Tholinomics-chemical analysis of nitrogen-rich polymers

Pascal Pernot, Nathalie Carrasco, Roland Thissen, Isabelle Schmitz-Afonso

Tholinomics – Chemical analysis of nitrogen-rich polymers

Pascal Pernot*

Laboratoire de Chimie Physique

UMR 8000, CNRS, Université Paris-Sud 11

91405 Orsay cedex, France

Email: pascal.pernot@lcp.u-psud.fr

Nathalie Carrasco

Laboratoire Atmosphères, Milieux, Observations Spatiales

Université de Versailles Saint-Quentin, UMR 8190

91371 Verrières le Buisson cedex, France

Roland Thissen

Laboratoire de Planétologie de Grenoble,

UMR 5109, CNRS, Université Joseph Fourrier

38041 Grenoble cedex 9, France

Isabelle Schmitz-Afonso

Centre de Recherche de Gif, Institut de Chimie des Substances Naturelles,

CNRS, avenue de la Terrasse,

91198 Gif-sur-Yvette cedex, France

Abstract

The polymeric composition of Titan's tholins – laboratory analogues of Titan's aerosols – is elucidated using high resolution mass spectrometry. This complex organic matter is produced by plasma discharge in a gaseous nitrogen-methane mixture and analyzed with a hybrid linear trap/orbitrap mass-spectrometer. The highly structured mass spectra are treated with tools developed for petroleomics (Kendrick and van Krevelen diagrams), with original adaptations for nitrogen-rich compounds. Our goal is to find the best chemical basis set to describe the compositional space that these polymers occupy, in order to shed light onto the chemical structure of tholins. We succeeded in assigning the molecules identified in the mass spectra of tholins to a small number of regularly distributed X-$(CH_2)_m(HCN)_n$ families, where the balanced copolymer ($m = n$) is found to play a central role. Within each family, the polymer lengths n and m present Poisson-type distributions. We also identify the smallest species of a subset of families as linear and cyclic amino nitrile compounds of great astrobiological interest: biguanide, guanidin, acetamidine, aminoacetonitrile, methylimidazole.

1 Introduction

Titan, the largest moon of Saturn, has a dense atmosphere mainly composed of nitrogen (98 %), methane, molecular hydrogen and traces of hydrocarbons. Chemical growth is initiated by the activation of N_2 and CH_4 by solar UV light and ionizing particles precipitations.[1,2] It leads to the production of macroscopic particles which sediment slowly in the atmosphere, forming the brownish haze surrounding permanently Titan.[3] Composition of these aerosols, their formation and growth pathways are poorly understood, because of the difficulty of Earth-based observations of Titan's upper atmosphere and the limited resolution of chemical analyzes performed by space probes.[4]

The knowledge about Titan's aerosols is improved by the study of analogous materials produced in the laboratory, the so-called Titan's *tholins*. These are most often synthesized by maintaining a continuous plasma discharge in gaseous N_2-CH_4 mixtures.[5–7] Some experiments have also been driven using UV photons as energy deposit,[8,9] but they did not allow a significant production of tholin. The PAMPRE reactor[7] used here is based on a Radio Frequency capacitively coupled plasma. This original setup is designed for atmospheric simulation, as the production and growth of tholin occur directly in levitation inside the confined reactive plasma, avoiding any contact of tholin with a solid support during the growth process. In this work two different tholin samples have been generated by changing the relative abundance of N_2/CH_4: either 98/2 % (SA98) or 90/10 % (SA90).

The aim of this paper is to understand the polymeric structure of tholins produced in the PAMPRE reactor and analyzed by high resolution mass spectrometry. The polymeric nature of tholin has already been suggested by Sarker *et al.*[10] and Somogyi *et al.*,[11] where species composing their samples are sorted in seventeen $X(CH_2)_nN_x$ series, defined by a hydrocarbon residue, X, and a number of nitrogen atoms, N_x (for instance Series 1 is $C_6H_8(CH_2)_nN_3$, $n = 0, 9$). Concerning the chemical signature of nitrogen, Sarker *et al.*[10] have shown that N is the dominant carrier of insaturation in their tholin material. Moreover, all elemental analysis performed on tholins have shown a large content of nitrogen, with N/C ratios sometimes reaching the value of one.[6,12,13] The series classification of Sarker *et al.*[10] uses the CH_2 unit, a classical brick in petroleum organic matter,[14] but it does not offer a systematic unit for N insertion. It appears to us essential to understand the molecular growth of tholin components from the nitrogen insertion point of view.

Carrasco *et al.*[15] have shown that the analysis of the soluble fractions dissolved in a polar solvent gave valuable insights on the bulk tholin composition. In this earlier work, the highly structured nature of PAMPRE tholins, characteristic of a polymeric matter, has also been shown by atmospheric pressure photoionization time-of-flight mass spectrometry (APPI-TOF). Further tandem mass spectrometry on a dozen of negatively charged ions with m/z varying from 66 to 310 allowed to identify one neutral common brick, NC-NH-CN, which unveiled the importance of the C / N alternation in Titan's tholins. However,

with Q-Tof instrument, the mass resolution, m/z 0.03 (FWHM) at m/z 150 (which corresponds to a resolving power of 5000) did not allow to unambiguously identify the chemical formula of all the molecules. The present work solves this undetermination. We have studied the chemical composition of methanol extracts of tholin by ESI-FT-Orbitrap mass spectrometry, at high resolving power ($\frac{M}{\triangle M}$ $\geq$100 000 at m/z $\leq$400).

In order to identify the main mass patterns constituting the tholin, we have supported our analysis of the mass spectra by extending visual and statistical representations developed for *petroleomics* – Kendrick and van Krevelen diagrams.[14,16–21] Standard van Krevelen diagrams have been used for tholin analysis by Imanaka and Smith.[22] The term *tholinomics* has been coined here to account for the specificity of such nitrogen rich $C_xH_yN_z$ polymeric species, which, beside Titan's tholins, encompass other species of astrochemical interest as polycyclic aromatic nitrogen heterocycles (PANHs),[23] or of industrial interest as carbon nitrides.[24,25]

2 Methods

2.1 Tholin production in a RF dusty plasma: the PAMPRE experimental setup

The apparatus used in this experiment was described in details previously[7] and is summarized here briefly. The plasma is a capacitively coupled plasma discharge produced at 13.56 MHz frequency in a N_2-CH_4 gaseous mixture. The plasma is confined in a cylindrical cage shaped by a metallic grid. In the plasma discharge, electrons dissociate and ionize N_2 and CH_4. This initiates chemical reactions and molecular growth producing hydrocarbons and nitrogen-bearing molecules which eventually end up forming solid particles. The produced solid particles are negatively charged and maintained in levitation between the electrodes by electrostatic forces. The N_2-CH_4 gaseous mixture is injected continuously into the plasma reactor, as a neutral flow oriented downward. This produces a neutral drag force, which ejects the solid particles out of the plasma discharge. The ejected tholin is trapped in a glass vessel surrounding the metallic cage. When sufficient amounts of tholin are produced (i.e. typically after 8 hours), the plasma is turned off; the reactor is thoroughly pumped to evacuate potential traces of HCN and other residual gases, then filled-up with N_2 to the atmospheric pressure and opened for solid sample collection. When operating in normal conditions, about 0.5 cm^3 ($\sim$ 100 mg) of solid material, is produced in a 8 hours run. The tholins have the appearance of a very fine orange to brown powder, deposited gently in the glass vessel without any interaction with the substrate. They are collected into micro-vials for ex-situ analyzes. Because of the plausible toxicity of tholin, gloves and nanofiltering masks are used during open air manipulation of the material.

In this work, the samples were obtained in the following operating conditions: a total pressure of 1 mbar, a flow rate of 55 ± 0.1 sccm and an absorbed radio frequency power of 30 ± 2 W. Two different samples were produced by adjusting the relative flux of N_2 *vs.* CH_4. The first sample (SA98) was produced in a mixture containing 2.00 ± 0.06 % of methane, while the second sample (SA90) was produced in a mixture containing 10.0 ± 0.2 % of methane.

