

CONF-9605771

UCRL-JC-123846

**A DATA ANALYSIS ROUTINE TO PROTECT CONFIDENTIAL
INFORMATION DURING GC-MS ANALYSIS**

RAYMOND R. MCGUIRE and WALTER H. MARTIN

*Arms Control and Treaty Verification Program
Lawrence Livermore National Laboratory, L-175*

P.O. Box 808

Livermore CA. USA 94551-9900

INTRODUCTION:

The Chemical Weapons Convention allows for the collection and analysis of samples at inspection sites, including demilitarization facilities, to verify compliance with treaty obligations. The Preparatory Commission to the OPCW, through its Expert Group on Inspection Procedures, PC-VIII/B/WP.2, has determined that the Technical Secretariat must have analytical equipment that, among other things, allows "identification and recording of the detection of chemicals relevant to the purpose of the inspection, but the suppression of the indication and recording of the detection of chemicals and other information unrelated to the purpose of the inspection."

This paper describes unfinished work at the Lawrence Livermore National Laboratory (LLNL) to satisfy this requirement for GC-MS instrumentation.

APPROACH:

Over a period of years, scientists at LLNL have been developing a mass spectral data analysis code that could use the output of the various types of instruments at the laboratory. The result is the "M CODE," designed for IBM PC compatible computers. The M Code will use the output from FINNEGAN, VARIAN (SATURN), and HP, (both RTE and Chemstation formats).

The M CODE has then been overlaid with a routine called "MASK" that limits what information is actually presented to the analyst on the computer screen.

The following is a demonstration of the MASK code using data from samples collected during joint experiments of LLNL, the Edgewood Research, Development, and Engineering Center (ERDEC) and the Chemical and Biological Defence Establishment (CBDE) at Porton Down, UK. This discussion will show the current state of the program development, including what is yet to be done to make the software usable for verification purposes.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

OR

DISCUSSION:

Since the purpose of masking data analysis software is to withhold from the inspector any information that is not relevant to the inspection, a limited reference library containing only spectra of CW related compounds was used. The library contained about 70 different spectra and was obtained from CBDE. The first set of data is taken from an experiment where VX was exposed to the environment on a soil sample at CBDE.

The total ion chromatogram (TIC) of the extracted sample is shown in Figure 1.

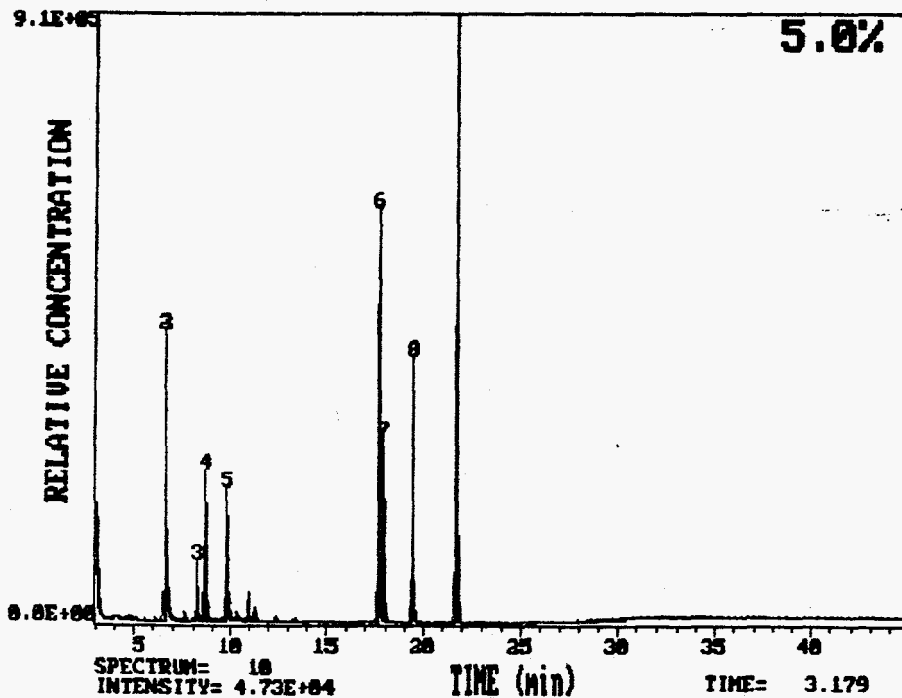


Figure 1. Total Ion Chromatogram of Extract of VX on Soil After Environmental Exposure

However, this screen is not shown to the inspector analyst. Figure 2. shows the first screen seen by the analyst.

