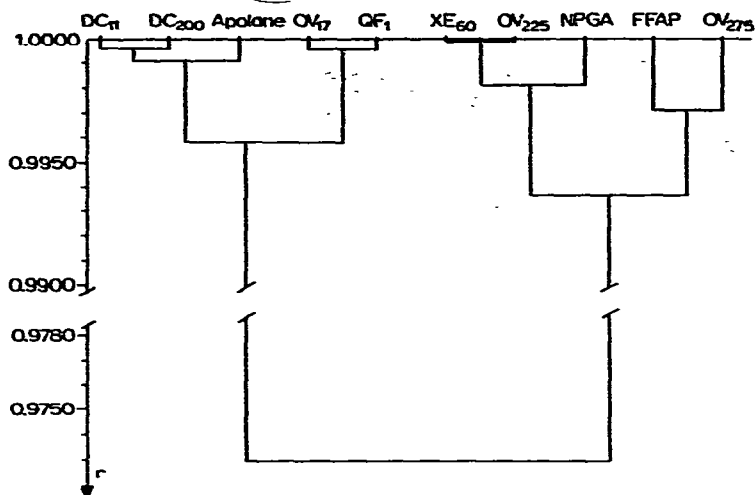


Fig. 1. Dendrogram plotted by means of the correlation coefficients of the PFB esters.

The methyl esters. The same methodology is followed for the methyl esters. The correlation coefficients between their retention indices on the ten GLC phases are given in Table V. This 10 × 10 matrix also can be reduced step by step. The highest correlation coefficients in each reduced matrix are:

6 + 7 → 6'	r = 0.9999	in 10 × 10 matrix
2 + 3 → 2'	r = 0.9999	9 × 9
4 + 5 → 4'	r = 0.9989	8 × 8
1 + 2' → 1'	r = 0.9986	7 × 7
6' + 8 → 8'	r = 0.9979	6 × 6
9 + 10 → 9'	r = 0.9969	5 × 5
1' + 4' → 4''	r = 0.9945	4 × 4
8' + 9' → 8''	r = 0.9883	3 × 3
1'' + 8'' → 1'''	r = 0.9669	2 × 2

The resulting dendrogram is plotted in Fig. 2. A similar pattern between the GLC phases is seen as for the PFB esters.

The distance coefficients

The PFB esters. In Table VI the distance coefficients between the retention indices of the ten PFB esters on each GLC phase are compiled in a 10 × 10 matrix. The smallest distance is found between phases XE-60 and OV-225. Stepwise reduction of each matrix using the "weighted pair group" method yields for each matrix the lowest "distance" coefficients:

6 + 7 → 6'	D = 34.76	in 10 × 10 matrix
2 + 3 → 2'	D = 50.42	9 × 9
8 + 9 → 8'	D = 76.05	8 × 8

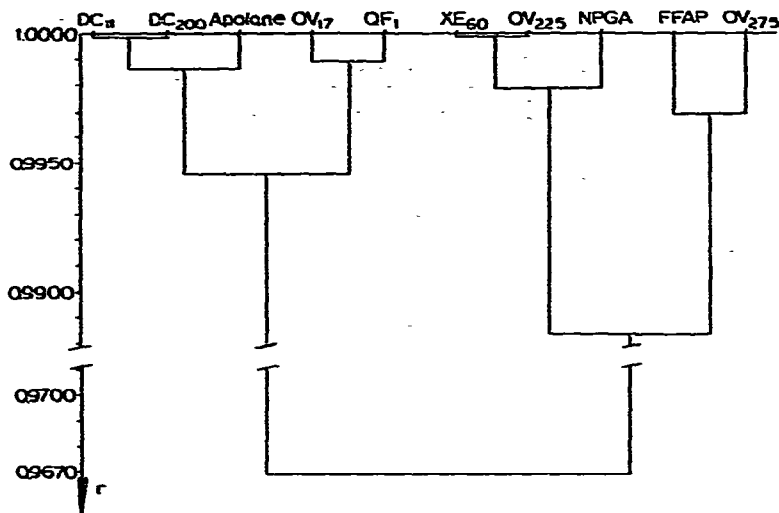


Fig. 2. Dendrogram plotted by means of the correlation coefficients of the methyl esters.

