Tetrahedron Letters No.11, pp. 647-653, 1965. Pergamon Press Ltd.
Printed in Great Britain.

# COMPUTER-AIDED INTERPRETATION OF HIGH RESOLUTION MASS SPECTRA.[1]

K. Biemann and W. McMurray[2]

Department of Chemistry, Massachusetts Institute of Technology
Cambridge, Massachusetts

(Received 22 January 1965)

The interpretation of mass spectra involves a well chosen mixture of simple arithmetic, experience and chemical intuition. In the case of high resolution mass spectrometry, the amount of arithmetic required is vastly increased, but so is the significance of the data because one is no longer merely dealing with integral mass units but with elemental compositions. It is for precisely this reason that we believe that high resolution mass spectrometry is the physical method most suited to computer-aided interpretation. The computer can handle the arithmetic involved much faster, more reliably and more exhaustively. If provided with the necessary instructions, it may also handle those steps requiring experience and thus arrive at a pre-interpretation which may represent the solution of the problem in simple cases, or the basis for the final

---

[1] Paper XXVI on the Application of Mass Spectrometry to Structure Problems. For paper XXV see K. Biemann, Lloydia, in press.

[2] Present address: Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut.

interpretation by the chemist who, being relieved by the computer from the bookkeeping portion of the work, may use his intuition more effectively and thus be able to solve more formidable problems.

One aspect of the interpretation of a mass spectrum is the deduction of the exact molecular weight and thus of the molecular composition of a compound from the data contained in a mass spectrum, and this problem is a good illustration of the above statement. In most instances the peak corresponding to the molecular ion is the one of highest mass in the spectrum with exception of the corresponding isotope peaks. There are two situations in which this simple assumption does not hold; first, if the compound has a high tendency for fragmentation molecular ions are not produced in detectable abundance, and second, if there are present impurities of a mass higher than the molecule under consideration. These two situations are, in general, recognized only after the spectrum is found uninterpretable when assuming that the peak of highest mass represents the molecular weight.

However, if the spectrum was obtained with a spectrometer of a resolving power that permits the determination of the elemental composition of all the ions, then it is possible to set a number of criteria which uniquely determine the molecular ion:

(1) the specie may not contain any heavy isotopes.

(2) the number of hydrogens must be even if the number of nitrogens is even or zero, and must be odd if the number of nitrogens is odd.

(3) the mass difference between the molecular ion and the ions of lower mass must correspond to a combination of atoms that can be lost in a reasonable fragmentation process.

(4) if the peak of highest mass and not containing heavy isotopes does not fulfill these criteria, the molecular composition of the compound must differ from these ions by a combination of atoms that can be lost in a simple fragmentation process, and

(5) the elemental composition of the ions of mass lower than the molecular ion must not exceed for any one element the composition of the molecular ion.

All these decisions can fast and easily be made by the computer with a completeness that would in most instances go beyond the patience of a human interpreter. Since the full exploitation[3] of a high resolution mass spectrum requires already the use of the computer, a small step added to the general program produces a reliable suggestion of the molecular ion if the available data are sufficient to permit such a conclusion at all. In addition, the process may reveal impurities or contaminants because of the presence of some ions that are not in agreement

---

[3] K. Biemann, P. Bommer, and D.M. Desiderio, Tetrahedron Letters, No. 26, 1725 (1964). K. Biemann, J. Pure Appl. Chem. **9**, 95 (1964).

with an otherwise very probable molecular ion.

The FORTRAN Program which we have written involves the following:

The elemental composition of the ions produced by the compound, starting with the heaviest ion, are checked whether or not they fulfill conditions (1) and (2) above. If so, from the accurate mass of that elemental composition of the consecutively lighter ions are subtracted one after the other and the mass difference is checked vs. a list of the accurate masses of atoms or groups that can be easily lost from a molecular ion, i.e., H, $CH_3$, $NH_2$, $NH_3$, OH, $H_2O$, F, HCN, CO, $C_2H_4$, CHO, $C_2H_5$, etc., and combinations thereof (about 50 are presently used). Of course, only a limited but consecutive number of ions is checked from the ion of highest mass down (see Tables 1 and 3). The presence of any such fragment ion is of course of value for the interpretation of the spectrum itself, as shall be briefly pointed out later. If the mass from which the ions were subtracted is in fact the molecular ion one will find that a considerable number of the mass differences fit the mass of these small common fragments in the list, and that there are only few which do not because the type of small groups that can be lost are rather limited. If, on the other hand, only a small fraction of the presumed fragment ions fit such a relationship to the assumed molecular composition, then one can be reasonably certain that that assumption was incorrect. If this is the case, it might have been an impurity of higher molecular weight and one will find a much better fit with one of the ions of lower mass.

