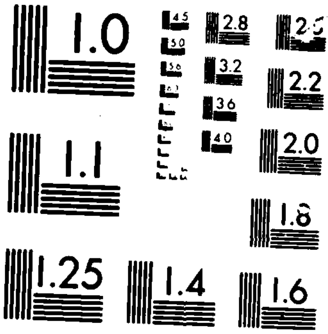
AD-A169 061 APPLICATION OF PATTERN RECOGNITION TO METAL ION CHEMICAL IONIZATION MASS SPECTRA(U) LAWRENCE LIVERMORE NATIONAL LAB CA R A FORBES ET AL. OCT 85 TR-4 1/1

UNCLASSIFIED N00014-86-F-0033 F/G 7/4 NL



MICROCOP CHART

(12)

OFFICE OF NAVAL RESEARCH

Contract No. N00014-86-F-0033

Task No. NR 051-854

Application of Pattern Recognition to Metal Ion Chemical Ionization Mass Spectra

R. A. Forbes, E. C. Tews, and B. S. Freiser*

Department of Chemistry
Purdue University
West Lafayette, IN 47907

M. B. Wise
Oak Ridge National Laboratory
Oak Ridge, TN 37831

S. P. Perone*
Lawrence Livermore National Laboratory
P. O. Box 808
Livermore, CA 94550

Published in Anal. Chem., 58, 684 (1986)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Technical Report No. 4

2. GOVT ACCESSION NO.
AD-A169061

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Application of Pattern Recognition to Metal Ion Chemical Ionization Mass Spectra

5. TYPE OF REPORT & PERIOD COVERED
Interim Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
R. A. Forbes, E. C. Tews, B. S. Freiser, M. B. Wise, and S. P. Perone

8. CONTRACT OR GRANT NUMBER(s)
N00014-86-F-0033

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Lawrence Livermore National Laboratory
P. O. Box 808, L-310
Livermore, CA 94550

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 051-854

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

12. REPORT DATE
October 1985

13. NUMBER OF PAGES
24

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
Published in Analytical Chemistry, 58, 684 (1986)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Pattern Recognition
Artificial Intelligence
Mass Spectrometry
Chemical Ionization Mass Spectrometry

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Pattern recognition methods were used to evaluate the information content of mass spectrometry data obtained using transition metal ions as an ionization source. Data sets consisting of the chemical ionization mass spectra for Fe^+ and Y^+ with 72 organics (representing the six classes alkane, alkene, ketone, aldehyde, ether, and alcohol) and 24 alkanes (representing the three subclasses linear, branched, and cyclic) were subjected to pattern recognition analysis using a k-nearest neighbor approach with feature weightings. The reactivites of Fe^+ and Y^+ toward the classes of compounds studied were characterized using

Technical Report No. 4 (continued)

20. classification accuracies as a measure of selectivity, and important chemical information was extracted from the raw data by empirical feature selection methods. A total recognition accuracy of 81% was obtained for the recognition of the six organic classes and 96% accuracy was obtained for the recognition of the three subclasses of alkanes.

BRIEF

A weighted k-nearest neighbor algorithm is used to analyze and extract chemical information from mass spectral data derived from Fe^+ and Y^+ chemical ionization of 72 organic compounds.

ABSTRACT

Pattern recognition methods were used to evaluate the information content of mass spectrometry data obtained using transition metal ions as an ionization source. Data sets consisting of the chemical ionization mass spectra for Fe^+ and Y^+ with 72 organics (representing the six classes alkane, alkene, ketone, aldehyde, ether, and alcohol) and 24 alkanes (representing the three subclasses linear, branched, and cyclic) were subjected to pattern recognition analysis using a k-nearest neighbor approach with feature weightings. The reactivites of Fe^+ and Y^+ toward the classes of compounds studied were characterized using classification accuracies as a measure of selectivity, and important chemical information was extracted from the raw data by empirical feature selection methods. A total recognition accuracy of 81% was obtained for the recognition of the six organic classes and 96% accuracy was obtained for the recognition of the three subclasses of alkanes.

INTRODUCTION

Electron impact (EI) ionization mass spectrometry has become a standard means for the classification of unknown compounds according to functionality or structure (1-5). The differentiation of isomeric molecules, however, remains a difficult problem and subtle differences in molecular structure often cannot be distinguished by electron impact. The need arises, therefore, for a more selective form of ionization. Chemical ionization (CI) has the potential for such an increased selectivity since it is possible to adjust the reactivity of the CI reagent for selectivity in a way that is not possible for the EI mass spectrometry experiment (6-8).

