195
1-24-78

1788

BERC-0020-9

**MASS SPECTROMETRIC ANALYTICAL SERVICES AND RESEARCH
ACTIVITIES TO SUPPORT COAL-LIQUID CHARACTERIZATION
RESEARCH**

Quarterly Report, June 9–September 30, 1977

By
Stuart E. Scheppele

MASTER

October 25, 1977

Work Performed Under Contract No. EY-76-S-02-0020

Department of Chemistry
The Oklahoma State University
Stillwater, Oklahoma

U. S. DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

MASS SPECTROMETRIC ANALYTICAL SERVICES AND RESEARCH ACTIVITIES
TO SUPPORT COAL-LIQUID CHARACTERIZATION RESEARCH

Quarterly Report for the Period
June 9, 1977 through September 30, 1977

Stuart E. Scheppele

The Oklahoma State University
Department of Chemistry
Stillwater, Oklahoma 74074

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Date Submitted — October 25, 1977

PREPARED FOR
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

Under Contract No. EY-76-S-02-0020-A001

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

ABSTRACT

The purpose of this programming project is automatically to assign reasonable molecular formulas to experimental masses obtained from high-resolution mass spectra.

The program which was developed reads experimental carbon-12 masses, converts them to corresponding Kendrick masses, and assigns a set of reasonable molecular formulas to each Kendrick mass within a given tolerance. The computational methods to do this are based on the tables previously developed by Guffey and Weber. The methods used to search these tables and generate the corresponding molecular formulas are discussed.

This program is organized now so that it is resident on a disk pack on the IBM 370/158. Data entry may be from cards or from ket-to-disk data entry (TSO) terminal.

We are now using this program to process data acquired from the CEC 21-110B mass spectrometer. We are planning to continue the development of this program to increase its analytical capabilities and to increase its ease of use.

Acquisition of high-resolution mass spectral data for 12 fractions obtained from separation of a Canadian Tar Oil Sand sample has been continued. Decisions have been made concerning the welding components needed to spot weld 8 μ m tungsten wire onto the tungsten posts in preparation of emitters for field-desorption mass spectrometry.

OBJECTIVE AND SCOPE OF WORK

The objective of this program is to obtain qualitative and quantitative information concerning the composition of alternate fossil energy derived materials using the technique of mass spectrometry. Special emphasis is placed upon the use of low- and high-resolution field-ionization mass spectrometry and high-resolution electron-impact-induced-mass spectrometry to obtain information relating to the compound types and/or compounds present in a sample. The technique of field-desorption mass spectrometry will be applied to the analysis of relatively nonvolatile high-molecular weight components present in these mixtures. Known mixtures of compounds structurally related to those either believed or known to be present in these samples is being investigated with the view of 1) determining sensitivities towards field ionization and field desorption and 2) ascertaining for sample quantification the accuracy obtained by various techniques. Gas chromatography in combination with electron-impact and field-ionization mass spectrometry is also used to acquire characterization data. Computer software is being developed to permit routine use of computer technology in the reduction, correlation, and interpretation of experimental mass spectral data.

In this research program, attention will, as required, be devoted to the modification of existing and/or the development of new techniques for the mass spectral analysis of materials derived from fossil fuels.

SUMMARY OF PROGRESS TO DATE

Field-ionization and electron-impact mass spectrometry is being used to analyze samples both qualitatively and quantitatively. Modification of combination FI/EI ion source to operate in field desorption mode is complete. Determination of field ionization sensitivity data and development of computer programs for interpretation of mass spectral data are on schedule, if not somewhat ahead.

During the month of June, the percentage of time devoted by S. E. Scheppele was 10%; for July, August and September, 15% of Dr. Scheppele's time was devoted to this project. The percentage of time devoted by G. E. Hedrick was 10%.

DETAILED DESCRIPTION OF TECHNICAL PROGRESS

Introduction

The purpose of this programming project is automatically to assign reasonable molecular formulas to experimental masses obtained from high resolution mass spectra.

