

# Database and Library Development of Organic Species using Gas Chromatography and Mass Spectral Measurements in Support of Sample Analysis at Mars

**Raul Garcia, Prabhakar Misra**  
Department of Physics & Astronomy  
Howard University  
Washington, DC, USA  
rsanchez@howard.edu, pmisra@howard.edu

**Paul R. Mahaffy, Inge L. ten Kate**  
Atmospheric Experiments Laboratory  
NASA Goddard Space Flight Center  
Greenbelt, MD, USA  
Paul.R.Mahaffy@nasa.gov, Inge.L.TenKate@nasa.gov

**Abstract** – *This research work involves the development of an organic contaminants database, which will allow the identification of compounds that are found here on Earth and would be inadvertently detected in the Mars soil and gaseous samples as impurities. The database will be used for the Sample Analysis at Mars (SAM) instrumentation analysis on the Mars Science Laboratory (MSL) rover scheduled for launch in 2011. Based on the analysis of the laboratory-based Gas Chromatography-Mass Spectrometry (GC-MS) data of pyrolyzed and baked samples, organic compounds were successfully identified in the spectral data and have been successfully incorporated into an initial target compound database that will aid SAM in determining whether the components being analyzed come from Mars or are contaminants from either the rover itself or the Earth environment.*

**Keywords:** GC-MS, organic contaminants, SAM, MSL, rover, follow the water, NASA, Mars.

## 1 Introduction

Sample Analysis at Mars (SAM) [1] is the largest payload of the MSL rover (dubbed “*Curiosity*”). It contains 3 instruments for chemical analysis: a gas chromatograph (GC), a quadrupole mass spectrometer (QMS) and a tunable laser spectrometer (TLS). With these instruments, SAM will analyze the composition of soil and rock on Mars with the intention of finding organic compounds (that are associated with life). Samples are introduced in one of 72 cups in a sample manipulation system, which is then either directly heated or first derivatized and subsequently heated, so that the gases that evolve can be analyzed using one or a combination of the above-mentioned instruments. Direct atmospheric sampling will be carried out through an atmospheric inlet. The QMS will directly measure gases sampled from the atmosphere or released from the solid samples by heating. The GC will separate out individual gases from a complex mixture into molecular components for QMS and GC stand-alone analyses. The TLS will implement a sensitive search for methane and make precision measurements of oxygen and carbon isotope ratios in carbon dioxide.

This paper describes the development an organic contaminants database for SAM utilizing GC-MS analysis tools; it will allow the identification of compounds that are found here on Earth and perhaps be inadvertently detected in the Mars soil and gaseous samples as impurities. In order to develop a comprehensive target database, the National Institute of Standards & Technology (NIST) Automated Mass Spectral Deconvolution and Identification System (AMDIS) and Ion Signature’s Ion Fingerprint Deconvolution (IFD) software were utilized to analyze data recorded in a controlled laboratory environment using GC-MS. The compound spectra that have been added to the library are used to identify compounds that might be found as impurities from potential Mars mission samples.

## 2 Organic contaminants

Contamination due to outgassing of spacecraft materials is an issue not only in the framework of *in situ* organic analysis on planetary surfaces, but also for other spacecraft and space-based telescopes. Several studies have been done on outgassing of polymeric materials [2-5]. Contamination control is also essential for several areas of terrestrial geochemical research, and lessons learned from ocean floor research and glacial ice cores, for example, can be useful for space mission planning [6, 7]. To address this contamination problem, NASA commissioned an Organic Contamination Science Steering Group (OCSSG), which identified four primary areas of focus prior to the announcement of opportunity for the MSL Mission. These strategies were developed to provide direction for the engineering and operations teams responsible for the design, fabrication, assembly, and operations of Mars landed systems and the science teams responsible for analysis of the data from Mars [8]. Some of these strategies were also employed during the Viking mission [9]. The strategies include the following:

1. Identify and quantify the contaminants of most concern with regard to their possible adverse impact on the goals of each Lander mission.
2. Understand methods of quantifying residual contamination and its abundance and transport during all stages of Lander development and operation.

