frontiers

Supplementary Material

Characterization of the colloidal properties of dissolved organic matter from forest soils

Contents (20 pages, 3 tables, 9 figures)

S1.1. Elemental analysis

Supplementary Table S1. Extended chemical characterization of DOM samples from the extraction series analyzed by DLS and SAXS. Total carbon (TOC), nitrogen (TN), iron (Fe) and phosphorus (P) are reported in mg/L. The samples analyzed in section 3.2 are grey-shaded.

	Total DOM[1]						Dialyzed DOM[2]						
Sample	**TOC**	**TN**	**Fe**	**P**	**C/N**	**pH**	**TOC**	**TN**	**Fe**	**P**	**C/N**	**pH**	**% of DOM[3]**
FH_{35}	1083	71	3.6	24.4	15	4.5	417	14	0.5	1.2	30	4.9	39
FH_{61}	964	37	1.5	12.1	26	3.5	368	5.7	0.5	1.2	65	4.2	38
FH_{90}	820	37	1.5	12.1	22	3.6	297	6.3	0.6	1.2	47	4.3	36
GH_{ref35}	500	55.4	3.6	4.0	9	5.0	219	14	2.1	1.3	16	4.9	44
GH_{ref61}	466	49.5	5.6	3.4	9	5.0	207	12.2	3.4	1.2	17	4.9	44
GH_{ref90}	523	42.5	5.0	1.6	12	6.0	210	8.3	2.8	0.6	25	4.6	40
FC_{35}	74	8.5	0.6	4.6	9	5.4	55	2.4	0.5	0.3	23	4.6	74
FC_{61}	48	2.5	0.3	0.2	19	5.7	24	0.8	0.2	0.1	30	4.7	50
FC_{90}	29.5	2.2	0.3	1.6	13	5.6	18	0.9	0.2	0.1	20	4.9	61
GC_{ref35}	9.5	8.3	0.5	0.9	1	6.3	5	0.4	0.4	0.2	13	4.5	53
GC_{ref61}	9.3	5.1	0.7	0.7	2	6.2	4	0.4	0.5	0.2	10	5.1	43
GC_{ref90}	16.6	1.6	1.2	0.5	10	6.2	9	0.3	0.9	0.2	30	5.0	54

[1] DOM obtained from the respective extraction procedure

[2] High-molecular weight fraction remaining after dialysis (>12-14 kDa).

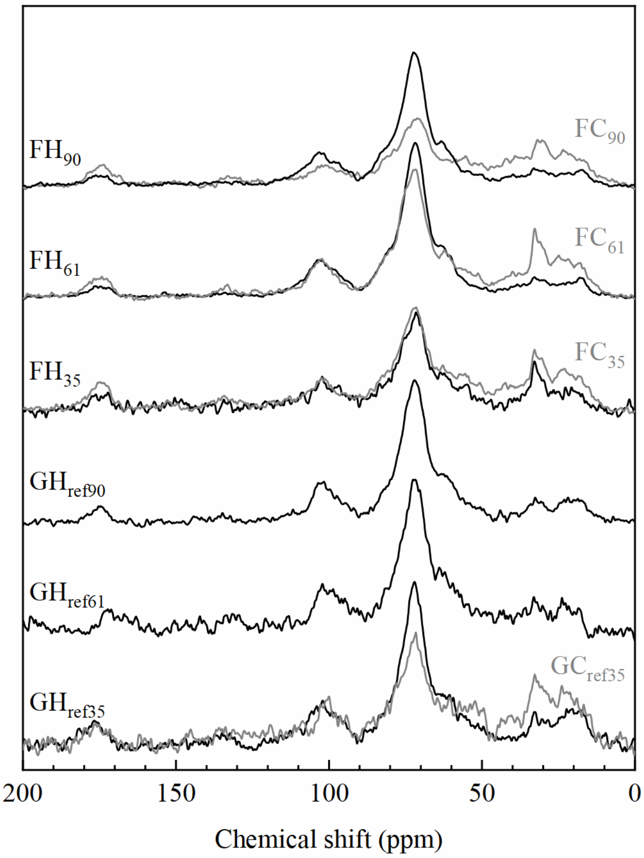
[3] % of total DOM based on TOC.

S1.2. Solid-state CP-MAS ^{13}C NMR

The CP-MAS ^{13}C NMR spectra obtained showed pronounced resonances at 50-80 ppm and around 100 ppm indicative of O-alkyl carbon atoms (Fig. 2A, Fig. S1). These resonances are in agreement with cellulose-like materials and thus suggested a predominance of carbohydrates in the DOM samples (Kamida et al., 1984; Gustavsson et al., 2014). The NMR spectra also contained resonances at 22 and 33 ppm that likely originated from aliphatic terminal carbons ($–CH_3$ and $(−CH_2−)_n$) of fatty acids and/or acetyl and alkyl groups in carbohydrates (Wilson, 1981; Krosshavn et al., 1992; Gilardi et al., 1995), while the peaks between 165 and 190 ppm originated from carboxylic acids, esters and/or amides ($COOH$, COO^-, $COOR$, $CONH_2$) (Wilson, 1981). The signals from aromatic compounds, such as lignin fragments, in the region 110-165 ppm were weak, but relatively more pronounced in FC_{61} than FH_{61} (Fig. 2A, Fig. S1). The weak aromatic signals can, of course, be due to relatively low concentrations of aromatics, but may also be the result of long relaxation times of aromatic carbon atoms compared to other functional groups. Therefore the fraction of aromatic carbon may be underestimated (Wilson, 1981).

It should be noted that spectral intensity from CP-MAS experiments is not directly proportional to the carbon concentration because it is dependent on the structural environment and the cross polarization from surrounding ^{1}H spins. Thus, the technique is generally considered to be semi-quantitative (Preston, 1996; Simpson et al., 2011; Mao et al., 2017). Still, we can make relative comparisons of the NMR peak intensities, and although the TOC concentrations of FH and FC samples were very different (Table S1) the overall NMR spectral profiles were similar indicating similar chemical composition and only minor temperature selectivity during extraction. However, some differences were observed namely that relative to the carbohydrates more aliphatic and slightly more oxidized and aromatic carbon atoms were present in FC than FH samples (Fig. 2A; Fig. S1).

Supplementary Figure S1. Solid-state CP-MAS ^{13}C NMR spectra of hot (black) and cold (gray) DOM extracts. The measurements were performed on freeze-dried DOM. Spectra of reference samples GC_{61} and GC_{90} were not recorded because of the very low total carbon concentrations.

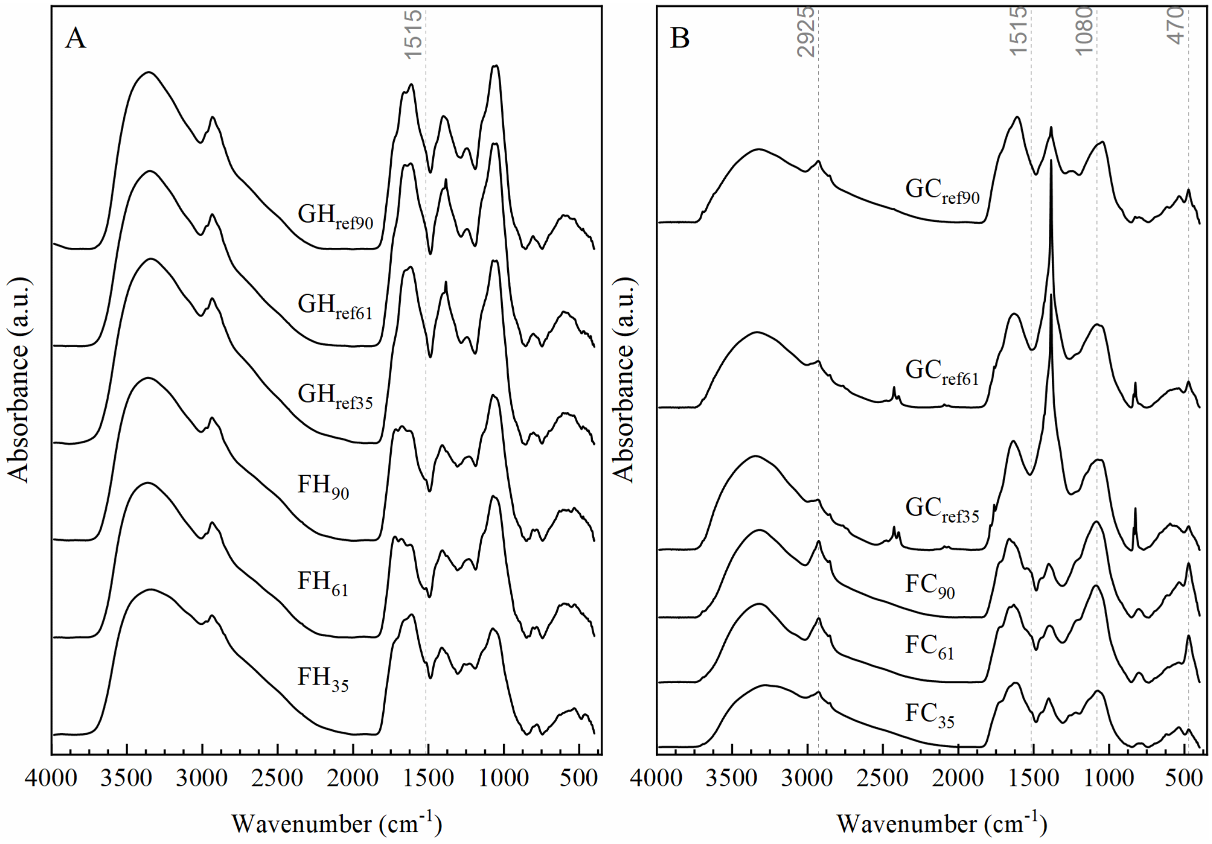


S1.3. IR spectroscopy

In agreement with the ^{13}C NMR results, the IR spectra were dominated by bands between 1000 and 1150 cm^{-1} indicative of carbohydrates (Fig. S2). The stretching mode of $–CH_2–$ at 2925 cm^{-1} indicated more aliphatic structures in FC than FH, except for the 35 year old soil (Fig. 2B;