In our laboratory the reactivities of laser generated transition metal ions toward various types of compounds have been studied for several years (9-13). A major goal of this work has been to evaluate the utility of metal ions as selective reagents for mass spectral identification of the functionality and structure of unknown compounds. In view of the potentially large data matrix generated from the reactions of different metal ions with various organic compounds, the application of pattern recognition techniques provides a particularly useful means for achieving these goals.

Pattern recognition has been applied to a wide variety of chemical problems and numerous reviews on the subject have been published (14-21). Some of its more recent applications include: recognition of organic compounds using Fourier transform infrared spectroscopy, interpretation of gas chromatography data, nuclear magnetic resonance spectral interpretation, and analysis of electrochemical systems (22-32). Many applications of pattern recognition to mass spectral data have recently appeared in the literature. Electron impact ionization is by far the most widely used ionization means and has been employed in studies ranging from the analysis of complex mixtures using gas chromatography/mass spectrometry, to the recognition of steroids, and the use of mass spectrometric data to predict the biological activity of antibiotics (33-36). Pattern recognition has also been applied to the experimental optimization of field-desorption and fast atom-bombardment mass spectrometry and for the location of homoconjugated triene and tetraene units in aliphatic compounds using NO chemical ionization (37,38).

Because pattern recognition is a well established tool for interpretation of mass spectral data, it was the goal of the work described here to use this tool to enhance our understanding of the information content of a new and important advance in chemical ionization mass spectrometry. This was a particularly efficient approach because of the potentially enormous data matrices. It also provided a unique opportunity to apply pattern recognition to an emerging data base where the scientist is near the bottom of the "learning curve." This work illustrates how such an application can enhance the rate of climbing that "learning curve."

One of the goals of pattern recognition is to minimize the number of features required to effect class separation within a data set while maximizing the recognition accuracy through the elimination of features detrimental to class separation. Thus, an empirical feature selection algorithm is often used to map a classification problem down from the space of all features to a space of smaller dimensionality which consists of only important, relevant features. Not only does this procedure enhance the ratio of patterns to features in order to have a statistically valid separation of classes (15), but it also can provide new chemical insight based on the feature set selected to achieve a given informational goal.

In this study the use of two metal ions, Fe^+ and Y^+, as chemical ionization reagents was evaluated. Two different empirical feature selection algorithms were used to extract important features from the data: successive subtraction of features with total recognition accuracy (SSTRA) as the selection criteria which employs constant weighting of the features, and forward addition of features using the nearest neighbor distance error (FANNDE) as selection criteria which performs weighting optimization. A complete description of these algorithms has been given (39). Using the recognition accuracies obtained by these algorithms with the metal ion data, the selectivities of the reagents for six organic functionalities as well as selectivities for linear, branched and cyclic alkanes were evaluated. The analytical utility of Fe^+ and Y^+, alone and used in combination, for the recognition of functionality and structure was also explored. Trends in reactivity have been inferred from the misclassified compounds.

EXPERIMENTAL

INSTRUMENTATION

A home-built capacitance bridge ion cyclotron resonance (ICR) mass spectrometer, under the control of an IBM 9000 laboratory computer (40,41), and a Nicolet prototype Fourier transform mass spectrometer (FTMS 1000) were used to generate the CI mass spectral data (42). Details of the ICR and FTMS experiment have been described elsewhere (39). For most of these experiments the magnetic field strength has been held constant at 0.9 Tesla, and the sample pressure has been maintained at approximately 2 X 10-7 torr, while the trapping times ranged from 100 to 500 milliseconds. The chemicals used for the collection of the data in these experiments were obtained commercially in high purity and were used as supplied except for application of multiple freeze-pump-thaw cycles to remove non-condensable gases.

The transition metal ions for chemical ionization were generated by focusing the fundamental or quadrupled beam of a Quanta Ray Nd:YAG laser (1.06u) onto a metal foil or rod located on one of the ICR cell plates. The details of this method for generating metal ions have been described elsewhere (11,13), as well as some of the problems associated with the generation of non-thermal ions (40,43). Laser power and beam diameter have been adjusted to predominantly form monopositive ions. A background pressure of nitrogen at approximately 5 X 10-6 torr has been found to stabilize the ICR signal for the capacitance bridge instrument during acquisition of slow scan data and has been used for most of the experiments using this instrument (40).

