Journal of Chromatography, 162 (1979) 495—506
Biomedical Applications

CHROMBIO. 287

APPLICATION OF PATTERN RECOGNITION AND FEATURE EXTRACTION TECHNIQUES TO VOLATILE CONSTITUENT METABOLIC PROFILES OBTAINED BY CAPILLARY GAS CHROMATOGRAPHY

MICHAEL L. McCONNELL*, GERALD RHODES, URSULA WATSON** and MILOS NOVOTNÝ***

Department of Chemistry, Indiana University, Bloomington, Ind. 47401 (U.S.A.)

(First received June 14th, 1978; revised manuscript received November 6th, 1978)

SUMMARY

The applicability of threshold logic units, a form of nonparametric pattern recognition, to the processing of metabolic profile data obtained by high-efficiency glass capillary column gas chromatography has been investigated. The test data included profiles of the volatile constituents of urine from normal individuals and from individuals with diabetes mellitus. A feature extraction algorithm allowed for dimensionality reduction and indicated the constituents most important in the normal versus pathological distinction. With an optimum number of dimensions, a normal versus pathological prediction rate of 93.75% was achieved. Gas chromatography—mass spectrometry was utilized to identify important profile constituents.

INTRODUCTION

Multicomponent chromatographic analyses have been developed for several of the classes of chemical constituents in various human physiological fluids [1—18]. Such analyses, generally termed metabolic profiles [1—2], provide qualitative and quantitative data that reflect the state of a variety of metabolic processes within the body. The primary utility of metabolic profiling is recognized as being in the study of the etiology of diseases through the biochemical elucidation of physiological and pathological processes. Eventually, selected profiling techniques may also serve as clinical screening or diagnostic aids.

Although the sampling and analytical aspects of certain profiling techniques are sufficiently refined for routine research use, biomedical correlations of pro-

*Present address: Chromatix, Inc., 560 Oakmead Pkwy, Sunnyvale, Calif. 94086, U.S.A.
**Present address: Medical Sciences Program, Indiana University, Bloomington, Ind. 47401, U.S.A.
***To whom correspondence should be addressed.

file data are rare. Several problems have limited the extraction of useful information from the complex profile data. Aside from pathological alterations, profiles vary considerably due to factors such as individual diet and genetic conformation. These individual variations can be of the same magnitude as the pathological variation. Ascertainment of the latter variations thus requires comparison of large sets of normal and pathological profiles, such that the nonpathological variations are effectively equalized in both pattern sets. Reliable manual comparison of large sets of profiles containing more than 200 constituents is simply not feasible. The alternative process of preselecting profile constituents for comparison, based on their probable metabolic activity, is not always desirable. Since much of metabolism is not well understood, preselection may discard potentially important information. Furthermore, preselection requires previous identification of essentially all profile constituents, for which the only practical technique is gas chromatography—mass spectrometry (GC—MS). Due to the small sample quantities involved, and the general lack of appropriate reference spectra, GC—MS identifications of volatile constituents are frequently arduous. Thus, a data analysis technique is required that must first be capable of distinguishing profiles of individuals in a particular diseased state from those of normal individuals. It must not require initial assumptions concerning the statistical distribution of the data. Secondly, but quite importantly, the analysis must indicate what components of the profile data are involved in the normal versus pathological distinction. The distinctive metabolites may then be identified, and their metabolic precursors can be determined with loading tests or other appropriate methodology.

In this report, we describe the development of threshold logic unit (TLU) techniques, a form of nonparametric pattern recognition [19], for the processing of the metabolic profile data. TLUs have been trained to distinguish normal and pathological patterns and, in combination with feature extraction algorithms, have designated the pattern components most significant in the distinction. Since TLU techniques function independent of assumptions of metabolic significance, preselection is not required, and identification need only be obtained for those compounds found to be significant in the normal versus pathological distinction.

Nonparametric pattern recognition techniques have been applied to several biomedical problems [20, 21] and their numerous applications in the chemical field have been recently reviewed [22, 23]. However, their prior utilization with multicomponent chromatographic data has been limited to studies of petroleum sample type identification [24], Wilcoxon-test correlations of human urinary amino acid analyses and initial correlations of volatile constituent analyses [6, 25].

