Mass spectrometry of humic substances of different origin including those from Antarctica
A comparative study

E.M. Peña-Méndez [a],[*],[1], D. Gajdošová [a], K. Novotná [a], P. Prošek [b], J. Havel [a]

[a] *Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic*
[b] *Department of Geography, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic*

Abstract

Mass spectra of humic acids (HA) from different sampling sites (Antarctica, Brazil, Czech Republic, Mexico and USA) and origin (plant, soil, peat, and coal derived) were obtained by laser desorption/ionization time of flight mass spectrometry (LDI-TOF MS). Optimisation of the experimental conditions are given as the optimal value of the laser energy at ∼20–30% higher than the threshold. Under these conditions, reproducible mass spectra of HA samples were obtained. In the mass spectra the majority of the peaks are observed in the m/z region 100–1000 Da. Mass spectra fingerprints of HA were analyzed and, in spite of the differences in their origin, a number of common features and profiles (patterns of peaks) were observed in most of the samples. Very similar structural groups (patterns) of the peaks are present in the m/z range 717–918 Da for HA samples of quite different origins, countries or continents.

The tandem LDI-TOF MS and multivariate statistical tools allowed us to extract and elucidate underlying information contained in the mass spectra of the HA samples under study. Applying principal components and cluster analysis, it was, e.g. demonstrated that most of the Antarctica HA samples show distinguishable differences when compared with humic acids from other continents and of different origin.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Mass spectrometry; Laser desorption ionization; Humic acids; Principal component analysis; Cluster analysis

1. Introduction

From the point of view of the global environment, the soil organic matter (SOM) represents an important reservoir of carbon, playing an important role in the carbon cycle. The nitrogen cycle is also indirectly affected by the humic matter through the biological properties of soil [1]. Humic substances (HS) are among the most important constituents of SOM and their importance, e.g. in agriculture is well known. The chemical and physical properties of soil are related to SOM yielding HS, being of great importance not only in

soil, but also in the plant growth (nutrition and physiology). There are many applications of HS in industry for various purposes, however most of this information is patented or considered as "industrial secrets". The idea of using humic matter for medicinal and pharmaceutical purposes comes from ancient time, i.e. the therapeutic properties of peat have been known for a long time in Europe [1]. Other important properties of humic acids (HA), such as antibacterial, antitoxic, immunomodulatory, anti-inflamatory, anti-arthritic, and anti-ulcerogenic have attracted the attention of medical and pharmaceutical communities. Nowadays, pharmaceutical companies devote much effort to investigating the medicinal drugs that can be obtained from humus matter [2]. HS are also considered to be quite important for solving environmental problems. HA are considered to be the most widely-spread natural complexing ligands in nature. Due to this property, HS play an important role in removing toxic metals, anthro-

---

* Corresponding author. Tel.: +34 922 31 80 47; fax: +34 922 31 80 03.
  E-mail address: empena@ull.es (E.M. Peña-Méndez).
[1] On leave from: Department of Analytical Chemistry, Nutrition and Food Science, University of La Laguna, Campus de Anchieta, 38071 La Laguna, Tenerife, Spain.

pogenic organic chemicals and other pollutants from water and soil [3]. The important and critical issues of humic matter, together with the latest development in the analytical instrumentation, explain the renaissance of HS research in recent years.

HS are widely spread in nature. They are present in soil and waters. It is estimated that they represent about 40–90% of the dissolved organic matter in soil and water environments. HS are formed by the association of substances of animal, vegetable and microbiological origins and, to a large extent also from lignin [4]. The major part of the research studying HS assumes plant residues as the main source of humus. HS have also been found in the Antarctic continent, where the humification process under the Antarctic conditions is very specific and different from other continents. Examples of the latest results concerning the presence of HA in the Antarctic environment are given elsewhere [6–9].

Part of the HS are the humic acids, which are supposed to be high molecular weight supramolecules, including polymeric aromatic structures with heterogeneous monomers and heterogeneous kinds of linkages formed by random events. HA present an enigma in that their structure still remains unknown. That is in spite of the fact that a considerable amount of information concerning their behaviour, chemical composition and properties has been gathered over the past years [5]. Lignin has also been included in these studies. Considerable effort has been made in research into the structural characterisation of HS. The presence of some compounds, such as sugars, polyphenols, aminoacids, among others has been confirmed in the HA structure and/or as HA constituents [1]. Practically all available analytical techniques and methods have been applied in the attempts to explain or, at least, to elucidate the composition and structure of these highly complex natural products. However, in spite of great efforts by different research groups, their complete structure still remains unknown.

Determination of the HA structure represents a quite complex "problem". Recently, a number of various soft ionization mass spectrometric techniques, like the electrospray ionization (ESI) [10,11] or the matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) and lately laser desorption ionization time of flight mass spectrometry (LDI-TOF MS) are being applied more and more to the characterisation of HS [4,8,12–17]. Application of MALDI-TOF MS or LDI-TOF MS to characterise HA gives complex mass spectra [8,9,12]. In the case of LDI-TOF MS, there is a danger of fragmentation which needs to be avoided, while MALDI-TOF MS also has limited use, not only due to fragmentation of HA, but also because of the formation of aggregates or clusters. Formation of clusters in MALDI has been reported by Remmler et al. [18], Haberhauer et al. [19] and Pacheco and Havel [20]. Therefore, for both techniques, careful optimisation of the parameters affecting measurement is necessary.

Mass spectra not only contain many structurally significant peaks, but also a number of peaks related to chem-

ical noise (i.e. contaminants), which represent a kind of rather complex HA fingerprint. This structural information can be used for the characterisation of HA, but the information from the spectra is not straightforward. Chemometrics methods have proved to be useful for the extraction and managing of such underlying and relevant information contained in the spectroscopic data [21–23] and, in the last decade, the use of multivariate techniques has been extended to the field of HS studies. Several papers describe the application of principal component analysis (PCA) and/or factor analysis (FA) to reveal underlying differences in HA, using the solid-state cross-polarisation/magic-angle spinning (CPMAS) ^{13}C NMR spectroscopy [24], pyrolysis-field ionization mass spectrometry (Py-FIMS) [24], gas chromatography–mass spectrometry (GC–MS) [25], thermogravimetry/fourier transform infrared spectroscopy (TG-FTIR) [26], capillary electrophoresis (CE) [27], vapour pressure osmometry [28–30], etc. However, the use of chemometrics for the investigation and characterisation of HA mass spectra obtained by MALDI-TOF MS and LDI-TOF MS has not been applied, as far as we know.

The aim of this work is to study the possibilities and limitations of LDI-TOF MS to characterise HA from various countries or continents (Antarctica, Brazil, Czech Republic, Mexico and USA), including samples from the International Humic Substances Society (IHSS) and of different origin through their mass spectra fingerprints. We have used LDI for simplicity, because ESI leads mostly to multiple charged ions while LDI gives single charged ions, and also because well reproducible results are obtained [4,9,12,15,17,31,32]. Also, the possibility of multivariate statistical methods to extract information contained in the HA mass spectra will be studied.