If none of the ions fit these criteria all ions **are fragments** of an undetectable molecular ion which, by definition, has to be of still higher mass. In this case, requirement (2) of the relationship between the hydrogen and nitrogen atoms must not necessarily hold because the fragment can have been formed either by loss of a radical or of a neutral molecule from the molecular ion. Thus, if the ion of highest mass, not being an isotopic specie, has an odd number of hydrogens and an even number of nitrogens, or none, it must be due to the loss of a radical and the molecule itself must have a composition that is larger by H, $CH_3$, OH, $C_2H_5$, $CH_3CO$, etc., (all groups that are known to be lost from compounds not showing a molecular ion). The mass of these are therefore added one at a time to the fragment ions of high mass, and after each addition, the resulting theoretical mass is again checked as above for the number of ions that would now correspond to the loss of an easily eliminated group.

If the ion of highest mass was however one that also fulfills requirement (2) it must have come about by the elimination of a neutral molecule and the computer therefore adds $NH_3$, $H_2O$, CO, $CH_2CO$, $CH_3COOH$, etc., before going again through the subtraction routine.

If the spectrum is at all of a quality that permits deducing the molecular composition of the compound, one of those three steps will lead to a result with a high number of positive conclusions for a particular ion which then is a highly probable choice for the molecular ion.

Table 1 is an example of the conclusions arrived at by the computer when processing the high resolution spectrum of androsterone acetate. It correctly represents the best fit out of five ions (Table 2) at the high mass end, all of which could be molecular ions based on their elemental composition alone.

For purpose of demonstration the two line positions corresponding to $M^{+\cdot}$ and $(M-1)^+$ were removed from the input data to simulate a complex molecule not exhibiting a molecular ion peak. The results (Table 3) clearly show that the computer still is able to deduce the molecular composition of the intact molecule. Twenty eight combinations were checked in this case.

The entries on the right of tables 1 and 3 indicate the kind of additional structural information concerning structural features which is obtained simultaneously. The predominance of ions involving the loss of acetic acid and methyl strongly suggests the presence of an acetoxy group as well as methyl attached to a highly substituted carbon atom; loss of CO in various steps is in agreement with a cyclic ketone.

While the approach and examples outlined above represent only a small part of what may be done with more elaborate programming even on compounds of greater complexity, it is clear that the objectives outlined at the outset of this paper can be achieved.

Acknowledgement. This work was supported by a grant from the National Institutes of Health (GM 09352) and from the National Aeronautics and Space Administration (NsG-211-62). The use of the facilities of the MIT Computer Center is gratefully acknowledged.

## TABLE 1

```
                                                MOLECULAR ION CHECK

ION UNDER CONSIDERATION IS-  C21  H32  N 0  O 3                       MOLECULAR WEIGHT  332.23496

NUMBER OF RINGS AND DOUBLE BONDS= 6

INT  FRAGMENT FRAGMENT LOST      C  H  N  O                          GROUP LOST
  1 331.22554   1.00942          0  1  0  0                      HYDROGEN
  5 317.20956  15.02540          1  3  0  0                      METHYL
  1 304.20377  28.03119          2  4  0  0                      ETHYLENE
  3 299.19945  33.03551          1  5  0  1                      H2O + CH3
 29 288.20701  44.02795          2  4  0  1                      CH2=CHOH
  9 276.20805  56.02692          3  4  0  1                      ETHYLENE + CO
  3 275.22254  57.01242          6  1  0 -1                             FRAGMENT IMPOSSIBLE FROM MOLECULAR ION
  2 275.20070  57.03427          3  5  0  1                      C2H5C=O
  8 273.18605  59.04892          3  7  0  1                      C2H5CHOH, (CH3)2COH, CH3 + CH2=CHOH
995 272.21140  60.02356          2  4  0  2                      ACETIC ACID
 17 271.20413  61.03083          2  5  0  2                      ACETIC ACID + HYDROGEN
  7 270.19782  62.03714          2  6  0  2
  1 262.19178  70.04318          4  6  0  1
  3 261.18409  71.05088          4  7  0  1
331 257.19028  75.04468          3  7  0  2                      ACETIC ACID + METHYL
  4 256.17993  76.05503          3  8  0  2
  4 255.17421  77.06075          3  9  0  2
 41 254.20159  78.03337          2  6  0  3                      ACETIC ACID + WATER
  4 253.19319  79.04177          2  7  0  3                      ACETIC ACID + WATER + HYDROGEN
  4 252.18680  80.04816          2  8  0  3
  1 250.17047  82.06449          2 10  0  3
 70 244.21810  88.01686          3  4  0  3                      ACETIC ACID + CO
 14 244.18035  88.05462          4  8  0  2                      ACETIC ACID+ETHYLENE, BUTYRIC ACID, ETHYL ACETATE
  3 243.20831  89.02665          3  5  0  3
 33 243.17235  89.06261          4  9  0  2

MOLECULAR ION SEARCH NO.  1     NO. OF FITS  15     NO. OF FRAGMENTS TESTED  25     NO. OF NONFITS  10

NO. WITH NO COMPOSITION AVAILABLE  1
```

