

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231173567>

Cluster analysis applied to the selection and combination of buffering electrolyte systems used for capillary electrophoresis of anions with water or methanol as solvents

ARTICLE *in* ANALYTICAL CHEMISTRY · MARCH 1990

Impact Factor: 5.64 · DOI: 10.1021/ac00204a003

CITATIONS

17

READS

18

2 AUTHORS, INCLUDING:



Ernst Kenndler

University of Vienna

258 PUBLICATIONS 6,364 CITATIONS

SEE PROFILE

- (16) *Handbook of Chemistry and Physics*; Weast, R. C., Ed.; CRC Press, Inc.: Boca Raton, FL, 1985.
 (17) Pospichal, J.; Deml, M.; Zemiova, Z.; Bocek, P. *J. Chromatogr.* **1985**, *320*, 139-146.
 (18) Kenndler, E.; Jenner, P. *J. Chromatogr.* **1987**, *390*, 185-197.
 (19) Kiso, Y.; Hirokawa, T. *Chem. Lett.* **1979**, 891-894.
 (20) Hirokawa, T.; Gojo, T.; Kiso, Y. *J. Chromatogr.* **1986**, *369*, 59-81.
 (21) Mikkers, F. E. P.; Everaerts, F. M.; Verheggen, Th. P. E. M. *J. Chromatogr.* **1979**, *169*, 11-20.
 (22) Mikkers, F. E. P.; Everaerts, F. M.; Verheggen, Th. P. E. M. *J. Chromatogr.* **1979**, *169*, 1-10.
 (23) Gross, L.; Yeung, E. S. *J. Chromatogr.* **1989**, *480*, 169-178.

- (24) Hogan, B.; Yeung, E. S. *J. Chromatogr. Sci.*, in press.

RECEIVED for review September 1, 1989. Accepted December 5, 1989. The Ames Laboratory is operated by Iowa State University for the U.S. Department of Energy under Contract W-7405-Eng-82. This work was supported by the Director of Energy Research, Office of Health and Environmental Research.

Cluster Analysis Applied to the Selection and Combination of Buffering Electrolyte Systems Used for Capillary Electrophoresis of Anions with Water or Methanol as Solvents¹

Ernst Kenndler* and Brigitte Gassner

Institute for Analytical Chemistry, University of Vienna, Währingerstrasse 38, A-1090 Vienna, Austria

The use of methanol as a solvent for buffering electrolyte systems used for the electrophoresis in capillaries was examined by cluster analysis. For this general approach, four methanolic electrolyte systems with different pH values were compared with eight aqueous systems. The similarity between these 12 systems was described by the Euclidian distances, calculated from the values of the electrophoretic mobilities of 55 anions. Clusters were constructed by use of a hierarchic algorithm and delineated by dendrograms. On the basis of the structures of the clusters, an appropriate selection is derived for the combination of electrolyte systems. It was found, that at least one methanolic system must be selected for the most favorable combination of three systems. Besides a physicochemical interpretation, the validity of the clusters was empirically proved by comparison with clusters formed by systems characterized by attributes with randomly generated values.

INTRODUCTION

Organic solvents are used for buffering electrolyte systems in capillary electrophoresis because they often influence the mobilities of the analytes in a direction, which can improve the selectivity in a single system. Furthermore, the combination of aqueous and organic (or mixed) solvents can be used in a multidimensional approach to increase the separability of analytes on the one hand, or to increase the probability of their identification on the other hand. This goal is better reached as the lower electrophoretic properties in the combined systems correlate.

Methanol is an often used organic solvent or solvent constituent in electrophoresis in capillaries, especially in isotachophoresis (1). Besides the effect of the enhanced solubility of organic solutes, the advantage of the application of this solvent as a constituent in mixed aqueous-organic solvents was demonstrated on some examples (2-4). No broader ap-

Table I. Electrolyte Systems for Isotachophoresis of Anions with Water or Methanol as the Solvent^a

code	solvent	pH of leading electrolyte	counterion
3	H ₂ O	3.0	β -alanine
4	H ₂ O	4.0	β -alanine
5	H ₂ O	5.0	creatinine
6	H ₂ O	6.0	histidine
7	H ₂ O	7.0	imidazole
8	H ₂ O	8.0	tris(hydroxymethyl)aminomethane
9	H ₂ O	9.0	2-amino-2-methyl-1,3-propanediol
10	H ₂ O	10.0	ethanolamine
7*	MeOH	7.76	triethanolamine
8*	MeOH	8.06	triethanolamine
9*	MeOH	9.22	tris(hydroxymethyl)aminomethane
10*	MeOH	10.3	cyclohexylamine

^a Leading ion: in aqueous systems chloride; in methanolic systems perchlorate. The methanol content is above 99.5% (w/w).

proach was given, however, to evaluate the utility of pure methanol in capillary electrophoresis.

