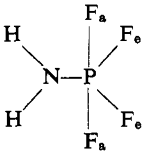
phorane is readily synthesized by the gas-phase reaction of PF_4Cl with NH_3. The establishment of the formulation F_4PNH_2 lies principally on mass spectral evidence (see Experimental Section). The observation of ^{19}F resonances of equal integrated intensity at -24.3 (F_a) and -1.33 ppm (F_e) (relative to internal $HCCl_2F$) is consistent with a trigonal bipyramidal structure in which the amino group occupies an equatorial site.



The complex appearance of the ^{19}F and ^1H nmr spectra (Figure 2) is ascribed to the fact that the molecule appears to constitute an $A_2KK'PTXX'$ spin system. The only satisfactory manner of extracting all the spin coupling parameters will be to perform iterative calculations and spectral simulations. At the present time, therefore, only approximate values are available for the coupling constants $J_{F_aPF_e}$ (~71 Hz), J_{F_aPNH} (~42 Hz), and J_{F_ePNH} (~1.6 Hz). However, the couplings $J_{PF_e} = 936$, $J_{PF_a} = 760$, $J_{^{15}NH} = 90.3$, $J_{^{15}NPF_a} = 22.8$, and $J_{PNH} = 17.7$ Hz can be recognized on a first order basis. The observation that the axial fluorines and amino protons are chemical shift equivalent (but spin coupling nonequivalent) implies that these nuclei are symmetrically disposed with respect to each other. Presumably the ground-state geometry of F_4PNH_2 involves a coplanar arrangement of the amino hydrogens and axial fluorine atoms since this would permit maximum $F \cdots H{-}N$ intramolecular hydrogen bonding. This suggestion is also in accord with the high barrier to intramolecular ligand exchange in F_4PNH_2. Although broadened, separate F_{axial} and $F_{equatorial}$ resonances are still discernible at 30°, thus placing $\Delta G^{\ddagger}$ in excess of 15 kcal/mol. Such an increase in barrier over, *e.g.*, $(CH_3)_2NPF_4$ (where $\Delta G^{\ddagger} \sim 9.0$ kcal/mol[34,75]), is presumably a consequence of intramolecular hydrogen bonding.

Acknowledgment. The authors are grateful to the National Science Foundation (Grant GP 28634) and the Robert A. Welch Foundation for generous financial support. Several stimulating discussions with Professor John W. Gilje are also appreciated.

Computer-Aided Interpretation of Mass Spectra. III. A Self-Training Interpretive and Retrieval System[1,2]

Kain-Sze Kwok,[3] Rengachari Venkataraghavan, and F. W. McLafferty*

Contribution from the Department of Chemistry, Cornell University, Ithaca, New York 14850. *Received October 26, 1972*

Abstract: A self-training system is described for computer interpretation of mass spectra which utilizes directly data of all available reference spectra, and does not require prior spectra-structure correlations of these data either by human or computer effort. The computer selects different classes of data known to have high structural significance, such as characteristic ions, series of ions, and masses of neutrals lost, from the unknown mass spectrum, and matches these against the corresponding data of all the reference spectra. The reference compounds of closest match in each data class are examined for common structural features; criteria have been determined so that such features can be identified with approximately 95% reliability. Tests with 110 "unknowns" not represented in the reference file showed that extensive to complete structural information can be obtained if spectra of related compounds are present in the reference file.

Compared to other spectroscopic methods commonly employed, a mass spectrum is unique in the large quantity of data it contains; the spectrum of a complex molecule can have hundreds of peaks, each discrete in mass. In sharp contrast to other methods, however, the chemist usually makes use of only a small percentage of these data for structure elucidation of an unknown. Although there has been substantial progress in mechanistic and empirical correlation of mass spectral data with molecular structure,[4,5] the large quantity of data itself apparently is a deterrent both to further progress by mass spectrometrists and to the acquiring of present knowledge by chemists with applicable problems. The average spectrum in our laboratory reference file representing 13,000 compounds contains over 100 peaks; searching such a data base for detailed similarities to an unknown spectrum is obviously a complex task. The rapidly growing utilization of gas chromatography-mass spectrometry, which can produce scores of mass spectra per hour, has greatly increased the need for such capabilities.

---

(1) Part II: R. Venkataraghavan, F. W. McLafferty, and G. E. Van Lear, *Org. Mass Spectrom.*, 2, 1 (1969).

(2) For preliminary communications, see K.-S. Kwok, R. Venkataraghavan, and F. W. McLafferty, 19th Annual Conference on Mass Spectrometry, Atlanta, Ga., May 1971; F. W. McLafferty, *Pure Appl. Chem.*, 7, 61 (1971).

(3) Taken from the Ph.D. Thesis of K.-S. K., Cornell University, 1973, which contains further details and examples.

(4) H. Budzikiewicz, C. Djerassi, and D. H. Williams, "Mass Spectrometry of Organic Compounds," Holden-Day, San Francisco, Calif., 1967.

(5) F. W. McLafferty, "Interpretation of Mass Spectra," 2nd ed, W. A. Benjamin, Reading, Mass., 1973.

Systems developed in response to this problem fall generally into three classes.[6-18] An early proposal of Zemany[6] utilizing keysort cards is the forerunner of many systems designed to identify a mass spectrum by matching it against those in a reference file;[6-14] the details of such "fingerprinting" techniques have been recently reviewed.[12b,13b] The specificity of the mass and abundance information in a mass spectrum usually leads to high selectivity in matching. It was recognized early, however, that the importance of particular peaks for such matching far outweighed their relative abundance,[7] and in some systems the peaks used in matching are not selected on abundance criteria alone.

A second class of systems can be defined in which the computer search of the reference file provides information concerning structural features of the molecule. It was noted even in the application of an early system[7] that often several of the reference spectra selected as closest matches were of compounds of close structural similarity to the unknown, even if the spectrum of the same compound were not in the file. The use of "learning machine" theory, as proposed by Isenhour and his coworkers,[15] offers a sophisticated method in which the computer performs specific structure-spectra correlations on the reference file in advance so as to be able to give "yes" or "no" answers to specific questions such as "Does the compound contain oxygen?" Development of a useful system utilizing this method has been slow, in part because the description of even a simple molecule requires answers to a very large number of such questions, and because the training required for one question using a large data base involves a very substantial investment of time of the computer and the chemist.

A third general class of systems attempts to utilize the present large body of knowledge concerning the mass spectral behavior of organic molecules, which is ignored in the above empirical approaches. In some systems[1,16,18a] identification of general structural moieties of the unknown is sought through characteristic spectral features such as the ion series (*e.g.*, $C_nH_{2n+1}^+$) present and neutral fragments lost, and then the specific features and their structural environments are investigated using subprograms incorporating detailed information on the mass spectral reactions to be expected. Systems have been described which utilize the latter for particular compound classes such as fatty acids,[10] aliphatic hydrocarbons,[10] and peptides.[17] In the "artificial intelligence" method developed by the Stanford group[18] possible structures for the unknown are generated by a sophisticated program (DENDRAL); information on spectral behavior is then used by the computer to predict the mass spectrum of each possibility, and these are matched against the unknown spectrum. In practice these methods appear to suffer not only from the substantial effort necessary to develop and update programs for each specific class of compounds, but also that they can only utilize correlative information extracted from reference spectra by human effort.

The "Self-Training Interpretive and Retrieval System" (STIRS) described here seeks to combine the advantageous features of the general approaches outlined above. "Training"[19] of the system by computer extraction of structural correlations from the reference data is done specifically for the spectral data of the unknown; the correlations involve computer searches for pattern similarities between the unknown spectrum and those in the reference file. Further, our present knowledge of mass spectral behavior is used to identify for the computer the classes of spectral data for correlation that should be of greatest structural significance; a separate search is carried out for each of these classes. To the extent that these data categories are sensitive to particular structural features, STIRS should be able to characterize the unknown by selecting compounds containing structural features of the unknown as the best matches in the various data categories.

