Target Organic Contaminant Library Development in Support of Sample Analysis at Mars (SAM)

Raul Garcia-Sanchez[1], Paul Mahaffy[2], Prabhakar Misra[1]
[1]Howard University, [2]NASA Goddard Space Flight Center Code 699

Introduction

The primary purpose of the research is to develop an organic contaminants database for the SAM project that supports the Mars Science Laboratory (MSL). Our work involves the development of an organic contaminants database that will allow us to determine what compounds are found here on Earth and would be inadvertently detected in the Mars soil and gaseous samples as impurities. In order to develop a comprehensive target database, we utilize the National Institute of Standards & Technology (NIST) Automated Mass Spectral Deconvolution and Identification System (AMDIS) and Ion Fingerprint Deconvolution (IFD) software to analyze the Gas Chromatography Mass Spectral (GCMS) data.

The MSL is NASA's Mars Exploration Program newest rover. It contains the most advanced suite of instruments to date. It is the biggest rover developed and will implement an innovative landing system.

Figure 1. Rover predecessor to MSL Size Comparison.

Sample Analysis at Mars (SAM) is the largest payload of the MSL rover. It contains equipment for chemical analysis such as a Gas Chromatograph (GC), mass Spectrometer (MS) and Tunable Laser Spectrometer (TLS). With these instruments, it will analyze the composition of soil and rock on Mars with the intention of finding organic compounds that are associated with life. The QMS detects gases sampled from the atmosphere or those released from solid samples by heating. The GC can separate out individual gases from a complex mixture into molecular components for QMS and GC stand alone analysis. The TLS implements a sensitive search for methane and makes precision measurements of oxygen and carbon isotope ratios in carbon dioxide.

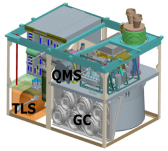


Figure 2. SAM Suite Configuration.

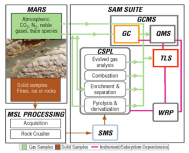


Figure 3. SAM Gas/Solid Sample Flow path.

Research Questions

- What organic compounds can be found in the rover structure? Where in the rover can they be found?
- How accurately can we determine the compounds from tests to those on the actual rover?
- How will identifying these compounds help when the time comes for the actual mission analysis?

Software

1. Automated Mass Spectral Deconvolution and Identification System (AMDIS): Used to extract spectra from the GCMS data we obtain through heating tests and creating a target compounds library.

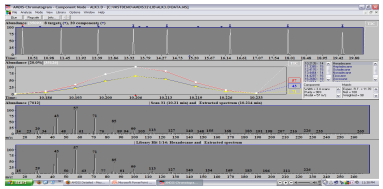


Figure 4. AMDIS GCMS analysis interface. Allows the use of a text-based target compound library used for deconvolution of the GCMS data.

2. National Institute for Standards and Technology (NIST) MS Search: Used for searching organic contaminant libraries for spectra to match the unknown compounds in our GCMS data runs.

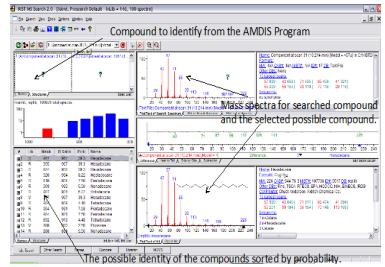


Figure 5. NIST Mass Spectral Search interface. Compares between the unknown compound and library compounds that are similar.

3. Ion Signature Quantitative Deconvolution (IFD): Used to analyze all the hidden peaks in the Total Ion Chromatograph. It is also used to create a database of compounds to search for after deconvolution.

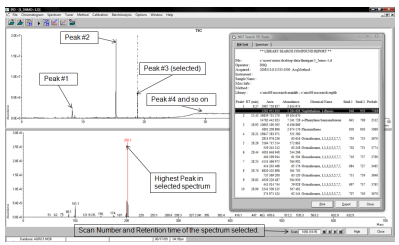


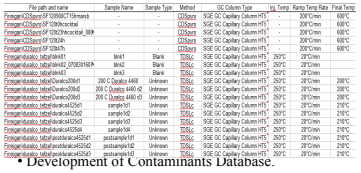
Figure 6. Ion Fingerprint Deconvolution software interface. Displays the GCMS file. Allows you to search the NIST library to help identify the unknown compound.

Methodology

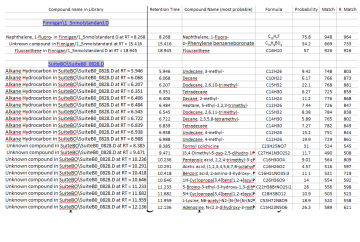
1. Utilize IFD to identify the TIC peaks, then select the most significant ones.
2. Run a similarity (based for highest match/rmatch ratio) and an identity search (based on probability) using IFD to see if different searches yield similar results. If they are different, compare all related spectra and interpret.
3. Export the TIC peak spectra to NIST.
4. Use NIST to identify the spectra and test if it agrees with IFD results.
5. Insert entry into contaminants spreadsheet.
6. Add the identified (or unidentified) spectrum into a user-created NIST Library. Also add to a target compound library using IFD or AMDIS.

Results

- Development of File Database.



- Development of Contaminants Database.



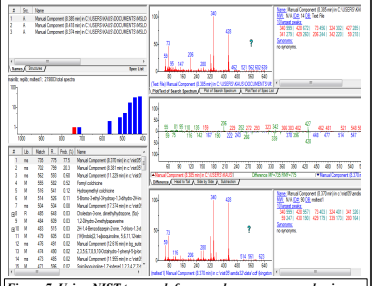


Figure 7. Using NIST to search for an unknown compound using a user-created library with compounds identified in other files.

Discussion

- **Development of File Database:** The file database is a spreadsheet file that contains file locations with information on the runtime parameters and environment in which the tests were taken. This database contains parameters such as: location of the file, the number by which it is identified in the material archive, location on rover, sample name and Type, the GC Column Type used, temperatures and the date the run was done. So far we have been able to obtain most of these parameters for over 1000 Finnigan Themo files. We are awaiting input from JPL to determine the material archive number and the location on the rover.
- **Development of Contaminants Database:** The contaminants database is a reference spreadsheet file that contains the compounds found in each file and how they can be identified by their library name. It lists the name and formula of the most likely compound according to the NIST search that matches an unknown compound. It also lists how similar the two compounds are in terms of their peaks (match/rmatch) and their abundance (percentage). Based on how similar they are, they are appropriately named in the library.
- **Development of Target Compound Library:** As seen in Figure 7, on the upper side of the figure we have the name, spectral and other information regarding our unknown spectrum, in the middle portion is a peak comparison and the lower part is the compound that most likely matches your unknown compound. It can be seen that our unknown spectrum is similar to another spectrum we had found in a previous file and had placed in the user-defined library. These library compounds have the format of: <Compound Name at RT = x.xxx in Filepath> and is how we are able to reference them in the previous database spreadsheets mentioned. We see that the spectra match very closely to each other and the peaks are similar as well in terms of abundance. After this, we go to our contaminants database, look for the component in the library by following the file path. Then we can determine what the known compound was identified as and how probable it is.

Future Work

- Continue analyzing more test data files as well as real-time data as it becomes available.
- Develop a database program prompt that will ease the retrieval of the file information instead of an indexed spreadsheet.
- Analyze Coldfinger (CF) data.

Reference

SAM MSL GSFC Atmospheric Experiments Laboratory Website: http://ael.gsfc.nasa.gov/marsSAM.shtml