Problems and Progress in High Resolution Mass Spectrometry*

By K. Biemann and P. V. Fennessey

Department of Chemistry, Massachusetts Institute of Technology, Cambridge (Massachusetts, U.S.A.)

During the last few years mass spectrometry has developed from a technique used mainly by the physicist and analytical chemist to a method that is now an integral part of the instrumentation used by the organic chemist.

The reason for this sudden importance was the realization that it provides specific structural information not available by other measurements and this makes mass spectrometry a valuable complement to other spectroscopic techniques. Furthermore, its extreme sensitivity permits one to obtain mass spectra from subtances available only in amounts insufficient for other methods such as infrared or nuclear magnetic resonance spectra.

The interpretation of the mass spectrum of an organic compound is based on a logical, but empirical, recognition of the nature of the various particules produced on electron impact, which represent either the positively charged molecule or fragments thereof. The physical data displayed in the mass spectrum is the mass and relative abundance of these ions.

This principle shall be outlined using a by now classical example. Based on the known structure of deacetylaspidosperine (I) its mass spectrum could be interpreted as outlined in Fig. 1.[1] Once this had been accomplished the structure of a number of related alkaloids could be determined from a comparison of their mass spectra with that shown in Fig. 1.

The mass spectra of a wide variety of compound types have been studied and the resulting data were very successfully used for the determination of the structure of

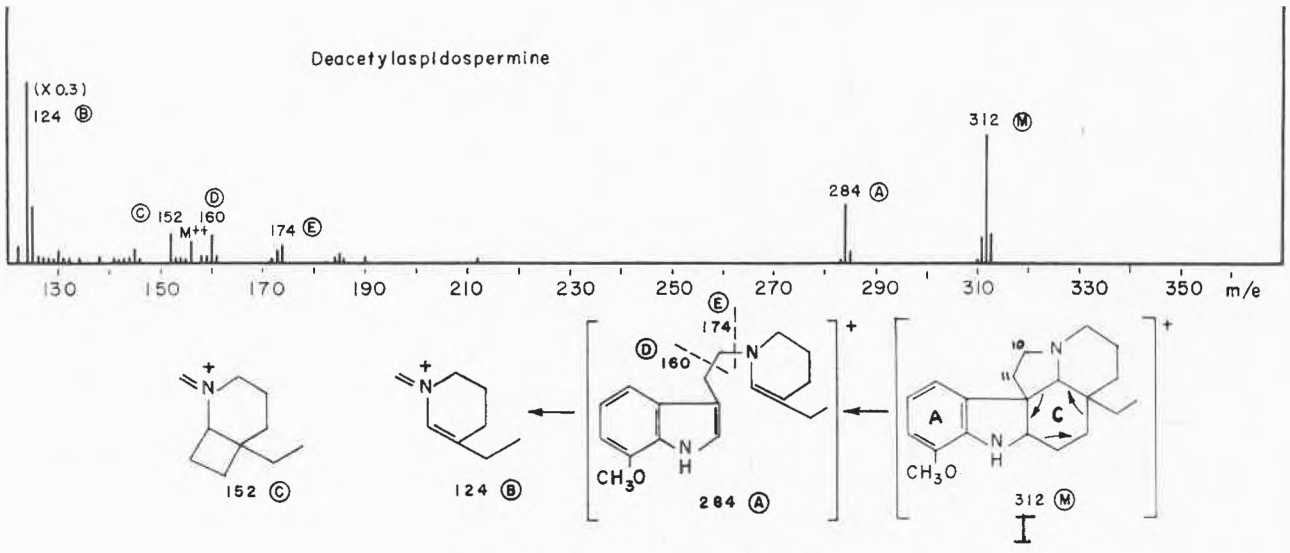


Fig. 1. Mass spectrum of an alkaloid

* Lecture delivered (by K.B.) at the 1st Swiss Chemical Convention on the occasion of the 3rd International Exhibition of Laboratory, Measurement and Automation Techniques in Chemistry, October 17th to 21st, Basle. Part V of the series «Computer-aided Interpretation of High Resolution Mass Spectra». For pare IV see ref. [11].

[1] K. Biemann, M. Spiteller-Friedmann and G. Spiteller, *J. Amer. Chem. Soc. 85* (1963) 631.

organic compounds.[2] Until very recently, all this work was done using conventional spectra of the type shown in Fig. 1.

It is obvious, however, that it would have been very difficult or impossible to interpret such a spectrum if absolutely nothing were known of the compound. Mass 124 can be composed of different sets of atoms, such as C_9H_{16} or $C_8H_{12}O$ or even $C_7H_{12}N_2$. For the ion at mass 312 many more combinations are possible and it is here where a more recent modification, namely high resolution mass spectrometry comes into play.

Table I

Masses of atoms		Masses of molecules or groups		Difference
^{1}H	1.007825	$C_{10}H_8$	128.0626	0.0939
^{12}C	12.000000	C_9H_{20}	128.1565	0.0126
^{13}C	13.003355	$C_8H_{18}N$	128.1439	
^{14}N	14.003037	$C_7H_{12}{}^{32}S$	128.0660	
^{16}O	15.994914	0.0034		
^{35}Cl	34.968853			
^{37}Cl	36.965903			

About eight years ago BEYNON showed[3] that it is practically possible to determine routinely the mass of an ion with an accuracy that enables one to calculate, within limits, its elemental composition. Table I illustrates this point. Summation of the accurate mass of 10 carbon atoms plus 8 hydrogen atoms gives 128.0626 while replacement of one carbon by 12 hydrogens leaves the nominal mass unchanged (128) but increases the fractional mass by 93.9 millimass units. Similar replacement of CH_2 by N changes the mass by a small amount, etc. Obviously, if one can determine the mass of an ion to better than $\pm$ 2 millimass units then one can unambiguously distinguish between the combinations listed in Table I. The advantage of high resolution mass spectrometry is thus very obvious, but it should not be overlooked that conventional mass spectra are equally useful or even superior, if the elemental composition is either irrelevant, such as in the comparison of two spectra for identity or simple structural differences, or if the composition is obvious from the nominal mass, for example in the case of hydrocarbons.

