



ELSEVIER

Contents lists available at ScienceDirect

Planetary and Space Science

journal homepage: www.elsevier.com/locate/pss

COSIMA calibration for the detection and characterization of the cometary solid organic matter



Léna Le Roy^{a,b,*}, Anais Bardyn^{b,a}, Christelle Briois^a, Hervé Cottin^b,
Nicolas Fray^b, Laurent Thirkell^a, Martin Hilchenbach^c

^a Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E), UMR 7328 CNRS – Université d'Orléans, 3A Avenue de la Recherche Scientifique, 45071 Orléans Cedex 2, France

^b Laboratoire Interuniversitaire des Systèmes Atmosphériques, LISA, UMR CNRS 7583, Université Paris Est Créteil (UPEC) et Université Paris Diderot (UPD), Institut Pierre Simon Laplace, 61 Avenue du Général De Gaulle, 94010 Créteil Cedex, France

^c Max Planck Institute for Solar System Research (MPS), Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

ARTICLE INFO

Article history:

Received 29 April 2014

Received in revised form

1 August 2014

Accepted 28 August 2014

Available online 10 September 2014

Keywords:

Organic compounds

Calibration

COSIMA

Rosetta

TOF SIMS cometary grains

Comet

ABSTRACT

On the orbiter of the Rosetta spacecraft, the Cometary Secondary Ion Mass Analyser (COSIMA) will provide new *in situ* insights about the chemical composition of cometary grains all along 67P/Churyumov–Gerasimenko (67P/CG) journey until the end of December 2015 nominally. The aim of this paper is to present the pre-calibration which has already been performed as well as the different methods which have been developed in order to facilitate the interpretation of the COSIMA mass spectra and more especially of their organic content. The first step was to establish a mass spectra library in positive and negative ion mode of targeted molecules and to determine the specific features of each compound and chemical family analyzed. As the exact nature of the refractory cometary organic matter is nowadays unknown, this library is obviously not exhaustive. Therefore this library has also been the starting point for the research of indicators, which enable to highlight the presence of compounds containing specific atom or structure. These indicators correspond to the intensity ratio of specific peaks in the mass spectrum. They have allowed us to identify sample containing nitrogen atom, aliphatic chains or those containing polyaromatic hydrocarbons. From these indicators, a preliminary calibration line, from which the N/C ratio could be derived, has also been established. The research of specific mass difference could also be helpful to identify peaks related to quasi-molecular ions in an unknown mass spectrum. The Bayesian Positive Source Separation (BPSS) technique will also be very helpful for data analysis. This work is the starting point for the analysis of the cometary refractory organic matter. Nevertheless, calibration work will continue in order to reach the best possible interpretation of the COSIMA observations.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The European Space Agency (ESA) Rosetta cometary mission could be a milestone in the exploration of the Solar System for the *in situ* detection of organic compounds. Previous space probes with instruments enabling direct *in situ* measurements, *i.e.* without teledetection, of carbonaceous molecules have proven how difficult and challenging is the conception and operation of an instrument tailored to detect them. The global context of the explored environment which is most of the time poorly

characterized, if not unknown, when the instrument is conceived, can have unexpected implications jeopardizing its science capabilities. It was the case for the NASA Viking landers for which the astrobiology experiments including gas chromatographs (GC) were not able to detect organic compounds (if any) due to the strong oxidizing conditions of the Martian environment (Klein, 1978). More recently, the first results released by the Sample Analysis at Mars (SAM) instrument on-board the NASA Curiosity rover are quite ambiguous regarding the origin of the carbon in the carbo-chlorinated molecules detected: either endogenous to Mars, or with carbon imported from Earth and reacting with oxidizing Martian perchlorates during the analytic process (Leshin et al., 2013; Ming et al., 2014). The *in situ* search for organic compounds in the atmosphere of Titan, thanks to the coupling of the Aerosol Collector Pyrolyser (ACP) and Gas Chromatograph–Mass

* Corresponding author. Present address: Center for Space and Habitability, University of Bern, Sidlerstrasse 5, 2012 Bern, Switzerland. Tel.: +41 31 631 34 43; fax: +41 31 631 44 05.

E-mail address: lena.leroy@csh.unibe.ch (L. Le Roy).

Spectrometer (GCMS) instruments on-board the ESA Huygens probe, has been less rewarding than expected. About 20 organic molecules have been detected in the atmosphere of this satellite of Saturn since the Voyager missions, and many other compounds were anticipated according to laboratory simulations both in the gaseous and in the solid phase (Coll et al., 1999). Nonetheless the *in situ* measurements during the probe descent through the atmosphere and landing have not revealed any new molecule that was not already detected with remote sensing techniques (Niemann et al., 2005; Niemann et al., 2010). This time, the lack of data was not due to any unexpected peculiar environmental conditions interfering in the analytic suite, but rather some issue at the ACP/GCMS coupling, which resulted in an extremely limited sample transfer from ACP to GCMS. Nevertheless it was shown that Titan aerosol particles include a solid organic N containing refractory core (Israël et al., 2005). Almost paradoxically, the most rewarding detection of organic compounds in the atmosphere of Titan was the utterly unexpected detection of high molecular weight organics (up to ~ 8000 u/e) in the upper atmosphere by direct mass spectrometry thanks to the Ion and Neutral Mass Spectrometer (INMS) instrument (Waite et al., 2007).

The exploration of comets has provided so far the most encouraging results concerning the analysis of organic matter using *in situ* techniques. In 1986, comet 1P/Halley has been visited by five space probes, Sakigake and Suisei, two Japanese probes, Vega 1 and 2, two Soviet probes, and Giotto, a European probe. The Sakigake and Suisei spacecrafts stayed rather far from the nucleus of the comet (7106 and 152,400 km respectively) and had no *in situ* capabilities, but the Vega 1 and 2 and Giotto spacecrafts were equipped with a series of mass spectrometers devoted to the analysis of the gaseous and the solid phase in the coma.

Comets are a key to understand the origin of the Solar System, and maybe the origin of life. Rosetta would be an unparalleled chance to make a real breakthrough into these questions only if the instruments analyzing the organic composition of comet 67P/Churyumov–Gerasimenko would work properly, but also if sufficient calibration campaigns (before, during and most probably after their operation in the vicinity of the comet) have been conducted. Indeed, a complex mixture of organic material is anticipated, and the interpretation of results should be extremely challenging. More than 20 volatile molecules have been detected in comets by spectroscopic surveys including H₂O, CO and CO₂ for the most abundant, but also a large range of small organic compounds such as CH₄, CH₃OH, HCN, H₂CS (Bockelée-Morvan et al., 2004; Mumma and Charnley, 2011). Among them, the distribution density in the atmosphere of molecules such as H₂CO cannot be explained only by its direct release from the nucleus, but rather through a distributed source that could be due to the degradation of polyoxymethylene (H₂CO polymer) in the solid phase (Cottin et al., 2008). Hence, complex organic compounds, *i.e.* with high molecular mass, should be present in the solid phase. This idea is also supported by measurements performed in 1986 with the mass spectrometers on-board the Giotto and Vegas spacecrafts. Spectra showed that both the gaseous (Mitchell et al., 1992) and the solid phase (grains) (Kissel and Krueger, 1987) were containing organic molecules with higher masses than those of the molecules detected by remote sensing techniques in the gaseous phase. Some of the grains analyzed in the atmosphere of comet 1P/Halley seem to be essentially made of a mixture of carbon, hydrogen, oxygen and nitrogen (Fomenkova, 1999). The precise nature of the organic content detected is still rather elusive and only tentative assignments of spectral features to actual molecules have been proposed (Kissel and Krueger, 1987) because the bulk matter was analyzed without any prior separation, and the resolving powers of the instruments were too low to avoid an important spectral confusion.

NASA Stardust mission was another opportunity to probe the composition of a comet (81P/Wild 2). A few *in situ* measurements by mass spectrometry (Cometary and Interstellar Dust Analyzer (CIDA) instrument) have confirmed an organic component of grains collected by this instrument (Kissel et al., 2004) but did not get our knowledge of the organic content of comets much beyond what was known since the post Halley era. On the other hand greatest expectations were placed in the samples returned to Earth. However, although those samples have been analyzed with the state-of-the-art ground analytical techniques (among them are some we cannot currently imagine to adapt for a space probe to perform *in situ* analyses) FTIR, HPLC, XANES (Sandford et al., 2006), the organic content of the cometary grains has been extremely difficult to distinguish from carbonaceous contamination in the aerogel (Sandford et al., 2010). To date, only glycine, the simplest amino acid, has been identified as an individual molecule present in the cometary material, unaltered by the collection process or contamination (Elsila et al., 2009).

The organic composition of comets can also be inferred from the analysis of natural proxies such as carbonaceous meteorites. Some of them, such as the Orgueil class CI meteorite, are suspected to be of cometary origin (Gounelle et al., 2006), and they have proven to contain organic material including soluble molecules such as amino acids and nucleobases (Botta et al., 2002; Martins et al., 2008, 2007; Sephton, 2002) as well as insoluble organic matter (IOM) which is an organic macromolecular material. Recent analysis by high resolution mass spectrometry of the Murchison class CM meteorite has shown an exceptional diversity of soluble organic compounds (Schmitt-Kopplin et al., 2010). Hundreds of thousands molecular formulae were revealed by those analyses, which means that most probably several millions of molecular structures are present in the meteorite. CI class meteorites are as rich in organic matter than CM (Glavin et al., 2010; Martins and Sephton, 2009), and such a complexity is to be expected in comets. The composition of micrometeorites is also seen as a good analog since some of them are thought to originate from comets. Amino acids have been detected in some of them (Matrajt et al., 2004) and a variety of them called Ultra Carbonaceous Antarctica Micrometeorites (UCAMMs), show very low metamorphism level (Dobrica et al., 2011), and a high content of organic matter (up to 70%) (Dartois et al., 2013; Duprat et al., 2010). The extremely low amount recovered so far prevents to conduct analyses similar to those performed on meteorites, but their discovery strengthens (if necessary) the interest of the study of comets in the frame of the study of organic matter in the Solar System.

Laboratory simulations complete the anticipated portrait of the organic content of comets that should be revealed by the Rosetta mission. They are based on the photolysis, irradiation and/or warming of icy mixtures made of the most abundant molecules detected in the gaseous phase: H₂O, NH₃, CH₃OH... (Cottin et al., 1999; Greenberg, 1982). The processing of these simple mixtures results in the production of a large amount of organic compounds which provides a glance on the anticipated complexity of comets with several thousands of molecules produced within a few hours of laboratory work (Danger et al., 2013). Such experiments are also crucial to understand how the pristine organic matter from a cometary nucleus can be processed when it is ejected on grains in the coma, and how measurements from the instruments on-board the Rosetta spacecraft can be linked to the composition of the nucleus (Briani et al., 2013; Le Roy et al., 2012). Moreover, laboratory simulations can provide essential material for the calibration of the instruments.

Even though little information is currently known about the exact nature of refractory organic matter present in comet's nuclei, some *in situ* measurements at 1P/Halley and some laboratory analysis of the Stardust samples and of some natural proxies have

been able to determine some elemental ratio. These analyses could give important clues concerning the origin and evolution of cometary materials. In particular the comparison of the cometary elemental ratios with the protosolar ones (Lodders, 2003) could constraint the origin and evolution of cometary materials.

The analysis of cometary grains of 1P/Halley by the mass spectrometers on board Giotto, Vega-1 and Vega-2 probes have shown that, in average, 1P/Halley has more or less protosolar abundances for almost all rock-forming elements. Nevertheless, H, C and N are significantly more abundant in 1P/Halley than in CI-chondrites and approach solar abundances (Jessberger et al., 1999; Jessberger et al., 1988). Moreover, it has also been shown that elemental ratios, especially the ones related to organic elements such as N/C and O/C, present large variations between individual grains (Jessberger et al., 1991). This point seems to have been confirmed thanks to the study of Stardust grains (Cody et al., 2008; Sandford et al., 2006). These analyses reveal that N/C ratio is ranging from 0.03 up to 0.2–0.3 in different grains (Cody et al., 2008; de Gregorio et al., 2011). This ratio presents also large variation in the insoluble organic matter of natural proxies (carbonaceous chondrites, IDPs and chondritic micrometeorites, UCAMM). For the carbonaceous chondrites, the N/C elemental ratio ranges from 0.01 to 0.05. The highest values correspond to the most primitive meteorites. Most IDPs and chondritic micrometeorites present bulk N/C ratio compatible with carbonaceous micrometeorites (Matrajt et al., 2003), whereas N/C ratio can span up to 0.1 or 0.2 in some micrometer sized inclusions. Recently, it has been shown that the UCAMMs are characterized by high nitrogen concentration with bulk values about 0.1 (Dartois et al., 2013). The highest N/C values could be tentatively associated with the most primitive and less altered materials. Thus, it is important to measure *in-situ* this ratio in comet 67P/CG to compare with the previous measurements. We could also try to correlate the N/C variations from grain to grain with the nature of organic and minerals contained in the grains.

COSIMA is a Time Of Flight–Secondary Ion Mass Spectrometer (TOF-SIMS) on-board the orbiter of the Rosetta spacecraft. The basic principle of the instrument can be described as follow. Grains ejected from the nucleus of the comet will be collected on metallic targets exposed outside the spacecraft. At regular time intervals, exposed targets will be replaced by new ones, and brought inside

the instrument to be pictured by a microscope (COSISCOPE) to locate the coordinates and the shape of any grains larger than about 15 μm . Selected locations on the targets can then be bombarded with an indium gun and secondary ions from the surface of the grains analyzed by time of flight mass spectrometry (Fig. 1). More details about the instrument can be found in Kissel et al. (2007). The resolving power of COSIMA ($M/\Delta M \sim 1400$ at 50% high of the peak at $m/z=100$) should enable to distinguish between the mineral and the organic components of grains thanks to the mass defect properties of elements. Nevertheless, this resolving power will not prevent spectral confusion among organic compounds. One peak on a mass spectrum measured by COSIMA can then be produced by several combinations of atoms, as well as several fragments of various compounds with similar masses can be seen as only one peak on the spectrum. Even though COSIMA is of course a wonderful tool with greater capabilities than an instrument such as PUMA on-board the Vegas spacecrafts (resolving power of ~ 1400 vs. ~ 100 , soft ionization vs. high velocity impact), which has analyzed grains in the atmosphere of comet 1P/Halley in 1986 (Kissel and Krueger, 1987), we cannot expect an analytical capacity at the scale of the molecular complexity we are expecting from comets. As current ground advanced techniques barely meet the analytical requirements, space borne instrumentation conceived in the 90s cannot be more advanced. A calibration of the instrument is therefore necessary to prepare the interpretation of the measurements and retrieve as much information as possible from the results. Mass spectra measured by TOF-SIMS are strongly influenced by the characteristics of the instrument such as the primary ions source, their energy and by the sample itself (cleanness and physical state). Therefore we cannot rely on existing databases. Over the last years, we have prepared the cometary rendezvous and characterize the instrument to learn its response to the analysis of various samples. This paper is focussing on the analysis of organic matter and calibration of the instrument to meet this goal. Calibration on COSIMA regarding mineral samples is presented in a separate paper (Krüger et al., in preparation).

Calibration of a mass spectrometer to prepare a cometary rendezvous is all but a straightforward process. It would be illusionary to pretend to constitute an exhaustive database with a reference spectrum for each organic compound expected to be

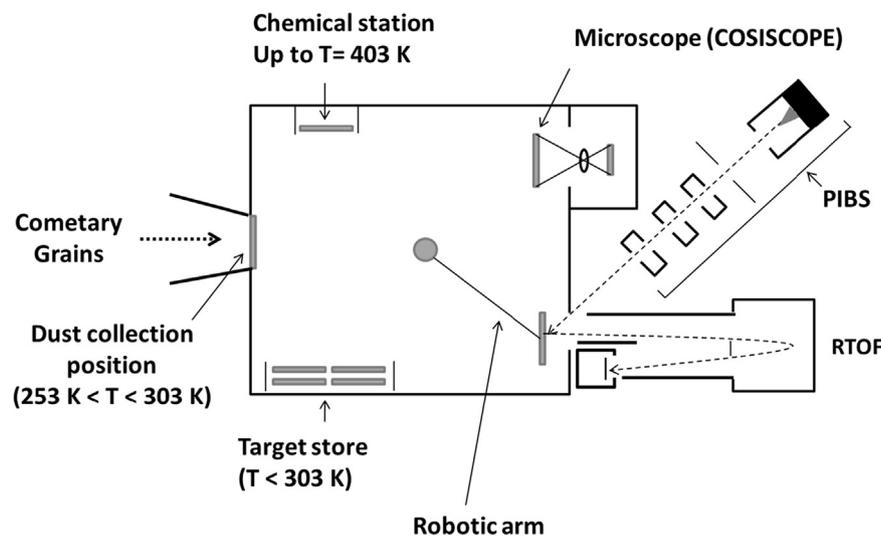


Fig. 1. Schematics of the COSIMA instrument on-board the ROSETTA spacecraft. The main instrument features are the dust collection position where metallic targets extracted from the target storage are exposed to space and should collect grains in the atmosphere of the comets, the microscope (COSISCOPE) in front of which exposed targets will be placed to detect the location of any grains, and the analysis position where grains will be bombarded with the Primary Ion Beam System (PIBS) and from which secondary ions will be accelerated and focalized into the Reflectron Time Of Flight (RTOF) mass spectrometer. A chemical station can also be used to warm the targets up to 403 K. This figure is adapted from Kissel et al. (2007).

present in comets. Such a reference TOF-SIMS library would have to include tens of thousands of molecules, not mentioning that we also have to get ready to analyze unexpected compounds. Henceforth, another strategy has been chosen to prepare the analysis of COSIMA measurements. Our study is focused on a selected number of organic compounds representative of various chemical families. They were selected not only on their likeliness to be actually detected on comets, but to cover a large panel of organic functions, (aliphatic and aromatic patterns, N-heterocycles, ethers, carboxylic acids and C, N, O, H ratios). We also have chosen to proceed step by step, from simple and pure compounds to learn how the instrument is 'responding' to organics, before studying complex mixtures of organics, analogs of cometary material, and interactions between organic and minerals. This paper presents the first step of the calibration of COSIMA regarding organic material and the systematic analysis strategy derived from our results on pure samples. Of course, we are aware that such samples are far from the anticipated nature of the samples COSIMA will analyze. But their analysis is the foundation of a more detailed work in progress on more realistic materials.

The purpose of this article is to present the calibration measurement done in order to prepare the data interpretation of COSIMA, more especially the organic content of the mass spectra. This work is divided in five sections.

The first one reviews the current knowledge about the cometary organic composition.

Section 2 presents the TOF-SIMS principle and the Orleans model instrument used for this study. The list, structure and choice of the samples are also given.

Section 3 deals with the fingerprint characterization of each chemical family analyzed in the positive and negative mode of the instrument. Mass spectra of 23 organic compounds have been measured. Features related to common contaminants for these analyses are also described.

Section 4 is about the different methods found in order to facilitate the COSIMA data interpretation. Elaborating on our

database, we will show that detection of polymeric structures, distinction between samples poor and rich in nitrogen, saturated aliphatic and aromatic chains, should be feasible.

Section 5 presents the Bayesian Positive Source Separation approach and its different utilities to interpret a large amount of mass spectra.

2. Methods

In the following section, the Time-Of-Flight Secondary Ion Mass Spectrometry principle, the description of the instrument used and its representativeness towards the flight model will be presented as well as a description of a typical analysis and the list of the sample.

