Pure & Appl. Chem., Vol. 50, pp. 197-203. Pergamon Press, 1978. Printed in Great Britain.

COMPUTER IDENTIFICATION OF MASS SPECTRA[a]

Fred W. McLafferty

Department of Chemistry, Cornell University, Ithaca, New York 14853, USA

Abstract - Automation of the interpretation of unknown mass spectra is a necessity in many laboratories. Two types of computer programs have been developed for this; "retrieval" systems compare the unknown against a file of reference spectra, while "interpretation" programs identify various molecular features to aid the interpreter. Particular emphasis here will be placed on systems of each of these types developed at Cornell, the "Probability Based Matching" and "Self-Training Interpretive and Retrieval System".

INTRODUCTION

The gas chromatograph/mass spectrometer/computer (GC/MS/COM) system has revolutionized the analyses of many kinds of complex organic mixtures. In this decade there has been an exponential growth in GC/MS/COM applications to areas such as pollutants, drug metabolites, insect pheromones, body fluids, and liquid fuels. With modern GC/MS/COM systems which produce hundreds of unknown mass spectra per day, an obvious bottleneck is the capability for molecular identification of these unknowns. Many laboratories have recognized the potential of computers as aids to this problem; a recent review lists 187 references (Ref. 2). The reader is referred to this review for details on the variety of algorithms proposed; this paper will emphasize the systems which are now most commonly used.

A number of factors govern the choice and design of such algorithms. These include type of computer (e.g., central versus "mini"), cost per run, data base quality and availability, unknown spectral quality, sample purity, degree of user interaction necessary, and identification confidence desired. Thus there is little chance that one system will answer all needs for mass spectral identification.

We perceive two distinct approaches, "retrieval" and "interpretation", to the identification of unknown mass spectra; these appear to be convenient ways to classify and design computer algorithms, although in many systems these approaches substantially overlap. A retrieval system attempts to find a match for the unknown mass spectrum by comparing it against each reference spectrum in a file. If a sufficiently good match is found, the problem is solved. Otherwise, the interpretive system attempts to find as much information as possible on the unknown's molecular features, such as elements, molecular weight, and substructures.

RETRIEVAL SYSTEMS

Most retrieval systems compare the unknown and reference spectra, ranking the latter according to the number of peaks in common and, in some cases, the similarity of the abundances of each pair. Most systems do not compare all of the peaks (many use <10 peaks), usually selecting the most abundant. The PEAK option of the Mass Spectral Search System (Ref. 3) operates in a "conversational" mode to allow user choice of peaks. Compilations are available (Ref. 4), however, so that matching of this kind can be done without the aid of the computer. Even with our present reference file representing over 30,000 different compounds, it is possible that a mass spectrum can be uniquely defined using only a few peaks because of the unusually high information content of a mass spectrum (Ref. 1). In such cases, very simple matching systems will suffice, and most GC/MS/COM configurations now have such options available. If their limitations are kept in mind, these can be of substantial help. However, computer capabilities are improving so rapidly that automatic routine screening of unknown mass spectra with the best matching systems appears feasible, and this discussion will concentrate on these.

Note a: Computer-Aided Interpretation of Mass Spectra, XXIV. For Part XXIII, see Ref. 1.

The MIT retrieval system

The system developed by Biemann and coworkers (Ref. 5), one of the most widely used and successful, utilizes the two most abundant peaks in each 14 mass unit range. This peak selection at least partially overcomes two distinct disadvantages of systems which choose only the most abundant peaks of the whole spectrum. (1) The most abundant peaks in the unknown mass spectrum are often the most common, such as those of low mass, and thus provide less selectivity than, for example, less abundant peaks at high mass; the MIT system forces the use of peaks in all mass ranges. (2) Experimental artifacts such as mass discrimination or changing sample concentration (such as conducting the MS scan on the side of a GC peak) often distort the relative abundances in the unknown mass spectrum as a function of mass; even if, for example, the high mass peaks have been reduced by an order of magnitude in this way, this system should still find the same two peaks as the largest in a particular 14 mass unit range. A "Similarity Index" is calculated for the comparison against each spectrum in the reference file, and the identity of the spectrum or spectra giving the highest values are listed for the user.