2. Experimental data and methods

2.1. Chemicals and reagents

All reagents were of analytical grade purity. Sodium hydroxide was from Merck (Darmstadt, Germany); lignin alkali was purchased from Aldrich (Milwaukee, USA). De-ionized water used to prepare all solutions was double distilled from the quartz apparatus by Heraeus Quartzschmelze (Hanau, Germany).

2.2. Humic acids

Chemapex HA standard was a gift from Chemapex s.r.o. (Chomutov, Czech Republic). Soil HA called Cananeia and Pindorama were from Brazil (Sao Paulo, Brazil) [33], Peat HA denominated "Thurfa" was from Brazil (Sao Paulo, Brazil) [34] and Leonardite was from Mexico (Grupo Bioquímico Mexicano, Mexico) [35]. The plant HA from stem, leaf and root of water hyacinth were isolated by Ghabbour and Davis [36]. HA from the Antarctic soil samples
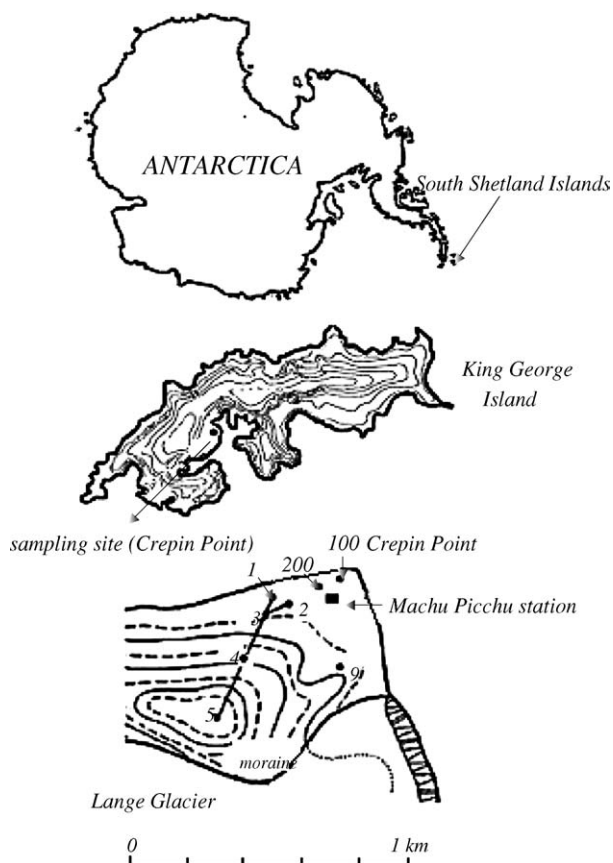
Fig. 1. Geographical location of sampling site in relation to King George Island and to the Antarctica. Codes for sampling sites: (1) Ant-1; (2) Ant-2; (3) Ant-3; (4) Ant-4; (5) Ant-5; (9) Ant-9; (100) Ant-100; (200) Ant-200; (moraine) Ant-m.

were extracted by Gajdošová et al. [8] from the Antarctic soil in Crepin Point (King George Island, South Shetland, Antarctica). The procedure is described in detail elsewhere [8]. Fig. 1 shows the sampling location in Antarctica.

Commercial HA: Fluka HA was from Fluka (Buchs, Switzerland) and Aldrich HA was purchased from Sigma–Aldrich (Steinheim, Germany). The standard HA from the International Humic Substances Society (IHSS) were Soil (1S102), Peat (1R103H), and Leonardite (1S104H) HA standards.

A review of the elemental composition and analysis of HA studied in this work is shown in Table 1.

2.2.1. Granularity and chemical constitution of the Antarctic soil related to the HA sampling site

Soil composition might also play its role in the specifics of the Antarctic humification process. Therefore, the soils related to various sampling sites of the Antarctica HA were also characterised in detail in order to follow any eventual differences due to the soil composition [37]. Grain size analysis was also made by sieving dry sediments. The Antarctic soil samples contain mainly sandy gravel—gravely sand character.

The chemical constitution of the clastic material was made by the primacy of $Na_2O > K_2O$ and $Na_2O + K_2O < CaO$. The studied clastic material is similar to the volcanic rock of the andesite and basalt type.

Analysis of the volcanic material of soil samples also allows us to classify the source rocks. Position of the analysed soil samples in the diagram TAS corresponds to basaltic andesite (soil samples Ant-1, Ant-2, Ant-3 and Ant-9) and tholeiitic basalts (sample Ant-4, Ant-5).

2.2.2. Procedure of humic acid dissolution

Stock solutions of HA were prepared by weighing a given amount of the humic acid sample, dissolving it in a small volume of 0.1 M NaOH and diluting to a concentration of 0.036 M NaOH using double-distilled water. All HA working solutions (1 mg ml^{-1}) were analysed a minimum of 10 days after preparation because it was found that, after this time, HA yield better and reproducible mass spectra patterns [38]. Most probably, this is due to a slow process of disaggregation of the humic acids supramolecular structures after they are dissolved, as has already been observed before [9,13,38].

2.3. Apparatus

The total content of carbon, nitrogen and hydrogen was determined on the Perkin-Elmer 2400 CHNS/O analyzer and the ash content was obtained by burning a certain amount of HA in an oven at 900 °C for 6 h.

All measurements of the mass spectra were carried out on the AXIMA CFR (Kratos Analytical, Manchester, UK) mass spectrometer, equipped with a nitrogen laser with a wavelength of 337 nm operated in the linear and positive ion mode. The energy of the laser was changed in the range 0–80 relative units. External calibration was used. Proof that we did precise calibration is supported by the fact that some of the *m/z* values of the peaks that we are reporting here were also observed in other papers and the values are confirmed by a different laboratory [17], where it is mentioned that their results were in good agreement with our previous results.