Structural features produce groups of spectral peaks of characteristic relative abundances and masses, or mass differences; the identification of data classes selective for the multitude of possible structural features is complicated by the concomitant number of possible mass combinations, and by the probability of overlap of spectral groups from different structural features of a complex molecule (the availability of exact mass data would obviously increase the selectivity). The choice and definition of these classes have evolved from testing and application of our earlier system,[1] which was based on a general method of spectral interpretation.[5] The data classes presently include specific ion series, characteristic ions, and neutral losses, which have been selected on the basis of abundance, mass range, and ion type (odd- or even-electron). The "ion series" class is designed to ensure collective consideration of particular peaks even if some have more abundant neighboring peaks. At present these include each of the 14 possible homologous series using peaks of up to *m/e* 100 (for example, *m/e* 15, 29, 43, 57,

---

(6) P. D. Zemany, *Anal. Chem.*, **22**, 920 (1950).

(7) F. W. McLafferty and R. S. Gohlke, *ibid.*, **31**, 1160 (1959).

(8) S. Abrahamsson, S. Ställberg-Stenhagen, and E. Stenhagen, *Biochem. J.*, 92, 2PB (1964); S. Abrahamsson, *Sci. Tools*, 14, 129 (1967).

(9) V. L. Talroze, V. V. Raznikof, and G. D. Tantsyrev, *Dokl. Akad. Nauk USSR*, **159**, 182 (1964); 170, 379 (1966).

(10) B. Pettersson and R. Ryhage, *Ark. Kemi*, **26**, 293 (1967); *Anal. Chem.*, **39**, 790 (1967).

(11) L. R. Crawford J. D. Morrison, *ibid.*, **40**, 1464 (1968).

(12) (a) R. A. Hites and K. Biemann, *ibid.*, 40, 1217 (1968); (b) H. S. Hertz, R. A. Hites, and K. Biemann, *ibid.*, **43**, 681 (1971).

(13) (a) B. A. Knock, I. C. Smith, D. E. Wright, R. G. Ridley, and W. Kelly, *ibid.*, 42, 1516 (1970); (b) R. G. Ridley in "Biochemical Applications of Mass Spectrometry," G. R. Waller, Ed., Wiley-Interscience, New York, N. Y., 1972, p 177.

(14) S. L. Grotch, *Anal. Chem.*, 42, 1214 (1970); **43**, 1362 (1971).

(15) P. C. Jurs, B. R. Kowalski, and T. L. Isenhour, *ibid.*, 41, 21 (1969); T. L. Isenhour and P. C. Jurs, *ibid.*, **43**(10), 21A (1971).

(16) L. R. Crawford and J. D. Morrison, *ibid.*, 40, 1469 (1968); **43**, 1790 (1971); D. H. Smith, *ibid.*, **44**, 536 (1972).

(17) M. Senn, R. Venkataraghavan, and F. W. McLafferty, *J. Amer. Chem. Soc.*, **88**, 5593 (1966); K. Biemann, C. Cone, B. R. Webster, and G. P. Arsenault, *ibid.*, **88**, 5598 (1966); M. Barber, P. Powers, M. J. Wallington, and W. A. Wolstenholme, *Nature (London)*, **212**, 784 (1966).

(18) (a) A. Buchs, A. B. Delfino, A. M. Duffield, C. Djerassi, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg, *Helv. Chim. Acta*, **53**, 1394 (1970); (b) D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, *J. Amer. Chem. Soc.*, **94**, 5962 (1972).

(19) The term "self-training" as used here does not imply some specific definition of computer semantics. When a human is faced with the task of interpreting the mass spectrum of an unknown compound of a new structural class, he trains himself to do this by studying the spectra of related compounds to correlate their spectral behavior. STIRS is said to be self-training because it can interpret an unknown spectrum using only the related reference spectra; no human correlation of spectral behavior is necessary. We thank a referee for pointing out this possible source of confusion.

71, 85, and 99), and special series such as those of peaks commonly found for aromatic compounds and aliphatic amines. The degree to which the total ion abundance in each of the five most abundant series of the unknown spectrum match those of the reference spectrum is calculated as Match Factor 1 (MF1). Three classes of "characteristic ions" include the most abundant peaks in the mass regions 1–89 (3 peaks each of odd and even masses), 90–149 (5 peaks), and 150 up to the molecular ion ($M\cdot^+$) (5 peaks); abundance comparisons for these peak classes give MF2, MF3, and MF4, respectively. The "neutral loss" data represent the mass differences between the molecular ion and more abundant peaks of higher mass; these data provide information complementing that from the characteristic ions, as the neutral losses represent those parts of the molecules which are less capable of stabilizing the positive charge. The most abundant peaks for losses of smaller (<65) and larger masses are used for MF5 and MF6, respectively. Neutral losses are also used to indicate the structures of the most abundant high-mass fragment ions formed by the loss of an odd- and of an even-mass neutral (MF7 and MF8, respectively). The latter $(M - \text{even})\cdot^+$ usually is an odd-electron ion; its neutral losses are also matched against those of $M\cdot^+$ for the reference spectra (MF9) to select reference molecules of similar structure. Finally, an overall match factor, MF11, is calculated from a combination of factors MF1 through MF6 to attempt to identify combinations of structural features.

STIRS also incorporates programs for preliminary examination of the unknown to identify the molecular ion,[1,5,20] to determine the elemental compositions of abundant ions using isotopic peaks,[5] and to check for the presence of a spectrum of the identical compound in the reference file. The latter technique matches the most abundant odd- and even-mass peaks in each 14 mass unit interval from *m/e* 90 to $M\cdot^+$ of the unknown with each reference spectrum to obtain the "fingerprint match factor," MF10.

In operation STIRS selects for each match factor the ten reference spectra of highest MF values; the presence of a particular structural feature in the unknown is indicated by its appearance in several of the ten selected spectra. Specific appearance criteria have been determined for each match factor. To aid in the identification of features common to the selected compounds, the structures for each reference molecule have been coded in Wiswesser Line Notation (WLN).[21] At the present time, for the final step the chemist must combine the information on structural features, molecular weight, and ion elemental compositions to determine the most probable structure of the unknown. Particular interactive capabilities have been developed for STIRS to aid in this.

Experimental Section

Reference File. The method was tested with 13,000 spectra of nearly 10,000 different compounds of average molecular weight 200; 6652 spectra were from the "Atlas of Mass Spectral Data,"[22a] 5000 from the Aldermaston file,[22b] 1025 from the "Archives of Mass Spectral Data,"[22c] and the rest from the literature[4,22d] and this laboratory. Each spectrum has been checked for errors and its structure coded in WLN.

Condensed File and Match Factors. Nine classes of spectral data have been extracted from each of the reference spectra and stored in a "condensed file;" these are compared with the corresponding data of the unknown spectrum to calculate 11 match factors (Table I). For the four classes of neutral loss data the peak abundances are normalized relative to that of the largest peak as 1000; for the other five classes of data the square roots of the peak abundances are used, normalized relative to the sum of these values for all peaks as 1000. Morrison has shown that peak intensities normalized in this way are less dependent on instrumental conditions.[11]

Table I. Spectral Data Classes and Match Factors Utilized by STIRS

Class of spectral data	Match factor
1. Ion series	MF1
2. Low mass characteristic ions	MF2
3. Medium mass characteristic ions	MF3
4. High mass characteristic ions	MF4
5. Small primary neutral losses	MF5
6. Large primary neutral losses	MF6
7. Secondary neutral losses from the most abundant odd mass loss	MF7
8. Secondary neutral losses from the most abundant even mass loss	MF8
Class 8 data of the unknown spectrum matched against Class 5 data of the reference spectrum	MF9
9. Fingerprint ions	MF10
Overall match factor (MF1 + MF2 + 2MF3 + + 2MF$_4$ + 4MF5 + 2MF6)/12	MF11

Ion Series (MF1). Peaks at *m/e* 15 and 27–99, excluding *m/e* 28 and 32, are utilized for the 14 possible homologous ion series. Additional series include the "low aromatic" (*m/e* 38, 39, 50, 51, 63, 64, 75, 76), and "high aromatic" (*m/e* 39, 40, 51, 52, 65, 66, 77, 78, 79) series. A separate "amine series" has been set up to distinguish the $C_nH_{2n+2}N^+$ ions from isobaric ions (such as rearrangement or isotopic ions) giving peaks at masses 30, 44, 58, 72, and 86. These ions are classified as the amine series if either the relative abundance of the peak at *m/e* 30 is $>5\%$ or the peak at *m/e* 44 is $>45\%$, and if the compound contains nitrogen (the latter is not required for the unknown spectrum).