[2] For reviews see (a) K. BIEMANN, *Mass Spectrometry, Applications to Organic Chemistry*, McGraw-Hill Book Company, New York 1962; (b) H. BUDZIKIEWICZ, C. DJERASSI and D. H. WILLIAMS, *Interpretation of Mass Spectra of Organic Compounds*, Holden-Day, San Francisco 1964; (c) H. BUDZIKIEWICZ, C. DJERASSI and D. H. WILLIAMS, *Structure Elucidation of Natural Products by Mass Spectrometry*, Vol. 1: *Alkaloids*, Holden-Day, San Francisco 1964; (d) H. BUDZIKIEWICZ, C. DJERASSI and D. H. WILLIAMS, *Structure Elucidation of Natural Products by Mass Spectrometry*, Vol. 2: *Steroids, Terpenoids, Sugars, and Miscellaneous Classes*, Holden-Day, San Francisco 1964.

[3] J. H. BEYNON, *Advances in Mass Spectrometry*, Vol. I, Ed. J. D. WALDRON, Pergamon Press, Oxford 1959, p. 328.

To obtain such accurate mass measurements one uses so-called double focusing mass spectrometers. Of the two commercially available systems, the NIER-JOHNSON and MATTAUCH-HERZOG geometries, respectively, we use the latter (CEC 21–110) which employs electrostatic energy focusing followed by a magnetic field which refocuses all ions in one plane. Placing a photographic emulsion in this focal plane permits one to record all ions simultaneously and continuously, without scanning any one parameter. The spectrum is then a set of lines, much like an emission spectrum, and the distance between lines is directly proportional to the difference between the square roots of the corresponding masses. The mass measurement is thus reduced to the problem of accurate distance measurements. Most importantly, it permits determining the accurate mass of all ions in the spectrum rather conveniently and this is the aspect in which our approach differs from the earlier one introduced by BEYNON.

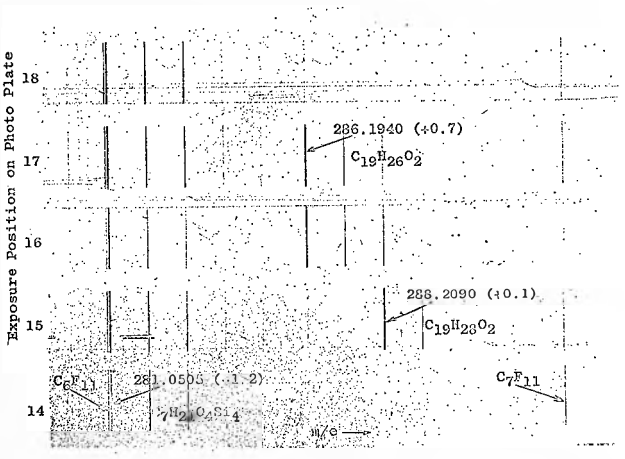


Fig. 2. Portion of photographic plate from a high resolution mass spectrometer (from ref. [4])

A small sector of such a plate is shown in Fig. 2. Each horizontal group is part of one spectrum and covers the region from mass 281 through 293. The line on the very left is due to C_6F_{11} and the one on the right to C_7F_{11} from a fluorocarbon which is always added to provide standards of known mass. After measuring the position of all lines one can calculate the mass of any ion between these standard lines simply from the relationship of distance to square root of mass. Some of these results are indicated in the figure, the value in parentheses giving the difference in millimass units between the mass found and that corresponding to the elemental composition assigned. The compounds in question are androstanedione and androstenedione; the spectra were taken during the time these compounds were eluted from a gas chromatograph which was directly coupled to the mass spectrometer.[4] Exposure 15 was the front and exposure 17 was the tail of an incompletely resolved

[4] J. T. WATSON and K. BIEMANN, *Anal. Chem. 37* (1965) 844.

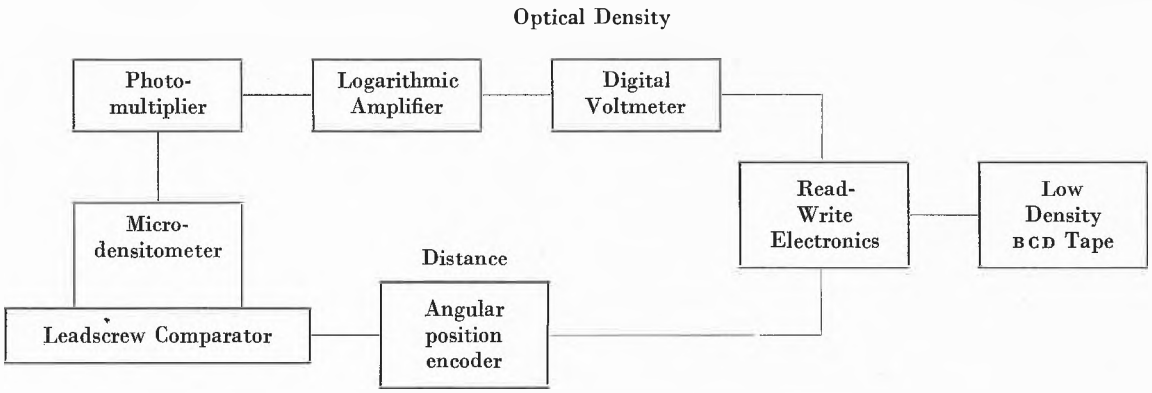


Fig. 3. Block diagram of automatic comparator-densitometer

peak containing these substances, illustrating the value of such measurements on gas chromatographic peaks. The doublet at mass 281 is due to contributions of the liquid phase of the gas chromatographic column, because its mass corresponds to $C_7H_{21}O_4Si_4$. It is present in all spectra and increases from exposure 14 to 18 because the temperature and thus the "bleeding" was increased.