In this study the TLU procedure has been tested with urinary volatile constituent profiles obtained by high-resolution (glass capillary) gas chromatography. Volatile constituent profiles are of interest [7—18] due to their complexity, frequently exceeding 200 detectable compounds, and due to their inclusion of by-products, intermediates, and terminal products of a wide variety of metabolic processes. The pathological condition chosen for testing purposes was diabetes mellitus, primarily because of sample availability. While the biochemical information extracted from the profiles in this study may be of some utili-

ty, the primary objective has been the development of a generally applicable methodology.

MATERIALS AND METHODS

Headspace sampling and chromatographic separation

The procedures and instrumentation utilized for the acquisition of the metabolic profiles have been previously described [16]. Volatiles in the heated (100°) urine headspace were adsorbed onto 2 mg of 2,6-diphenyl-*p*-phenylene oxide porous polymer (Tenax GC; Applied Science Labs., State College, Pa., U.S.A.). The porous polymer was contained in platinum microbaskets which were subsequently encapsulated for injection.

All chromatographic functions were automated [16]. Reproducibly prepared glass capillary columns [16] (60 m × 0.29 mm I.D.) coated with GE SF-96 methylsilicone fluid were employed. The column effluent was split between a flame ionization detector (FID) and a nitrogen-sensitive thermionic detector. Only the FID data were incorporated into the pattern recognition studies; the nitrogen-sensitive detector data were used in the identification of profile constituents.

In the only modification of prior procedures, a reference for relative retention time calculations was provided by the addition of two internal standards to the physiological fluid prior to headspace sampling. Full scale peaks (corresponding to approximately 25 ng by direct injection) resulted from the introduction of 0.3 μg of 6-undecanone and 20 μg of 10-nonadecanone in 5 μl acetone. The standards were added to the urine samples immediately prior to the transferring of them into the sampling vessel. As acetone elutes in the first peak of the profile and has a short retention time on the Tenax adsorbent, it had no adverse effect on the chromatograms.

Sample collection

Diabetic urine samples were collected from 29 individuals, including outpatients and patients at two hospitals. Of these, 15 were diagnosed as diabetics with further complications. Normal samples were obtained from 35 individuals, including those hospitalized for 24 h solely for the sample collection. Medical histories were maintained, but samples were not excluded on this basis. No diet regulation was involved.

Twenty-four hour urine samples were collected; interim samples were frozen over dry ice. The samples were then brought to room temperature, filtered, divided into 50 ml aliquots and refrozen at −20°. No preservatives or diluents were added.

Data analysis

For use in TLU processing, each metabolic profile was represented as a pattern vector of the form:

$$\vec{X} = (x_1, x_2, \ldots\ldots x_n)$$

where each vector component, x_i, was calculated from peak areas in the profile

chromatogram. Using a training set of known patterns, representing normal and pathological profiles, an n-dimensional hyperplane, or classification surface, was developed which classified, or linearly separated, the patterns into the two known groups. The classification criteria were:

$\vec{W} \cdot \vec{X} < 0$ for $\vec{X}$ representing normal profiles

$\vec{W} \cdot \vec{X} > 0$ for $\vec{X}$ representing pathological profiles

where $\vec{W}$ is the weight vector, a normal vector from the classification surface. The TLU training procedures are quite simple and have been described elsewhere [19]. Once developed, the validity of the weight vector was tested by predicting, according to the above criteria, the classification of patterns not present in the training set.

In any pattern recognition process, it is important that the ratio (R) of the number of training patterns to the number of pattern dimensions be as large as possible; cases where $R \leqslant 2$ must be carefully interpreted [26]. As described below, a leave-out-one algorithm was employed in this study to allow operation with an initial $R \leqslant 2$. In addition, to control the pattern dimensionality, each pattern component was calculated as the sum of the peak areas in a small interval of the profile chromatogram. Such combining of peaks yields an apparent loss of information, in that differences in more than one peak in an interval may partially cancel, and observed differences cannot necessarily be assigned to a single peak. In fact, many intervals contained only one peak. Further, once an interval was found to be significant, it was subdivided and re-evaluated. This process can be continued until the significance can be assigned to a single constituent.