2.2 High resolution mass spectrometry of the tholins by the orbitrap technique

Methanol soluble part of the tholin samples were prepared based on the following procedure: samples were dissolved in methanol at a concentration of 1 mg.mL^{-1}, mixed using ultrasonic bath, centrifuged and finally the supernatant recuperated for further analysis. The choice of methanol as a solvent is consistent with our previous specific study on tholins solubility.[15] In this polar solvent, the soluble mass fraction of SA98 and SA90 is about 19 and 35 % respectively. Samples were infused in the electrospray (ESI) ionization source in MeOH/water (80/20, v/v) at 10 μL.min^{-1}. The advantage of the ESI ionization technique is that it doesn't fragment the tholin molecules: the species are only protonated and detected as $[M+H]^+$, when operated in the positive ionization mode.[26] A hybrid linear trap/orbitrap mass spectrometer (LTQ orbitrap) was used for measurements. Mass spectra were acquired using the following conditions: needle voltage 5 kV; heated capillary temperature 275 °C; tube lens voltage 65 V. In such an instrument, ions are collected in the linear ion trap, and can be axially ejected and collected in a C-shaped ion trap (so called C-trap) to be finally injected in the Fourier Transform Orbitrap (FT-orbitrap).[27,28] In the orbitrap, ions assume circular trajectories around a center electrode and their axial oscillations along this center electrode produce a periodic signal on the outer electrodes, which is detected and converted into a spectrum using a Fourier transform algorithm. The mass spectrometer was externally calibrated using a frequency/mass Lmethionine-arginyl-phenylalanyl-alanine, caffeine and ultramark standards mixture. Data were acquired from 50 to 800 u, integrated on 27 scans, processed using XCalibur 2.0 software and saved to ASCII file for further analysis. Data were acquired with a resolution chosen at 100 000. Experimental resolution was around $\frac{M}{\triangle M} = 200\,000$ at m/z 150 and around 100 000 at m/z 450. In these conditions, we estimate that no reliable assignment can be done above m/z 500.

2.3 Data analysis

Peak maxima were located by parabolic interpolation over three peak-bracketing data points. Only peaks above a threshold intensity I_{thresh} were considered (see 3.1). A reference composition database was generated containing all m/z values for combinations $C_xH_yN_zO_w$ within the spectrum mass range, with the following restrictions: (1) the "nitrogen rule" for positive ions generated by ESI: an ion of even

(odd) mass must contain an odd (even) number of nitrogen atoms; and (2) in order to limit the size of the reference database, a maximum number $w = 5$ of O atoms, found sufficient to assign all peaks, was used. For each peak, its position was compared to the database elements, and the composition with the nearest m/z value was retained. Peaks with residuals from the nearest composition above 5 ppm were left unassigned and excluded from the analysis. The window of $\pm$ 5 ppm corresponds to the instrument specifications in external calibration mode, which was used here to calibrate the instrument. Attributions were picked randomly and checked successfully against other formula attribution algorithms.

The presence of oxygenated species in our samples might result from water exposure during the experiment[11] or from the instantaneous oxidation of the tholins when opening the reactor for collection.[29–31] As there is strong evidence that the formation of oxygenated compounds is due to air exposure during the harvest of tholin, these compounds were left out from the present study. Identification of ^{13}C-containing species was also tested and resulted into a negligible contribution: very few peaks reached an intensity level above the chosen threshold intensity and those peaks were rejected.

2.3.1 Kendrick plots and van Krevelen diagrams

Kendrick plots and van Krevelen diagrams are two graphical representations which have been intensively used in the past years for the rational analysis of mass spectra of complex organic matter.[14,16–21,32–35] We present them briefly as a starting point for alternative representations developed in the present article.

Kendrick plots do not require chemical formula assignment. The principle is to choose a basic pattern (e.g. CH_2) and to calculate the mass defects of all peaks with regard to this pattern. Plotting of the Kendrick Mass Defect of each peak as a function of its mass reveals series of homologous compounds as horizontal lines. For a chosen pattern M, one defines the "Kendrick mass" as

$$\text{Kendrick mass} = \text{IUPAC mass} \times (\text{M nominal mass}/\text{M IUPAC mass}) \tag{1}$$

In this new mass scale, compounds series with formula X-M_n, with the same X root, will have identical Kendrick Mass Defect (KMD)

$$KMD = (\text{nominal Kendrick mass} - \text{exact Kendrick mass}) \times 1000 \tag{2}$$

The CH_2 pattern is frequently used for crude oil analysis,[33] but other patterns, as carboxylate[36,37] and oxygen[38] have been used in the study of natural organic matter. For tholin, nitrogen containing groups seem more appropriate and are tested in the following.

The van Krevelen plot requires composition assignment to all peaks and provides a convenient representation of elemental compositions by plotting, for instance, the H/C ratio vs. N/C ratio for each

molecule.[16] As it normalizes compositions to the carbon number, the standard 2D van Krevelen plot discards a lot of information, notably the molecular mass. Three dimensional (3D) van Krevelen plots have been proposed to introduce additional information. The third axis might represent O/C ratios,[19] or peak intensities.[33] In the following, we show that using the molecular mass as the third dimension can also be very helpful to reveal underlying structures of the samples.

3 Analysis of the orbitrap mass spectra of tholins

3.1 Orbitrap mass spectra for SA98 and SA90 samples

The spectra of SA98 and SA90 samples are shown in Fig. 1. The differences between the two spectra are similar to the ones observed for APPI-TOF spectra in Carrasco *et al.*:[15] the solid sample produced in a reactive medium with more methane (SA90) gives rise to a larger abundance of heavy molecules (>100 u). The high resolution of the orbitrap reveals a large number of fine peaks for each mass unit (Fig. S-1) and enables us to assign a $C_xH_yN_zO_w$ formula to mass peaks, the contribution of isotopes being negligible, as stated previously. Such assignments have also been made on other Titan's aerosol analogues produced in a AC plasma discharge with ESI-FT-ICR mass spectrometry for m/z above 120.[10,11]

Considering the limits in resolution and accuracy of the acquired mass spectra, the analysis of the spectra was limited to values below m/z 500. The threshold is set to $I_{thresh} = 4000$, corresponding to about 5σ of baseline noise. This value is a compromise, chosen to limit the assignments to $^{12}C_x{}^{1}H_y{}^{14}N_z$ isotopic species, without discarding too many peaks of interest. The threshold is to be compared to a maximum peak intensity of about 6×10^5 for SA90 and 3.5×10^6 for SA98. For SA90, these limits enable to locate 2465 peaks, amongst which 2404 (98 %) were identified, corresponding to 99 % of the cumulated peak intensity (CPI; the CPI is the sum of intensities of a set of peaks). For the SA98 sample, 1321 peaks were located, 1133 (86 %) assigned, corresponding to 96 % of the CPI. The aim of this exploratory work being to outline the major $C_xH_yN_z$ features of the mass spectra, we did not attempt a more exhaustive assignment.

Peaks attributed to O or ^{13}C bearing species are not considered in the following and are reported for reference with negative intensities in Fig.1.

[Figure 1 about here.]

The same global shape and structure of the spectrum is observed for the present orbitrap and previous FT-ICR spectra.[10,11] The broad distribution found in Sarker *et al.*[10] for a solid sample produced in a 2 % of methane gaseous mixture is surprisingly more in agreement with the SA90 orbitrap spectrum than

with the SA98 one. However the information given by the relative intensities of the peaks in the mass spectra is not directly correlated with the molecules concentrations and has to be considered with caution. In such a complex mixture, different ionization efficiencies occur for the various compounds.[39] Effects of solubility, proton affinity, and instrument transmission function might also greatly affect the general aspect of spectra recorded on different instruments. Moreover, the two studies used tholins produced in different plasma setups, leading probably to different methane dissociation ratios within the reactive medium.

In their samples, Sarker *et al.*[10] observed mass clusters regularly spaced by 14 u. SA90 and SA98 present also regular clusters, but with a 13.5 u spacing. This large difference reveals an important discrepancy between the chemical compositions of tholin produced by both methods that will be discussed in a forthcoming paper. This 14 u interval allowed Sarker *et al.*[10] to identify a polymeric CH_2 growth pattern in their tholins, a pattern well-known in the mass spectrometry of organic matter.[14] This CH_2 pattern is also observed in our case (Fig. S-1), but, considering the 13.5 u spacing, it is not the unique inter-cluster transition pattern.

