Rurale. Adriatico. Un ragionamento sulle interfacce

In preparation for ***Journal of Hazardous Materials***

Date: *November 12, 2022*

Characteristics of dissolved organic matter in complex shale gas wastewater analyzed with ESI FT-ICR MS

Xuanyu Ji [a,b], *Alberto Tiraferri* [c], *Xiaofei Zhang* [d], *Peng Liu* [e], *Zhiwei Gan* [a], *John C. Crittenden* [f], *Jun Ma* [g], *Baicang Liu* [a,b,*]

[a] State Key Laboratory of Hydraulics and Mountain River Engineering, College of Architecture and Environment, Institute of New Energy and Low-Carbon Technology, Sichuan University, Chengdu, Sichuan 610207, PR China

[b] Yibin Institute of Industrial Technology, Sichuan University Yibin Park, Section 2, Lingang Ave., Cuiping District, Yibin, Sichuan 644000, PR China

[c] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

[*]Corresponding author. Tel.: +86-28-85995998; fax: +86-28-62138325; E-mail: bcliu@scu.edu.cn; baicangliu@gmail.com (B. Liu).

[d] State Key Laboratory of Petroleum Pollution Control, CNPC Research Institute of Safety and Environmental Technology Co., Ltd, Beijing 102206, P R China

[e] Wuxi Research Institute of Petroleum Geology, Petroleum Exploration and Production Research Institute, SINOPEC, Wuxi 214000, PR China

[f] Brook Byers Institute for Sustainable Systems, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

[g] School of Environment, Harbin Institute of Technology, Harbin 150090, PR China

Abstract: Knowledge on the composition and characteristics of dissolved organic matter (DOM) in complex shale gas wastewater (SGW) is critical to evaluate environmental risks and to determine effective management strategies. Herein, five SGW samples from four key shale gas blocks in the Sichuan Basin, China, were comprehensively characterized. Specifically, FT-ICR MS was employed to provide insights into the sources, composition, and characteristics of SGW DOM. Organic matter was characterized by low average molecular weight, high saturation degree, and low aromaticity. Notably, the absence of correlations between molecular-level parameters and spectral indexes might be attributed to the high complexity and variability of SGW. The unique distribution depicted in van Krevelen diagrams suggested various sources of DOM in SGW, such as microbially derived organics in shales and biochemical transformations. Moreover, linear alkyl benzene sulfonates, as well as associated biodegraded metabolites and coproducts, were identified in SGW, implying the distinct anthropogenic imprints and abundant microbial activities. Furthermore, high DOC removal rates (31.42-79.23%) were achieved by biological treatment, fully supporting the inherently labile nature of SGW and the feasibility of biodegradation for SGW management. Therefore, we conclude that DOM in SGW is a complex but mostly labile mixture reflecting both autochthonous and anthropogenic sources.

Keywords: Shale gas wastewater (SGW); Dissolved organic matter (DOM); FT-ICR MS; Molecular lability; Biological treatment

Environmental Implication: Global concerns on the enormous quantity of SGW have raised substantially in recent years due to the adverse environmental risks associated with this hazardous material. However, the need of efficient management of SGW is largely hindered by the poor understanding of its organic composition. This study provides a comprehensive and in-depth investigation on the composition and characteristics of DOM in SGW. Especially, the inherently labile nature of SGW is further supported by the high DOC removal rates during biological treatment. The rich information of this study would facilitate optimizing management strategies and mitigating environmental risks associated with shale gas exploitation.

Graphical Abstract:

1. Introduction

Extraction of shale gas has expanded significantly in recent years and has profoundly affected the global energy landscape (Hou et al., 2012; Vidic et al., 2013; Zhong et al., 2021). Especially, China possesses the largest shale gas reserves (25.08 trillion m^3) in the world and the shale gas industry is booming (Zhong et al., 2021). Nevertheless, concerns about the water management issues associated with shale gas extraction have concurrently grown (Qin et al., 2018; Vengosh et al., 2014). Horizontal drilling and hydraulic fracturing techniques, implemented for the economic exploitation of shale gas, are very water-intensive (Butkovskyi et al., 2017; Chang et al., 2019). Generally, more than 20000 m^3 water is needed for each well. Immediately after hydraulic fracturing, flowback water, at a flow rate up to 1000 m^3/d, returns to the surface usually within two weeks. During gas production, produced water continues to be generated (2-8 m^3/d) over the lifetime of the well. In practice, flowback water and produced water are commonly impounded together in open-air holding ponds, namely shale gas wastewater (SGW) here, for subsequent disposal, treatment, and/or reuse. A variable amount of SGW is produced for each well, ranging from 5200 to 26000 m^3 (Chang et al., 2019), and it is reported that the volume of SGW could reach 499-3585 million m^3 by 2030 (Kondash et al., 2018). SGW is a mixture of injected chemical additives, connate water, in-situ transformation products, and constituents leaching from the shale formation (Coonrod et al., 2020; Ferrer and Thurman, 2015). As a result, the composition of SGW is highly complex and variable. According to previous studies,

the concentrations of total dissolved solids (TDS) and total organic carbon (TOC) are in the ranges of 6906-31090 mg/L and 78-1975 mg/L, respectively, in SGW from the Sichuan Basin, which is the most productive shale gas region in China (Chang et al., 2019; Xie et al., 2022). Moreover, SGW exhibits high levels of radioactivity, with gross α activity of 3.71-83.4 Bq/L and gross β activity of 1.62-18.7 Bq/L (Xie et al., 2022). Therefore, the effective management of SGW is of vital importance to mitigate associated environmental risks and to ensure the sustainable development of the shale gas industry.