The Probability Based Matching (PBM) system

PBM was actually developed for real-time identifications with a microprocessor-controlled GC/MS system (Ref. 6), and has been extended to utilize a large data base (Ref. 7). PBM incorporates two unique features, (a) data "weighting", proven valuable for document retrieval from libraries (Ref. 8), and (b) "reverse search", which has been proposed independently by Abramson (Ref. 9).

The weighting involves the two principal types of data in low resolution mass spectra, masses and abundances (Ref. 10). The probability of occurrence of particular abundances (based on 100% for the most abundant peak) should follow a log normal distribution (Ref. 11). This was shown to be true for a data base of 18,806 different compounds (Ref. 10); abundance ranges differing by a factor of two in their occurrence probability are $\geq$0.24%, $\geq$1.0%, $\geq$3.4%, $\geq$9.0%, $\geq$19%, $\geq$38%, and $\geq$73%. The probability of occurrence of the different mass values also varies widely. For example, because larger molecular fragments tend to decompose to give smaller fragments, higher m/e values are less common in mass spectra, the probability decreasing by a factor of 2 approximately every 130 mass units. This is a surprisingly smooth function at high m/e values, but in the lower mass range there can be much larger differences in the probability of occurrence. Thus, although m/e 39, 41, and 43 ions of 1% or greater abundance are found in more than two-thirds of all reference spectra, peaks such as at m/e 34 and 340 occur less than 10% as frequently. In an unknown mass spectrum if abundant ions at m/e 34 and 340 are matched by comparably abundant ions in a reference spectrum, this is far more significant in indicating that a correct retrieval has been made than if the unknown's m/e 39, 41, and 43 peaks were matched in a reference spectrum. These abundance and mass uniqueness weightings are used to calculate the probability that the match occurred by chance and is thus a "false positive"; the reciprocal of this probability (log base 2 "confidence index, K") is used to rate the degree of match.

The second unique feature of PBM, "reverse searching", is valuable for the identification of components in mixtures. In this, PBM ascertains whether the peaks of the reference spectrum are present in the unknown spectrum, not whether the unknown's peaks are in the reference. Thus the reverse search in effect ignores peaks in the unknown which are not in the reference spectrum, as they could be due to other components of the mixture. Although reverse searching should thus reduce the capabilities of PBM for matching unknown spectra of pure compounds, this apparently is more than offset by the increased capabilities resulting from the data weighting. The system has been tested with over 800 "unknown" mass spectra taken from a large collection of mass spectra from diverse sources (Ref. 12), and for the spectra of pure compounds its performance has shown to be generally superior (Ref. 7 & 13) to the MIT system, and thus surely to those systems matching only the more abundant peaks in the spectrum. For mixtures, as expected, the reverse search procedure of PBM results in a substantially improved performance over these other systems. However, many of the advantages of the MIT and PBM systems are complementary, so that for important unknowns it is often wise to try both. Of course, the user must accept the final responsibility for the identification.

Improvements to PBM

Recent research at Cornell has focussed on several improvements to PBM. Revisions are being made to weight the data more accurately. For example, extra credit is justified if the reference molecular ion is present in the unknown; some peaks often occur together (e.g., m/e 43 and 57), so that their combined weight should be reduced; credit should also be reduced for "flagged peaks" (Ref. 7) and similar artifacts. As has been suggested previously (Ref. 14), the subtraction of the reference spectrum of an identified compound from the mixture spectrum should produce a residual unknown spectrum which is easier to identify. This approach has been implemented (Ref. 15) so that at the end of the PBM run the computer automatically subtracts the best-matching reference spectrum (or, at user command, some

other reference spectrum) from the unknown spectrum, notifies the user of any significant residual peaks, and on command does a PBM match on the residual spectrum. Although PBM is a reverse search system, its recall is lower for components in lower proportion in the mixture; thus subtracting out the contribution of an abundant component improves the PBM recall for other mixture components.