The problem of measurement is simple if one were only interested in the molecular ion. The entire spectrum which is on the plate and of which Fig. 2 shows only a very small part, consists, however of many hundred lines. Their masses represent valuable information which should not be neglected but this requires the measurement of all these line centers as well as their density. Manual measurement is practically impossible if one considers that each single measurement represents a six or seven digit number. Over the past few years we have evolved a fully automatic system which records the density profile in digital form on magnetic tape and leaves it to the computer to find the center of the line.[5] A similar system is now also available commercially.[6] Most recently we have even eliminated the intermediacy of the magnetic tape by reading the density data directly into the computer.

The density profile of a single line is nearly gaussian. In order to avoid processing density data between lines, the densitometer senses density continuously but activates the read-write electronics only when a certain threshold is reached. From there on density is recorded every 1/2 micron, dictated by a digital encoder on the precision screw of the comparator. When the threshold is reached again, the distance is read off the encoder onto the tape thus defining the absolute position of each one of the equidistant data points.

The computer finds the maximum either by interpolation of the three highest points or by using the half width at various density values. A combination of these data also reveals whether or not one is dealing with an unresolved multiplet lacking distinct maxima, in which case a more complex curve-fitting routine is required.[6] A block diagram of this system is shown in Fig. 3.

It is important to realize that defining the line positions is only the first step in the immediate goal, the determination of the elemental compositions of the corresponding ions. First, distances have to be converted to masses using a set of lines of known mass distributed over the entire mass region recorded. These are due to a standard compound whose spectrum is recorded simultaneously with that of the sample. A fluorocarbon is used to prevent coincidence with the lines of the latter. The spectrum of this compound (intensities and exact masses of all ions) is a part of the computer program and one needs only to indicate the position of the first two such standard lines. This permits the computer to identify the entire set of standard lines by extrapolation from one to the next and to calculate the values of all masses falling between each pair of standard lines. It is then a matter of merely eliminating those masses which correspond to that part of the fluorocarbon spectrum which had not been used as standard lines. This is easily done in the computer which had been fed the complete mass spectrum of the standard.

It remains to mention that during the process of evaluating the density profile all data points had been converted to intensities, using a predetermined blackening curve of the emulsion, and then integrated. To avoid the necessity of exactly reproducing conditions of emulsion treatment an intensity standard, e.g. xenon, is also added. When its lines are encountered by the computer in the region between mass 122 and 136 it uses the known abundance ratio of the various xenon isotopes to correct all intensities once more, i. e. it functions as an internal standard for each individual spectrum. These are some of the additional advantages of the use of a computer, as it permits employing procedures too time consuming to be done manually.

[5] (a) D. DESIDERIO and K. BIEMANN, *12th Annual Conference on Mass Spectrometry*, Montreal, June 1964; (b) D. DESIDERIO, Ph. D. thesis, MIT, December 1965; (c) K. BIEMANN, P. V. FENNESSEY and J. M. HAYES, *Proceedings of the Filmed Data and Computers Seminar*, Society of Photo-optical Instrumentation Engineers, Boston, June 1966, p. II-1.

[6] R. VENKATARAGHAVAN, F. W. MCLAFFERTY and J. W. AMY, *Anal. Chem. 39* (1967) 178.

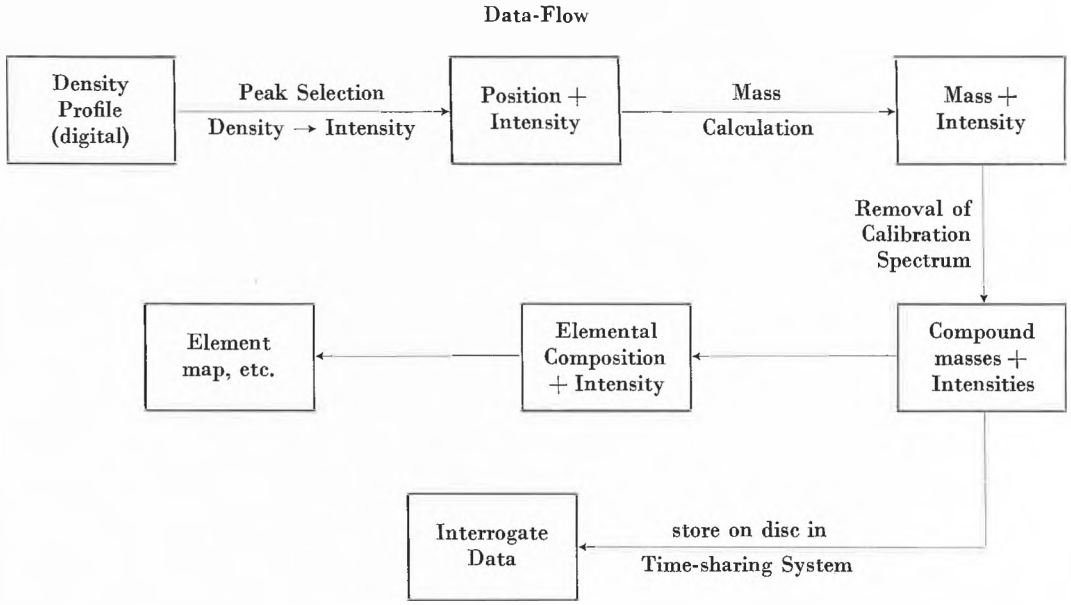


Fig. 4. Data Flow within the computer

Table II. Section of a high resolution mass spectrum

Intensity *	Mass spectrum
0	100.0892
0	101.0610
0	107.0861
0	108.0940
46	109.1017
111	110.1092
0	111.0816
10	111.1134
0	112.1252
29	113.0969
1	114.1014
0	123.1177
43	127.1131
29	138.1419
5	139.1443
0	141.1281
0	142.1351
0	143.1069
0	155.1431
0	169.1584
177	170.1668
273	171.1751
26	172.1782

* Intensities are all relative to most abundant ion = 999 units. Thus an intensity of 0 means that an ion has an abundance of less than 0.1% of the most abundant.