Advanced treatment of SGW for beneficial external reuse (such as livestock watering and irrigation) has attracted increasing attention, and has been considered as the best option for SGW management (Coonrod et al., 2020; Robbins et al., 2022). Membrane-based technologies (including reverse osmosis, forward osmosis, electrodialysis, and membrane distillation) and mechanical vapor compression have been studied for the desalination of SGW, according to the level of salinity (Chang et al., 2019; Tong et al., 2019). However, they have not been widely employed on a large scale, as these processes are technically and economically challenging. Advanced oxidation, adsorption, and biological treatment have been evaluated for the removal of organics in SGW (Liu, Y. et al., 2022; Tang et al., 2022). Whereas, remarkable negative influence of the high salinity in SGW on these methods has been reported, and more toxic by-products could be generated during the oxidation processes (Butkovskyi et al., 2017). At present, due to the critical knowledge gaps in the composition of SGW, it is

difficult to unravel the mechanisms of pollutant removal during treatment, and to establish a reasonable regulatory framework for reuse and discharge of the treated effluents. Thus, in order to design effective treatment processes and determine appropriate management strategies, it is critical to reveal the composition of SGW comprehensively.

In comparison with inorganic components whose composition and latent regulation have been extensively disclosed (Barbot et al., 2013; Ni et al., 2018), research on the complex organic constituents in SGW has been relatively limited, especially for SGW from China (Butkovskyi et al., 2017; Zhong et al., 2021). Current understanding has been heavily biased toward volatile and semi-volatile compounds, such as hydrocarbons, studied with gas chromatography paired with mass spectrometry (Luek and Gonsior, 2017), while non-volatile and more polar organic compounds remain largely uncharacterized due to the challenges caused by the intrinsically extreme complicacy and variability of SGW combined with unavailable standard methods of analysis (Oetjen et al., 2017). On the other hand, most studies have merely been devoted to targeting known chemical additives in flowback water and fracturing fluid (Oetjen et al., 2017), which is fairly insufficient as SGW also comprises the so-called produced water, rich with compounds originating from the shale formation. Hence, further investigations involving nontargeted profiling of organic compounds in SGW are needed to better understand the complex composition of SGW.

Nontargeted Fourier transform ion cyclotron resonance mass spectrometry (FT-

ICR MS) is well known as the ultimate technique for probing the composition, source, and evolution of organic matters in various ecosystems based on its unrivaled resolving power and mass accuracy (D'Andrilli et al., 2020; Li et al., 2021; Marshall et al., 1998). Particularly, FT-ICR MS coupled with electrospray ionization (ESI) has been most widely employed to identify polar DOM in aquatic environments and has extended encyclopedic knowledge of DOM much further than any former techniques, at molecular level (Chen et al., 2020; He et al., 2021; Kellerman et al., 2015; Sleighter and Hatcher, 2011). However, to the best of our knowledge, limited publications have applied this technique to characterize DOM in SGW so far (Luek et al., 2018; Luek et al., 2017; Sun et al., 2021). Two articles uncovered the possible subsurface reactions and origins of halogenated organic compounds (Luek et al., 2018; Luek et al., 2017). And Sun *et al*. (2021) concerned more about providing information on the environmental fate of three specific low molecular weight compounds (Sun et al., 2021). Their pioneering research has advanced the understanding of organic components in SGW and demonstrated the potential of FT-ICR MS for characterizing DOM in SGW, whereas, apparently, constituents they investigated were only a small fraction of the DOM pool.

Therefore, in this study five SGW samples collected in the four key shale gas blocks of the Sichuan Basin, China, were investigated. A suite of analysis techniques was combined to provide a comprehensive and in-depth understanding of the SGW DOM. Specifically, ESI FT-ICR MS was employed to reveal the possible sources,

molecular composition, and characteristics of DOM in SGW. Correlations between molecular-level parameters and spectral indexes were evaluated to reveal the complexity of this wastewater. Furthermore, biological treatment experiments were conducted to support the inherently labile nature of SGW as indicated by FT-ICR MS and the feasibility of biodegradation for SGW management. The rich information of this study is critically important to optimize management strategies and to reduce environmental risks associated with shale gas exploitation.

2. Materials and methods

2.1. Sites and sampling.

Five SGW samples analyzed in this study were collected between January 2020 and June 2021 from the four key zones for shale gas exploration in the Sichuan Basin, China: Weiyuan, Changning, Fushun-Yongchuan, and Fuling shale gas blocks (Yu et al., 2016), as shown in **Fig. 1**. In Weiyuan shale gas block, two sampling sites, located in Macaoyan village and Xinlong village, respectively, were used. SGW samples were obtained from each site and labelled as WY1 (at day 148 after hydraulic fracturing) and WY2 (at day 82 after hydraulic fracturing), respectively. In Changning block, sample labeled as CN was obtained at day 53 after hydraulic fracturing from a well in Leijiawan village. Sample FY at day 225 after hydraulic fracturing was obtained from Zhongxin village in Luxian county (Fushun-Yongchuan block), and sample FL at day 123 was from Jiaoshi village in Fuling county (Fuling block). All the hydraulic fracturing-stimulated shale gas wells were horizontally drilled into organic-rich areas of the Lower Silurian Longmaxi Formation and were in production while sampling. Samples were collected from the open-air holding ponds at each shale gas station and were then shipped to the laboratory in hermetically sealed brown high-density polyethylene plastic buckets with limited headspace. They were stored at 4 °C in the dark, and associated experiments and analyses were performed as quickly as possible. Organic characterizations were carried out immediately upon sample arrival to minimize composition variation, followed by other parameters. Moreover, field blanks of

deionized water, processed and analyzed with the same methods of the SGW samples, were used to control possible contamination during sampling and sample preparation.