2.1. TOF-SIMS technique principle

Time-Of-Flight Secondary Ion Mass Spectrometry (TOF-SIMS) is a surface sensitive analytical method used to analyze the very outermost surface of a solid sample. The TOF-SIMS technique principle consists in focusing an ion beam (called primary ion) on the solid surface of a sample that will gently bombard it in order to extract chemical species (called secondary particles) characteristic of the sample. When striking the solid surface, the primary ion transfers energy and momentum to a first atom that which will also interact with its surrounding and so on, generating a so-called collision cascade. Ions and neutrals (atoms and molecules) are by then sputtered from the outermost mono-atomic layers of the sample. Only a small fraction ($\sim 1\%$) of these secondary particles is charged (positively and negatively) when escaping the surface. These sputtered secondary ions are then accelerated to the same electrical field into a flight tube (with a well-defined length) to sort them by their mass, and they finally reach the mass analyzer detector. The time each ion of mass m and charge z takes to reach the detector is correlated to its mass m to charge ratio. Number of

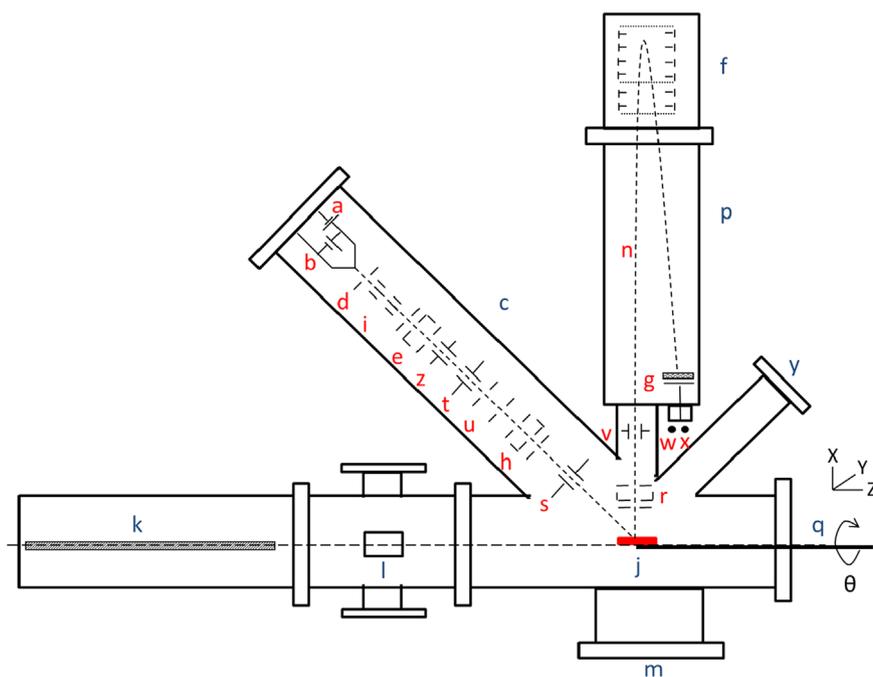


Fig. 2. Functional blocks of Orléans TOF-SIMS instrument (OM) with (a) ion emitter, (b) heating, (c) Primary Ions Beam Subsystem (PIBS), (d) extraction electrode, (e and h) lens, (f) 2 stages ion reflector, (g) microchannel plate (detector), (i and s) primary deflection plates, (j) target holder shelf, (k) transfer target holder system, (l) introduction chamber of target holder, (m) high vacuum connection zone, (n) secondary ion beam, (p) flight tube, (q) manipulator X, Y, Z, θ system, (r) secondary lens, (t and u) chopper/buncher, (v) secondary deflection plates, (w) analog output, (x) digital output, (y) camera, (z) analog to digital switch.

counts vs. m/z histogram gives mass spectrum of atoms or molecules constituent of the surface analyzed. TOF-SIMS leads almost exclusively to the formation of singly ionized species. Consequently the mass to charge ratio can almost always be read as the mass of the ion.

2.2. Description of the calibration TOF-SIMS and its representativeness towards the flight model (COSIMA)

There are two ground models of COSIMA flight model: (i) an exact replica of flight COSIMA instrument, the Reference Model (hereafter RM) located at Max Planck Institute of Göttingen (Germany) and, (ii) a conventional homemade instrument of TOF-SIMS especially build up to test and to qualify the primary ion beam subsystem (PIBS) of COSIMA experiment, named OM for Orléans Model and located at LPC2E (Orléans, France).

The COSIMA flight model (XM), RM and OM instruments have the same PIBS system, which is one of the key elements of a TOF-SIMS instrument to determine its mass resolution power. The OM TOF-SIMS has similar instrumental characteristics than COSIMA, and therefore can mimic COSIMA flight model and COSIMA Reference Model. OM instrument is therefore of good support for all the calibration campaigns that RM could perform.

For this study, the measurements have been performed with the OM.

2.2.1. Description of the Orleans model (OM) TOF-SIMS

The schematic view of the Orléans TOF-SIMS instrument (OM) functional elements is shown in Fig. 2. Several subunits are necessary to perform analysis of samples:

- a sample holder to put the sample at the optical axis of the lenses,
- a primary ion ($^{115}\text{In}^+$) emitter to ionize the sample,
- a system to provide a pulsed primary ion beam,
- an ion optic to provide a focused primary ion beam on the sample,
- an ion optic to focus secondary ions (sputtered from the sample) on detector,
- a time-of-flight mass analyzer to separate ions by their mass to charge ratios m/z ,
- an ion reflector to tune the initial ion energy compensation and minimize time dispersion of ions of the same mass at the detector,
- a micro-channel plate as detector of secondary ions.

The OM instrument is composed of two distinct areas: the sample introduction and the analysis area. They are hermetically divided, which allows to introduce and remove sample at atmospheric pressure without altering the high vacuum condition ($P < 1.10^{-8}$ mbar) highly recommended for the PIBS and detector (located in the analysis area). Moreover, to optimize the time for

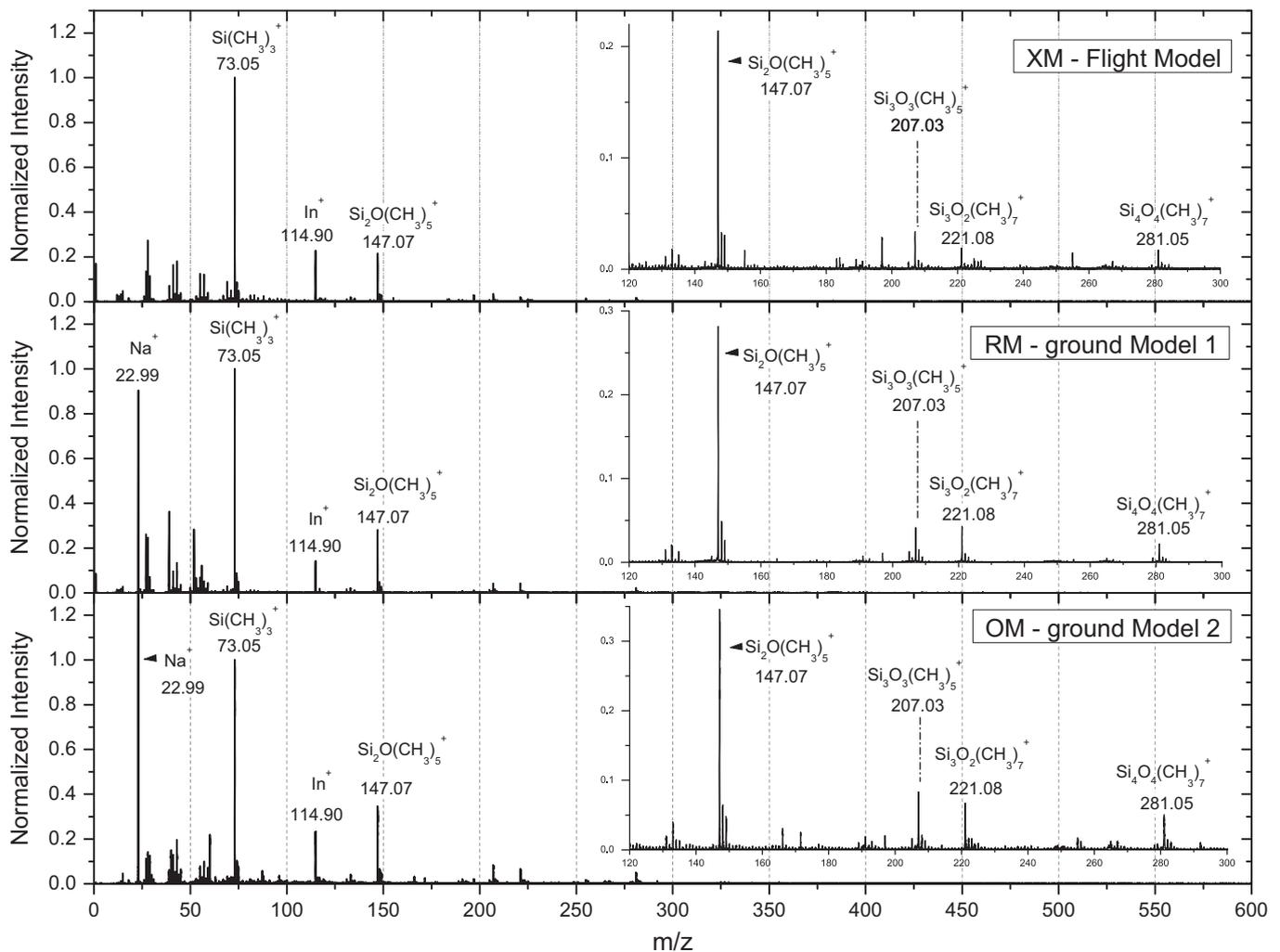


Fig. 3. Positive ion mode mass spectra of polydimethylsiloxane (PDMS) on gold blank target measured with the flight model (XM), the ground Reference Model (RM), and the Orleans Model (OM).

the introduction process several vacuum units are used. The introduction chamber is connected to a dry primary vacuum pump (Scroll pump, GVSP30) coupled to a turbo vacuum pump (Pfeiffer Vacuum, D35614-Asslar), bringing the pressure from atmospheric to $\sim 10^{-6}$ mbar. The analysis chamber is held at pressure around 10^{-8} to 10^{-9} mbar by two cryogenic vacuum pumps (CT8, Brooks Automation).

The PIBS is the LPC2E (Orléans, France) hardware contribution to COSIMA/Rosetta experiment, for which its specifications and general concept have been presented in Kissel et al. (2007). The primary ion beam is generated by a liquid metal ion source (LMIS) of indium ($^{115}\text{In}^+$) beam pulsed within 2–3 ns of about 2000 ions per pulse at a repetition rate of up to 1000 pulses per second. The focal spot of the beam on the target is of elliptical shape of around 50 μm major axis.

As mentioned above with a TOF mass analyzer there is a proportional relation between the root square of m/z ratio of the particles and their time of flight. The flight distance is fixed, but particles of same mass can arrive at slightly different times due to their initial energy, their temporal and space dispersions, which contribute to a degradation of the mass resolving power ($m/\Delta m$) of the analyzer. A two stages ion reflector device offsets these time-related dispersions. Thanks to this reflectron, the OM TOF-SIMS instrument has a mass resolving power of 1400 at the full width at half maximum (FWHM) of peak at $m/z=100$ u/e. Reflected secondary particles travel the field-free region before reaching the two stages micro-channel plate (Hamamatsu, model F4292-06) detector.

2.2.2. Representativeness of the OM towards the flight model (COSIMA)

The OM and the RM have been built to have similar instrumental characteristics than the COSIMA flight model (XM) and so to give similar results *i.e.* similar mass spectrum. In order to check the representativeness of the OM data towards those obtained by the flight model, three mass spectra containing features of a same compound have been compared (Fig. 3). Actually there is few cross-linked data between the XM, the RM and the OM but a lot between the RM and the OM. The RM and the OM were then used for the calibration of the XM. It is not an issue because measurement with the RM and the XM leads to the exact same result. The only difference comes from their internal own contaminants. Currently the only organic molecule that has been measured with these three instruments is the polydimethylsiloxane (PDMS). The PDMS, a silicon oil, is a very common contaminant for TOF-SIMS analysis. For more details on COSIMA contamination see Hilchenbach et al. (in preparation).

This comparison will only be qualitative and not quantitative. First we do not know the concentration of the PDMS on each target and it is most likely not the same because this contaminant tends to migrate easily and quickly inside the instrument (Leggett et al., 1992). Moreover, even if blank gold target and indium as primary ion beam have been used for each analysis, the top layered surface properties of the targets are not controlled for these experiments. Thus different matrix effects can be expected.

Two forms of PDMS ($\text{CH}_3\text{-}[\text{Si}(\text{CH}_3)_2\text{O}]_n\text{-Si}(\text{CH}_3)_3$) and ($\text{CH}_3[\text{Si}(\text{CH}_3)_2\text{O}]_n\text{-Si}_2\text{O}_2(\text{CH}_3)_3$) are found inside the instruments. The more characteristic and intense peaks of these contaminants are summarized in Table 1 (Briggs et al., 1989). As shown in Fig. 3, all these features are observed at different relative intensities for each mass spectrum measured. We also detected others characteristic peaks at m/z 27.97 and 43.00 u/e relative to Si^+ , and SiCH_3^+ ions respectively. From a qualitative point of view, we found the same PDMS features for each instrument as expected. So it can be

Table 1

List of some specific fragments obtained during the TOF-SIMS analysis of two PDMS forms.

Mother molecule	Fragment	m/z
$\text{Si}(\text{CH}_3)(\text{OSi}(\text{CH}_3)_2)_n$	$\text{Si}(\text{CH}_3)_3^+$	73.05
	$\text{Si}_2\text{O}(\text{CH}_3)_5^+$	147.07
	$\text{Si}_3\text{O}_2(\text{CH}_3)_7^+$	221.08
$\text{Si}_2\text{O}_2(\text{CH}_3)_3(\text{OSi}(\text{CH}_3)_2)_n$	$\text{Si}_3\text{O}_3(\text{CH}_3)_5^+$	207.03
	$\text{Si}_4\text{O}_4(\text{CH}_3)_7^+$	281.05

concluded that analysis of a same compound leads to the same fingerprint with the OM, the RM and the XM instruments.

2.3. Limitations to the data interpretation (from the peak centroid to the molecular formula assignment)

Two parameters limit our data interpretation: the mass resolving power (or mass resolution) of the instrument and the precision on the mass to charge ratio measurement due to the mass calibration.

The mass resolution is the ability of an instrument to separate peaks with very close mass to charge ratio. Indeed COSIMA has the best mass resolving power ever achieved in space for a Time-Of-Flight mass spectrometer, making currently Rosetta the space mission hosting the two space borne MS with highest mass resolution. But, as slightly mentioned in the first chapter, compared to the state-of-the-art analytical instrument on Earth, space borne *in situ* instrument are not able to offer same abilities. One has also to keep in mind that Rosetta's mission concept started in the nineties. Thus, COSIMA analytical capacities are not sufficient to avoid the overlap of peaks with very close mass to charge ratio as for example ^{12}CH ($m/z=13.00782$ u/e) and ^{13}C ($m/z=13.00335$ u/e). We can only give the abundance of ^{13}C bearing species to ^{12}C bearing species with an error of 50% at least.

Another important parameter to take into account for the data interpretation is the precision of peak centroid measurement. The precision is relative to the error made on the determination of the mass to charge ratio. This parameter depends mainly on the mass calibration. For this study the mass calibration has been performed thanks to the knowledge of the time of flight of specific ions. In our case as the samples were known, these ions were chosen to be mainly coming from our sample. If the sample is unknown the peaks related to common features found in our mass spectra can also be chosen, as example the peaks related to the metallic target (Au^+) and the primary ion beam (In^+) or also to some contamination peaks like Na^+ . If none of these peaks are detected then unfortunately some hypothesis about the nature of the peaks and so the ion should be done. For the molecular formula assignment, it has been observed that more confidence should be given to the formula related to molecular ions compared to radical ions, which are less stable and so less abundant.

In this context it is hardly difficult to assign a unique molecular formula for each peak in the entire mass spectrum. The common Van Krevelen diagram procedure currently use in the MS field is unusable. The Kendrick mass defect and the mass defect vs. exact mass diagram give both limited results depending on the precision of the Peak's centroid determination. So we had to find other methods more pertinent for non-high resolution mass spectrum.

2.4. Description of a typical sample analysis

Before any sample deposition, the gold targets were rigorously cleaned in ultrasonic baths of n-hexane and then acetone to

remove any kind of contamination. For each samples, three sets of high statistics of positive and negative mass spectra were taken in order to get molecular structure information in the framework of COSIMA spectra interpretation database. The first set acquired before sample deposition monitors the level of the surface pollution. For each sample, tiny amount was then gently crushed with agate pestle on the cleaned targets. The second and third series of mass spectra were then measured: sample spectra and contamination control spectra on the gold foil outside the sample area.

2.5. List and choice of samples

The samples and their structures are summarized in Table 2 and in Fig. 4. These compounds have been analyzed for different purposes. The first aim was to measure compounds which might

be present in comets and/or which have a high astrobiological or planetological interest. Therefore we have focused our choice on molecules that have been tentatively detected during the *in situ* measurement of PUMA on-board the VEGA-1 mission. We have also chosen molecules which have been detected from the analysis of cometary organic proxies (natural samples like carbonaceous chondrites as well as molecules coming from the synthesis in laboratory of cometary or interstellar organic analogs). The objective of these analyses is to establish a database of reference spectrum. Then as it is quite obvious that such a database will never be exhaustive enough, our work have also been focus on the analysis of molecules, which have no direct link with comets but which belong to the same chemical family than the expected compounds in cometary grains. Their TOF-SIMS measurements will help us to determine the fingerprint of the different chemical families.

Table 2

List of the analyzed compounds.

Molecule		Molecular formulae	Exact mass (amu)	Supplier	Purity (%)
Eicosane	Cos	C ₂₀ H ₄₂	282.3287	Aldrich	99
Tétracosane	Tcos	C ₂₄ H ₅₀	330.3287	Aldrich	99
Octacosane	Ocos	C ₂₈ H ₅₈	394.4539	Aldrich	99
Hexatriacontane	HTC	C ₃₆ H ₇₄	506.5791	Aldrich	99
Anthracene	Ant	C ₁₄ H ₁₀	178.0783	Prolabo	97
Pyrene	Pyr	C ₁₆ H ₁₀	202.0783	Prolabo	95
Coronene	Cor	C ₂₄ H ₁₂	300.0939	Fluka	97
Imidazole	Imi	C ₃ H ₄ N ₂	68.0374	Avocado	99
2-Aminopyrimidine	2amiP	C ₄ H ₅ N ₃	95.0483	AcrosOrganics	99
4-Aminopyrimidine	4amiP	C ₄ H ₅ N ₃	95.0483	Alfa Aesar	98
Adenine	A	C ₅ H ₅ N ₅	135.0545	Aldrich	99
Hexamethylenetetramine	HMT	C ₆ H ₁₂ N ₄	140.1062	Flukabiochemika	> 99.5
Glycolic acid	GA	C ₂ H ₄ O ₃	76.0160	AcrosOrganics	99
Benzoic acid	BA	C ₇ H ₆ O ₂	122.0368	Aldrich	> 99.5
Stearic acid	SA	C ₁₈ H ₃₆ O ₂	284.2715	Merck	97
Polyoxymethylene	POM	HO-[CH ₂ O] _n H (8 < n < 100)		Prolabo	> 99
Cytosine	Cyt	C ₄ H ₅ N ₃ O	111.0433	Alfa Aesar	98
Isocytosine	ICyt	C ₄ H ₅ N ₃ O	111.0433	AcrosOrganics	99
Uracile	U	C ₄ H ₄ N ₂ O ₂	112.0273	AcrosOrganics	99
Thymine	T	C ₅ H ₆ N ₂ O ₂	126.0429	Fluka	97
Hypoxanthine	H	C ₅ H ₄ N ₄ O	136.0385	Aldrich	> 99
Guanine	G	C ₅ H ₅ N ₅ O	151.0494	Aldrich	98
Xanthine	X	C ₅ H ₄ N ₄ O ₂	152.0334	Aldrich	> 99

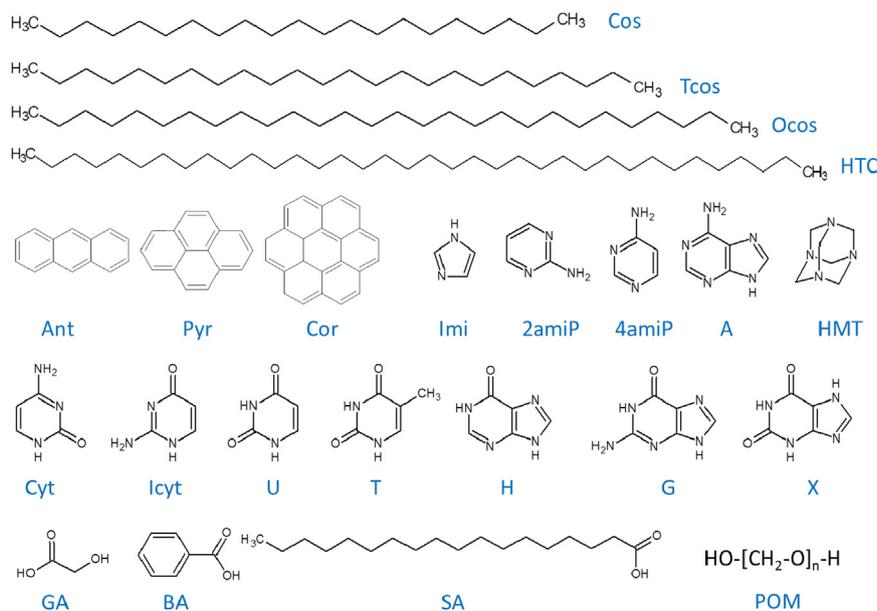


Fig. 4. Structure of the analyzed compounds.