3. Suggest possible contamination mitigation options.
4. Examine the use of controls and facility-provided standards to be analyzed by Lander instruments after arrival at Mars.

Each sufficiently volatile organic material that becomes part of the rover or the MSL aeroshell, by way of design or flight hardware processing or environment, is a potential source of organic molecules. These molecules can be in the form of gasses (outgassing from molecules within the bulk material or molecular films on the surface of the material), films (deposition and condensation of gaseous molecules or contact transfer of contaminants), or particles. The goal in designing and building the MSL has been to minimize the overall quantity of these contaminants and have knowledge of them. To assure that the quantity of contaminants arriving at the sensitive inlets are within acceptable standards, the contamination control plan relies on (1) identification of potential sources through rigorous review of drawings and processes, (2) predictions of contaminant levels at sensitive surfaces, (3) elimination or reduction of sources (through cleaning, bake-out, pyrolysis, material substitutions, directed venting, filters, molecular absorbers, etc.), and (4) addition of design elements to prevent transport (barriers, seals, relocating components, directed venting, etc.). A detailed overview of the MSL mitigation of contamination efforts can be found in [10].

Even with the best efforts at mitigation, some residual contamination is likely to be observed by sensitive instruments on the surface of Mars. The use of controls is important for achieving definitive scientific conclusions, as is the characterization of terrestrial organic material transported to Mars that might credibly mix with atmospheric or solid phase Martian samples. The study described in this paper is focused directly on the materials used on MSL and their impact on the analytical instruments in the SAM suite. The TLS and GC-MS instruments in the SAM suite are designed to detect methane and other organic molecules with sub-parts-per-billion sensitivity. This high sensitivity demands that contaminant background levels be very low. Target levels for this background contamination in the atmospheric samples and the solid samples, respectively, must be evaluated. Cross contamination of organic material between different samples is also of concern.

One goal of the contamination control plan is to have *a priori* knowledge of the materials used; especially those that occupy large areas of the GC-MS data space and produce multiple chromatographic and mass spectral peaks. To achieve this, a materials testing program has been put into place. GC-MS data is collected for organic-releasing materials used on the exterior of the rover and for materials used in large quantities in the interior of the rover. Assembled parts of the instruments or the rover are being tested when material sampling cannot be obtained. The data obtained will ultimately be compared to flight data to

facilitate identification of terrestrial contaminants within a SAM measurement. Identification of the source may permit the elimination or reduction of its effect through changes to the mission operations plan (changing temperatures, orientation with respect to wind, time of sampling, etc.). Additionally, samples of materials are kept and stored in an archive for testing during the mission – should the need for future testing arise.

The identification of these compounds will allow – when the time for the actual mission analysis comes – comparison of the compounds that were found on the rover during its preparation phase and filtering out those spectra that are likely to have arisen from potential contaminants, and thereby reduce the likelihood of detecting terrestrial impurities erroneously on Mars because of SAM's sensitive measurements.

### 3 Contamination Analysis

This section discusses the tools utilized for the analysis of the GC-MS files, as well as the analytical method used in this research.

#### 3.1 Summary of software

AMDIS is free software used to extract clean spectra from complex GC-MS data and allows one to build one's own target compound library. This in turn allows one to analyze the samples and match them with the components one wants. It is especially useful when one has noisy Total Ion Current (TIC) spectral files. AMDIS allows deconvolution of the data, matching of compounds and the finding of target compounds from the selected library and subsequently enables one to search for components in the NIST library through the NIST MS Search program.

The IFD software features a set of advanced deconvolution algorithms that identify target compounds present even in difficult and complex mixtures. This software extracts each compound's characteristic signature, disentangling it from the surrounding sample matrix.

The NIST library search program identifies and returns the compounds that are the most likely compound being searched for. Given the MS for a compound at a certain retention time in the GC, one can run a search command to see if the compound matches an entry in the NIST library.