The M Code has numbered each of the peaks that are a certain percentage above the baseline. This percentage is an adjustable parameter. The TIC is divided into a number of time increments. The increment chosen for this discussion is 5 minutes. The code compares the mass spectrum of each of the numbered peaks within the time increment with the reference library and displays a shaded area for the increment if there is a fit between the sample and reference spectra to the degree that is prespecified. The

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, make 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.**

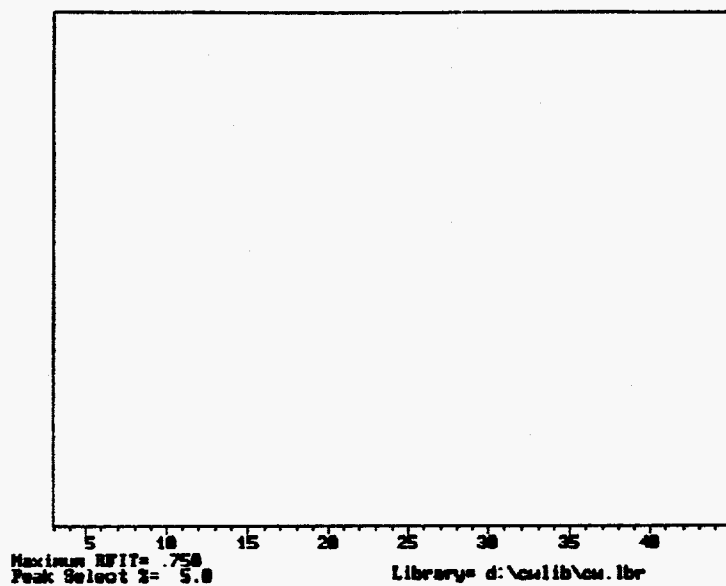


Figure 2. Computer Screen as Seen by Analyst at the Beginning of Data Analysis

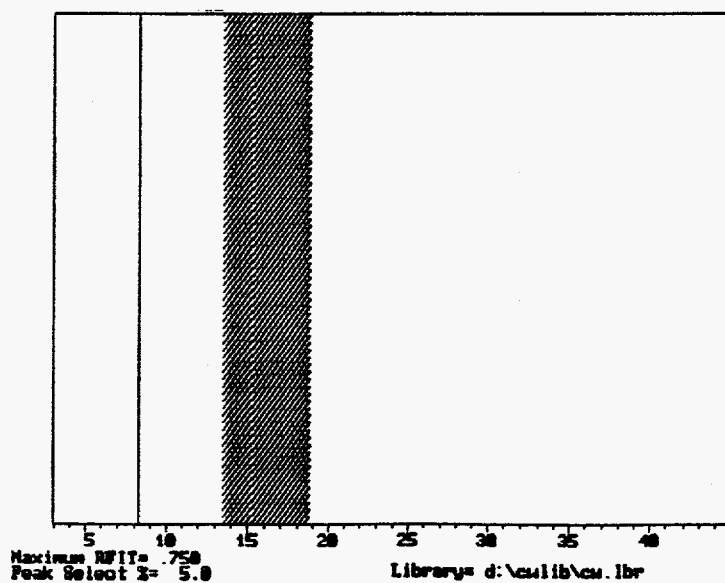


Figure 3. Screen Display Showing the Shaded Area Denoting a "Hit"

specification of the required fit coefficient is also adjustable. A 0.750 matching coefficient was chosen. The increment area is not shaded if there are no peaks within it that match the reference library. Figure 3 illustrates this.

At this point the TIC of the relevant time increment can be displayed along with the mass spectrum of the peak which gave an adequate match to the reference library. This is shown in Figure 4.