DATA

The metal ion CI mass spectra for each of the compounds listed in Table 1 has been obtained as described above for both Fe^+ and Y^+. The data used for the recognition of the organic compounds consist of the branching ratios for the primary products generated by the initial reaction of the metal ion of interest with the organic neutral sample. Subsequent reactions of these product ions with the neutral (ie. secondary, tertiary, etc. reactions) were not considered. The reaction time and pressure were adjusted such that predominantly primary products were observed. The precursors of any product peaks in question have been confirmed by double resonance techniques and data has been collected under a variety of different conditions (trapping times and pressures) to confirm the primary product intensity ratios. Much of the data collected has been repeated using both the conventional ICR and FTMS to test the reproducibily of the data. Under these conditions the relative intensities of the products do not vary widely and have been found to be reproducible to better than 10%.

Six different training sets were generated from the data collected. The first two contain the data for the reactions of 72 compounds representing the six organic classes (alkane, alkene, ketone, aldehyde, ether and alcohol) with Fe^+ and Y^+, respectively. The third contains the combination of the data in the first two. The fourth and fifth training sets contain the data for the reactions of 24 alkanes of the three subclasses (linear, branched and cyclic) with Fe^+ and Y^+, respectively.

And the sixth data set contains the combination of the data in the fourth and fifth.

COMPUTER SOFTWARE

All of the pattern recognition programs have been written for the IBM 9000 lab computer using IBM version CS 9000 FORTRAN 77 and employ the k-Nearest Neighbor algorithm (KNN) (44). A classifier using this algorithm predicts the class of an unknown to be the same as that of the majority of its k-nearest neighbors. Due to the relatively small size of the data sets, only the first nearest neighbor (k = 1) was used to effect classification, and the distance measure employed was the Euclidian distance in an N-dimensional feature space. A scaling factor or weighting of the features has been found to improve clustering of the classes and a scheme has been developed to weight the features differently for each class (39). Using this weighting scheme, it is possible to assess the importance of a given feature for each class since the uniqueness of a feature can be inversely related to its weighting factor. The leave-one-out (LOO) algorithm (45,46) has been used to generate recognition accuracies used by the pattern recognition algorithms.

RESULTS AND DISCUSSION

Functional Group Recognition

The optimal feature selection searches for the two algorithms with the three organic data sets are compared in Table

II. From an examination of the individual class recognition accuracies for Fe^+ with the organics, it appears that iron ion can distinguish alcohols most readily from the other compounds. In the reaction of iron with alcohols, $FeOH_2^+$ is often observed and is unique to this class. Alkanes are separated from the other classes with 75% accuracy, and a very low recognition accuracy for the ketone class indicates that iron cannot distinguish this class from the others.

The nearest neighbors of the compounds that have been misclassified in the best of the two trials reported in Table II are identified in Table IV. In accordance with the high recognition accuracies, fewest misclassifications of the iron data set occur for the alkanes and the alcohols. The alkenes are misclassified as alkanes and ketones, two of the misclassified alkenes being closest to cyclic ketones and two closest to branched alkanes. In the reaction of ketones with Fe^+, the oxygen is most often lost as CO neutral, and often accompanied by H_2 and other hydrocarbons. Thus the corresponding ions observed are similar to those observed with alkenes. Two of the ketones are closest to alkenes while the others are misclassified as aldehyde, alcohol, alkane, and ether. Aldehydes are also poorly separated from the other classes, having misclassifications as alcohols, ketones, alkene, and alkane. Only three misclassifications occur for ethers, two as alcohols and one as ketone.

In contrast to iron, yttrium ion is particularly well suited for differentiating hydrocarbons from oxygen-containing species.

Due to a very strong yttrium-oxygen bond strength the major primary product for virtually all oxygen-containing functionalities studied has been YO^+. With this single unique feature it is possible to linearly separate oxygen-containing from non-oxygen-containing species.