To make this category more sensitive to ion sequences, the abundance of a single peak in a series is not counted as $>40\%$ of the total abundance of the series, and no ion series is considered that does not contain at least three peaks of relative abundances $>0.3\%$. The five most abundant ion series arranged in decreasing order of importance are extracted from each spectrum, along with the sum of all series peak intensities relative to the total sum of all series, plus a mass centroid of the peak abundances within the series.

The ion series match factor, MF1, is computed using a formula similar to that of Knock[13] (eq 1), where m and n are the total num-

$$\mathrm{MF1} = \frac{2000}{5(m+n)} \sum_{k=1}^{A} (5 - |i - j|_k) \tag{1}$$

bers of ion series found in the two spectra being compared, A is the number of agreements, and i and j are the positions of kth pair of ion series with respect to the others. In cases where MF1 = 1000 (a perfect match), further differentiation is made by calculating a centroid match factor, CMF (eq 2), where c_k and

$$\mathrm{CMF} = \mathrm{MF1}/5 \sum_{k=1}^{A} (15 - |c_k - d_k|)/15 \tag{2}$$

---

[20] K. Biemann, W. McMurray, and P. V. Fennessey, *Tetrahedron Lett.*, 3997 (1966).

[21] E. G. Smith, "The Wiswesser Line Formula Chemical Notation," McGraw-Hill, New York, N. Y., 1968.

[22] (a) E. Stenhagen, S. Abrahamsson, and F. W. McLafferty, "Atlas of Mass Spectral Data," Interscience, New York, N. Y., 1969; (b) Mass Spectrometry Data Centre, AWRE, Aldermaston, Berkshire, England; (c) *Arch. Mass. Spectral Data*, 1–3 (1970–1972); (d) A. Tatematsu and T. Tsuchiya, Ed., "Structural Indexed Literature of Organic Mass Spectra," Academic Press of Japan, Tokyo, 1966–1968.

d_k are the centroids of the kth pair of ion series. If $(15 - |c_k - d_k|) < 0$, the value is reset to zero.[3]

Characteristic Ions (MF2–MF4). The three most abundant odd-mass ions and three most abundant even-mass ions below m/e 90 are used to calculate separately MF_{odd} and MF_{even} using eq 3 where

$$\text{MF} = 1000\left[\sum_{k=1}^{A}(I_i/I_j)_k(I_i + I_j)_k\right] \Bigg/ \left[\sum_{i=1}^{n} I_i + \sum_{j=1}^{n} I_j\right] \tag{3}$$

I_i and I_j are the intensities of the two peaks in the kth match, and n is the number of odd- or even-mass characteristic ions in each spectrum. If $(I_i/I_j)_k > 1$, its reciprocal is used in the equation. MF2 is taken as the mean of MF_{odd} and MF_{even}. Equation 3 is also used to calculate MF3 and MF4 for the medium and high mass characteristic ions using the five most abundant peaks from m/e 90 to 149 [or to $(\text{M} - 1)^+$ if $\text{M}\cdot^+ < 150$] and from m/e 150 to $(\text{M} - 1)^+$, respectively.

Neutral Losses (MF5–MF9). Selection of "neutral loss" data for a compound must be based on its monoisotopic spectrum; the method used for approximating this from the normal spectrum is relatively straightforward.[3] The "small primary neutral losses" correspond to the six most abundant peaks produced by losses of 0 (*i.e.*, $\text{M}\cdot^+$), 1, 2, 15 to 64, and 80. Their match factor, MF5, is calculated by eq 4, for which n and m are the respective numbers of

$$\text{MF5} = 2000\sum_{k=1}^{A} w_k \Bigg/ \left[\sum_{i=1}^{n} w_i + \sum_{j=1}^{m} w_j\right] \tag{4}$$

neutral losses present in the two spectra, and w_i, w_j, and w_k are weighting factors to increase the importance of intense peaks. The value of w_i (or w_j) = 6 if the ith (or jth) neutral loss in one (or the other) spectrum is >10% relative abundance; otherwise w_i (or w_j) = 2. For the kth match of the two spectra $w_k = 2$ except when the peaks are >10% in both spectra, in which case $w_k = 6 - |i - j|$, where i and j are the respective relative abundance positions of the neutral loss peaks in the two spectra.

The large primary neutral losses (MF6) represent the most abundant ion in each 15 mass unit interval starting downward from mass (M − 65). The lower mass limit is either (M − 64 − 90), in which case six neutral loss values will be used, or M/2, whichever comes first. MF6 is calculated in the same manner as MF5.

Secondary neutral losses are used to characterize $(\text{M} - \text{odd})^+$ and $(\text{M} - \text{even})\cdot^+$, the most abundant ions formed by the loss of an odd-mass fragment and an even-mass fragment of m/e <65. (Thus a secondary neutral loss of mass 28 from the fragment ion $(\text{M} - 15)^+$ corresponds to the $(\text{M} - 43)^+$ peak.) For MF7 the six most abundant secondary neutral losses from $(\text{M} - \text{odd})^+$ are selected using the same method as that used for the small primary neutral losses from the molecular ion M. $\text{MF7} = 2000A/(n_1 + n_2)$, where n_1 and n_2 are the numbers of secondary neutral losses in the two spectra being matched; note that the match factor is independent of the masses of the ions from which the secondary losses originate. For MF8 the four most abundant secondary neutral losses from $(\text{M} - \text{even})\cdot^+$ are selected. MF8 is calculated by the same method as that for MF7, except that it is required that the neutral fragments lost in forming $(\text{M} - \text{even})\cdot^+$ in the two spectra be the same. The secondary neutral losses from $(\text{M} - \text{even})\cdot^+$ of the unknown are also matched against the primary neutral losses of $\text{M}\cdot^+$ of each reference spectrum to give MF9; $\text{MF9} = 2500A/(n_1 + n_2)$ where n_1 is the number of secondary neutral losses in the unknown spectrum and n_2 is the number of primary neutral losses in the reference spectrum.

Fingerprint Ions (MF10). The most abundant odd-mass ion and the most abundant even-mass ion in each 14 mass unit interval from m/e 90 up to $\text{M}\cdot^+$ are used to calculate MF10 values using eq 3.

The Overall Match Factor. MF11 is computed as an average of the values for MF1 through MF6 using weighting factors of 1, 1, 2, 2, 4, 2, respectively, except that the magnitude of a factor is proportionally decreased if the corresponding class of data has less than the maximum specified number of entries.

Analysis of an Unknown Spectrum. The fingerprint ions of the unknown spectrum are used first for a preliminary search of the reference file. If a reference compound is retrieved with MF10 >900, this is usually considered to be a sufficient identification of the unknown, and the analysis is terminated. For unknown spectra having less than five fingerprint ions, the same criterion is applied to a weighted average of MF2 and MF10. If MF10 <900, the remaining match factors are calculated (*vide infra*); if MF11 >800 for a reference compound, this is also considered to be a sufficient identification. A random sampling of pairs of reference spectra of the same compound showed 85% had either MF10 >900 or MF11 >800, or both (in a majority of the remaining cases MF11 was the highest match factor).