The project began with a computer program which was developed originally by George Stewart and T. D. Marriott at Oklahoma State University (1). It was modified and adapted for this project by Yaduraj Kapoor in the Department of Computing and Information Sciences at Oklahoma State University. The analysis required for the modifications, implemented by Mr. Kapoor, were performed by Drs. Marriott and Hedrick. This analysis and programming are the subjects of this report.

The program was written to read an experimental mass (on the carbon-12 scale), convert that mass to a corresponding Kendrick mass (on the scale where $\text{CH}_2 = 14$ amu) and produce chemical formulas corresponding to the Kendrick mass.

At the present time this program can be run in either of two modes. Both modes are applicable to the IBM 370 computer only. This program may be run in batch mode with either punched card or disk input, or it may be run in a remote mode via any of the time sharing option (TSO) terminals which are attached to the IBM 370 computer at Oklahoma State University. In the latter case all of the input to the program must reside on disk prior to the program's being executed. There are still several changes which need to be made in order to increase the flexibility and the usefulness of this program. Also, the program needs to be restructured slightly to increase the ease of program maintenance and to increase the ease of use. The program also must be modified in such a way that it is possible to run it by either introducing data to the IBM 370 or via the Nova 3/12. One additional item that needs to be done with this program is to produce its documentation including a user's manual. Part of this manual is included as a part of this report.

In order to discuss the manner in which the assignments of molecular formulas are performed, it is necessary to define some of the terminology used. These definitions are either chemical, computational, mathematical, or some combination thereof, in form.

Definition 1. The floor function of a real value, x , denoted by $\lfloor x \rfloor$ is the greatest integer value contained in x . Examples: $\lfloor 5.01 \rfloor = 5$, $\lfloor 15.99 \rfloor = 15$, $\lfloor -5.01 \rfloor = -6$, $\lfloor -15.99 \rfloor = -16$, $\lfloor 15 \rfloor = 15$, $\lfloor -5 \rfloor = -5$.

Definition 2. The ceiling function of a real value, x , denoted by $\lceil x \rceil$, is the smallest integer value not contained in x if x is not an integer and is x otherwise. Examples: $\lceil 5.01 \rceil = 6$, $\lceil 15.99 \rceil = 16$, $\lceil -5.01 \rceil = -5$, $\lceil -15.99 \rceil = -15$, $\lceil 15 \rceil = 15$, $\lceil -5 \rceil = -5$.

Definition 3. The quantity $i \bmod n$ for integer values i and n is defined to be the remainder after finding the integer quotient of dividing i by n . If $i \div n = q + r$ for quotient, q , and remainder, r , then $i \bmod n = r$.

Definition 4. A transformation, t , from set Q to set R , denoted by $t: Q \rightarrow R$ is a mapping which maps any element of Q into some element of R ; that is, for every $q \in Q$, there exists an $r \in R$ such that $t(q) = r$. A complete specification of t includes the rule or rules by which elements of Q are mapped into R .

Definition 5. A Kendrick mass, M_K , is a mass on the scale where $\text{CH}_2 = 14$ amu (2).

The following discussion explains the symbolism used in definitions 6 and 7 below. Guffey and Weber (3) defined a set of tables which were used for the assignation of molecular formulas performed in this programming project. These tables which relate the Kendrick mass to molecular formulas, are tabulated as a function of the nominal hydrocarbon Z-value (Z_N). The nominal hydrocarbon Z value is defined by Weber and Guffey (3). We repeat their definition here for convenience.

Compounds containing the most abundant light isotopes of carbon, hydrogen, nitrogen, and oxygen can be represented by the formula $C_nH_{2n+Z}N_aO_bS_c$. For any given value of Z in this formula it is possible to determine Z_N .