Fig. S2) (Gu et al., 1994; Silverstein et al., 2005). Moreover, a band at ca. 1390 cm^{-1} and some of the complex features around 1600 cm^{-1} likely originated from the symmetric and asymmetric carboxylate modes, respectively. The IR spectra also contained evidence of carbonyl-like groups between 1690-1750 cm^{-1} and C–OH vibrations of protonated carboxylics and/or phenolics between 1200-1300 cm^{-1} (Stevenson and Goh, 1971; Artz et al., 2008). In contrast to the NMR data the IR spectra showed a more obvious sign of aromatic, lignin-like structures indicated by the pronounced shoulder at 1515 cm^{-1}. In line with the ^{13}C NMR results, this shoulder was relatively stronger in all FC as compared to FH samples (Fig. S2B). The spectra of FC samples also displayed notable IR spectral contributions from clay minerals such as kaolinite (Nayak and Singh, 2007; Chukanov and Chervonnyi, 2016) e.g. at 3695 and 470 cm^{-1} (Fig. 2B and S2B). Moreover, the shift of the main peak of FC_{61} at ca. 1080 cm^{-1} as compared to FH_{61} was likely caused by overlap with the strong Si-O modes originating from clays.

Supplementary Figure S2. Diffuse reflectance IR spectra of (A) hot and (B) cold water extracted DOM. Noticeable is the strong contribution at 1378 cm^{-1} from inorganic nitrate in some of the reference samples. The dotted vertical line at 1515 cm^{-1} highlights the presence of aromatic lignin-like compounds. The DOM samples were lyophilized prior to measurements. Tentative peak assignments are present in Table S2.



Supplementary Table S2. Tentative assignments of the principal IR absorption bands in DOM samples (Stevenson and Goh, 1971; Gu et al., 1994; Silverstein et al., 2005; Nayak and Singh, 2007; Artz et al., 2008; Chukanov and Chervonnyi, 2016).

Wavenumber (cm^{-1})	**Assignments**	**Compound or functional group**
3340	O–H stretching	Polysaccharides, alcohols
2925/2940	C–H stretching of $–CH_2$ and $–CH_3$ groups	Fats, wax, lipids
1720-1725	C = O stretch of COOH or COOR	Carboxylic acids, ketones, aromatic esters
1675-1600	Aromatic C=C stretching and/or asymmetric C–O stretch in COO–	Lignin and other aromatics, or aromatic or aliphatic carboxylates
1515	Aromatic C=C stretching	Lignin/phenolic backbone
1460, 1378	C–H deformations of $–CH_3$ and $–CH_2$	Phenolic (lignin) and aliphatic structures
1410	Symmetric C–O stretch from COO– or stretch and OH deformation (COOH)	Carboxylate/carboxylic structures
1378	N=O stretching	Nitrate
1265	C–O stretching of phenolic OH	Indicative of lignin backbone
1150-1000	Combination of C–O stretching and O–H deformation	Carbohydrates,polysaccharide-like substances, phosphate anions
825	Out of phase bending of C–H	Polysaccharides
500-700	Out of plane N–H wagging (broad band) and Si–O–H vibrations	Primary and secondary amides and clays
470	Silica bending of Si-O-Si	Kaolinite ($Al_2Si_2O_5$) together with a band at 3695 cm^{-1} (O–H stretch)

S1.4. ^{1}H NMR experiments

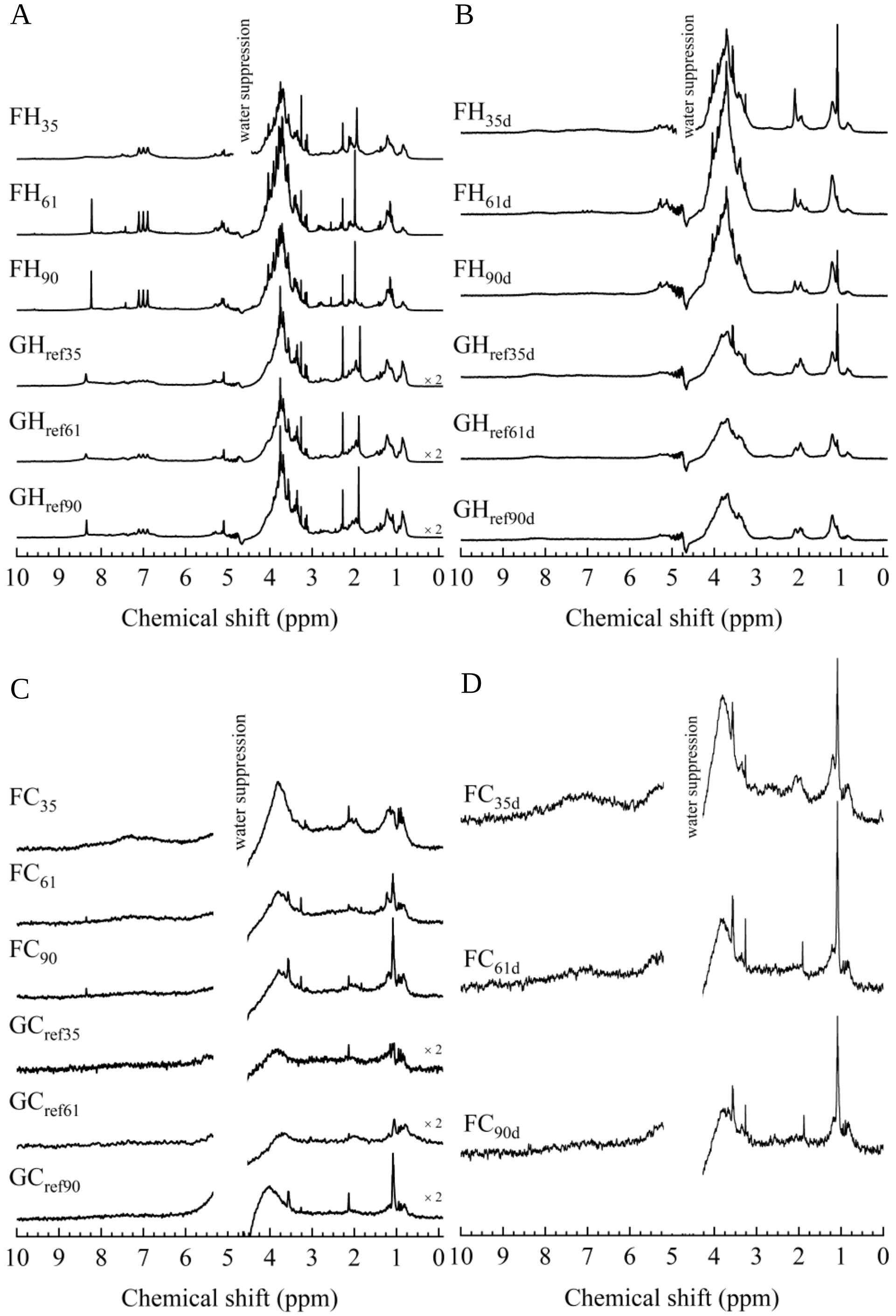
The appearance of solution ^{1}H NMR spectra depends strongly on the molecular mobility. Hence, small or flexible molecules typically give rise to narrow sharp resonances while resonances associated with large and slowly tumbling aggregates can experience significant line broadening that in the extreme case cause the signal to be inseparable from the baseline. For solid-like aggregates, one may also expect residual dipolar couplings in the kHz regime making resonances from these aggregates essentially undetectable. Moreover, any signal that overlaps with the main water peak is removed by the excitation sculpting technique. It follows that although resonances in the ^{1}H NMR spectra are proportional to proton concentrations the technique may be blind to some of the compounds in the DOM samples.