If the unknown is not identified by MF10 as a compound in the reference file, the program attempts to determine the elemental compositions of abundant ions using isotopic peaks and tests to see if the highest mass peak can be $\text{M}\cdot^+$; the algorithms follow schemes proposed earlier.[1,5,20]

For each match factor the ten highest values are found; these compounds are printed out in decreasing MF order giving for each the name, WLN, serial number, molecular weight, formula, and peak of lowest mass in the spectrum (used to indicate the completeness of the spectral data).

A structural feature common to several of the selected compounds is indicated as also present in the unknown if particular criteria are met;[23] these are summarized in Table II.[3] For example, a structural feature is indicated by MF1 when it is present in at least seven different reference compounds in the top ten list which have MF1 >800.

Table II. Criteria for Structural Feature Identification

	Minimum no.	Minimum MF value	No. of WLN symbols	Minimum no. of entries
MF1	7/10	800		3 ion series
MF2	5/10	500		1 odd-mass and 1 even-mass ion
MF3, MF4	2/10	700	≥5	3 ions
	3/10	700	3, 4	
MF5	5/10	500	≥3	3 neutral losses
	6/10	500	2	
	7/10	500	1	
MF6	4/10	600		3 neutral losses
MF10	2/10	600	≥4	6 ions
	3/10	600	3	
	4/10	600	2	
MF11[a]	2/5	350	≥4	
	3/5	350	3	
	4/5	350	2	

[a] For MF11 the identified structural feature may be an isomer or homolog of the true feature.

Instrumentation. The condensed file is created from the master file of original reference spectral data using an IBM 360/65 computer; new reference data can be added easily to the file. Analysis of unknown mass spectra is carried out on either the IBM 360/65 or a DEC PDP-9 computer (16K words memory, 262K words disk storage, 300 lpm line printer and a 9 track, 800 bpi magnetic tape drive). The average computer time required per unknown is 1.3 min on the IBM 360 and 20 min on the PDP-9.[3] Progress in programming this for a laboratory PDP 11/45 computer indicates that a time requirement of only a few minutes can be achieved.

Interactive Search. Three special search capabilities have been incorporated to enable the human interpreter to interact with STIRS and tailor the search strategies to his needs.

Molecular Weight Determination. If the molecular ion finder program or the STIRS results indicates that $\text{M}\cdot^+$ has been incorrectly assigned, the interpreter may propose other molecular weight values and run the resulting sets of primary neutral loss data separately through STIRS to see which trial value gives the highest MF5 and MF6 results and the spectral features most consistent with those indicated by other match factors.[3]

Feedback Option for the Overall Match Factor. The interpreter can vary the weightings of the various individual match factors used to compute MF11 to see if a particular functionality indicated to

---

(23) At present, the human interpreter applies the criteria. Work is in progress to make the computer carry out this task utilizing connection tables generated from the WLN codes of the selected reference compounds.

some extent by several individual match factors can be confirmed by the new MF11 value.

The Shift Technique.[24] In many cases the addition of a small functional group to a large molecule changes the mass spectrum mainly by increasing the masses of the particular ion fragments which contain this functional group, without changing the relative abundances of these ions to a great extent. In the "shift technique" feature of STIRS the mass spectrometrist can specify the masses (*e.g.*, x, y) of additional small functional groups which are suspected to be present (or to be missing) in the unknown molecule. If an ion of mass m present in the unknown spectrum is not found in the reference spectrum, the program will look for peaks at $(m \pm x)$ and $(m \pm y)$ in the reference spectrum. Thus when $x = 14$ and $y = 28$, the system will search in addition for higher and lower homologs of the unknown; this has been tested for MF3 and MF4 using eq 5 where i_k and j_k are the relative abundance positions of the two ions m_{ik} and m_{jk} in the kth match; $w_k = 100$ when $m_{ik} = m_{jk}$, 60 when $m_{ik} = m_{jk} \pm 14$, and 40 when $m_{ik} = m_{jk} \pm 28$, except $w_k = 100$ when the same mass difference has been observed in any previous match.[3]

$$\mathrm{MF} = {}^1/_3 \sum_{k=1}^{A} w_k(6 - |i_k - j_k|) \qquad (5)$$

Results and Discussion

The ability of STIRS to identify structural features was tested using 110 mass spectra of a wide variety of organic compounds (average molecular weight = 208) obtained on a variety of mass spectrometers.[3] For each of these spectra all possible match factors were calculated by comparison to each of the reference spectra, eliminating any spectra of the 110 unknown compounds.

The compounds giving the top ten values for each match factor are used to indicate the structural features of the "unknown" according to the criteria given above. It should be emphasized that these data are used as *positive* indicators only; a structural feature not found in a sufficient number of the top ten reference compounds is not necessarily absent in the unknown structure. For each match factor examples will be given of the types of structural features retrieved.

Ion Series (MF1). The most abundant ion series in the spectrum of methyl nonane-1,9-dioate are 27, 41, ···; 15, 29, ···; 31, 45, ···; "low aromatic;" and 42, 56, ···. More then ten reference spectra gave a perfect match, so that such spectra were ranked additionally according to their mass centroid values (Table III). The presence of these particular ions probably would have led a trained mass spectrometrist to conclude only that the unknown is an oxygenated compound containing a small degree of unsaturation. On the other hand, Table I indicates in high probability that the unknown is a saturated methyl ester; further, three of the top ten compounds are dimethyl dialkanoates. Note also that these identifications by the STIRS program were accomplished without specific pretraining to recognize these or other structural features.[25]

Table III. Ion Series (MF1)[a] Selections for Methyl Nonane-1,9-dioate

Compd	WLN	Centroid MF
Methyl heptane-1,7-dioate	1OV5VO1	797
Methyl octane-1,8-dioate	1OV6VO1	753
Methyl 3,5-dimethylheptanoate	2Y&1Y1VO1	719
Methyl *n*-nonanoate	8VO1	705
Methyl *n*-hexanoate	5VO1	666
Methyl *n*-octanoate	7VO1	665
Methyl 11-bromoundecanoate	E10VO1	660
Methyl hexane-1,6-dioate	1OV4VO1	644
Methyl *n*-decanoate	9VO1	637
Isobutyl methacrylate	1Y&1OVYU1	73

[a] MF1 = 1000 for all listed compounds.

The structural feature criteria for MF1 (Table II) predicted the presence of 122 structural features from the 110 spectra tested (Table IV); 115 of these predictions (94%) were correct.[3] Although this compares very favorably with rates reported for learning machine[15,26] and computer interpretation[1,16–18] predictions, a quantitative comparison is not particularly meaningful because the other systems employed much smaller reference files, searched for different structural features, and often were limited to a particular compound class. It should be possible to improve the prediction accuracy and selectivity of MF1 by better utilization of the mass centroid data[27] and by incorporating other ion series specifically chosen to identify structural features poorly selected by other match factors.

Table IV. Structural Features Selected by Matching Low Mass Ions

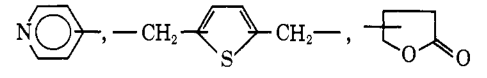
Structural feature	MF1	MF2	Structural feature	MF1	MF2
Oxygen (no N)	17/17[a]	28/29[a]	Amines	4/4	5/5
>1 oxygen	6/6	11/11	-S-	2/2	2/2
atoms	11/11	22/23	Phenyl group	13/14	24/25
-O-	5/5	12/13	Heterocyclic	2/2	3/3
-CO-	3/3	7/7	aromatic	34/37	32/32
-COO-	2/2	2/2	Nonaromatic	3/3	6/6
-COOCH$_3$	6/6	12/12	Cycloalkyl	3/3	2/2
Nitrogen	1/1	2/2	Cycloalkenyl	1/2	7/7
>1 nitrogen			Alkyl chain		
atoms			>C$_4$		

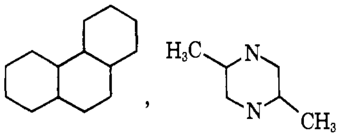
Other features predicted correctly by MF2:

$—CH(CH_3)_2$, $—C(CH_3)_3$, $—CH_2CH{=}C(CH_3)_2$, $—OH$, $>CHCH_2OH$,

$—O\text{-}n\text{-}C_4H_9$, $>NH$, $C_6H_5C{<}$, [b]$C_6H_5CH—$, $C_6H_5O—$,

$C_6H_5CHOH—$, $HOC_6H_4—$, $C_6H_5CO—$, $ClC_6H_4—$, $C_6H_5NH—$,





[a] Number of times the particular structural feature was selected correctly/total number of times selected. [b] C_6H_5 and C_6H_4 indicate mono- and disubstituted phenyl, respectively.