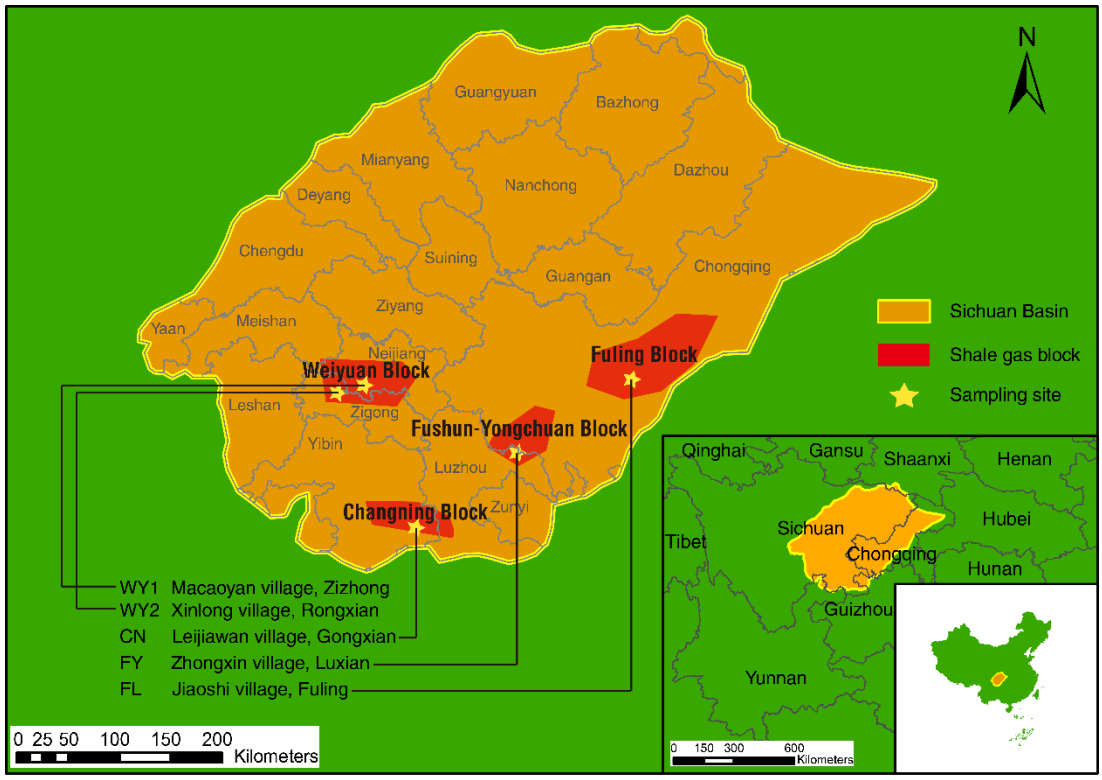


Fig. 1. Map displaying the major shale gas block locations and sampling sites within each block in the Sichuan Basin, China.

2.2. General quality parameters of SGW.

All the samples were analyzed for total dissolved solids (TDS), electrical conductivity (EC), pH, turbidity, dissolved organic carbon (DOC), total dissolved nitrogen (TDN), zeta potential, mean particle size, and inorganic ion concentrations. And we further examined the geochemical characteristics of the SGW samples by absorbance and fluorescence spectroscopy techniques. A wide set of spectral indexes,

including absorbance at 254 nm (UV_{254}), specific UV absorbance (SUVA), spectral slope (S) (Helms et al., 2008), fluorescence index (FI) (Kellerman et al., 2015; McKnight et al., 2001), biological index (BIX) (Parlanti et al., 2000; Wilson and Xenopoulos, 2009), and humification index (HIX) (Ohno, 2002; Zsolnay et al., 1999), were calculated to reflect the sources and structural characteristics of DOM in SGW. The detailed procedures for the above analyses can be found in **Text S1** of the Supporting Information **(SI)**.

2.3. Solid phase extraction and comparative study.

Previous investigations have shown that the extraction efficiency of DOM by solid phase extraction (SPE) is largely affected by SPE sorbents and DOM composition (Dittmar et al., 2008; Li et al., 2017). However, the selectivity of different SPE cartridges for isolating DOM in SGW has not yet been investigated, which might hinder an in-depth understanding of the composition and characteristics of SGW. Therefore, three widespread SPE cartridges (Bond Elut PPL, Oasis HLB, and Sep-Pak C18) were chosen in this study for isolation of SGW DOM following identical extraction procedure to obtain representative extracts. DOC recovery, UV–visible absorbance spectrometry, and FTIR spectroscopy were employed to compare the selectivity of PPL, HLB, and C18. The SPE procedure and analysis methods, as well as the corresponding results, are presented in **Text S2** and **Fig. S1**. Briefly, for the purpose of acquiring representative results, PPL was finally selected due to the consistently highest DOC recoveries and the excellent capacity to concentrate aromatic compounds and molecules

with oxygen-containing functional groups.

2.4. Analytical methods.

The molecular compositions of the PPL extracts of DOM from the five SGW samples were analyzed using a 9.4 T Bruker Apex Ultra FT-ICR mass spectrometer equipped with an Apollo II electrospray ionization (ESI) source. Based on the characterization obtained from FTIR (**Text S2**), suggesting the main occurrence of functional groups that will facilely lose a proton (such as carboxylic acids and alcohols), negative ESI mode was employed for this work (Sleighter and Hatcher, 2011). Ahead of the FT-ICR MS measurement, deuterated stearic acid ($C_{18}D_{35}H_1O_2$, 10^{-6} mol/L) was equally added to the PPL extracts diluted with methanol as an internal standard to obtain semi-quantitative results. The optimized instrumental parameters, procedures for mass calibration, data acquisition and processing can be found in **Text S3**.

The highly abundant class species indicated by FT-ICR MS were identified with UHPLC-QTOF-MS in electrospray negative ion mode. The linear alkyl benzene sulfonates (LAS) standard (C_{10}-C_{13} alkyl homologues, i.e., C_{10}-LAS—C_{13}-LAS) was obtained from Aladdin BioChem Technology Co., Ltd. (Shanghai, China) and three kinds of commercial LAS were supplied by Usolf Chemical, ChuangCheng Washing, and DongRun Chemical.