The PBM algorithm has also been applied (Ref. 16) to predict the number of bromine and chlorine atoms in ions of unknown mass spectra from the relative abundances of isotopic peaks. The theoretical isotopic patterns for 36 Br/Cl combinations are matched against peak "clusters" in the unknown spectrum with a modified PBM algorithm. The confidence values of the best Br/Cl assignments are further reinforced if they are consistent with the mass differences found for other peak clusters. In a test using the spectra of 2,670 compounds in which 814 contained bromine and chlorine, 90% of the predictions were correct.

INTERPRETIVE SYSTEMS

Basically, interpretive systems attempt to emulate the human interpreter. Their procedures may include tabulation of information from other sources (few samples are "total unknowns"), seeking special information from the spectrum (isotopic abundances, molecular weight, "nitrogen rule", exact mass measurements), and then applying structure-spectra correlations to elucidate the structural details. This last step, which is the one most commonly found in interpretive systems, has been approached in several ways. These again resemble ways taken by the human interpreter, who may obtain the relevant structure-spectra correlations by inspecting reference mass spectra to find related patterns, from tables available (possibly in the interpreter's memory) on this type of compound, or by applying mechanistic rules which detail the spectral relationships to be expected in more general structural situations. In this discussion the types of interpretive systems will be classified rather arbitrarily as "Pattern Recognition", those that resemble the first approach; "Artificial Intelligence", those that resemble the latter two; and the Cornell "Self-Training Interpretive and Retrieval System" (STIRS), which attempts to combine these.

Pattern Recognition

Even in the use of early retrieval systems (Ref. 17) it was noted that the best matching spectra often were of compounds containing similar structural features to that of the unknown. For example, in the PBM system four "classes of match" have been defined: I, identical compound or stereoisomer; II, class I or ring position isomer; III, class II or homolog; and IV, class III or an isomer of class III compound formed by moving only one carbon atom. The recall/reliability performance of PBM was evaluated separately for these four matching criteria; at the 50% recall level 65% of the compounds (low MW set) selected matched within the class I criteria, but over 95% matched within the class IV criteria. Thus most of the "wrong" answers which are retrieved by a matching system are of closely related molecular structures, so that a matching system can have interpretive value for the unknown spectrum for which a corresponding reference is not in the file.

Probably the most elegant systems for such automated correlations are those which incorporate "pattern recognition" or "learning machine" principles. Extensive research on development and applications of these methods has been carried out, in particular by Isenhour and his students (Ref. 2 & 18), but these do not appear to have been used appreciably as yet for actual unknowns. "Training" of the algorithm prior to use must be carried out for each functionality sought, which requires extensive computer runs if done on a large data base. A recent study (Ref. 19) comparing a "K-Nearest Neighbor" pattern recognition system to STIRS indicated that at present the latter is generally superior for substructure identification (vide infra).

Artificial Intelligence

A number of interpretive systems have been developed which utilize known structure/spectral correlations. Probably the most sophisticated of these is the Stanford "Artificial Intelligence" (AI) method of Lederberg, Djerassi and coworkers (Ref. 2 & 20). In this, heuristic DENDRAL is used to generate all possible molecular structures with prescribed limitations (Ref. 21); spectra for these are predicted from empirical mass spectrometry rules. Such rules can be generated with the aid of a special program Meta-DENDRAL (Ref. 22), which provides an algorithm for the discovery of MS fragmentation rules from empirical data of known compounds. To date complete AI programs have been generated only for a relatively few special compound classes, such as estrogenic steroids (Ref. 20). These programs are among a variety of special chemistry/molecular structure algorithms available on the Stanford computer to outside users (Ref. 23).