The ^{1}H solution NMR spectra of FH and FC extracts were dominated by carbohydrate resonances in the range from 3.0 to 4.3 ppm (Fig. 3A and B; Fig. S3) (Wilson, 1981), thus corroborating the ^{13}C NMR results. In addition we observed aliphatic $–CH_3$ and $–CH_2–$ protons at 0.8 and 1.2 ppm, respectively, as well as the presence of low molecular weight organic acids presumably formic acid at 8.23 ppm and acetic acid 1.98 ppm (Sleighter et al., 2015). It is notable that we extracted substantially more of the low molecular weight acids in FH than FC extracts despite the fact that these acids should be soluble at room temperature and the slightly acidic pH value of the latter sample. This indicates that it is not only pure solubility that controls the concentrations in the DOM extracts but also the disruption of structures such as plant cells and soil aggregates. At 7 ppm we detected a characteristic triplet displaying three equally intense peaks separated by 0.105 ppm (52.5 Hz). This spectral feature is consistent with the presence of ammonium ions, NH_4^+, where there is a direct J-coupling to the spin 1 ^{14}N nucleus. Typically, aqueous ammonium and amine protons are not detectable due to their short spin-spin relaxation times, which is caused by rapid proton exchange with water at neutral pH resulting in a very broad signal (Preece and Cerdan, 1993). At acidic pH values, however, the proton exchange slows down. We observed the NH_4^+ triplet in FH_{61} and only a broad signal in FC_{61} reflecting the lower pH in the former (Table S1). The presence of ammonium in soil extracts has been reported previously but its source remains unclear (Kaiser et al., 2003). Proposed explanations include the formation of ammonium through the ammonification of organic nitrogen or atmospheric deposition of ammonium salts (Aarnes et al., 2007).

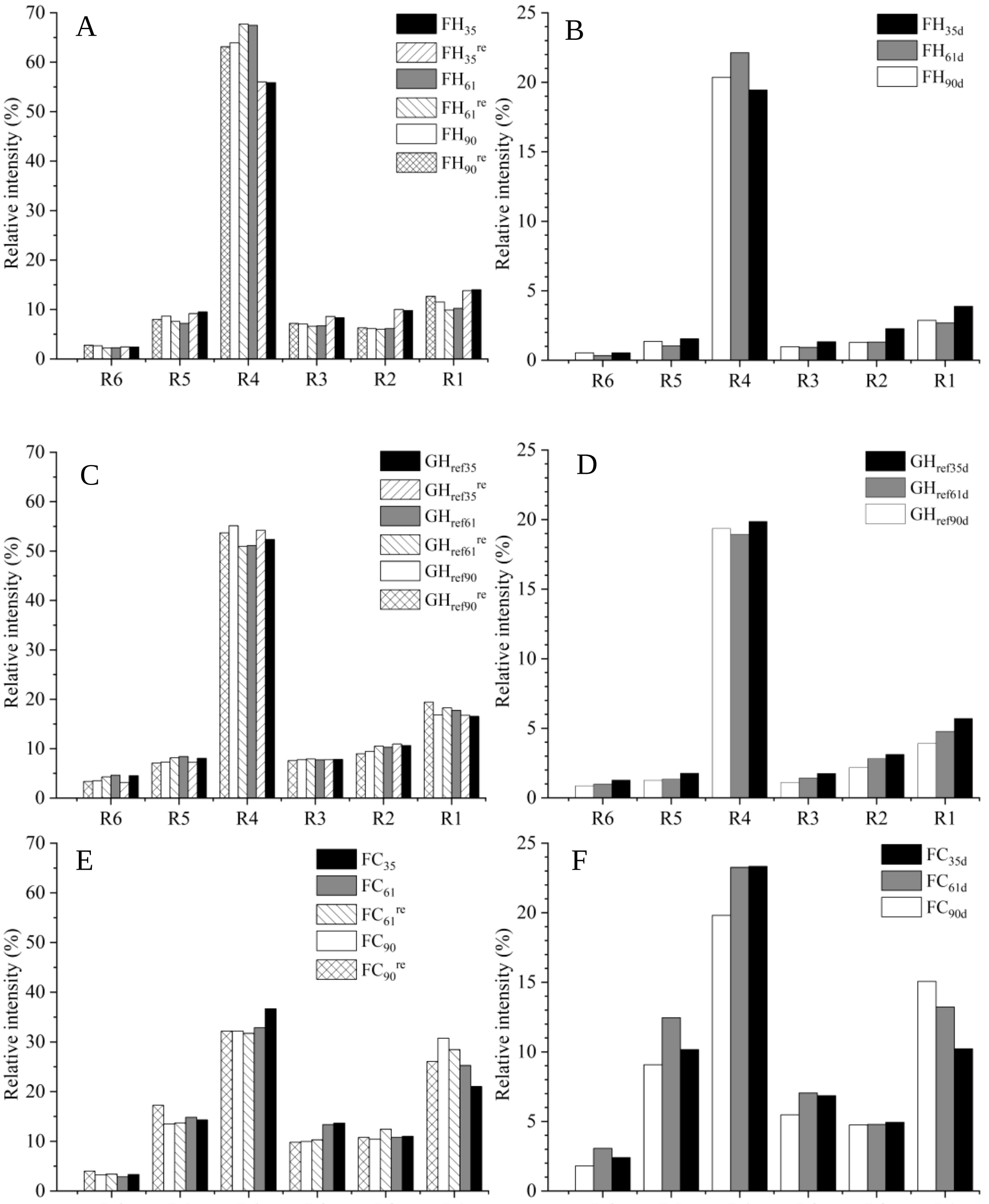
The total area under the ^1H NMR spectrum of FC_{61} was 3 % of the total area of FH_{61}, which is comparable to the difference in TOC (Table S1). The relative concentrations of DOM compounds observed in the spectra were quantified by dividing the spectra into six resonance intervals (Fig. 3): R1) 0.6–1.65 ppm (alkyl, $R–CH_3$, $R–CH_2–R$); R2) 1.65–2.2 ppm (acetate, CH_3COO^- and $CH_3COO–R$); R3) 2.2–3.0 ppm (HC–C=Y, α protons to carboxyl, carbonyl groups, and aromatic ring, HC–X, protons next to nitrogen or halogen atom); R4) 3.0–4.3 ppm (HC–O–R, α protons to carbon attached to oxygen groups, carbohydrates, alcohols, ether); R5) 6.0–7.8 ppm (Ar–H, aromatic protons including quinones, phenols and ammonium protons, N–H); R6) 7.8–9.0 ($HCOO^-$ (formate), sterically hindered protons of aromatics, nitrogen heteroaromatics) (Wilson, 1981; Whitty et al., 2019). These calculations confirmed that carbohydrates were a major component in FH and FC extracts corresponding to 56-68 % and 32-37 %, respectively, of the total spectrum integral. Also, aliphatic compounds (R1 + R2) were a common DOM component, especially in FC_{61} and FC_{90}.

Dialysis of FH and FC caused a decrease in intensities of the ^1H NMR spectra and the sharp peaks from small organic acids and ammonium disappeared completely (Fig. 3 and Fig.S3). Still, several of the characteristic features remained after dialysis, which indicated that both low molecular weight compounds and structures/aggregates larger than the 4 nm cut-off of the dialysis membrane contributed to the ^1H NMR spectra of the un-dialyzed DOM samples. The total area of the FH_{61} and FC_{61} decreased by 64 % and 28 %, respectively, and this change was similar to the change in TOC (Table S1) suggesting a rough correlation between TOC and the NMR resonances and that a substantial fraction of the DOM was recorded in the ^1H NMR spectra. Evaluation of the relative intensity changes showed that the carbohydrate content of dialyzed FH_{61} decreased from 68 % to 22 % (Region 4, Fig. S4) while the effect was less pronounced for FC_{61} with a decrease from 33 % to 23 %. Moreover, dialysis caused a change in the shape of the spectral carbohydrate features of FH_{61} removing most of the sharp resonances that likely originated from low molecular weight carbohydrates. In contrast, the broad carbohydrate resonances of FC_{61} remained practically unchanged. These results point to that FH_{61} contained a wide size range of carbohydrates from small molecules to large colloids whereas FC_{61} contained mainly high molecular weight carbohydrates. We observed similar results in all our FH and FC extracts (Table S1, Fig. S3 and S4).

Supplementary Figure S3. High-resolution ^{1}H-NMR spectra of total DOM hot (A) and cold (C) and the corresponding dialyzed DOM samples (B) and (D). GH and GC spectra were magnified for better comparison.



Supplementary Figure S4. Comparison of peak area integrations of ^{1}H-NMR spectra of DOM shown in Fig. S3. The integrated spectra of dialyzed DOM (B, D, F) were normalized to the corresponding total (undialyzed) DOM spectra area. The regions are: 1 – from 0.6 to 1.65 ppm, 2 - from 1.65 to 2.2 ppm, 3 – from 2.2 to 3.0 ppm, 4 – from 3.0 to 4.3 ppm, 5 – from 6.0 to 7.8 ppm, 6 – from 7.8 to 9.0 ppm. The subscript “re” denotes a replicate.