Low-Mass Characteristic Ions (MF2). The unexpected specificity achieved by STIRS utilizing the low mass (m/e <90) peaks is shown even more clearly by MF2, in which the computer seeks the best matches

---

[24] K. Biemann, "Mass Spectrometry: Organic Chemical Applications," McGraw-Hill, New York, N. Y., 1962, p 305.

[25] Thus the presence of the so-called "low aromatic ion series," obviously a misnomer here, does not require that STIRS give weight to aromatic compounds.

[26] For example, learning machine prediction rates of 73.1, 75.3, 92.5, and 94.6% are reported for carbonyl, oxygen linkage, amine, and aromatic compounds, respectively.[15]

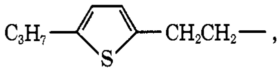
[27] This is compromised, however, by the fact that many reference spectra have not been recorded at lower masses.

for the most abundant odd- and even-mass ions. For example, in matching with the mass spectral data of farnesol[13] (*m/e* 69, 41, 43, 68, 70), six of the ten compounds yielding the highest MF2 values are hydroxymonoterpenes ($C_{10}H_{18}O$) or their acetate or formate derivatives. The structural feature criteria for MF2 (Table II) gave 186 predictions (Table IV) for the 110 unknowns, for which 181 were correct (97%). These results are clearly superior to those of MF1 in quantity, specificity, and accuracy of the structural information; however, MF1 matching did identify structural features not found by MF2 in 25% of the cases of Table IV. The number of cases in which MF2 matching identified more complex structural moieties is surprisingly large; many are of higher mass than the *m/e* 89 upper limit of the MF2 characteristic ions.

Medium- and High-Mass Characteristic Ions (MF3, MF4). It is expected that matching these higher mass data should identify more complex structural features. Thus, it is possible to make the selection criteria less strict; for example, selection requires that a feature describable with three WLN symbols be present in only three compounds of MF >700. Though this specificity restricts the number of structural moieties identified (Table V), all of the predictions fulfilling the criteria were correct. The usefulness of these match factors should increase markedly as more reference spectra of high molecular weight compounds are added to the search file. We also plan to test match factors for peaks covering overlapping mass ranges (*e.g.*, *m/e* 75–119, 120–200).

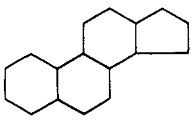
Table V. Structural Features Selected by Matching Medium- and High-Mass Characteristic Ions

MF3: $CH_3OCO(CH_2)_9$—, $CH_3CO(CH_2)_{6\text{ or }7}COOCH_3$, Cl_2CHCOO—, $HOCH_2CH_2NHCO(CH_2)_6$—, Cl_3C—,

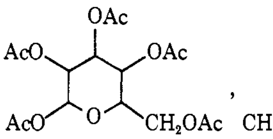


, ClC_6H_4CO—, $H_2NC_6H_4$—,

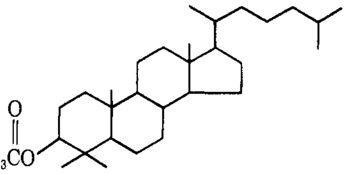
$C_6H_5C(CH_3)_2O$—, $CH_3C_6H_3(CN)$—, C_6H_5CO—,

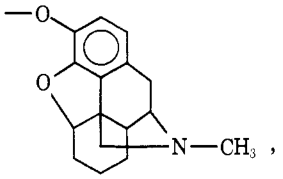


MF4: $CH_3OCO(CH_2)_{13}$—, $CH_3OCOC_7H_{14}COOCH_3$,

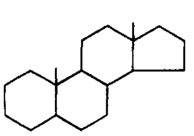


,





,



Although the "fingerprint ions" match factor (MF10) was designed to retrieve from the file reference spectra of the same compound as the unknown, MF10 does give matching of a larger number of characteristic ions over a broader mass range than does MF3 or MF4. The criteria of Table II gave 22 structural feature identifications by MF10 for the 110 unknowns; however, two were incorrect. Several of these 22 features were also identified by MF3 or MF4; the other features found were $(CH_3)_2C{=}CHCH_2CH_2$—, $(CH_3)_2$—piperazine—, C_6H_5COO—, $C_6H_5CH(OH)$—, $C_6H_5CH{=}CHCO$—, $C_6H_5CH{=}CHCH_2O$—, $CH_3OC_6H_4$—, $CH_3OC_6H_4NH$—, HOC_6H_4COO—, $C_6H_5C_6H_4CO$—, $C_6H_5OC_6H_3(CH_3)$—, and $(CH_3)CS$—pyridyl—.

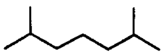
Primary Neutral Losses (MF5, MF6). The usefulness of the mass of the neutral lost to indicate certain small electronegative functional groups is well known;[4,5] this is reflected in the types of features identified by MF5 and MF6 (Table VI). However, STIRS often identifies more complex moieties from the combination of neutral loss peaks used for MF5. For example, all ten of the MF5 selections for 3-phenylthiophene (neutral losses of 0, 45, 32, 1, 26, 58) contain at least one thiophene ring, and seven contain phenyl groups. An even more surprising example of the capability of STIRS without prior training is the matching of the low-mass neutral loss data of cholesterol (losses of 18, 0 ($M\cdot^{+}$), 33, 15, 17, and 61); the six compounds formed with the highest MF5 values are the steroidal derivatives allopregnanol-3α-one-20, pregnenolone alcohol, pregnenolone, 7β-hydroxycholestanyl 3β-acetate, 16α-methylpregnenolone alcohol, and 16α-methylpregnenolone (a large number of other oxygenated carbocyclic compounds are present in the data base). Although the loss of these simple neutral species from cholesterol is quite *consistent* with present knowledge of the mass spectral behavior of such molecules, it is doubtful that these losses were known to be *characteristic* of such molecules.

Table VI. Structural Features Selected by Matching Neutral Loss Data

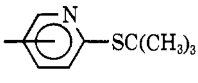
Structural feature	MF5	MF6	Structural feature	MF5	MF6
Oxygen (no N)	20/21[a]	7/7[a]	$COOCH_3$	4/4	
>1 oxygen	9/10	5/5	Nitrogen	5/5	2/2
atoms	22/22	7/7	Nonaromatic	20/23	1/1
-O-	16/18	4/4	Phenyl group	6/8	4/4
-CO-	10/13	2/2	Heterocyclic	3/3	
-COO-			aromatic	3/3	
			Chlorine		

Other features predicted correctly by MF5:

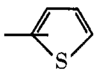
—OH, —CH_2OH, —OCH_3, —$O(CH_3)_3$, —$OC(CH_2)_2$—, —CN, —CH_2CH_2CN, —CCl_2, $>C-CH_3$, —C_2H_5, —$COOC_2H_5$, —CH_2COCH_3, —$C{=}CCOOC_2H_5$, —NO_2, R—S—R, —$SC(CH_3)_3$, —$HOOCC_6H_4$—, $CH_3OOCC_6H_4$—, ClC_6H_4—,



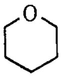
,



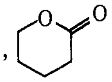
,



,



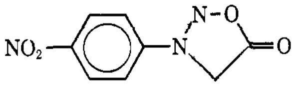
,



[a] See footnotes of Table IV.