At this stage one now has essentially a mass spectrum, namely mass and abundance of each ion, listed as shown in Table II. Intensity is relative to the most abundant compound ion, equalling 999 units. The spectrum obviously consists not only of the few shown in Table II but of many hundred ions. It becomes immediately apparent that the display of a high resolution mass spectrum in tabular form, listing intensity vs. mass represents the information in a form almost impossible to grasp. One suddenly realizes that the mass of the ion is of no interest at all and that the corresponding elemental composition is the data actually sought. This is the point where the philosophy of conventional and high resolution mass spectrometry start to deviate.

The entire data-flow is shown in Fig. 4. After peak center selection and conversion of density to integrated intensity to give position-intensity pairs; mass calculation to give mass-intensity pairs, subtraction of the calibration spectrum leaving the contribution of the compound itself; the elemental composition must be assigned to each mass and the results presented in a suitable form or, alternately, the mass data stored for further treatment as shall be discussed later.

To assure maximum flexibility, the elemental compositions are computed for each ion separately as schematically illustrated in Fig. 4.[5c] By division of the mass by that of CH_2 and of the remainder by that of hydrogen one can recognize ions C_nH_{2n+x} where $x = 2$, 1 or 0 (i. e. alkane, alkyl or olefin ions). If that remainder did not equal 0, 1 or 2 hydrogens within the experimental confidence limits* (e. g. 2 millimass units below mass 500, increasingly larger above that), the number of carbons is increased by one, multiplied by 12, the result subtracted from the total mass and the rest divided by the mass of hydrogen. If the remainder is now within the limits one has found the number of carbons and hydrogens composing the ion. If not, the cycle continues until the hydrocarbon combinations are exhausted. Then the mass of the first hetereoatom is subtracted and the process is repeated to exhaustion, the next heteroatom subtracted, etc. Carbon-13 can be treated as a heteroatom.

This particular approach leads to a list of the elemental compositions of all the ions by increasing mass. For certain aspects of interpretation which will be discussed later, it is more useful to display selectively the elemental composition of a certain type of ions, i. e. homologous

* Not to be confused with average error which is of course smaller.

series. An entirely different method of calculation has been developed for this purpose and consists of consecutively adding the exact mass of CH_2 to that of $C_0H_xN_aO_bS_c$ ... and searching the list of accurate masses of the ions in that spectrum for the masses representing that series.[7] Continuous incrementing of the values x, a, b, c ... eventually covers all ions. The choice between the above two forms of computation depends on the final form of data presentation (i. e. element map of specific ion types).

This procedure would appear to be lengthy but is surprisingly fast. It requires about 1/2 minute of computer time for about 500 ions using an IBM 7094 computer.

Table III. Some pairs of combinations of elements differing by less than 5.0 millimass units (m.m.u.) and occurring frequently in the spectra of organic molecules

Compositions	Difference (m.m.u.)
C_5—N_2O_2	4.02
C_3N—H_2O_3	2.68
C_2H_2O—N_3	1.34
C_3—$^{32}SH_4$	3.3
HFO—C_3	1.1
$^{13}CHNO$—C_5	4.0
CH—^{13}C	4.47
$^{13}CF_2$—$CH_4{}^{35}Cl$	0.009

One of the major difficulties in assignment of a unique elemental composition is due to the coincidence of the masses of certain combinations of elements. Table III lists some of the more commonly encountered pairs of small mass differences. Obviously, assignment of combinations differing by C_2H_2O vs. N_3 requires reliable accuracy of about ± 0.5 millimass units and this is still difficult to achieve routinely. Thus the computer will always provide two answers in such cases. Here as in many other cases, the decision is often possible considering other features of the spectrum, because all these pairs differ greatly in heteroatom composition. The notable exception is ^{12}CH vs. ^{13}C, which is difficult to resolve above mass 150. Fortunately it does not complicate the interpretation because the presence or absence of a ^{13}C-isotope peak can be deduced from the ^{12}C-species one mass unit lower.

One can now ask the computer to print the results. A very small part of a typical spectrum is shown in Table IV. Comparing it with Table II the progress is clearly evident as it contains in addition to intensities and accurate masses also the elemental compositions, their theoretical mass and the difference between that and the found mass (in millimass units). But a new problem becomes also evident, namely extracting useful information out of all the elemental compositions. In the present form they are ordered according to increasing mass which places totally unrelated entries next to each other. If one sees that there is $C_7H_{13}O_1$ present, one wishes to know whether there is $C_8H_{15}O_1$ also, and if so whether it is more or less abundant and so forth, with increasing hydrogen content. Similarly, the abundant $C_{11}H_{23}O_1$ ion leads to the question whether this homologous series is continuous and at what carbon number it begins and where it ends.