Supplementary Table S3. Summary of neutral sugars after acid hydrolysis (values given in polymeric form) of cold DOM[**]

Sample	TOC	TN	**Concentration (mg/L)**[*1,2,3,5]									
			Fucose	**Galactose**	**Glucose**	**Xylose**	**Mannose**	**Rhamnose**	**Arabinose**	**Total (polymeric)**	**(rha+fuc) /(ara+xyl) ratio**	**(gal+man)/ (ara+xyl) ratio**
FC$_{35}$	63	15	8.7 (0.3)	18.5(1.3)	19.6 (1.7)	7.6(0.2)	23.3(1.6)	11.7(1.3)	7.5(0.6)	96.9	1.35	2.77
FC$_{61}$	139	11	6.9(0.2)	16.0 (1.1)	18.6(1.5)	7.8(0.8)	18.6(1.4)	9.6(0.3)	6.3(0.3)	83.8	1.17	2.45
FC$_{90}$	106	7	4.9(0.6)[*4]	14.3(1.1)	14.5(1.3)	5.8 (0.8)	13.8(1.2)	7.1(0.8)	3.7(0.5)	64.1	1.26	2.95
GC$_{ref35}$	20	42	0.42(0.07)	3.6(0.4)	4.1(0.3)	1.5(0.2)	1.6(0.3)	2.4(0.3)	2.8(0.2)	16.4	0.66	1.21
GC$_{ref61}$	29	33	0.4(0.1)	3.1(0.3)	3.0(0.3)	2.2(0.3)	0.4(0.1)	1.0(0.2)	1.2(0.1)	11.3	0.41	1.03
GC$_{ref90}$	15	29	0.2(0.07)	2.1(0.2)	2.7(0.2)	1.3(0.3)	0.5(0.1)	1.5(0.1)	1.4(0.05)	9.7	0.63	0.96

[*1] Analysis using HPAEC-PAD with CarboPac-PA20, calculations based on calibration standards.

[*2] All the values are represented in the polymeric form of carbohydrates after anhydro correction.

[*3] Forest samples contained suspended material.

[*4] The values in the parenthesis indicate the standard deviation of triplicates of the same extract.

[*5] Ribose was not detected by HPAEC-PAD.

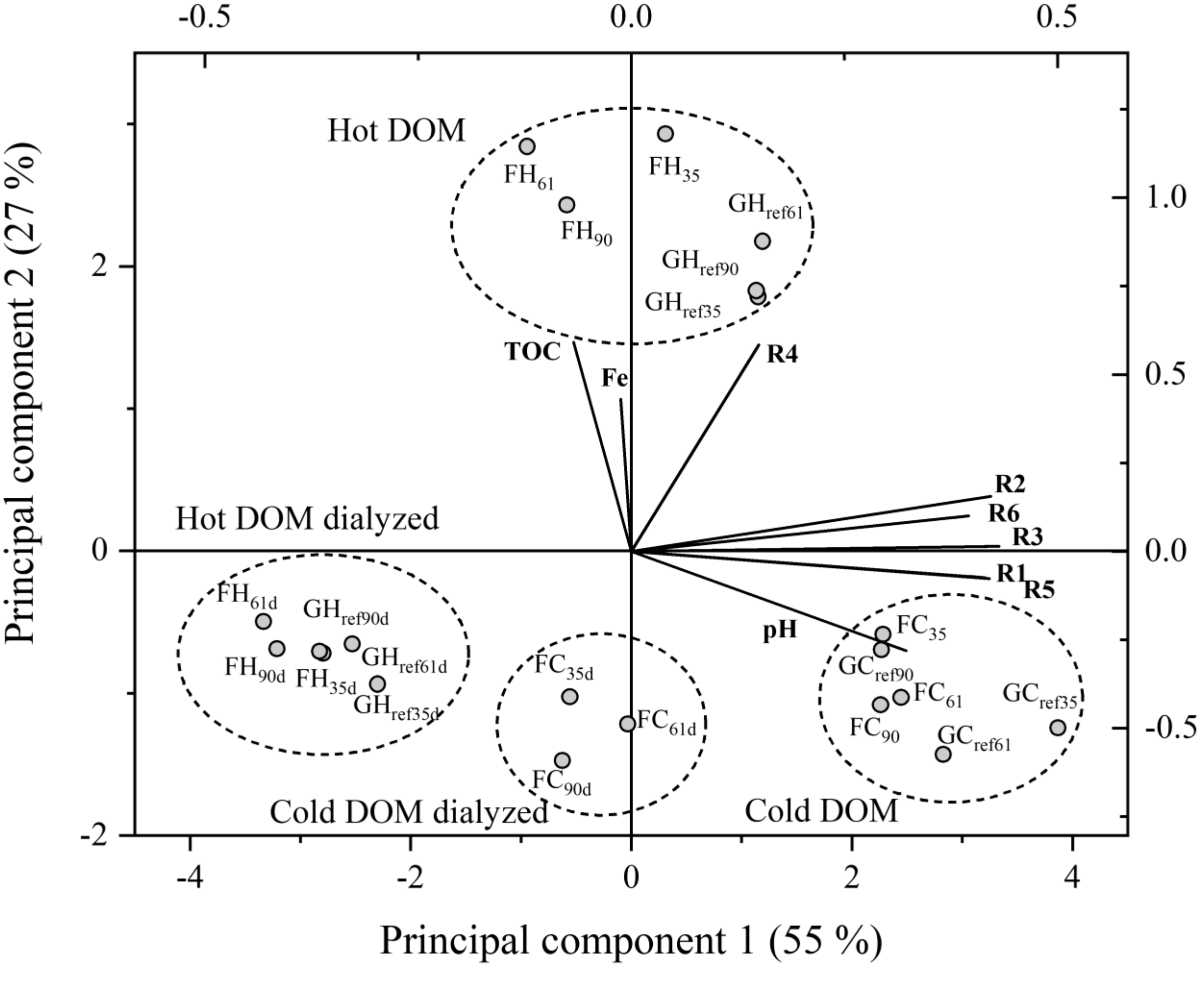
(rha+fuc)/(xyl+ara) = ratios of rhamnose plus fucose to xylose plus arabinose.

(man+gal)/(xyl+ara) = ratios of mannose plus galactose to xylose plus arabinose.

**For the sugar analysis we prepared separately fresh cold DOM extracts from the soils that were collected in the same area in April 2019. The soil to water ratio was changed to 1:2 w/v to obtain higher concentration of extracts needed in the analysis.

S.1.5. Principal Component Analysis (PCA)

Figure S5. Principal component analysis biplot of TOC, Fe and relative intensity in the six ^{1}H NMR regions in DOM samples. The left and bottom axes are the PCA scores of the samples, while the right and top axes display the loadings of the principal components. Abbreviations are as follows: F and G = forest and grassland reference plots; H and C = hot or cold water extractions; numbers correspond to forest stand age; subscript *d* indicates whether the sample was dialyzed, e.g., FH_{35d} = dialyzed forest 35 hot DOM. The four groups of samples are highlighted schematically with circles (note: the circles do not correspond to confidence ellipses).



S2.1. Analysis of DLS data

The use of DLS to determine particle size has increased significantly in recent years, partly due to the introduction of simple compact benchtop instruments, with user-friendly interfaces. The DLS experiments yield the so-called autocorrelation function $g_2(q, \tau)$, where q is the magnitude of the scattering vector and τ the lag time (delay time), and typically this function is analyzed to obtain a hydrodynamic radius, R_H, or a distribution of radii in the case of polydisperse solutions. The relative concentrations of different particles in a distribution are weighted by their scattered intensity. Hence, an average hydrodynamic radius corresponds to the intensity-weighted average, often referred to as the z-averaged hydrodynamic radius, $<R_H>_z$. As the scattering power of a particle typically depends on the square of the particle volume or weight, $<R_H>_z$ is strongly biased towards the larger sizes in a broad distribution. In addition, in case of a chemically heterogeneous mixture, different species may have different scattering power (different contrast), which further increases the number of parameters affecting the analysis of DLS data.