This study has been focused on the analysis of alkanes, polyaromatic hydrocarbons (PAHs), carboxylic acids, nucleobases and other N-heterocyclic compounds. The following paragraph explains the choice of these families.

Many attempts were made to demonstrate the presence of PAHs in comets. COSIMA may raise the ambiguity of their presence in the grains ejected by 67P/CG. The first report was made by Baas et al. (1986) with the detection of a weak infrared feature at 3.28 μm in the coma of 1P/Halley after its perihelion, at heliocentric distances of 1.6 AU and 2.0 AU. Three Channel Spectrometer on-board the Vega-2 probe took spectra in which Moreels et al. (1994) and later Clairemidi et al. (2004) tentatively attributed respectively specific emissions bands to the fluorescence of phenanthrene and pyrene. Nevertheless their extraterrestrial origin was discussed. The presence of PAHs has also been suggested in the cometary grains collected by the Stardust spacecraft. However their extraterrestrial origins have been a matter of controversy (Sandford et al., 2006, 2010; Spencer and Zare, 2007). On the other hand, PAHs with up to 7 rings and their alkyl derivatives are detected in carbonaceous chondrites (Sephton et al., 2004). They are also found in extraterrestrial interplanetary dust particles (Clemett et al., 1993) and in carbonaceous Antarctic Micrometeorites (Clemett et al., 1998). Another interesting family is the carboxylic acids family. Indeed numerous analysis of the soluble matter of the carbonaceous chondrites like Murchison, Orgueil or the Tagish lake meteorite have highlighted that this chemical family is quite abundant in these meteorites (Botta and Bada, 2002; Herd et al., 2011; Pizzarello et al., 2001; Sephton, 2002). Moreover some of them are synthesized during the photolysis of cometary/interstellar ice analogs (Despois and Cottin, 2005). Among the N-heterocyclic compounds, our choice has been focused on the nucleobases. These molecules, elementary building block of DNA and RNA, which have obviously a high astrobiological interest, have been detected in the soluble matter of numerous carbonaceous chondrites since the 60s (Hayatsu, 1964; Hayatsu et al., 1968, 1975; Stoks and Schwartz, 1979, 1981; Van der Velden and Schwartz, 1977). These analyses have highlighted the presence of adenine, guanine hypoxanthine, xanthine and uracil. Nonetheless their extraterrestrial origin was discussed. The ^{13}C isotopic analysis on Murchison made by Martins et al. (2008) and the recent experiment on eleven meteorites made by Callahan et al. (2011) have also resulted to the identification of nucleobases and some derivatives, which are consistent with an extraterrestrial origin. Nitrogen-heterocyclic compounds have also been suggested as being part of the Halley grains with PUMA-1 mass spectrometer on-board Vega 1 spacecraft. The results of CIDA on-board also tentatively proposed the presence of molecules containing nitrogen (Kissel et al., 2004). The last chemical family is the alkanes. Indeed light alkanes have been detected in various astrophysical medium: molecular cloud, interstellar medium, comets. Nevertheless the presence of heavy alkanes, which have been detected in several carbonaceous chondrites, probably come from terrestrial contamination (Sephton, 2002). The conditions on the early earth were not favorable to the accumulation of heavy acyclic hydrocarbons. Therefore an exogenous source as the comets has been considered (Marcano et al., 2003).

3. The mass spectra database

The aim of this paper is to facilitate the interpretation of COSIMA mass spectra and more especially of their organic part. One of the COSIMA ultimate purposes is to highlight with no ambiguity the presence of organic compounds in the cometary grains collected and to identify them. A prerequisite for the identification of molecules or family of molecules in a mass spectrum of an unknown sample is the knowledge of their specific features. Therefore this work has started with the elaboration of a library of mass spectra in order to determine the fingerprint of

these compounds and of their chemical families. The determination of the sensitivity factors (i.e. factors which enables the quantification) will not be addressed in this paper. To perform a quantitative analysis, a perfect knowledge of the surface state of the sample is needed. The aim of our measurements was to obtain the best spectral feature and not to study exhaustively the sample deposition since we cannot fairly mimic the organics at the surface of cometary grains.

The following section presents first the common species, which are found in the reference mass spectra and then the specific features related to each family (alkanes, PAH, N-heterocyclic compounds, carboxylic acids).

Table 3 summarizes for each analysis the nature of the highest peak and it also informs about the intensity of the molecular or quasi-molecular peaks. All the values for the number of counts (including the number of count for the peak's maximum) are normalized to 1 min of TOF-SIMS analysis. During operation with the XM, TOF SIMS measurement will mainly last around 1 min in order to localize the grain position (several spectra on the same position in order to have better statistics are obviously also planned).

It should be mentioned that as the number of counts depends on the sample deposition, it is very hazardous to derive from this table and to apply to other analysis any kind of quantitative information.

3.1. Typical features in a reference mass spectrum

As mentioned before, TOF-SIMS is a technique sensitive to surface contamination. The mass spectra measured for this study are no exception to that rule. Some contaminants peaks are often detected.

This section aims to present these species and some common features found in our mass spectra in order to that the reader will not be confused by the presence of these peaks in the following figures.

3.1.1. Typical features in the positive ion mode

Generally alkali ions as, Na^+ and K^+ ions are often observed and respectively localized at $m/z=22.99$ and 38.96 u/e. Presence of phthalates and azelates have also been sometimes detected. These internal contaminants of the OM instrument lead to the presence of small hydrocarbons peaks as C_2H_3^+ , C_2H_5^+ , C_3H_5^+ , C_3H_7^+ , C_4H_7^+ and C_4H_9^+ , respectively at $m/z=27.02$, 29.04 , 41.04 , 43.05 , 55.05 and 57.07 u/e. These peaks can be a source of problems during the data interpretation. Indeed numerous organic compounds produce also these ions when they are ionized with the indium ion beam. The presence of peaks related to PDMS is also often observed. Its characteristic peaks have been described in paragraph 2, Section 2.2. Peaks associated to the indium (from the primary ion beam) and gold (from the target) ions are also observed.

3.1.2. Typical features in the negative ion mode

The major peak in the negative ion mode is most of the time related to the presence of H^- ion. But other peaks are systematically detected as O^- , OH^- , CN^- and CNO^- ions respectively at $m/z=15.99$, 17.00 , 26.00 and 42.00 u/e. Their relative intensities vary from one reference spectrum to another as a function of the target pollution. The C^- ($m/z=12.00$ u/e), CH^- ($m/z=13.01$ u/e), C_2^- ($m/z=24.00$ u/e) and C_2H^- ($m/z=25.01$ u/e) ions, as well as the chloride isotopes ($m/z=34.97$ u/e and $m/z=36.96$ u/e) are classically detected. The presence of sulfur ions as SO_3^- ($m/z=79.96$ u/e) and HSO_4^- ($m/z=96.96$ u/e) can also be highlighted in the majority of the reference mass spectra.

Table 3

Summary for each sample analysis of the total number of counts, the nature of the highest peak and it informs also about the intensity of the molecular or quasi-molecular peaks. All the value for the number of counts (including the columns “total number of counts” and “peak’s maximum”) are normalized to 1 min. X means that no molecular (M^+) or quasi molecular ion ($M \pm xH^+$) have been detected. [–] means that the molecular or quasi molecular ion is the peak with the highest intensity in the mass spectrum.

Molecule	Mode	Total number of counts	Highest peak			Quasi-molecular ion (if not major peak)		
			Ion	<i>m/z</i>	Peak's maximum (counts)	Ion	<i>m/z</i>	Peak's maximum (counts)
Eicosane	Positive	3403	$C_{20}H_{42}^+$	29.04	719	$[M+CH]^+$	295.34	35
	Negative	2077	H^-	1.01	646	X	X	X
Tetracosane	Positive	785	$C_{24}H_{50}^+$	43.06	104	$[M-2H]^+$	336.38	1
	Negative	5178	H^-	1.01	1089	X	X	X
Octacosane	Positive	1381	$C_{28}H_{58}^+$	43.06	324	$[M-2H]^+$	392.46	3
	Negative	5120	H^-	1.01	1318	X	X	X
Hexatriacontane	Positive	3746	$C_{36}H_{74}^+$	43.06	866	$[M-4H]^+$	502.56	4
	Negative	7768	H^-	1.01	1409	X	X	X
Pyrene	Positive	5607	$[M+2H]^+$	204.09	432	[–]	[–]	[–]
	Negative	9197	H^-	1.01	2320	X	X	X
Anthracene	Positive	1002	$C_{14}H_{10}^+$	57.07	110	$[M+2H]^+$	180.09	38
	Negative	1416	H^-	1.01	4359	X	X	X
Coronene	Positive	2577	$C_{24}H_{18}^+$	29.04	239	M^+	300.09	182
	Negative	12,863	H^-	1.01	3401	X	X	X
Glycolic acid	Positive	9873	CH_3O^+	31.02	359	X	X	X
	Negative	6556	H^-	1.01	2239	$[M-H]^-$	75.01	70
Benzoic acid	Positive	2509	$C_7H_6O_2^+$	43.05	218	X	X	X
	Negative	3339	O^-	15.99	626	$[M-H]^-$	121.03	29
Stearic acid	Positive	1075	$C_{18}H_{34}O_2^+$	43.05	146	$[M+H]^+$	285.28	30
	Negative	1752	O^-	15.99	431	$[M-H]^-$	283.22	1
POM	Positive	1932	$[(H_2CO)_2+H]^+$	61.03	243	X	X	X
	Negative	2066	H^-	1.01	451	X	X	X
Imidazole	Positive	1964	$[M+H]^+$	69.05	243	[–]	[–]	[–]
	Negative	5854	CN^-	26	940	$[M-H]^-$	67.03	79
2-Aminopyrimidine	Positive	2000	$C_4H_5N^+$	29.04	229	$[M+H]^+$	96.06	84
	Negative	7877	CN^-	26	1394	$[M-H]^-$	94.04	29
4-Aminopyrimidine	Positive	2296	$C_4H_5N^+$	43.04	213	$[M+H]^+$	96.06	185
	Negative	9148	CN^-	26	1740	M^-	95.03	475
Adenine	Positive	1940	$[M+H]^+$	136.07	214	[–]	[–]	[–]
	Negative	5999	CN^-	26	978	$[M-H]^-$	134.05	102
HMT	Positive	30,366	$C_2H_4N^+$	42.03	4464	$[M+H]^+$	141.11	626
	Negative	29,799	H^-	1.01	8524	X	X	X
Cytosine	Positive	16,813	$[M+H]^+$	112.05	1283	[–]	[–]	[–]
	Negative	2297	O^-	26	445	$[M-H]^-$	110.03	13
Isocytosine	Positive	9798	$[M+H]^+$	112.05	1713	[–]	[–]	[–]
	Negative	4831	H^-	1.01	1288	$[M-H]^-$	110.03	2
Uracil	Positive	1437	$[M+H]^+$	113.04	124	[–]	[–]	[–]
	Negative	4125	H^-	1.01	1860	$[M-H]^-$	111.01	19
Thymine	Positive	6498	$[M+H]^+$	127.05	601	[–]	[–]	[–]
	Negative	5708	H^-	1.01	242	$[M-H]^-$	125.04	12
Hypoxanthine	Positive	6641	$[M+H]^+$	137.05	589	[–]	[–]	[–]
	Negative	11,608	H^-	1.01	3720	$[M-H]^-$	135.03	8
Guanine	Positive	5240	$[M+H]^+$	152.06	435	[–]	[–]	[–]
	Negative	2255	CN^-	26	413	$[M-H]^-$	150.04	6
Xanthine	Positive	5794	$[M+H]^+$	153.04	847	[–]	[–]	[–]
	Negative	4125	H^-	1.01	657	$[M-H]^-$	151.03	5

3.2. Family fingerprint in the positive ion mode of the instrument

3.2.1. Alkanes

Four alkanes have been measured: eicosane ($C_{20}H_{42}$), tetracosane ($C_{24}H_{50}$), octacosane ($C_{28}H_{58}$) and hexatriacontane ($C_{36}H_{74}$), all of them being made of a long aliphatic saturated chain.

Common characteristics have been found in the positive ion mode mass spectrum of these alkanes (Fig. 5).

First, dominant peaks measured around the molecular weight of the alkane can be attributed to $[M-xH]^+$ with M the mass of the analyzed compound and xH the number of hydrogen atom loss. This series of peaks due to a lack of hydrogen(s) is presented

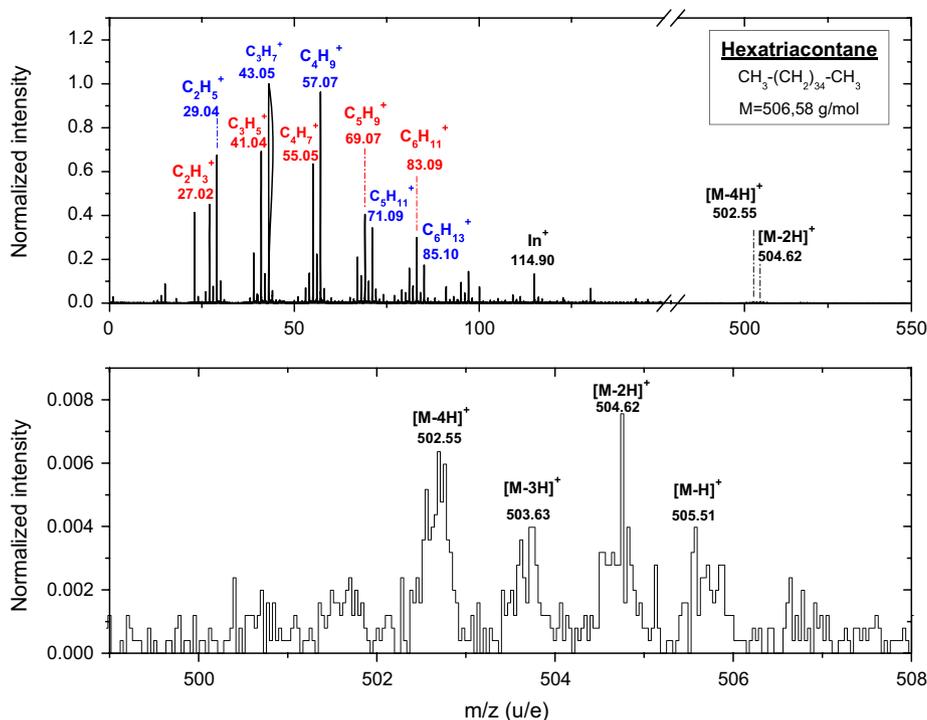


Fig. 5. Positive ion mode mass spectrum of hexatriacontane. The alternation of $[C_nH_{2n-1}]^+$ and $[C_nH_{2n+1}]^+$ ions within a pattern is represented in red and blue respectively. The intensity is normalized to the major peak ($m/z=43.05$ u/e). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

in the spectrum of the hexatriacontane (Formula weight (FW)=506.58 u/e): $[M-H]^+$ at $m/z=505.51$ u/e, $[M-2H]^+$ at $m/z=504.62$ u/e, $[M-3H]^+$ at $m/z=503.63$ u/e and $[M-4H]^+$ at $m/z=502.59$ u/e. This hydrogen loss would come from an unimolecular dissociation in the vacuum because of an internal excess of energy (Delcorte and Bertrand, 1998). The peak presenting the major intensity is either $[M-2H]^+$ or $[M-4H]^+$. Because of its high number of carbons, isotopical ratios have been calculated in order to determine the potential contribution of ^{13}C at high masses. Nevertheless, ^{13}C isotopes cannot completely explain the presence of the peaks $[M-H]^+$ and $[M-3H]^+$, meaning that they are part of the alkane's features.

A second characteristic is found in the low range of the positive ion mode mass spectrum (below 100 u/e), also called the hydrocarbons envelope. This domain is defined by hydrocarbon fragments which, in the case of alkane, forms well defined pattern dominated by the alternation of $[C_nH_{2n-1}]^+$ and $[C_nH_{2n+1}]^+$ ions, as observed in Fig. 5 in red and blue respectively. These peaks result from the fragmentation of the alkane chain and a maximum is usually observed for $n=3$.

The last feature comes from a dominant mass separation of 14.02 u/e between the peaks. Because of its long alkane chain, this repeated mass separation can easily be attributed to the fragment CH_2 (see Section 4.3.2).

3.2.2. Polyaromatic hydrocarbons (PAHs)

Polycyclic aromatic Hydrocarbons (PAHs) are benzene-based multi rings molecules. Three of them were analyzed for this study: a catacondensed PAH, anthracene ($C_{14}H_{10}$), and two pericondensed ones, pyrene ($C_{16}H_{10}$) and coronene ($C_{24}H_{12}$). The positive mass spectra of these PAHs exhibit common characteristic features.

For the three PAHs analyzed, the molecular ion $[M]^+$ (with M the mass of the analyzed compound) is detected by TOF-SIMS in the positive ion mode: at $m/z=178.08$ u/e for anthracene at

$m/z=202.13$ u/e for pyrene and at $m/z=300.08$ u/e for coronene. In the high mass range the most intense peak is the $[M]^+$ ion for coronene, and the secondary ion relative to $[M+2H]^+$, at $m/z=180.08$ u/e and $m/z=204.17$ u/e respectively for anthracene and pyrene. Toporski and Steele (2004) have obtained similar results for pyrene dropped on silicon wafer. For each compound, around their intense peak in the mass range of the intact molecule, a cluster of peaks is also observed relative to successive addition or loss of hydrogen atom on the molecule (series $[M \pm xH]^+$). This series of peaks is presented in the positive mass spectrum of coronene (FW=300.35 u) in Fig. 6.

In this figure, it can also be noticed other series of clusters peaks surrounding the molecular peak series. Similar features were observed by Fister et al. (1995) on their studies of benz(a)anthracene and benzopyrene cast on Si wafer. On each mass spectrum, these series of cluster peaks can be attributed to successive removal or addition of (CH) group, leading to several cluster series of $[M \pm xH-(CH)_y]^+$ ions and $[M \pm xH+(CH)_y]^+$ ions above and below the exact mass.

Au-cationised molecules (cluster series $[M+xH+Au]^+$) are clearly observed for coronene, but are more fainter for anthracene and for pyrene. It can be noticed by the ratio $I_{(M+Au)^+}/I_{(Au)^+}$ that the signal of the gold-cationised coronene is more intense than the gold signal itself, while this trend is different for anthracene and pyrene. Dimers and/or their metallic cationised homologs are clearly detected for anthracene and pyrene compounds. For coronene several cluster peaks of dimer series of $[2M-xH-(CH)_y]^+$ ions resulting of successive removal of (CH) group are dimly observed. On the other hand, due to the mass range chosen for the analysis, no metallic cationised dimers of coronene are detected in the mass spectrum. For anthracene, it is interesting to note that the intensity of dimers cationised molecules $[2M+Au]^+$ is much stronger than the monomer cationised ones $[M+Au]^+$. Table 4 summarizes the ions observed in the higher mass range for the 3 PAH analyzed.

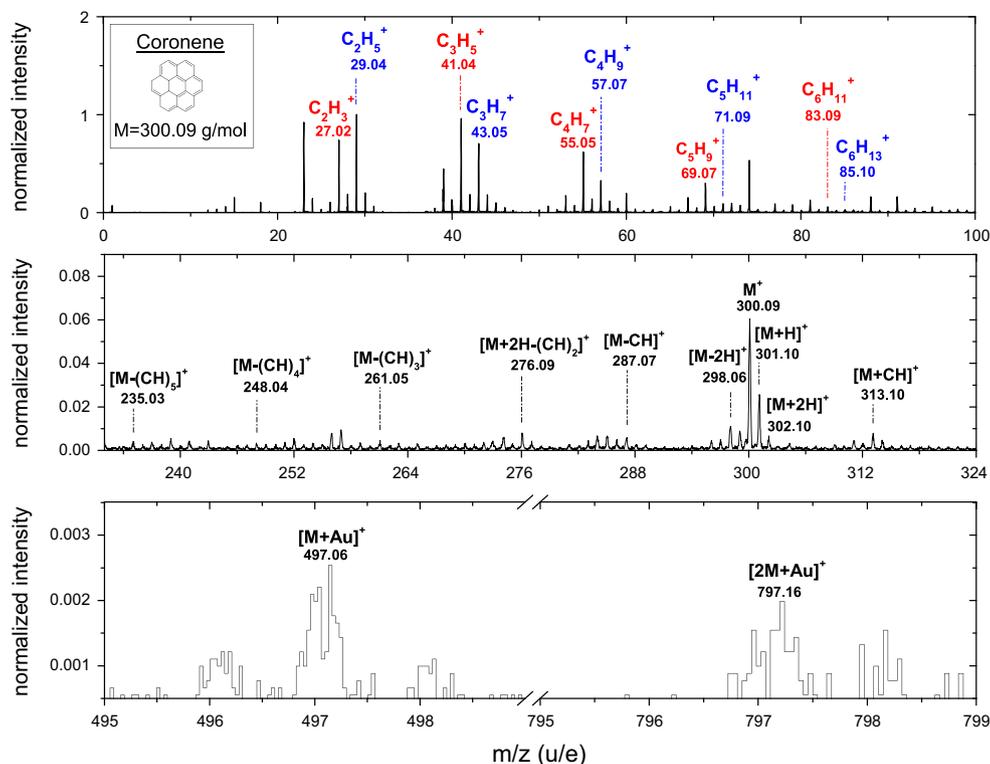


Fig. 6. Positive ion mode mass spectrum of coronene. The intensity is normalized to the major peak ($m/z=29.04$ u/e).