6' + 8' → 6''	D = 126.84	7 × 7
5 + 6'' → 5'	D = 191.19	6 × 6
1 + 2' → 1'	D = 210.29	5 × 5
5' + 10 → 10'	D = 638.86	4 × 4
1' + 4 → 4'	D = 1009.97	3 × 3
4' + 10' → 10''	D = 2270.93	2 × 2

The resulting dendrogram is plotted in Fig. 3.

The methyl esters. In Table VII the 10 × 10 matrix is given for the distance coefficients between the retention indices of the ten methyl esters on each GLC phase. Here the smallest distance is found between DC-200 and apolane (C₈₇). After step-wise reduction of this matrix, the lowest distance coefficients for each matrix are:

1 + 2 → 2'	D = 43.31	in 10 × 10 matrix
7 + 8 → 8'	D = 77.98	9 × 9
2' + 3 → 3'	D = 80.78	8 × 8
6 + 8' → 6'	D = 195.15	7 × 7
5 + 6' → 5'	D = 318.89	6 × 6
9 + 10 → 9'	D = 319.08	5 × 5
5' + 9' → 5''	D = 719.45	4 × 4
3' + 4 → 4'	D = 874.31	3 × 3
4' + 5'' → 4''	D = 1551.40	2 × 2

The resulting dendrogram is plotted in Fig. 4.

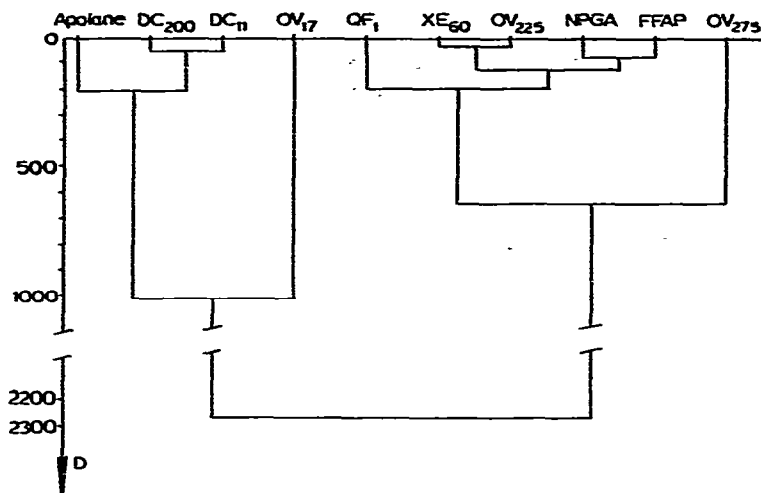


Fig. 3. Dendrogram plotted by means of the distance coefficients of the PFB esters.

DISCUSSION

Similar GLC phases can easily be recognized by observing the position of the chlorophenoxy acetic esters on the chromatograms with respect to the other chlorophenoxy esters. In contrast, the "2-propionic" and "butyric" esters cannot be employed to elucidate any relationship between GLC phases, based on their correlation coefficients.

The resulting dendrograms, plotted by means of either the correlation coefficients or the distance coefficients, divide the examined GLC phases into two main groups which can be designated as "apolar" and "polar", reflecting the strongly divergent separation patterns. The significance of the correlation coefficients is questionable since the number of examined solutes chromatographed on the different GLC phases is rather small and the correlation coefficients between their retention indices on these phases are rather high. However, when the two dendrograms are considered together, the resulting classification of the GLC phases is exactly as predicted by classical theories. In both dendrograms four groups of phases can clearly be distinguished:

- (1) strictly apolar phases: Apolane (C_{87}), DC-11 and DC-200
- (2) intermediate apolar phases including OV-17 and QF-1
- (3) intermediate polar phases including XE-60, OV-225 and NPGA
- (4) extremely polar phases including FFAP and OV-275

The dendrograms, obtained from the distance coefficients, exhibit a similar phase classification. It is noteworthy however that QF-1 and OV-17 in both dendrograms are classed in a different cluster and are considered as not related. OV-17 is in the same cluster as apolane (C_{87}), DC-11 and DC-200; QF-1 on the other hand is related to XE-60, OV-225 and NPGA. In these dendrograms there is also only one group of intermediate phases, without any distinction between intermediate apolar and intermediate polar phases.