3.2 Identification of HCN and C_2H_3N as possible nitrogen growth units

Our aim was to identify additional representative polymeric patterns involving nitrogen. The tandem MS/MS analysis of Somogyi *et al.*[11] and Carrasco *et al.*[15] on a few compounds highlighted recurrent HCN (-27) and C_2H_3N (-41) neutral losses. These unsaturated patterns would be compatible with the observation, made by Sarker *et al.*,[10] of nitrogen as unsaturation carrier. It has been proposed by Coll *et al.*[6] that poly-(HCN), commonly identified as prebiotic and extraterrestrial macromolecules,[40] could be a base for Titan's tholins. Quirico *et al.*[41] have shown that despite infrared and Raman spectroscopic differences, poly-(HCN) and tholin present interesting similarities. We considered thus HCN and C_2H_3N as possible polymeric patterns and tested them through Kendrick Mass Defect analysis.

The standard Kendrick plot based on the CH_2 pattern was compared to the alternative patterns, HCN and C_2H_3N (Fig. 2). The CH_2 and HCN patterns produce Kendrick Mass Defects (KMD) of opposite signs and with iso-KMD families of comparable lengths. An interesting point about the C_2H_3N pattern, is that it produces a rather symmetric KMD distribution, with a much smaller dispersion than the other two. The KMD represents a distance between the observed ions and a potential polymeric backbone. We are unaware of any study about the optimal choice of Kendrick patterns, but we propose that compacity and symmetry properties of the KMD distribution are indicative of a better pattern, *i.e.* one which is more representative of the global sample's composition and readily highlight slight variations between samples.

The main conclusion of the KMD analysis is that a combination of CH_2 and HCN plays an important

role in the structural composition of the studied samples.

[Figure 2 about here.]

3.3 Validation of a copolymer X-$(CH_2)_m(HCN)_n$ structure

The 2D ($N/C; H/C$) van Krevelen diagram of sample SA90 summarizes the compositional analysis (Fig. 3). Nearly all compounds contain nitrogen and most are quite unsaturated ($H/C \lesssim 2$), with a few exceptions. This diagram is rather symmetric, with two interesting features: the (0;2)-(1;1) and (0;1)-(1;2) orthogonal axes. The first axis defines the pure copolymer $(CH_2)_m(HCN)_n$ family and the point (0.5;1.5) corresponds to $m = n$, confirming the role of this family evidenced in the KMD analysis. Globally, there are only three salient groups in the van Krevelen plot, the $(CH_2)_m(HCN)_n$ family and one group on each side of it. One observes also converging beams toward the (0;2) point representing pure $(CH_2)_m$ species and toward the (1;1) point representing the pure $(HCN)_n$ ones, an additional hint to the role of both units. In the spirit of Principal Components Analysis of multidimensional data,[42,43] the symmetry of the van Krevelen plot suggests that a rotation of -45 degrees of the ($N/C; H/C$) axes would probably provide a more pertinent basis for the compositional analysis of SA90 (see next section).

We tried to ease the dense 2D van Krevelen diagram by using a 3D representation ($N/C; H/C; m/z$), which was instrumental in revealing the underlying structure of the compositional space (Fig. 3). The representation of the sample in the 3D space, with a viewing line along the $(CH_2)_m(HCN)_n$ family, lets appear a stratified structure. We clearly observe a splitting of the compositions into distinct groups at the lower masses. These groups lie on hypersurfaces, which converge towards the $(CH_2)_m(HCN)_n$ family as mass increases. In the 3D representation, the pure copolymer family is a vertical plane which is projected on the (0;2)-(1;1) axis of the 2D van Krevelen plot. Other families are projected in the groups on either side of this line. In each family, the heavier species are located near the pure copolymer line, and the lighter species farther apart. This convergence is the source of the apparent complexity of the 2D van Krevelen plot.

These observations are diagnostic of the presence of a set of families of the type X-$(CH_2)_m(HCN)_n$, which converge into van Krevelen compositional space as $n + m$ increases.

[Figure 3 about here.]

3.4 Decomposition of $C_xH_yN_z$ into C_α-$(CH_2)_m(HCN)_n$ families

All N-bearing species $C_xH_yN_z$ can formally be expressed as C_α-$(CH_2)_m(HCN)_n$. This representation is particularly interesting to codify our previous observations on the symmetric role of CH_2 and HCN in SA90 and SA98. It is done in three steps:

1. n is the number of N atoms, which are systematically attributed to HCN units;

2. $m = (H - N)/2$ is obtained considering that all H atoms but in $(HCN)_n$ belong to CH_2 subunits (the "nitrogen rule" ensures the evenness of this remaining number of H atoms);

3. α provides the number of carbon atoms completing the pure $(CH_2)_m(HCN)_n$ copolymer: $\alpha = C - (H + N)/2$, with α taking positive (carbon excess) or negative (carbon defect) values.

A new type of representation is proposed to implement this copolymeric decomposition. The first axis is the Carbon Excess Index (CEI, symbol α). It can be seen on Fig. 4 that the whole sample can be discriminated into a small number of families in terms of the CEI (horizontal axis). Using the balanced copolymer as a reference, we can now define a second axis to represent the excess (defect) in methylene in the copolymer, $\beta = m - n = (H - N)/2 - N = (H - 3N)/2$. We refer to it as the "Methylene Excess Index" (MEI, symbol β).

A geometrical analogy is based on a 135 degrees rotation of the 2D van Krevelen axes around the $(0.5; 1.5)$ "focal point" characterizing the balanced copolymer $(CH_2)_{m=n}(HCN)_n$. This rotation defines new axes: $(1 - (N/C + H/C)/2; H/C - N/C - 1)$. However, this representation still suffers of the "mass convergence" issue, which is relieved by using absolute compositions instead of carbon-relative ones: $(C - (N + H)/2; H - N - C)$. The CEI appears as the first axis, and to recover the MEI, one transforms the second axis from $H - N - C$ to $H - N - (C - \alpha) = (H - 3N)/2 = \beta$.

All the identified compounds $C_xH_yN_z$ of our samples can therefore be attributed to a (α, β, n) triplet or C_α-$(CH_2)_{n+\beta}(HCN)_n$ composition. Fig. 4 reports the species distribution within CEI/MEI (α/β) classes; each class containing all occurrences of n values. The cumulated peak intensity within each class is calculated and normalized to the total CPI, which defines a 2D probability function $P(\alpha, \beta)$. The marginal distribution $P(\alpha) = \sum_\beta P(\alpha, \beta)$ and $P(\beta) = \sum_\alpha P(\alpha, \beta)$ are unimodal and very smooth (Fig. S-2).

The (α, β, n) representation is elegant and compact for the studied samples. It gives a central role to the balanced copolymer $(CH_2)_{m=n}(HCN)_n$ $(\alpha = \beta = 0)$, which is justified *a posteriori* by the good symmetry of the samples within this representation, and was also foresighted from the Kendrick plots (Fig. 2). An interesting point is also the uniformity of the polymeric pattern across the C_α families, which is further developed below.

[Figure 4 about here.]

3.5 The CH_2/HCN copolymer families

To build a 2D representation of the whole mass spectrum using the CEI (α) and MEI (β) indexes, we designed a composite index $\alpha + 0.5 * \beta/max(|\beta|)$. This enables to present a full attribution of the mass spectrum to α-families in a single view (Fig. 5). These families are the same that were observed in the 3D van Krevelen diagram (Fig. 3), vertically separated by $\Delta\alpha = 1$. Within each family, β is used to discern between individual peaks. A close-up (Fig. 5, inset) illustrates the internal structure of the families, showing the CH_2 and HCN increase directions. One can observe again that the C_α families are compact, with a regular and smooth peak intensity distributions.