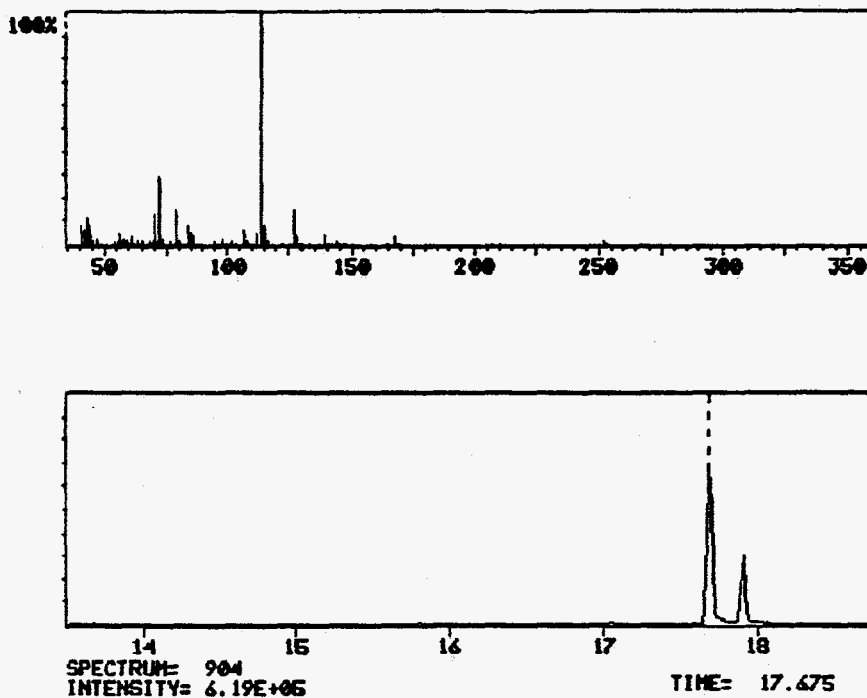
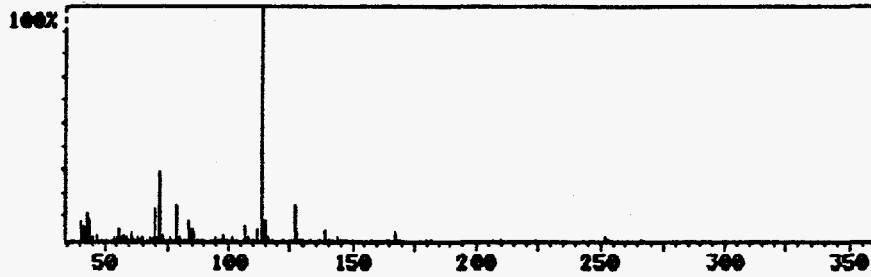


Figure 4. Screen Display of Selected (Matching) Peak and its Mass Spectrum

The matching reference spectrum and identification is then displayed as in Figure 5 where the first peak is identified as VX.

In this particular time increment both of the peaks matched the library. The second peak is then identified (Figure 6) as dicyclohexylcarbodiimide, an additive to VX.



VX (Ethyl 8-2-diisopropylamino ethyl methyl phosphonothioate)
 Not Given
 CAS ID.... - -0 ;\mask\
 MOL. WT= 190 RANK= 4 INDEX= 104 RFIT= .642 MATCHES= 35/102
 MASS RANGE= 35 - 361 MH SEARCH RANGE= 50 - 400 BASEPEAK HEIGHT ON

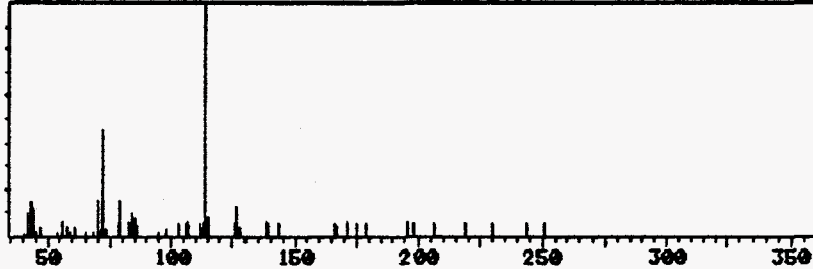
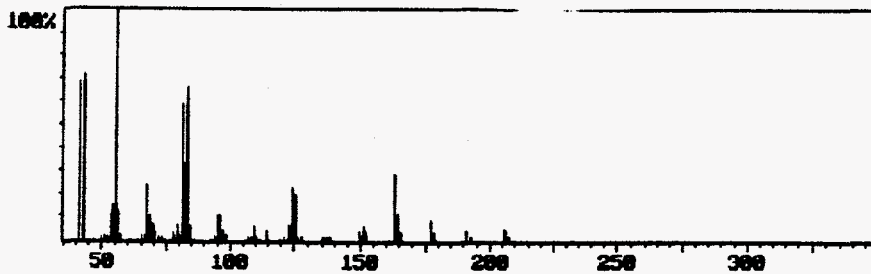