The data analysis software of DLS instruments often offer the option to transform the intensity weighted distribution into a volume weighted distribution or a number distribution. The idea is to reduce the heavy bias towards larger particles. This option, however, involve some significant assumptions. It is assumed that the sample is chemically homogeneous (same contrast for all particles) and that the particles are homogeneous spheres with R_H equal to a physical radius and thus have particle volumes given by $v = (4\pi/3)R_H^3$. However, applying this to complex samples such as DOM from soils or aquatic systems that possibly contain particles of different chemistry and various unknown shapes that exhibit unknown relationships between R_H and v, is questionable and may lead to erroneous interpretations. Therefore, we have analyzed and interpreted the autocorrelation function $g_2(q, \tau)$ conservatively. From a cumulant analysis (Schurtenberger and Newman, 1993) we obtained the average diffusion coefficient and the polydispersity index, which is a measure of the heterogeneity of the colloidal size distribution in the sample. In this procedure the scattered field correlation function $g_1(q,\tau)$ is analyzed, which is obtained from $g_2(q, \tau)$ through the Siegert relation (Schurtenberger and Newman, 1993)

$$g_2(q,\tau) = 1 + \beta|g_1(q,\tau)|^2 \qquad (1)$$

where β is an experimental constant close to unity. Assuming non-interacting polydisperse colloids the electric field correlation function is given by a sum of exponentials, one for each species, i, with decay rate $\Gamma_i = q^2 D_i$. For systems with narrow size distributions, the exponentials can be expanded around an average decay rate $\langle\Gamma\rangle = q^2\langle D\rangle$ (cumulant expansion), giving (Schurtenberger and Newman, 1993)

$$g_1(q,\tau) = e^{-\langle\Gamma\rangle\tau}\left[1 + \frac{\sigma}{2}\langle\Gamma\rangle^2\tau^2 + \cdots\right] \qquad (2)$$

where $<D>$ is the average diffusion coefficient and $\sigma = \langle\Gamma - \langle\Gamma\rangle\rangle^2/\langle\Gamma\rangle^2$ is the relative variance, often is referred to as the polydispersity index (PDI). Subsequently, the z-averaged hydrodynamic radius, $<R_H>_z$, is calculated from the Stokes-Einstein equation

$$\langle D\rangle = \frac{k_B T}{6\pi\eta\langle R_H\rangle_z} \qquad (3)$$

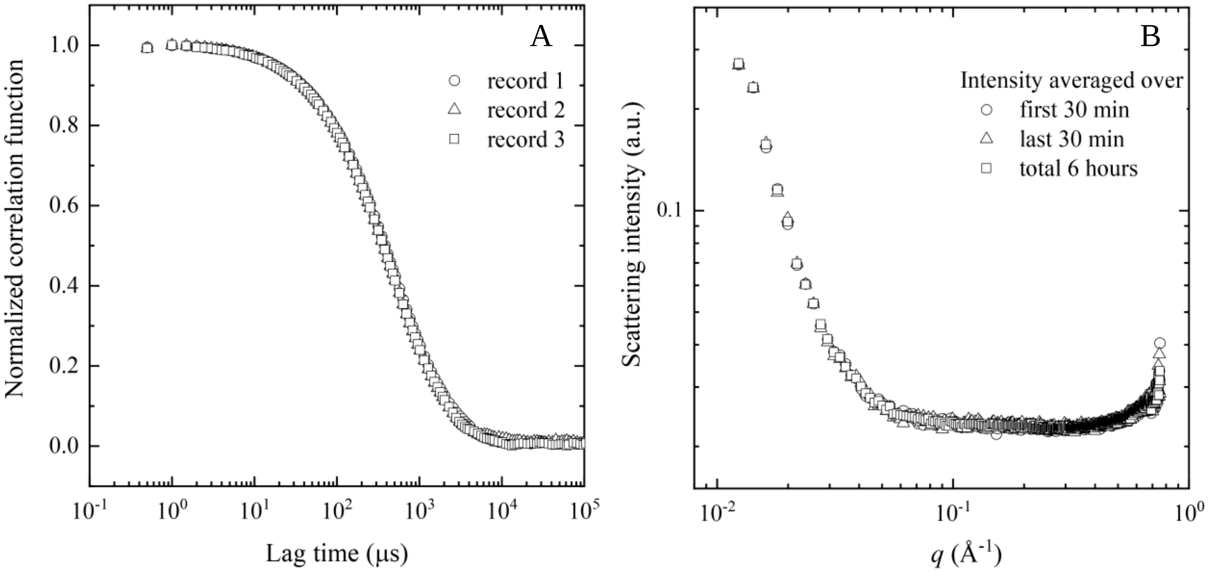
where k_B is the Boltzmann's constant, T the absolute temperature and η the solvent viscosity.

S2.2. Stability of DOM

To check whether DOM is stable during the DLS measurements the same sample was measured three times at 2 min intervals. The corresponding correlation functions ($g_2(\tau)$) are displayed in Fig. S6A and showed that there were no sample sedimentation or aggregation.

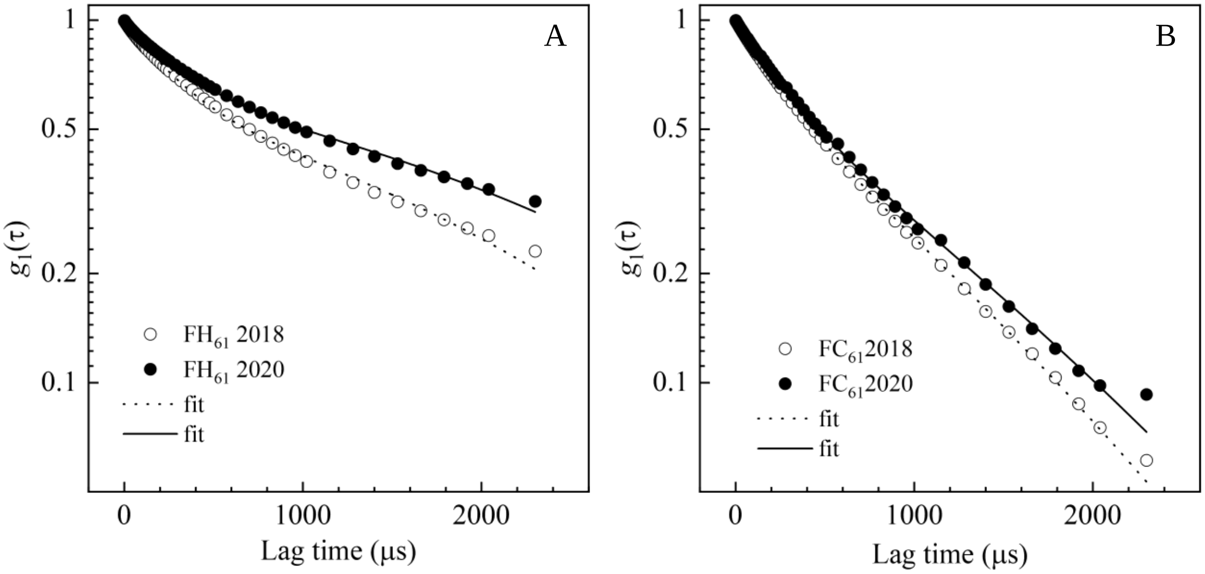
A typical SAXS experiment in the present study using a laboratory X-ray source (SAXSLab Ganesha) lasts 4 to 10 hours due to low DOM concentration. We have checked the stability of sample during these measurements by averaging the scattered intensity over different time intervals (see Fig. S6B). The result show that the scattering intensity (not corrected for the solvent scattering) averaged over first 30 min of measurement and the last 30 min overlapped with the intensity averaged over 6 hours implying that DOM was stable during the experiments.

Supplementary Figure S6. DOM stability during DLS and SAXS experiments: (A) three consecutive measurements of the FH_{61} correlation function (i.e. 2 min/record); (B) comparison of X-ray scattered intensity of FC_{61} in arbitrary units (not corrected for the solvent scattering) averaged over the first 30 min, last 30 min and over total 6 hours of measurement. Overlapping scattering profiles indicate absence of sedimentation or aggregation.



Over much longer times (up to 2 years) hot DOM extracts start to aggregate while the corresponding cold extracts remains quite stable. Fig. S7 shows the DLS results from FH_{61} and FC_{61} measured just after extraction and 2 years after extraction. From the cumulant fit the average hydrodynamic radius of FH_{61} increased $\langle R_H \rangle_z$ from 113 nm to 141 nm, while a smaller increase from 90 nm to 98 nm was observed for FC_{61}. The polydispersity index, σ, of FH_{61} increased from 0.69 to 0.89, but remained 0.40 in FC_{61}. This indicates that compounds that were extracted at high temperature start to aggregate when cooled down but that aggregation is slow.

Supplementary Figure S7. DOM stability over long time-scales. Correlation functions, $g_1(\tau)$, of hot (A) and cold (B) DOM and the corresponding cumulant fits to the data of DOM measured in December 2018 (hollow circles) and repeated in October 2020 (filled circles).