As with iron, high recognition accuracy (100% here) for the alcohols is observed. Three unique peaks are observed for Y^+ with alcohols, YO^+, YOH^+, and YOH_2^+ allowing easy distinction from the other oganics. Alkanes are also recognized with high accuracy using yttrium, and in contrast to iron, the ketones are recognized with high accuracy from the yttrium data. The alkenes are the most poorly recognized class at 50% accuracy.

The contrast in the reactivity of Y^+ versus Fe^+ becomes apparent from the data presented in Table IV. The two cyclic alkanes misclassified with yttrium are closest to cyclic alkenes. Y^+ show a preference for dehydrogenation as opposed to C-C bond cleavage and thus dehydrogenation of the cyclic alkanes produces products similar to those found with alkenes.

With yttrium, as opposed to iron, a large improvement in the separation of the oxygen-containing organics from the hydrocarbons is observed. No misclassification of the alkenes as oxygenated species is observed for Y^+ which is in contrast to iron; the misclassified alkenes are all closest to alkanes, and similarly, cyclopentene and cyclohexene are closest to cyclopentane and cyclohexane respectively. For ketones with yttrium, only methylcyclopropyl ketone is misclassified and appears closest to hexanal. Three of the aldehydes are misclassified and are closest to butanone for which the only observed peak is YO^+. Since the major peak for the ketones and aldehydes is YO^+, it is not surprising that there is some difficulty in distinguising these two classes. Three of the misclassified ethers appear closest to butanone, on the basis of YO^+ intensity, and propylene oxide is closest to cyclohexanol.

When the data for the CI mass spectra using the two metals are combined, an improvement in the individual class accuracies is observed as well as an improvement in the total recognition accuracies for the data sets of the six organic classes. Alkanes and alcohols are still recognized with high accuracy while a large improvement in the recognition of alkenes is noted. Alkenes are recognized with a maximum of 83% accuracy (10 of the 12 compounds) as compared to a maximum of 67% accuracy with iron alone and 50% with yttrium alone. Using the data for iron alone, it had been difficult to distinguish the alkenes from the alkanes and oxygen containing classes; using yttrium alone it had been difficult to separate the alkanes from the alkenes. By adding the yttrium features to those of iron, oxygen-containing organics are now distinguishable from alkenes by the presence of YO^+. Thus additional and complementary information for the alkenes is obtained by combining the features of the two metals. Recognition of the ketones has improved from 33% accuracy using iron features to 67% accuracy at worst, and 83% at best when the yttrium data is included. Using yttrium features alone, however, a 92% recognition accuracy for ketones is possible. Thus some iron features which are detrimental to the classification of the ketones may have been included indicating that in feature selection for a multicategory classification problem there is often a trade-off between individual class recognition accuracies. A similar trend is noted for the recognition of aldehydes. No substantial improvement in the recognition of the ethers is observed by combining the data for both metals.

Structural Identification

Table III lists the recognition accuracies found for the three data sets of the alkane subclasses by the selection algorithms. A high total recognition accuracy is obtained with either metal ion, indicating that either can distinguish the three subclasses with ease. The nearest neighbors of the misclassifications for the best trial of SSTRA or FANNDE performed on these data sets are shown in Table V. Few misclassifications occur since a good separation of the three subclasses is observed for both metals. By combining the data for the two metals reacted with the 24 alkanes, a total recognition accuracy of 96% is possible, as opposed to a maximum of 92% for iron alone and 92% for yttrium alone. Only cyclobutane is misclassified as its nearest neighbor is propane. The reaction of Fe^+ with cyclobutane yields 95% $FeC_2H_4^+$ while propane reacts with iron to form 76% this ion. In examining the feature weightings for the three subclasses, this feature is most important for and unique to the linear subclass. The addition of the yttrium features to those of iron are not sufficient in this case to allow the correct classification of cyclobutane.

In examining the misclassifications presented in both Tables IV and V, the cyclic compounds are most often misclassified. For example, the three cyclic compounds, cyclopentene, cyclohexene, and vinylcyclohexane are misclassified as cyclic alkanes in almost every trial in Table IV. The cyclic subclass of alkanes are most difficult to distinguish and are also misclassified most often as evidenced in Table V.

In the reactions of Fe^+ and Y^+ with cyclic organics, often fewer products are observed than in the reactions with linear and branched species which makes the cyclics very difficult to distinguish. Thus, the cyclic compounds within an organic class react differently than the linear and branched compounds in the same class and often appear more similar to the cyclics of other classes.