Self-Training Interpretive and Retrieval System (STIRS)

STIRS (Ref. 24) combines a knowledge of mass spectral fragmentation rules with an empirical search for correlations of reference spectra. To accomplish the former 15 classes of mass spectral data have been selected which are indicative of particular types of structural features; for example, low mass ions can be characteristic of electron-donating species such as the amino group and aromatic rings, while the masses of neutrals lost from the molecular ion can indicate electronegative functionalities. However, there are no predesignated spectra/structure correlations; instead, STIRS identifies particular molecular features by matching the unknown's spectral data in each of the 15 classes against the corresponding data of all the reference file spectra and computes a match factor (MF) indicating the degree of similarity. In each data class the fifteen reference compounds of highest MF values are saved. If a particular substructure(s) is found in a significant proportion of these compounds, its presence in the unknown is probable. Absence of a substructure is not predicted, as the mass spectral features of one substructure can be made negligible by the presence of a more powerful fragmentation-directing group. The data base for the system includes information from 29,468 different organic compounds containing the common elements H, C, N, O, F, Si, P, S, Cl, Br, and/or I. All structures of these compounds have been coded in Wiswesser Line Notation (WLN) to facilitate computer handling of structure data.

To utilize the information provided by the STIRS system, the results for each data class are examined and the common structural features identified. To aid this process, in a recently implemented system (Ref. 25), the computer examines the data for the presence of 179 frequently found substructures. The probability for the presence in the unknown of each substructure is predicted using a random drawing model. Knowing the frequency of occurrence of a specific substructure in the file, this method indicates the probability that the prediction of its presence in the unknown occurred at random. From this probability the confidence for each prediction is calculated. For example, in the STIRS data base the phenyl substructure is found to be present in 28% of the compounds. Statistically on the average this substructure would occur in 4 of any 15 compounds in the data base, including the top 15 compounds selected in a STIRS data class. On the other hand if phenyl is found in 10 of the 15 compounds, the probability that this occurred by chance is only 1 in 113, so that the confidence in the phenyl prediction is >99%, or a false positives value of <1%.

Peak selection for a number of STIRS data classes has been improved by statistical testing on a large group of unknowns (Ref. 26 & 27). Recently the performance of the improved STIRS system has been examined for the identification of the 101 most common substructures using ~600 "unknown" spectra selected at random from the data base (Ref. 28). For the "overall match factor", a weighted combination of the individual data class results, 58% of the substructures present were correctly identified (58% "recall") with only 1.0% of the postulated identifications being incorrect (1% "false positives"). Not surprisingly, the best performance is found for substructures which are known to give characteristic peaks or neutral losses. For example, the carbonyl shows only 33% recall at the 1% false positives level, as compared to recalls of 95% for trimethylsilyl, 89% for the steroid skeleton, and 87% for thiazole. In the case of a carbonyl group there is no distinct peak associated with the substructure, but the visibility of the group increases in carbonyl-containing substructures such as acetyl and benzoyl. A current research program is testing an enlarged list of such substructures chosen for their ability to correlate with spectral features (Ref. 29), so that STIRS should show a substantial improvement in its recall ability for these substructures.

STIRS versus other interpretive systems

STIRS is the only generally-available interpretive system applicable to the mass spectra of a wide variety of compounds. It has been available to outside users for nearly four years, and its rapidly increasing use is indicative of utility for a variety of problems. Probably the most important reason for the general applicability of STIRS is that it requires no pretraining for the mass spectral behavior of particular structural moieties, in contrast to the pattern recognition and Artificial Intelligence schemes described above. For STIRS, the 15 classes of mass spectral data have been selected to be indicative of different compound types or substructures, but without designating what these types or substructures are. STIRS then, in effect, trains itself to interpret the unknown mass spectrum by matching the unknown data in each of these classes against the corresponding data for all of the reference spectra; particular substructures or types of compounds which are found in a substantial proportion of the best-matching reference compounds have a correspondingly high probability of being present in the unknown. Thus STIRS does not have to be pretrained to recognize the presence of any particular structural feature; the thousands of research papers that have already appeared on mass spectra/structure correlations testify to the enormity of the task of developing an Artificial Intelligence type system applicable to a wide variety of multisubstituted compounds. On the other hand, pattern recognition systems must be pretrained to recognize particular structural features. A recent comparison of the K-Nearest Neighbor (KNN) pattern recognition system with STIRS (Ref. 19) which used the same data base and

substructure assignments showed (Table 1) that at comparable recall (RC) levels STIRS gave a much smaller number of false positives (FP). It was concluded that the application of mass