Let x be the number required for a solution of the equation

$$y \cdot \text{Mass}(\text{CH}_2) + x \cdot \text{Mass}(\text{H}_2) = 14a + 16b + 32c$$

where $\text{Mass}(\text{CH}_2) = 14$, $\text{Mass}(\text{H}_2) = 2$, a, b, c are as given in the formula above, and x, y are subject to the constraints

$$2 \cdot x < \text{Mass}(\text{CH}_2)$$

$$y = \lfloor (14a + 16b + 32c) / \text{Mass}(\text{CH}_2) \rfloor$$

Definition 6. The nominal hydrocarbon Z-value, Z_N , of a compound is the number Z_N in the formula $Z_N = Z + 2x$ where x and y are computed as above.

The occurrence of molecules containing one or more heavy isotopes needs to be considered in computation and formula assignation. For the compounds of present interest and the molecular weight range covered, it is pragmatically necessary to consider only the heavy isotopes ^{13}C and ^{34}S . Thus, the general molecular formula consistent with the tables of Guffey and Weber (3) becomes $^{12}\text{C}_{n-e}^{13}\text{C}_e\text{H}_{2n+Z}^{14}\text{N}_a^{16}\text{O}_b^{32}\text{S}_c^{34}\text{S}_d$ where e = 0 or 1.

Definition 7. Let M_K be the Kendrick mass of a compound of the form $C_nH_{2n+Z}^{14}\text{N}_a^{16}\text{O}_b^{32}\text{S}_c^{34}\text{S}_d$. Further let $p = M_K - \lfloor M_K \rfloor$. A Z-table for a given Z_N is a tabulation of all of the combinations of the elements of the set, {a,b,c,d,z,p} which yield the specified Z_N value with $a+b+c+d \leq 3$, according to the method outlined previously. Table I is an example of such a table where $Z_N = 0$.

Definition 8. The nominal mass, M_n , is defined to be $M_n = \lfloor M_K + 0.5 \rfloor$ or equivalently $M_n = \lfloor M_K - 0.5 \rfloor$ except that M_n is undefined when $.4994417 \leq M_K - \lfloor M_K \rfloor \leq .500559$.

The nominal mass is undefined in the region where $.4994417 \leq M_K \leq .500559$. The rational for this is as follows.

When $M_E - \lfloor M_E \rfloor = .5$ then the nominal mass computed from M_E is not the same as the nominal mass computed from M_K .

Similarly when $M_K - \lfloor M_K \rfloor = .5$ the nominal mass computed from M_K is not the same as the nominal mass computed from M_E . This is true in the entire interval $[\lfloor .9988834(.5) \rfloor, \lfloor 1.00111784(.5) \rfloor] = [.4994417, .500559]$ and may be verified by hand computations. This region where the nominal mass is undefined is of no concern to our present experiments since the experimental masses do not appear in the regions. This may be of concern in future experiments where higher masses are encountered.

TABLE I

TABLE FOR DETERMINING THE ELEMENTAL COMPOSITION
OF IONS FOR THE $Z_N = 0$ SERIES

abcd	Z	p	Z	p	Z	p	Z	p
0000	00	.0000	-14	.9062	-28	.8124	-42	.7186
1000	00	.9874	-14	.8936	-28	.7998	-42	.7060
2000	00	.9749	-14	.8811	-28	.7873	-42	.6935
100	-2	.9637	-16	.8699	-30	.7761	-44	.6823
3000	00	.9623	-14	.8685	-28	.7747	-42	.6809
1100	-2	.9511	-16	.8573	-30	.7635	-44	.6697
2100	-2	.9385	-16	.8447	-30	.7509	-44	.6571
200	-4	.9273	-18	.8335	-32	.7397	-46	.6459
1200	-4	.9147	-18	.8209	-32	.7271	-46	.6334
10	-4	.9096	-18	.8158	-32	.7220	-46	.6282
1010	-4	.8970	-18	.8032	-32	.7094	-46	.6156
300	-6	.8910	-20	.7972	-34	.7034	-48	.6096
1	-6	.8897	-20	.7959	-34	.7021	-48	.6083
2010	-4	.8844	-18	.7906	-32	.6968	-46	.6031
1001	-6	.8772	-20	.7834	-34	.6896	-48	.5958
110	-6	.8732	-20	.7794	-34	.6856	-48	.5919
2001	-6	.8646	-20	.7708	-34	.6770	-48	.5832
1110	-6	.8607	-20	.7669	-34	.6731	-48	.5793
111	2	.8568	-12	.7630	-26	.6692	-40	.5754
101	-8	.8534	-22	.7596	-36	.6658	-50	.5720
1101	-8	.8408	-22	.7470	-36	.6532	-50	.5594
210	-8	.8369	-22	.7431	-36	.6493	-50	.5555
30	2	.8225	-12	.7287	-26	.6349	-40	.5411
20	-8	.8191	-22	.7253	-36	.6316	-50	.5378
201	-10	.8170	-24	.7232	-38	.6295	-52	.5357
1020	-8	.8066	-22	.7128	-36	.6190	-50	.5252
21	00	.8027	-14	.7089	-28	.6151	-42	.5213
11	-10	.7993	-24	.7055	-38	.6117	-52	.5179
1011	-10	.7867	-24	.6929	-38	.5992	-52	.5054
120	-10	.7828	-24	.6890	-38	.5952	-52	.5014