Information from Feature Extraction

Along with the examination of misclassifications for the evaluation of the selectivities and reaction trends within organic classes, information is also contained in the features which are chosen to maximize the recognition accuracy of the data. The feature corresponding to the attachment of H_2O to Fe^+ is chosen for the best total recognition of the iron data set. This feature is unique to the reaction of iron with alcohols which is evident from the feature weightings. Its occurence in the set of features which best separate alcohols from the other compounds in the data set indicates a difference in the reaction of iron with other organics. With this information a reaction mechanism involving the initial insertion of Fe^+ into the HO-R bond of an alcohol, with subsequent shift of a β-hydrogen to the metal, and loss of the corresponding alkene neutral could be postulated. Further study of these reactions could produce information about the relative bond strengths of iron to various alkenes versus water. Thus an important piece of chemical information concerning chemical reactivities is extracted and highlighted using a purely empirical pattern recognition approach.

The peak corresponding to YOH^+ is chosen for the recognition of the organics using yttrium. This feature is distictive for alcohols. An important difference in the reactivity of Y^+ toward alcohols is indicated by the selection of this feature, since no $FeOH^+$ is observed for the reactions of iron with alcohols. The initial insertion of Y^+ into the R-OH bond of the alcohol must be much more exothermic than that of Fe^+ in order to provide enough internal energy to the reaction complex to fragment to YOH^+ and a hydrocarbon radical. Thus an important reactivity difference between the two metals is highlighted.

Much of the chemical information extracted through pattern recognition could be inferred by a detailed manual analysis of the data. The same analysis could be accomplished virtually instantaneously with the aid of a super-computer. It is encouraging that the same important features are extracted by empirical algorithms as would be selected by intuition of an experienced analyst. It is believed however, that the pattern recognition approach will uncover information which is more subtle yet important, and which is difficult to detect without the aid of this analysis.

CONCLUSION

Pattern recognition provides an objective means whereby trends in reactivity of the two metal ions along with their selectivities and differences are examined. The selectivities of the metals can be quantitated by examination of the individual class recognition accuracies as well as the misclassifications. Trends in reactivity and information regarding reaction mechanisms can be discovered through analysis of the features which have been extracted empirically by feature selection.

In this study 72 organics representing six classes have been recognized with an overall accuracy of 81%, and 24 alkanes, representing linear, branched and cyclic subclasses, have been recognized with 96% total accuracy using the combined CI mass spectral data of yttrium and iron (random guess classification would produce accuracies of only 16% and 33%, respectively).

Thus metal ions can be very useful for analytical applications in unknown analysis. The data from several metal ions can be combined for general unknown identification or a single selective metal may be used for identification of a specific organic.

The speed of the analysis and the fact that the unknown does not have to be present in the data set make pattern recognition attractive for online applications. Some further applications for our study will be to expand the technique to more complex molecules of interest, such as multiple functionalities, isomeric compounds, and biologically active samples, and the range of metal ions will be expanded. Another long range goal is the application of artificial intelligence to the optimization of experimental parameters involved in the metal ion FTMS experiment such that the maximum information can be obtained.

ACKNOWLEDGMENT

Acknowledgment is made to the Division of Chemical Sciences in the Office of Basic Energy Sciences in the United States Department of Energy (DE-AC02-80ER10689) for supporting the metal ion chemical ionization research and to the National Science Foundation (CHE-8310039) for providing funds for the advancement of FTMS methodology. R. A. Forbes would like to acknowledge the support of Lawrence Livermore National Laboratory and the Office of Naval Research.

The authors would also like to thank fellow research group members for their help in collecting the metal ion data: Denley Jacobson, William Stanton, Leo Lech, and Beth Stanko.

LITERATURE CITED

(1) Justice, J. B.; Isenhour, T. L. Anal. Chem. 1974, 46, 223.

(2) Abe, H.; Jurs, P. C. Anal. Chem. 1975, 47, 1829.

(3) Domokos, L.; Henneberg, D.; Weimann, B. Anal. Chim. Acta 1983, 150, 37.

(4) Dromey, R. G. Spectra 1984, 10, 3.

(5) Domokos, L.; Henneberg, D.; Weimann, B. Spectra 1984, 10, 11.

(6) Munson, M. S.; Field, F. H. J. Am. Chem. Soc. 1966, 88, 2621.