#### 3.2 File/data types

Database files are currently in Microsoft Excel spreadsheet format. Once all the pertinent information from the files is obtained, it will be possible to develop a more interactive and user-friendly database that will be searchable through an interface rather than having to access it directly from the file.

The spectra are added to a library in MSP format and the parameter information edited based on what the NIST library search determines (i.e. identify the unknown compound, name it and give it the corresponding compound element composition). Once done, they are imported into the NIST MS Search program and a NIST-formatted library created from the exported spectra.

The data is in the form of GC-MS files that have been run on a Finnigan SSQ7000 Thermal Desorption Model 2000 Gas Chromatograph (equipped with an SGE Forte GC Capillary Column HT5 of length: 25m, I.D: 0.22 mm, Film: 0.1µm). The data contains information about possible compounds in the rover material samples that have been supplied by the Jet Propulsion Laboratory (JPL) and have been pyrolyzed using this instrument. Materials such as polymers, paint, and o-rings, are examples of some of the material samples that have been pyrolyzed in this process.

### 3.3 Description of analytical method

IFD is utilized to identify the Total Ion Current (TIC) peaks, and subsequently select the most significant ones. A similarity (based on highest match/rmatch ratio) and an identity search (based on probability) are run using IFD. Matching is a peak comparison between the unknown and the known spectra, while reverse matching (rmatch) is the comparison between known and unknown spectra & with emphasis on peak weighting and abundance levels. Hence, rmatch values will always be higher (both go up to 999), unless a perfect match is found. The most likely spectrum of one search is compared with the one from the other search, and the most probable one determined. The spectrum for each of the peaks identified in the NIST TIC peak search is obtained and exported into MSP text files.

The NIST Search Library is used to read the spectra and obtain information regarding the most likely compound to match the spectra. This is done for all the spectra found; they are then combined and exported into a single MSP library file. In cases in which the search does not provide an accurate compound to match the unknown compound, it is left as an unknown compound of unknown composition. When obtaining an accurate matching spectrum, AMDIS or a text editor is used to rename the unknown spectra into the name format that will be used throughout the entire library. The modified library is then imported to the NIST MS Search program and a NIST-formatted library created for comparing other spectra against those that have already been found.

An example of this method follows. The results from the analysis done on the Hysol EA9360 GC-MS are presented. Table 1 summarizes the TIC Peak identification using the Ion Fingerprint Deconvolution (IFD) software.

Table 1. TIC Peaks for HYSOL EA9360.

Ret.Time [min.]	Chemical Name	Match	Prob
7.02	Phenol	0.996845	7130
7.6	4-Cyanocyclohexene	0.994505	4654
8.03	Benzyl Alcohol	1	7849
11.17	5-Amino-1,3,3-trimethylcyclohexanemethylamine	1	8783
11.33	5-Amino-1,3,3-trimethylcyclohexanemethylamine	1	9684
12.411	1-Dodecanol	0.995708	2425
15.36	Phenanthrene-D10	0.998921	7627
16.048	Hexadecanoic acid, methyl ester	0.996703	7266
17.29	9-Octadecenoic acid, methyl ester, (E)-	0.990142	1952
17.41	Octadecanoic acid, methyl ester	0.996746	6365
17.792	Hexadecanoic acid, butyl ester	0.964871	6455
19.185	O-Benzyl-L-tyrosine	0.961436	2425
19.415	2,2'-Ethylidenebis(4,6-di-tert-butylphenol)	0.969061	9570
20.172	Eicosanoic acid, phenylmethyl ester	0.887425	8098
22.62	Oxirane, 2,2'-[(1-methylethylidene)bis(4,1-phenyleneoxymethylene)]bis-	1	9818
24.77	Chromone, 2-[2-[3,5-dihydroxyphenyl]ethenyl]-5-hydroxy-	0.927536	5320
43.22	2-Phenyl-4,6-di(4-acetylaminophenyl)pyrimidine	0.85025	2312

As can be seen from the table, some of the spectra obtained from these files have very high match/rmatch ratio and probabilities. The green cells indicate a very high match, to the point where it can be said that the unknown spectrum and the NIST spectrum match are nearly identical (as seen in Figure 1). The yellow cells indicate a good match, but with a higher amount of peak differences. In both cases, however, the matched spectrum retains the general shape of the unknown spectrum. For the red cells, a match is not properly found, as the compound NIST identifies as the most likely one does not share similar properties, as can be seen in Figure 2.