Table IV. Section of high resolution mass spectrum displaying the elemental compositions of the ions

Intensity	Determined	Calculated	Difference	C-12	C-13	H	N	O
0	100.0892	100.0888	0.38	6	0	12	0	1
0	101.0610	101.0602	0.74	5	0	9	0	2
0	107.0861	107.0860	0.02	8	0	11	0	0
0	108.0940	108.0939	0.09	8	0	12	0	0
46	109.1017	109.1017	−0.02	8	0	13	0	0
111	110.1092	110.1095	−0.35	8	0	14	0	0
0	111.0816	111.0809	0.60	7	0	11	0	1
10	111.1134	111.1129	0.49	7	1	14	0	0
0	112.1252	112.1252	−0.00	8	0	16	0	0
29	113.0969	113.0966	0.26	7	0	13	0	1
1	114.1014	114.0999	1.40	6	1	13	0	1
0	123.1177	123.1173	0.32	9	0	15	0	0
43	127.1131	127.1122	0.81	8	0	15	0	1
29	138.1419	138.1408	1.04	10	0	18	0	0
5	139.1443	139.1442	0.09	9	1	18	0	0
0	141.1281	141.1279	0.16	9	0	17	0	1
0	142.1351	142.1357	−0.66	9	0	18	0	1
0	143.1069	143.1072	−0.30	8	0	15	0	2
0	155.1431	155.1435	−0.49	10	0	19	0	1
0	169.1584	169.1592	−0.83	11	0	21	0	1
177	170.1668	170.1670	−0.26	11	0	22	0	1
273	171.1751	171.1748	0.21	11	0	23	0	1
26	172.1782	172.1782	−0.04	10	1	23	0	1

The obvious answer lies in correspondingly rearranging of the data, to group them according to heteroatom content. This compressed presentation has been termed "element map"[8] and an example is shown in Fig. 5. It represents the complete high resolution mass spectrum of adenosine[9] ordered according to heteroatom content, as indicated on the top line. The first group represents only those ions containing C and H, the next C, H and N and so on to the last, C, H, O_4N_5. Carbon and hydrogen content increases downward. Much can be deduced about the distribution of the heteroatoms in this molecule from this presentation. All oxygen containing ions retaining all five nitrogen atoms contain up to ten C-atoms indicating that three oxygens can be lost from the periphery without losing carbon or nitrogen. The oxygen containing ions lacking nitrogen have only up to five carbons, indicating that all oxygen is part of a C_5 moiety. All five nitrogens are found with as few as five carbon atoms, which must be the other five, recalling that the total number of carbons is ten. The lower right corner of Fig. 5 reveals that the molecular composition

[7] K. Biemann, W. J. McMurray and P. V. Fennessey, *Tetrahedron Letters 1966*, 3997.

[8] K. Biemann, P. Bommer and D. M. Desiderio, *Tetrahedron Letters 1964*, 1725.

[9] S. Tsunakawa, to be published.

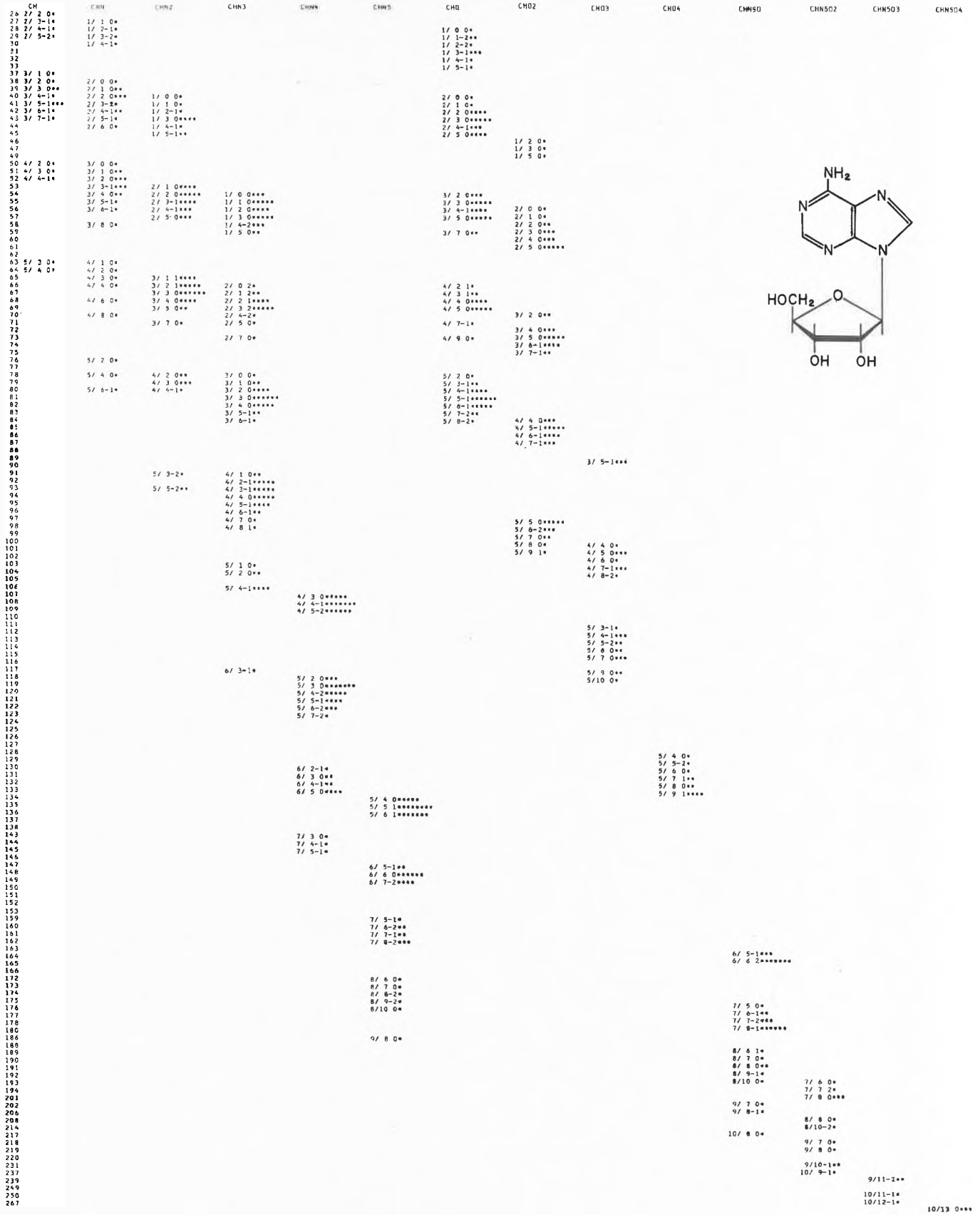


Fig. 5. Element map of Adenosine

is $C_{10}H_{13}O_4N_5$. The next number is the difference, in millimass units, of found and theoretical mass and the asterisks are a semi-logarithmic indication of the intensity, eight asterisks being the maximum. The absence of any O_4N_5 ions with less than 10 C-atoms indicates that there is no methyl or alkyl group present, because its loss would result in such an ion. One oxygen can be lost as OH, H_2O or CH_3O, i. e. one of the hydroxyl groups or C-5 of the ribose. Further loss of H_2O leads to ions with progressively fewer oxygen atoms.