(7) Munson, B. Anal. Chem. 1977, 49, 772A.

(8) Jennings, K. R. in "Gas Phase Ion Chemistry"; vol. 2, Bowers, M. T., ed.; Academic Press, New York, 1979, 123.

(9) Burnier, R. C.; Byrd, G. D.; Freiser, B. S. J. Am. Chem. Soc. 1981, 103, 4360.

(10) Byrd, G. D.; Burnier, R. C.; Freiser, B. S. J. Am. Chem. Soc. 1982, 104, 3565.

(11) Jacobson, D. B.; Freiser, B. S. J. Am. Chem. Soc. 1983, 105, 5197.

(12) Jacobson, D. B.; Freiser, B. S. J. Am. Chem. Soc. 1983, 105, 7484.

(13) Jacobson, D. B.; Freiser, B. S. J. Am. Chem. Soc. 1983, 105, 7492.

(14) Jurs, P. C.; Isenhour, T. L. "Chemical Applications of Pattern Recognition"; John Wiley & Sons, New York/London/Sydney/Toronto, 1975.

(15) Varmuza, K. "Pattern Recognition in Chemistry"; Springer-Verlag, New York, 1975.

(16) Veress, G. E. Trends in Analytical Chemistry 1982, 1, 374.

(17) Habbema, J. D. F. Anal. Chim. Acta. 1983, 150, 1.

(18) Hippe, Z. Anal. Chim. Acta. 1983, 150, 11.

(19) Leegwater, D. C.; Leegwater, J. A. Trends in Analytical Chemistry 1984, 3, 66.

(20) Ritter, G. L.; Lowry, S. R.; Wilkins, C. L.; Isenhour, T. L. Anal. Chem. 1975, 47, 1951.

(21) Kowalski, B. R.; Bender, C. F. J. Am. Chem. Soc. 1972, 94, 5632.

(22) Frankel, D. S. Anal. Chem. 1984, 56, 1011.

(23) Parrish, M. E.; Good, B. W.; Jeltema, M. A.; Hsu, F. S. Anal. Chim. Acta. 1983, 150, 163.

(24) Dunn, W. J., III,; Stalling, D. L.; Schwartz, T. R.; Hogan, J. W.; Petty, J. D.; Johansson, E.; Wold, S. Anal. Chem. 1984, 56, 1308.

(25) Jellum, E.; Bjornson, I.; Nesbakken, R.; Johansson, E.; Wold, S. Journal of Chromatography 1981, 217, 231.

(26) Kowalski, B. R.; Bender, C. F. Anal. Chem. 1972, 44, 1405.

(27) Bink, J. C. W. G.; Van't Klooster, H. A. Anal. Chim. Acta. 1983, 150, 53.

(28) Thomas, Q. V.; Perone, S. P. Anal. Chem. 1977, 49, 1369.

(29) Bugard, D. R.; Perone, S. P. Anal. Chem. 1978, 50, 1366.

(30) Schachterle, S. D.; Perone, S. P. Anal. Chem. 1981, 53, 1672.

(31) Byers, W. A.; Perone, S. P. Anal. Chem. 1983, 55, 615.

(32) Byers, W. A.; Freiser, B. S.; Perone, S. P. Anal. Chem. 1983, 55, 620.

(33) Chien, M. Anal. Chem. 1985, 57, 348.

(34) Rotter, H.; Varmuza, K. Anal. Chim. Acta. 1978, 103, 61.

(35) Brent, D. A.; Roth, B.; Johnson, R. L.; Brunner, T. R. Biomed. Mass Spectrom. 1981, 8, 440.

(36) Tsao, R.; Voorhees, K. J. Anal. Chem. 1984, 56, 368.

(37) Van Der Greef, J.; Tas, A. C.; Bouwman, J.; Ten Noever Brauw, M. C.; Schreurs, W. H. P. Anal. Chim. Acta 1983, 150, 45.

(38) Brauner, A.; Budzikiewicz, H.; Boland, W. Org. Mass Spectrom. 1982, 17, 161.

(39) Forbes, R.A.; Tews, E.C.; Freiser, B.S. Journal of Computational Chemistry, submitted, 1985.

(40) Wise, M. B. Ph. D. Thesis, Purdue University, West Lafayette, 1984.