This representation provides also a good way to compare samples composition: spectra of both samples have an overlap area between $\alpha = -3$ and $\alpha = 0$, but the spectrum of SA90 presents many peaks at higher values of the α, up to 5-6 instead of 2 for the SA98 one. One can also appreciate on the C_0 family how SA90 is richer in CH_2 than SA98. This point is further discussed below.

[Figure 5 about here.]

The regularity of the peak intensity distribution within families suggests a "statistical copolymer" pattern,[44] with a low correlation between the CH_2 and HCN units numbers (Fig. 5, inset). Analytical artefacts mentioned above (solubility differentiation, proton affinity and transmission function of the instrument) do not enable us to assess quantitatively this point. Still, a semi-quantitative picture can be obtained by plotting the marginal distributions of both units for each family.[44] A probability is assigned to each peak from it's relative intensity within the family, which defines a 2D probability function $P(m, n)$ over the $(CH_2)_m(HCN)_n$ members within each C_α family. The marginal distribution for one unit is obtained by summing over all occurrences of the other unit, $P(m) = \sum_n P(m, n)$ for CH_2 length distribution, and $P(n) = \sum_m P(m, n)$ for HCN length distribution.

Fig. 6 shows the corresponding length distributions. For the families shared between both samples (top and middle row of Fig. 6), the shape of the distributions is remarkably regular, Poisson-like. The HCN distributions for α = -3 and -2 are quite irregular, but the distributions for $\alpha = -1$, 0 and 1 are practically identical between samples and families, and peaked at $n \sim 4$. On the contrary, the CH_2 length distributions shift and widen regularly with the family index, with different positions and width for both samples. The SA90 sample is notably richer in CH_2 than the SA98, but has identical HCN distribution. This supports the idea of a poly-(HCN) base for Titan's tholins, with CH_2 enrichment increasing with the CH_4 concentration in the gaseous mixture.

The bottom row of Fig. 6 displays the $C_{\alpha>1}$ families, observed only in SA90. For HCN, one sees a regular shift of the distribution peak of one unit, with conservation of the width. This shift is consistent with observation, from C_0 upwards, that the first members of the families have their masses increasing with

α (Fig. 5). The CH_2 distributions are in the continuity of their previously described behavior, with an additional deformation of the distributions at the upper end of the m scale, probably due to the solubility limits of tholin and the transmission function of the instrument. The CH_4 concentration in the gaseous mixture seems to be intimately linked to the formation of new, carbon rich, families as observed in SA90.

[Figure 6 about here.]

3.6 Comparison with the alternative C_2H_2/HC_3N copolymer scheme

The C_2H_2/HC_3N pair of units has been invoked in the literature as a possible structural basis for Titan's tholins.[45] This pair can also be used to formally express all $C_xH_yN_z$ species, the only difference with the CH_2/HCN scheme being the CEI value α. The distribution of peak intensity amongst the CEI families, based on this copolymer scheme is compared with CH_2/HCN in Fig. 7. The C_2H_2/HC_3N schemes provides a much wider distribution (more C_α families) which is strongly biased towards the negative α values. Due to the larger number of families, each family contains notably less peak number/intensity. The C_2H_2/HC_3N copolymer scenario is thus much less parsimonious than the CH2/HCN scenario.

Existing copolymer formation schemes implemented in models of Titan atmosphere[46–49] are essentially based on Clarke and Ferris[45] laboratory simulations and reaction scheme involving HC_3N and small hydrocarbons, C_2H_2 and C_2H_6 . The present study suggests additional routes for aerosols formation and nitrogen inclusion, involving HCN and/or CH_3CN.

[Figure 7 about here.]

3.7 Identification of copolymer family roots

The C_α classification provides an homogeneous representation of the different families, but the C_α residue has no chemical meaning in itself. We have thus identified the smallest detected species of each family in the mass spectrum, in order to highlight possible chemical growth trees starting from these roots, as suggested by Fig. 5 (inset) for the C_0 family. These roots are reported in Table 1. For $\alpha < 2$, the first molecule retrieved for each family has a mass lower than 100 u and tentative structures can be reasonably proposed (Table 1). Linear amino nitrile compounds with a great astrobiological interest are identified as roots for $-4 \leq \alpha \leq -1$, whereas a very probable cyclization towards aromatic azoles is observed for $0 \leq \alpha < 2$. Validation of these tentative root structures will require further MS/MS analyzes. Nevertheless, they conform with the C/N alternation pattern observed by MS/MS analysis of a set of tholin components.[15] For $\alpha \geq 2$, there are too many possible isomers to risk an identification.

The first neutrals that have been identified are not always ideal roots, in the sense that they do not enable to generate the whole family by addition of copolymer units. For instance, in the C_0 family, the second smallest member $C_4H_5N_3$ is obtained by a theoretical +N-H transition, resulting from a +HCN -CH_2 substitution. For a simple growth process, there should be a parent molecule at a lower mass ($C_3H_4N_2$), which is not visible in our spectrum (low mass molecules might have been pumped during the production and the thorough pumping that precedes the chamber opening). A unit loss mechanism is thus probably active in this heterogeneous chemistry.

[Table 1 about here.]

4 Conclusion

This work was focused on identifying and visualizing the polymeric structure of Titan's tholins produced in the PAMPRE setup, with statistical-graphical representations developed for hydrocarbons in petroleomics and originally adapted here for nitrogen-rich organic compounds.

In addition to the classical hydrocarbon CH_2 monomer, we identified the HCN monomer with an original Kendrick Mass Defect analysis based on a poly-$(CH_2)_m(HCN)_n$ hypothesis for Titan's tholins. We have also shown that a 3D $\{H/C; N/C; m/z\}$ van Krevelen diagram is an interesting tool to reveal families of copolymers differing by their root (or terminator).

New diagrams were developed with geometrical transformations on the classical H/C and N/C van Krevelen axes. A first axis, the "Carbon Excess Index" $\alpha = C - (H + N)/2$, ranks the species detected in the mass spectra according to their residue in comparison with a pure poly-$(CH_2)_m(HCN)_n$ copolymer. A second axis, the "Methylene Excess Index" $\beta = (H - 3N)/2$, characterizes the excess of methylene in comparison with the corresponding balanced copolymer ($m = n$). This scheme enables us to assign all compounds to a small number of families with very regular/statistical polymeric distributions. The roots of some of these families (*i.e.* the smallest members detected in the mass spectra) are identified as linear and cyclic amino-nitriles. Precise identification of the roots of the remaining families, involving too many possible isomers, requires further structural investigations by tandem MS/MS studies.

This work confirms the importance of the nitrile chemistry for Titan's aerosol production, in agreement with the large nitrogen incorporation observed by the ACP instrument of the Huygens probe.[4] More specifically, our analyzes are compatible with a poly-(HCN) basis of tholin, with a CH_2 enrichment directly depending on the concentration of CH_4 in the gaseous mixture. New routes are proposed for Titan's aerosols formation, compatible with the present knowledge on Titan's atmospheric reactivity. They involve HCN and CH_2 as patterns, besides the C_2H_2/HC_3N pair presently favored in most Titan's atmospheric chemistry models.

However the polymeric compositions of tholin evidenced by high resolution mass spectrometry, if it allows detailed and extremely sensitive comparative studies, do not enable to understand details of the chemical mechanisms involved in their formation. Further studies of these mechanisms are necessary to characterize the roots formation processes, the polymeric growth and the possible interactions and pathways between the families highlighted in this work.

Acknowledgments

O. Laprévote is warmly acknowledged for enabling the mass-spectrometry measurements at the Institut de Chimie des Substances Naturelles. We thank the PRES UniverSud-Paris (Contract 2008-53), the CNRS/OPV Interdisciplinary Program, the Programme National de Planétologie, and French National Agency for Research (Contracts ANR-07-BLAN-0123 and JC09_464733) for their financial support.

Supporting Information Available

Two figures (Fig. S-1 and Fig S-2) are provided separately. This material is available free of charge via the Internet at http://pubs.acs.org.

References

[1] Vuitton, V.; Yelle, R. V.; Anicich, V. G. *The Astrophysical Journal* **2006,** *647,* L175–L178.

[2] Fox, J.; Galand, M.; Johnson, R. *Space Science Reviews* **2008,** *139,* 3–62.