A close look at the nitrogen free ions in Fig. 5 reveals the pentose, 4 oxygens require 5 carbons, while 3 oxygens can do with four or even three, etc. Finally, of the nitrogen containing, oxygen free ions those with five nitrogens contain up to 8 C-atoms (three of the sugar after elimination of all oxygens) but a minimum of 5 C-atoms is necessary to retain all five nitrogens. $C_5H_5N_5$ is in fact, one of the most abundant ions which is indicative of the presence of an adenine derivative. It is clear that the data presented in this form are structurally extremely informative and another example will be discussed later in a slightly different context.

It should be pointed out that although the above described data acquisition system was developed in the authors' laboratory using a spectrometer of the MATTAUCH-HERZOG geometry, it was recently shown[10] that the same approach can also be used with spectrometers of the NIER-JOHNSON type. While the latter geometry has only a focal point rather than a focal plane, the above described function of the photographic plate must be performed by high-speed recording of the output of the mass spectrometer's electron multiplier on an analog magnetic tape which is then digitized and processed by the computer. From that point on, the data flow is basically the same as shown in Fig. 4, except that it starts with an intensity vs. time profile.

The above discussion emphasized that the elemental composition of the prominent and frequently occurring ions becomes an integral part of the interpretation as it tells much about the distribution of the hetereoatoms within the molecule and its degree of saturation in the various parts. For example $C_5H_9O_4$ is a rather saturated system but $C_5H_5N_5$ is a highly unsaturated ion. This factor can only be realized when considering all ions and their distribution, but not by measuring very few selected ones as was the earlier concept of high resolution mass spectrometry. It has to be kept in mind, however, that the element map shows only the elemental composition of the ions produced and if rearrangements of groups take place some of these compositions may at a first glance not be compatible with the structure of the original molecule. Here again, consideration of all the ions, rather than a selected few, prevents one from being misled by a single one, which might be such a rearrangement ion. Rearrangement of hydrogen, a rather frequent process, does of course not change the heteroatom content and is thus rather inconsequential.

At this point one has reached a stage at which one can visually inspect the vast amount of information contained in a high resolution mass spectrum and can interpret it sensibly.

Once one has done this for some time one begins to realize that many of the steps taken in the interpretation of the data are of a numerical nature, such as substraction of one elemental composition from another, weighing the relative abundance of one heteroatom type to another, checking the largest as well as smallest number of carbon atoms associated with a certain set of heteroatoms, etc. Those are, of course, processes which a computer can easily do in a few milliseconds and with a thoroughness and reliability far exceeding human patience. It is thus only natural to try to have the computer, which had to be used to generate the original data in the first place also take over this task.

At first we had attempted to do this in the usual "batch processing mode" but it soon became apparent that it is exceedingly difficult to write all these arguments and possibilities into a single program and expect the computer to provide a single answer. For a given set of data there are too many possibilities which have to be considered and evalutated; in short, we know far too much about chemistry and mass spectrometry to be able to incorporate all this into a single program capable of interpreting the spectra of a wide variety of compounds. This is only possible if limited to a specific group of compounds whose mass spectrometric behavior is strictly predicitable and we[11], and recently also others[12], have accomplished this successfully for the determination of the amino acid sequence in small peptides which the computer can deduce directly given only the mass spectrum, i. e. all accurate masses and intensities.

For more general problems we have found it necessary to combine the computer's capacity of quickly and exhaustively performing many numerical comparisons and evaluations, with the human mind's enormous capacity of recognition and decision-making based on knowledge and experience.[13] For this purpose the person has to be in *direct contact* with the computer to ask it questions, select the most plausible answer if more than one is given and decide what to do next. This can be economically done using a so-called "time-shared computer" which permits many investigators, apparently simultaneously, to use a very large and highly sophisticated computer system, since it takes only a fraction of a second to perform the computations necessary to ans-

[10] W. J. MCMURRAY, B. N. GREENE, and S. R. LIPSKY, *Anal. Chem. 38* (1966) 1194.

[11] K. BIEMANN, C. CONE, B. WEBSTER and G. P. ARSENAULT, *J. Amer. Chem. Soc. 88* (1966) 5598.

[12] M. SENN, R. VENKATARAGHAVAN and F. W. MCLAFFERTY, *J. Amer. Chem. Soc. 88* (1966) 5593.

[13] K. BIEMANN and P. V. FENNESSEY, *14th Annual Conference on Mass Spectrometry*, Dallas (Texas), May 1966.

wer a question and serve many others while the first user reads the answer and decides how to proceed. Questions are asked and answers received via an electric typewriter, linked to the computer via an ordinary telephone line. Such a typewriter can, therefore, be located in the laboratory quite far from the computer or even a few thousand kilometers away.