Two main methods are proposed in the literature to evaluate the resemblance or the dissimilarity of analytical systems in a general mode: information theory and cluster analysis (5-8). Both methods were applied to isotachophoresis in previous papers (9, 10). In the present paper, cluster analysis is used to characterize and compare electrolyte systems with methanol or water as solvents for the aim of a rational selection of combinations of such systems.

The cluster analysis follows the general scheme (11-13): selection of the taxonomic units (or data units), selection of the variables (or properties, attributes, features), definition of the measure of similarity, construction of the dissimilarity matrix, selection of a clustering procedure for the agglomeration of the units and visualization, e.g. in a dendrogram, and finally the interpretation of the grouping.

According to this scheme eight electrolyte systems with water and four systems with methanol as solvents at different pH values were selected as the taxonomic units. These systems were proposed for the isotachophoresis of anions (14,

¹ This work is dedicated to Professor J. F. K. Huber on the occasion of his 65th birthday.

Table II. List of Anions Used for the Characterization of the Electrolyte Systems by Cluster Analysis

acetate	2,3-dibromobutyrate	lactate
acrylate	2,3-dibromopropionate	levulinate
benzoate	dichloroacetate	mandelate
bromate	enanthate	methacrylate
bromide	fluoride	2-naphthalenesul-
bromoacetate	fluoroacetate	fonate
2-bromopropionate	formate	nicotinate
5-bromosalicylate	glucuronate	nitrate
butyrate	glutamate	nitrite
caproate	glycerate	pelargonate
caprylate	glycolate	picrate
chlorate	2-hydroxybutyrate	propionate
chloroacetate	3-hydroxybutyrate	pyrovate
<i>m</i> -chlorobenzoate	2-hydroxyisobutyrate	salicylate
<i>o</i> -chlorobenzoate	iodate	sorbate
<i>p</i> -chlorobenzoate	iodide	thiocyanate
2-chloropropionate	iodoacetate	trichloroacetate
cinnamate	isovalerate	trifluoroacetate
crotonate		valerate

15). The values of the relative electrophoretic mobilities of 55 anions in each system were chosen as the features for the calculations. Similarity was described by the Euclidian distances between the points, by which the electrolyte systems could be represented in the corresponding (55-dimensional) pattern space. The grouping was finally visualized by a hierarchic clustering method. The interpretation of the grouping found was based on the physicochemical parameters, which influence the effective mobilities of the analytes in the electrolyte systems. The existence of a "real" structure of the data was proved by comparison with systems characterized by features with randomly chosen values.

EXPERIMENTAL SECTION

The values of the relative mobilities (R_E values) of 55 anionic compounds, shown in Table I, were used as variables for the cluster analysis of the 12 operational taxonomic units (eight aqueous and four methanolic electrolyte systems, given in Table II). The R_E values are the ionic mobilities, m_i , of the analytes, i , related to the mobility of the leading ion, m_L : $R_{E,i} = m_L/m_i$. They were taken from data published by Hirokawa et al. (14, 15). The data in water were calculated from known absolute mobilities and pK_a values according to an algorithm described in ref 14. The R_E values in methanol (purity > 99.5% (w/w)) were measured by isotachopheresis. The pK_a values were calculated based on these data (15). The leading ions were chloride in aqueous and perchlorate in methanolic buffers.

Only those anions were selected for the present investigation, for which a complete data set, concerning the R_E and the pK_a values, was available. For five anions (2,3-dibromobutyrate, 5-bromosalicylate, trifluoroacetate, glutamate, and glucuronate), some of the R_E values were not given in the literature but could be obtained by extrapolation. In a similar way, pK_a values not given in refs 14 and 15 could be approximated from data measured in similar solvent systems (2, 16), whereby a data set based on 55 solutes remained.

The clustering procedures were carried out by the statistical package SPSS-X 3.1 (SPSS, Inc., Chicago, IL). The clustering algorithms (average and single linkage) were sequential, agglomerative, hierarchic, and nonoverlapping. The Euclidian distances were calculated from the scaled R_E values (zero mean, unit variance). The random numbers were generated by using the NAG Fortran Library Mark 12 (Numerical Algorithms Group, Ltd., NAG Central Office, Oxford, U.K.). An IBM 3090/400E-VF was used for the calculations.