[3] Waite, J.; Young, D.; Cravens, T.; Coates, A.; Crary, F.; Magee, B.; Westlake, J. *Science* **2007,** *316,* 870–875.

[4] Israel, G. *et al. Nature* **2005,** *438,* 796–799.

[5] Sagan, C.; Khare, B. *Nature* **1979,** *277,* 102–107.

[6] Coll, P.; Coscia, D.; Gazeau, M.-C.; Ramirez, S.; Cernogora, G.; Israel, G.; Raulin, F. *Planetary and Space Science* **1999,** *47,* 1331–1340.

[7] Szopa, C.; Cernogora, G.; Boufendi, L.; Correia, J.; Coll, P. *Planetary and Space Science* **2006,** *54,* 394–404.

[8] Imanaka, H.; Smith, M. *Geophysical Research Letters* **2007,** *34,* L02204.

[9] Tran, B.; Force, M.; Briggs, R.; Ferris, J.; Persans, P.; Chera, J. *Icarus* **2008,** *193,* 224–232.

[10] Sarker, N.; Somogyi, A.; Lunine, J.; Smith, M. *Astrobiology* **2003,** *3,* 719–726.

[11] Somogyi, A.; Oh, C.-H.; Smith, M.; Lunine, J. *J. Am. Soc. Mass Spectrom.* **2005,** *16,* 850–859.

[12] McKay, C. *Planetary and Space Science* **1996,** *44,* 741–747.

[13] Imanaka, H.; Khare, B.; Elsila, J.; Bakes, E.; McKay, C.; Cruikshank, D.; Sugita, S.; Matsui, T.; Zare, R. *Icarus* **2004,** *168,* 344–366.

[14] Kendrick, E. *Analytical Chemistry* **1963,** *35,* 2146–2154.

[15] Carrasco, N.; Schmitz-Afonso, I.; Bonnet, J.-Y.; Quirico, E.; Thissen, R.; Dutuit, O.; Bagag, A.; Laprévote, O.; Buch, A.; Giulani, A.; Adandé, G.; Ouni, F.; Hadamcik, E.; Szopa, C.; Cernogora, G. *Journal of Physical Chemistry A* **2009,** *113,* 11195-11203.

[16] Krevelen, W. V. *Fuel* **1950,** *29,* 269–284.

[17] Hughey, C.; Hendrickson, C.; Rodgers, R.; Marshall, A.; Qian, K. *Analytical Chemistry* **2001,** *73,* 4676–4681.

[18] Marshall, A.; Rodgers, R. *Accounts of Chemical Research* **2004,** *37,* 53–59.

[19] Wu, Z.; Rodgers, R.; Marshall, A. *Analytical Chemistry* **2004,** *76,* 2511–2516.

[20] Meija, J. *Analytical and Bioanalytical Chemistry* **2006,** *385,* 486–499.

[21] Reemtsma, T. *Journal of Chromatography A* **2009,** *1216,* 3687–3701.

[22] Imanaka, H.; Smith, M. *American Astronomical Society, DPS meeting #40, #20.02; Bulletin of the American Astronomical Society* **2008,** *40,* 421.

[23] Hudgins, D.; Bauschlicher, Jr., C.; Allamandola, L. *The Astrophysical Journal* **2005,** *632,* 316–332.

[24] Majumdar, A.; Schï¿œfer, J.; Mishra, P.; Ghose, D.; Meichsner, J.; Hippler, R. *Surface & Coatings Technology* **2007,** *201,* 6437–6444.

[25] Majumdar, A.; Scholz, G.; Hippler, R. *Surface & Coatings Technology* **2009,** *203,* 2013–2016.

[26] Yamashita, M.; Fenn, J. *The Journal of Physical Chemistry* **1984,** *88,* 4451–4459.

[27] Makarov, A. *Analytical Chemistry* **2000,** *72,* 1156–1162.

[28] Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Analytical Chemistry* **2006,** *78,* 2113–2120.

[29] Morosoff, N.; Crist, B.; Bumgarner, M.; Hsu, T.; Yasuda, H. *Journal of Macromoleculecular Science* **1976,** *A10(3),* 451–471.

[30] Gegenbach, T.; Chatelier, R.; Griesser, H. *Surface and Interface Analysis* **1996,** *24,* 271–281.

[31] Jiang, H.; Grant, J.; Enlow, J.; Su, W.; Bunning, T. *Journal of Materials Chemistry* **2009,** *19,* 2234–2239.

[32] Rodgers, R.; Schaub, T.; Marshall, A. *Analytical Chemistry* **2005,** *77,* 20–27.

[33] Rodgers, R.; Marshall, A. Asphaltenes, Heavy Oils, and Petroleomics. In ; Mullins, O.; Sheu, E.; Hammami, A.; Marshall, A., Eds.; Springer New York: 2007; Chapter Petroleomics: Advanced Characterization of Petroleum-Derived Materials by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS), pages 63–93.

[34] Marshall, A.; Rodgers, R. *Proceedings of the National Academy of Sciences* **2008,** *105,* 18090.

[35] Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. *Analytical Chemistry* **2008,** *80,* 8908–8919.

[36] Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. *Geochimica et Cosmochimica Acta* **2006,** *70,* 2990–3010.

[37] Kramer, R.; Kujawinski, E.; Hatcher, P. *Environmental Science & Technology* **2004,** *38,* 3387–3395.

[38] Tremblay, L.; Dittmar, T.; Marshall, A.; Cooper, W.; Cooper, W. *Marine Chemistry* **2007,** *105,* 15–29.

[39] Hughey, C.; Hendrickson, C.; Rodgers, R.; Marshall, A. *Energy & Fuels* **2001,** *15,* 1186–1193.

[40] Liebman, S.; Pesce-Rodriguez, R.; Matthews, C. *Advances in Space Research* **1995,** *15,* 71–80.

[41] Quirico, E.; Montagnac, G.; Lees, V.; McMillan, P.; Szopa, C.; Cernogora, G.; Rouzaud, J.-N.; Simon, P.; Bernard, J.-M.; Coll, P.; Fray, N.; Minard, R.; Raulin, F.; Reynard, B.; Schmitt, B. *Icarus* **2008,** *198,* 218–231.

[42] Pearson, K. *Philosophical Magazine* **1901,** *2,* 559–572.

[43] Wold, S.; Esbensen, K.; Geladi, P. *Chemometrics and Intelligent Laboratory Systems* **1987,** *2,* 37–52.

[44] Wilczek-Vera, G.; Danis, P.; Eisenberg, A. *Macromolecules* **1996,** *29,* 4036–4044.

[45] Clarke, D.; Ferris, J. *Icarus* **1997,** *127,* 158–172.

[46] Lebonnois, S.; Bakes, E.; McKay, C. *Icarus* **2002,** *159,* 505–517.

[47] Wilson, E.; Atreya, S. *Planetary and Space Science* **2003,** *51,* 1017–1033.

[48] Lebonnois, S. *Planetary and Space Science* **2005,** *53,* 486–497.

[49] Lavvas, P.; Coustenis, A.; Vardavas, I. *Planetary and Space Science* **2008,** *56,* 67–99.

[50] Khare, B.; Sagan, C.; Thompson, W.; Arakawa, E.; Suits, F.; Callcott, T.; Williams, M.; Shrader, S.; Ogino, H.; Willingham, T.; Nagy, B. *Advances in Space Research* **1984,** *4,* 59–68.

List of Tables

Family	First neutral	Mass (u)	Tentative Structure	Name
C_{-4}	$C_2H_7N_5$	101.0701	H2N-C(=NH)-NH-C(=NH)-NH2	biguanide
C_{-3}	CH_5N_3	59.0483	(NH2)2C=NH	guanidin
C_{-2}	$C_2H_6N_2$	58.0825	CH3-C(=NH)NH2	acetamidine
C_{-1}	$C_2H_4N_2$	56.0666	NC-CH2-NH2	aminoacetonitrile
C_0	$C_4H_6N_2$	82.1039		methylimidazole
C_1	$C_5H_6N_2$	94.1146		vinylimidazole
				methylpyrazine, *cf.* Khare *et al.*[50]
C_2	$C_7H_7N_3$	133.0640	too many isomers...	
C_3	$C_{10}H_{11}N_3$	173.0953		
C_4	$C_{13}H_{14}N_4$	226.1218		
C_5	$C_{17}H_{20}N_4$	280.1688		

Table 1: Putative chemical roots of the main formal C_α-$(CH_2)_m(HCN)_n$ families.