## TABLE 2

```
SUMMARY

SEARCH NO.  1  FITS 15  NO. TESTED 25  NO. NONFITS 10  IMPOSSIBLE COMBINATIONS  1

SEARCH NO.  2  FITS  7  NO. TESTED 22  NO. NONFITS 15  IMPOSSIBLE COMBINATIONS  2

SEARCH NO.  3  FITS  7  NO. TESTED 20  NO. NONFITS 13  IMPOSSIBLE COMBINATIONS  1

SEARCH NO.  4  FITS  3  NO. TESTED 19  NO. NONFITS 16  IMPOSSIBLE COMBINATIONS  8

SEARCH NO.  5  FITS  7  NO. TESTED 15  NO. NONFITS  8  IMPOSSIBLE COMBINATIONS  9


MOLECULAR WEIGHT  332.23438  COMPOSITION  C21  H32  N 0  O 3
```

## TABLE 3

```
ION UNDER CONSIDERATION-  C21  H32  N 0  O 3                         MOLECULAR WEIGHT,CALC 332.23490DETM 332.23496
DERIVED FROM SUM OF  317.21150  AND   15.02340

NUMBER OF RINGS AND DOUBLE BONDS= 6

INT  FRAGMENT  FRAGMENT LOST      C  H  N  O                 GROUP LOST
  5 317.20956  15.02534           1  3  0  0                     METHYL
  1 304.20377  28.03113           2  4  0  0                     ETHYLENE
  3 299.19945  33.03545           1  5  0  1                     H2O + CH3
 29 288.20701  44.02789           2  4  0  1                     CH2=CHOH
  9 276.20805  56.02686           3  4  0  1                     ETHYLENE + CO
  3 275.22254  57.01236           6  1  0 -1                            FRAGMENT IMPOSSIBLE FROM MOLECULAR ION
  2 275.20070  57.03421           3  5  0  1                     C2H5C=O
  8 273.18605  59.04885           3  7  0  1                     C2H5CHOH, (CH3)2COH, CH3 + CH2=CHOH
995 272.21140  60.02350           2  4  0  2                     ACETIC ACID
 17 271.20413  61.03077           2  5  0  2                     ACETIC ACID + HYDROGEN
  7 270.19782  62.03708           2  6  0  2
  1 262.19178  70.04312           4  6  0  1
  3 261.18409  71.05082           4  7  0  1
331 257.19028  75.04462           3  7  0  2                     ACETIC ACID + METHYL
  4 256.17993  76.05497           3  8  0  2
  4 255.17421  77.06069           3  9  0  2
 41 254.20159  78.03331           2  6  0  3                     ACETIC ACID + WATER
  4 253.19319  79.04171           2  7  0  3                     ACETIC ACID + WATER + HYDROGEN
  4 252.18680  80.04810           2  8  0  3
  1 250.17047  82.06443           2 10  0  3
 70 244.21810  88.01680           3  4  0  3                     ACETIC ACID + CO
 14 244.18035  88.05456           4  8  0  2                     ACETIC ACID+ETHYLENE, BUTYRIC ACID, ETHYL ACETATE
  3 243.20831  89.02659           3  5  0  3
 33 243.17235  89.06255           4  9  0  2
 50 242.20173  90.03317           3  6  0  3

MOLECULAR ION SEARCH NO. 1  ADDITION SEQUENCE 2

NO. OF FITS  14  NO. OF FRAGMENTS TESTED  25  NO. OF NONFITS  11  NO. WITH NO COMPOSITION AVAILABLE  1
```