# Classification analysis of humic substances based on data of $^{13}C$ NMR spectroscopy, elemental analysis, and size exclusion chromatography

Andrey I. Konstantinov[1], Alexey V. Kudryavtsev[1], Irina V. Perminova[1], Elena Yu. Belyaeva[1], and Norbert Hertkorn[2]

[1]Department of Chemistry, Lomonosov Moscow State University, Leninskie Gory 1-3, 119992 Moscow, Russia, e-mail: konstant@org.chem.msu.ru,

[2]GSF-Research Center for Environment and Health, Institute of Ecological Chemistry, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany.

Being the products of stochastic synthesis, humic substances (HS) are composed of mixtures of structurally close macromolecules. This leads to non-stoichiometric elemental composition and heterogeneous structure. As a result, HS are still “operationally defined”. Nevertheless, it has been numerously reported that humic fractions (humic and fulvic acids, HA and FA) isolated from the same source (coal, peat, fresh or marine waters) display striking structural similarities [1]. The objectives of this study were, first, to generate structural descriptors using data of elemental analysis, $^{13}C$ NMR spectroscopy and size exclusion chromatography (SEC), and, second, to evaluate their discriminating power upon classifying a large set of humic fractions isolated from different sources.

## Materials and methods

*Eighty humic materials* were used isolated from three different sources: soil, peat and coal. Soil humic materials included 7 FA and 27 HA. Peat humic materials included 9 FA, 11 HA, and 13 non-fractionated HS. Coal humic materials included 5 isolated HA and 8 commercial coal humates. *$^{13}C$ solution state NMR spectra* were acquired using Bruker AC-400 and Bruker Avance-400 spectrometers operating at 400 MHz proton frequency. The spectra were recorded on HS samples dissolved in 700 μl 0.1 M NaOD at concentrations of 80-100 mg/ml. $^{13}C$ NMR spectra were acquired with a 5 mm broadband probe, using CPMG pulse program (relaxation delay: 8 s at $90^o(^{13}C) = 9.4$ μs). Nine partials integrals were calculated using following assignments, ppm: C=O (220-187); COO (187-165); $C_{AR}$-O,N (165-145); $C_{AR}$-C,H (145-110); O-C-O,N (110-90); CH-O,N (90-64); $CH_2$-O,N (64-58); $CH_3O$ (58-48); and $CH_n$ (48-0). The integrals were used as structural descriptors. *Elemental analysis* and *SEC* conditions were as described in [2]. Atomic ratios, average molecular weights ($M_w$) and polydispersity ($M_w/M_n$) were used as structural descriptors. *Classification* analysis was conducted using linear discriminant analysis (LDA) and “*K* nearest neighbours” (KNN) methods. Software “Statistica” (StatSoft company) was used for LDA, and “Regression” software (© A.V. Kudryavtsev) was used for KNN analysis.

**Results and Discussion**

The results of classification analysis are presented in Table 1. As it can be seen from the table, both classification techniques yielded very similar results. However, given the shorter calculation times, the preference could be given to a use of LDA technique.

**Table 1.**

Results of classification analysis using descriptors generated from the data of $^{13}$C NMR, elemental analysis and SEC

| **Data** | **Linear Discriminant Analysis** | | ***K* nearest neighbours** | |
|---|---|---|---|---|
| | Classification of test samples, % | Descriptors* | Classification of test samples, % | Descriptors* |
| $^{13}$C NMR | 62 | CHX, $CH_3O$, COO, $C_{AR}X$, $C_{AR}$, $CH_2X$, C=O | 67 | $CH_n$, $CH_3O$, CHX, $C_{AR}X$, COO |
| $^{13}$C NMR +SEC | 68 | CHX, $M_w$, $C_{AR}X$, COO, $CH_3O$, $C_{AR}$, OC-X, $CH_2X$, $M_w/M_n$, $CH_n$ | 68 | $CH_n$, $CH_3O$, CHX, $C_{AR}X$, COO, $M_w$ |
| $^{13}$C NMR +Elemental analysis | 84 | N, $C_{AR}$, O, O/C, CHX, $C_{AR}X$, COO, C=O, $CH_2X$, H, $CH_3O$ | 84 | CHX, $C_{AR}$, C=O, N, O, O/C |
| $^{13}$C NMR + EA+SEC | 89 | N, CHX, O, O/C, $C_{AR}X$, Mw, COO, $CH_3O$, OC-X, $CH_n$, $CH_2X$, C=O | 89 | $CH_3O$, $C_{AR}$, N, O/C, $M_w$, $M_w/M_n$ |

*X in structural groups means N or O.

The $^{13}$C NMR descriptors used in this study did not reveal high discriminating power in classifying HS according to source and fractional composition. Extension of $^{13}$C NMR descriptors through SEC data did not improve quality of classifications. At the same time, a substantial improvement was achieved with a use of elemental-composition descriptors. LDA classifications show that nitrogen content had the highest discriminating power among all structural descriptors used. This could explain low discriminating power of $^{13}$C NMR data, which do not distinguish between O- and N-containing structural fragments. Determination of N speciation in HS can provide a promising tool for classification analysis of HS.

**Acknowledgement**: This research was supported by ISTC (project KR-964) and DOE (project RUC2-20006).

**References**

1. Stevenson, F. Humus Chemistry. John Wiley and Sons. New York. 1994.
2. Perminova, I.V., Frimmel, F.H., Kovalevskii, D.V., et al. *Wat. Res.* 1998, 32, 872-881.