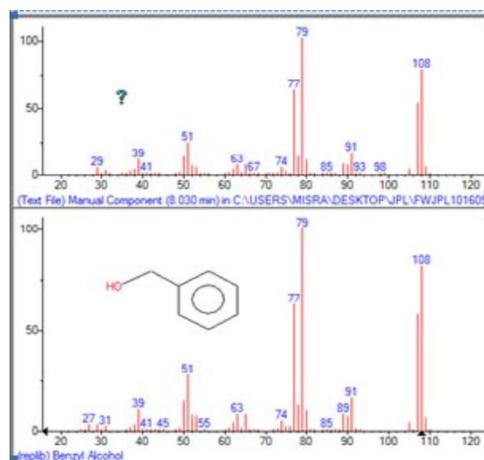


Figure 1. Good (green) library spectrum match.

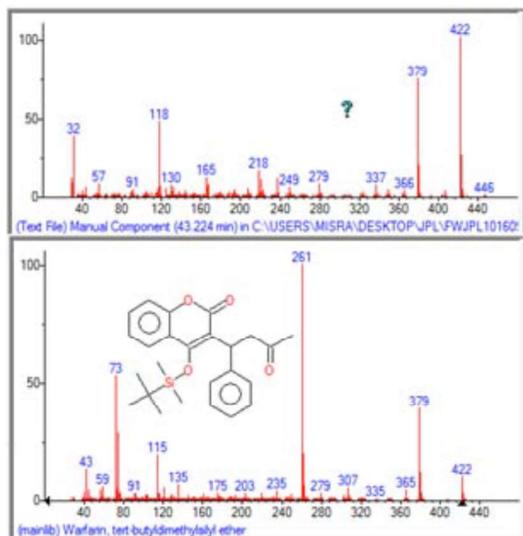


Figure 2. Poor (red) library spectrum match.

Now we place the spectra in the above portion of each of these figures into the library under their corresponding nomenclature names: 8.030RT\_Hysol\_EA9360 and 43.224RT\_Hysol\_EA9360.

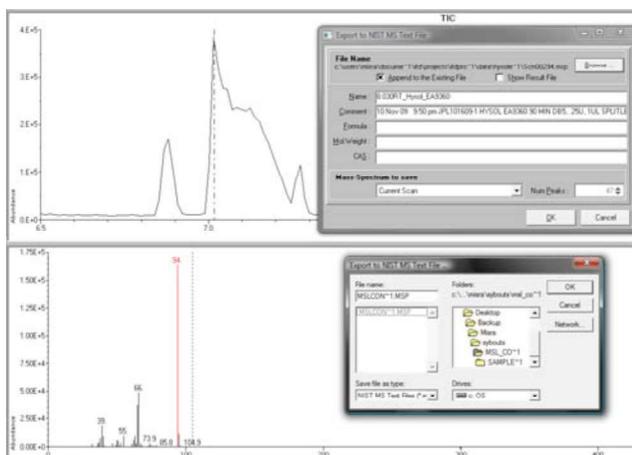


Figure 3. Exporting spectrum from IFD to MSP file.

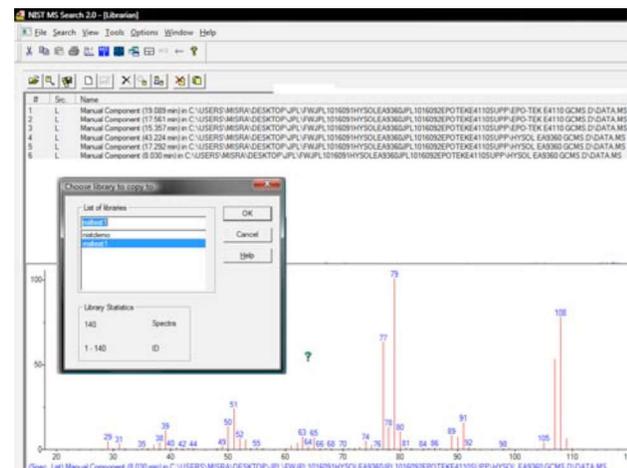


Figure 4. Using Librarian utility in NIST MS Search to add to a NIST-formatted library.