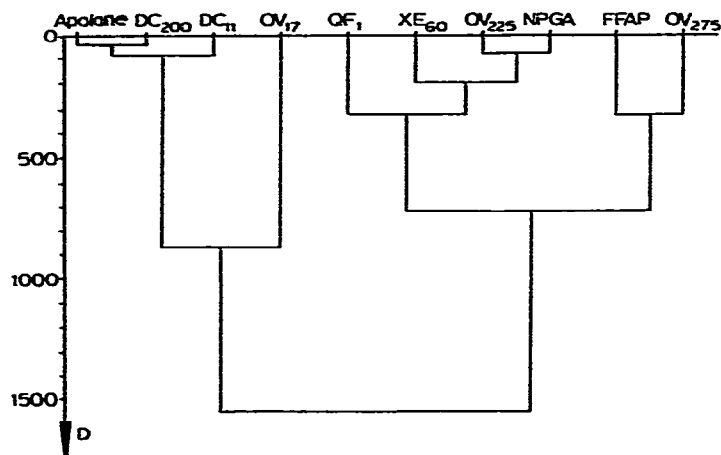


Fig. 4. Dendrogram plotted by means of the distance coefficients of the methyl esters.

The strong polar character of OV-275 is confirmed. In the dendrogram of the methyl esters, FFAP must be considered as a strong polar phase clustered with OV-27.5, but in that of the PFB-esters FFAP and NPGA are linked with XE-60 and OV-225. Here OV-275 appears as an isolated extremely polar phase.

The "distance" between polar and apolar phases is more clearly expressed by the retention indices of the PFB esters than those of the methyl esters. We believe, however, that the phase classification is more exactly reproduced by the correlation coefficients, perhaps because the whole separation patterns are compared. Nevertheless, the dendrograms plotted by means of the distance coefficients allow one to select those GLC phases, whose mutual distances, expressed by the respective difference between the retention indices of each test solute, differ most. Numerical taxonomy using both the correlation and distance similarity coefficients is an important mathematical tool, enabling the classification and combination of chromatographic systems.

The obtained results allow to consider two extreme points of view: the first believes in the use of a limited set of phases to solve all GLC problems; the second states that complicated separations cannot be achieved using a restricted set of GLC phases. We conclude that the selection of an optimal set of liquid phases and the chromatographic information they provide remains influenced by the studied series of chemically related compounds. It is not the aim of numerical taxonomy to develop a classification system for GLC phases, but to select those GLC phases which yield identical or divergent separations of a group of compounds.

REFERENCES

- 1 J. de Beer, C. van Peteghem and A. Heyndrickx, *J. Chromatogr.*, 157 (1978) 97.
- 2 D. L. Massart and H. de Clercq, *Anal. Chem.*, 46 (1974) 1988.
- 3 H. de Clercq and D. L. Massart, *J. Chromatogr.*, 115 (1975) 1.
- 4 A. Eskes, F. Dupuis, A. Dykstra, H. de Clercq and D. L. Massart, *Anal. Chem.*, 47 (1975) 2168.
- 5 H. de Clercq, D. van Oudheusden and D. L. Massart, *Analysis*, 3 (1975) 500.
- 6 D. L. Massart, M. Lauwereys and P. Lenders, *J. Chromatogr. Sci.*, 12 (1974) 617.
- 7 D. L. Massart and H. L. de Clercq, *Advan. Chromatogr.*, 16 (1978) 75.
- 8 P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, CA, 1973.
- 9 A. C. Moffat, K. W. Smalldon and C. Brown, *J. Chromatogr.*, 90 (1974) 1.
- 10 A. C. Moffat and K. W. Smalldon, *J. Chromatogr.*, 90 (1974) 9.