Figure 5. The First Selected Peak is Identified as VX



Dicyclohexylcarbodi-imide
 Not Given
 CAS ID.... - -0 ;\cm\cm.lib
 MOL. WT= 190 RANK= 1 INDEX= 24 RFIT= .785 MATCHES= 18/ 81
 MASS RANGE= 35 - 350 MH SEARCH RANGE= 50 - 400 BASEPEAK HEIGHT ON

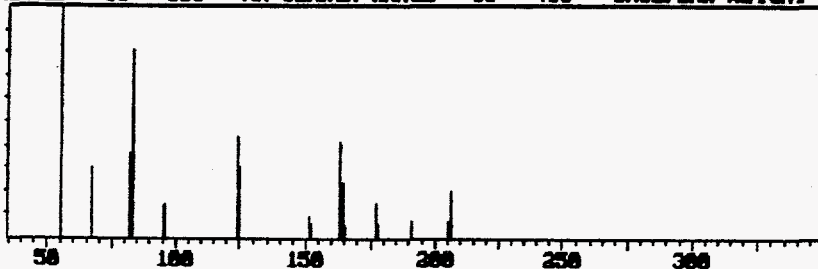


Figure 6. The Second Peak is Identified as Dicyclohexylcarbodiimide

The program will then move on to the next increment. In this particular analysis, there is another "hit" in the following increment as seen in Figure 7.

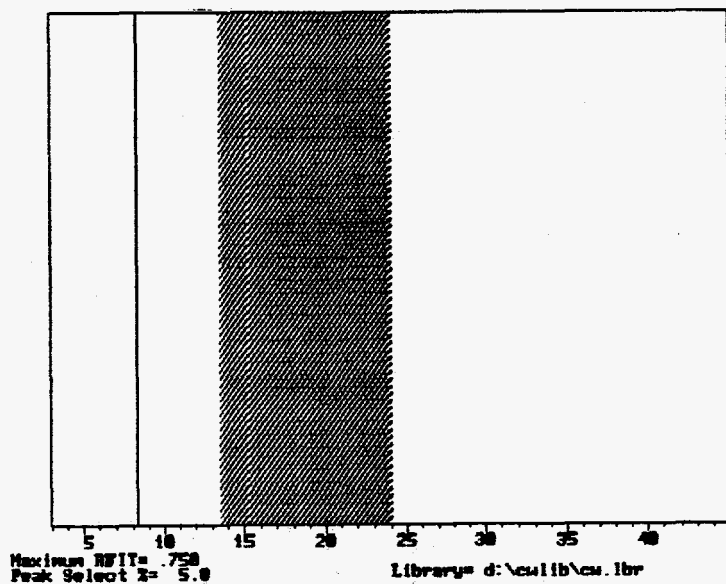


Figure 7. Screen Display Showing "Hits" in Consecutive Time Increments

Figures 8 and 9 show that the first peak in the interval is identified as bis-(2-diisopropyl-aminoethyl) disulphide. A degradation product of VX. The second peak does not give a match to the library and is not identified.