Table 4

Organic and organometallic molecular ions observed in the mass spectra of PAH sample deposited on Au substrate.

Samples deposition on Au	Major molecular ions (in bold the dominant one of the series)	$I(M+Au)/I(M)$	$I(M+Au)/I(Au)$
Anthracene	$[M+2H]^+$ (series $[M \pm xH]^+$) $[M+2H+CH]^+$ (series $[M \pm xH+CH]^+$) $[M-CH]^+$ (series $[M \pm xH-CH]^+$) $[M-(CH)_2]^+$ (or $[M-H-(C_2H)]^+$) (series $[M \pm xH-(CH)_2]^+$) $[M+Au]^+$ (series $[M \pm xH+Au]^+$) $[2M+Au]^+$ (series $[M+xH+Au]^+$)	0.05	0.07
Pyrene	$[M+2H]^+$ (series $[M \pm xH]^+$) $[M+3H+CH]^+$ (series $[M \pm xH+CH]^+$) $[M+4H-CH]^+$ (series $[M \pm xH-CH]^+$) $[M-(CH)_2]^+$ (or $[M-H-(C_2H)]^+$) (series $[M \pm xH-(CH)_2]^+$) $[M-(CH)_3]^+$ (or $[M-2H-(C_3H)]^+$) (series $[M \pm xH-(CH)_3]^+$) $[M+H+Au]^+$ (series $[M \pm xH+Au]^+$) $[2M+2H+Au]^+$ (series $[M+xH+Au]^+$) $[2M]^+$ (series $[2M \pm xH]^+$) $[(2M+4H)-CH]^+$ (series $[2M \pm xH+CH]^+$) $[(2M+4H)+CH]^+$ (series $[2M \pm xH-CH]^+$)	0.33	0.28
Coronene	$[M]^+$ (series $[M \pm xH]^+$) $[M+CH]^+$ and $[M+2H+CH]^+$ (series $[M \pm xH+CH]^+$) $[M-CH]^+$ (series $[M \pm xH+CH]^+$) $[M+Au]^+$ and $[M+H+Au]^+$ $[M+2H-(CH)_2]^+$ (or $[M+H-(C_2H)]^+$) (series $[M \pm xH-(CH)_2]^+$) $[M-(CH)_3]^+$ (or $[M-2H-(C_3H)]^+$) (series $[M \pm xH-(CH)_3]^+$) $[M-(CH)_4]^+$ (or $[M-3H-(C_4H)]^+$) (series $[M \pm xH-(CH)_4]^+$) $[M-(CH)_5]^+$ (or $[M-4H-(C_5H)]^+$) (series $[M \pm xH-(CH)_5]^+$) $[M-2H+(CH)_2]^+$ (series $[M \pm xH+(CH)_2]^+$) $[M-2H+(CH)_3]^+$ (or $[M+(C_3H)]^+$) (series $[M \pm xH+(CH)_3]^+$) It can also be slightly observed some series of dimer molecular peak: series $[2M-xH-(CH)_5]^+$; series $[2M-xH-(CH)_4]^+$; series $[2M-xH-(CH)_3]^+$; series $[2M-xH-(CH)_2]^+$; series $[2M-xH-CH]^+$; series $[2M-xH]^+$	0.10	1.27

3.2.3. N-heterocyclic compounds

N-heterocyclic compounds are molecules made of a closed chain of atoms, one of them at least being a nitrogen atom. The size of ring typically varies from 3 to 6 atoms. Several rings can be

fused together and they can contain a large variety of substituent. So the N-heterocyclic compounds are a wide class of molecules.

For this study, we have focused the analysis on nucleobases and other N-heterocyclic molecules. We have measured unsaturated

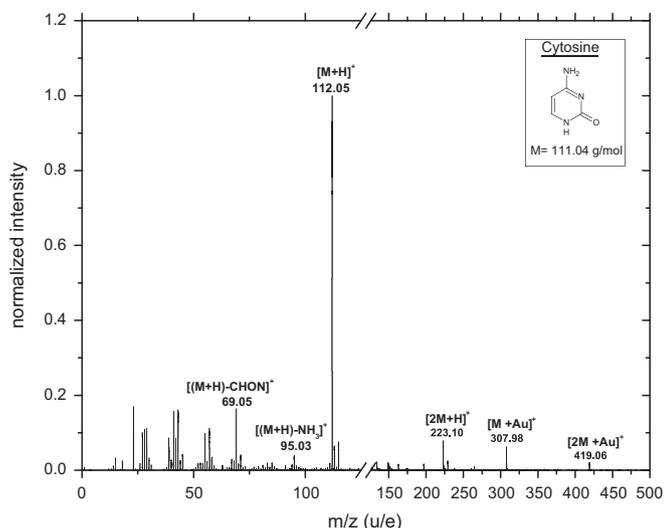


Fig. 7. Positive ion mode mass spectrum of cytosine. The intensity is normalized to the major peak ($m/z=112.06$ u/e). For graphical purpose, there is a break in the scale between 125 and 126 u/e.

N-heterocyclic compounds: some pyrimidine and purine derivatives, the imidazole and one saturated N-heterocyclic compound the hexamethylenetetramine (HMT). The general behavior of these compounds towards a TOF-SIMS analysis will be first discussed and then the specificities of some subclasses will be presented.

A typical positive mass spectrum of the N-heterocyclic compounds is presented in Fig. 7. For all analyzed molecule, a quasi-molecular ion, $[M+H]^+$ (the analyzed molecule (M) enriched with a proton (H)) is observed and easily identified. For almost all the unsaturated compounds, this quasi-molecular ion is the major peak of the mass spectrum. These high intensities can be explained by the stability of this peak induced by the resonance effect of the ring. For almost all the N-heterocyclic compounds, we can also identified peaks related to dimers $[2M+H]^+$ and to organo-metallic clusters $[M+Au]^+$ and or $[2M+Au]^+$ (see Table 5).

It can be also mentioned that as in Varmuza et al. (1999) some alkali clusters of the quasi-molecular ion ($[(M+H)+Na]^+$ or $[(M+H)+K]^+$) can be observed.

For the analyzed unsaturated compounds, we can distinguish three subclasses: the imidazole; the pyrimidine and the purine derivatives

The pyrimidine derivatives (cytosine, isocytosine, uracil, thymine, 2- and 4-aminopyrimidine) are six-membered ring compounds with two nitrogen atoms. At least one substituent (amino, carbonyl and/or methyl group) is linked to the pyrimidine ring. Some effects on the fragmentation mechanisms related to the presence of the different functional group substituent are observed. The fragmentation of the quasi-molecular ion leads often to the loss of NH_2/NH_3 , O or HCNO.

The purines derivatives (adenine, guanine, xanthine and hypoxanthine) are made of a pyrimidine derivative fused with an imidazole ring. All the pyrimidine derivatives are part of the analyzed purines. Same kind of fragmentation than for the pyrimidine derivatives can be observed: the loss from the quasi-molecular ion of NH_2/NH_3 , O or HCNO and also of CH_2N_2 . Observations expected because the analyzed pyrimidine derivatives are substructures of purines. Moreover, peaks related to each of the fused rings minus one or two hydrogen have been observed.

The case of the unsaturated N-heterocyclic compound, HMT, will be discussed in details in a following paper.

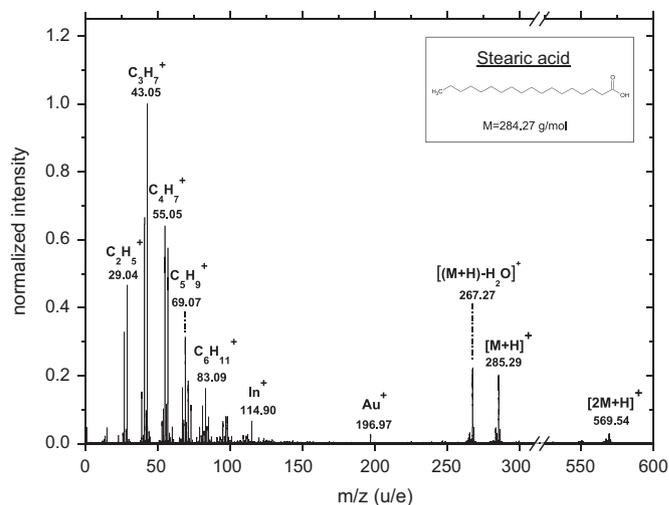


Fig. 8. Positive ion mode mass spectrum of stearic acid. The intensity is normalized to the major peak ($m/z=43.05$ u/e).

3.2.4. Carboxylic acid

Three carboxylic acids have been currently analyzed: a small hydroxycarboxylic acid (glycolic acid), an aromatic acid (benzoic acid), and a fatty acid (stearic acid). From these measurements, the fatty acid and the two others lighter carboxylic acids seem to behave completely differently towards the different mode of the instrument.

The stearic acid, a long aliphatic chain containing a carboxylic function, showed a strong quasi-molecular peak and even some dimer peaks on the positive ion mode mass spectrum (Fig. 8). The loss of a water fragment from the quasi-molecular ion leads also to an important peak. Below 100 u/e, several peaks can be linked to the presence of hydrocarbon fragments, as its mass spectrum presents the characteristic peaks of the alkanes C_nH_{2n+1} and C_nH_{2n-1} . These hydrocarbons fragments stem from the scission of two sp^3 carbon atoms between the numerous CH_2 functions that contain the fatty acid. Spool (2004) has suggested that the initialization of this carbon cleavage start at the chain end of the stearic acid.

On the other hand, the glycolic and benzoic acid positive ion mode mass spectra show just a few peaks, which are characteristic to the analyzed sample. For the benzoic acid, no quasi-molecular ion is observed contrary to Mahoney et al. (2006). At least, the fragment $C_6H_5CO^+$ at mass 105.04 has been identified but at low intensities yield according to the background (Fig. 9). For the glycolic acid, no quasi-molecular ion is observed either. The positive ion mode mass spectrum obtained is very similar to the control spectrum (spectrum of the target without the sample). Only the peak at $m/z=31.02$ u/e could be attributed to the sample. This mass corresponds to the molecular formula H_3CO^+ and can result from the loss of the HOCHO fragment from the sample.

It appeared that light carboxylic acids would not be easy to detect in the positive ion mode of COSIMA. Further analyses of other carboxylic acids are required to see if it is a common trend.

3.2.5. Polyoxymethylene

The positive ion mass spectrum of polyoxymethylene have already been discussed in Le Roy et al. (2012). In this section a brief summary of these previous results will be presented.

POM is the polymer of formaldehyde (H_2CO). As function of the terminal group and of monomer's number, POM has different chemical properties. The sample analyzed in this study is a

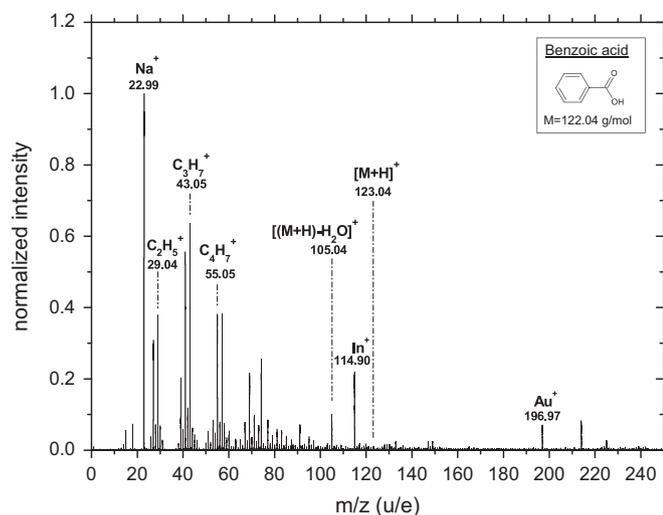


Fig. 9. Positive ion mode mass spectrum of benzoic acid. The intensity is normalized to the major peak ($m/z=22.99$ u/e).

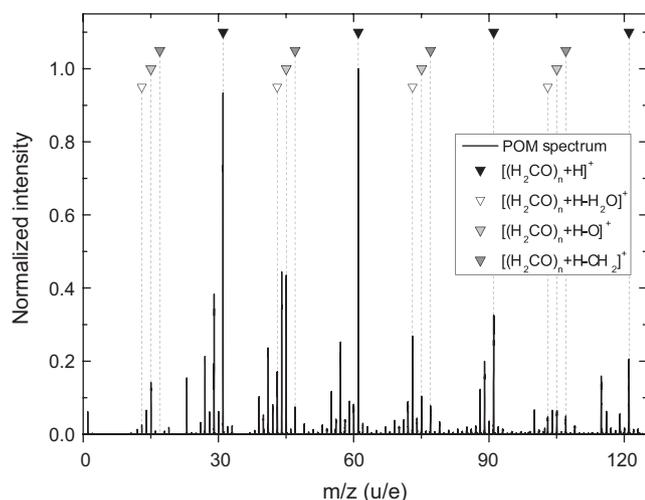


Fig. 10. Positive ion mode mass spectrum of POM. The triangles as well as the vertical gray lines represent the position of the main fragments of POM which can be attributed to $[(H_2CO)_n+H]^+$, $[(H_2CO)_n+H-H_2O]^+$, $[(H_2CO)_n+H-O]^+$ and $[(H_2CO)_n+H-CH_2]^+$.

paraformaldehyde sample (Molecular formula: $HO-[CH_2]_n-H$ with $8 < n < 100$).

Fig. 10 shows a close-up of a positive ion mode mass spectrum for mass to charge ratio lower than 125. Most of the intense peaks can be attributed to fragments of POM, such as $[(H_2CO)_n+H]^+$, $[(H_2CO)_n+H-H_2O]^+$, $[(H_2CO)_n+H-O]^+$ and $[(H_2CO)_n+H-CH_2]^+$. Such fragmentation pattern is observed up to $m/z=600$ but after $m/z=300$ molecular formula attributions are very difficult due to the limited number of secondary ions and also to the broad shape of these peaks.

3.3. Family fingerprint in the negative ion mode of the instrument

3.3.1. Alkanes

The negative ion mode spectra of alkane do not show as many peaks as in the positive ion mode. In particular, no peaks are observable in the high mass range, especially near the exact mass of the alkane, making it impossible to distinguish an alkane from another. It is worth to point out that alkanes are difficult to ionize.

Negative ion mode mass spectra can be characterized in the low mass range by dominance of ions made of hydrogen and/or

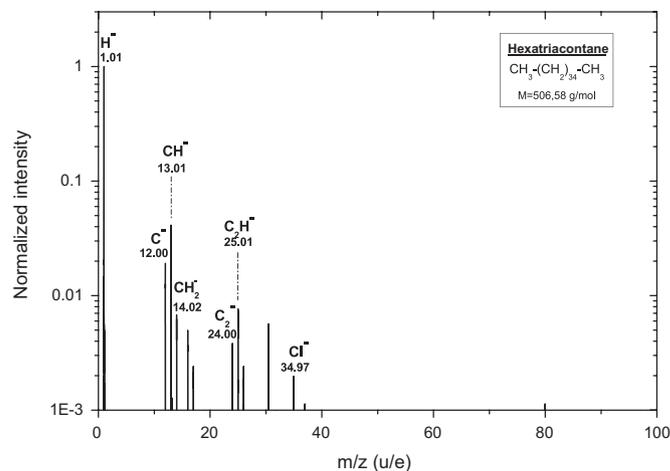


Fig. 11. Negative ion mode mass spectrum of hexatriacontane. The intensity is normalized to the major peak ($m/z=1.01$ u/e).

carbon such as H^- , C^- , CH^- , CH_2^- , C_2^- and C_2H^- (Fig. 11). These peaks come clearly from the fragmentation of the aliphatic saturated chain. Nevertheless some distinctions appear for these ions within the four alkanes. The common feature is that H^- , C^- and CH^- are the ions for which the intensity increases the most from the blank. On the contrary, the increase of CH_2^- , C_2^- and C_2H^- is lower, and seems correlated to the chain length: these ions are faintly present for eicosane ($C_{20}H_{42}$) although they are clearly present for hexatriacontane ($C_{36}H_{74}$) as figured on its negative ion mode mass spectrum (Fig. 11).

3.3.2. Polyaromatic hydrocarbons

For the three PAHs analyzed, no quasi-molecular peaks relative to the different molecules have been observed. Consequently it would be rather difficult to be able to identify a PAH molecule from another one in mass spectrum of an unknown sample by using only the negative ion mode.

Nonetheless characteristic fingerprints of PAHs that resulted from the fragmentation of molecular or quasi-molecular ion of PAHs under the TOF-SIMS primary ion beam are observed in the lower mass range: H^- , C^- , CH^- , C_2^- and C_2H^- . These peaks are the major peaks of the PAHs negative ion mode mass spectra. It should also be mentioned that series of peaks relative to C_n^- and C_nH^- are observed for all the PAHs analyzed (Fig. 12). For anthracene, these series are detected till $n=6$. For the pyrene, these series are not really obvious after $n=4$. Further analyses should be performed to determine if these series are also observed for other PAHs but they seem to be a very good indicator to highlight the presence of PAHs in the negative ion mode.

3.3.3. N-heterocyclic compounds

For the negative ion mode mass spectra of the N-heterocyclic compounds, the quasi-molecular peak ($[M-H]^-$) is always observed but at rather low intensity compared to the positive ion mode (Fig. 13). For some compounds it can be 50 times lower than the major peak of the spectrum.

The other characteristic peaks of the N-heterocyclic compounds are CN^- or O^- for compounds including oxygen atom at $m/z=26.00$ and 15.99 u/e respectively. These ions have among the most important intensities of the negative mass spectrum Fig. 13. The peak related to CNO^- at $m/z 42.00$ u/e can also give an important contribution to the total amount of counts of the mass spectrum. Nevertheless these peaks are not very specific of the analyzed compounds. Therefore we cannot use only them as indicators in order to identify with no ambiguity the presence of

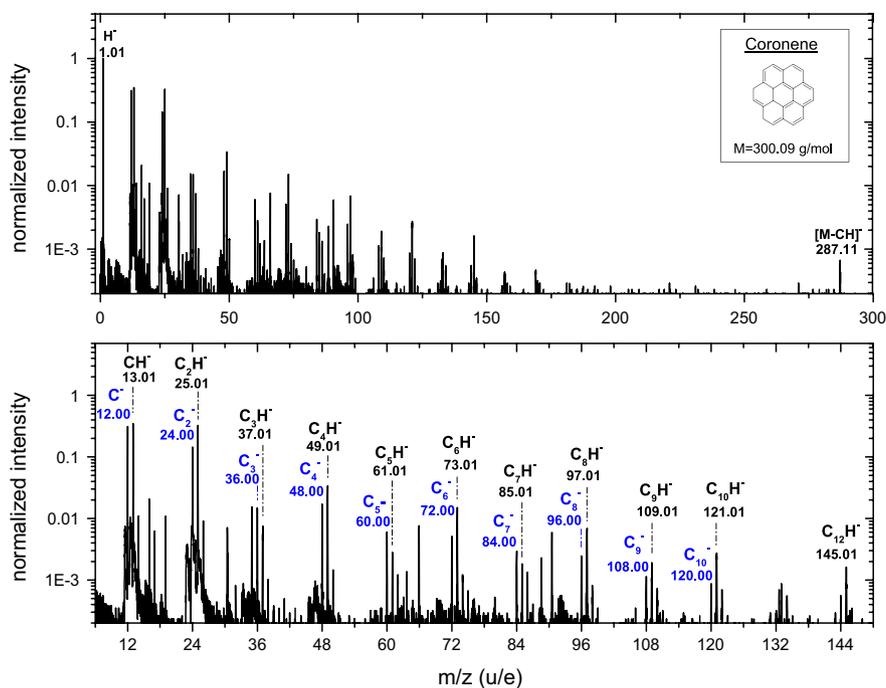


Fig. 12. Negative ion mode mass spectrum of coronene. The intensity is normalized to the major peak ($m/z=1.01$ u/e).