RESULTS AND DISCUSSION

Comparison of Aqueous Electrolyte Systems Based on Different Numbers of Solutes. A first step was to clarify whether the actual data units based on the attributes of the 55 anions selected in the pure aqueous solvent can be considered to be representative for the larger data set, consisting

Table III. Similarity Matrix of the Aqueous Electrolyte Systems Given by the Euclidian Distances^a

	relative Euclidian distance							
system	3	4	5	6	7	8	9	10
3	0.000							
4	0.165	0.000						
5	0.643	0.486	0.000					
6	0.919	0.785	0.321	0.000				
7	0.975	0.846	0.391	0.067	0.000			
8	0.978	0.849	0.396	0.073	0.011	0.000		
9	0.981	0.852	0.400	0.103	0.076	0.069	0.000	
10	1.000	0.872	0.449	0.240	0.231	0.224	0.156	0.000

^a The values of the Euclidian distances are related to the largest distance (between the systems 3 and 10). The numbers of the systems correspond with Table I.

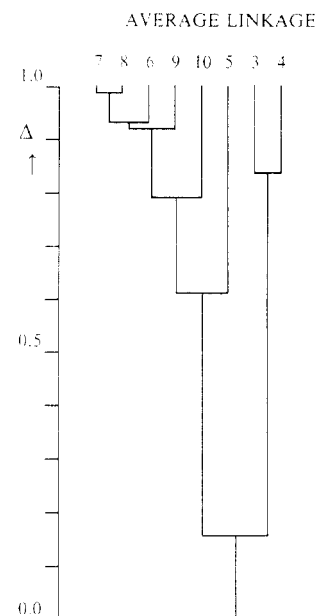


Figure 1. Dendrogram as the result of the average linkage algorithm based on the R_E values of 55 solutes in aqueous electrolyte systems. The symbols of the systems correspond to their pH values, as given in Table I. Δ = normalized distance.

of 263 anions, which was investigated in a previous paper (10).

For this comparison the similarity matrix was calculated and is given in Table III. It is based on the Euclidian distances between the points (in a 55-dimensional pattern space), which represent the aqueous buffering systems given in Table II. It can be seen from Table III that systems 3 and 10 show the largest value of 1.000 for the relative distance. This is in agreement with the result previously derived from the larger data set consisting of the values of 263 anions. Accordance is also observed for the sequences of, e.g., the five most similar pairs of systems: the smallest distances are found between (7,8), (6,7), (8,9), (6,8), and (7,9) in the present paper, and between (8,9), (9,10), (7,8), (7,9), and (6,7) in our previous investigation.

The high agreement of the results can also be visualized clearly by comparing the respective dendrograms. The dendrogram based on the data set from 55 anions was calculated with the same average linkage algorithm as the one demonstrated previously and is shown in Figure 1. The Δ values in the dendrograms are the Euclidian distances, d , normalized in the interval [0,1], according to $\Delta = (d_{\max} - d)/d_{\max}$, d_{\max} being the largest distance.

If this dendrogram is used for the selection of three systems as discussed (10) a grouping into {3,4}, {5}, and {7,8,6,9,10} is found. This is exactly the same result that was obtained

Table IV. Similarity Matrix of the Aqueous and the Methanolic Electrolyte Systems Given by the Euclidian Distances^a

system	Euclidian distance											
	3	4	5	6	7	8	9	10	7*	8*	9*	10*
3	0.00											
4	1.06	0.00										
5	4.07	3.12	0.00									
6	5.90	5.04	2.08	0.00								
7	6.26	5.43	2.51	0.43	0.00							
8	6.28	5.45	2.54	0.47	0.07	0.00						
9	6.30	5.47	2.57	0.66	0.49	0.44	0.00					
10	6.42	5.60	2.88	1.54	1.48	1.44	1.00	0.00				
7*	2.19	2.35	4.26	5.82	6.14	6.16	6.18	6.29	0.00			
8*	2.21	2.25	4.07	5.65	5.97	5.99	6.02	6.15	0.54	0.00		
9*	5.06	4.61	3.57	3.79	3.94	3.96	4.12	4.61	4.56	4.34	0.00	
10*	6.09	5.59	4.07	3.70	3.73	3.75	3.91	4.39	5.65	5.43	1.51	0.00

^aThe numbers of the systems correspond with Table I. The methanolic systems are indicated by an asterisk.

when cluster analysis was based on the values from 263 anions.