TABLE 1. Comparison of KNN and STIRS

Substructure	% in file	KNN 3/5		KNN 3/3		STIRS ≤2% FP		STIRS ≤0.5% FP	
		RC	FP	RC	FP	RC	FP	RC	FP
Ester, anhydride	19	29	4.3	19	0.8	55	2.1	50	1.4
Phenyl	28	71	11.6	51	3.3	75	5.0	67	3.4
Chlorine	8	54	0.8	47	0.5	74	0.3	74	0.0
Doubly-branched C	11	41	4.1	25	1.0	44	3.4	42	1.5
20 Substr. average	14	43	7.5	27	2.0	42	1.9	36	1.2

spectral knowledge in the selection of parameters employed in the KNN classifier could give KNN results comparable to those of STIRS.

Other STIRS improvements

The list of substructures specifically sought by STIRS must of necessity be limited to the more simple and common molecular fragments. However, it would be obviously advantageous if the computer could inspect the molecules representing the best-matching spectra for the maximal substructures which they have in common. For an algorithm recently developed at Cornell to do this (Ref. 30), Dot-Plot symbols based on Wiswesser Line Notations are used to generate binary occurrence vectors representing the molecules. A Compatibility Table is constructed from these vectors; "k-cover" incompatibility positions are cleared as well as positions not in agreement with "neighbor lists" and "degree lists" representing the connectivities of the molecules. A special analysis of this optimized Compatibility Table then yields the longest node-string paths, which represent the optimal substructures. Computational time requirements vary from 0.2 seconds for simple aliphatic and monocyclic compounds to 100 seconds for steroids; implementation on a routine basis will obviously require improving the algorithm's efficiency.

A key problem in mass spectral interpretation is the assignment of molecular weight (MW); for STIRS this is critical, as the neutral loss assignments for the unknown spectrum must be based on an assumed MW value. Approximately 15% of mass spectra have no molecular ion, and in others its assignment may be confused by impurities or ion-molecule reactions. An effective algorithm for MW assignment has been published by Dromey and coworkers (Ref. 31), based on the assumption that primary neutral losses from the molecular ion will match secondary neutral losses from the fragment ions. STIRS already incorporates several data classes for primary (MF 5, 6) and secondary neutral losses; a Cornell MW algorithm under development (Ref. 32) ranks the trial MW values by the degree to which the substructures found in their corresponding MF 5, 6 results match those indicated by the other data classes, whose results are MW independent. This MW program finds the best trial MW values by first attempting to identify elements such as Br and Cl from their isotopic patterns (Ref. 16), by predicting the odd- or even-parity of the MW value (the nitrogen rule), and restrictions according to logical neutral losses weighted by the secondary losses observed. Preliminary results of this program appear promising.

System performance evaluation

Examples of PBM and STIRS applications will be presented in the talk; a variety of these will be published elsewhere (Ref. 33-35), and will not be repeated here. The user in need of such methods would prefer more than just examples, however, as evidence on which to select a method. Because of the wide variation in needs, the user should try a variety of spectra representing his problems. For development of new algorithms, or improvements to present ones, however, it would appear that evaluations should be quantitative and statistically valid, and that any claims for improved performance should be accompanied by such data.