The volume of 1 μl of sample solution (1 mg ml^{-1} HA) was pipetted onto a sample plate and dried at room temperature before the analysis. Mass spectra were measured in various modes depending on the analysed sample (linear positive, linear negative or reflectron negative modes), but generally, the spectra obtained with the most signals were those done in the linear positive mode. The protocol of the International Humic Substances Society (IHSS) extraction procedure was used in this work. Because of the high sodium content in HA solutions, cationization is observed during mass spectrometry analysis and most of the ions are adducts with Na$^+$ [14]. Thus, under such conditions, the ionization leads to cationization and positive spectra are more intensive. Spectra observed in negative mode do not contain sodium adducts and therefore, under our conditions, were less intensive and so they were not used. Higher intensity of spectra was also observed by Mugo and Bottaro [17]. The fact that for LDI, predominantly

Table 1
Review of elemental analysis and ash content concerning HA studied

Humic acid	Abbreviation	Type	C (%)	H (%)	N (%)	O (%)	C/N	H/C	Ash (%)	Reference
Steam (water hyacinth)	ps	Plant	48.15	2.25	5.12	–	10.97	0.56	<1.00	This work
Leaf (water hyacinth)	pl	Plant	49.60	3.20	9.46	–	6.13	0.77	<1.00	This work
Root (water hyacinth)	pr	Plant	45.31	2.69	5.13	–	10.43	0.70	<1.00	This work
Soil IHSS standard	sS	Soil	58.13	3.68	4.14	34.08	16.37	0.76	0.88	www.ihss.gatech.edu
Pindorama	sP	Soil	51.26	4.82	4.95	38.97	12.08	1.13	8.76	Sao Paulo University, Brazil
Cananeia	sC	Soil	53.75	3.58	2.54	40.13	24.68	0.80	2.26	Sao Paulo University, Brazil
Garden soil	sCZ	Soil	–	–	–	–	–	–	60.00	This work
Leonardite IHSS standard	cdLst	Coal	63.81	3.70	1.23	31.27	60.50	0.69	2.58	www.ihss.gatech.edu
Leonardite HA (Mexico)	cdL	Coal	–	–	–	–	–	–	–	[32]
Aldrich	CdA	Coal	39.03	4.43	0.68	–	11.60	1.45	–	Sigma–Aldrich
Fluka	cdF	Coal	49.01	4.37	0.71	–	80.50	1.06	–	Sigma–Aldrich
Chemapex HA standard	cdC	Coal	50.25	4.90	1.23		47.60	1.16	7.30	Chemapex
Peat IHSS standard	P	Peat	56.37	3.82	3.69	37.34	24.40	0.81	1.12	www.ihss.gatech.edu
Peat (Thurfa)	Ptf	Peat	53.26	3.30	1.98	41.46	–	0.74	6.00	Sao Paulo University, Brazil
Lignin (Norway spruce)	L		47.57	4.97				1.25		Sigma–Aldrich
Antarctica sampling site										
1	Ant-1	Soil	43.99	3.76	5.29		9.70	1.02	–	This work
2	Ant-2	Soil	43.51	2.44	6.54	–	7.76	0.67	–	This work
3	Ant-3	Soil	33.87	2.18	5.55	–	7.12	0.77	68.86	This work
4	Ant-4	Soil	37.48	1.04	5.64	–	7.75	0.33	–	This work
5	Ant-5	Soil	–	–	–	–	–	–	–	This work
9	Ant-9	Soil	–	–	–	–	–	–	–	This work
Ant 100 m	Ant-100	Soil	17.00	2.60	2.00	–	9.82	1.81	76.30	[8]
Ant 200 m	Ant-200	Soil	–	–	–	–	–	–	–	This work
Moraine	Ant-m	Soil	4.60	2.40	0.60	–	8.33	6.37	47.00	[8]

(–) Non available.

positive ions are formed has previously been described by [9,12–14]. Positive spectra were also observed and analysed by Kujawinski et al. [14,39].

2.4. Multivariate data analysis

The data analysis was performed on a Pentium-based IBM compatible personal computer. STATISTICA V.6 (STAT soft Inc., USA) and MATLAB® were used. The data set for all humic acid samples is an $m \times p$ matrix of m samples where each row corresponds to the relative intensities obtained for certain selected m/z values and the intensity values are the features. The total number of rows corresponds to the total number of samples. The intensity of the column (at selected m/z values) was accepted for PCA analysis only if the intensity of the peak was observed at higher than noise for at least three or more samples. Otherwise such a column was rejected.

PCA summarises all the variations in matrix M into a few new variables called principal components. The method is based on the decomposition of the data set matrix, M. The decomposition of the matrix is:

$$M = SL + E$$

where S are the scores and represent trends down the columns of M, and L are the loadings and represent trends across the rows of M, and E represents the residual errors. In cases such as the ones studied here, where the rows of M contain mass spectra of different HA, the scores provide information on trends in the tendency of grouping, whilst the loadings correspond to trends in the m/z values. Data pre-processing is often applied in chemometrics, especially when it is necessary to eliminate differences (variability) related either to variables and/or samples. In this work, each of the three different HA mass spectra were recorded under the same experimental conditions and in the same mode. Before the data pre-processing, the average of these three mass spectra was calculated and the intensities of the peaks were ranged between 0 and 100. The average value keeps the underlying natural information and variability [40]. When different HA fingerprints are compared using chemometrics tools, each mass-to-charge value is considered to be an independent variable. Then it is important that there are no shifts of the peaks. In order to avoid shifts of the peaks, precise calibration must be done. According to the profiles of the mass spectra of the HA samples described before, the m/z values in the range of 120–1000 have been considered for the multivariate analysis (Table 2).

Principal component analysis was applied to reduce the dimensionality of the data and as a display technique of the underlying structure of the data. PCA calculations were carried out on the correlation matrix. The cluster analysis was performed using the Euclidean distance and the Ward's linkage method available in the STATISTICA software.

Table 2
Review of selected peaks in the m/z region between 700 and 900

m/z $[M+H]^+$	Soil IHSS standard	Ant 100 m	Chemapex standard	Peat IHSS standard
717	+	+	+	+
733	+	+	−	+
745	+	+	+	+
758	+	+	+	+
773	+	+	+	+
785	+	+	+	+
799	+	+	+	−
801	−	−	−	+
813	−	+	+	+
815	+	−	−	−
827	+	+	+	−
829	−	−	−	+
841	+	+	+	+
853	−	+	−	−
855	+	−	+	+
867	+	+	−	−
870	+	−	+	+
881	+	+	+	+
901	−	−	+	−
907	+	+	−	+
918	−	−	+	−

(+) Present; (−) not present.

3. Results and discussion

3.1. TOF mass spectrometry

First of all, we have focused on the study of the conditions required to reach reproducible mass spectra. The use of several matrices (MALDI) and also the LDI mode were examined. In spite of the fact that Mugo and Bottaro [17] recently succeeded in ionising fulvic acids (FA), again we have found in HA that all the common matrices (sinapinic, dithranol, α-cyano-4-hydroxycinnamic, gentisic, 2-(4-hydroxyphenylazo)-benzoic (HABA) and, ferulic acids, etc.) are not suitable. These matrices form aggregates themselves (e.g. ferulic acid) or form aggregates with the HA.

Therefore, we have again turned to the LDI mode as already described [9] and used in several works [12,32]. In this work, various parameters, such as laser power, mode, concentration and age of the HA solutions were studied in detail.