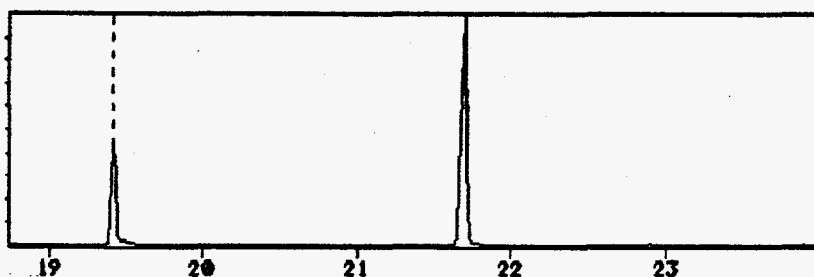
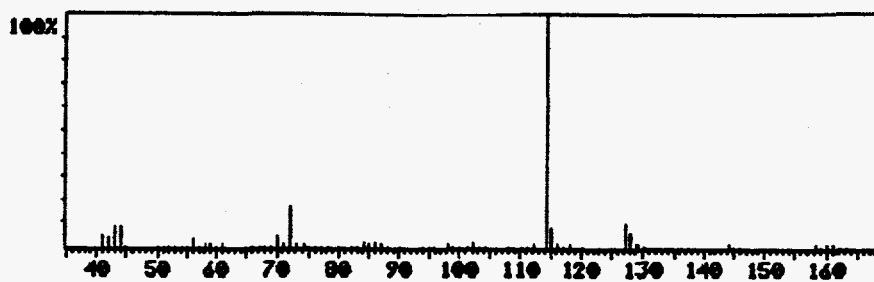
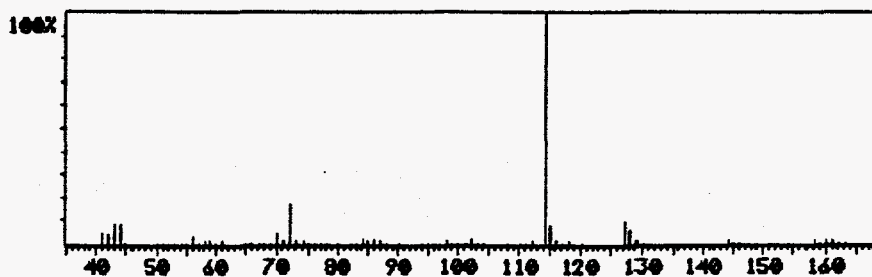


Figure 8. TIC of the Second Increment with the Mass Spectrum of the First Peak



Bis (2-diisopropylaminoethyl) disulphide
 Not Given
 CAS ID... - -0 : \mask\
 MOL. WT= 100 RANK= 1 INDEX= 33 RFIT= .693 MATCHES= 44/ 74
 MASS RANGE= 35 - 168 MH SEARCH RANGE= 50 - 400 BASEPEAK HEIGHT ON

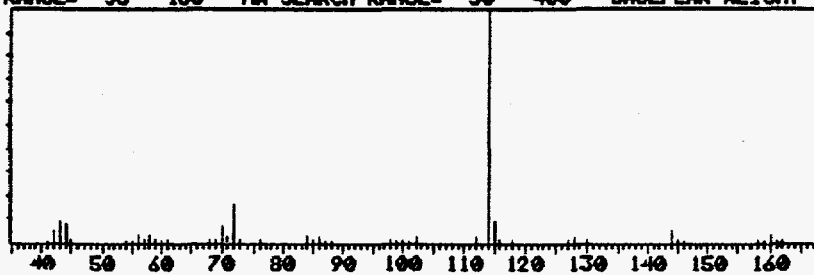


Figure 9. Identification of Selected Peak as Bis-(diisopropylaminoethyl) disulphide

Since there are no further "hits" in this analysis, no further spectral information is displayed. (See Figure 10.) Keep in mind that the complete TIC (Figure 1) is never seen by the inspector analyst.

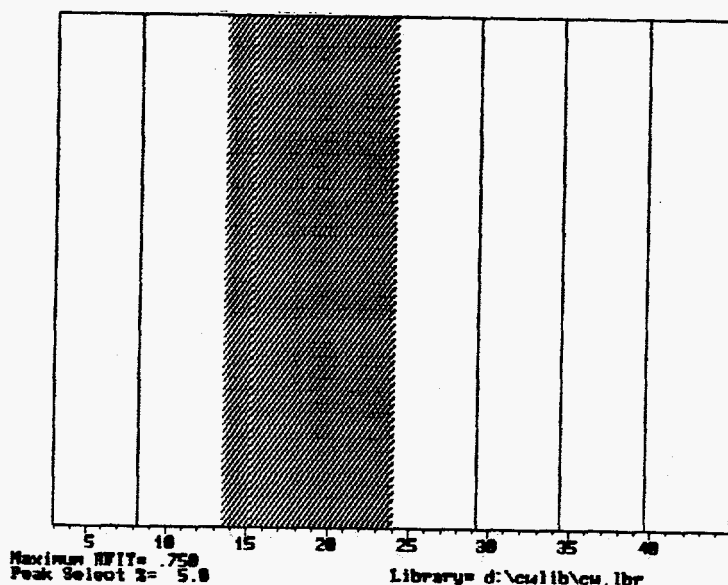


Figure 10. Screen Display Showing that no Peaks in Subsequent Time Increments Matched the Reference Library.