Recently we proposed (Ref. 36) that two types of evaluations are needed for automated systems used for the retrieval and interpretation of unknown spectra. The basic performance of such systems should be measured by their recall (RC), the proportion of structures predicted correctly (I_c) relative to the total possible correct predictions "present" (P_c), and by their false positives (FP), the number of cases of incorrect "present" predictions (I_f)

relative to the possible total number of such false predictions (P_f). Increasing the strictness of the identification criteria should make FP more favorable but decrease RC; thus the evaluation should include a range of RC/FP pairs to cover the potential needs of the user. Note, however, that when the system is applied to an unknown spectrum, the reliability (RL) of the resulting prediction is determined by the probability that it is correct, $RL = I_C/(I_C + I_f)$. It follows that $RL = P_C \cdot RC/(P_C \cdot RC + P_f \cdot FP)$, so that RL depends, in addition, on the occurrence probability of the particular structure in the unknown and in the reference file.

System availability over computer networks

At present it appears that most sophisticated retrieval and interpretive systems with the most comprehensive and accurate data bases are the ones that are also available over extensive computer networks, CYPHERNET and TYMNET. In addition, Artificial Intelligence methods for particular compound classes are available through telephone line connection to the Stanford computer (Ref. 23).

The Mass Spectrometry Data Centre (MSDC) in Aldermaston, England, has set up the Mass Spectral Search System (MSSS) in cooperation with NIH and EPA in the United States; this is being operated through the Cyphernetics Division of ADP Network Services, Inc. (Ref. 5). This system currently operates with a data base containing some 30,000 mass spectra of different organic compounds. Charges include an initial fee of $400, an annual subscription fee of $300 per institution, charges for computation, and $10 - $15 per hour for connect time. Algorithms available include the user interactive PEAK search with average charges of $3 - $6 per spectrum and the Biemann/MIT search, costing $6 per spectrum of $2 in overnight batch.

The Cornell TYMNET system (Ref. 37) has available the PBM retrieval and STIRS interpretive algorithms; both are substantially improved versions, updated at regular intervals to incorporate research improvements as described above. The present Cornell PBM system employs a data base of 41,429 different spectra of 32,403 different compounds, including the spectra of 2,365 compounds unique to the Cornell collection. There are no initial or annual fees for use of the Cornell/TYMNET system. Individual PBM searches are $10 per unknown spectrum, or $2.50 for overnight batch runs for users who have prepaid the data base royalties. STIRS runs are $15 per unknown spectrum, and lower prices are also available for STIRS batch runs. The only other charge is a $5 per hour TYMNET connect time assessment.

Acknowledgement - This report was made possible by the contributions of many coworkers: for PBM these included Dr. G. M. Pesyna, Dr. R. H. Hertel, R. D. Villwock, B. L. Atwater, and I. K. Mun; for STIRS these included Dr. K.-S. Kwok, Dr. H. E. Dayringer, and K. S. Haraki; while Dr. R. Venkataraghavan was active in all aspects of the research. The author is also indebted to the Environmental protection Agency, grant R804509, for generous financial support of this work, and to the National Science Foundation for partial funding of the DEC PDP-11/45 computer used in this research.

REFERENCES

1. D. D. Speck, R. Venkataraghavan, and F. W. McLafferty, Org. Mass Spectrom., submitted.
2. G. M. Pesyna and F. W. McLafferty, Determination of Organic Structures by Physical Methods, F. C. Nachod, J. J. Zuckerman, and E. W. Randall, Eds., Academic Press, New York, 6, 91-155 (1976).
3. S. R. Heller, G. W. A. Milne, and R. J. Feldman, Science, 195, 253-259 (1977).
4. Eight Peak Index of Mass Spectra, MSDC, Aldermaston, Reading, Britain (1974).
5. H. S. Hertz, R. A. Hites, and K. Biemann, Anal. Chem., 43, 681-691 (1971).
6. F. W. McLafferty, R. H. Hertel, and R. D. Villwock, Org. Mass Spectrom., 9, 690-702 (1974).
7. G. M. Pesyna, R. Venkataraghavan, H. E. Dayringer, and F. W. McLafferty, Anal. Chem., 48, 1362-1368 (1976).
8. G. Salton, Automatic Information Organization and Retrieval, McGraw-Hill, New York (1968).
9. F. P. Abramson, Anal. Chem., 47, 45-49 (1975).
10. G. M. Pesyna, F. W. McLafferty, R. Venkataraghavan, and H. E. Dayringer, Anal. Chem., 47, 1161-1164 (1975).
11. S. L. Grotch, 17th Annual Conference on Mass Spectrometry, p. 459-466, Dallas, May (1969).
12. E. Stenhagen, S. Abrahamsson, and F. W. McLafferty, Registry of Mass Spectral Data, Wiley-Interscience, New York (1974).
13. G. M. Pesyna, Ph.D. Thesis, Cornell University, 1975.
14. R. A. Hites and K. Biemann, Adv. Mass Spectrom., 4, 37-54 (1968).
15. B. L. Atwater, R. Venkataraghavan, and F. W. McLafferty, in preparation.