3.1.1. Optimisation of LDI experiments

During LDI-TOF MS analysis, there are several crucial factors, such as laser energy, concentration and age of humic acid solutions. The most important thing was to find the optimal energy of the laser. At low energy levels of the laser, no ionization of HA was observed. When a certain value of laser energy (threshold) was used, the ionization started. It is interesting to note that mostly compounds with low m/z values were observed in the mass spectra (m/z 69–400 Da) at or slightly higher than the threshold energy, while the majority of these peaks were the same for all studied HA. During a moderate increase of the laser energy ($\sim$10–20% higher than the threshold) no fragmentation was observed, but a group

of peaks with m/z values $\sim m/z$ 800 Da were formed. At a much higher energy of the laser (almost twice of the optimal value), the spectra were destroyed. Many fragments with low m/z values were observed and the high resolution of the specific group of peaks at m/z 800 Da was lost as well.

To conclude: by applying the laser energy only slightly higher ($\sim$20%) than the threshold value, reproducible mass spectra are obtained. The optimal energy is $\sim$50–60 (arbitrary units; maximum is 180). We also studied the effect of the HA concentrations and it was found that, at quite low HA concentrations (less than 0.1 mg ml^{-1}), it was hard to reach ionization and get reproducible mass spectra. The optimal concentration was 0.5–1 mg ml^{-1}.

In total, 24 samples of HA, including those of lignin, were analysed. The procedure for the analysis is given in Sections 2.2 and 2.3 above. Fig. 2A–D shows the complete mass spectra obtained for different HA samples. Labelled peaks in the mass spectra (Fig. 2A–D) are those observed with relatively high intensities. The mass spectra apparently show no significant differences between plant, soil, peat, and coal derived HA. Most of the peaks are observed in the m/z region ranging between 120 and 1000 Da. This region seems to be representative for the majority of the HA samples and it is the same as that observed by Gajdošová et al. for other HA samples [8]. It can be observed from the mass spectra that in the m/z range between 717 and 918, similar structural groups (patterns) of the peaks are observed (Fig. 2A–D). We are again proving here that the HA components are of relatively low molecular weight. These results are in agreement with those already published [9].

From the analysis of the mass spectra and from the mass spectra modeling (Fig. 3A and B) concerning the family of peaks around m/z 830 Da, it is obvious that there are at least 10–12 families of peaks with the difference most similar to the one of 14 Da. Similar patterns with $\Delta = 14$ were observed for many other HA samples of another origin [8,9,13,14,20]. Such difference can be associated with the presence of aliphatic carboxylic groups in the structure of HA and/or with a different length of alkyl $-(CH_2)-$, multiple $-(CH_2)_n-$ (e.g. waxes), etc. This finding is in agreement with one published elsewhere [9,7]. Peaks separated by 1–2 Da spacing (c.f. an example of magnification at Fig. 3A) can be associated with ring structures and different degrees of saturation. Based on the experimental results, possible empirical formulas of, e.g. $[M+H]^+$ peak m/z 813.41 Da were calculated and the simulation of mass spectra for possible HA constituents was performed (Fig. 3B) via computer modeling using the AXIMA Kratos software. Several empirical formulas constituted by carbon, oxygen and hydrogen were suggested and theoretical mass spectra of such compounds were compared to those observed experimentally (Fig. 3A). We have found that the empirical formula of the compound at m/z $\sim$813.41 Da (in tolerance $m/z = 0.05$ Da) could be: $C_{32}H_{61}O_{23}^+$, $C_{33}H_{65}O_{22}^+$, $C_{342}H_{69}O_{21}^+$, $C_{36}H_{61}O_{20}^+$, $C_{37}H_{65}O_{19}^+$, $C_{38}H_{69}O_{18}^+$, $C_{40}H_{61}O_{17}^+$, $C_{41}H_{65}O_{16}^+$, $C_{43}H_{57}O_{15}^+$, $C_{44}H_{61}O_{14}^+$, $C_{45}H_{65}O_{13}^+$, $C_{47}H_{57}O_{12}^+$,

$C_{48}H_{61}O_{11}^{+}$, $C_{49}H_{65}O_{10}^{+}$, $C_{50}H_{53}O_{10}^{+}$, $C_{51}H_{57}O_{9}^{+}$, $C_{52}H_{61}O_{8}^{+}$, $C_{54}H_{53}O_{7}^{+}$, $C_{32}H_{61}O_{23}^{+}$, $C_{55}H_{57}O_{6}^{+}$ $C_{56}H_{61}O_{5}^{+}$, $C_{58}H_{53}O_{4}^{+}$, $C_{59}H_{57}O_{3}^{+}$, $C_{61}H_{49}O_{2}^{+}$, $C_{62}H_{53}O^{+}$, $C_{32}H_{61}O_{23}^{+}$, $C_{63}H_{57}^{+}$ (in total 25 possible formulas and another 22 including sodium). However, there are many other possible formulas and isomeric forms which it is not possible to distinguish because of the instrument's capabilities. In order to distinguish between them, a higher resolution and further research is needed. In addition, a group of the peaks (cf. Fig. 3A and B) around m/z $[M + H]^{+}$ 813.4 Da cannot correspond to a single compound but must mean overlapping of at least five compounds with a similar structure (Fig. 3B). However, there is another model which can also explain this group of peaks (around m/z 813.4 Da) with the same m/z tolerance. For example, such a model is the overlap of the peaks concerning the mixture of ions: $C_{42}H_{62}O_{14}Na^{+}$, $C_{42}H_{63}O_{14}Na^{+}$, $C_{42}H_{64}O_{14}Na^{+}$,

$C_{42}H_{60}O_{14}Na^{+}$, $C_{42}H_{59}O_{14}Na^{+}$. Such Na^{+} adducts were also suggested by Kujawinski et al. [14]. However, from the achieved resolution it is not possible to distinguish between $C_{42}H_{62}O_{14}Na^{+}$ and $C_{45}H_{65}O_{13}^{+}$, etc. The total number of possible compounds in the group between m/z 745.4 and 881.4 Da can be up to 2820 ($12 \times 5 \times 47$). This phenomenon (family of the peaks at ∼800 Da) was observed for all of the HA samples (soil, peat, coal derived) independent of the country or continent. It means that all these very different HA most probably contain the same compounds.