2.5. Biological treatment.

Biological treatment experiments were carried out in 300 mL glass bottles with 200 mL of liquid and 100 mL of headspace using sequencing batch reactors (SBR).

Five reactors were conducted, with one for each SGW sample. They were inoculated with activated sludge obtained from a municipal wastewater treatment plant (Chengdu, China) and aerated via aeration disks at the bottom to maintain the dissolved oxygen level at 4–6 mg/L. The reactors were operated in two-day cycles and 150 mL was replaced by new feed water at the end of each cycle. During the acclimation period, the feed water (a mixture of the municipal wastewater and SGW) TDS concentration was increased by ~3000 mg/L (by adding SGW) for every stage until it consisted of 100% SGW in each reactor. After acclimation, the reactors were operated for two additional cycles, with the influent and effluent samples taken for DOC analysis. In addition, control experiments were conducted without inoculation. The detailed steps of the biological treatment experiments are provided in **Table S1**.

3. Results and discussion

3.1. Geochemical characteristics of SGW.

The general quality parameters are summarized in **Table S2** to inform the bulk physical and chemical characteristics of SGW. It is obvious that the concentrations of the main constituents were highly variable spatially. TDS of the five SGW samples varied by a factor of 3, from 10880 to 33650 mg/L, and DOC by a factor of ~5, from 12.86 to 59.75 mg/L, representing the wide spectrum of characteristics observed for SGW globally (Chang et al., 2019; Tang et al., 2020a; Zhong et al., 2021). Especially, FI values, commonly used to reflect the sources of DOM, were all greater than 2.30, suggesting the predominant microbial sources of organic material in SGW (Kellerman et al., 2015; McKnight et al., 2001). BIX values that served as a tool to estimate biological activity were greater than 1.04, indicating the strong autogenetic characteristics of SGW DOM (Parlanti et al., 2000; Wilson and Xenopoulos, 2009). Moreover, HIX was developed to evaluate the humification extent of DOM, and the extremely low values (0.53-1.97) implied that SGW DOM was mainly derived from metabolic processes of microorganisms and was of low humification extent (Ohno, 2002; Zsolnay et al., 1999). Actually, the above analyses were mainly based on the differences in the values of the spectral indexes (FI, BIX, and HIX) between SGW and other water sources (such as freshwater and seawater), and the limitations of these indexes would be discussed in detail in **Section 3.2**.

3.2. Molecular characteristics of SGW.

The broadband mass spectra of the SGW samples showed several thousand peaks between m/z 200 and 700, and peaks were generally located below m/z 450 (**Fig. S2**). As opposed to the symmetric distribution for the mass spectra of DOM from typical water sources, such as Suwannee River Fulvic Acid (SRFA, a widely used DOM reference) (D'Andrilli et al., 2013), the SGW DOM distribution was quite irregular and dispersive. Nevertheless, the regular mass spacing patterns common to other types of water sources were also observed for SGW, including typical 14.0156 Da for CH_2 groups, 0.9953 Da (NH vs. CH_2), and 0.0364 Da (CH_4 vs. O) (Stenson et al., 2003). The CH_4 vs. O pattern is highlighted by showing the expanded spectra at m/z 275 (**Fig. S2**), and all assigned formulas at this nominal mass, sorted into different molecular series according to the substitution of CH_4 for O, are presented in **Table S3**. Most of the common formulas for the five SGW samples at this nominal mass were CHO formulas, whereas differences among the samples largely stemmed from compounds with more heteroatoms. And these trends were present along with the spectra.

The total number of assigned formulas and molecular-level parameters of the five SGW samples analyzed in this work are summarized in **Table S4**, together with those of samples from various aquatic ecosystems (including rivers, lakes, oceans, etc.) obtained from the literature (He et al., 2019; Li et al., 2017; Wang et al., 2020; Xu et al., 2020). Overall, SGW samples had apparently higher H/C ratios (1.36-1.56), and significantly lower average molecular weight (295.55-319.66), O/C ratios (0.253-

0.402), double bond equivalent (4.56-6.38) (D'Andrilli et al., 2013), and aromaticity index (0.027-0.197) (Koch and Dittmar, 2006) compared to other more traditional water samples (including freshwater, seawater, river water, and lake water). Therefore, DOM in SGW seemed to contain more aliphatic compounds with high levels of saturation and low oxidation degree. In addition, the lower N/C and S/C ratios might be associated with the narrow distribution range of the two heteroatoms. The remarkable differences observed between SGW with respect to other types of water sources, underline the importance of understanding the organic composition of SGW for its effective and safe management.

To investigate the similarity among the five SGW samples, principal component analysis (PCA) was performed based on the molecular-level parameters and spectral indexes, as shown in **Fig. S3**. The samples WY1 and WY2 from the same shale gas block were indeed similar, while FL was dissimilar to all the other samples. Furthermore, the relationships between molecular-level parameters and spectral indexes were determined by Spearman rank correlation and the color correlogram is presented in **Fig. S4** to provide a more systematic understanding of the nature of SGW DOM. In general, molecular-level parameters correlated well with each other. DBE_w, $(DBE/C)_w$, and AI_w were positively correlated with each other, while $(H/C)_w$ was negatively correlated with them, indicating the excellent power of the four parameters to predict the aromaticity and unsaturation degree of DOM. That is, DOM with lower DBE_w, $(DBE/C)_w$, and/or AI_w, and higher $(H/C)_w$ has higher degree of saturation, and