16. I. K. Mun, R. Venkataraghavan, and F. W. McLafferty, Anal. Chem., in press.
17. F. W. McLafferty and R. S. Gohlke, Anal. Chem., 31, 1160-1163 (1959).
18. T. L. Isenhour, B. R. Kowalski, and P. C. Jurs, Critical Review Anal. Chem., 4, 1-30 (1974).
19. T. L. Isenhour, S. R. Lowry, J. B. Justice, Jr., F. W. McLafferty, H. E. Dayringer, and R. Venkataraghavan, Anal. Chem., in press.
20. D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, J. Am. Chem. Soc., 94, 5962-5971 (1972).
21. L. M. Masinter, N. S. Sridharan, J. Lederberg, and D. H. Smith, J. Am. Chem. Soc., 96, 7702-7714 (1974).
22. B. G. Buchanan, D. H. Smith, W. C. White, R. J. Gritter, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, J. Am. Chem. Soc., 98, 6168-6178 (1976).
23. Dr. D. H. Smith, Department of Genetics, Stanford University Medical Center, Stanford, California 94305.
24. K.-S. Kwok, R. Venkataraghavan, and F. W. McLafferty, J. Am. Chem. Soc., 95, 4185-4194 (1973).
25. H. E. Dayringer, G. M. Pesyna, R. Venkataraghavan, and F. W. McLafferty, Org. Mass Spectrom., 11, 529-542 (1976).
26. H. E. Dayringer and F. W. McLafferty, Org. Mass Spectrom., 11, 543-551 (1976).
27. H. E. Dayringer, F. W. McLafferty, and R. Venkataraghavan, Org. Mass Spectrom., 11, 895-900 (1976).
28. K. Haraki and F. W. McLafferty, unpublished results.
29. F. W. McLafferty, Mass Spectral Correlations, Advances in Chemistry Series No. 40, American Chemical Society, Washington (1963).
30. M. M. Cone, R. Venkataraghavan, and F. W. McLafferty, J. Am. Chem. Soc., in press.
31. R. G. Dromey, B. G. Buchanan, D. H. Smith, J. Lederberg, and C. Djerassi, J. Org. Chem., 40, 770-774 (1975).
32. I. K. Mun, R. Venkataraghavan, and F. W. McLafferty, in preparation.
33. F. W. McLafferty, H. E. Dayringer, and R. Venkataraghavan, Ind. Res., 18, 78-83 (1976).
34. R. Venkataraghavan, H. E. Dayringer, G. M. Pesyna, B. L. Atwater, I. K. Mun, M. M. Cone, and F. W. McLafferty, ACS Symposium Series Computer-Assisted Structure Elucidation, D. H. Smith, Ed. (1977).
35. F. W. McLafferty, B. L. Atwater, I. K. Mun, and R. Venkataraghavan, J. Chromatogr. Sci., in preparation.
36. F. W. McLafferty, Anal. Chem., 49, 1441-1443 (1977).
37. Office of Computer Services, Cornell University, Ithaca, New York 14853.