Description of the Program and the Methods Used

The purpose of the program is to assign molecular formulas to a given set of high resolution mass spectral data using the method developed by Guffey and Weber (3) which uses the tables corresponding to the 14 hydrocarbon nominal Z series.

The program converts experimental carbon-13 based exact masses obtained from the high-resolution mass spectra to corresponding Kendrick masses. The Kendrick exact mass is rounded to obtain an integer mass. A Z_N value is then computed from the integer mass. After the appropriate Z_N number is determined, that number (Z_N) is used as an index to start a table search to find the molecular formula that best fits the given mass. Due to inherent inaccuracy in the experimental data, and due to finite precision arithmetic, the fit is not exact but rather it is within some tolerance which can be specified by the user of the program. At the present time, the only heteroatoms permitted in the molecular formulas are nitrogen, oxygen, sulfur-32, sulfur-34, and/or 1 carbon-13.

Let the measured experimental carbon-12 based experimental mass be represented by M_E and let the corresponding Kendrick mass be represented by M_K . The Kendrick mass M_K can be found from the experimental mass by

$$M_K = .0099934 M_E$$

Methods Used

The program determines a y number from the Kendrick mass by using the formula of equation 1 which was developed by Dr. Marriott (5).

$$y = 14 \left[\frac{M_K}{14} - ([M_K/14] + 1) \right] \quad [1]$$

Equation 1 produces integer y values which range from 0 through -13. These y values are transferred into nominal hydrocarbon Z-values (Z_N) via the transformation h below. The Kendrick masses are then grouped according to hydrocarbon Z number (Z_N). For each array of Kendrick masses, the program searches the table containing data corresponding to the appropriate nominal Z value, Z_N , in order to determine the calculated mass defects which fit, within a specified tolerance, the measured mass defect (based upon $CH_2 = 14$ amu) for each member in the array.

The tables used in this searching routine are a modified version of the tables compiled at LERC (3). Each of the tables used in our program contains the mass defects for compounds containing zero and one carbon-13 atom. Associated with each mass defect in the various tables is a general formula code containing the number of nitrogen (a), oxygen (b), sulfur-32 (c), sulfur-34 (d), and carbon-13 (e) atoms in the molecular formula and a value Z which is used to obtain the number of hydrocarbons. Consequently, the general formula has the form $C_{n-e}^{13}C_eH_{2n+Z}NaOb^{32}Sc^{34}Sd$. The value of n is calculated from equation 2 in which mass is the nominal mass of the ion; a, b, c, d, and

$$n = \frac{M_n - a(14) - b(16) - c(32) - d(34) - e(13) - Z}{14} \quad [2]$$

e are the values of the subscripts defined by the general formula; and Z equals the Z-number from the general formula.