Initially, patterns of 100 dimensions were used. The section of the chromatogram between injection and the elution of the 6-undecanone internal standard was divided into 50 intervals of equal length in time. Similarily, the section between the elution of 6-undecanone and the 10-nonadecanone internal standard was divided into 50 uniform intervals. The 10-nonadecanone peak was the last to elute in nearly all profiles. The program located the two internal standards by examining an absolute retention time "window" (2 min for 6-undecanone and 3 min for 10-nonadecanone) and designating the largest peak in the "window" as the internal standard. For this study, the designation was also manually verified; no errors occurred.

Prior to the TLU training procedure, the pattern components were autoscaled [23] according to the formula:

$$x'_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i}$$

where x'_{ij} is the autoscaled i^{th} component of the j^{th} pattern, x_{ij} is the initial component, μ_i the mean and σ_i the standard deviation of the i^{th} component in all patterns. The autoscaling operation equalizes the initial weighting of the intervals. Autoscaling improved the prediction rate of the subsequently trained

TLUs by as much as 15%. A normalization procedure, in which each component peak area was expressed as a fraction of the total chromatogram peak area according to:

$$x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^{100} x_{ij}}$$

was found to reduce prediction rates, both with and without autoscaling [27].

Training and testing of TLUs was executed with the 100-dimensional patterns and reduced dimension patterns derived therefrom. In each case, training was performed with a leave-out-one algorithm [28] operated in a rotating basis. In the leave-out-one algorithm, all available patterns except one were used to train the TLU. The validity of the TLU was then tested by predicting the classification of the omitted pattern. The pattern set was then rotated to omit a different pattern from the training set, and the training and prediction were repeated. The rotation was repeated until each pattern had served as the prediction test pattern. The percentage of the test patterns that was correctly predicted was recorded for each cycle. A feature extraction algorithm, to designate pattern components most significant in the classification, was developed for the leave-out-one technique. Since the pattern components were autoscaled, the components of the weight vector were indicative of the significance of the respective components. An average weight vector was determined for a complete rotation cycle by calculating the vector sum of each of the individual weight vectors trained. At the condition of a rotation, the magnitude of the components of the average weight vector where thus indicative of the average significance of each pattern component. A preselected number of the least significant components was then discarded and another rotation cycle was initiated. The program was continued until the training process was unable (in 500 iterations) to find a hyperplane that linearly separated the profile classes.

The leave-out-one algorithm offers two significant advantages, as compared to the use of discrete training and prediction pattern sets. First, it yields a maximum value of R for any available data set. Second, it facilitates operation with $R \leqslant 2$, provided the ratio of the number of patterns to the number of intrinsic dimensions is greater than two. The intrinsic dimensions are defined [29] to be those dimensions that contain information significant in the classification process. In standard training studies with $R \leqslant 2$, a weight vector may be obtained that provides a linear separation of the pattern classes, but has no valid predictive capability. Such a separation may be based on the non-intrinsic dimensions. The non-intrinsic dimensions can be regarded as noise mixed with the intrinsic dimensions. The leave-out-one technique, in calculating an average weight vector, effectively filters out the non-intrinsic dimensions in a manner analogous to ensemble averaging, thus allowing the extraction of the intrinsic pattern components.

Peak areas and retention times were recorded with a commercial chromatographic data system (Model PEP-2; Perkin Elmer, Norwalk, Conn., U.S.A.). All subsequent processing was performed on a CDC 6600 computer. The training

and autoscaling programs were developed in part from the program set ARTHUR, created by Kowalski and Drewer [30]. Our programs are available upon request.

Gas chromatography—mass spectrometry

Volatile urine constituents were sampled as described above, except that the Tenax adsorbent was contained in a glass precolumn, rather than a platinum microcapsule. Details of this sampling technique have been previously reported [14]. Electron-impact spectra were obtained with a gas chromatograph-mass spectrometer (Hewlett-Packard Model 5980A). The glass capillary column was interfaced to the ion source with an all-glass, single-stage jet separator (Scientific Glass Engineering, Melbourne, Australia), maintained at 270°. Identifications were confirmed by comparison of spectra and retention indices of authentic samples recorded with the same instrument.