Peaks at the same $[M + H]^{+}$ m/z values were observed in the mass spectra of all humic acids, such as 129.0, 145.03, 164.92, 198.98, 360.19, 413.01, 501.19, 557.23, 799.09, 813.09, 827.03, 853.03, and 881.07. Some of these peaks were observed in quite different HA, especially the peak at m/z 360.19. All these peaks (according to the m/z differences
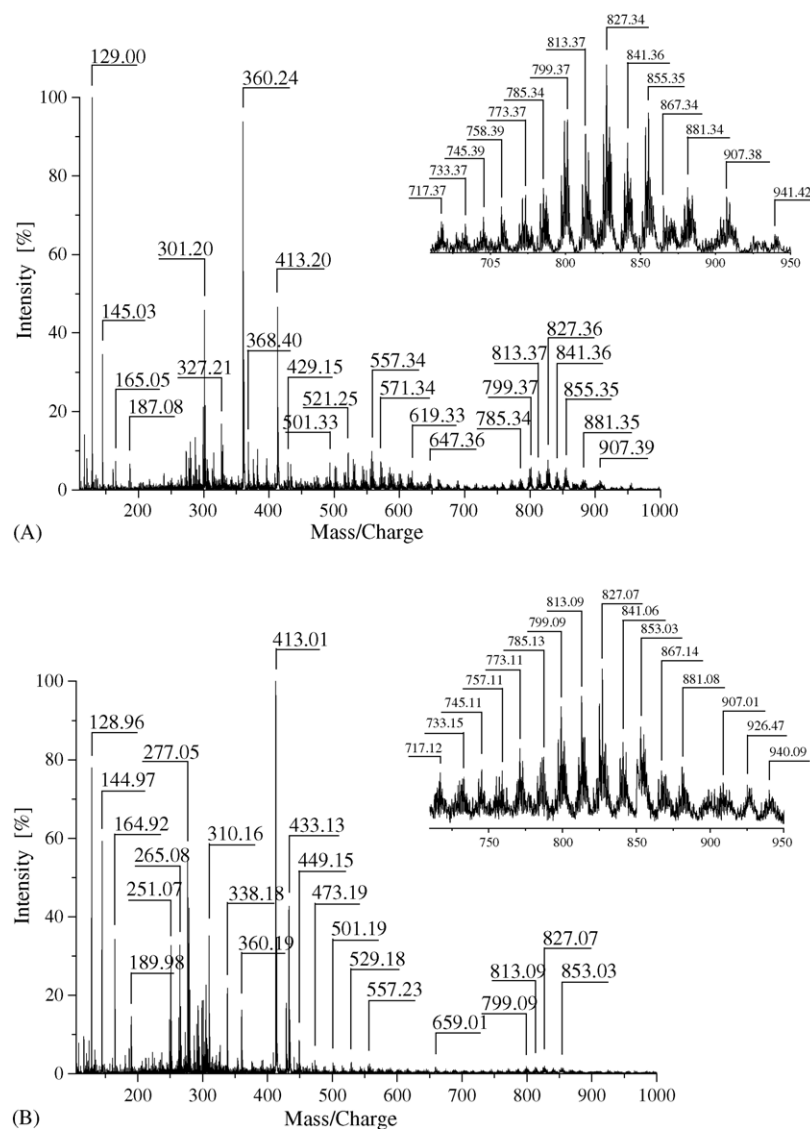

Fig. 2. Examples of LDI-TOF mass spectra obtained for different HA. (A) Soil IHSS standard; (B) Ant 100 m; (C) Chemapex standard; (D) Peat IHSS standard.
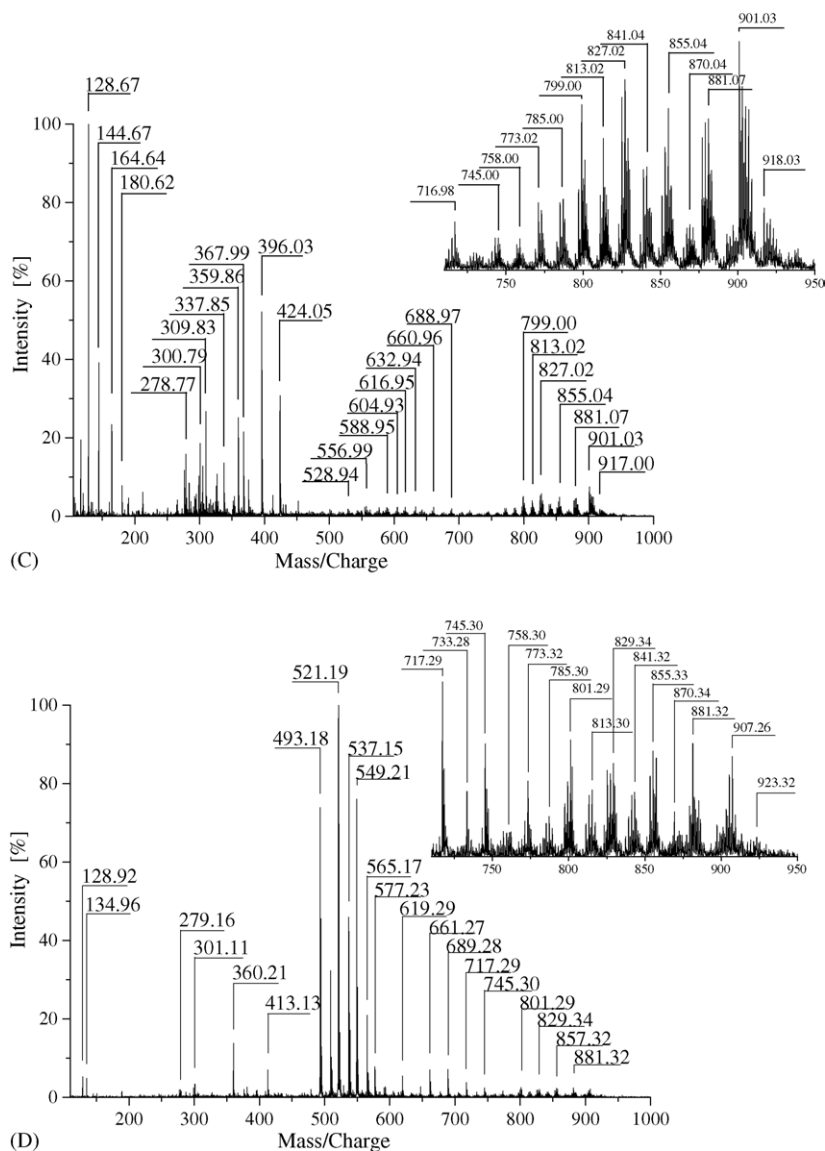
Fig. 2. (*Continued*).

in ^{13}C isotopic patterns) show that the ions have the charge equal to +1, so the *m/z* values correspond to the real molecular weight. It is interesting to note that some of the peaks with *m/z* [*M*+H]$^+$ 165, 327, 368, 360, 494, 495, 522 and 551 Da observed in the mass spectra of the HA samples studied in this work were described by Mugo and Bottaro as fulvic acids constituents [17]. This finding is also in agreement with the observation of Conte and Piccolo [41] showing disaggregation in GPC columns under different conditions. Thus, fulvic acids are very similar to HA and are probably partly constituents of HA.

Because of the high similarity of the mass spectra, where the intensities of the individual peaks only varied a little, we tried to apply chemometrics in an attempt to extract the underlying information contained in the spectra. Also, multivariate data analysis was applied to follow the tendencies in the mass spectra patterns with the aim of recognising changes in the grouping and eventually to distinguish between different HA samples.

3.2. Principal component analysis and cluster analysis of HA fingerprints

One of the best techniques to elucidate differences in humic acid structural analysis (functional detection groups) is NMR as recently reviewed by Conte et al. [42]. The difference in this work is that we are using relative peak intensities for selected *m/z* values (where the distribution is different for different HA samples and thus reflecting similarities and dissimilarities) to perform principal component analysis.