List of Figures

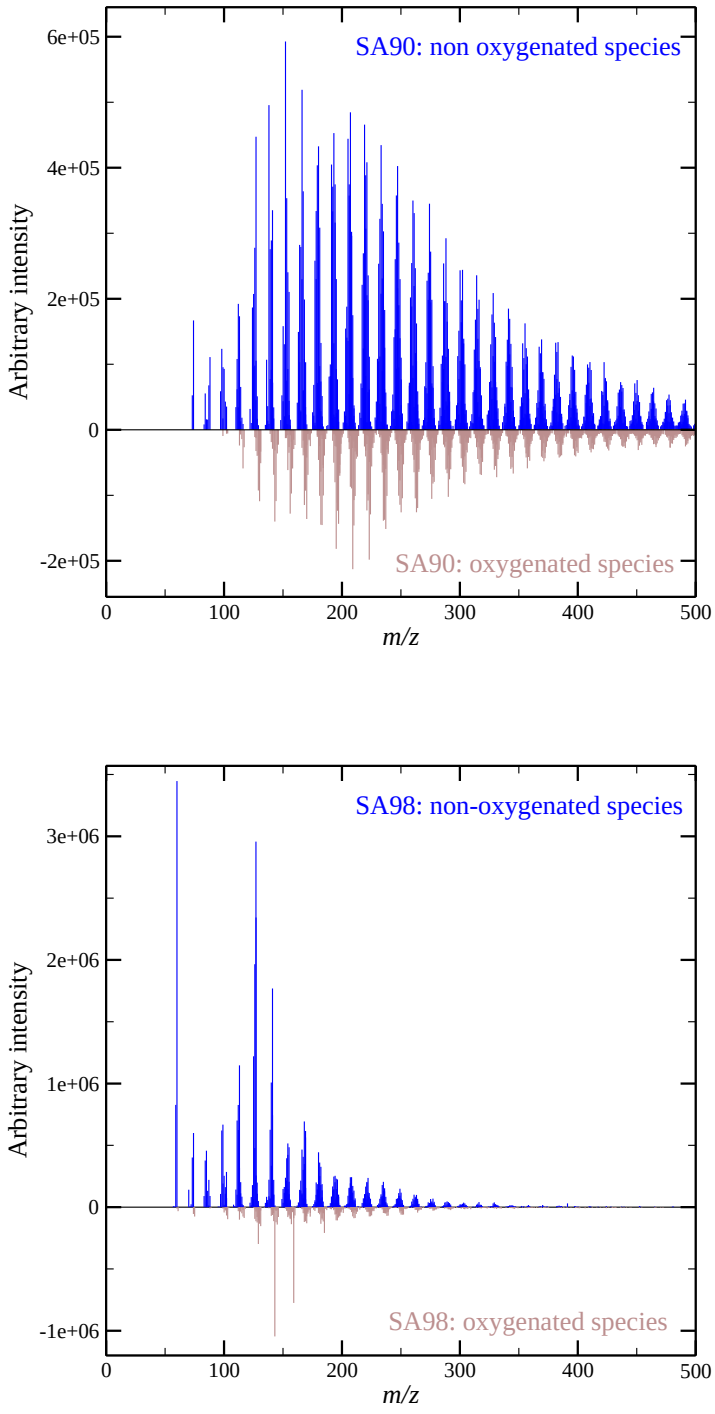

Figure 1: ESI/Orbitrap mass spectra for SA90 (top) and SA98 (bottom) samples. Oxygenated/^{13}C species are shown with negative intensities.

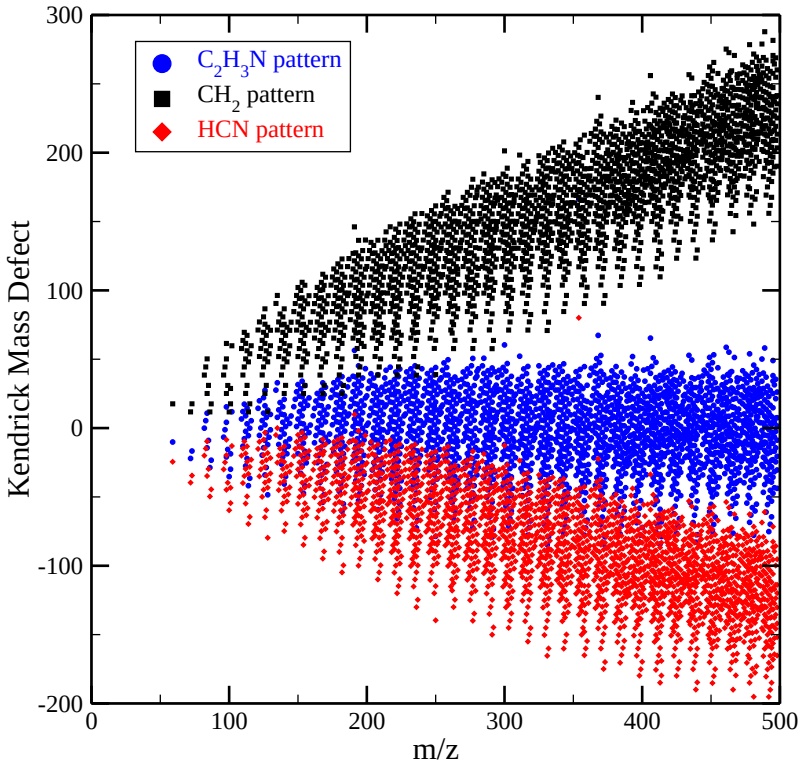

Figure 2: Kendrick plots for the SA90 sample using different patterns: CH_2, HCN and C_2H_3N.