Let W be the set of all compounds with the general formula $C_nH_{2n+Z}NaOb^{32}S_c^{34}S_d^{13}Ce$. In order to use the tables supplied by Guffey and Weber (3), we partition W according to nominal Z values. Based on Dr. Marriott's technique (equation 1), each nominal Z value, Z_N , is computed by first obtaining a ζ value, $\zeta = M_n \bmod 14$, and mapping that ζ value to the corresponding Z_N value by the transformation $h:\{0,\dots,13\} \rightarrow \{-11,-10,\dots,0,1,2\}$, where $h(i) = \begin{cases} i & \text{if } i \leq 2 \\ i-14 & \text{if } i > 2 \end{cases}$. Each $Z_N = h(\zeta)$ identifies the proper partition of W which in turn identifies the proper table from Guffey and Weber (3). This table is searched using the decimal fractional Kendrick mass, M_K , corresponding to the input mass, M_E , in order to determine appropriate molecular formulas. Any formula which matches within the input tolerance, δ , is assumed to be a candidate for assignment. (Table I shows an example of one such Z table.) The table is searched using a binary search and any formula in which $|p - (M_K - |M_K|)| \leq \delta$ is considered a match.

Program Description

The program reads (from cards or from disk) all the available Z series tables, one table at a time; sorts the masses on each table in decreasing order; and stores the sorted tables in an internal array, ZTABLE. the sorted tables are then used in searching for the best fit. In searching for a fit, a maximum allowable mass defect error (in mmu) is allowed, and this is read as IDELTA. Each successful fit has an error less than or equal to IDELTA.

Experimental ^{12}C masses obtained from high-resolution mass spectral data are read into the program and converted to Kendrick masses. Nominal masses are computed as the rounded integer for the Kendrick masses. The proper Z table to be searched and its Z series number is then determined for each nominal mass. A binary search is performed on the selected Z table using the mass defect as the key and all the fits within the tolerance are found. A code (equation 2) is associated with each fit which is used to determine the heteroatoms and the number of hydrogen atoms that will occur in the final formula. The molecular formula is then computed and assigned using the best fit information. If no fit can be found, a message to this effect is printed.

Description of Program Input Requirements

Input data for this program may either reside on a computer disk or it may be on punched cards. Even when the data resides on disk, it is in card image format. For purposes of discussion, the term, card, will refer either to a punched card or to an image of a punched card on disk.

There are three separate sections of input. The first section consists of heading and tolerance information, the second section consists of the Z tables from Guffey and Weber (3) and the third consists of the carbon-12 based experimental input masses. Input sections one and three must reside on the same data set (both must be on disk or both must be on cards), but the middle

section may reside on a possibly different data set. The data input in each section is shown in Table II.

TABLE II
INPUT SECTIONS

Input Section	Input Quantities
1	heading defect tolerance, δ
2	Z-tables
3	experimental carbon-12 masses

Data in input section 1 normally starts with a blank card, followed by a card which contains the maximum allowable mass defect, δ , expressed in mmu, in an I2 format.

The Z-tables are input in the second input section. A card which contains the Z_N value (in a 4X,I7 format) for a given table starts each new input table. The last Z-table to be input contains the letters "FINI" in card columns one through four.

The data entered into each Z-table have the format shown in Table III. A -1 in card columns 1 and 2 terminates a given table.

Normally, the program user need not be concerned with the table entry as it will be performed by the program automatically, except when the user wants to have section 2 input from cards.

The experimental mass data is input in section 3. All of the input required for this section is input one value per card in a F8.4 format.

The effective mass data is terminated by -01.0000 punched in columns 1 through 8. This terminates all of the input to the program.

TABLE III
FORMATS FOR INPUT SECTION 2

Columns	Format	Data
1,2,3,4	I4	p
5	IX	blank
6,7,8	{I8}	Z
9,10,11,12,13		abcde

Description of the Output

The output of the Z series program consists of the assigned molecular formulas and some appropriate messages. The first message gives us the maximum allowable error in a fit used in the program. If some of the Z series tables are not present during program execution, a message is printed for each missing table. In assigning a molecular formula the message "NO MATCH FOUND" is printed if a successful reasonable fit within the allowable error range has not been found.