vice versa. $NOSC_w$ presented high level significant positive correlation with $(O/C)_w$. Stubbins *et al*. (2014) have suggested that fluorescent DOM (FDOM) would have the potential to track the bulk pool of DOM in freshwater (Stubbins et al., 2014). Many previous studies have also devoted to interpreting the broader DOM pool by utilizing chromophoric DOM (CDOM) and FDOM as prisms, and they have revealed the significant correlations between spectral indexes as well as between spectral indexes and molecular-level parameters: the S parameter is negatively correlated with $(m/z)_w$, and HIX is negatively correlated with BIX and positively correlated with $(O/C)_w$ (Singer et al., 2012; Wang et al., 2021; Zhang, B. et al., 2019). However, these relationships were not observed in this study, which would be attributed to the complexity of SGW. That is, the percentage and chemistry of CDOM and FDOM in SGW might be highly variable, also highlighting the limitations regarding the interpretation of DOM optical signatures in SGW. It should be noted that the molecular-level parameters are subjected to the selectivity of SPE and ESI in negative mode (D'Andrilli et al., 2020; Li et al., 2017), also probably leading to discrepancies from expected correlations. Besides, the limited number of samples should be cautioned here, and collecting more SGW samples for correlation analysis would be important in future efforts.

3.3. Molecular composition and distribution.

All SGW samples mainly contained CHO, CHNO, CHOS, and CHNOS categories, and their semi-quantitative relative abundance decreased consistently in the order CHO

(> 50%) > CHOS > CHNO > CHNOS (< 6%), except for sample CN that contained 53.75% CHOS compounds, followed by CHO, CHNO, and in which no CHNOS compounds were detected (**Fig. S5**). To further investigate the elemental composition, a total of 43, 56, 37, 44 and 28 classes were assigned for WY1, WY2, CN, FY, and FL, respectively, and the semi-quantitative relative abundance of each class for each sample is shown in **Figs. 2** and **S6-S9** in detail. The molecular composition varied significantly among the samples, and 23 classes in total including $O_{2\text{-}10}$, $N_1O_{2\text{-}8}$, $O_{3\text{-}9}S_1$ were assigned in all samples. The most abundant class species in the CHO category were $O_{2\text{-}6}$ for FL and $O_{4\text{-}8}$ for the other four samples, with much lower oxygen atom numbers than that of other traditional water types, which agreed well with the low $(O/C)_w$ values found for SGW in this study. N_nO_o always consisted of the largest number of classes, except for sample FL that had no molecular formulas containing more than one N heteroatom. Only O_oS_1 and $N_1O_oS_1$ class species were determined in CHOS and CHNOS, respectively, for all the samples.