RESULTS AND DISCUSSION

Two training sets were prepared. One included the entire set of 35 normals and 29 diabetics (Data Set A), the other included the normals and only the 14 diabetics diagnosed to be free of complications (Data Set B). The results of TLU training and feature extraction with these two data sets are presented in Table I. Features were extracted at the rate of five or ten per cycle, as indicated in the Table.

At the optimum number of dimensions, 92.2% of the patterns in Data Set A and 83.7% of Data Set B were correctly predicted. These percentages are significantly greater than the 54.7% and 71.4%, respectively, that would be obtained by simply classifying each pattern as a normal. With the leave-out-one algorithm, satisfactory results were obtained despite the low ratio of patterns to initial dimensions. The increase in prediction with decreasing number of features substantiates the importance of the intrinsic dimensionality. As the insignificant dimensions were discarded, there was less chance of a change in an insignificant constituent adversely affecting the pattern classification. The improved prediction with increasing training set size was expected, as the weight vector was exposed to a more complete subset of the possible profiles and was thus more likely to be trained for the test data.

To utilize the information obtained by feature extraction, it is necessary to associate the retained dimensions with the original profiles. In Fig. 1, a diabetic profile (A) and a normal profile (B) are shown together with scale markings indicating the retained intervals in the test of 35 normals and 14 diabetics (Intervals Retained A) and in the test of 35 normals and 14 diabetics (Intervals Retained B). To describe the two chromatograms as typical of normals and diabetics would falsely imply that the pattern chromatograms can be generalized. However, the chromatograms shown are not unlike others obtained from such individuals.

The components of the average weight vector and the original pattern dimensions corresponding to the retained intervals are given in Table II. The initial assignment of patterns designated a negatively valued dot product of the pattern and weight vector to be normal, and a positive dot product to be

TABLE I

TLU PREDICTION RATE FOR DATA SETS A AND B WITH EITHER 10 OR 5 FEATURES EXTRACTED ON A ROTATION CYCLE

Data Set A		Data Set B	
No. dimensions retained	Correct (%)	No. dimensions retained	Correct (%)
100	67.2	100	57.1
90	70.1	90	57.1
80	71.9	80	57.1
70	71.6	70	69.4
60	79.7	60	71.4
50	81.3	50	73.5
40	79.7	40	77.6
35	84.4	35	81.6
30	79.7	30	79.6
25	84.4	25	81.6
20	82.8	20	83.7
15	92.2	15	83.7
10	DNC*	10	78.7
		5	DNC*

*DNC = Did not converge after 500 training iterations, i.e., could not linearly separate the training set data.

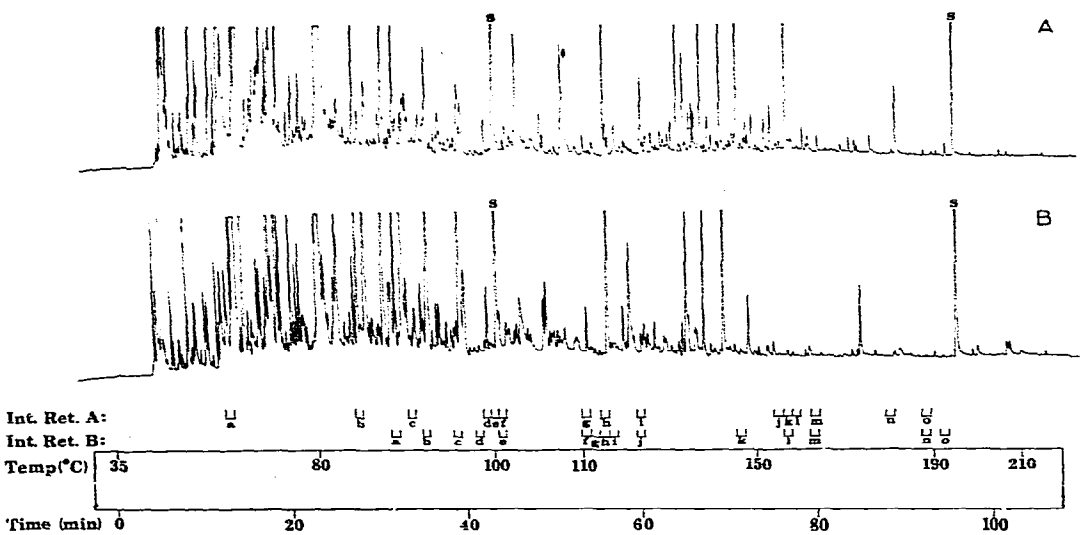

Fig. 1. (A) Chromatogram of the urinary volatiles of a diabetic male. (B) Chromatogram of the urinary volatiles of a normal female. Scale markings indicate intervals selected as significant in the distinction of normal and diabetic profiles, as reported in Tables III and IV. S= internal standard peaks.