After exporting these spectra from IFD (Figure 3), we can import them into NIST MS Search or directly from AMDIS, which allows us to correlate the data by comparison between AMDIS and IFD results; then we run NIST Librarian to generate a NIST-formatted library (Figure 4). Now these spectra are in our organic compounds library and we can compare these spectra to other unknown spectra that are found in the GC-MS files and are similar to the spectra in the library.

## 4 Contamination database

In this section the data analysis results and how the contaminant spectra found in the data have been organized and summarized into a database are discussed.

### 4.1 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. The purpose of this database is to keep track of the conditions under which each file was run, so that it is possible while comparing with data taken in the Mars environment, to better assess its similarity with the library.

The file 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 most of these parameters for over 1000 Finnigan Thermo files have been obtained. The database is useful because it allows one to track each file that has been analyzed and implemented into the library. The file path and name are based on how the data is organized in the file system. For example, duralco\_tefzel\postduralco4525d4.D indicates that this GC-MS file is found in the duralco\_tefzel folder in the file system.

In Table 2 a small portion of the Microsoft Excel spreadsheet database is shown. As can be seen, it has a listing of each file and its path in the file directory, and it is linked directly to each file, so that it can be readily opened (without the need to search for it). The sample name, where the sample is located on the rover, and the JPL materials archive number (not shown) are also included. Such information will allow the identification of what type of material each file corresponds to and how likely it is to be a source of contamination (e.g. contamination is more likely on materials along the GC-MS flow path and the arm that delivers the sample versus an outside plate of the rover). Temperature information on the pyrolyzation/bakeout process for each of these files, allows comparison of this approach with any other method that might be used in the actual mission; and such information will aid in assessing the probability and relevance of the match-mismatch in identifying potential terrestrial contaminants.

Table 2. Sample File Database.

File path and name	Sample Name	Final Temperature
duralco_tefzel\Duralco200d1.D	200 C Duralco 4460	200°C
duralco_tefzel\postduralco4525d3.D	postsample1d3	210°C
duralco_tefzel\postduralco4525d4.D	postsample1d4	210°C
duralco_tefzel\posttefzelm22759d1.D	postsample2d1	200°C
duralco_tefzel\posttefzelm22759d2.D	postsample2d2	200°C
duralco_tefzel\tefzelm22759d3.D	sample3d3	200°C
duralco_tefzel\tefzelwire.D	Tefzel wire	50°C
duralco_tefzel\tefzelwired2.D	Tefzel wire 48 hrs	50°C
duralco_tefzel\tefzelwired3.D	Tefzel wire 72 hrs	50°C
duralco_tefzel\tefzelwired4.D	Tefzel wire day 4	50°C
EcothecR2\ecotheccpaintd3.D	sample1d3x1	70°C
EcothecR2\ecotheccpaintd4.D	sample2d4x1	70°C
EcothecR2\post ecotheccpaintd1R.D	sample3d1Rx2	70°C
EcothecR2\post ecotheccpaintd2.D	sample2d2x2	70°C
Insulation\HCTrap\306blackd1.D	Chemglaze306blackd1	100°C

## 4.2 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 names. The purpose of the contaminants database is to identify each compound that has been found in the files and to correlate it to a material sample.

The database identifies each compound found by its name in the compound library. It also lists the name and formula of the most likely compound according to the NIST search that matches an unknown compound and 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.

The contaminants database is particularly useful for determining which compounds can be found on each of the GCMS files analyzed and is closely related to the organic compound library since it contains all the summarized information found in the library.