S2.3. SAXS modelling of the structure of colloidal DOM

In SAXS experiments, the scattered intensity, $I(q)$, inform on the size, structure and shape of the colloidal particles (Lindner and Zemb, 2002). For N particles within a volume V, the scattered intensity can be written as

$$I(q) = \frac{N}{V}\Delta\rho^2 V_p^2 P(q)S(q) \tag{4}$$

where $P(q)$ is the normalized single particle scattering function, the so called “form factor” of the particle; $S(q)$ is the structure factor, which is approximated to unity, as interparticle interactions can be neglected because of the low concentrations; V_p is the particle volume; $\Delta\rho$ is the scattering length density difference between particle and solvent, i.e. the contrast. We modeled the SAXS data assuming that the colloidal DOM consisted of two fractions. Linear polymer chains and mass fractal aggregates (clusters). It is further assumed that both fractions chemically are identical and correspond to carbohydrates, having the chemical formula $(C_5H_8O_4)_n$ and a mass density of 1.5 g/cm^3. Neglecting interparticle interactions, the total scattered intensity, $I(q)$, can be written as a sum of the two contributions

$$I(q) = I_{SFC}(q) + I_{agg}(q) \tag{5}$$

The linear polymer-like objects are described as semi-flexible cylinders and their scattered intensity can be written as

$$I_{SFC}(q) = c_{cyl}\frac{\Delta\rho_{cyl}^2 M_{cyl}}{d_{cyl}^2\, N_A} P_{SFC}(q) \tag{6}$$

where $\Delta\rho_{cyl} = \rho_{cyl}-\rho_{solv}$ is the scattering length density difference between the cylinder (*cyl*) and water solvent (*solv*), $\Delta\rho_{cyl} = 4\times10^{10}$ cm^{-2}; $M_{cyl} = 10^4$ g/mol is the cylinder weight averaged molar mass; $d_{cyl} = 1.5$ g/cm^3 is the cylinder mass density; N_A is Avogadro's number; c_{cyl} is the concentration that is expressed as mass per unit volume (*e.g.*, mg/l). The particle volume, V, appearing in Eq. (4) is being replaced by mass divided by mass density. The semi-flexible cylinder form factor, $P_{SFC}(q)$, taking into account excluded volume interactions, was calculated numerically, following the work of Pedersen and Schurtenberger (Pedersen and Schurtenberger, 1996; Chen et al., 2006). The overall $P_{SFC}(q)$ is characterized by a contour length, *L*, a cylinder cross section radius, *r*, a Kuhn length, λ_K, describing the flexibility, and a concentration *c*. The model includes intramolecular excluded volume effects, so that chain configurations resembles a self-avoiding random walk, leading to $I(q){\sim}q^{-5/3}$ (de Gennes, 1979). Assuming cylinder geometry, *r* is for simplicity fixed to 3.0 Å that corresponds to the effective cross section radius of hemicellulose chains. The Kuhn length, λ_K, which characterizes the stiffness of a hemicellulose chain was fixed to 20 Å equaling to 4 glucose units. The contour length, defined as $L = n_b b$, with *b* being the effective length of an anhydroglucose units (ca. 5 Å in solution (Gubitosi et al., 2016)) and $n_b = M_{cyl}/M_{C5H8O4}$ is the degree of polymerization. The gyration radius of the semi-flexible cylinders assuming aforementioned contour length is 3.8 nm giving an effective diameter of about 10 nm.

The scattering at lower q-values is dominated by the fractal aggregates. The resulting scattered intensity is written as

$$I_{agg}(q) = c_{agg}\frac{\Delta\rho_{agg}^2 M_{agg}}{d_{agg}^2 N_A} P_{agg}(q) \tag{7}$$

where $\Delta\rho_{agg}$ is the scattering length density difference, $\Delta\rho_{agg} = \Delta\rho_{cyl} = 4\times10^{10}$ cm^{-2}; $M_{agg} = 10^8$ g/mol is the aggregates weight average molar mass; $d_{agg} = d_{cyl} = 1.5$ g/cm^3 is the mass density of the aggregates and c_{agg} is the aggregates concentration. We assumed that the aggregates can be modelled as mass fractals characterized by a radius of gyration, $R_{g,agg}$, and a fractal dimension, *D*. We used the corrected Beaucage model (Hammouda, 2010) where the normalized form factor can be written as

$$P_{agg}(q) = exp\{-q^2 R_{g,agg}^2/3\} + \frac{C}{q^D}\left(erf\{qR_{g,agg}/6^{1/2}\}\right)^{3D} \tag{8}$$

with *C* being given by

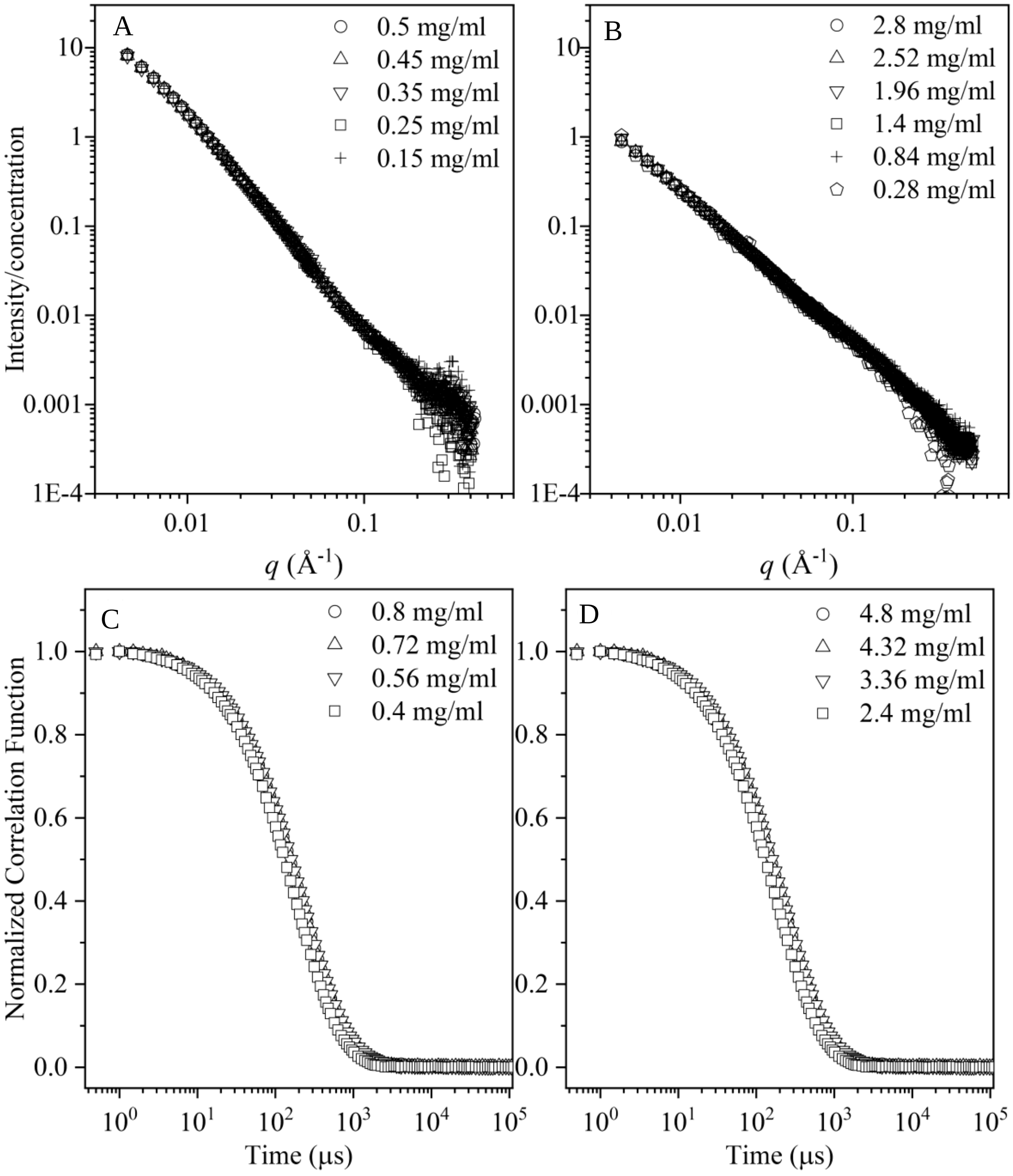
$$C = \frac{D}{R_{g,agg}^D}\left(\frac{6D^2}{(2+D)(2+2D)}\right)^{D/2}\Gamma(D/2) \tag{9}$$

where *erf*(x) is the error function and $\Gamma(x)$ is the gamma function. The mass fractal dimension was deduced from the cold DOM scattering profile and then fixed to 2.8 in all extracts. This value is similar to what has been observed in previous small angle scattering studies of soil organic matter (Österberg and Mortensen, 1992; Senesi et al., 1996; Rice et al., 1999).

Obviously, the parameter space of the model was large. To constrain the model, we make approximations and fix some of the parameters. As concluded above, the main component of DOM, including in the colloidal fraction, are carbohydrates presumably hemicelluloses. Hence, we assume that both the semi-flexible cylinders and the fractal aggregates are carbohydrates and we use the mass density value available for cellulose (Gubitosi et al., 2016), which should be similar to *e.g.* hemicellulose (Ehrnrooth, 1984). For the semi-flexible cylinders we have also fixed the cross section radius and the Kuhn length to values consistent with cellulose (Lindner and Zemb, 2002). Remaining free parameters are then the effective molar mass and concentration of the semi-flexible cylinders, and concentration, molar mass and radius of gyration of the fractal aggregates. We note that concentration and molar mass appear as a product, and we obtain a concentration by assuming a molecular weight. Moreover, the intensity in the Porod (power law) regime of the scattering pattern depends also on $R_{g,agg}$. We have therefore assumed that the gyration radius should not exceed the average size of DOM, $R_g \leq \langle R_H \rangle_z$. For globular and dense systems relationship $R_g = 0.775 \langle R_H \rangle_z$ has been well known in the literature (Richards, 1980). We therefore fixed $R_{g,agg}$ to 70 nm. This is also a reasonable number considering the aggregates are expected to have an effective radius of around 100 nm, set by the final filter pore size of 200 nm, and also confirmed by DLS ($\langle R_H \rangle_z \approx 100$ nm).