TABLE II

COMPONENTS OF AVERAGE WEIGHT VECTOR AT 15 RETAINED DIMENSIONS AND IDENTIFIED COMPOUNDS FOR DATA SETS A AND B

Data Set A			Data Set B		
Interval*	Component	Identified component	Interval**	Component	Identified component
a—15	2.03	4-Heptanone	a—37	- 9.39	
b—32	- 9.68		b—41	- 9.24	
c—39	-12.96		c—45	5.51	Carvone
d—49	-13.27		d—48	7.96	
e—50	6.41	Indole	e—51	9.18	
f—51	6.56		f—60	4.90	
g—60	7.19		g—61	-10.82	
h—62	8.59		h—62	7.76	
i—66	- 5.78		i—63	4.29	
j—81	7.34		j—66	- 9.18	
k—82	11.56		k—77	9.18	
l—83	- 7.97		l—82	16.33	
m—85	7.19		m—85	6.12	
n—93	7.66		n—97	5.51	
o—97	7.03		o—99	- 8.57	

*Fig. 1 Int. Ret. A.

**Fig. 1 Int. Ret. B.

diabetic. Because the pattern components were autoscaled, a given pattern component may have either a negative or positive value. Thus, a peak of greater than average area that occurs in an interval with a negative weight vector component is indicative of a normal profile; a peak of less than average area is indicative of a diabetic profile. The converse is true for positive weight vector components. This consideration comprises an important aspect of autoscaling. Without this pre-processing, all pattern components would have positive values, and the presence of a given peak could indicate only one condition, regardless of the size of the peak. Thus, autoscaling is advantageous when significant peaks are present in both classes, but with differing areas. Autoscaling would be disadvantageous when the simple presence of certain peaks is indicative of the profile class.

In several intervals (15, 32, 39, 41, 49, 50, 63, 77, 82, 85, 93) a peak may be recognized in the chromatograms in Fig. 1 that varies in area as predicted by the sign of the weight vector components. Confirmation of the results in other intervals required examination of several more chromatograms. It is of interest that many of the significant intervals fall in the latter portion of the chromatogram and thus involve urinary constituents not observed in the previous volatile profile studies. Further, many of these intervals involve only small peaks which might be neglected in manual data analysis. The selection of several sets of adjacent intervals may simply reflect the locality of several important peaks, or may arise from certain significant peaks shifting between intervals due to

very slight variations in relative retention times.

Due to the interaction and frequent nonspecificity of metabolic processes, the information in certain pattern dimensions may be diagnostically equivalent. Such redundance may result in none of the dimensions achieving a high significance in the classification until one or more is discarded. Rapid feature extraction may discard a set of redundant dimensions in a single cycle, eliminating the information completely. This is demonstrated by the improved prediction rates obtained by slow feature extraction reported in Table III. Ten features were extracted per training cycle for five cycles, and one per cycle thereafter; results are given only for selected cycles. A trade-off between computation costs and classification results is also evident.

Another example of redundance may be observed in Fig. 1. Seven intervals (51, 60, 62, 66, 82, 85, 97) are clearly significant, in that they are retained in both weight vectors. The information content of the other intervals is likely to be redundant, with respect to either the intervals retained in both vectors or with respect to other periodically retained intervals. Because of this redundance, the significance of the retained and discarded dimensions must be carefully interpreted. While profile constituents in a retained dimension are clearly important in the pattern classification, a discarded dimension may also contain constituents involved in the aberrant metabolism.

To investigate further the redundance in the profile data and the effect of increasing the value of R, TLUs were trained and tested with abbreviated patterns. Data sets were prepared that contained only twenty intervals (1—20, 21—40, etc.). On the first training cycle, only the set composed of intervals 41—60 was linearly separable. Feature extraction resulted in a maximum pre-

TABLE III

TLU PREDICTION RATE FOR DATA SET A

With 10 features extracted each of first 5 rotations and 1 feature extracted each rotation thereafter.