A rather serious analytical problem, leading to possible false negatives, arises from the fact that the analyst cannot view and operate on the entire TIC. This problem is illustrated in the next series of figures.

Figure 11 shows the display screen from the analysis of a liquid sample collected during a second cooperative (US-UK) experiment at CBDE. No library matches were made. Yet the sample was known to contain DMMP, DEMP and DIMP. (It was spiked with them and they were identified during a "normal analysis where the MASK program was not used.) The required fit coefficient was reduced from 0.750 to 0.150. (This would, of course, not be done during a real inspection.) The analysis software now finds two time increments where matches to the reference library occur. (See Figure 12.)

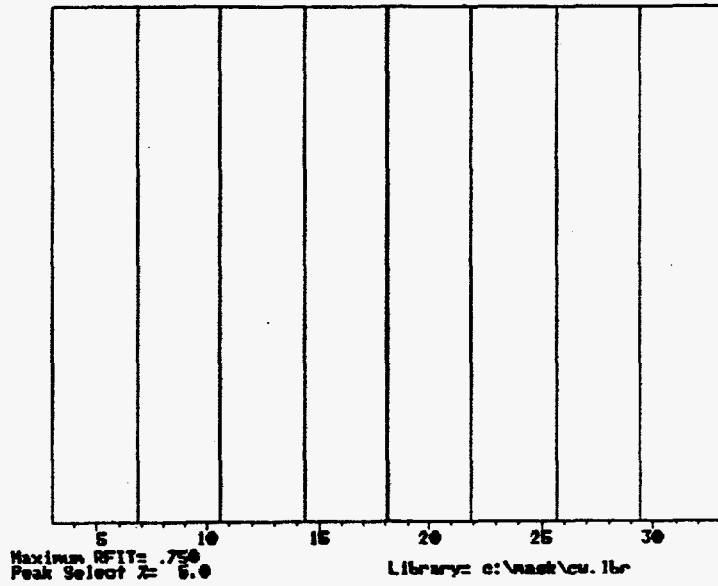


Figure 11. Screen Display of Sample Containing DMMP, DEMP and DIMP Showing no "Hits."

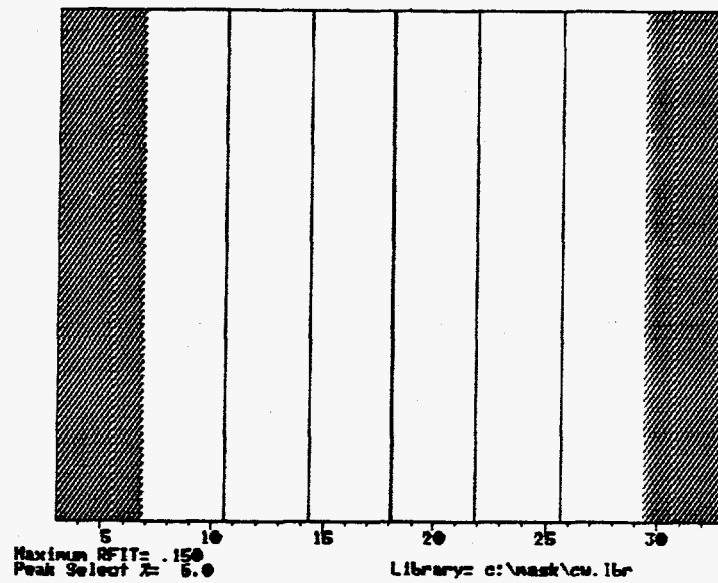


Figure 12. "Hits" are Obtained in Two Increments When the Fit Coefficient is Reduced.

The display now shows that there are three small peaks, very near the baseline. Figure 13 and 14 show that the first peak is identified as DMMP. (The other two are DEMP and DIMP respectively.)