Under the current instrument capabilities and resolution achieved, pattern recognition can be applied and, with the use of PCA, the information can be extracted to highlight similarities or non-similarities in the mass spectra of HA.
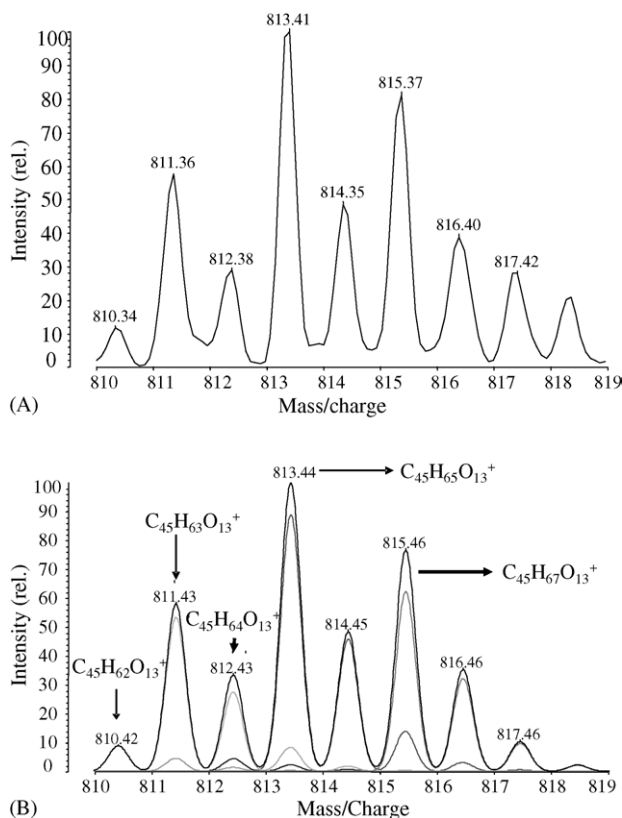
Fig. 3. (A) Amplification of LDI-TOF MS mass spectra $\sim$830 m/z for soil IHSS standard HA measured after 7 days of preparation; (B) model of mass spectra concerning the peak 813.14 presents in the range $\sim$830 m/z $[M+H]^+$ values. Overlapping of $C_{45}H_{63}O_{13}^+$, $C_{45}H_{64}O_{13}^+$, $C_{45}H_{65}O_{13}^+$, $C_{45}H_{66}O_{13}^+$, and $C_{45}H_{67}O_{13}^+$ ions is assumed (resolution 1500).

The use of PCA at even lower resolution is justified in the literature, e.g. [43,44].

PCA is an unsupervised pattern recognition technique that represents the *n*-dimensional data set in a smaller number of dimensions. Moreover, the PCA method allows the study of data sets where the number of features (variables) is greater than the number of samples and the variables are highly correlated. Generally, the results from PCA are shown as loadings and scores plots. The problem of PCA analysis of HA is to ascribe loadings, on PC1, PC2 and the other PCs, to the structural information that influence most of the *m/z* peaks. This is quite difficult because of the lack of information concerning the structure of HA. Since the eigenvalue is proportional to the percentage of variance captured by each eigenvector, each principal component accounts for a certain percentage of the variability. The first ten principal components (PCs) with eigenvalues greater or equal to 1 describe 90% of the cumulative variance, Fig. 4A, where the PCs are shown in decreasing order of their eigenvalues. The scores plot for LDI mass spectra of HA projected on the first three PCs, shows the relationship between the HA mass spectra in the data set, Fig. 4B. The PCs represent new variables that keep all the information about the data set. This means that the number of variables required to explain the variance of the data has
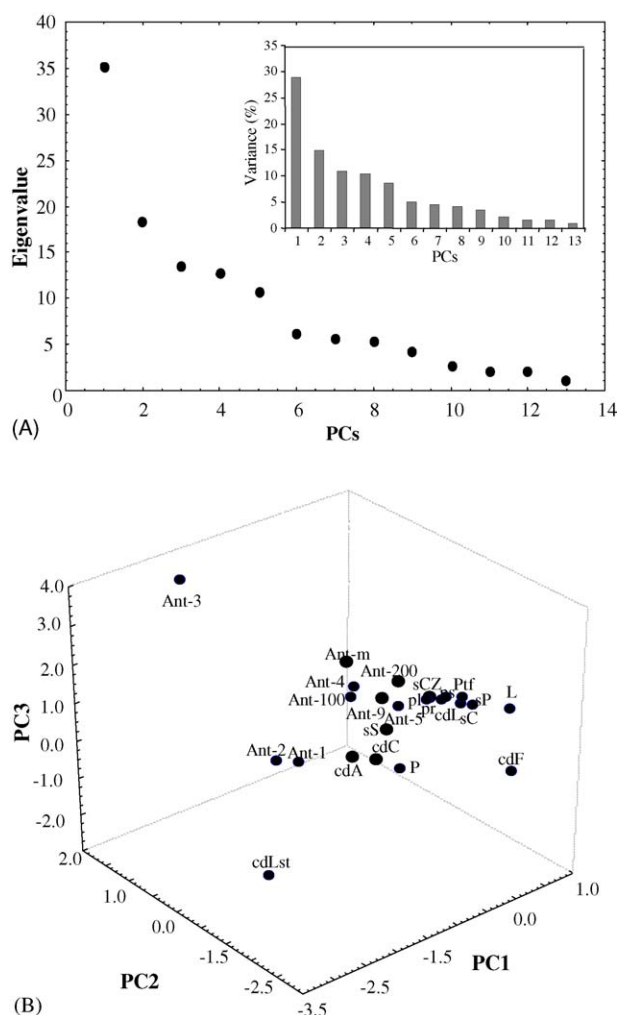

Fig. 4. (A) Eigenvalues and distribution of variance (%) on the PCs; (B) the scores plot for LDI-TOF mass spectra of HA projected on the first three principal components.