No. dimensions retained	Cycle No.	Correct (%)
100	1	67.19
70	4	76.56
50	6	81.25
35	21	81.25
26	30	90.63
20	36	90.63
18	38	93.75
15	41	92.18
13	43	89.06
10	46	89.06
7	49	90.63
6	50	DNC*

*DNC= Did not converge after 500 training iterations.

diction rate of 82.8% with 13 of the original 20 dimensions retained. The seven most significant of these dimensions were then combined with the 13 most significant intervals (as determined from the test reported in Table I) of the remaining original 80 dimensions. Training and feature extraction with these 20 dimension patterns resulted in a maximum prediction of 93.8% correct with 11 retained dimensions. All eleven dimensions were included among those listed in Table I.

These results substantiate the previous conclusion that redundance exists within the profile information, but that several dimensions contain essential information. Further, within the range examined, the results are not significantly improved by manually selecting the initial dimensions and thus increasing the value of R.

Referring back to Fig. 1, several retained intervals are seen to contain more than one constituent peak. Thus, the significance of the interval cannot be assigned to a single constituent. To reduce this ambiguity, each of the 15 most significant intervals (Table II) was split into two equal intervals and the resulting 30 dimension patterns were used to train a new TLU. In the case of Data Set A, 17 of the 30 dimensions were retained at an optimum prediction rate of 92.2%, identical to the rate prior to the breakdown. A prediction rate of 91.8% was obtained with Data Set B with 13 of the dimensions retained. The signs of the weight vector components following the split invariably agreed with those corresponding to the same interval prior to the division, although the relative magnitude of the weight vector components did vary. A few of the subdivided intervals contained more than one constituent peak. It would be possible to continue the division until each interval contained only one peak or until small variations in relative retention times precluded the reliable assignment of a peak to a given interval.

The high prediction rates achieved by the TLUs are indicative of their potential as diagnostic or screening aids. However, their biomedical potential is more immediately discernable, and requires the identification of the significant constituents. Several of the constituents found to be significant in the pattern recognition studies have been tentatively identified, including both sulfur- and nitrogen-containing compounds. However, only three of these compounds have been confirmed by comparison to authentic spectra at present.

Once reference spectra can be obtained, the other identifications will be reported. The three identified and confirmed compounds and the corresponding intervals are listed in Table II. The weight vector components of each of these compounds was positive. Thus, increasing concentrations of these compounds were indicative of a diabetic classification. Increasing concentrations of 4-heptanone in the urine of diabetics has been previously reported by Liebich and Al-Babbili [7].

One reservation must be cited concerning the interpretation of the results of the TLU training and prediction. All but one of the diabetic patients were receiving various medications for diabetes and in some cases for other conditions. The one individual not receiving medication was misclassified in both TLU rotations. Contraceptives were the only medication known to have been taken by any of the normal individuals. Thus, the possibility that the pattern distinction was based on metabolites of the medication or compounds induced

by it rather than those related to the disease cannot be excluded. The analysis of more diet-controlled diabetics would clarify the interpretation of the pattern recognition results. However, the capability of the TLU technique to identify the profile components most significant in the class distinction is not diminished by this ambiguity.

CONCLUSIONS

The results of this study indicate the utility of threshold logic unit and feature extraction techniques in the determination of constituents important in the classification of urinary volatile constituent metabolic profiles. It is likely that these techniques can be readily extended to profiles of other complex fractions in which pathological changes cannot otherwise be distinguished from normal individual variations. By increasing the efficiency of the extraction of information from the profiles, pattern recognition techniques should enhance the utility of metabolic profiles in many biochemical studies.

ACKNOWLEDGEMENTS

We thank Drs. R. Robinson, J. Galloway, B. Obermeyer and R. Gurd for their help in obtaining urinary samples of various patients. Valuable comments of Dr. P.C. Jurs concerning the pattern recognition aspects are also acknowledged. This work was supported by the Grant No. PHS RG. GM 23668-01 from the NIGMS, U.S. Public Health Service. One of us (M.L. McConnell) was a recipient of the summer fellowship from the American Chemical Society, Division of Analytical Chemistry, sponsored by Carle Instruments, Inc.