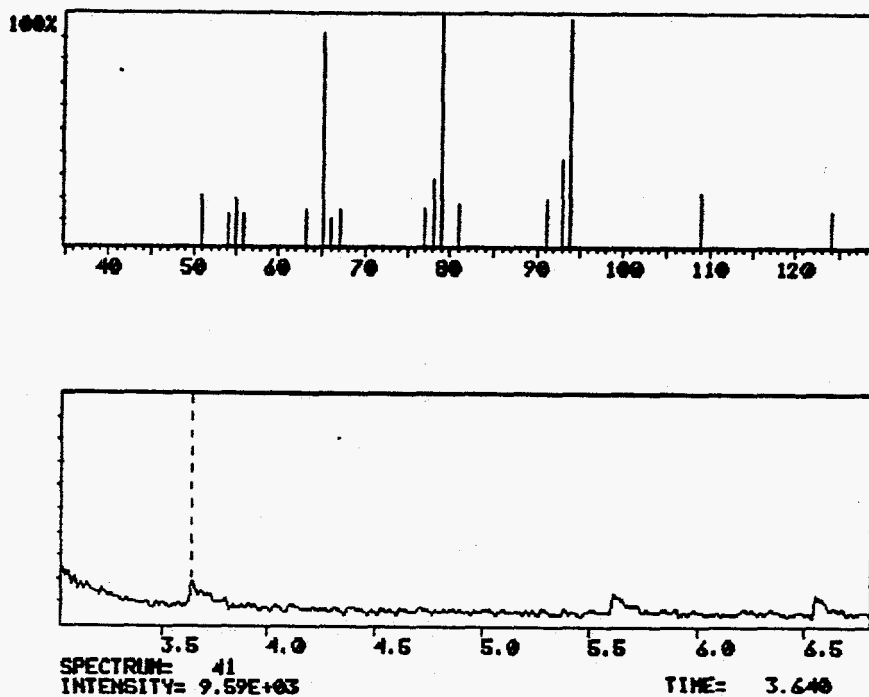


Figure 13. Three Small Peaks are Detected in the First Increment

The reason that these compounds were not originally detected is that, since the analyst could not view them, no background subtraction could be performed. In this case, there is a major background "spike" which degrades the fit coefficient. (See Figure 15.)