For the work described below we use the facilities of Project MAC at MIT which has been instrumental in the development of time-shared computers.[14] Simply, it consists of a central processing unit, the actual computer (an IBM 7094) with one memory containing the "supervisor program" which controls the entire process and lets one user after the other enter the "working" memory. All programs and data of the more than one hundred individuals working with the system are stored on discs. Data and programs can be written on these discs via cards or tapes, large volumes of results can be printed or displayed but the direct access is provided via typewriters (IBM 1050) of which more than two hundred are located throughout the institute and some of them scattered over the entire USA. One example shall illustrate the question and answer dialog between the chemist and the computer which has stored the data (masses and intensities) of all the spectra waiting to be interpreted. In the following figures commands or questions typed by the chemist appear in small letters while the answers or requests of the computer appear in capital letters. Numbers can, unfortunately, not be so differentiated.

```
start program inter

WHICH SPECTRUM DO YOU WANT TO INTERPRET
505-55-2
505-55-2     FOUND
  237 LINES BEGINNING AT  25 AND ENDING AT 336  TOTAL IONIZATION = 25593

T=  .205E 01 SEC.
```

Fig. 6

First one specifies to the computer what program is to be used, in this case "interpretation" (Fig. 6). The first question is answered with the code number of the spectrum to be interpreted, which causes the computer to look through its files, find it on the disc and to transfer the data into the working memory. It also tells the mass range of the spectrum and total intensity. All this takes two seconds (the second number is the exponent of ten). It should be noted that this spectrum is that of a compound of unknown structure, obtained upon hydrolysis of a nucleic acid.[15]

The next question is shown in Fig. 7: What is the elemental composition of the compounds? This can be deduced on the basis of the argument that if nitrogen or oxygen were present, ions below mass 100 would surely

[14] F. J. CORBATO and R. M. FANO, *Scientific American 215* (September 1966) No. 3, p. 129.

[15] K. BIEMANN, S. TSUNAKAWA, J. SONNENBICHLER, H. FELDMANN, D. DUTTING and H. G. ZACHAU, *Angew. Chem.* (Int. Ed. in English) *5* (1966) 590.

```
test c10h20n2o2

HET.  S-INT  PERCENT
  N    4660  27.161
  O    6559  38.229

T=  .180E 01 SEC.
```

Fig. 7

contain N or O. Upon the request "test $C_{10}H_{20}N_2O_2$" the computer calculates the elemental compositions of all ions below mass 100 and adds the relative intensity of all those containing one or two nitrogens and/or one or two oxygen atoms and computes the percentages in terms of the total below mass 100. Obviously if as much as 27% contain nitrogen and 38% oxygen, the compound must contain N and O. It took 1.8 second to arrive at this answer which is instantly typed out.

```
locate hal (chlorine,bromine) and sulfur

MASSES DIFFERING BY   1.99800
 INT     MASS  INT     MASS   RATIO
   0  52.0318    1  54.0337   0. / 1
   4  70.0416    0  72.0443   0. / 1
   0  90.0471  149  92.0496   0. / 1
   7  94.0416    0  96.0368   0. / 1
   1 100.0687    0 102.0689   0. / 1
   0 102.0476    7 104.0494   0. / 1
   0 103.0549    4 105.0560   0. / 1
   0 103.0757    1 105.0713   0. / 1
   0 103.0757    2 105.0712   0. / 1
   4 105.0560    0 107.0504   0. / 1
   1 105.0713    3 107.0721   .3/ 1
   2 105.0712    3 107.0721   .7/ 1
   0 115.0537    5 117.0535   0. / 1
   0 116.0604    4 118.0633   0. / 1

NO CHLORINE, BROMINE, OR SULFUR PRESENT.

T-   .78 SEC.
```

Fig. 8

Next we wish to know whether there is chlorine, bromine or sulfur present. Fig. 8 shows both question, method and answer. Since the heavy and light isotopes of Cl, Br and S as well as silicone differ by slightly less than two mass units, the computer searches for such pairs differing by 1.998 mass units ± 0.002 and computes their intensity ratios. If they are between 1:1 and 20:1, the elemental composition of the lighter component of the pair is calculated involving these elements. In this manner the presence or absence of these particular hetero-elements is detected. In this particular example no such pairs are present and the compound thus consists only of C, H, N and O.

Next question is how many of each are in the molecule? Asking the computer to print the ten highest masses it produces the answer shown in Fig. 9. Considering the intensities it appears that of these mass 335 is the most probable candidate for the molecular ion and that mass 336 represents the C-13 isotope peak.

```
print top 10

 INT      MASS
  83 230.10370
 180 232.11960
  49 233.12330
 130 246.13370
  28 247.13910
  32 292.10460
  44 305.14750
  35 320.13810
 273 335.16010
  59 336.16290

T=  .200E 00 SEC.
```

Fig. 9

```
call c99h99n8o12 ,335-336

CMAX  S-INT   X  DB-R  HETEROATOMS/CONCLUSIONS
  18    273 -17  12F   N= 6   O= 1
  15    273  -9   8M   N= 5   O= 4
  17    273 -11   7F   N= 2   O= 5
ION INTENSITY =  273   TOTAL INTENSITY = 25593

T=  .458E 01 SEC.
```

Fig. 10

This can easily be tested by asking the computer for the elemental composition of mass 335 (see Fig. 10). For a single mass one can permit up to 8 nitrogens and 12 oxygens, using the command "call". The computer presents the possible compositions in a highly informative manner, namely carbons, nitrogens and oxygens, but instead of the number of hydrogens it lists the number of rings and/or double bonds; it also notes whether it has a free valence and thus must be a fragment, or not, and thus can represent a molecule. Obviously the best candidate is $C_{15}N_5O_4$ with 8 rings or double bonds.