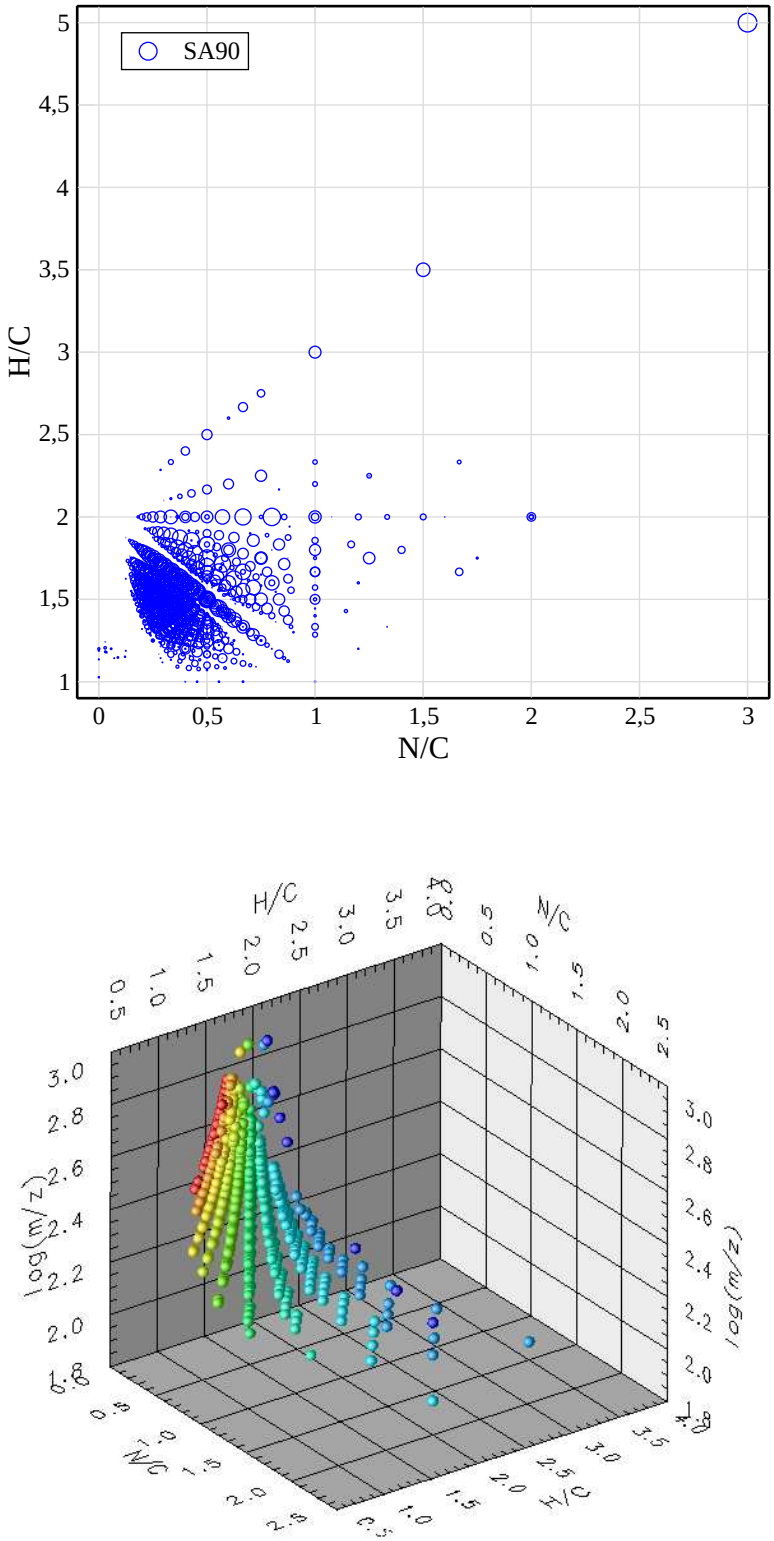

Figure 3: Van Krevelen diagrams of the SA90 sample: (top) 2D (N/C; H/C), symbol size is proportional to $\log(I/I_{thresh})$; (bottom) close-up on the central part of the 3D plot (N/C; H/C; $\log(m/z)$), families identified in Fig. 5 have been colored *a posteriori* for better legibility.

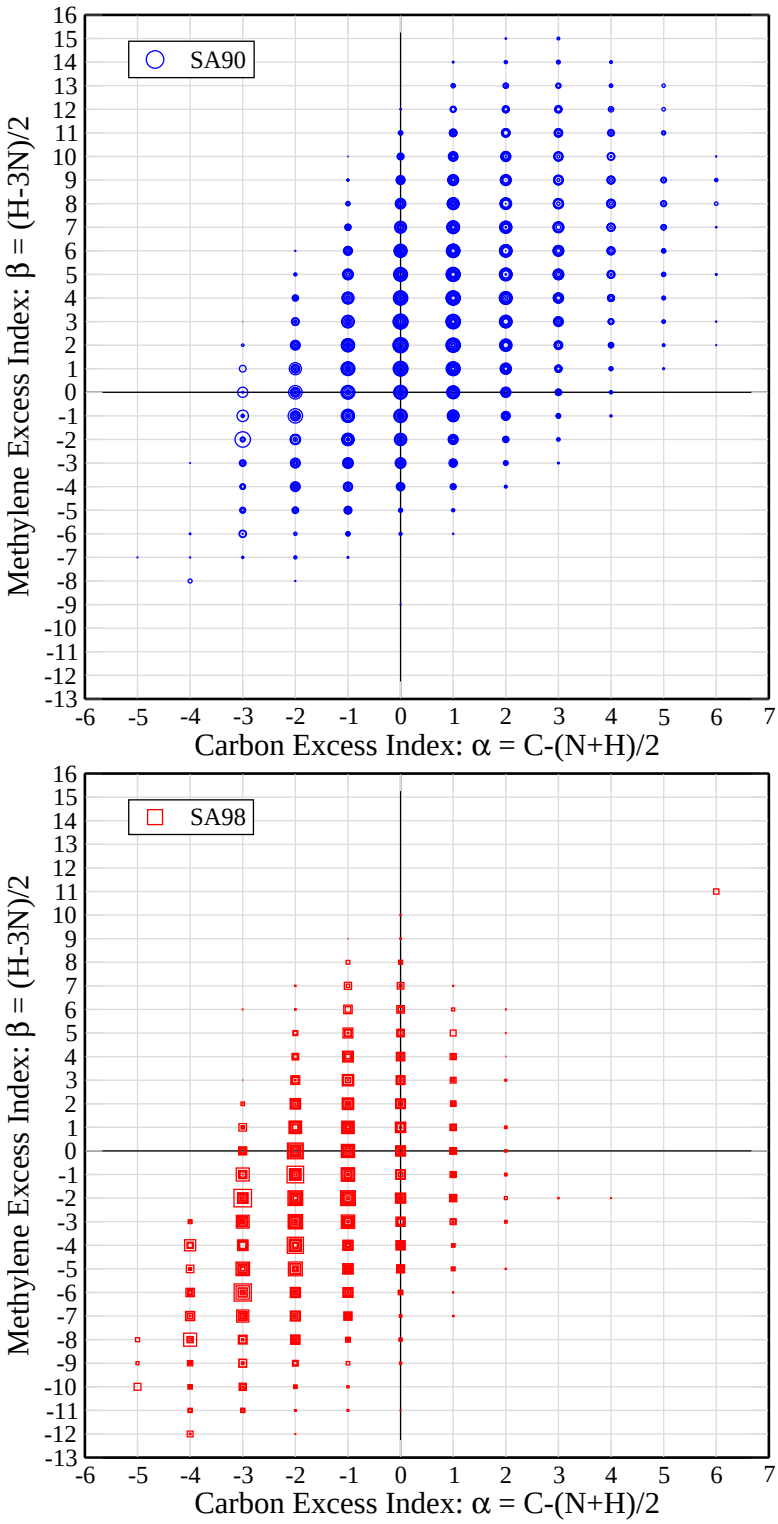

Figure 4: CEI/MEI representation of the SA90 (top) and SA98 (bottom) samples. Symbol size is proportional to $\log(I/I_{thresh})$.