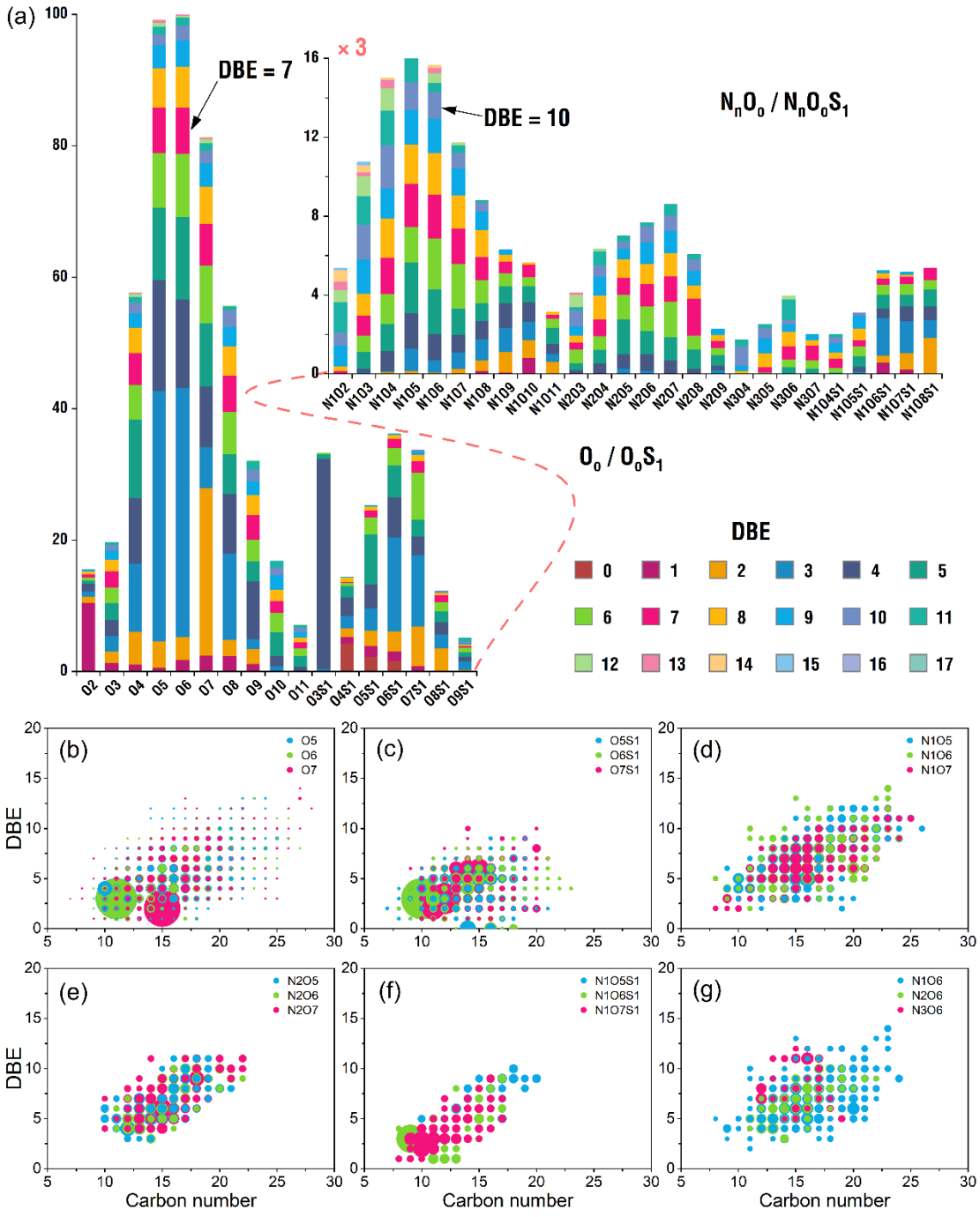


Fig. 2. Molecular composition and distribution of sample WY1. (a) Relative abundance of the assigned classes. The horizontal axis represents heteroatom classes and is split into 2 segments. The vertical scale for the N_nO_o and $N_nO_oS_s$ segment is enlarged 3 times. Columns with different colors correspond to compound with different DBE. Diagrams of DBE versus carbon numbers for O_5-O_7 (b), O_5S_1-O_7S_1 (c), N_1O_5-

N_1O_7 (d), N_2O_5-N_2O_7 (e), $N_1O_5S_1$-$N_1O_7S_1$ (f), and N_1O_6-N_3O_6 (g) classes. Bubble size represents the relative abundance of subclasses in each diagram of DBE versus carbon numbers.

The DBE_w distribution for each category is illustrated more clearly in **Table S5**. CHNO compounds had the highest DBE_w (5.973-7.825), consistently for all samples. Meanwhile, CHOS always had the lowest DBE_w, except for sample FY associated with a slightly lower DBE_w for the CHO category (4.352). Also, the consistently higher DBE_w of CHNO and CHNOS suggested that molecules with no N heteroatom were much more saturated. This phenomenon might result from the more complex core structures of nitrogen-containing compounds. Particularly, the fully saturated compounds with DBE=0 were mainly composed of CHOS and there were no CHNO compounds, see **Figs. 2** and **S6-S9**. Furthermore, according to the DBE vs. carbon number data, classes with higher relative abundance extended to wider DBE and carbon number ranges, and all classes showed the common tendency of increasing DBE as the number of carbon atoms increased. The main O_5-O_7 classes species in sample WY1 had comparatively low DBE values of 2-4 (**Fig. 2b**), indicating that hydroxyl, carbonyl, and/or carboxyl might be contained in these compounds. The O_6S_1 class in WY1 had DBE values of 0-9 and carbon numbers of 8-23 (**Fig. 2c**). The most abundant O_6S_1 species exhibited DBE=3 and the peak with corresponding neutral formula of $C_{10}H_{16}O_6S_1$ had the highest relative abundance. In general, the relative abundance of

N-containing formulas was less variable (**Fig. 2d-g**), without obviously abundant subclasses (such as O_6 species with 3 DBEs and 11 carbon numbers in **Fig. 2b**). More importantly, it could be found that DBE increased with higher number of N atoms (**Fig. 2g**), suggesting more aromatic structures, and less obvious trends were also observed for most classes with increasing the number of O atoms (**Fig. 2b-f**).

Evidently, O_3S_1 class species showed an unusually high relative abundance in O_oS_1 and were characterized by the complete dominance of compounds with 4 DBEs, especially for sample CN (**Fig. S7**). For detailed examination, primary formulas common to all the samples, accounting for more than 95% of the subclass (O_3S_1 with 4 DBEs) by abundance, are listed in **Table S6**. And compounds with these assigned formulas were further confirmed as linear alkyl benzene sulfonates (LAS), the most widely used anionic surfactant, with alkyl chains in the range of 10 to 13 carbon atoms in length (C_{10}-LAS—C_{13}-LAS), based on the perfect matching of accurate mass, MS/MS spectrum, and retention time with LAS standard analyzed by UHPLC-QTOF-MS. Moreover, major compounds with formulas in subclasses of O_5S_1 with 5 DBEs and O_3S_1 with 5 DBEs were identified via MS/MS analysis as sulfophenyl carboxylic acids (SPC) and dialkyl tetralin sulfonates (DATS), respectively (**Table S6**). Note that quite similar fragmentation behaviors of the same sulfonates have been reported previously (Gonsior et al., 2011; Lara-Martín et al., 2010). Detailed descriptions and structures of the above compounds are given in **Fig. S10** and exemplary MS/MS spectra of some selected analytes are shown in **Fig. S11**.

LAS, SPC, and DATS constituted a considerable proportion of the dissolved organic sulfur (DOS) pool. It should be noted that these compounds were absent from the blanks. More importantly, the identification of LAS as fracturing additives was verified with the engineers responsible for the wells in the present study, despite the fact that the composition of hydraulic fracturing fluids has not been fully disclosed in China. Generally, the concentrations of fracturing additives in SGW decline quickly as the flowback time increases (Sun et al., 2019). Thus, it makes sense that LAS exhibited an extremely high relative abundance in sample CN (**Fig. S7**), as CN was collected in the relatively early stage of the well (at day 53 after hydraulic fracturing). Nonionic surfactants (such as polyethylene glycols and polypropylene glycols identified in electrospray positive ion mode) have been primarily reported surfactant additives that act as multifunctional chemicals (corrosion inhibitors, friction reducers or even biocides) in hydraulic fracturing in North America (Sun et al., 2019; Thurman et al., 2017). On the other hand, alcohol sulfate and alcohol ethoxysulfate were considered to be the dominant DOS classes in SGW samples from Morgantown, WV, USA (Luek et al., 2019). The results fully demonstrate the significant differences of composition and characteristics of SGW between China and North America (Zhong et al., 2021), putatively stemming from formation-specific geogenic constituents and various fracturing fluid chemistry designed as a function of well depth and geology, as well as operator discretion (Coonrod et al., 2020; Ferrer and Thurman, 2015; Stringfellow et al., 2014). Additionally, LAS would represent sensitive organic tracers of injected

fracturing fluid in such complex matrixes due to their very high ionization efficiencies in negative ESI (Luek et al., 2019).