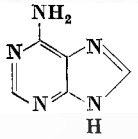
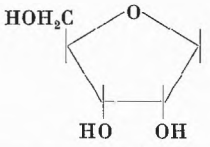
Another source of useful information are the intense peaks in the spectrum, which may be of particular structural significance. The information stored in the program contains a list of masses of ions which are known to be characteristic of certain structural features. The command "search intense ions" causes the computer to compare this list with the intense ions in the spectrum. The result (Fig. 11) may be puzzling at first, because it

```
search intense ions

THE FOLLOWING SUGGESTIONS ARE BASED ON THE CHARACTERISTIC INTENSE IONS

 INT      MASS
 230  31.01870    -CRR-O-R  (R)= H OR -C-CH-
 364  43.01820    R-CHH-CO- (R)= H OR -C-CH-
 221  43.05450    PROPYL
 268  44.02550    -CRR-CHO  (R)= H OR -C-CH-
 273  45.03420    -CRR-OR   (R) 1= ME, 3= H OR C-CH-
 287  57.03350    ME-CRR-CO (R)= H OR C-CH-
 164  59.01300    -CO-OME
 221  60.02090    RCRR-CO-OR (R)= H OR -C-CH-
 176  94.03960    PHENOXY + H
 111 133.05080    PENTOSIDE
 999 135.05530    ADENINE DERIVATIVE

T=  .600E 00 SEC.
```

Fig. 11

contains both trivial and, as we shall see later, unlikely (e.g. phenoxy + H or COOMe) suggestions. The last two entries are very suggestive, however, as they are based on ions of rather unique elemental compositions. They should therefore be considered first. Most of the others are due to contaminants in the specimen (last purification step was electrophoresis). The selection of the most significant pieces of information contained in Fig. 11 is one of the best examples for the need of human decision making in computer-aided interpretation of mass spectra.

Adenine has five carbons, five nitrogens and six double bonds and rings. A pentose has five carbons, five oxygens and one ring. Thus we lack a total of five carbons and one ring or double bond. To decide whether these are connected with the purine part, it is necessary to search for ions representing five nitrogens and seven double bonds or rings.

Adenosine: 	C_5N_5	6 DB-R
Pentose: 	C_5O_4	1 DB-R
M =	$C_{15}N_5O_4$	8 DB-R
Substituent	C_5	1 DB-R
On Adenine	$C_{10}N_5$	7 DB-R
On Pentose	$C_{10}O_4$	2 DB-R

```
call c99h99n5x-8,-7

S-INT  X DB-R INT/ Y INT/ Y INT/ Y INT/ Y INT/ Y    HETEROATOMS
 1293 -8   7F  97/ 6 552/ 7 354/ 9 290/10   0/ 0    N= 5
  559 -7   7M 559/10   0/ 0   0/ 0   0/ 0   0/ 0    N= 5

T=  .400E 00 SEC.


find c10h17o4

 MASS =201.11268  NOT PRESENT

T=  .000E 00 SEC.
```

Fig. 12

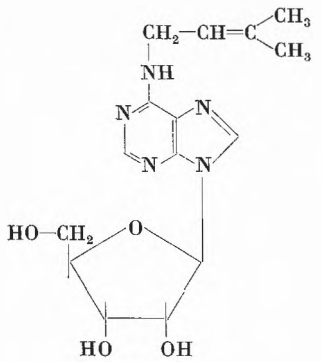
The question and answer is presented in Fig. 12, calling for a listing of the homologous ions of the type $C_yH_{2y+x}N_5$ where $x = -8$ for simple bond cleavage, or -7 for addition of one hydrogen. In the latter group one ion is found with 10 carbon atoms (symbol Y) indicating that all 5 carbons of the substitutent are on the purine. The fragment ion series shows loss of one carbon (C_9) but not of two (C_8 is missing), pointing to a geminal dimethyl group. To be certain we check whether the five carbons are indeed not on the pentose part by asking for $C_{10}H_{17}O_4$, which the computer finds to be absent.

```
call c99h99n1x0,1

S-INT  X DB-R INT/ Y INT/ Y INT/ Y INT/ Y INT/ Y    HETEROATOMS
  725  0   1F 247/ 1  97/ 2  64/ 3  73/ 4 244/ 5    N= 1
  137  1   1M  64/ 1  73/ 2   0/ 0   0/ 0   0/ 0    N= 1

T=  .400E 00 SEC.
```

Fig. 13

It now remains to see how the C_5H_9-unit is attached to the adenine ring. If it is on the amino group we would expect it to give a fragment containing this group plus NH. Searching for corresponding ions (Fig. 13), namely the type $C_yH_{2y+x}N$, for $x = 0$ or 1 tells us that the fragment ions are in fact present with up to five carbons on nitrogen which permits, in consideration of biogenetic factors, to write a plausible structure (II). The position of the double bond was deduced from an NMR spectrum which in spite of the very low concentration permitted one to recognize the two methyls at a double bond.[15]

This last point illustrates one other welcome aspect of this flexible combination of man and computer, as it



II

permits continuing use of other data and information to be included without actually incorporating information concerning other spectral data, origin or chemical properties into the computer program. It is well known how important it is to take all this information continuously into account.

The foregoing discussion illustrates the present stage of high resolution mass spectrometry and where it might go in the near future. It is abundantly clear that the exploitation of its great potential requires not only the acquisition of the rather expensive instrument itself but a large manpower pool to generate and digest the data which in turn requires a considerable amount of auxilliary equipment and the use of a modern and fast, although not necessarily very large computer. It also provides an occasion for the organic chemist to become accustomed to the use of computers which will play an increasingly important role also in other areas of chemistry over the next few years.

The authors would like to acknowledge the enthusiastic collaboration of many of their associates, particularly Drs. DESIDERIO, CONE and HAYES as well as thank Professor ZACHAU (Köln) for the nucleoside discussed above. The work was done in part with the computing facilities of the MIT Computation Center and of Project MAC and financially supported by grants from the National Science Foundation (GP-3734), National Institutes of Health (GM 09352) and the National Aeronautics and Space Administration (NGR 22-009-102) as well as an NIH-predoctoral Fellowship to P.V.F.