A small deviation of the results (concerning the sequence within the subcluster containing the systems 6 to 10) is not relevant in practice, because it was found—at least for aqueous systems—that the combination of more than three electrolyte systems out of the properly chosen subclusters does not enhance the identification power of the electrophoretic method significantly (10).

It can be concluded from the results derived that the aqueous electrolyte systems are characterized equivalently by the reduced set of data (obtained from 55 solutes) compared to the larger data set (obtained from 263 solutes).

Similarity of Electrolyte Systems with Water or Methanol as Solvents. In order to evaluate the similarities of the electrolyte systems with water or methanol as solvents, again the Euclidian distances (based on the R_E values) were used as a measure of resemblance. The resulting similarity matrix is given in Table IV. The corresponding dendrograms were calculated with an average linkage and with a single linkage algorithm. They are shown in Figure 2. It can be seen that both dendrograms are well structured; they consist of subclusters with high similarity of the units on the one hand, e.g. with Δ values larger than 0.9 for the groups {7,8,6,9} and {7*,8*}, but also connections of the main subclusters at low Δ values (less than 0.2 or 0.5, respectively), on the other hand. Both algorithms connect, however, two different groups on the last level of agglomeration. The average linkage algorithm leads to two main subclusters, both consisting of aqueous as well as methanolic systems: {3,4,7*,8*} and {7,8,6,9,10,5,9*,10*}. From the single linkage algorithm the methanolic systems 9* and 10* are found together in one and the other aqueous and methanolic systems in another large subcluster. (It should be mentioned that the single linkage algorithm seems to be the least successful in many cases, compared with other hierarchic clustering algorithms (19).)

If electrolyte systems out of three subclusters were selected in a further step, identical results are obtained with both algorithms: the three subclusters consist of the systems {9*,10*}, {7*,8*,3,4}, and {5,6,7,8,9,10}. From these results it can be seen that the electrolyte systems with the lowest pH values in both solvents—{3,4} and {7*,8*}—behave similar. The aqueous systems with pH values from 5 to 10 are very similar, too, which is in accordance with our previous results. The methanolic systems 9* and 10* are forming separate clusters in any case.

If these results are used to select combinations of three electrolyte systems are proposed (10), one system out of each of the three following groups must be chosen: {3,4,7*,8*}, {9*,10*}, and {5,6,7,8,9,10}. From this result it can be concluded that the application of methanol as solvent of the buffering electrolyte systems leads indeed to an improvement

of the selectivity in the electrophoresis of anions.

It should be mentioned that a variation of the extrapolated R_E values (see Experimental Section) even by 20% did not influence the structure of the clusters.

Plausibility of the Results. The observation of a structure in a set of data by cluster analysis does not imply, per se, that this structure originates by any real relation between the distinct attributes. The resulting structure which is found always has to be critically interpreted based on plausible physicochemical relations.

If this critical sight is applied to the present case, it can be assumed that the most important influence on the changes of the electrophoretic behavior of the analytes is caused by the degree of dissociation. It determines in a very pronounced way the effective mobility, m_i^{eff} , which depends on the actual mobility m_i (the mobility of the fully dissociated species of the analyte, i , at the given concentration), the pH value of the sample zone, and the pK_a value of the analyte, given by the relation $m_i^{\text{eff}} = m_i / (1 + 10^{pK_a - \text{pH}})$.

It is well-known that organic solvents can influence the dissociation constants of neutral acids of the type HA, often by several orders of magnitude (e.g. ref 2 and the literature cited there). Addition of solvents like methanol, acetonitrile, or dimethyl sulfoxide to water increases the pK_a values of these acids. In fact, an increase of the pK_a values of the analytes is found in methanol compared with water, as shown in Figure 3. It can be seen that the pK_a values in water range from about -3 to +5, whereas the corresponding range in methanol is 3 to 4 pK units higher. The extent of this increase is not identical, considering the particular analytes, but fluctuates within a certain range. This different influence is reflected by the relatively low value of 0.911 for the linear correlation coefficient for the correlation between the pK_a values in methanol and water, respectively.