The loss of larger neutrals can give specific structural information. *N*-(*p*-Methoxyphenyl)-*N'*-phenylurea gives peaks corresponding to losses of 119, 93, and 108; these give the highest MF6 values [value in brackets] for *N*-(*o*-methoxyphenyl)-*N'*-phenylurea [1000], *N*-(*m*-chlorophenyl)-*N'*-phenylurea [800], phenothioxinsul-

Table VII. Overall Match Factor (MF11) Selections for 3-*p*-Nitrophenylsydnone



T5NNOVJ AR DNW

Compd	WLN	MF11	MF1	MF2	MF3	MF4	MF5	MF6
3-(2-Naphthyl)sydnone	L66J C-AT5NNOVJ	523	666	508	0	0	800	555
3-(*p*-Carboxymethylphenyl)-sydnone	T5NNOVJ AR D1VQ	476	466	146	0	0	857	416
4-Methoxycarbonylbenzo[*c*]-cinnoline	T B666 HNNJ FVO1	424	666	401	0	145	666	0
4-Chloro-3-phenylsydnone	T5NNOVJ AR EG	424	500	317	111	0	714	0
1,3,5-Cyclooctatriene-tricarbonyltungsten	L8U CU EUTJ 0-WO-/CO 3	424	266	160	54	0	833	0
5-Methylundecane	6Y4	414	0	0	0	0	857	500
4-Methyl-3-phenylsydnone	T5NNOVJ AR & E	389	400	279	0	0	714	0
5-Methyldecane	5Y4	378	0	0	0	0	857	0
4,4′-Dinitroazobenzene	WNR DN 2U	365	800	827	326	345	222	0
N-Acetylbenzamide	1VMVR	359	500	257	0	0	625	0

fone [666], 1,4-dimethyl-1,4-diphenylbutene [631], *N*-phenyl-*N*′-(1-naphthyl)urea [625], *N*-(*o*-chlorophenyl)-*N*′-phenylurea[615], 1,4-di(4′-phenylbutyl)benzene [571], *N*-(2-chloro-6-methylphenyl)-*N*′-phenylurea [571], 1,1,3-trimethyl-*p*-3-phenylindan [526], and 1,3,5-triphenylbiuret [500]. Note that six of the compounds contain the *N*-phenylurea substructure and nine of the compounds contain more than one aromatic ring.

For the 110 unknown spectra tested by STIRS, 179 of the 191 structural features identified by MF5 were correct (94%), and all 35 of the structural features identified by MF6 were correct.[28] It should be noted that most of the incorrect selections are for structural functionalities for which MF1 and MF2 are sensitive, and thus selection of these functionalities by MF11 should be both sensitive and accurate (*vide infra*).

Secondary Neutral Losses (MF7, MF8, MF9). Derivation of "neutral losses" data concerning a fragment ion is more ambiguous than derivation for a molecular ion because a fragment ion is not necessarily the precursor for all ions below it in mass in the spectrum.[29] Although positive results in the testing of the 110 spectra were insufficient to derive criteria for Table II, particular examples indicate that valuable information can be obtained, especially for abundant fragment ions of masses near that of $M\cdot^+$. In the mass spectrum of 3-methyl-4-chloroacetophenone the base peak at *m*/*e* 153 corresponds to $(M - 15)^+$, and is thus the most abundant $(M - \text{odd})^+$ ion; lower mass peaks correspond to losses from it of 28, 64, 54, 63, 62, and 36 mass units, in order of decreasing abundance. The highest MF7 value was found when these data were matched against *o*-, *m*-, and *p*-chlorobenzaldehyde, *o*-, *m*-, and *p*-chloroacetophenone, and *o*-, *m*-, and *p*-chlorobenzoic acid; for all of these $(M - \text{odd})^+$ is *m*/*e* 139, ClC_6H_4CO–. Thus, MF7 gives strong evidence that the *m*/*e* 153 ion of the "unknown" contains the chlorobenzoyl moiety.

For δ-laurolactone the most abundant $(M - \text{even})\cdot^+$ peak is $(M - 18)\cdot^+$, and lower mass peaks correspond to further losses of 18, 44, 80, and 53 mass units. The highest MF8 value, 500, is found for the data from ten compounds; four of these are lactones of C_{15}–C_{18} fatty acids.

For ethyl isobutyl sulfide the most abundant $(M - \text{even})\cdot^+$ is $(M - 62)^+$, *m*/*e* 56. Matching the data for further neutral losses from this fragment against the *primary* neutral loss data of the reference file gave the highest MF9 values for 1-, 2-, and isobutene, consistent with the true nature of the $(M - 62)\cdot^+$ ion.[3]

Overall Match Factor (MF11). This factor often indicates structural features that are too complex to be selected clearly by individual match factors. For the data of methyl nonane-1,9-dioate (Table III), MF1 selected three dimethyl dialkanoates; all ten of the best matches for MF11 are dimethyl dialkanoates. The compounds of highest MF11 values found using spectral data of 3-*p*-nitrophenylsydnone as the "unknown" are shown in Table VII; four selections have indicated an aromatic sydnone, and two others have also indicated nitrogen atoms attached to an aromatic ring. The MF11 values have given a much more positive indication of the true structure than any of the individual match factors; only two of the sydnone spectra were in the top ten for MF5, and none for MF1 – MF4 or MF10.

Further examples are given in Table VIII, which lists the two compounds selected with the highest MF11 values for a number of unknowns. Table IX lists results for 17 randomly selected compounds of higher molecular weight (average = 325).[30] These and other examples show that the compounds giving the highest MF11 values can provide useful detailed information on the unknown structures; however, the criteria of Table II are often satisfied by a structural feature that is slightly different than that of the unknown, such as an isomer or homolog.[31] Within these limits, the MF11 predictions were 93% correct. Another limitation is illustrated by the results for

---

(28) The lower number of selections made by MF6 is due in part to the lower number of unknown and reference compounds of high molecular weight; the criterion (Table II) of three or more large-neutral-loss peaks requires a mol wt >218.

(29) It is possible to reduce this ambiguity by the use of elemental compositions obtained from the exact mass values of the unknown spectrum. A greater restriction of the allowed neutral losses may also help; this is currently being tested.

(30) The spectra of the estrones (6th and 7th compounds of Table IX) have also been interpreted using the "Artificial Intelligence" method.[18b]

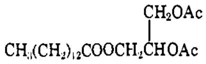
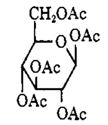
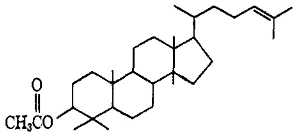
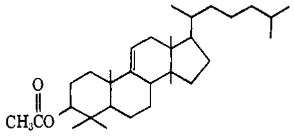
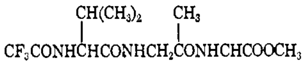
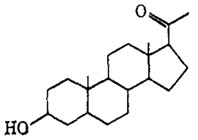
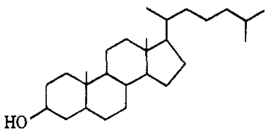
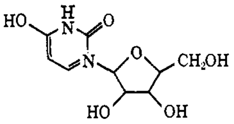
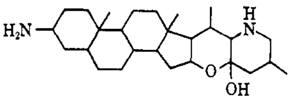
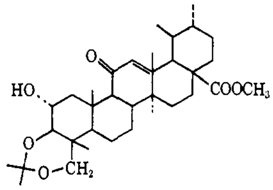
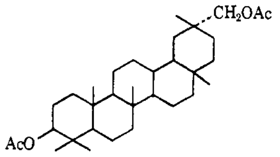
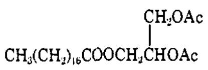
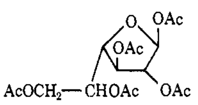
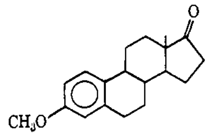
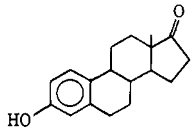
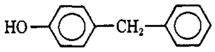
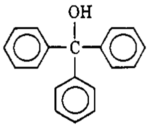
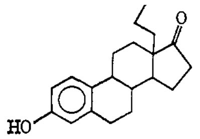
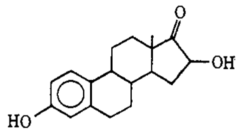
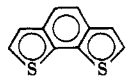
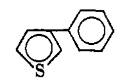
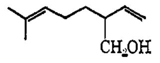
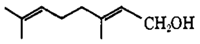
(31) For example, the moiety $CH_2CONHCH_2COOCH_3$ was indicated by MF11 for the spectrum of $CF_3CONHCH(C_3H_7)CONHCH_2CONHCH(CH_3)COOCH_3$.[3]