been greatly reduced. This allows observation of trends in the grouping of the HA samples which can define the structure of the data set. According to the PCA theory, the samples that have similar scores on all PCs are either identical or similar or, simply, indistinguishable. Distance is used as a measure of similarity. Those samples similar to each other in all principal components keep the same information. To simplify and improve the information, the scores plot was projected on PC1–PC2 and PC2–PC3 planes (Fig. 5A and B). Projections of the mass spectra on the first two PCs (PC1–PC2) allowed a rough separation of the HA studied by their origin. Nevertheless, the main group of samples is distributed along the PC1 where HA from Antarctica (Ant-1, Ant-2, Ant-5, Ant-9, Ant-100, Ant-200, Ant-m) are all well distinguishable from the other samples along PC2. However, one sample from Antarctica (Ant-3) looks to be quite different from the others from the same sampling place. Coal derived HA (cdC, cdA and cdF) are not grouped by the origin of HA. In Fig. 5A, peat HA (P) and coal derived Leonardite HA stan-
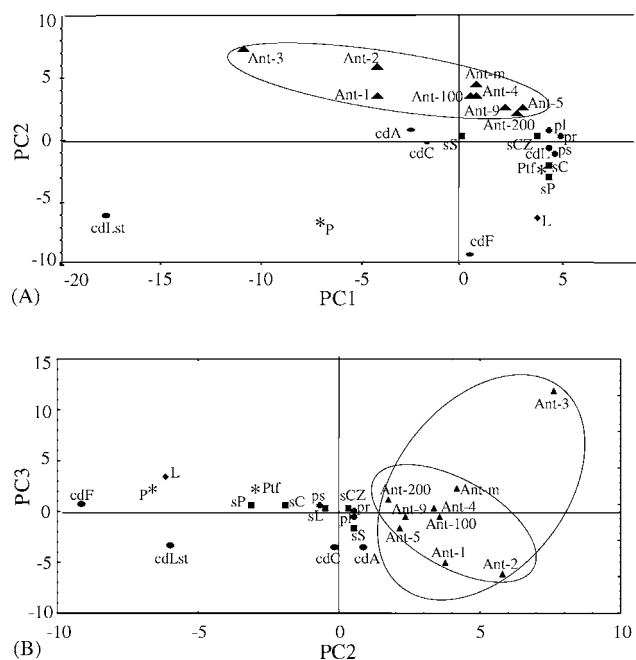
Fig. 5. Scores plot of H samples projected on (A) PC1–PC2, (B) PC2–PC3. (■) Soil HA; (●) coal derived HA; (▲) Antartica HA; (*) Peat HA; (♦) lignin.

dard (cdLst) are localised in the plane PC1 versus PC2 and coal derived HA Fluka (cdF) on PC1 versus PC2 plane. These samples do not follow the trend of the other HA, and may possibly be considered as outliers. Fig. 5B confirms the previous results and proves the category as outliers for cdLst and Ant-3.

As for lignin (L), the average sample of lignin is localised in the plane PC1 versus PC2, near the group of samples constituted by most of the plant, soil and peat HA. On the other hand, lignin is not close to the groups of HA from Antarctica. This is very significant and the finding is in agreement with the fact that there are no lignin containing plants in Antarctica (with the exception of two rare species of flowering plants, namely hair grass *Deschampsia antarctica* and pearlwort *Colobanthus quitensis*, which are, however, very rare). If plant residues are considered to be the main source of humus, the most chemically resistant lignin part of the plant becomes the matrix for programming the chemical and structural features of raising the humic substances. This structure yields different alterations with aging over a geological time scale, different botanical taxons and alteration to the soil sediment [11,40]. With respect to the Antarctica samples, Wilson et al. [3] and Beyer et al. [4] have already suggested formation of HS without lignin precursors from plants, and our findings confirm this.

There was also an attempt, taking into account the chemical composition of HA samples, to apply PCA to the LDI-TOF mass spectra information. However, no improvement in the PCA results has been achieved in comparison with those previously described above.
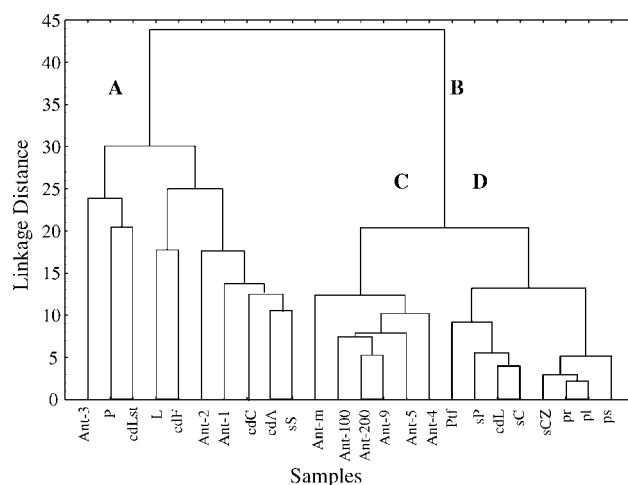

Fig. 6. Cluster analysis dendrogram for LDI-TOF mass spectra of HA samples.

3.3. Cluster analysis

Results obtained from CA were then plotted as a dendrogram in Fig. 6. In the dendrogram, two clusters, A and B, are formed. The first one (A) includes lignin, peat HA, coal derived Fluka HA, coal derived Leonardite HA standard, Soil HA standard, coal derived Chemapex HA, Ant-1, Ant-2, and Ant-3. The second one includes all the other HA samples and is divided into two subgroups, C and D. Subgroup C includes most of the samples from Antarctica (Ant-4, Ant-5, Ant-9, Ant-100, Ant-200, and Ant-m). Subgroup D includes most of the samples which belong to the categories of soil HA (Brazilian Cananeia HA, and Pindorama HA), Brazilian peat, and plant derived HA (water hyacinth leaves, roots and stems). This association of HA extracted from plant with those extracted from soils and peat may be related by the fact that plants containing HA are considered to be the sources for soil origin through composting [36].

4. Conclusions

LDI-TOF MS proved to be a suitable ionization technique to obtain reproducible and characteristic mass spectra of HA. It was found that the mass spectra of humic acids present similar patterns for all the studied HA. This means that even HA of different origin have the same structural units or that they contain the same components. We can also conclude that LDI-TOF MS under optimised conditions produces mass spectra where several of the low *m/z* values are identical to those observed by MALDI and ESI [41,45]. It was also proved that many constituents of HA from different continents and of various origins are of the same or quite similar structure.

However, we have to realise that not all compounds in such complicated mixtures (as HA) can be ionized or also that

ionisation of some of them can be depressed by the presence of other compounds.

By applying multivariate techniques (PCA, CA) it was found that these methods are useful for obtaining information out of the mass spectra recorded in the LDI-TOF MS mode for the different HA studied, and that the HA samples were classified. In this way, by applying chemometrics, the differences between the samples can be elucidated. It was also found that, e.g. the Antarctica HA mass spectra fingerprints differ to some extent from the others. This finding is in agreement with different humification processes in the arctic conditions and also confirms that most of the present-day Antarctica humic substances studied from sampling sites might mostly come from lichens and lower plants. This is also in agreement with the mineralogical analysis of soils showing that the sediments are rather young.