The fact that the main part of the analytes have pK_a values of about 3–5 in water and 6.5–9 in methanol (Figure 3) supports the plausible explanation of the results about the similarities of the electrolyte systems obtained by cluster analysis. The most pronounced variations of the effective mobilities must be expected in buffering systems whose pH values are in the pK_a region of the main part of the analytes. The systems 3, 4, 7*, and 8* are found indeed in the same subcluster. It can be assumed that in pure aqueous electrolyte systems with pH values larger than 5 the variation of the pH value has no relevant influence on the effective mobilities: indeed systems 5 to 10 are found in the same subcluster. In a first view a similar argument should be valid for systems 9* and 10*. These systems are not found, however, to be closely related with systems 5 or 6 as expected. A plausible explanation for this deviation can be derived from the different solvation effects of water and methanol on the anions. These

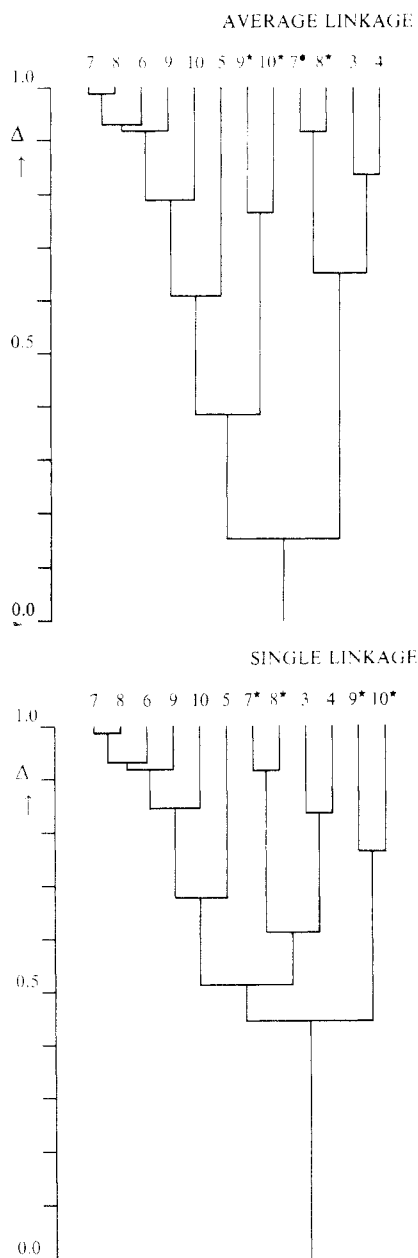


Figure 2. Dendrograms as the results of the average linkage and the single linkage algorithm. The calculations were based on the R_E values of 55 anions in aqueous and methanolic electrolyte systems. The symbols of the systems are according to Table I. The methanolic systems are indicated by asterisks. Δ = normalized distance.

different effects can additionally lead to decisive changes of the actual mobilities in both solvents. This assumption is supported by the evaluation of the correlation between the (absolute) mobilities of the (obviously totally dissociated) analytes in water and methanol, respectively (14, 15); a low value of 0.882 is obtained for the linear correlation coefficient. This result reflects that different solvation leads to relevant differences of the electrophoretic properties of the analytes in the buffering systems of higher pH values. Thus, the dissimilarity between the corresponding systems in water and methanol can be plausibly explained.

Similarity of Systems Characterized by Attributes Based on Random Numbers. One has to decide from case to case whether the variables of the data units show clustering tendencies that reflect meaningful relations of the systems. The problem of the validity of the structures of clusters is treated in the mathematical literature in several ways (17–22). Nevertheless there is no straightforward algorithm, which delivers a decision criterion on the validation of the resulting

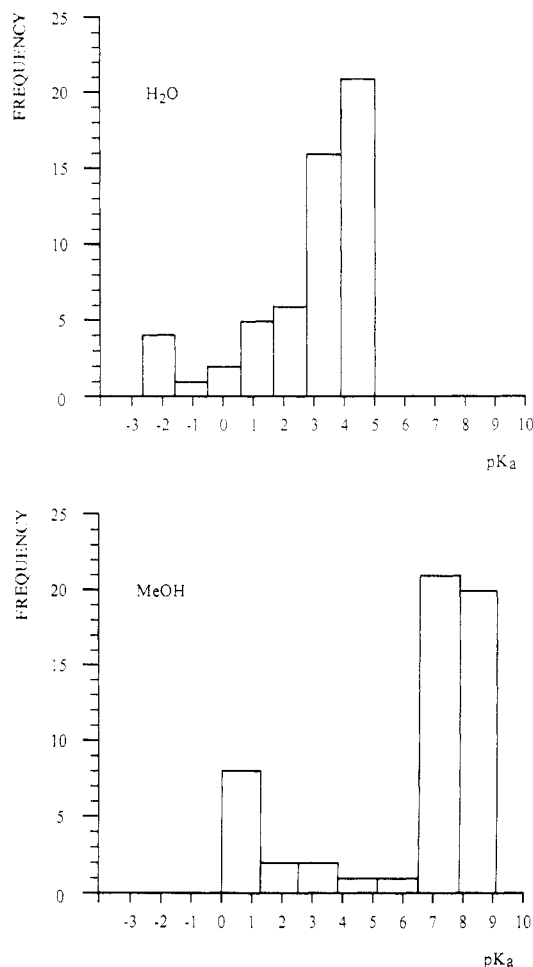


Figure 3. Distribution of the pK_a values of the solutes in water and methanol.