Operating the Program in Batch Mode

When operating this program in batch mode, all of the experimental masses are presented on the program via punched cards. Normal operation (on the IBM 370) requires the following structure for the input deck:

1. The control cards to initiate execution of this program.
2. A //GO.SYSIN DD * control card.
3. A blank card.
4. The maximum error to be permitted in formula matching.
5. The experimental masses in the format described in Section III above. There is one input card for each experimental mass.
6. A -01.0000 card also described in Section III above.
7. A // (null) control card.

The user has the option of reading additional heteroatoms prior to the input of the maximum allowable error. Because not all of the Z tables are presently on disk, the user has the option of providing as many additional Z tables as he desires. At the present time tables for Z = -10, Z = -11, and carbon-13 are not available on disk. These will be placed on disk for automatic use after they have been checked. The user should contact one of the program authors when he wants to use these options.

Operating the Program in Remote Mode

When operating this program in remote mode, the experimental masses are presented to the program via a disk data set. In order to operate in remote mode the user must adhere to the following steps:

1. Make certain that the required data is on a disk data set. The disk data set will be created by transmitting the data from the Data General Nova 3 to the IBM 370 after the data communications software is in place (5). At this time a separate job must be run to place the data on disk.
2. Follow the time sharing option (TSO) logon procedure as specified in the OSU computer Center User's Guide (6).
3. Type EXEC ZSERIES on the TSO terminal.
4. After the program requests "specify INDSNAME" (input dataset name), type the name of the disk data set mentioned in item 1.

After following these steps, there will be a short delay and the output will appear at the TSO terminal.

Mass spectrometric analysis (7) of the fractions obtained from separation of a Canadian Tar Oil Sand sample by the Separations and Characterization Group at the Bartlesville Energy Research Center has continued. High-resolution data were recovered off of photographic plates for 12 fractions and the line position data were converted to exact ion masses. Activities in regard to this mass spectrometric analysis reflect the following factors. First, the research associate who will be a major participant in these activities was unavoidably delayed in his arrival. Second, acquisition and processing of photoplate data was accomplished by Mr. K. N. Detwiler who, as an undergraduate research participant can devote only a limited number of hours to the project. Third, activities of Dr. S. E. Scheppelle were curtailed by his extensive involvement in implementation of the CEC 21-110B mass spectrometer/AEI DS-50S data acquisition system combination. In this regard, it is important to note that, although the DS-50S was purchased with funds supplied jointly by ERDA under contract EX-76-S-01-2537 and Oklahoma State University, the CEC 21-110B/DS-50S combination will be routinely used on this ERDA project. The benefits which

will accrue to this project in terms of the reduction in the time required to acquire and reduce mass spectral data dictated this concentration of effort.

Acquisition of a spot welder is a prerequisite for preparation of field-desorption emitters. Consequently, the welding components and the associated welding jig appropriate for spot welding the 8 μ m tungsten wire onto the tungsten posts was discussed with Dr. D. F. Barofsky. The availability of the requisite components from vendors other than the one from which Dr. Barofsky acquired his was assessed. Consequently, a requisition is being issued for acquisition of the appropriate spot welding components.

References

- 1) George Stewart and T. D. Marriott, unpublished program notes, 1977.
- 2) E. Kendrick, *Anal. Chem.*, 35, 2146 (1963).
- 3) F. D. Guffey and J. H. Weber, "A Tabular Method for Obtaining Empirical Formulas from High-Resolution Mass Spectral Data," (Report of Investigations, Laramie Energy Research Center, Laramie, Wyoming).
- 4) Quarterly Progress Report FE 2537-2 submitted under ERDA contract EX-76-S-01-2537.
- 5) Quarterly Progress Report FE 2537-3 submitted under ERDA contract EX-76-S-01-2537.
- 6) Oklahoma State University Computer Center User's Guide, September, 1977.
- 7) Quarterly Progress Report BERC-0020-8 submitted under ERDA contract EX-76-S-02-0020 MOD #1.