Table VIII. Examples of Overall Match Factor (MF11) Selections[a]

First choice	MF11	Second choice	MF11	Actual Compd
2-Chlorooctane	616	1-Chloroheptane	504	2-Chloroheptane
2-Thianonane	602	2-Thiaoctane	523	2-Thia-3-methyloctane
Phthalimide	581	1,2,3-Trioxoindan	554	Phthalic anhydride
γ-Decalactone	614	γ-Octalactone	608	γ-Nonalactone
Di-*n*-butoxymethane	785	Di-*sec*-butoxymethane	499	Diisobutoxymethane
Cinnamyl acetate	516	1-Methoxy-4-propenylbenzene	475	Cinnamyl propyl ether
Ethyl 2,2-diethyl-3-ketobutyrate	506	Ethyl 2-*n*-butyl-3-ketobutyrate	495	Ethyl 2-isopropyl-3-ketobutyrate
Adiponitrile	598	Pimelonitrile	567	2-Methyladiponitrile
1,2,3,4-Tetramethoxybutane	317[b]	1-Phenyl-1,2,3-trimethoxy-propane	309[b]	1,2,3-Trimethoxy-4-phenylbutane

[a] The spectrum of the actual compound was *not* in the reference file searched. [b] These unusually low match factors result, at least in part, from the fact that these spectra have a few ions of substantial relative abundance. However, all of the compounds selected (top ten MF11 values) had similar structural features.

Table IX. Higher Molecular Weight Examples of MF11 Selections[a]

First choice[b]	MF11	Actual compd	First choice	MF11	Actual compd
$CH_3OOC(CH_2)_6COOCH_3$	706	$CH_3OOC(CH_2)_7COOCH_3$			
$CH_3(CH_2)_{15}COOCH_3$	862	$CH_3(CH_2)_{14}COOCH_3$	 $CH_3(CH_2)_{12}COOCH_2\overset{CH_2OAc}{\overset{\vert}{C}}HOAc$	539	
	758		$CH_3CONHCH_2COOCH_3$	444	 $CF_3CONH\overset{CH(CH_3)_2}{\overset{\vert}{C}}HCONHCH_2\overset{CH_3}{\overset{\vert}{C}}ONHCHCOOCH_3$
	405			407	
	499		 $CH_3(CH_2)_{16}COOCH_2\overset{CH_2OAc}{\overset{\vert}{C}}HOAc$	486	
	638			542	
	579			492	
	771		Codeine(methyl morphine)	670	morphine
$(BrC_6H_4NH)_2CHCOC_6H_5$	272	 $(C_2H_5S)_2\overset{S}{\overset{\Vert}{P}}CH_2\overset{S}{\overset{\Vert}{P}}(SC_2H_5)_2$	1-Acetyllysergic acid diethylamide	541	Lysergic acid diethylamide

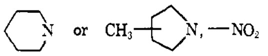
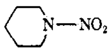
[a] See footnote *a*, Table VIII. [b] For identification by the criteria of Table II, a second compound of the top five choices must also contain the structural moiety if it requires at least four WLN symbols for its description.

tetraethyl-*S*,*S*′-methylenebisphosphorodithioate in Table IX; no closely similar compounds were found by STIRS because none was present in the reference file, a fact which is indicated by the low MF11 value.

Structure Elucidation Utilizing All Match Factors. Ten unknown spectra were chosen at random, and the criteria of Table II were applied by a chemist who was given the match results but no other spectral or structural information. The major structural features identified by the chemist[23] using STIRS are shown in Table X; note that use of even the molecular ion information (which was *not* given to the chemist) would have allowed near completion of the structure in several cases.

Obviously the reliability of the structural feature identifications can be increased by increasing the stringency of the criteria of Table II; this has the offsetting disadvantage of reducing the amount of information

Table X. Structural Features Found by STIRS from Individual Match Factor Criteria[a]

Predicted	Actual
$-C_6H_4NHCONHC_6H_5$	p-$CH_3OC_6H_4NHCONHC_6H_5$
$CH_3C_6H_4Cl$	$(CH_3)_3C(HO)C_6H_3Cl$
Pyridyl—$OCOCH_3$	4-Pyridyl—CH_2OCOCH_3
CH_3—pyridyl—$SC(CH_3)_3$	C_6H_5—pyridyl—$SC(CH_3)_3$
$ClC_6H_4COOCH_3$	o-$ClC_6H_4COOCH_3$
t-Bu—phenyl—OH, —NH, —Me	2-NH_2-4-Me-6-t-Bu-phenol
$HOCH_2CH_2NHCO(CH_2)_n$–	$HOCH_2CH_2NHCO(CH_2)_{10}CH_3$
CCl_2, -COO-, $-CH(CH_3)_2$	$CHCl_2COOCH(CH_3)_2$
C_6H_5O–, phenyl, $-CH_3$	$C_6H_5OC_6H_4OCH_2CH_2OCH_3$
	

[a] No molecular weight nor elemental composition information was utilized for these predictions.

Table XI. Match Results for 1-Octadecanol as Unknown with Various Trial Molecular Weights

	Top 10 MF5 Values			Top 10 MF6 Values			Top 10 MF11 Values	
Mol wt	Alkanols	1-Alkanols $>C_{12}$	MF5	Alkanols	1-Alkanols $>C_{12}$	MF6	1-Alkanols $>C_{12}$	MF11
252	1	0	666–833	0	0	421–571	9	421–584
253	0	0	666–727	0	0	750–800	8	420–620
267	1	0	571–666	0	0	666–800	7	426–620
270	10	5	666–800	8	6	727–1000	10	590–695
279	0	0	375–444	0	0	400–545	9	402–600
280	4	0	375–500	0	0	363–600	9	402–600
281	0	0	500–666	0	0	833	8	406–633
282	0	0	444–666	1	0	428–727	9	396–666
283	0	0	545–666	2	0	461–800	10	402–600
284	0	0	500–666	9	8	800–909	10	541–812

retrieved. On the other hand, additional structural *possibilities* can be obtained by relaxing the criteria. The data of Tables IV and VI indicate that much higher reliability is possible for particular structural features with selected match factors; note the variability of prediction rates of MF1, MF2, and MF5 for "nonaromatic" (92, 100, and 87%, respectively), and for saturated oxygen (100, 96, and 100%). Reliability is also increased if elemental composition information on the unknown spectrum is available from exact mass measurements. At present this is only applied to check the STIRS results; however, the STIRS system could be modified to utilize elemental compositions instead of masses.

Interactive Search. The program for molecular weight determination was tested using 1-octadecanol, molecular weight 270. Its spectrum does not contain an $M\cdot^+$ ion; the $(M - 18)\cdot^+$ ion at m/e 252 actually fulfills the requirements of the molecular ion identification program. Table XI summarizes the top ten MF5, MF6, and MF11 selections for a variety of trial molecular weights. 1-Alkanols are strongly indicated by all trial molecular weights for the MF11 values, but only by the homologous mass 270 and 284 choices for MF6 (large neutral losses). Finally, 1-alkanols are uniquely indicated by the correct mass 270 choice for MF5 (small primary neutral losses).[3] Only preliminary evaluations have been made of the MF11 feedback option and the shift technique, but the results are promising.[3]

Retrieval of Reference Spectra. If the reference file contains a spectrum of the unknown compound, STIRS will usually assign to it both the highest fingerprint (MF10) and overall (MF11) match factors. The main exception involves compounds such as m- and p-xylene which give closely similar spectra. This capability was tested with spectra for which another spectrum of the same compound, measured on a different instrument or under different conditions, was present in the file. For 28 spectra of a wide variety of compounds, the MF10 and MF11 values of the same compound were the highest for 26 and 27 cases, respectively. Neither retrieved the reference spectrum of 2-monolaurin, but this had been obtained at a nominal electron energy of 6 eV *vs.* 70 eV for the unknown; however, all compounds giving the top of 10 MF11 values were fatty acid monoglycerates, and the highest value was that of 1-monolaurin. Using examples given for other matching systems[8–14] STIRS appeared to perform as well or better, but again quantitative comparisons are not meaningful. In a number of cases the retrieval capabilities of MF11 also appear to be independent of errors in the spectral data to a surprising degree.