structures of clusters, comparable, e.g., to the criteria for the null hypothesis in statistics. Therefore the problem of the aggregation of entities into clusters was investigated for the case that these entities or units are defined by variables which have randomly chosen values. For this approach the taxonomic units were characterized by a series of random numbers, based on a normal distribution, and a Poisson distribution, respectively. These types of distributions were selected because they often occur in the statistical treatment of analytical chemistry.

Values of 2.58 and 5.0 were taken for the mean and for the standard deviation, respectively, of the normal distribution for the random numbers, since the same values were found taking all relative mobilities into account. For the same reason, the Poisson distribution was characterized by a value of 2.58 for the parameter λ . Both distributions of the random numbers led to nearly identical results. Therefore the discussion is focused to the clusters derived from the Gaussian distribution.

Cluster analysis was applied based on different numbers, N , of attributes. N covers the range from $N = 5$ to $N = 500$ for each of the 12 units. Small numbers for N were also considered, because in analytical investigations often only relatively few experimental data are available.

The results of the clustering procedure, based on the average linkage algorithm, are shown in Figure 4, where the units are symbolized in the dendrograms by the letters A to L. One can see very well structured clusters derived from $N = 5$ variables. Even from $N = 10$ variables a well-structured cluster is obtained, although random numbers are the source of the interrelation of the data. From 50 variables on, the dendrograms lead to less structured images, where the par-