Acknowledgements

References

[1] K.H. Tan, Humic Matter in Soil and the Environment Principles and Controversies, Marcel Dekker, New York, USA, 2003, pp. 127–168.
[2] I. Schepetkin, A. Khlebnikov, B.S. Kwon, Drug Develop. Res. 57 (2002) 140.
[3] E.M. Peña-Méndez, J. Havel, J. Patočka, J. Appl. Biomed. 3 (2005) 13.
[4] M. Remmler, A. Georgi, F.-D. Kopinke, Eur. J. Mass Spectrom. 1 (1995) 403.
[5] G. Davies, E.A. Ghabbour (Eds.), Humic Substances, Structures, Properties and Uses, The Royal Society of Chemistry, Cambridge, 1999.
[6] M.A.K.M. Wilson, P.J. Collin, L.G. Greenfield, Org. Geochem. 9 (1986) 225.
[7] L. Beyer, H.-P. Blume, C. Sorge, H.-R. Schulten, H. Erlenkeser, D. Schneider, Arctic Alpine Res. 29 (1997) 358.
[8] D. Gajdošová, L. Pokorná, K. Láska, P. Prošek, J. Havel, in: G. Davies, E.A. Ghabbour (Eds.), Humic Substances Structures, Models and Functions, The Royal Society of Chemistry, Cambridge, 2001, pp. 121–131.
[9] J. Havel, D. Fetsch, E.M. Peña-Méndez, P. Lubal, J. Havliš, Recent Developments in Humic Acid Characterization. Acidobasic and Complexation Properties, Separation and Reliable Fingerprints by Capillary Electrophoresis and MALDI-TOF Mass Spectrometry. Ninth International Meeting of the IHSS, Adelaide 1998, in: R.S. Swift, K.M. Spark (Eds.), Understanding and Managing Organic Matter in Soils, Sediments, and Waters, International Humic Substances Society, Hyde Park Press Atlanta, 2001, pp. 77–82.
[10] A.C. Stenson, W.M. Landing, A.G. Marshall, W.T. Copper, Anal. Chem. 74 (2002) 4397.
[11] A.C. Stenson, A.G. Marshall, W.T. Copper, Anal. Chem. 75 (2003) 1275.
[12] D. Gajdošová, L. Pokorná, J. Havel, in: G. Davies, E.A. Ghabbour (Eds.), Humic Substances Versatile Components of Plants, Soil and Water, The Royal Society of Chemistry, Cambridge, 2000, pp. 133–149.
[13] L. Pokorná, D. Gajdošová, J. Havel, Characterization of humic acids by capillary zone electrophoresis and matrix assisted laser desorption ionisation time of flight mass spectrometry, in: E.A. Ghabbour, G. Davies (Eds.), Understanding Humic Substances: Advanced Methods, Properties and Applications, RSC, December, 1999, pp. 107–119.
[14] E.B. Kujawinski, M.A. Freitas, Xu Zang, P.G. Hatcher, K.B. Green-Church, R.B. Jones, Org. Geochem. 33 (2002) 171.
[15] M.L. Pacheco, J. Havel, Electrophoresis 23 (2002) 268.
[16] T.L. Brown, J.A. Rice, Org. Geochem. 31 (2000) 627.
[17] S.M. Mugo, C.S. Bottaro, Rapid Commun. Mass Spectrom. 18 (2004) 2375.
[18] M. Remmler, A. Georgi, F.D. Kopinke, Eur. Mass Spectrom. 1 (1995) 403.
[19] G. Haberhauer, W. Bednar, M.H. Gerzabek, E. Rosenberg, Royal Society of Chemistry 259 (2000) 143 (special publication).
[20] M.L. Pacheco, J. Havel, New results from Capillary Electrophoresis and MALDI TOF MS studies of humic acids interactions with various compounds and xenobiotics, in press.
[21] V. Schoojans, D.L. Massart, J. Pharm. Biomed. Anal. 26 (2001) 225.
[22] H. Yoshida, R. Leardi, K. Funatsu, K. Varmuza, Anal. Chim. Acta 446 (2001) 485.
[23] G. Mazeroless, M.F. Devaux, E. Dufour, E.M. Qannari, Ph. Courcoux, Chemom. Intell. Lab. Syst. 63 (2002) 57.
[24] H. Wilcken, C. Sorge, H.-R. Schulten, Geoderma 76 (1997) 193.
[25] I. Rozenbaha, G. Odham, U. Järnberg, T. Alsberg, M. Klarins, Anal. Chim. Acta 452 (2002) 105.
[26] M. Geyer, F.A.-H. Hemidi, L. Brüggemann, G. Hanschmann, Thermochim. Acta 361 (2000) 139.
[27] P.K. Egeberg, S.O. Bergli, J. Chromatogr. A 950 (2002) 221.
[28] D. Fetsch, M. Hradilová, E.M. Peña-Méndez, J. Havel, J. Chromatogr. A 817 (1998) 313.
[29] Y.P. Chin, G. Aiken, E. O'Loughlin, Environ. Sci. Technol. 28 (1994) 1853.
[30] G.R. Aiken, R.L. Malcom, Geochim. Cosmochim. Acta 51 (1987) 2177.
[31] L. Pokorná, D. Gajdošová, S. Mikeska, P. Homoláč, J. Havel, in: G. Davies, E.A. Ghabbour (Eds.), Humic Substances Structures, Models and Functions, The Royal Society of Chemistry, Cambridge, 2001, pp. 133–152.
[32] A. Fievre, T. Solouki, A.G. Marshall, W.T. Cooper, Energy Fuels 11 (1997) 554.
[33] M.D. Landgraf, R.C. Alonso, R. Oliveira, J. Cap. Elec. 5 (1998) 193.
[34] M.L. Pacheco, E.M. Peña-Méndez, J. Havel, Chemosphere 51 (2003) 95.
[35] GBM, Grupo Bioquímico Mexicano, S.A. de C.V., Blvd. Dr. Jesus Valdes Sanchez no. 2369, Saltillo, Coahuila, C.P. 25290, Mexico.
[36] E.A. Ghabbour, G. Davies, G.-Y. Yuen Lam, M.E. Vozzella, Environ. Pollut. 131 (2004) 445.
[37] Personal communication of J. Štelcl and P. Kadlec from Department of Geography, Faculty of Science, Masaryk University, Brno, Czech Republic.
[38] L. Pokorná, D. Gajdošová, S. Mikeska, P. Homoláč, J. Havel, in: G. Davies, E.A. Ghabbour (Eds.), Humic Substances Structures, Models and Functions, The Royal Society of Chemistry, Cambridge, 2001, pp. 133–149.

[39] E.B. Kujawinski, P.G. Hatcher, M.A. Freitas, Anal. Chem. 74 (2002) 413.
[40] J.I. Hedges, D.C. Mann, Geochim. Cosmochim. Acta 43 (1979) 1803.
[41] P. Conte, A. Piccolo, Chemosphere 38 (1999) 517.
[42] P. Conte, R. Spaccini, A. Piccolo, Prog. Nucl. Magn. Reson. Spectrosc. 44 (2004) 215.
[43] J. Arunachalam, S. Gangadharan, Anal. Chim. Acta 157 (1984) 245.
[44] P.J. de Groot, G.J. Posma, W.J. Melssen, L.M.C. Buydens, V. Deckert, R. Zenobi, Anal. Chim. Acta 446 (2001) 71.
[45] P. Tinoco, G. Almendros, F.J. Gonzáles-Vila, J. Anal. Appl. Pyrolysis 64 (2002) 407.