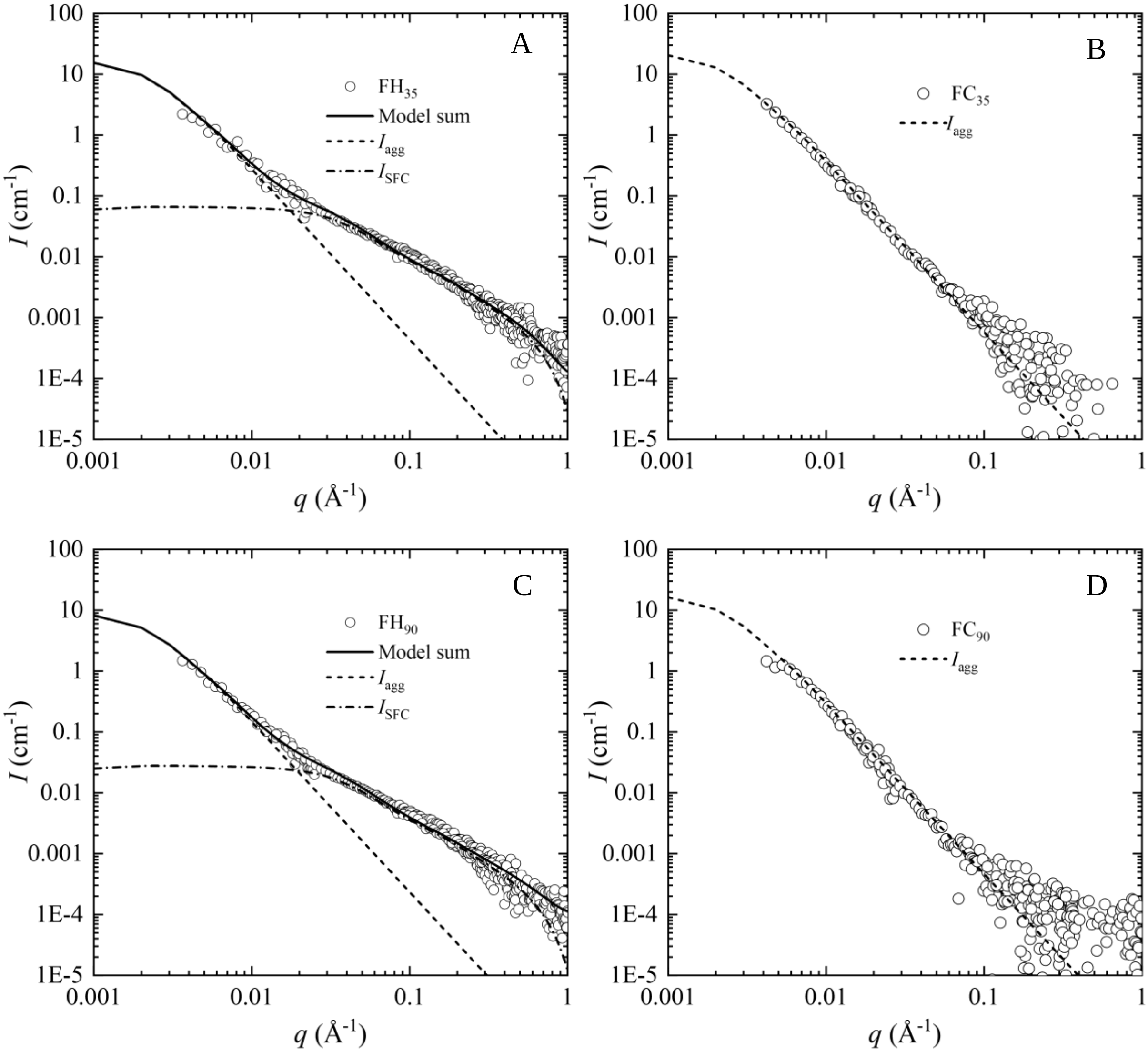
The molecular weights M_{cyl} and M_{agg} were fixed to 10^4 g/mol and 10^8 g/mol, respectively. M_{agg} was estimated from modeling the SAXS pattern of dialyzed cold DOM using the relationship c_{agg}=2.2TOC. We thus assume that all carbon left after dialysis is present in the aggregates. To obtain the molecular weight of semi-flexible cylinders present in hot DOM extracts we fixed M_{agg} and then adjusted M_{cyl} to fit the experimental data. In such a way, the experimental q-range is in the Porod regime, and we adjusted the concentrations c_{cyl} and c_{agg} in all simulations of the SAXS data: the error was roughly estimated to ± 100 mg/l for c_{cyl} and ± 10 mg/l for c_{agg}. This procedure may result in a systematic error, but it facilitates relative comparisons between the samples.

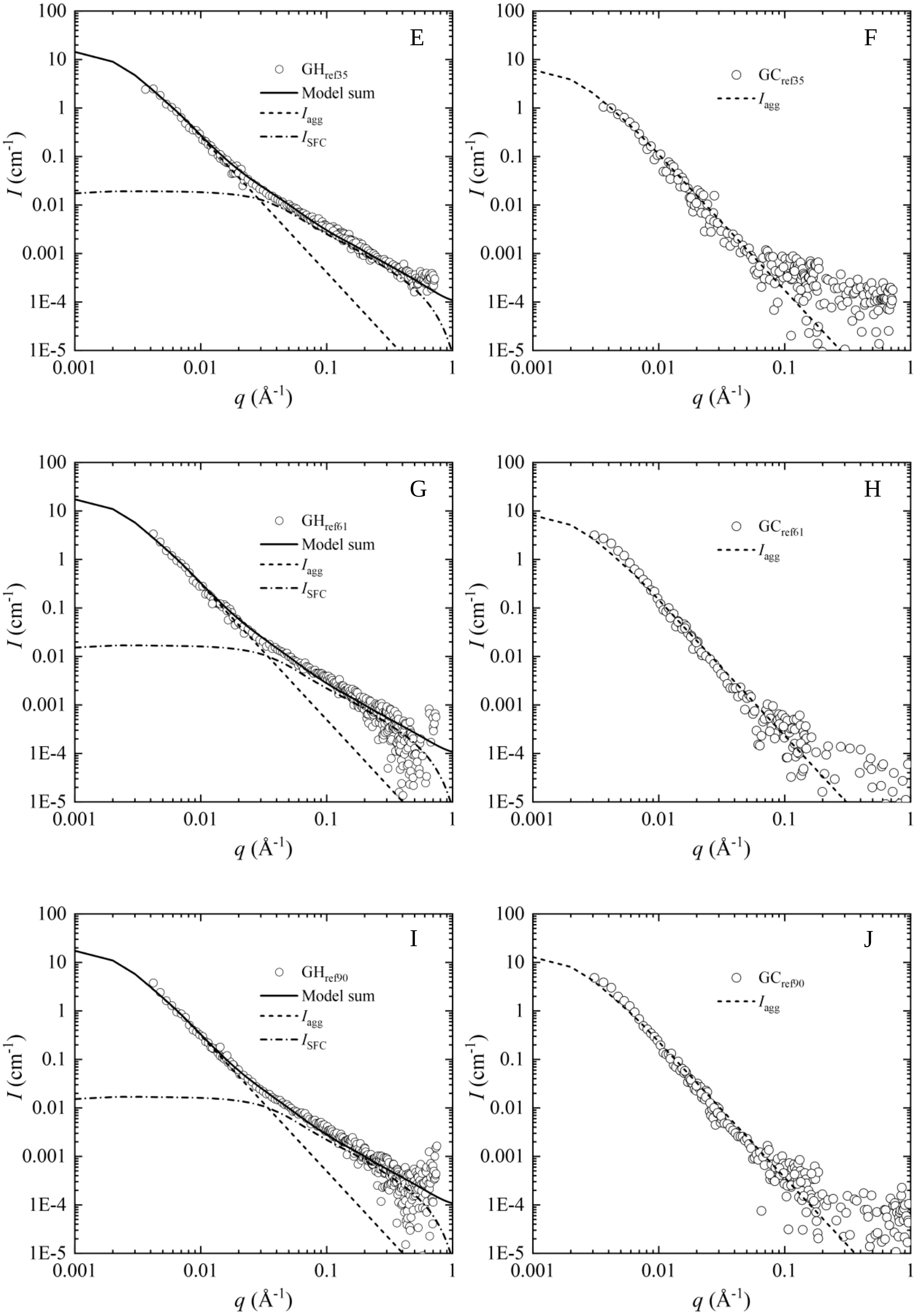
Supplementary Figure S8. Top row: SAXS scattering profiles of cold DOM (A) and hot DOM (B) as a function of concentration. The data were obtained by joining two sample-detector distances. The SAXS profiles were normalized for the concentration. Bottom row: Normalized correlation function from DLS measurements of cold DOM (C) and hot DOM (D) of the same dilution series. The correlation functions were normalized to their maximum. All experiments were performed on freeze-dried DOM that was re-dissolved in order to precisely control the TOC concentrations of the DOM samples.