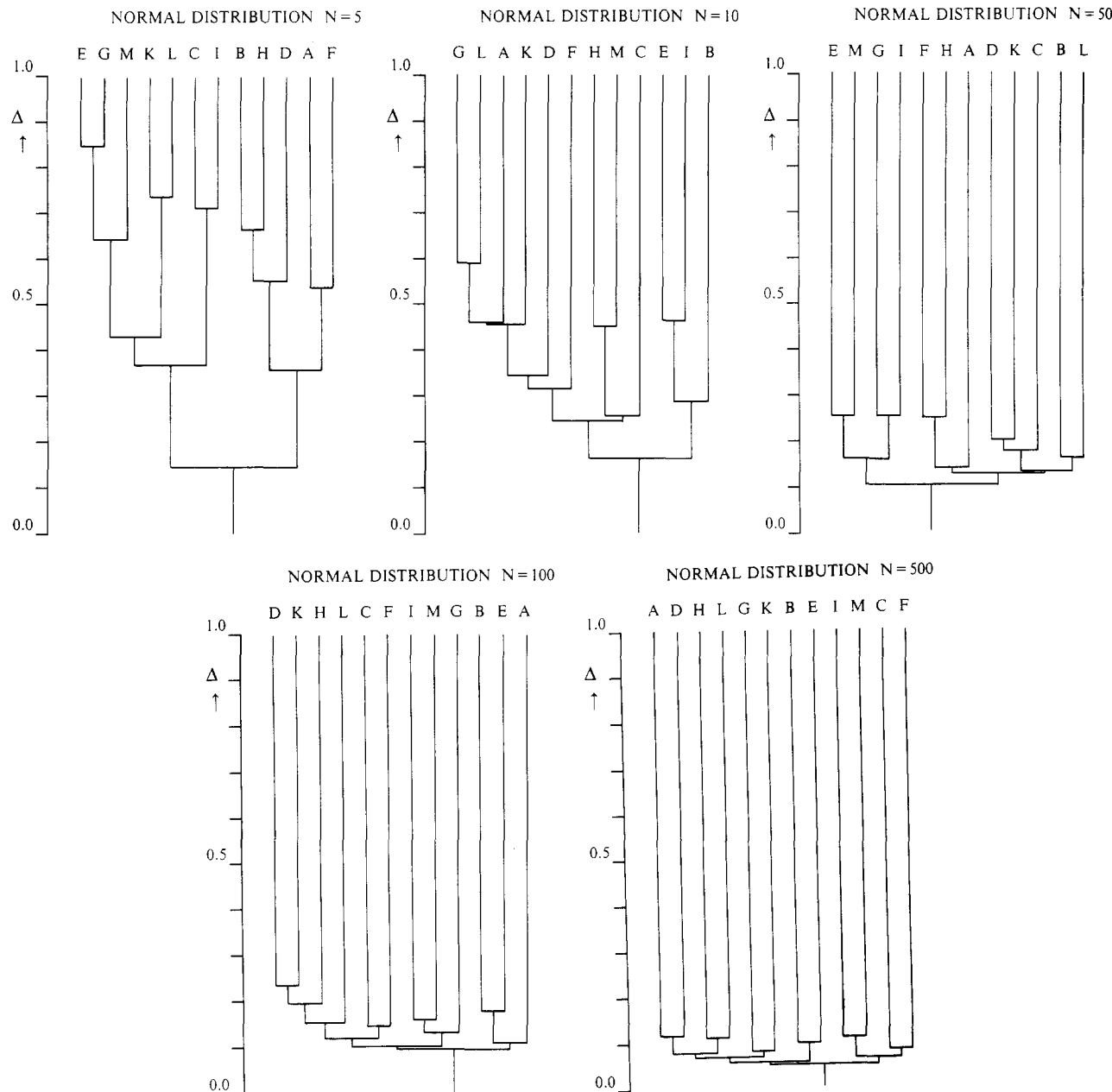


Figure 4. Dendrograms illustrating the similarity of taxonomic systems (A to L) characterized by randomly chosen values of the features. The random numbers were generated according to a normal distribution. N = number of taxonomic attributes per system. Δ = normalized distance.

ticular pairs of systems agglomerate at a low level of similarity (Δ less than 0.3). This picture is enforced for high numbers of variables where the dendrograms approach the image of the cluster, which is obtained, when the data are in fact totally uncorrelated, and where all systems would be linked together at a Δ level of zero.

On comparison of the clusters obtained from $N = 50$ variables with the randomly chosen values in Figure 4 with that formed by the R_E values of the electrolyte systems shown in Figure 2, which was derived from a similar number of attributes, two totally different images can be observed. All former units are highly dissimilar (the maximum value for Δ is smaller than 0.3). In contrast, the clusters obtained from the electrolyte systems show pairs with high similarities on the one hand—(7,8), (7*,8*), etc.—and units aggregating at low values on the other hand. It can therefore be assumed that the similarities between the electrolyte systems, which are elucidated by cluster analysis, are indeed based on their physicochemical properties as discussed above and not by random characters.

Our results show that the structures obtained by clustering

procedures must be interpreted very carefully, because high similarities can originate even from a random character of the attributes, especially when the procedure is based on only few values of the attributes.

LITERATURE CITED

- (1) Everaerts, F. M.; Beckers, J. L.; Verheggen, Th. P. E. M. *Isotachopheresis; Theory, Instrumentation and Applications*; Elsevier: Amsterdam, 1976.
- (2) Kenndler, E.; Jenner, P. J. *Chromatogr.* **1987**, *390*, 169.
- (3) Wielders, J. P. M. Thesis, Eindhoven, 1978.
- (4) Koval, M.; Kaniansky, D.; Hutta, M.; Lacko, R. J. *Chromatogr.* **1985**, *325*, 151.
- (5) Massart, D. L.; Dijkstra, A.; Kaufman, L. *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*; Elsevier: Amsterdam, 1978.
- (6) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: a Textbook*; Elsevier: Amsterdam, 1988.
- (7) Massart, D. L.; De Clercq, H. In *Advances of Chromatography*; Giddings, J. C., Keller, R. A., Eds.; Marcel Dekker: New York, 1978; Vol. 16.
- (8) Kaufman, L.; Massart, D. L. In *Chemometrics, Mathematics and Statistics in Chemistry*; Kowalsky, B. R., Ed.; Reidel: Dordrecht, The Netherlands, 1984.
- (9) Kenndler, E. *Anal. Chim. Acta* **1985**, *173*, 239.
- (10) Kenndler, E.; Reich, G. *Anal. Chem.* **1988**, *60*, 120.