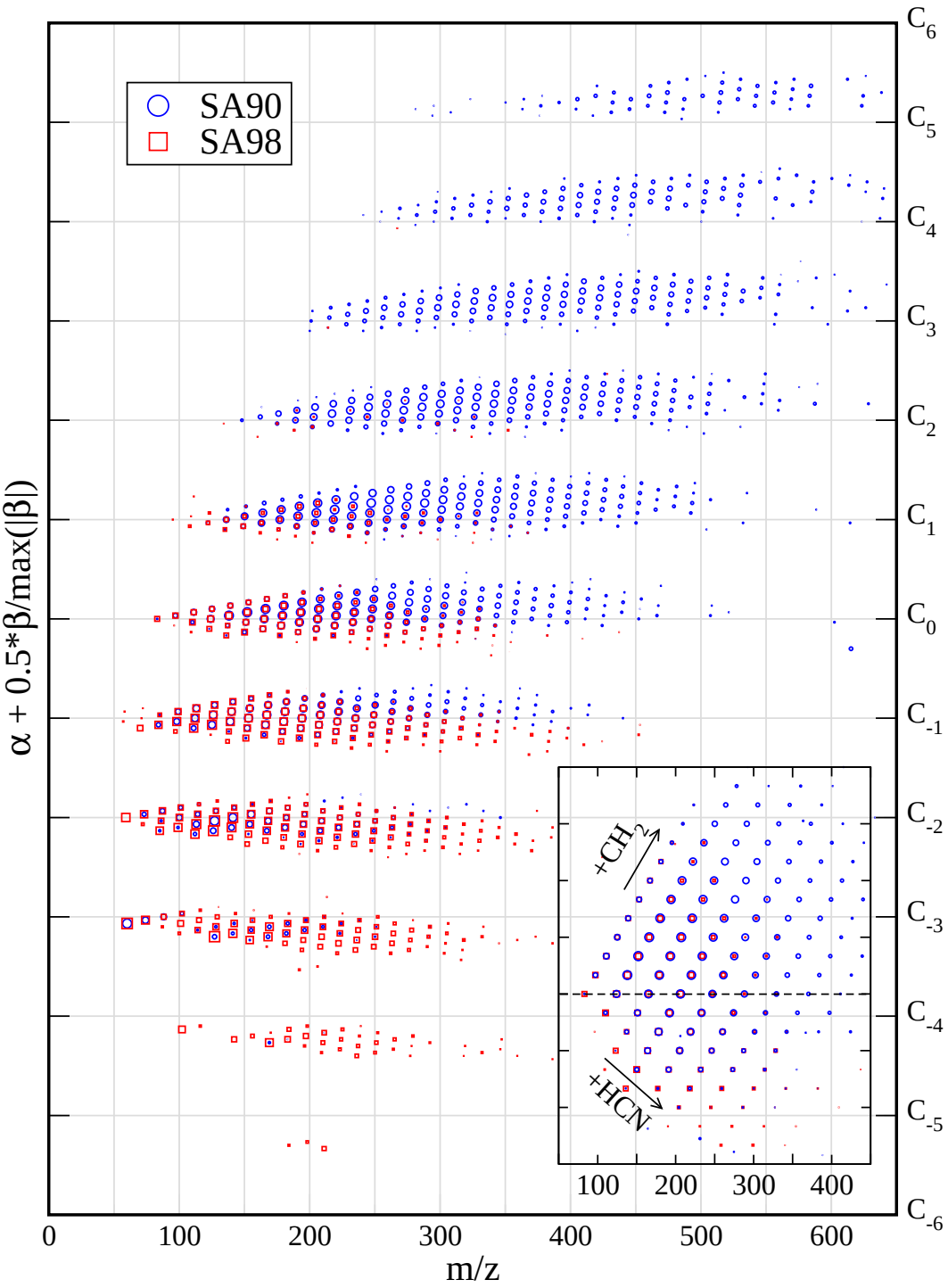

Figure 5: Composite (m/z; CEI/MEI) representation of the SA90 (blue circles) and SA98 (red squares) samples, showing the attribution of all peaks within the copolymeric scheme. Symbol size is proportional to $\log(I/I_{thresh})$. Inset: close-up of the C_0 family, showing the directions of increase of CH_2 and HCN units within the family.

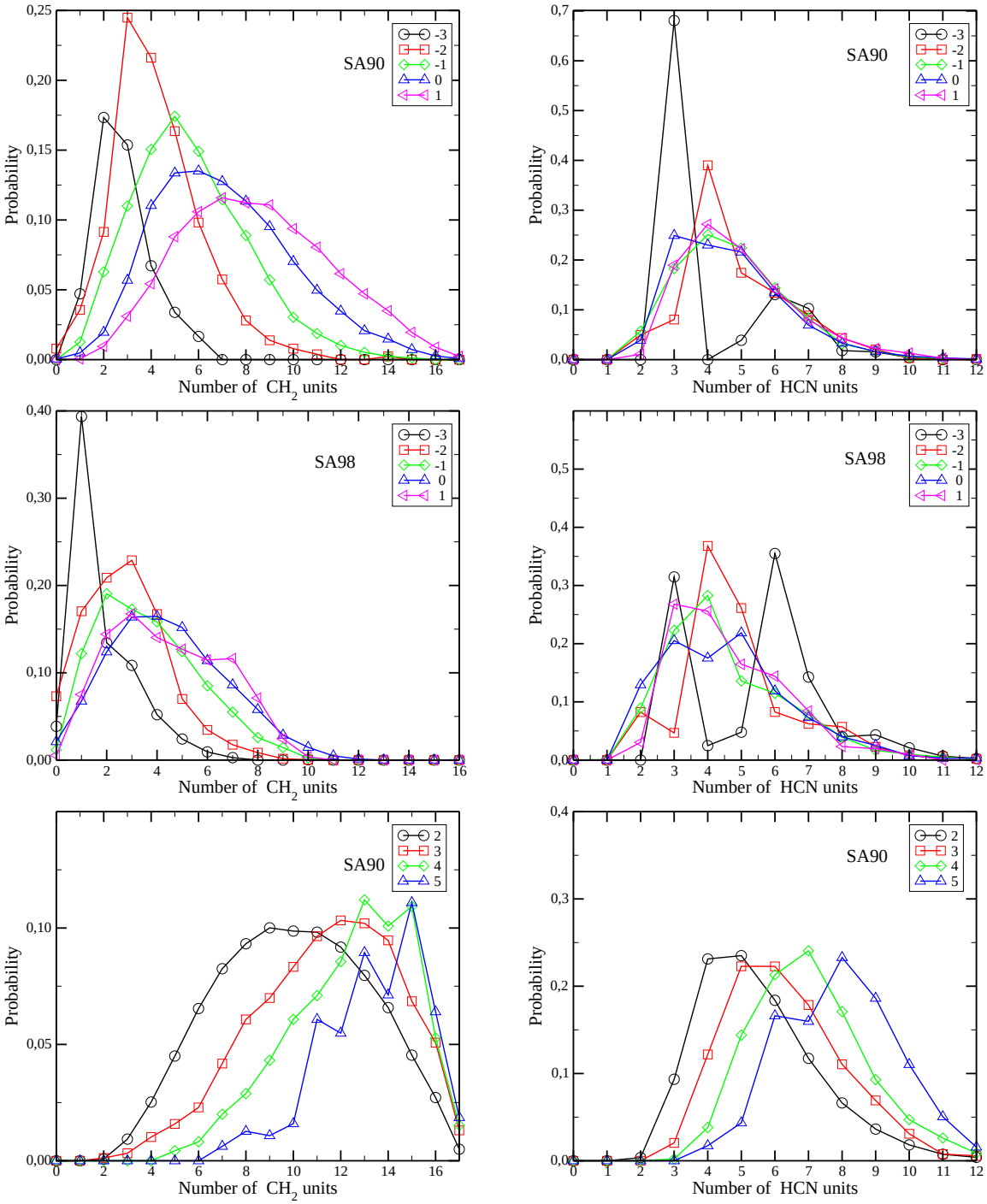

Figure 6: Experimental marginal probabilities of "length distributions" for the CH_2 and HCN units within the shared (from $\alpha = -3$ to 1) C_α families for SA90 (top row) and SA98 (middle row) samples. For clarity, the bottom row shows the $C_{\alpha>1}$ (from $\alpha = 2$ to 5) families, present only in SA90.

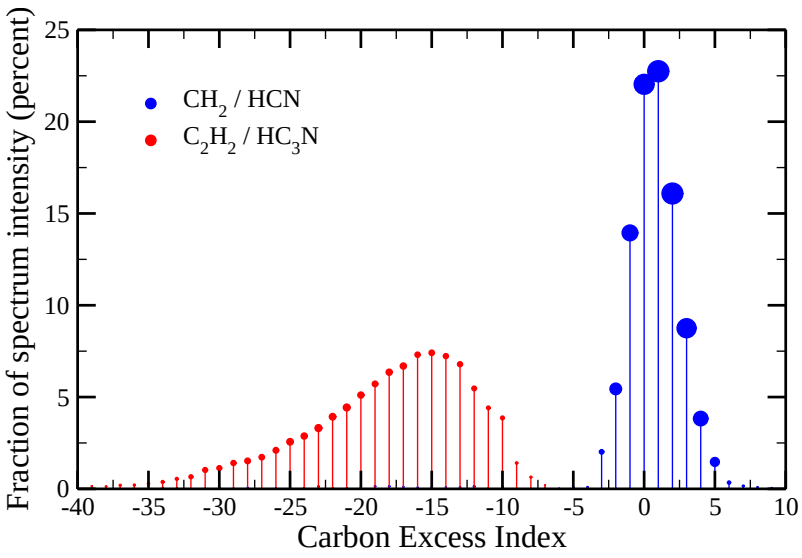

Figure 7: Distribution of peak intensity of SA90 mass spectrum within the C_α families, based on two hypothetical copolymer patterns: (blue) CH_2/HCN and (red) C_2H_2/HC_3N. The symbol size is proportional to the percentage of peak numbers represented by each family.