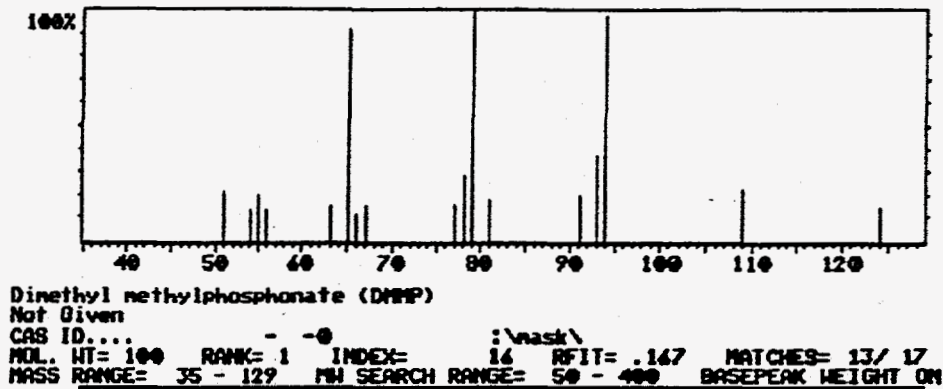


Figure 14. Identification of the First Small Peak as DMMP

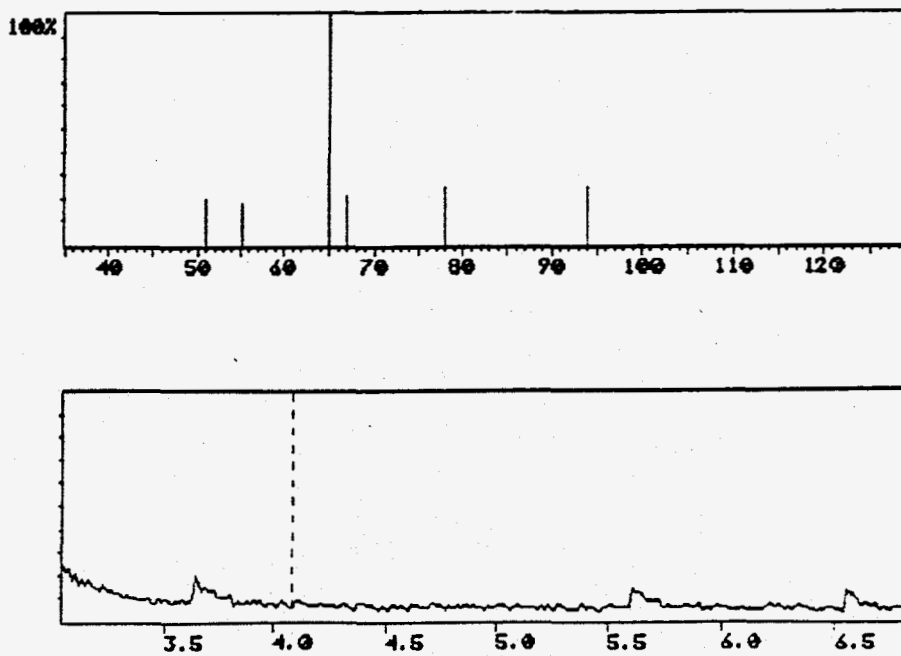


Figure 15. Background "Noise" which Degrades Fit Coefficient

A scheme to perform a blind, general background subtraction has been designed but has not been programmed into the code as yet. Additional subroutines to perform peak deconvolution have been discussed but have not been implemented.

The "hit" in the final increment is spurious, caused by the low fit coefficient. (See Figure 16.)

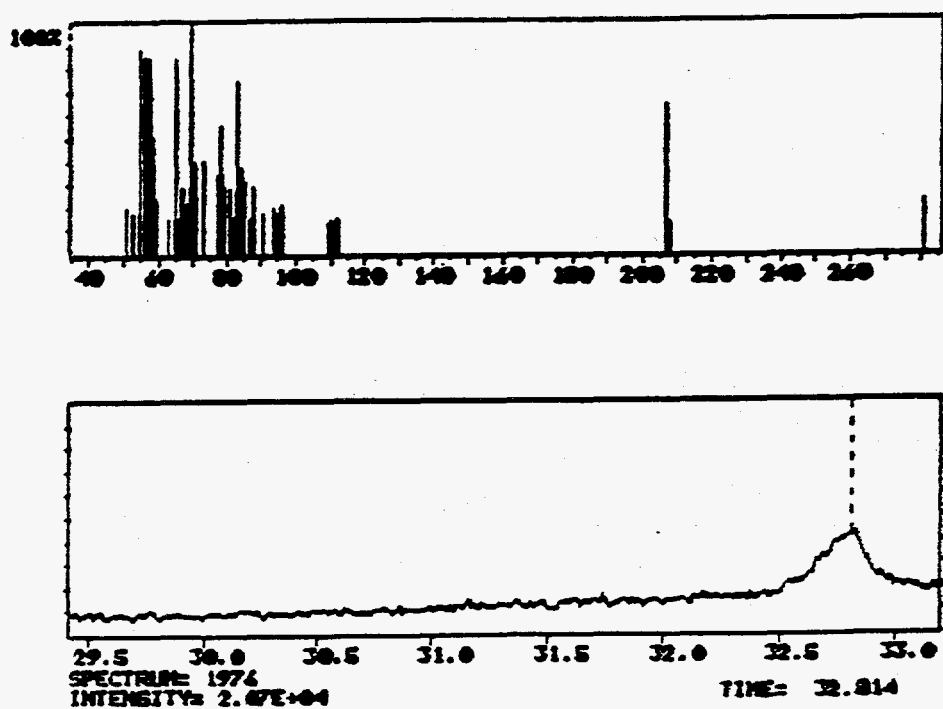


Figure 16. Spurious "Hit" Caused by Low Fit Coefficient

SUMMARY:

A shell program, "MASK" has been added to the LLNL standard GC-MS data analysis code "M CODE" to protect the confidentiality of information not relevant to the CWC. This program has been demonstrated to work well where high total ion counts are available. A subroutine for "blind" background subtraction needs to be added for cases where the peak is only slightly above baseline.

The authors wish to thank all those who have contributed to the support of this project. This work was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.