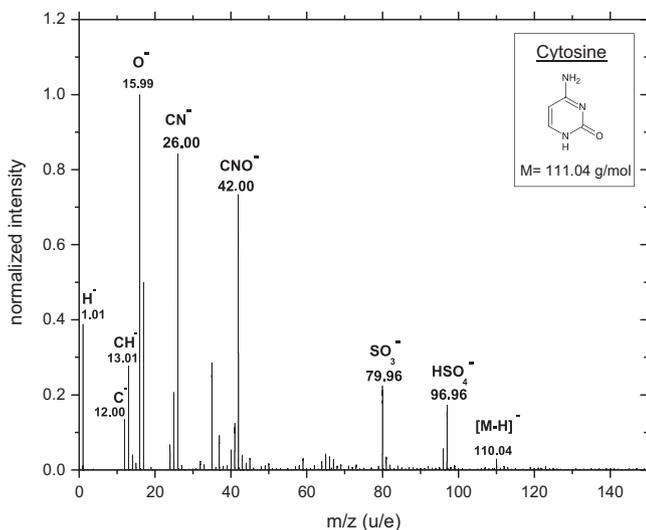


Fig. 13. Negative ion mode mass spectrum of cytosine. The intensity is normalized to the major peak ($m/z=15.99$ u/e).

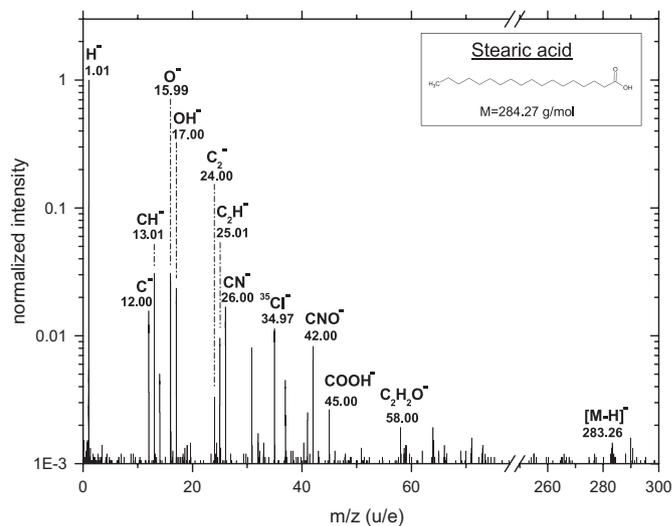


Fig. 14. Negative ion mode mass spectrum of stearic acid. The intensity is normalized to the major peak ($m/z=1.01$ u/e).

N-heterocyclic compounds in the analyzed sample. Moreover these ions (CN^- , O^- , CNO^-) are usual contaminant peaks in the negative mode. Their origin (contamination or sample) can be resolved by comparing their relative intensities to the control spectrum. Some others peaks related to the fragmentation of the quasi-molecular ion with the loss of NH_3 , HCN , HNCO or H_2O but also at low intensity compared to the major peak can also be found.

3.3.4. Carboxylic acid

In the negative ion mode, for the fatty acid (Fig. 14), the most intense peaks in decreasing order are relative to the following ions: H^- (at $m/z=1.008$ u/e), CH^- (at $m/z=13.01$ u/e), O^- (at $m/z=15.99$ u/e), OH^- (at $m/z=17.00$ u/e), C^- (at $m/z=12.00$ u/e). As in the positive ion mode, some similarities can be observed between the negative ion mode mass spectrum of the fatty acid and the ones of alkanes with the relative high intensity of the peak

associated to the CH^- and C^- ion (compare to the other major peaks). Some other peaks more specific to the carboxylic function have also been detected. A peak at $m/z=45.00$ u/e relative to COOH^- ion has been observed. Some peaks ($\text{C}_2\text{H}_2\text{O}_2^-$ and $\text{C}_2\text{H}_3\text{O}_2^-$ at $m/z=58.00$ and 59.01 u/e) which come probably from the fragmentation of the aliphatic chain are also observed. Nevertheless the quasi-molecular ion $[\text{M}-\text{H}]^-$ of the stearic acid is barely discernible from the background.

The analyses of the glycolic and benzoic acid revealed that these compounds were more easily identifiable in negative than in positive ion mode (Fig. 15). For both molecules, their quasi-molecular ions ($[\text{M}-\text{H}]^-$) at $m/z=121.03$ and 75.01 u/e, respectively for the benzoic and the glycolic acid, have been slightly detected. For the glycolic acid we have identified peaks linked to C_2HO_3^- ($[(\text{M}-\text{H})-2\text{H}]^-$ at $m/z=72.99$ u/e), COOH^- ($[(\text{M}-\text{H})-\text{H}_2\text{CO}]^-$ at $m/z=44.99$ u/e), CHO^- ($[(\text{M}-\text{H})-\text{CH}_2\text{O}_2]^-$ at $m/z=29.00$ u/e) ions. For the benzoic acid, some peaks related to

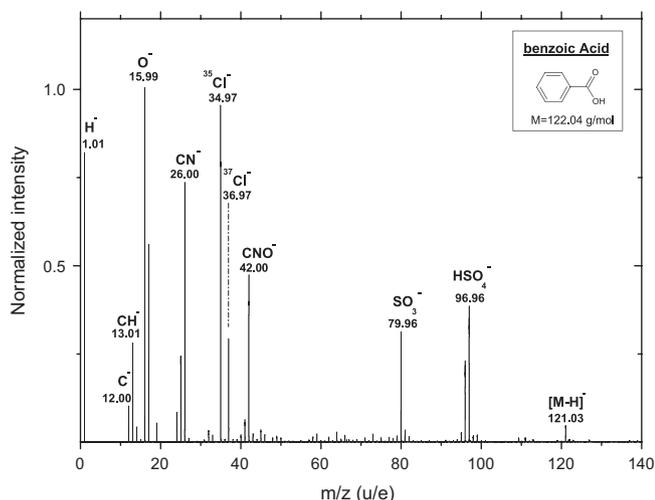


Fig. 15. Negative ion mode mass spectrum of benzoic acid. The intensity is normalized to the major peak ($m/z=15.99$ u/e).

the fragmentation of the quasi-molecular ion have also been detected: $C_6H_5^-$ ($[(M-H) - CO_2]^-$ at $m/z=77.03$ u/e), $C_4H_2O^-$ ($[(M-H) - C_3H_3O]^-$ at $m/z=66.01$ u/e), C_4HO^- ($[(M-H) - C_3H_4O]^-$ at $m/z=65.01$ u/e). Nevertheless, for both glycolic and benzoic acid, it should be mentioned that all these specific peaks, have rather low intensities yield according to the background.

The highly electronegative COOH group of these samples makes these compounds potentially rather observable in negative ion mode (except for the stearic acid, which contains a long aliphatic chain). This aliphatic chain, which is ionized more easily in the positive ion mode, seems to be the better explanation for these behavior differences toward the instrument mode.

Further analyses of other carboxylic acids are required to see if it is a common trend.

3.3.5. Polyoxymethylene

Fig. 16 displays the negative ion mode mass spectrum of the polyoxymethylene sample. As discussed in Le Roy et al. (2012) few peaks related to the sample have been detected. These peaks are found at $m/z=1.01, 12.00, 13.01, 15.99, 17.00, 44.99$ u/e respectively relative to the presence of $H^-, C^-, CH^-, O^-, OH^-, CH_2OH^-$ ions. The simultaneous presence of these peaks is not sufficient to claim the POM detection because they can be found in the mass spectrum of other oxygenated compounds.

4. How can we facilitate the interpretation of the future COSIMA mass spectra?

Mass spectrometry results of the previous *in situ* cometary space missions have taught us how difficult can be the interpretation of complex organic matter data with mass spectra of relatively low mass resolution. The nature of the refractory organic component is inferred from the results of previous missions, from laboratory analysis of carbonaceous chondrites, of interplanetary dust particles (IDPs), of micrometeorites, of synthesized cometary analogs or from the studies of distributed sources. These inputs have been the starting point for the database establishment but it was also clear that the results that we will get could be quite different from what we expect. Therefore the aim of our database has never been thought to be exhaustive. Based on its data processing, we have tried to develop different methods in order to characterize and get as much information as possible about the chemical nature of the sample than just comparing COSIMA future

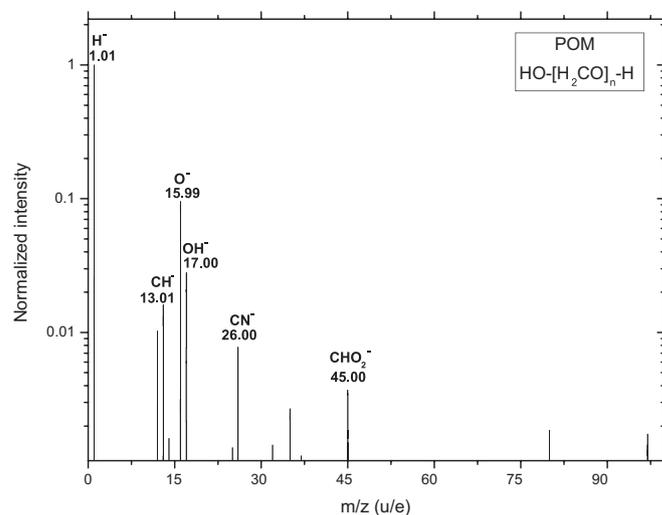


Fig. 16. Negative ion mode mass spectrum of polyoxymethylene. The intensity is normalized to the major peak ($m/z=1.01$ u/e).

data to our library. Numerous approaches have been developed for the characterization of complex organic mixture (Danger et al., 2013; Hughey et al., 2001; Kendrick, 1963; Kim et al., 2003). These methods are based on the accurate determination of peak centroid thanks to ultra-high resolution mass spectrometer and then to their different representations. Nevertheless due to the mass resolution of COSIMA and the precision of the peak's centroid determination, unique molecular formula cannot be assigned for each peak in the entire mass spectrum. Therefore we cannot really use the Van Krevelen approach. The Kendrick mass defect and the mass defect vs. exact mass diagram give some results when we have a good precision in the peak's centroid determination. Thus we had to find other methods, more suitable for non-high resolution mass spectrum.

In this section we will present the different procedures as proposed to be used for COSIMA flight data analysis.

4.1. How to identify quasi-molecular ion in a mass spectrum of an unknown sample?

The most specific feature related to a compound is the peak related to the molecular ion (M^+) or to the quasi-molecular ion ($[M \pm xH]^+$). The detection of these peaks is the best way to identify a compound. By comparing all the mass spectra measured for the database, we have tried to find a way to identify these molecular or quasi-molecular ions in a forest of other peaks.

In the positive mode, we have observed a simultaneous presence of some peaks related to quasi-molecular ions, $[M+H]^+$, to dimers $[2M+H]^+$ and to organometallic clusters $[M+Au]^+$ and or $[2M+Au]^+$. Table 5 summarizes features that can be observed in the mass spectrum of analyzed sample. It appears that the research of specific mass differences (ΔM) associated to the simultaneous presence of these signatures is a good way to highlight the presence of quasi-molecular ions.

For all the N-heterocyclic compounds, for the fatty acid, for all the PAHs analyzed, we can look for ΔM_1 ($\Delta M_1 = M_{Au} - M_H = 195.96$ relative to the simultaneous presence of $[M+H]^+$ and $[M+Au]^+$ or $[MH+H]^+$ and $[MH+Au]^+$ and/or $[2M+H]^+$ and $[2M+Au]^+$) and ΔM_2 ($\Delta M_2 = M$ relative to the simultaneous presence $[M+H]^+$ and $[2M+H]^+$, $[M+Au]^+$ and $[2M+Au]^+$).

For the alkanes, no simultaneous observations of the organometallic cluster and of the quasi-molecular ion have been detected. So we will not be able to highlight the presence of molecular or quasi-molecular ions by using this method.

Table 5
Molecular, quasi-molecular and organometallic ions detected in the positive ion mode mass spectrum of the analyzed compounds. “+” means that the considered ion has been detected.

Molecule (M)	M ⁺	[M - nH] ⁺	[M - H] ⁺	[M - 2H] ⁺	[M - 3H] ⁺	[M - 4H] ⁺	[M + nH] ⁺	[M + H] ⁺	[M + 2H] ⁺	[M + CH _n] ⁺	[M + CH ₂] ⁺	[M + CH ₃] ⁺	[mM + H] ⁺	[2M + H] ⁺	[3M + H] ⁺	[mM + Me] ⁺	[M + Au] ⁺	[2M + Au] ⁺	[mMH + Me] ⁺	[MH + Au] ⁺	[2MH + Au] ⁺	
	Eicosane																					
Tetracosane																						
Octacosane																						
Hexatriacontane																						
Anthracene																						
Pyrene																						
Coronene																						
Cytosine																						
Uracile																						
Thymine																						
Adenine																						
Hypoxanthine																						
Guanine																						
Imidazole																						
2-Aminopyrimidine																						
4-Aminopyrimidine																						
Isocytosine																						
Xanthine																						
HMT																						
Stearic acid																						
Glycolic acid																						
Benzoic acid																						

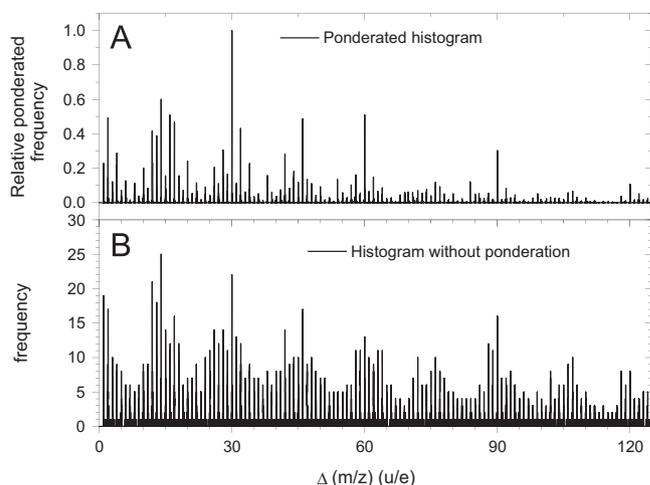


Fig. 17. Histogram of the mass difference presents in the positive ion mode mass spectrum of POM. The panel A presents the frequency of each mass difference with a class interval of 0.005 u/e. The panel B presents a histogram in which all mass differences have been weighted by the intensities product of both considered peaks.

By using this method implemented into a code, several peaks couple can be found. Some of these solutions can be false positive. So in order to decrease the number of false positive, we have used a filter on the peak intensity because the intensity of the molecular/quasi-molecular peak is higher than the one of the clusters.

4.2. How to characterize a polymeric structure in a mass spectrum of an unknown sample?

One of the high molecular weight organic compounds which is suspected to be present in cometary grains is the polyoxymethylene, a polymer of formaldehyde (H_2CO). POM positive ion mass spectrum presents a peculiar pattern: an alternation of peaks with a mass difference of 30.011 u/e corresponding to a H_2CO molecule as already presented in Fig. 10. In order to reveal this specific mass difference between peaks, the method already presented in Le Roy et al. (2012) can be applied.

This approach consists in the determination of the exact position and the intensity at maximum of each peaks of the mass spectrum (only peaks having an intensity higher than 100 counts at their maximum have been considered). Then, for each pair of peaks, the mass difference have been calculated and then plotted as a histogram with a class interval of 0.005 u/e. This first histogram is represented in Fig. 17a. Moreover, in Fig. 17b a second histogram represents each mass difference weighted by the intensities product of both considered peaks. It can be noted that this second histogram is identical to the one presented in Figure 7a of Le Roy et al. (2012). Both histogram reveal clearly the alternation of peaks with a mass difference of 30.010 ± 0.005 u/e.

As already shown in Le Roy et al. (2012) (see there Figure 8), the precision on the mass difference is better than on the position of each peaks in the original mass spectrum. This increase of precision in the mass determination is the main advantage of this technique based on the mass differences. Moreover in both histogram (see Fig. 17a and b), it can be observed some multiple of the main mass difference, i.e. 30.010 ± 0.005 u/e, as well as 60.025 ± 0.005 u/e and 90.030 ± 0.005 u/e, characteristic of a polymeric structure with a monomer having a mass of 30.010 ± 0.005 u/e. Furthermore, mass difference of 14.015 ± 0.005 u/e and 15.995 ± 0.005 u/e (Fig. 17a) are also observed. The first one can be due to pair of peaks linked to $[(H_2CO)_n + H]^+$ and $[(H_2CO)_n + H - CH_2]^+$ (difference of CH_2 FW=14.0156 amu) and the second one at 15.995 ± 0.005 u/e to pair of peaks linked to

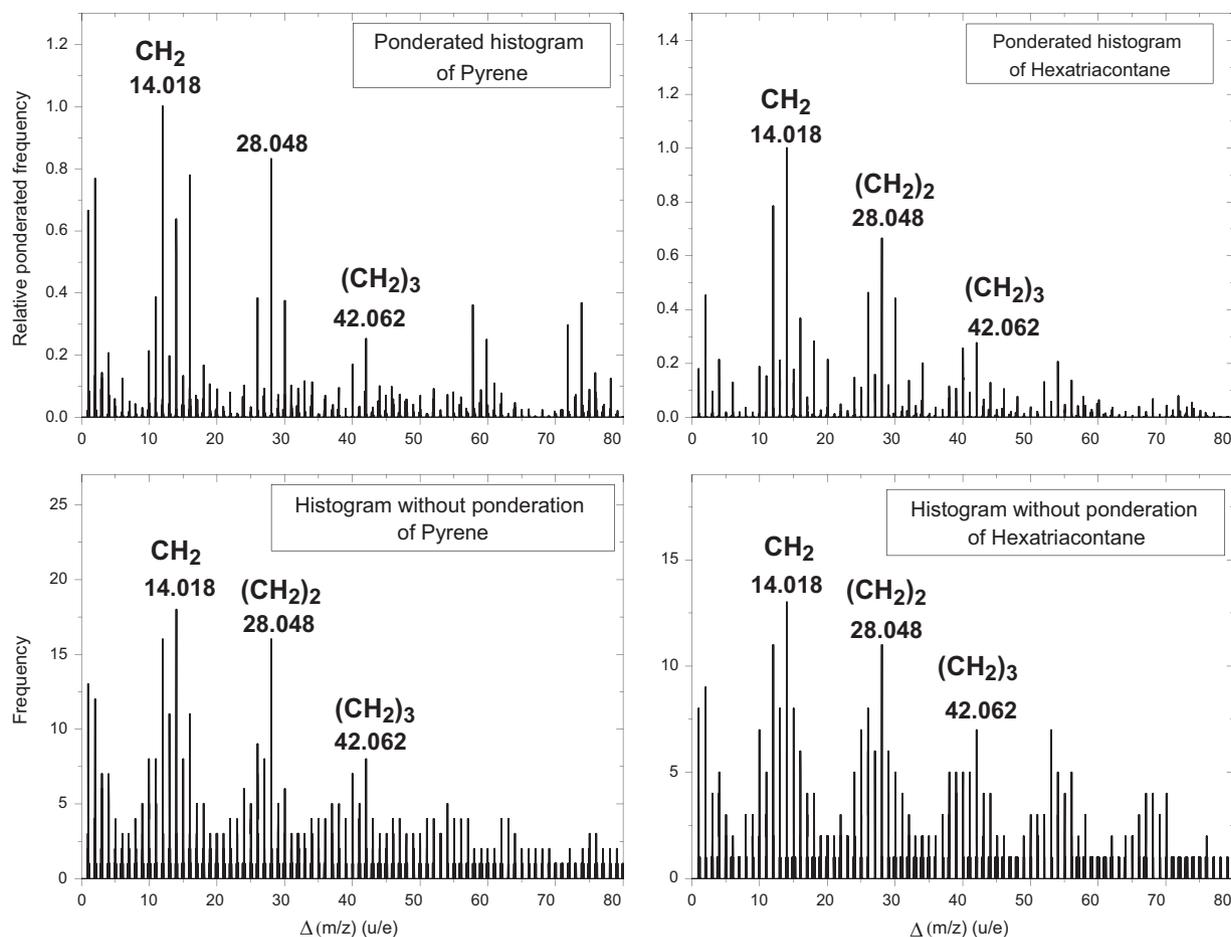


Fig. 18. Histograms of the mass differences in the positive ion mode mass spectrum of hexatriacontane and pyrene. The top panel presents the frequency of each mass difference with a class interval of 0.005 u/e. The bottom panel presents a histogram in which all mass differences have been weighted by the intensities product of both considered peaks.