Conclusions

Implementation of the system improvements suggested in the text and more extensive testing are in progress here. However, we feel that the present results establish the general validity of the STIRS concept, and suggest its application to other kinds of spectra. The system can provide useful structural information on virtually any type of unknown utilizing only reference spectral data; no training of the system is necessary, and no correlation information need be supplied by the interpreter. Computer time requirements are kept within reasonable limit by utilizing present knowledge of mass spectral reactions in the design of the match factors. In the application of STIRS to 110 spectra, a large number and variety of structural functionalities were reliably identified. In a number of cases, STIRS gave information which, we feel, a trained mass spectrometrist could not have deduced. For complete unknowns, STIRS saves the expert as well as the novice substantial time in identifying the compound class and structural features; for maximum information recovery from the spectral data, however, the scientist should do the final analysis using the STIRS results with any other available data.

Acknowledgments. We are deeply indebted to Drs. S. Abrahamsson, E. Stenhagen, J. W. Serum, A. Tatematsu, R. G. Werth, and I. Sakai for help in collecting and checking the reference data and for WLN assignments, to Miss G. Pesyna for assistance with programs, to Dr. John McGuire for supplying challenging unknowns, and to the National Institutes of Health (GM 16609) for generous financial support.

Activation and Stabilization of Small Molecules by a Chelating Triphosphine–Rhodium(I) Complex. A Correlation of the Chemistry with the Structures of Chlorobis(3-diphenylphosphinopropyl)phenylphosphinerhodium(I) and Its Nitrosyl Derivative[1]

Thomas E. Nappier, Jr.,[2a] Devon W. Meek,*[2a]
Richard M. Kirchner,[2b] and James A. Ibers[2b]

Contribution from the Departments of Chemistry,
The Ohio State University, Columbus, Ohio 43210, and
Northwestern University, Evanston, Illinois 60201. *Received October 31, 1972*

Abstract: The tridentate ligand $C_6H_5P[CH_2CH_2CH_2P(C_6H_5)_2]_2$, L, forms RhLX complexes (X = Cl, Br, I), which are dissociatively stable in nonaqueous solvents. The chloride complex, RhLCl, reacts readily with a variety of small molecules to form five-coordinate, nonionic derivatives, RhLCl·A, where A = BF_3, CO, SO_2, and O_2. The S_2 adduct, RhLCl·S_2, was prepared by treating RhLCl with S_8 in benzene. Cationic complexes of RhLCl were made by two general methods: (1) addition of a positively charged ligand (electrophile), or (2) displacement of the chloride ion by a neutral ligand. The most important difference between the two methods is that rhodium remains Rh(I) in the latter case, whereas in the first case oxidation formally to Rh(III) occurs. For example, the five-coordinate, cationic complexes $[RhLCl\cdot A]^+$ were prepared by treating RhLCl with the cationic reagents H^+, NO^+, N_2Ph^+, and CH_3CO^+. Four- and five-coordinate cations $[RhL(CH_3CN)]^+$, $[RhL(CO)]^+$, and $[RhL(CO)_2]^+$ can be obtained by chloride displacement from RhLCl in polar solvents. The interconversions and chemical reactions of some of the above complexes are presented. The structures of the parent compound, RhLCl, and its nitrosyl derivative, $[RhLCl(NO)][PF_6]$, have been determined from three-dimensional X-ray data collected by counter methods. The yellow parent compound crystallizes in space group C_{2h}^2-$P2_1/m$ of the monoclinic system with two molecules in a unit cell of dimensions $a = 9.312$ (1), $b = 21.187$ (2), $c = 8.851$ (3) Å, and $\beta = 105.39$ (1)°. The observed and calculated densities are 1.36 (1) and 1.38 g cm^{-3}, respectively. Full-matrix least-squares refinement gave a final value of the conventional R factor (on F) of 0.035 for the 2566 reflections having $F^2 > 3\sigma(F^2)$. The structure consists of discrete monomers, and the coordination geometry around the metal is square planar. The molecule has imposed m symmetry. Some important bond lengths in the parent compound are as follows: Rh–P(terminal), 2.288 (1); Rh–P(central), 2.201 (2); and Rh–Cl, 2.381 (2) Å. The nitrosyl derivative crystallizes in space group D_{2h}^{16}-$Pnma$ with four molecules in a unit cell of dimensions $a = 24.350$ (5), $b = 15.196$ (2), and $c = 9.914$ (2) Å. The observed and calculated densities are 1.57 (1) and 1.58 g cm^{-3}, respectively. Full-matrix least-squares refinement gave a final value of the conventional R factor (on F) of 0.055 for the 1282 reflections having $F^2 > 3\sigma(F^2)$. The structure consists of discrete monomers. Addition of a nitrosyl ligand to the parent compound, RhLCl, gives a tetragonal pyramidal coordination around the rhodium atom of $[RhLCl(NO)][PF_6]$, with imposed mirror symmetry as in the parent compound. The nitrosyl ligand itself is bent, with a Rh–N–O bond angle of 131 (1)°. Some important bond lengths in the nitrosyl derivative are as follows: Rh–P(terminal), 2.374 (3); Rh–P(central), 2.282 (4); Rh–Cl, 2.408 (4); and Rh–N, 1.91 (2) Å. The relationship between structure and reactivity is discussed.

Complexes containing a transition metal with a d^8 electronic configuration have been used in many important and diverse chemical reactions involving (1) the homogeneous hydroformylation,[3] oxidation,[4] and hydrogenation[5] of olefins; (2) oxidative-addition reactions;[6] (3) the Lewis basicity of metal complexes;[7] (4) the coordination of small molecules such as O_2 and P_4;[8] and (5) the stabilization of small

---

[1] Presented in part at the Winter Meeting of the American Crystallographic Association, Albuquerque, N. M., April 1972, Abstracts, p 45; and at the XIV International Conference on Coordination Chemistry, Toronto, June 1972, Proceedings, p 649. A preliminary communication on part of this chemistry has appeared: T. E. Nappier, Jr., and D. W. Meek, *J. Amer. Chem. Soc.*, **94**, 306 (1972).

[2] (a) The Ohio State University; (b) Northwestern University.

[3] (a) C. W. Bird, *Chem. Rev.*, **62**, 283 (1962); (b) G. Yagupsky, C. K. Brown, and G. Wilkinson, *J. Chem. Soc. A*, 1392 (1970); (c) C. K. Brown and G. Wilkinson, *ibid.*, 2753 (1970).

[4] R. Jira, S. Sedlmeier, and J. Smidt, *Justus Liebigs Ann. Chem.*, **693**, 99 (1966).

[5] J. A. Osborn, F. H. Jardine, J. F. Young, and G. Wilkinson, *J. Chem. Soc. A*, 1711 (1966).

[6] J. Halpern, *Accounts Chem. Res.*, **3**, 386 (1970), and references contained therein.

[7] R. N. Scott, D. F. Shriver, and D. D. Lehman, *Inorg. Chim. Acta*, **4**, 73 (1970); D. F. Shriver, *Accounts Chem. Res.*, **3**, 231 (1970).

[8] A. P. Ginsberg and W. E. Lindsell, *J. Amer. Chem. Soc.*, **93**, 2082 (1971).