(41) Wronka, J.; Ridge, D. P. Rev. Sci. Instrum. **1982**, 53, 107.

(42) For a review of ICR see:

- a) Lehman, T. A.; Bursey, M. M. "Ion Cyclotron Resonance Spectrometry"; John Wiley & Sons, New York, 1976.
- b) Beauchamp, J. L.; Ann. Rev. Phys. Chem. **1971**, 22, 527.

For a review of FTMS see:

- c) Comisarow, M. B.; in Lecture notes in chemistry **1982**, 31, 484.
- d) Wanczek, K. P.; Int. J. Mass Spectrom. Ion Phys. **1984**, 60, 11.

(43) Kang, H.; Beauchamp, J. L. J. Phys. Chem. in press.

(44) Cover, T. M.; Hart, P. E. IEEE Trans. on Info. Theory 1967, IT-13, 21.

(45) Thomas, Q. V.; DePalma, R. D.; Perone, S. P. Anal. Chem. 1977, 49, 1376.

(46) Fukunaga, K. "Introduction to Statistical Pattern Recognition"; Academic Press, New York, 1972.

Table I. Compounds used for recognition experiments.

ORGANICS FOR RECOGNITION OF SIX CLASSES

ALKANE	ALKENE	KETONE
butane	1-butene	butanone
pentane	1-pentene	2-pentanone
hexane	1-hexene	2-hexanone
heptane	E-3-hexene	3-heptanone
2-methylpentane	3-methyl-1-butene	4-heptanone
3-methylpentane	2-methyl-1-pentene	3-methyl-2-butanone
2,3-dimethylbutane	4-methyl-1-pentene	3,3-dimethyl-2-butanone
2,3-dimethylpentane	2,3-dimethyl-1-butene	2,4-dimethyl-3-pentanone
cyclopentane	2,3-dimethyl-2-butene	cyclopentanone
1-methylcyclopentane	cyclopentene	methylcyclopropyl ketone
cyclohexane	cyclohexene	3-methylcyclopentanone
1-methylcyclohexane	vinylcyclohexane	cyclohexanone

ALDEHYDE	ETHER	ALCOHOL
propanal	ethyl ether	ethanol
butanal	methyl butyl ether	1-propanol
pentanal	ethyl propyl ether	2-propanol
hexanal	propyl ether	1-butanol
heptanal	ethyl butyl ether	2-butanol
octanal	butyl ether	1-heptanol
2-methylbutanal	isopropyl ether	1-octanol
3-methylbutanal	methyl-t-butyl ether	2-methyl-2-propanol
2,2-dimethylpropanal	sec-butyl ether	2-methyl-2-butanol
benzaldehyde	ethylene oxide	2,2-dimethyl-1-propanol
cyclohexanecarboxaldehyde	propylene oxide	cyclopentanol
cyclooctanecarboxaldehyde	tetrahydrofuran	cyclohexanol

ALKANES FOR RECOGNITION OF THREE SUBCLASSES

LINEAR	BRANCHED	CYCLIC
propane	methylpropane	cyclopropane
butane	3-methylpentane	cyclobutane
pentane	2,2-dimethylpropane	cyclopentane
heptane	2,3-dimethylbutane	methylcylopentane
octane	2,3-dimethylpentane	cyclohexane
nonane	2,4-dimethylpentane	1-methylcyclohexane
decane	2,2,4-trimethylpentane	ethylcyclopentane
dodecane	2,2,3,3-tetramethylbutane	1,4-dimethylcyclohexane

Table II. Recognition accuracies for six organic classes with two feature selection algorithms with three data sets.

	FE^+		Y^+		FE^+ AND Y^+	
	SSTRA	FANNDE	SSTRA	FANNDE	SSTRA	FANNDE
ALKANE	75	75	83	92	92	92
ALKENE	67	50	50	50	83	75
KETONE	33	25	92	83	67	83
ALDEHYDE	50	50	75	50	67	67
ETHER	75	50	67	50	75	42
ALCOHOL	92	92	100	100	100	92
TOTAL	65	50	78	71	81	75
P/F	4.2	9	5.1	12	3.4	14.4

Table III. Recognition accuracies for three alkane subclasses using two feature selection algorithms with three data sets.