Table 3 shows a segment of the contaminants database. The database is structured so that each compound is organized per file, with the file name linked to the file database, which in turn contains information regarding the sample type and the temperatures used in the sample runs.

Table 3. Sample Contaminants Database.

Compound name in Library	Retention Time	Formula	Probability
<a href="#">Finnigan\1_5nmolstandard.D</a>			
Naphthalene, 1-fluoro-	8.268	C <sub>10</sub> H <sub>7</sub> F	75.8
Unknown compound	15.416	C <sub>12</sub> H <sub>9</sub> BO <sub>2</sub>	34.2
Fluoranthene	18.945	C <sub>16</sub> H <sub>10</sub>	57
<a href="#">SuiteBO\SuiteBO_0828.D</a>			
Alkane Hydrocarbon	5.946	C <sub>12</sub> H <sub>26</sub>	9.42
Alkane Hydrocarbon	6.068	C <sub>10</sub> H <sub>22</sub>	6.17
Alkane Hydrocarbon	6.207	C <sub>15</sub> H <sub>32</sub>	22.1
Alkane Hydrocarbon	6.351	C <sub>14</sub> H <sub>30</sub>	6.27
Alkane Hydrocarbon	6.938	C <sub>12</sub> H <sub>26</sub>	15.2
Unknown compound	8.385	C <sub>23</sub> H <sub>25</sub> NO <sub>7</sub>	31
Unknown compound	9.471	C <sub>27</sub> H <sub>41</sub> N <sub>3</sub> O <sub>2</sub> S <sub>2</sub>	11.7

## 4.3 Development of Target Compound Library

The primary purpose of the research is to develop an organic contaminants database for the SAM project that supports MSL. It contains the spectra for all the compounds found in each of the files that have been analyzed. In cases where the library returns clear results, each spectrum is identified with its corresponding most likely compound. In the cases where it is not found in the library, it is left as an unknown compound, since it is still possible that its corresponding spectrum might be found in the Mars analysis, even if it cannot be properly identified due to GC-MS noise.

The target organic compounds library will allow one to organize the potential organic contaminant compounds and their associated spectra and will enable comparison with results from the actual Mars mission – given certain conditions (such as temperature, matching retention time, etc.) that can be determined using the file and contaminant databases. These library compounds have the format of <0.00RT\_File> and it is how one can reference them in the previous database spreadsheets mentioned.

In Figure 5, in the top half of the figure, the name, spectral and other information regarding the unknown spectrum is present, while in the lower half is shown the compound that most likely matches the unknown compound.

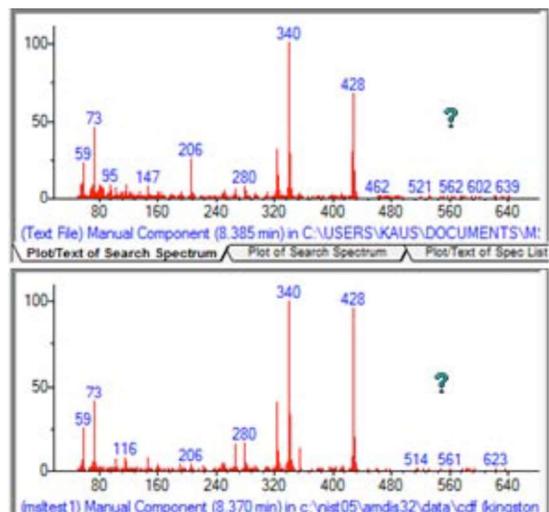


Figure 5. Comparison of unknown spectrum in file A versus spectrum found in file B and placed in the library.

## 5 Conclusions

In previous Mars missions, there has been no concrete and comprehensive reliable Earth compound library that could be run to test for accuracy of sample detection and identification. However, due to the sensitivity of the SAM instrumentation, it is of great importance that necessary steps are taken to deal with possible impurities, no matter how small, that may compromise the validity of the sample analysis in actual Mars data.

The databases developed are useful from both organizational and practical points of view. Through these databases it is possible to organize large volumes of GC-MS data, while at the same time breaking down the components that each material sample is made off. This in turn allows easy and fast access to information that will be critical when performing analysis on the data that the SAM instrumentation will record.