$[(\text{H}_2\text{CO})_n + \text{H}]^+$ and $[(\text{H}_2\text{CO})_n + \text{H} - \text{O}]^+$ (difference of O FW = 15.995 amu). The only compound, which can explain all these observations, is a H_2CO polymer.

The same approach could also be employed for non-polymeric compounds, such as hydrocarbons. Histograms of the mass differences present for the positive ion mode mass spectra of pyrene ($\text{C}_{16}\text{H}_{10}$) and hexatriacontane ($\text{C}_{36}\text{H}_{74}$) are represented in Fig. 18. In these histograms, major peak is located at 14.018 ± 0.005 u/e. This mass difference results from the alternation of $[\text{C}_n\text{H}_{2n-1}]^+$ and $[\text{C}_n\text{H}_{2n+1}]^+$ ions in the low mass spectra and can be attributed to CH_2 . Multiples are also observed at 28.048 ± 0.005 and 42.062 ± 0.005 supporting the presence of a repetitive unit within the mass spectrum. The same characteristics are observed for the four alkanes, the three PAHs and the stearic acid studied in this paper. The latter does not belong to the hydrocarbon family, but its long aliphatic saturated chain ($\text{CH}_3(\text{CH}_2)_{16}\text{CO}_2\text{H}$) confers the behavior of an alkane on it.

4.3. How to distinguish compounds made of different kind of atoms (absence or presence of nitrogen) or made of different structures (aliphatic and aromatic hydrocarbons) in an unknown mass spectrum?

4.3.1. Tentative of quantitative measurements of the N/C elemental ratio.

This section is focused on the N/C elemental ratio. In order to measure it, a study on the intensity of peaks which can be observed

in all the analyzed samples and which are not specific to some of them have been achieved: peaks at mass to charge ratio lower than 30 u/e for instance. Indeed, at these low masses, few overlappings occur and convincing attributions could be performed for most peaks. More specifically four peaks have been chosen: peaks at nominal masses 24 and 26 u/e in negative ion mode and peaks at nominal masses 28 and 29 u/e in positive ion mode. These four peaks have been adjusted by a Gaussian function to precisely determine the exact position of their maximum and to measure their intensities, *i.e.* the number of counts. We have considered the spectra, in negative and positive modes, of the 23 different pure organic compounds presented in Table 1, as well as “blank” spectra for each compound and mode, leading to a total of 92 considered spectra. Fig. 19 presents the ratio of the peak’s intensity at nominal masses 28 and 29 u/e ($I(28)/I(29)$) as a function of their maximum’s position. The maximum of the peak at nominal mass 29 u/e is quite stable and is precisely located at $m/z = 29.039 \pm 0.004$ u/e (see Fig. 19B), thus this peak can be attributed to C_2H_5^+ ($m/z = 29.039$ u/e) for all the positive ion mode mass spectra. The exact position of peak at the nominal mass 28 range from $m/z = 28.017$ up to 28.033 u/e and it is correlated to the intensity ratio $I(28)/I(29)$ (see Fig. 19A). Thus, an overlap between HCNH^+ ($m/z = 28.019$ u/e) and C_2H_4^+ ($m/z = 28.031$ u/e) is observed. When, the maximum of this peak is close to 28.031 u/e, the intensity ratio $I(28)/I(29)$ is about 0.1. In that case, this peak is attributed to C_2H_4^+ which is a radical ion, thus less stable than C_2H_5^+ which is a molecular ion explaining the low intensity ratio (as for alkanes and PAHs). On the other hand, when the maximum of the peak is close to 28.019 u/e,

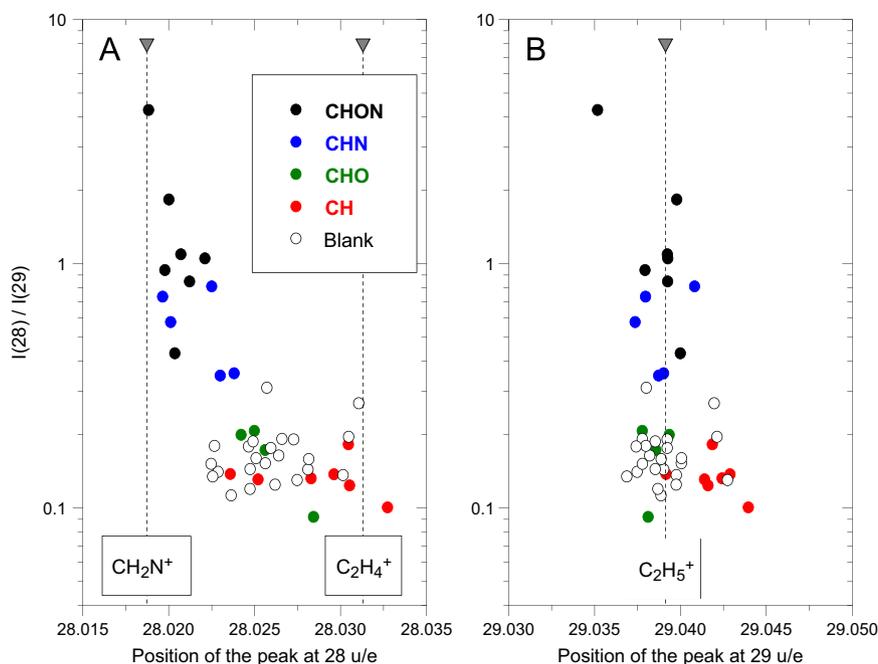


Fig. 19. Intensity, i.e. number of counts, ratio for peaks at nominal masses 28 and 29 u/e as a function of the exact position of peak at m/z 28 (Panel a) and 29 (Panel b) in positive ion mode. Spectra of molecules for the different molecular families “CHON”, “CHN”, “CHO” and “CH” have been represented in black, blue, green and red respectively, whereas the blank spectra have been represented by open symbols. Vertical dashed lines represents the exact position of HCNH^+ , C_2H_4^+ and C_2H_5^+ ions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the intensity ratio $I(28)/I(29)$ is about 1. In this second case, this peak could be attributed to HCNH^+ which is a molecular ion as C_2H_5^+ . Both ions have similar stability and the $I(28)/I(29)$ intensity ratio is about 1.

Unfortunately, the correlation between the variation of the $I(28)/I(29)$ intensity ratio and the N/C elemental ratio of the studied samples is very low. Thus, this intensity ratio cannot be used for quantitative estimation of the N/C elemental ratio in samples of unknown nature such as cometary grains. Nevertheless, it can be noted that the exact position of the peak at mass 28 as well as the $I(28)/I(29)$ intensity ratio could give clues to the presence of nitrogen atoms in a sample of unknown nature.

In negative ion mode, the exact positions of peaks at nominal masses 24 and 26 u/e are stable and are precisely located at 24.000 ± 0.003 u/e and 26.002 ± 0.003 u/e. They could be unambiguously attributed to C_2^- ($m/z=24$ u/e) and CN^- ($m/z=26.003$ u/e). The non-nitrogenated compounds as well as the blank present a quite large dispersion of the intensity ratio $I(26)/I(24)$ (Fig. 20), due to the presence of nitrogen atom in the background of the instrument and the weakness of the peaks related to the C_2^- ions. Nevertheless, the non-nitrogenated and nitrogenated compounds can be clearly distinguished (Fig. 20).

Considering only the spectra of nitrogenated compounds, a calibration line can be calculated as well as the prediction bands at 68% of confidence (Fig. 21). These prediction bands could be used to estimate the uncertainties on the determined N/C elemental from an unknown sample. Whatever the elemental ratio is, the errors bar is about ± 0.2 .

This error bar is high compared to the plausible cometary N/C elemental ratio which ranges from almost zero to 0.25. It is clear that more work is needed before to perform a rigorous analysis of the spectra of cometary grains. The present work could be seen as a guideline for a future calibration based on samples containing macromolecular carbonaceous compounds such as insoluble organic matter (IOM) extracted from carbonaceous meteorites, meteorites samples and kerogen which present low N/C elemental

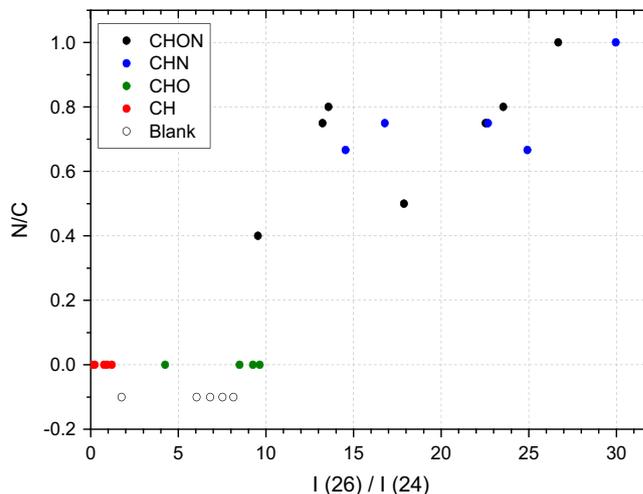


Fig. 20. N to C elemental ratio as a function of the intensity ratio of peaks at masses 26 and 24 in negative ion mode. We arbitrary assign a negative N/C ratio from the “blank” spectra represented by open symbols. Spectra of molecules for the different molecular families “CHON”, “CHN”, “CHO” and “CH” have been represented in black, blue, green and red respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ratio representative of the ones which could be present in the grains of comet 67P/CG

4.3.2. How to highlight the presence of aliphatic chains or aromatic structures in a sample of unknown nature?

Alkane and PAH families are both only made of hydrogen and carbon atoms but through a different structure: a long aliphatic saturated chain and a polyaromatic structure. Similarly, the hydrogen content differs, being higher for alkanes than for PAHs.

As mentioned in Section 4.2 alkanes and PAHs are compounds showing, in the low mass range of positive mass spectra, a significant alternation of $[\text{C}_n\text{H}_{2n-1}]^+$ and $[\text{C}_n\text{H}_{2n+1}]^+$ ions.

Moreover, in the negative ion mode, low mass fragments are also found such as H^- , C^- , CH^- , CH_2^- , C_2^- and C_2H^- . Among them C_2^- and C_2H^- ions belonging to the PAH family have a very high intensity in regard to the other analyzed compounds. This tendency, normalized to the CH_2^- ion is pointed out in Fig. 22. This figure can be an attempt to highlight the presence of aliphatic chains and polyaromatic structure in an unknown mass spectrum. It could also be seen as a tentative distinction of low atomic hydrogen to carbon ratio and high atomic hydrogen to carbon ratio. The organic compounds with a “low” content of hydrogen are the PAHs and those with a higher content of hydrogen atom to carbon atom are the alkanes. It should be obviously mentioned that our database is not exhaustive and therefore these results can evolve with the analyses of new chemical family. This result will obviously come as additional support to the analysis of the

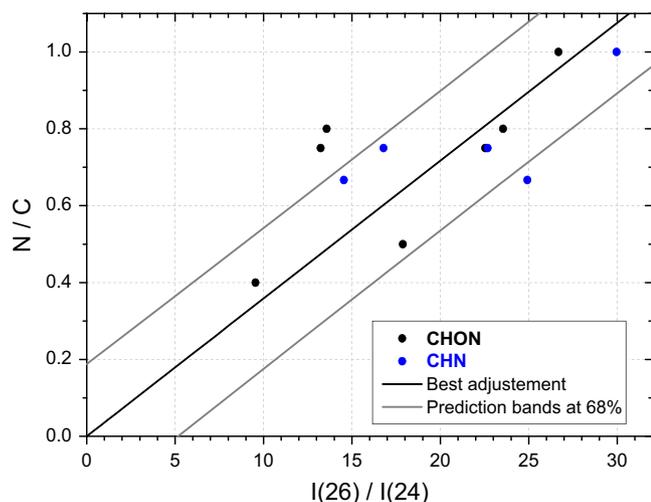


Fig. 21. N/C elemental ratio for nitrogenated compounds as a function of the intensity ratio at masses 26 and 24 in negative mode. The black line is the best adjustment and both gray lines are the prediction bands at 68% of confidence. Spectra of molecules for “CHON” and “CHN” molecular families have been represented in black and blue respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

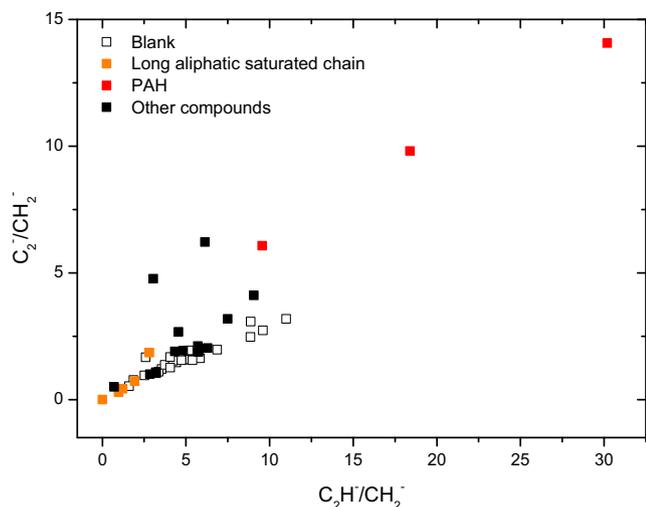


Fig. 22. $\text{C}_2^-/\text{CH}_2^-$ intensity ratio as a function of $\text{C}_2\text{H}^-/\text{CH}_2^-$ intensity ratio in negative mode. Long aliphatic saturated chain compounds in orange refer to the four alkanes and the stearic acid; PAH compounds are represented in red and blank spectra by open symbols; “Other compounds” in black refer to remaining analyzed molecules from Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

positive mass spectra. Nevertheless it clearly supports the fact that negative ion mode will also bring us some information on the interpretation of the future COSIMA mass spectra.

5. Applying the Blind Source Separation approach to interpret COSIMA mass spectra

As a first step for COSIMA data handling strategy, methods implementing data reduction will be extremely useful. As part of regular COSIMA operations on board Rosetta, after a preliminary set of TOF-SIMS analysis on a target holder, the COSIMA team will have to quickly (within 48 h) optimize the COSIMA commands sequence for more detailed measurements. Processing all the spectra one by one is time-consuming, and not feasible in a short time period. A data reduction method will provide clues to choose the most interesting cometary grains for which long analyses should be foreseen.

Moreover, COSIMA will provide a relatively high number of mass spectra as cometary grains found on the target and some target area free of visible sample will be analyzed by line or matrix scanning. A dataset of mass spectra can provide a huge amount of useful data that may be difficult to understand without a preliminary step of reduction. In order to extract useful and likelihood information, our goal is then to study a large number of spectra without any assumptions about the samples composition, through a limited data volume that should be representative of the full data set.

All these COSIMA data will have to be interpreted for scientific analyses of the collected cometary matter. It is expected that the data we may record would be mixtures of pure compounds, including minerals and organics from the grain, as well as a contribution from the target itself (as discussed in Section 2). Thus the data recorded could be seen as a linear combination of pure components spectra. It is then a source separation problem where the unknown pure components spectra would be the sources and the concentrations the mixing coefficients. For the scientific data analyses, again, data reduction methods could also provide clues to identify pure compounds hidden into the expected complex mixtures.

To process multivariate data, many numerical statistical techniques are available, such as the most popular one: the Principal Component Analysis (PCA). Blind Source Separation (BSS) approach (Comon and Jutten, 2010) is more constraining and is suitable to reveal hidden factors underlying on sets of measured signals. Through the Bayesian Positive Source Separation method (BPSS), it provides a decomposition of multivariate data into positive sources through positive mixing coefficients.

In this paper, we attempt to address the capability of the BPSS technique as a valuable method to interpret future COSIMA data. At first for an easier understanding of our work, a description of the method is briefly given in Section 5.1 (for more details see for instance Moussaoui et al. (2006a, 2006b, 2006c)). Section 5.2 presents the three significant use features that can bring BPSS technique to interpret COSIMA dataset. Section 5.3 recalls main preliminary results and further perspectives of the use of this technique.

5.1. Bayesian Positive Source Separation (BPSS)

A typical separation problem entails to extract individual sources emitting any kind of signal from the overall recording of a detector. Bayesian Positive Source Separation (BPSS) formalism generally assumes N independent sources $s(t) = (s_1(t), \dots, s_N(t))$ and P independent mixtures $x(t) = (x_1(t), \dots, x_P(t))$ ($x(t)$ being the observation vectors or mass spectra in this study). The mixture

is assumed to be a linear superposition of all sources, with an error $n(t)$, and can be expressed by the following compact equation:

$$x(t) = As(t) + n(t)$$

with A the matrix of mixing coefficients and t the observation index (being the mass in our study).

As an infinity of solutions is possible, some additional assumptions or prior information have to be introduced. The BPSS approach assumes that both the sources and mixing coefficients are positive. An inferential reasoning is then used with a Bayesian formalism. In this logical approach, the possibility of an event can be calculated or revised. Since there is a relationship between the data, the mixing matrix, and the source signals, it is not necessary to solve for both A and $s(t)$. The degree to which we believe the assumption is correct relies on the likelihood of the data and additional prior knowledge.

In practice, the N sources $s_i(t)$ resulting from the method would be comparable to one of the real mass spectra observed but would in fact be produced by computation. This is also true for A and $n(t)$ matrices.

The parameter N has to be defined by the user, and is optimized as follows: N is too high if many sources are too similar to each other, N is too low (and an increment value can be tested) if no twin sources are identified.

After a run of the BPSS method, the following issues must be addressed: in which extent are these sources representative of the dataset, and what is the stability of these sources and mixing coefficients as a function of a given computation? To validate the process and define a criterion of repeatability, several runs have to be implemented with the simplification advantage in mind. A source should appear in most of the runs to be considered as consistent.

5.2. COSIMA organic calibration data set analysis by BPSS

5.2.1. Estimated sources determination

Each mass spectrum is made of two vectors, one for the mass per charge ratios and one for the intensities. The mass spectra are reduced to lists of peaks with a peak recognition step. Each peak is

replaced by its maximum intensity and its mass per charge ratio. The mass calibration is performed through a software developed by the COSIMA team. The peak recognition is achieved using the Matlab function “mspeaks”. The BPSS treatment step is obtained through a Matlab function written by Moussaoui et al. (2006a, 2006b, 2006c). Most of the data treatment is made with the Matlab software. A matrix of mass spectra intensities, a vector of mass per charge ratios and a vector of spectra names, is then built during the data treatment.

In the dataset used in this paper, 48 mass spectra in the positive ion mode (p observations, $x_1(t), \dots, x_p(t)$) are used. They include blank spectra measured on the targets and the 23 organic pure compounds mass spectra of the laboratory calibration COSIMA data. Hence, the number of potential sources (N) should be at least 23. 2 sources dedicated to contaminant compounds are also taken into account as sources, giving N to be 25.

After BPSS runs, from this set of data, 16 organic samples are associated to a single source and 7 are not (those 7 samples will be discussed in Section 5.2.3). For several many BPSS runs, the fingerprints of these 16 sources are reproducible for the most intense peaks and these sources are representative at least of the dominant part of the dataset. The 16 samples show specific peaks of high relative intensities (including most of the time the quasi-molecular ion $[M+H]^+$). As an example, positive ion mode mass spectrum of adenine is shown in Fig. 23A. The closest source relative to adenine mass spectrum calculated by the BPSS method, the source number 8 ($s_8(t)$), is shown in Fig. 23B.

5.2.2. Test of correlation of a single source with one referenced mass spectrum

An interesting benefit of the BPSS approach could be to highlight positive high correlation between unknown mass spectra from a dataset with one referenced mass spectrum.

To test this feature, among the 16 samples which were associated to a single source, the source related to adenine ($s_8(t)$) was picked up to highlight the capability of BPSS to identify sources and estimate the correlation between these sources and all the mass spectra of the dataset.