S2. Colloidal properties of DOM from boreal forest stands of different age.

Supplementary Figure S9. SAXS profiles of DOM: (A) FH_{35}, (B) FC_{35}, (C) FH_{90}, (D) FC_{90}, (E) GH_{ref35}, (F) GC_{ref35}, (G) GH_{ref61}, (H) GC_{ref61}, (I) GH_{ref90}, (J) GC_{ref90}. Solid lines correspond to combined model calculations $I(q) = I_{SFC}(q) + I_{agg}(q)$. Dashed-dotted lines represent model scattering from semi-flexible cylinders, $I_{SFC}(q)$ and dashed lines - scattering from large fractal clusters $I_{agg}(q)$.




E
GH$_{ref35}$
Model sum
I_{agg}
I_{SFC}
F
GC$_{ref35}$
I_{agg}
G
GH$_{ref61}$
Model sum
I_{agg}
I_{SFC}
H
GC$_{ref61}$
I_{agg}
I
GH$_{ref90}$
Model sum
I_{agg}
I_{SFC}
J
GC$_{ref90}$
I_{agg}
I (cm^{-1})
100
10
1
0.1
0.01
0.001
1E-4
1E-5
q (Å^{-1})
0.001
0.01
0.1
1

References

Aarnes, H., Eriksen, A., Petersen, D., and Rise, F. (2007). Accumulation of ammonium in Norway spruce (Picea abies) seedlings measured by in vivo 14N-NMR. J. Exp. Bot. 58**,** 929-934. doi: 10.1093/jxb/erl247.

Artz, R.R.E., Chapman, S.J., Jean Robertson, A.H., Potts, J.M., Laggoun-Défarge, F., Gogo, S., et al. (2008). FTIR spectroscopy can be used as a screening tool for organic matter quality in regenerating cutover peatlands. Soil Biol. Biochem. 40**,** 515-527. doi: 10.1016/j.soilbio.2007.09.019.

Chen, W.R., Butler, P.D., and Magid, L.J. (2006). Incorporating intermicellar interactions in the fitting of SANS data from cationic wormlike micelles. Langmuir 22**,** 6539-6548. doi: 10.1021/la0530440.

Chukanov, N.V., and Chervonnyi, A.D. (2016). Infrared spectroscopy of minerals and related compounds. Cham: Springer International Publishing.

de Gennes, P.-G. (1979). Scaling concepts in polymer physics. Ithaca: Cornell university press.

Ehrnrooth, E.M.L. (1984). Change in pulp fibre density with acid-chlorite delignification. J. Wood Chem. Technol. 4**,** 91-109. doi: 10.1080/02773818408062285.

Gilardi, G., Abis, L., and Cass, A.E.G. (1995). Carbon-13 CP/MAS solid-state NMR and FT-IR spectroscopy of wood cell wall biodegradation. Enzyme Microb. Technol. 17**,** 268-275. doi: 10.1016/0141-0229(94)00019-n.

Gu, B., Schmitt, J., Chen, Z., Liang, L., and McCarthy, J.F. (1994). Adsorption and desorption of natural organic matter on iron oxide: mechanisms and models. Environ Sci Technol 28**,** 38-46. doi: 10.1021/es00050a007.

Gubitosi, M., Duarte, H., Gentile, L., Olsson, U., and Medronho, B. (2016). On cellulose dissolution and aggregation in aqueous tetrabutylammonium hydroxide. Biomacromolecules 17**,** 2873-2881. doi: 10.1021/acs.biomac.6b00696.

Gustavsson, S., Alves, L., Lindman, B., and Topgaard, D. (2014). Polarization transfer solid-state NMR: a new method for studying cellulose dissolution. RSC Adv 4**,** 31836-31839. doi: 10.1039/c4ra04415k.

Hammouda, B. (2010). Analysis of the Beaucage model. J. Appl. Crystallogr. 43**,** 1474-1478. doi: 10.1107/S0021889810033856.

Kaiser, E., Simpson, A.J., Dria, K.J., Sulzberger, B., and Hatcher, P.G. (2003). Solid-state and multidimensional solution-state NMR of solid phase extracted and ultrafiltered riverine dissolved organic matter. Environ Sci Technol 37**,** 2929-2935. doi: 10.1021/es020174b.

Kamida, K., Okajima, K., Matsui, T., and Kowsaka, K. (1984). Study on the solubility of cellulose in aqueous alkali solution by deuteration IR and 13C NMR. Polym. J. 16**,** 857-866. doi: 10.1295/polymj.16.857.

Krosshavn, M., Kögel-Knabner, I., Southon, T.E., and Steinnes, E. (1992). The influence of humus fractionation on the chemical composition of soil organic matter studied by solid-state 13C NMR. J Soil Sci 43**,** 473-483. doi: 10.1111/j.1365-2389.1992.tb00153.x.

Lindner, P., and Zemb, T. (2002). Neutron, X-rays and light. Scattering methods applied to soft condensed matter. Mater. Today 5**,** 38-38. doi: 10.1016/S1369-7021(02)01143-4.

Mao, J., Cao, X., Olk, D.C., Chu, W., and Schmidt-Rohr, K. (2017). Advanced solid-state NMR spectroscopy of natural organic matter. Prog. Nucl. Magn. Reson. Spectrosc. 100**,** 17-51. doi: 10.1016/j.pnmrs.2016.11.003.

Nayak, P.S., and Singh, B.K. (2007). Instrumental characterization of clay by XRF, XRD and FTIR. Bull. Mater. Sci. 30**,** 235-238. doi: 10.1007/s12034-007-0042-5.

Österberg, R., and Mortensen, K. (1992). Fractal dimension of humic acids. Eur. Biophys. J. 21**,** 163-167. doi: 10.1007/BF00196759.

Pedersen, J.S., and Schurtenberger, P. (1996). Scattering functions of semiflexible polymers with and without excluded volume effects. Macromolecules 29**,** 7602-7612. doi: 10.1021/ma9607630.

Preece, N.E., and Cerdan, S. (1993). Determining 15N to 14N ratios in biofluids by single-pulse 1H nuclear magnetic resonance. Anal. Biochem. 215**,** 180-183. doi: 10.1006/abio.1993.1573.

Preston, C.M. (1996). Applications of NMR to soil organic matter analysis: History and prospects. Soil Sci 161**,** 144-166. doi: 10.1097/00010694-199603000-00002.

Rice, J.A., Tombácz, E., and Malekani, K. (1999). Applications of light and X-ray scattering to characterize the fractal properties of soil organic matter. Geoderma 88**,** 251-264. doi: 10.1016/S0016-7061(98)00108-6.

Richards, E.G. (1980). An introduction to physical properties of large molecules in solution. Cambridge: Cambridge University Press.

Schurtenberger, P., and Newman, M.E. (1993). "Characterization of biological and environmental particles using static and dynamic light scattering," in Environmental Particles. Lewis Publishers: Boca Raton, FL), 37-115.

Senesi, N., Rizzi, F.R., Dellino, P., and Acquafredda, P. (1996). Fractal dimension of humic acids in aqueous suspension as a function of pH and time. Soil Sci. Soc. Am. J. 60**,** 1773-1780. doi: 10.2136/sssaj1996.03615995006000060023x.

Silverstein, R.M., Webster, F.X., and Kiemle, D.J. (2005). Spectrometric identification of organic compounds. USA: John Wiley & Sons, Inc.

Simpson, A.J., McNally, D.J., and Simpson, M.J. (2011). NMR spectroscopy in environmental research: from molecular interactions to global processes. Prog. Nucl. Magn. Reson. Spectrosc. 58**,** 97-175. doi: 10.1016/j.pnmrs.2010.09.001.

Sleighter, R.L., Caricasole, P., Richards, K.M., Hanson, T., and Hatcher, P.G. (2015). Characterization of terrestrial dissolved organic matter fractionated by pH and polarity and their biological effects on plant growth. Chem Biol Technol Ag 2**,** 9-9. doi: 10.1186/s40538-015-0036-2.

Stevenson, F.J., and Goh, K.M. (1971). Infrared spectra of humic acids and related substances. Geochim Cosmochim Ac 35**,** 471-483. doi: 10.1016/0016-7037(71)90044-5.

Whitty, S.D., Waggoner, D.C., Cory, R.M., Kaplan, L.A., and Hatcher, P.G. (2019). Direct noninvasive 1H NMR analysis of stream water DOM: Insights into the effects of lyophilization compared with whole water. Magn. Reson. Chem. 1-14. doi: 10.1002/mrc.4935.

Wilson, M.A. (1981). Applications of nuclear magnetic resonance spectroscopy to the study of the structure of soil organic matter. J Soil Sci 32**,** 167-186. doi: 10.1111/j.1365-2389.1981.tb01698.x.