REFERENCES

1 C.E. Dagliesh, E.C. Horning, M.G. Horning, K.L. Knox and E. Yarger, Biochem. J., 101 (1966) 792.
2 E.C. Horning and M.G. Horning, J. Chromatogr. Sci., 9 (1971) 129.
3 C.D. Scott, D.D. Chilcote, S. Katz and W.W. Pitt, Jr., J. Chromatogr. Sci., 11 (1973) 96.
4 M. Novotny, M.P. Maskarinec, A.T.G. Steverink and R. Farlow, Anal. Chem., 48 (1976) 468.
5 E. Jellum, O. Stokke and L. Eldjarn, Anal. Chem., 45 (1973) 1099.
6 A.B. Robinson and L. Pauling, Clin. Chem., 20 (1974) 961.
7 H.M. Liebich and P. Al-Babbili, J. Chromatogr., 112 (1975) 539; H.M. Liebich, J. Chromatogr., 112 (1975) 551.
8 H.M. Liebich, O. Al-Babbili, A. Zlatkis and K. Kim, Clin. Chem., 21 (1975) 1294.
9 A. Zlatkis and H.M. Liebich, Clin. Chem., 17 (1971) 592.
10 L. Pauling, A.B. Robinson, R. Teranishi and P. Cary, Proc. Nat. Acad. Sci. U.S., 68 (1971) 2374.
11 R. Teranishi, T.R. Mon, A.B. Robinson, P. Cary and L. Pauling, Anal. Chem., 44 (1972) 18.
12 A. Zlatkis, W. Bertsch, H.A. Lichtenstein, A. Tishbee, F. Shunbo, H.M. Liebich, A.M. Coscia and N. Fleischer, Anal. Chem., 45 (1973) 763.
13 K.E. Matsumoto, D.H. Partridge, A.B. Robinson, L. Pauling, R.A. Flath, T.R. Mon and R. Teranishi, J. Chromatogr., 85 (1973) 31.
14 M. Novotny, M.L. McConnell, M.L. Lee and R. Farlow, Clin. Chem., 20 (1974) 1105.
15 A. Zlatkis, W. Bertsch, D.A. Bafus and H.M. Liebich, J. Chromatogr., 91 (1974) 379.

16 M.L. McConnell and M. Novotný, J. Chromatogr., 112 (1975) 559.
17 I.R. Politzer, S. Githens, B.J. Dowty and J.L. Laseter, J. Chromatogr. Sci., 13 (1975) 378.
18 E. Stoner, D. Cowburn and L.C. Craig, Anal. Chem., 47 (1975) 344.
19 P.C. Jurs and T.L. Isenhour, Chemical Applications of Pattern Recognition, Wiley-Interscience, New York, N.Y., 1975.
20 A.F. Vagnucci, and C. Li, Compt. Biomed. Res., 4 (1971) 291.
21 R.A. Nordyke, C.A. Kulikowski and C.W. Kulikowski, Compt. Biomed. Res., 4 (1971) 374.
22 P.S. Schoenfeld and J.R. DeVoe, Anal. Chem., 48 (1976) 403R.
23 B.R. Kowalski and C.F. Bender, J. Amer. Chem. Soc., 94 (1972) 5632.
24 H.A. Clark and P.C. Jurs, Anal. Chem., 47 (1975) 374.
25 H. Dirren, A.B. Robinson and L. Pauling, Clin. Chem., 21 (1975) 1970.
26 N.A. Gray, Anal. Chem., 48 (1976) 2035.
27 M.L. McConnell, Doctoral Thesis, Indiana University, Bloomington, Ind., 1977.
28 B.R. Kowalski, in C.L. Wilkins and C.E. Klopfenstein (Editors), Computers in Chemical and Biomedical Research, Vol. II, Academic Press, New York, N.Y., 1974, p. 2.
29 W.S. Meisel, Computer-Oriented Approaches to Pattern Recognition, Academic Press, New York, N.Y., 1972.
30 B.R. Kowalski and D.L. Drewer, Documentation for ARTHUR (Batch), Chemometrics Society Report No. 1, Department of Chemistry, University of Washington, Seattle, Wash.