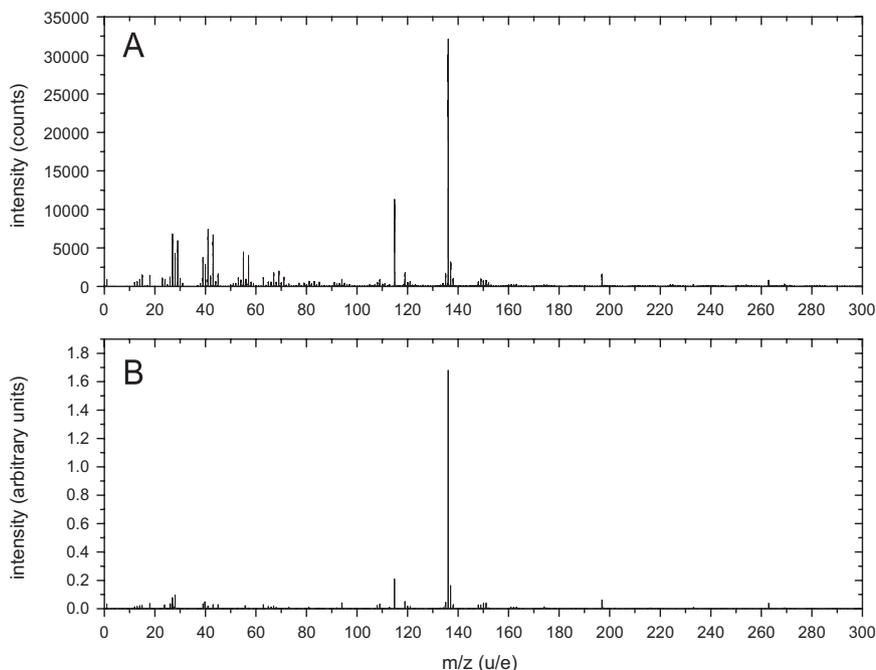


Fig. 23. Original positive ion mode mass spectrum of an adenine deposition on gold target sputtered with an Indium ion beam (Panel A) and the Source N8 (S_8) associated to an adenine fingerprint obtained by BPSS data treatment from a set of spectra including adenine compound (Panel B). For both, the main peak is at $m/z=136$ u/e.

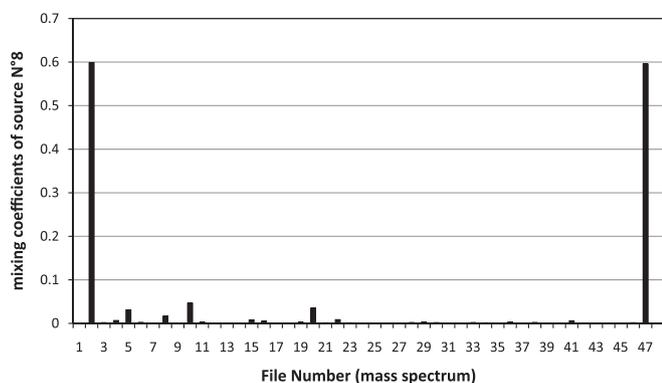


Fig. 24. Mixing coefficients of source $s_8(t)$ for all mass spectra of the data set. Only samples $x_2(t)$ and $x_{47}(t)$ are composed of adenine.

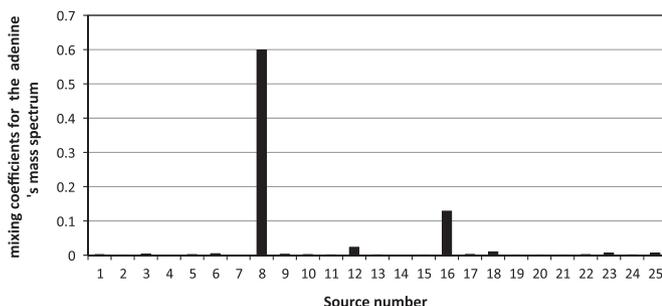


Fig. 25. Sources mixing coefficients for the mass spectrum number 2 (adenine sample). Source $s_8(t)$ is associated to adenine but sources $s_{12}(t)$ and $s_{16}(t)$ are also present. The source intensities are respectively 0.59, 0.13 and 0.02 for source $s_8(t)$, $s_{16}(t)$ and $s_{12}(t)$.

Therefore, the mixing coefficients of source $s_8(t)$ linked to adenine in a matrix made of all mass spectra are plotted to highlight which file numbers are relative to the analysis of the adenine. As shown in Fig. 24, for our test case the files number 2 and 47 (observations $x_2(t)$ and $x_{47}(t)$) (which are indeed actual adenine spectra) are correlated to the analysis of adenine. Thus as expected, we found two high mixing coefficients (above 0.5) for these file numbers. Thanks to the BPSS, we can also determine if the source $s_8(t)$ is actually the only main source that contributes to the adenine positive ion mode mass spectrum. To that purpose on Fig. 25 is displayed the mixing coefficients for an adenine spectrum ($x_2(t)$) as a function of the source number. A strong contribution of source $s_8(t)$ is expected, as well as minor contributions from other sources especially from sources associated to contaminants. From Fig. 25, it can be deduced that sources $s_{12}(t)$ and $s_{16}(t)$ are also significantly contributing to the mass spectrum file number 2 (adenine mass spectrum). Sources $s_{12}(t)$ and $s_{16}(t)$ are associated to contamination.

Finally, a mixture of these three sources ($s_8(t)$, $s_{16}(t)$ and $s_{12}(t)$) has then been computed (Fig. 26) and compared to the real adenine mass spectrum (Fig. 23A). Excellent fit is found with a correlation coefficient of 0.9988.

Data reduction process performed with BPSS treatment for adenine positive ion mode mass spectra, shows that adenine mass spectra are deconvoluted in one single sample source and two contamination sources. The mixing coefficients of the contaminants sources indicate the level of cleanliness of the mass spectra.

As a summary, BPSS treatment tests confirm the capability of the BPSS method to help the data interpretation procedure. BPSS provides good results for most of the major peaks and for many minor peaks. It seems to be a powerful tool for deconvolution of spectra.

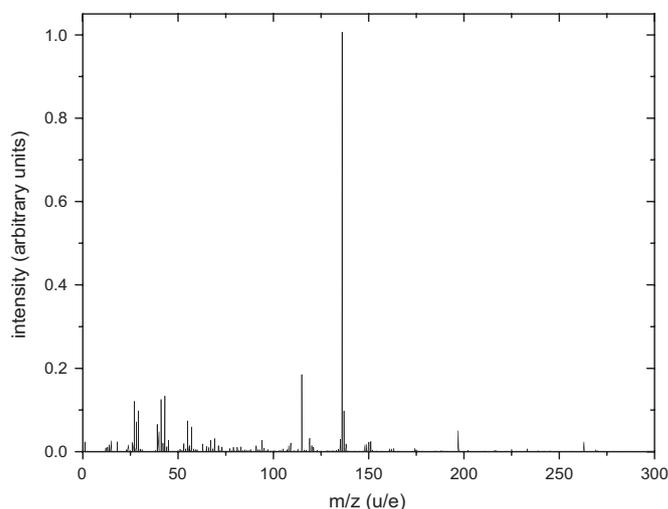


Fig. 26. Linear mixture of 3 sources: 0.59 times source $s_8(t)$ +0.13 times source $s_{16}(t)$ +0.02 times source $s_{12}(t)$. This mixture is a good model of adenine spectra $N2x_2(t)$ shown in Fig. 23. The main peak is at mass 136 u/e.

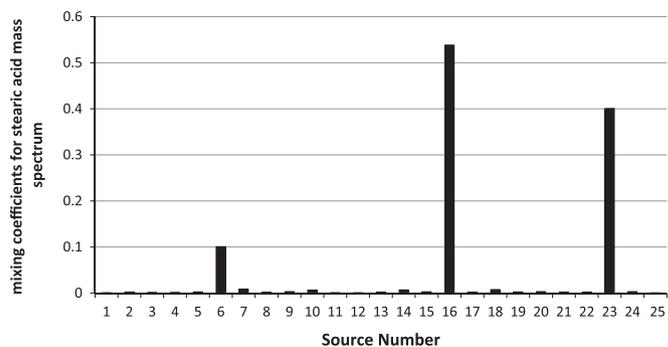


Fig. 27. Mixing coefficients of the 25 sources provided by the BPSS data treatment for the stearic acid mass spectrum.

5.2.3. Test of correlation from several sources associated to a sample

After BPSS runs, 7 organic samples analyzed in this study have not been associated to a single source: 2-aminopyrimidine, benzoic acid, stearic acid and the four alkanes. The peak intensity of their quasi-molecular ion is in fact low or not detected. These samples provide mass spectra dominated by C_xH_y fragments at low mass range ($m/z \approx 1-150$ u/e) and specific peaks about their mother molecular weight at low relative intensities (few % or less) (see Section 3.2).

2-Aminopyrimidine is even peculiar as it is linked to the source associated to 4-aminopyrimidine and also to other non-specific sources with similar values of concentrations.

All the four alkanes present a high concentration value for source $s_{23}(t)$, with strong mixing coefficient for octacosane and hexatriacontane (respectively 0.8 and 0.9). Source $s_{23}(t)$ could be a possible fingerprint for alkanes detection but some other samples have also a non-negligible concentration of source $s_{23}(t)$ (benzoic acid, stearic acid and anthracene).

Benzoic acid and stearic acid have not been associated to specific sources. This tendency is shown in Fig. 27 for stearic acid. In which major mixing coefficients for stearic acid are associated to sources $s_{16}(t)$, $s_{23}(t)$ and $s_6(t)$, all non-specific sources. For these acids the BPSS treatment was not able to provide significant result.

Clearly with this kind of samples, a more detailed treatment has to be performed. For example one can try to understand the composition and origin of source $s_{16}(t)$ (Fig. 28). This source is mainly associated to compounds including C_xH_y chain. Plotting its

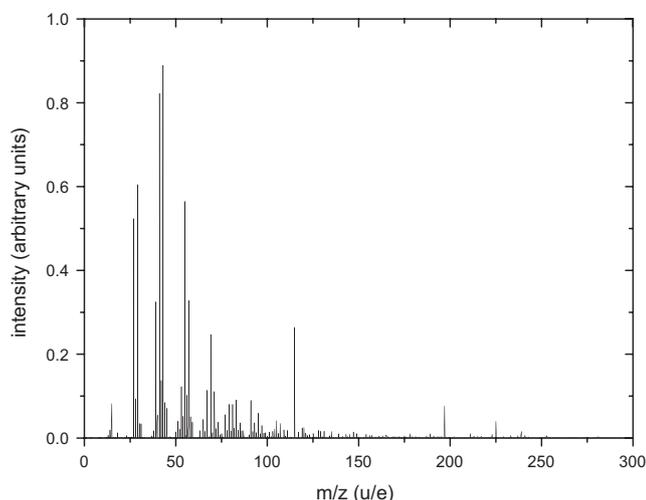


Fig. 28. Source $s_{16}(t)$ from BPSS treatment from a set of 48 mass spectra including 23 organic samples. These data could be associated to an organic compound as C_xH_y with a contribution of indium (mass 115 u/e, from the ion beam) and gold (mass 197 u/e, from the target).

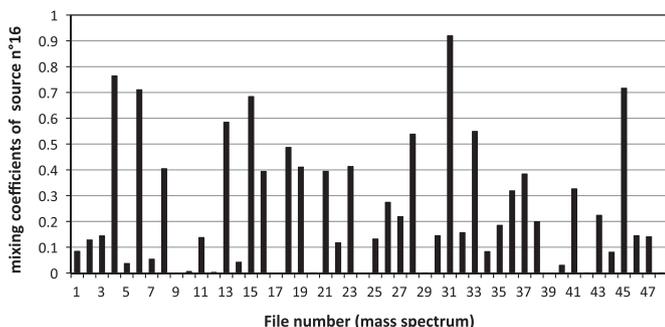


Fig. 29. Mixing coefficients of source $s_{16}(t)$ for all mass spectra of the dataset. This graph indicates probably that source $s_{16}(t)$ is a regular contaminant.

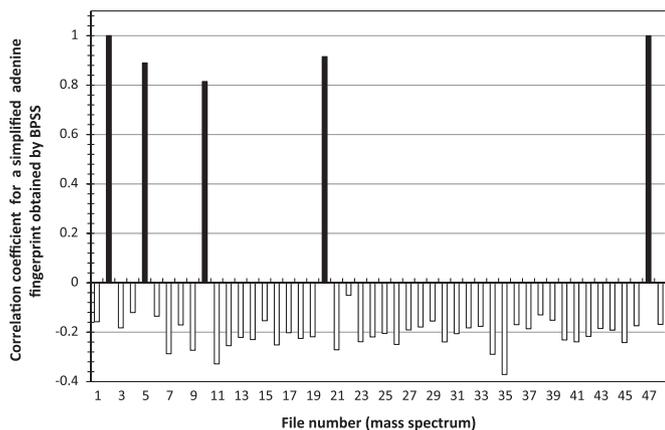


Fig. 30. Correlation between a simplified fingerprint of BPSS source associated to adenine and the dataset mass spectra. The two adenine mass spectra ($x_2(t)$ and $x_{47}(t)$) provide a coefficient about 1, but other spectra provide good coefficients about 0.8 and 0.9. These spectra are guanine, imidazole and xanthine which are molecules of the same chemical family.

mixing coefficient for all the spectra would give the level of contribution of this source to the observations. Fig. 29 which pinpoints that source $s_{16}(t)$ is present in almost all mass spectra of the dataset. It may be an indication of sample contamination, a conclusion already drawn in Section 5.2.2. Its origin is inherent to the OM instrument vacuum chamber that is slightly contaminated by some phthalates molecules.

5.2.4. Is BPSS technique able to provide fingerprints of sample?

Another user-friendly tool of BPSS technique would be the capability of BPSS to provide fingerprints of samples. This possibility has been tested with adenine (Fig. 30). A list of 10 major peaks from BPSS adenine data was selected to get a simplified fingerprint for this compound. They have been selected for their good correlation with the adenine's quasi-molecular ion $[M+H]^+$ observed in various measurements. Then we have made direct correlation between this simplified fingerprint and all mass spectra in the database. The two adenine mass spectra $x_2(t)$ and $x_{47}(t)$ provide a coefficient of about 1, but other mass spectra provide good coefficients about 0.8 and 0.9. These mass spectra are those of guanine, imidazole and xanthine that which are all nitrogen-heterocyclic molecules.

With this example we show that the estimated fingerprint may help to sort samples as a function of their chemical group. This has to be confirmed with an extended data set. Simplified computed fingerprints could then be defined for the pure organic samples of the mass spectra library in order to be used as an automatic trigger to search specific features in the future COSIMA spectra measured at the comet. By limiting the number of selected peaks in the computed fingerprint, the correlation number between this fingerprint and the mass spectra can be determined in a short time. Nevertheless this number can be high because of the false positive induced by the less specific computed fingerprint. The selected peaks should then be chosen carefully. In conclusion computed fingerprints could probably help the separation process.

5.3. Discussion of the BPSS results

In most of the cases, for a simple dataset of mass spectra, BPSS technique is able to provide sources similar to real mass spectra and linked to a pure compound. Some samples were not perfectly processed due to a lack of specific features in their mass spectra. BPSS technique is quite efficient to point out the main compounds and spectral features associated to contamination. And in a similar way BPSS treatment was able to provide a common source between samples sharing common characteristics. BPSS technique will probably help the data analysis of the COSIMA mass spectra by providing computed fingerprints, assumptions, hints and data reduction. It will not prevent from a deep and detailed analysis of all mass spectra but we expect to gain time in decision process and to get clues to guide the data treatment with other tools.

It is not expected that the mass spectra fingerprints from *in situ* cometary data recorded by COSIMA XM model will be similar to the mass spectra of pure compounds provided by the laboratory instrument. There are huge differences between sample preparation in the laboratory and the natural state of cometary matter. The TOF-SIMS library of pure samples presented in this paper is a good starting point but is probably not sufficient to describe the cometary samples. To take these differences into account the strategy will be to prepare new samples with the objective of producing new reference laboratory mass spectra as similar as possible to the one that will be recorded by the COSIMA instrument in space

6. Conclusion

The detection and the identification of high molecular weight species which remain solid in cometary grains is a major interest of the Rosetta mission. COSIMA, which will collect and analyze the grains ejected by the nucleus of 67P/CG, is probably one of the best instruments in the Rosetta scientific payload to detect and characterize the solid organic matter. For an easier identification of these organic molecules present in the future mass spectra of

COSIMA, a TOF-SIMS library of targeted molecules was initiated by the COSIMA team. The analyses of several families of compounds have started with one of the ground models of COSIMA: the Orleans Model (OM). In general, the interpretation of the positive ion mode mass spectra (except for two carboxylic acids) facilitates the identification of the analyzed compounds. Indeed, these organic molecules have a larger number of characteristic signatures (molecular, quasi-molecular ions, dimer, metal adducts etc.). In this mode, the quasi-molecular ion is most of the time, the predominant peak of the mass spectrum. However negative ion mode mass spectra provide also important information on the nature of the sample and so will not be neglected. The mass spectra of cometary grains collected by COSIMA will be most likely a mixture of different minerals and organic molecules. It could be a challenge to identify a peak related to a molecular or a quasi-molecular ion in the numerous peaks of a mass spectrum. The method that has been developed to address this issue is to look for specific mass differences relative to the simultaneous presence of some specific ions (molecular, quasi-molecular ions, dimer, metal adducts). Furthermore we are totally aware that our database is not exhaustive and that the intensity of the organic molecules could be very low in the future COSIMA mass spectra, which could make very difficult the identification of the samples in cometary grains. Therefore other procedures have been developed in order to characterize the nature of the sample: a software dedicated to the research of polymeric structure has been developed; by measuring the intensity ratio of some peaks in the different mode of COSIMA, the presence of compounds with specific atom (molecules containing nitrogen atom or not) or with a specific structure (aliphatic chains vs. polyaromatic structure) can be highlighted. A first step for the quantification of N/C has also been presented by calculating the intensity ratio of peak at nominal masses 24 and 26 u/e in the negative mode. The Bayesian Positive Separation Source seems to be also very valuable to help us to study quickly the large amount of data that we will received and in order to send the most relevant command to the instrument during the operational phase. This technique can be also very useful for the data interpretation because some correlation between simplified fingerprint and all COSIMA mass spectra can be done in a short time. In any case, this work is just the starting point for the COSIMA mass spectra interpretation. The calibration work will continue until the end of Rosetta mission. The post calibration will play an important role for the good evaluation of the data.

Acknowledgments

COSIMA was built by a consortium led by the Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany in collaboration with Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Orléans, France, Institut d'Astrophysique Spatiale, CNRS/INSU and Université Paris Sud, Orsay, France, Finnish Meteorological Institute, Helsinki, Finland, Universität Wuppertal, Wuppertal, Germany, von Hoerner und Sulger GmbH, Schwetzingen, Germany, Universität der Bundeswehr, Neubiberg, Germany, Institut für Physik, Forschungszentrum Seibersdorf, Seibersdorf, Austria, Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz, Austria and is led by the Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany.

The support of the national funding agencies of Germany (DLR), France (CNES), Austria and Finland and the ESA Technical Directorate is gratefully acknowledged.

We thank the Rosetta Science Ground Segment at ESAC, the Rosetta Mission Operations Centre at ESOC and the Rosetta Project

at ESTEC for their outstanding work enabling the science return of the Rosetta Mission.

L.L.R. was the recipient of a CNES/Region Centre Ph.D. fellowship, and A.B. is the recipient of a CNES/LABEX ESEP Ph.D. fellowship in the frame of the ROSETTA COSIMA project.

The authors would like to thanks S. Moussaoui (Ecole centrale de Nantes) to have provided us the BPSS code.