The three kinds of commercial LAS shared similar distribution of the homologous series as measured and extracted ion chromatograms (EIC) of C_{10}-LAS—C_{13}-LAS in sample CN and in a representative commercial LAS (Usolf Chemical) are presented in **Fig. S12**. In accordance with previous studies (Di Corcia, A. et al., 1999; Di Corcia, Antonio et al., 1999), absorption to particles/soil and biodegradation processes are probably responsible for the shift of LAS distribution to homologues with shorter alkyl chain length in sample CN (**Fig. S12**). And the ubiquitous SPC, known as the most common biodegradation metabolites of LAS, indicate the abundant microbial activities in SGW during exploitation processes, simultaneously (Cluff et al., 2014; Zhou et al., 2022). However, it is unknown whether the biodegradation of LAS occurred in the deep subsurface (anaerobic), in open-air holding ponds after returning to the surface (aerobic), or both, as SPC could be generated under both aerobic and anaerobic conditions with distinct degradation pathways (Di Corcia, Antonio et al., 1999; Lara-Martín et al., 2010). It is worth noting that, although the unique reaction metabolites (4-methyl sulfophenyl dicarboxylic acids, Me-SPdC) of LAS under anaerobic conditions (Lara-Martín et al., 2010) were not identified in the five SGW samples, this observation alone was not sufficient to indicate the absence of anaerobic degradation, as we collected samples subjected to aerobic conditions for a long time in open-air holding ponds, which would also lead to the degradation of Me-SPdC (Lara-Martín et al., 2010).

Collecting and analyzing a time series of samples that newly return to the surface would help to draw convincing conclusions. In fact, it is highly undesirable when the biotransformation takes place underground as attenuation of injected additives potentially affect the well production efficiency (Evans et al., 2019; Luek et al., 2019) and stimulated microbial processes could lead to gas souring and infrastructure corrosion (Sirivedhin and Dallbauman, 2004). Therefore, biodegradation sites and associations between hydraulic fracturing additives and microbial growth need to be systematically investigated for reasonably designing fracturing fluid formulation and guaranteeing effective hydraulic fracturing engineering. Moreover, recent studies have reported the endocrine disrupting potential of LAS, SPC, and DATS (Geng et al., 2018), it is also vital to reveal the environmental fate of these compounds upon unintended release.

3.4. Van Krevelen diagram analysis.

The use of van Krevelen diagram is a prominent approach to visualizing such huge datasets and interpreting the respective chemical properties in DOM, based on the characteristic H/C and O/C ratios possessed by each major compound groups (Kim et al., 2003). In this work, detected molecules tended to be divided into seven compound groups with detailed classification boundaries discussed in **Text S3** (Hertkorn et al., 2006; Kim et al., 2003; Xu et al., 2020). Overall, the DOM in SGW distributed centrally and corresponded reasonably well with their $(H/C)_w$ and $(O/C)_w$ (**Fig. 3**). Clearly, more than 96% of the compounds with elemental ratios (semi-quantitative relative abundance)

located in lipids, aliphatic/proteins, and lignins/carboxylic rich alicyclic molecules (CRAM)-like structures, for all the samples (**Table S7**). The lignins/CRAM-like structures group was the most abundant in samples WY1 (42.74%), CN (46.19%), FY (51.67%) and FL (68.88%), while molecular signals in aliphatic/proteins region exhibited the highest percentage in WY2 (46.67%). Moreover, the excessively low proportion of unsaturated hydrocarbons, carbohydrates, tannins, and condensed aromatic structures is tabulated in **Table S7**.