	FE^+		Y^+		FE^+ AND Y^+	
	SSTRA	FANNDE	SSTRA	FANNDE	SSTRA	FANNDE
LINEAR	100	100	88	88	100	100
BRANCHED	100	100	100	88	100	100
CYCLIC	63	75	75	100	88	88
TOTAL	88	92	88	92	96	96
P/F	2.7	5.6	3.4	4	3	4.8

Table IV. Nearest neighbors of the misclassifications occurring in the best trials of the two feature search algorithms with the six organic classes.

MISCLASSIFIED COMPOUND	IRON	YTTRIUM	IRON AND YTTRIUM
ALKANE			
2,3-dimethylbutane	2-methylbutanal		
cyclopentane	cyclopentene	cyclopentene	cyclopentene
methylcyclopentane	3,3-dimethyl-2-butanone		
cyclohexane		cyclohexene	
ALKENE			
1-butene	methylcyclopropyl ketone		
1-hexene	2,3-dimethylbutane		
E-3-hexene	2,3-dimethylbutane	3-methylpentane	
3-methyl-1-butene	3-methylcyclopentanone	cyclopentane	
2-methyl-1-pentene		2-methylpentane	
4-methyl-1-pentene		2,3-dimethylbutane	
cyclopentene		cyclopentane	cyclopentane
cyclohexene		cyclohexane	methylcyclopentane
KETONE			
butanone	butanal		
2-pentanone	cyclopentanol		pentanal
2-hexanone	2,3-dimethyl-2-butene		1-pentene
3-methyl-2-butanone	2,3-dimethylpropanal		propanal
3,3-dimethyl-2-butanone	methylcyclopentane		
2,4-dimethyl-3-pentanone			pentane
cyclopentanone	E-3-hexene		
methylcyclopropyl ketone	ethyl butyl ether	hexanal	
3-methylcyclopentanone	2-methyl-2-butanol		
ALDEHYDE			
propanal	ethanol	butanone	butanone
butanal	butanone	butanone	butanone
pentanal	1-octanol	butanone	2-pentanone
hexanal	2-hexanone		
3-methylbutanal			3-methylpentane
benzaldehyde	cyclopentene		
cyclooctanecarboxaldehyde	cyclopentane		
ETHER			
ethyl butyl ether	methylcyclopropyl ketone		
isopropyl ether	2-propanol	butanone	butanone
ethylene oxide		butanone	butanone
propylene oxide		cyclohexanol	
tetrahydrofuran	1-propanol	butanone	1-pentene
ALCOHOL			
ethanol	propylene oxide		

Table V. Nearest neighbors of the misclassifications occurring in the best trials of the two feature search algorithms with the three alkane subclasses.

MISCLASSIFIED ALKANE	IRON	YTTRIUM	IRON AND YTTRIUM
LINEAR			
butane		2,2-dimethylpropane	
BRANCHED			
2,2,3,3-tetramethylbutane		octane	
CYCLIC			
cyclopropane	methylpropane		
cyclybutane	propane		propane

DL/413/83/01
GEN/413-2

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Office of Naval Research Attn: Code 413 800 N. Quincy Street Arlington, Virginia 22217	2
Dr. Bernard Douda Naval Weapons Support Center Code 5042 Crane, Indiana 47522	1
Commander, Naval Air Systems Command Attn: Code 310C (H. Rosenwasser) Washington, D.C. 20360	1
Naval Civil Engineering Laboratory Attn: Dr. R. W. Drisko Port Hueneme, California 93401	1
Defense Technical Information Center Building 5, Cameron Station Alexandria, Virginia 22314	12
DTNSRDC Attn: Dr. G. Bosmajian Applied Chemistry Division Annapolis, Maryland 21401	1
Dr. William Tolles Superintendent Chemistry Division, Code 6100 Naval Research Laboratory Washington, D.C. 20375	1
Dr. David Young Code 334 NORDA NSTL, Mississippi 39529	1
Naval Weapons Center Attn: Dr. Ron Atkins Chemistry Division China Lake, California 93555	1
Scientific Advisor Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380	1
U.S. Army Research Office Attn: CRD-AA-IP P.O. Box 12211 Research Triangle Park, NC 27709	1
Mr. John Boyle Materials Branch Naval Ship Engineering Center Philadelphia, Pennsylvania 19112	1
Naval Ocean Systems Center Attn: Dr. S. Yamamoto Marine Sciences Division San Diego, California 91232	1

END

DTIC

1-86