References

- Baas, F., Geballe, T.R., Walther, D.M., 1986. Spectroscopy of the 3.4 μm emission feature in comet Halley. *Astrophys. J. Lett.* 311, L97–L101.
- Bockelée-Morvan, D., Crovisier, J., Mumma, M.J., Weaver, H.A., 2004. The composition of cometary volatiles. In: Festou, M.C., Keller, H.U., Weaver, H.A. (Eds.), *Comets II*. University of Arizona Press, Tucson, USA, pp. 391–423.
- Botta, O., Bada, J.L., 2002. Extraterrestrial organic compounds in meteorites. *Surv. Geophys.* 23, 411–467.
- Botta, O., Glavin, D.P., Kminek, G., Bada, J.L., 2002. Relative amino acid concentrations as a signature for parent body processes of carbonaceous chondrites. *Orig. Life Evol. Biosph.* 32, 143–163.
- Briani, G., Fray, N., Cottin, H., Benilan, Y., Gazeau, M.-C., Perrier, S., 2013. HMT production and sublimation during thermal process of cometary organic analogs. Implications for its detection with the ROSETTA instruments. *Icarus* 226, 541–551.
- Briggs, D., Brown, A., Vickerman, J.C., 1989. *Handbook of Static Secondary Ion Mass Spectrometry*. John Wiley & Sons Ltd, Chichester.
- Callahan, M.P., Smith, K.E., Cleaves, H.J., Ruzicka, J., Stern, J.C., Glavin, D.P., House, C.H., Dworkin, J.P., 2011. Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci.* 108, 13995–13998.
- Clairemidi, J., Bréchnignac, P., Moreels, G., Pautet, D., 2004. Tentative identification of pyrene as a polycyclic aromatic molecule in UV spectra of comet P/Halley: an emission from 368 to 384 nm. *Planet. Space Sci.* 52, 761–772.
- Clemett, S.J., Chillier, X.D.F., Gillette, S., Zare, R.N., Maurette, M., Engrand, C., Kurat, G., 1998. Observation of indigenous polycyclic aromatic hydrocarbons in 'Giant' carbonaceous antarctic micrometeorites. *Orig. Life Evol. Biosph.* 28, 425–448.
- Clemett, S.J., Maechling, C.R., Zare, R.N., Swan, P.D., Walker, R.M., 1993. Identification of complex aromatic molecules in individual interplanetary dust particles. *Science* 262, 721–725.
- Cody, G.D., Ade, H., O'D Alexander, C.M., Araki, T., Butterworth, A., Fleckenstein, H., Flynn, G., Gilles, M.K., Jacobsen, C., Kilcoyne, A.L.D., Messenger, K., Sandford, S.A., Tyliczszak, T., Westphal, A.J., Wirick, S., Yabuta, H., 2008. Quantitative organic and light-element analysis of comet 81P/Wild 2 particles using C-, N-, and O- μ -XANES. *Meteorit. Planet. Sci.* 43, 353–365.
- Coll, P., Coscia, D., Smith, N., Gazeau, M.-C., Ramirez, S.I., Cernogora, G., Israël, I.G., Raulin, F., 1999. Experimental laboratory simulation of Titan's atmosphere: aerosols and gas phase. *Planet. Space Sci.* 47, 1331–1340.
- Comon, P., Jutten, C., 2010. *Handbook of Blind Source Separation: Independent Component Analysis and Independent Component Analysis and Application*. Elsevier, Oxford.
- Cottin, H., Coll, P., Coscia, D., Fray, N., Guan, Y.Y., Macari, F., Raulin, F., Rivron, C., Stalport, F., Szopa, C., Chaput, D., Viso, M., Bertrand, M., Chabin, A., Thirkell, L., Westall, F., Brack, A., 2008. Heterogeneous solid/gas chemistry of organic compounds related to comets, meteorites, Titan and Mars: in laboratory and in lower Earth orbit experiments. *Adv. Space Res.* 42, 2019–2035.
- Cottin, H., Gazeau, M.C., Raulin, F., 1999. Cometary organic chemistry: a review from observations, numerical and experimental simulations. *Planet. Space Sci.* 47, 1141–1162.
- Danger, G., Orthous-Daunay, F.R., Marcellus, P.D., Modica, P., Vuitton, V., Duvernay, F., Flandinet, L., d'Hendecourt, L.L.S., Thissen, R., Chiavassa, T., 2013. Characterization of laboratory analogs of interstellar/cometary organic residues using very high resolution mass spectrometry. *Geochim. Cosmochim. Acta* 118, 184–201.
- Dartois, E., Engrand, C., Brunetto, R., Duprat, J., Pino, T., Quirico, E., Remusat, L., Bardin, N., Briani, G., Mostefaoui, S., Morinaud, G., Crane, B., Szwec, N., Delauche, L., Jamme, F., Ch, S., Dumas, P., 2013. UltraCarbonaceous Antarctic micrometeorites, probing the Solar System beyond the nitrogen snow-line. *Icarus* 224, 243–252.
- de Gregorio, B.T., Stroud, R.M., Cody, G.D., Nittler, L.R., David Kilcoyne, A.L., Wirick, S., 2011. Correlated microanalysis of cometary organic grains returned by Stardust. *Meteorit. Planet. Sci.* 46, 1376–1396.
- Delcorte, A., Bertrand, P., 1998. Sputtering of parent-like ions from large organic adsorbates on metals under keV ion bombardment. *Surf. Sci.* 412–413, 97–124.
- Despois, D., Cottin, H., 2005. Comets: potential sources of prebiotic molecules for the early earth. In: Heidelberg, S.B. (Ed.), *Lectures in Astrobiology*.
- Dobrica, E., Engrand, C., Quirico, E., Montagnac, G., Duprat, J., 2011. Raman characterization of carbonaceous matter in CONCORDIA Antarctic micrometeorites. *Meteorit. Planet. Sci.* 46, 1363–1375.
- Duprat, J., Dobrica, E., Engrand, C., Aléon, J., Marrocchi, Y., Mostefaoui, S., Meibom, A., Leroux, H., Rouzaud, J.-N., Gounelle, M., Robert, F., 2010. Extreme deuterium excesses in ultracarbonaceous micrometeorites from Central Antarctic snow. *Science* 328, 742–745.

- Elsila, J.E., Glavin, D.P., Dworkin, J.P., 2009. Cometary glycine detected in samples returned by Stardust. *Meteorit. Planet. Sci.* 44, 1323–1330.
- Fister, T.F., Strossman, G.S., Willett, K.L., Odom, R.W., Linton, R.W., 1995. In situ analysis of organic monolayers and their reactivity on single micrometer-sized particles by time-of-flight secondary ion mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* 143, 87–111.
- Fomenkova, M.N., 1999. On the organic refractory component of cometary dust. *Space Sci. Rev.* 90, 109–114.
- Glavin, D.P., Callahan, M.P., Dworkin, J.P., Elsila, J.E., 2010. The effects of parent body processes on amino acids in carbonaceous chondrites. *Meteorit. Planet. Sci.* 45, 1948–1972.
- Gounelle, M., Spurn, P., Bland, P.A., 2006. The orbit and atmospheric trajectory of the Orgueil meteorite from historical records. *Meteorit. Planet. Sci.* 41, 135–150.
- Greenberg, J.M., 1982. What are comets made of? A model based on interstellar dust. In: Wilkening, L.L. (Ed.), *Comets*. University of Arizona Press, Tucson, pp. 131–163.
- Hayatsu, R., 1964. Orgueil meteorite: organic nitrogen contents. *Science* 146, 1291–1293.
- Hayatsu, R., Studier, M., Oda, A., Fuse, K., Anders, E., 1968. Origin of organic matter in the early solar system-II. Nitrogen compounds. *Geochim. Cosmochim. Acta* 32, 175–190.
- Hayatsu, R., Studier, M.H., Moore, L.P., Anders, E., 1975. Purines and triazines in the Murchison meteorite. *Geochim. Cosmochim. Acta* 39, 471–488.
- Herd, C.D.K., Blinova, A., Simkus, D.N., Huang, Y., Tarozo, R., Alexander, C.M.O.D., Gyngard, F., Nittler, L.R., Cody, G.D., Fogel, M.L., Kebukawa, Y., Kilcoyne, A.L.D., Hilt, R.W., Slater, G.F., Glavin, D.P., Dworkin, J.P., Callahan, M.P., Elsila, J.E., De Gregorio, B.T., Stroud, R.M., 2011. Origin and evolution of prebiotic organic matter as inferred from the Tagish Lake Meteorite. *Science* 332, 1304–1307.
- Hilchenbach, M., Fischer, H., Krüger, H., Thirkell, L., Rynö, J., 2014. Compounds observed in-flight by secondary ion mass spectrometry with ROSETTA-COSIMA before the encounter with comet 67P/Churyumov-Gerasimenko. *Nucl. Instrum. Methods Phys. Res. B* (in preparation).
- Hughley, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., Qian, K., 2001. Kendrick Mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* 73, 4676–4681.
- Israel, G., Szopa, C., Raulin, F., Cabane, M., Niemann, H.B., Atreya, S.K., Bauer, S.J., Brun, J.-F., Chassefière, E., Coll, P., Condé, E., Coscia, D., Hauchecorne, A., Millian, P., Nguyen, M.-J., Owen, T., Riedler, W., Samuelsen, R.E., Siguier, J.-M., Steller, M., Sternberg, R., Vidal-Madjar, C., 2005. Complex organic matter in Titan's atmospheric aerosols from in situ pyrolysis and analysis. *Nature* 438, 796–799.
- Jessberger, E.K., 1999. Rocky cometary particulates: their elemental, isotopic and mineralogical ingredients. *Space Sci. Rev.* 90, 91–97.
- Jessberger, E.K., Cristoforidis, A., Kissel, J., 1988. Aspects of the major element composition of Halley's dust. *Nature* 332, 691–695.
- Jessberger, E.K., Kissel, J., Neugebauer, M., Rahe, J., 1991. Chemical properties of cometary dust and a note on carbon isotopes. In: Newburn, R.L. Jr. (Ed.), *IAU Colloq. 116: Comets in the post-Halley era*, pp. 1075–1092.
- Kendrick, E., 1963. A mass scale based on $\text{CH}_2 = 14.0000$ for high resolution mass spectrometry of organic compounds. *Anal. Chem.* 35, 2146–2154.
- Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Anal. Chem.* 75, 5336–5344.
- Kissel, J., Altwegg, K., Clark, B.C., Colangeli, L., Cottin, H., Czempel, S., Eibl, J., Engrand, C., Fehringer, H.M., Feuerbacher, B., Fomenkova, M., Glasmachers, A., Greenberg, J.M., Grün, E., Haerendel, G., Henkel, H., Hilchenbach, M., von Hoerner, H., Höfner, H., Hornung, K., Jessberger, E.K., Koch, A., Krüger, H., Langevin, Y., Parigger, P., Raulin, F., Rüdener, F., Rynö, J., Schmid, E.R., Schulz, R., Silén, J., Steiger, W., Stephan, T., Thirkell, L., Thomas, R., Torkar, K., Utterback, N.G., Varmuza, K., Wanczek, K.P., Werther, W., Zschege, H., 2007. Cosima: high resolution time-of-flight secondary ion mass spectrometer for the analysis of cometary dust particles onboard Rosetta. *Space Sci. Rev.* 128, 823–867.
- Kissel, J., Krueger, F.R., 1987. The organic component in dust from comet Halley as measured by the PUMA mass spectrometer on board Vega 1. *Nature* 326, 755–760.
- Kissel, J., Krueger, F.R., Silén, J., Clark, B.C., 2004. The cometary and interstellar dust analyzer at Comet 81P/Wild 2. *Science* 304, 1774–1776.
- Klein, H.P., 1978. The Viking biological experiments on Mars. *Icarus* 34, 666–674.
- Krüger, H., Baklouti, D., Briois, C., Engrand, C., Hilchenbach, M., Merouane, S., Siljeström, S., Stephan, T., Evaluation of inorganic compounds with the laboratory instrument of COSIMA, this issue, (in preparation).
- Le Roy, L., Briani, G., Briois, C., Cottin, H., Fray, N., Thirkell, L., Poulet, G., Hilchenbach, M., 2012. On the prospective detection of polyoxymethylene in comet 67P/Churyumov-Gerasimenko with the COSIMA instrument onboard Rosetta. *Planet. Space Sci.* 65, 83–92.
- Leggett, G.J., Vickerman, J.C., Briggs, D., Hearn, M.J., 1992. Surface studies by static secondary ion mass spectrometry: cluster ion formation studied by tandem mass-spectrometric techniques. *J. Chem. Soc. Faraday Trans.* 88, 297–309.
- Leshin, L., Mahaffy, P., Webster, C., Cabane, M., Coll, P., Conrad, P., Archer, P., Atreya, S., Brunner, A., Buch, A., Eigenbrode, J., Flesch, G., Franz, H., Freissinet, C., Glavin, D., McAdam, A., Miller, K., Ming, D., Morris, R., Navarro-González, R., Niles, P., Owen, T., Pepin, R., Squyres, S., Steele, A., Stern, J., Summons, R., Sumner, D., Sutter, B., Szopa, C., Teinturier, S., Trainer, M., Wray, J., Grotzinger, J., Team, M.S.L.S., 2013. Volatile, isotope, and organic analysis of martian fines with the Mars Curiosity rover. *Science (N.Y.)* 341, 1238937.
- Lodders, K., 2003. Solar System abundances and condensation temperatures of the elements. *Astrophys. J.* 591, 1220–1247.
- Mahoney, C.M., Gillen, G., Fahey, A.J., 2006. Characterization of gunpowder samples using time-of-flight secondary ion mass spectrometry (TOF-SIMS). *Forensic Sci. Int.* 158, 39–51.
- Marcano, V., Benitez, P., Palacios-Prü, E., 2003. Acyclic hydrocarbon environments $\geq n\text{-C18}$ on the early terrestrial planets. *Planet. Space Sci.* 51, 159–166.
- Martins, Z., Botta, O., Fogel, M.L., Sephton, M.A., Glavin, D.P., Watson, J.S., Dworkin, J.P., Schwartz, A.W., Ehrenfreund, P., 2008. Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* 270, 130–136.
- Martins, Z., Hofmann, B.A., Gnos, E., Greenwood, R.C., Verchovsky, A., Franchi, I.A., Jull, A.J.T., Botta, O., Glavin, D.P., Dworkin, J.P., Ehrenfreund, P., 2007. Amino acid composition, petrology, geochemistry, 14C terrestrial age and oxygen isotopes of the Shիր 033 CR chondrite. *Meteorit. Planet. Sci.* 42, 1581–1595.
- Martins, Z.M., Sephton, A., 2009. Extraterrestrial amino acids. In: Hughes, A.B. (Ed.), *Amino Acids, Peptides and Proteins in Organic Chemistry*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 1–42.
- Matrajt, G., Pizzarello, S., Taylor, S., Brownlee, D., 2004. Concentration and variability of the AIB amino acid in polar micrometeorites: implications for the exogenous delivery of amino acids to the primitive Earth. *Meteorit. Planet. Sci.* 39, 1849–1858.
- Matrajt, G., Taylor, S., Flynn, G., Brownlee, D., Joswiak, D., 2003. A nuclear microprobe study of the distribution and concentration of carbon and nitrogen in Murchison and Tagish Lake meteorites, Antarctic micrometeorites, and IDPs: implications for astrobiology. *Meteorit. Planet. Sci.* 38, 1585–1600.
- Ming, D., Archer, P., Glavin, D., Eigenbrode, J., Franz, H., Sutter, B., Brunner, A., Stern, J., Freissinet, C., McAdam, A., Mahaffy, P., Cabane, M., Coll, P., Campbell, J., Atreya, S., Niles, P., Bell, J., Bish, D., Brinckerhoff, W., Buch, A., Conrad, P., Des Rauxes, D., Ehlmann, B., Fairén, A., Farley, K., Flesch, G., Francois, P., Gellert, R., Grant, J., Grotzinger, J., Gupta, S., Herkenhoff, K., Hurowitz, J., Leshin, L., Lewis, K., McLennan, S., Miller, K., Moersch, J., Morris, R., Navarro-González, R., Pavlov, A., Perrett, G., Pradler, I., Squyres, S., Summons, R., Steele, A., Stolper, E., Sumner, D., Szopa, C., Teinturier, S., Trainer, M., Treiman, A., Vaniman, D., Vasavada, A., Webster, C., Wray, J., Yingst, R., Team, M.S.L.S., 2014. Volatile and organic compositions of sedimentary rocks in Yellowknife Bay, Gale crater, Mars. *Science (N.Y.)* 343, 1245267.
- Mitchell, D.L., Lin, R.P., Carlson, C.W., Korth, A., Rème, H., Mendis, D.A., 1992. The origin of complex organic ions in the coma of comet Halley. *Icarus* 98, 125–133.
- Moreels, G., Clairemidi, J., Hermine, P., Brechignac, P., Rousselot, P., 1994. Detection of a polycyclic aromatic molecule in comet P/Halley. *Astron. Astrophys.* 282, 643–656.
- Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C., 2006a. Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. *IEEE Trans. Signal Process.* 54, 4133–4145.
- Moussaoui, S., Carteret, C., Brie, D., Mohammad-Djafari, A., 2006b. Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods. *Chemom. Intell. Lab. Syst.* 181, 137–148.
- Moussaoui, S., Carteret, C., Brie, D., Mohammad-Djafari, A., 2006c. Bayesian analysis of spectral mixture data using MCMC methods. *Chemom. Intell. Lab. Syst.* 81, 137–148.
- Mumma, M.J., Charnley, S.B., 2011. The chemical composition of Comets – emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* 49, 471–524.
- Niemann, H.B., Atreya, S.K., Bauer, S.J., Carignan, G.R., Demick, J.E., Frost, R.L., Gautier, D., Haberman, J.A., Harpold, D.N., Hunten, D.M., Israel, G., Lunine, J.I., Kasprzak, W.T., Owen, T.C., Paulkovich, M., Raulin, F., Raaen, E., Way, S.H., 2005. The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438, 779–784.
- Niemann, H.B., Atreya, S.K., Demick, J.E., Gautier, D., Haberman, J.A., Harpold, D.N., Kasprzak, W.T., Lunine, J.I., Owen, T.C., Raulin, F., 2010. Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini-Huygens probe gas chromatograph mass spectrometer experiment. *J. Geophys. Res. – Planets* 115, E12006.
- Pizzarello, S., Huang, Y., Becker, L., Poreda, R.J., Nieman, R.A., Cooper, G., Williams, M., 2001. The organic content of the Tagish Lake Meteorite. *Science* 293, 2236–2239.
- Sandford, S.A., Aleon, J., Alexander, C.M.O.D., Araki, T., Bajt, S., Baratta, G.A., Borg, J., Bradley, J.P., Brownlee, D.E., Brucato, J.R., Burchell, M.J., Busemann, H., Butterworth, A., Clemett, S.J., Cody, G., Colangeli, L., Cooper, G., D'Hendecourt, L., Djouadi, Z., Dworkin, J.P., Ferrini, G., Fleckenstein, H., Flynn, G.J., Franchi, I.A., Fries, M., Gilles, M. K., Glavin, D.P., Gounelle, M., Grossemy, F., Jacobsen, C., Keller, L.P., Kilcoyne, A.L.D., Leitner, J., Matrajt, G., Meibom, A., Mennella, V., Mostefaoui, S., Nittler, L.R., Palumbo, M.E., Papanastassiou, D.A., Robert, F., Rotundi, A., Sneed, C.J., Spencer, M.K., Stadermann, F.J., Steele, A., Stephan, T., Tsou, P., Tylliszczak, T., Westphal, A.J., Wirrick, S., Wopenka, B., Yabuta, H., Zare, R.N., Zolensky, M.E., 2006. Organics captured from Comet 81P/Wild 2 by the Stardust spacecraft. *Science* 314, 1720–1724.
- Sandford, S.A., Bajt, S.A., Clemett, S.J., Cody, G.D., Cooper, G., Degregorio, B.T., de Vera, V., Dworkin, J.P., Elsila, J.E., Flynn, G.J., Glavin, D.P., Lanzirrotti, A., Limerio, T., Martin, M.P., Sneed, C.J., Spencer, M.K., Stephan, T., Westphal, A., Wirrick, S., Zare, R.N., Zolensky, M.E., 2010. Assessment and control of organic and other contaminants associated with the Stardust sample return from comet 81P/Wild 2. *Meteorit. Planet. Sci.* 45, 406–433.
- Schmitt-Koppin, P., Gabelica, Z., Gougeon, R.D., Fekete, A., Kanawati, B., Harir, M., Gebeffugi, I., Eckel, G., Hertkorn, N., 2010. High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc. Natl. Acad. Sci.* 107, 2763–2768.
- Sephton, M.A., 2002. Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep.* 19, 292–311.
- Sephton, M.A., Love, G.D., Watson, J.S., Verchovsky, A.B., Wright, I.P., Snape, C.E., Gilmour, I., 2004. Hydrolysis of insoluble carbonaceous matter in the

- Murchison meteorite: new insights into its macromolecular structure. *Geochimica et Cosmochimica Acta* 68, 1385–1393.
- Spencer, M.K., Zare, R.N., 2007. Comment on “organics captured from Comet 81P/Wild 2 by the Stardust spacecraft”. *Science* 317, 1680.
- Spool, A.M., 2004. Interpretation of static secondary ion spectra. *Surf. Interface Anal.* 36, 264–274.
- Stoks, P.G., Schwartz, A.W., 1979. Uracil in carbonaceous meteorites. *Nature* 282, 709–710.
- Stoks, P.G., Schwartz, A.W., 1981. Nitrogen-heterocyclic compounds in meteorites: significance and mechanisms of formation. *Geochim. Cosmochim. Acta* 45, 563–569.
- Toporski, J., Steele, A., 2004. Characterization of purified biomarker compounds using time of flight-secondary ion mass spectrometry (ToF-SIMS). *Org. Geochem.* 35, 793–811.
- Van der Velden, W., Schwartz, A.W., 1977. Search for purines and pyrimidines in Murchison meteorite. *Geochim. Cosmochim. Acta* 41, 961–968.
- Varmuza, K., Werther, W., Krueger, F.R., Kissel, J., Schmid, E.R., 1999. Organic substances in cometary grains: comparison of secondary ion mass spectral data and californium-252 plasma desorption data from reference compounds. *Int. J. Mass Spectrom.* 189, 79–92.
- Waite, J.H., Young, D.T., Cravens, T.E., Coates, A.J., Crary, F.J., Magee, B., Westlake, J., 2007. The process of tholin formation in Titan's upper atmosphere. *Science* 316, 870–875.