- (11) Sneath, P. H. A.; Sokal, R. R. *Numerical Taxonomy*, Freeman: San Francisco, CA, 1973.
- (12) Anderberg, M. R. *Cluster Analysis for Applications*; Academic Press: New York, 1973.
- (13) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1982.
- (14) Hirokawa, T.; Nishino, M.; Aoki, N.; Kiso, Y. *J. Chromatogr.* **1983**, *271*, D1.
- (15) Hirokawa, T.; Tsuyoshi, T.; Kiso, Y. *J. Chromatogr.* **1987**, *408*, 27.
- (16) Kenndler, E.; Schwer, Ch.; Jenner, P. *J. Chromatogr.* **1989**, *470*, 57.
- (17) Adamson, G. W.; Bawden, D. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 204.
- (18) Dubes, R.; Jain, A. K. *Pattern Recognition* **1979**, *11*, 235.
- (19) Panayirci, E.; Dubes, R. *Pattern Recognition* **1983**, *16*, 433.
- (20) Zeng, G.; Dubes, R. C. *Pattern Recognition* **1985**, *18*, 191.
- (21) Willett, P. J. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 78.
- (22) Jain, N. C.; Indrayan, A.; Goel, L. R. *Pattern Recognition* **1986**, *19*, 95.

RECEIVED for review August 8, 1989. Accepted November 14, 1989.

Experimental Verification of Parameters Calculated with the Statistical Model of Overlap from Chromatograms of a Synthetic Multicomponent Mixture

Scott L. Delinger and Joe M. Davis*

Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois 62901

A synthetic mixture of 56 hydrocarbons was prepared and partially resolved by capillary gas chromatography to experimentally test the predictions of the statistical model of overlap. Various statistical parameters characteristic of both the mixture and the chromatographic process, including the expected number of detectable mixture components and the expected numbers of singlet and multiplet chromatographic peaks, were theoretically estimated from several dozen chromatograms of this mixture by using two procedures. These theoretical parameters were then compared to their experimental counterparts, which were known for the synthetic mixture. Excellent agreement between experiment and theory was found at low chromatographic saturations. This agreement affirms earlier predictions based on this model and supports the argument that peak overlap can be addressed in some instances by theoretical means. At higher saturations, systematic departures of experiment from theory were found. In particular, the experimental numbers of singlet and multiplet peaks are smaller and larger, respectively, than those predicted, when the peak capacity is less than roughly twice the number of components.

INTRODUCTION

A simple statistical theory was developed some years ago (1), in parallel with the work of others (2, 3), to quantify the severity of peak overlap in multicomponent chromatograms. This theory was restricted to the limited but surprisingly common case in which mixture components elute in a random manner from the chromatographic column. The principal conclusion drawn from this theory was that overlap in such chromatograms is surprisingly high and that the number of peaks (especially singlet peaks) is considerably smaller than intuitively expected. A procedure was then suggested by which several statistical parameters characteristic of both the mixture and the chromatographic process could be estimated directly from the chromatogram. These parameters include the expected number of detectable components in the mixture and

the expected numbers of singlet and multiplet peaks in the chromatogram. The validity of this procedure, or slightly modified versions thereof, was verified in several groups by detailed analyses of computer-simulated chromatograms, for which these statistical parameters were known (4-8). These procedures have since been used to estimate these and other statistical parameters from a limited number of experimental gas (6, 9-11) and liquid (8) chromatograms of petroleum, environmental, and natural-product mixtures. Although these parameters were not verified by other types of measurement (and indeed most could not have been), by and large they were consistent with those calculated from computer-simulated chromatograms. The original theory has been reinterpreted (12) and extended by others, particularly by Martin, who has proposed an analogy between the models of peak overlap and polymer degradation (13), developed a theory of peak overlap applicable to chromatograms of both simple and complex mixtures (14), and significantly extended the working range over which statistical parameters can be reliably evaluated from multicomponent chromatograms (15).

Although these studies affirmed that peak overlap is a serious problem in multicomponent chromatograms, our opinion is that the practical implications of this problem are not fully appreciated. One serious implication is that determinations based on the measurement of peak area or height in such chromatograms can often be erroneously high, because one fails to account for the likelihood of overlap. Such errors may have serious consequences in legal issues (11). Similar problems are found in the preparative chromatography of complex mixtures. Unless one deals with simple mixtures, the likelihood of isolating a pure component is vanishingly small, and contamination to some degree is almost inevitable.

One reason that the problem of overlap is not fully appreciated may well be the dearth of experimental evidence directly supporting the predictions of theory. Skeptics can argue that previous studies of overlap in simulated chromatograms are only partially relevant to our understanding of overlap in real-world chromatograms, which behave far less ideally than the simulated ones. They can also point out that the true values of parameters previously estimated by theory from experimental chromatograms were unknown and that theory and experiment could consequently not be compared.

* Author to whom correspondence should be addressed.