The above research methodology indicates conclusively that the library is successful in identifying compounds using a multi-pronged approach utilizing both AMDIS and IFD software. Additional information, such as retention indexing and retention times, are being used to further weight compounds that are similar to each other, but occur under different conditions, such as different temperatures because of the nature of the material being tested.

Future work is being directed towards making access to the information contained in these database spreadsheets easier via a command prompt file search routine. The data from the cold finger analysis will also be integrated into the library. Simulations of possible situations that may surface during the mission can also be analyzed and studied in order to improve the effectiveness of the database and library. In addition, one can utilize the MSP library file as the deconvolution library for AMDIS and determine its accuracy.

## Acknowledgements

The authors (R.G.S. and P.M.) would like to acknowledge with appreciation the financial support from the DC Space Grant Consortium (DCSGC) provided to Raul Garcia-Sanchez and the award of a Faculty Summer Fellowship to Prabhakar Misra that allowed them to spend the 2009 summer in the Atmospheric Experiments Laboratory (Code 699/GSFC). The assistance of other members of Paul Mahaffy's research group on this project, especially Therese Errigo, Kiran Patel and Chris Johnson, is also gratefully acknowledged.

## References

- [1] P.R. Mahaffy, "Exploration of the habitability of Mars: Development of analytical protocols for measurement of organic carbon on the 2009 Mars Science Laboratory," *Space Science Reviews*, Vol.135, Nos. 1-4, pp. 255-268, 2007.
- [2] P.R. Young and W.S. Slemp, "The performance of selected polymeric materials exposed to low Earth orbit," *Polymers for Advanced Technologies*, Vol. 9, No. 1, pp. 20-23, 1998.

[3] R. Rampini, L. Grizzaffi and C. Lobascio, "Outgassing kinetics testing of spacecraft materials," *Materialwissenschaft Und Werkstofftechnik*, Vol. 34, No. 4, pp. 359-364, 2003.

[4] C.H.W. Chang, "Examination of material outgassing in the air and purging flow environments," Proc. of SPIE Vol. 5526, Optical Systems Degradation, Contamination and Stray Light, SPIE, Denver, CO, Aug. 2004.

[5] R.M. Villahermosa and P.L. Joseph, "Characterization of outgassed contaminants from polymeric spacecraft materials," Proc. of SPIE Vol. 5526, Optical Systems Degradation, Contamination and Stray Light, SPIE, Denver, CO, Aug. 2004.

[6] B.C. Christner, J.A. Mikucki, C.M. Foreman, J. Denson and J.C. Priscu, "Glacial ice cores: A model system for developing extraterrestrial decontamination protocols," Proc. 3rd International Conference on Mars Polar Science and Exploration, Academic Press Inc., Elsevier Science, Lake Louise, Canada, Oct. 2003.

[7] J. Eigenbrode, L.G. Benning, J. Maule, N. Wainwright, A. Steele, H.E.F. Amundsen and A. Team, "A field-based cleaning protocol for sampling devices used in life-detection studies," *Astrobiology*, Vol. 9, Issue 5, pp. 455-465, 2009.

[8] P.R. Mahaffy, D. Beaty, M. Anderson, et al, "Science priorities related to the organic contamination of martian landers," [unpublished white paper] MEPAG Topical Analysis Reports, Mars Exploration Program Analysis Group, Internet URL: <http://mepag.jpl.nasa.gov/reports/index.html>.

[9] D.A. Flory, J. Oró and P.V. Fennessey, "Organic contamination problems in the Viking molecular analysis experiment," *Origins of Life and Evolution of the Biosphere*, Vol. 5, pp. 443-455, 1974.

[10] I. L. ten Kate, J.S. Canham, P.G. Conrad, T. Errigo, I. Katz and P.R. Mahaffy, "Mitigation of the impact of terrestrial contamination on organic measurements from the Mars Science Laboratory," *Astrobiology*, Vol. 8, Issue 3, pp. 571-582, 2008.