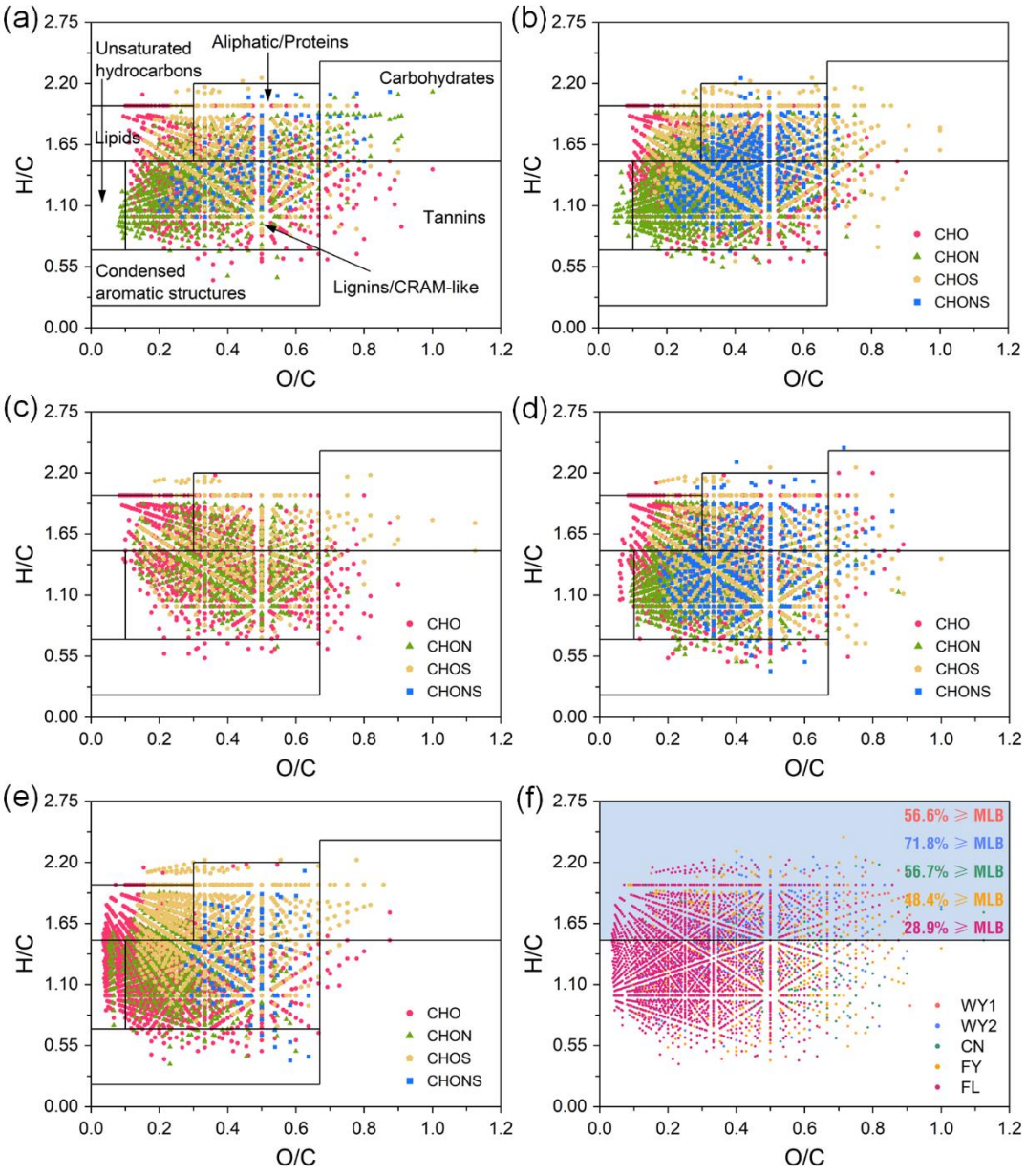


Fig. 3. Van Krevelen diagrams of sample WY1 (a), WY2 (b), CN (c), FY (d), and FL (e), constructed for each category (CHO, CHNO, CHOS, and CHNOS). The black lines depicting the boundaries corresponding to primary groups of DOM constituents are expounded in the VK diagram for WY1 (**Fig. 3a**). (f) Van Krevelen diagram for the five samples, with the proportion of intensity-weighted labile component indicated. The black line depicts the location of MLB, and the shaded area corresponds to labile region.

460 The chemical composition and nature of DOM in SGW would be associated with
461 its unique sources (He et al., 2019). Especially, a considerable proportion of lipids- and
462 aliphatic/proteins-like species, commonly used to reflect microbially influenced DOM
463 (D'Andrilli et al., 2013), were assigned in SGW (WY1: 53.61%, WY2: 69.27%, CN:
464 52.36%, FY: 46.2%, FL: 27.9%). Combined with the high FI and BIX (**Section 3.1**),
465 we suggested that DOM in SGW could partly originate from the organic material
466 trapped within shales in the deep subsurface (Krumholz et al., 1997), a microbially
467 dominated environment with abundant detritus and secretions of microorganisms .
468 Although numerous research efforts have also found microbially derived protein signals
469 in SGW (Riley et al., 2018; Tang et al., 2020b; Wang et al., 2019), it is important to
470 note that the N-containing formulas in aliphatic/proteins region were not merely related
471 to the proteins, given that N-containing formulas were concentrated mainly in
472 lignins/CRAM-like region which implied various sources responsible for them.
473 In order to extend FT-ICR MS analysis toward further prediction of
474 biogeochemical properties, D'Andrilli *et al*. (2015) have proposed the employment of
475 the molecular lability boundary (MLB) at H/C=1.5 to determine the extent of DOM
476 lability (D'Andrilli et al., 2015). Labile constituents in DOM refer to bioavailable
477 carbon utilized in heterotrophic activity (Battin et al., 2008). Compounds above MLB
478 at H/C≥1.5, including lipids, aliphatic/proteins, and carbohydrates in the VK diagram,
479 exhibit a labile character (shaded region in **Fig. 3f**), whereas constituents below the
480 MLB, H/C<1.5, correspond to generally refractory material. The normalized

intensities (MLB_L) and number (MLB_{nL}) of assigned labile molecules, as well as the relative proportion of them (MLB_L(%) and MLB_{nL}(%)) are summarized in **Table S8**, with their detailed explanations presented in **Text S3**. Interestingly, both MLB_L(%) (28.89-71.83%) and MLB_{nL}(%) (33.51-43.84%) indicated the much more labile nature of SGW compounds, compared to most other ecosystems including the glacial environments which are thought to contain a mass of labile component derived microbially (D'Andrilli et al., 2015; D'Andrilli et al., 2013). The MLB_L(%) and MLB_{nL}(%) were usually higher in CHO and CHOS compounds than in CHNO and CHNOS (**Table S8**). Furthermore, the distribution of different categories (CHO, CHNO, CHOS, and CHNOS) and groups (lipids, aliphatic/proteins, and carbohydrates) in labile region (H/C≥1.5) is illustrated more clearly in **Table S9**. As expected, CHO and CHOS compounds accounted for a considerable proportion (> 87% by peak intensity) in labile constituents. In addition, the majority (> 85% by peak intensity) of labile molecules were located in lipids and aliphatic/proteins regions, with much less contributions from molecules in carbohydrates.

The lignins/CRAM-like structures in SGW shared common features with different types of DOM from a variety of ecosystems. And this region was made up of various chemical species, including lignins, CRAM, and poly phenolic compounds. Lignin species are complicated and variable substances sourcing from higher plants and widely distributed in waters and soils (Stenson et al., 2003). The freshwater used for preparing fracturing fluid and leaching from solids during water-soil interactions might be

potential sources for this fraction of DOM in SGW. However, considering the low molecular signals in the tannins region that instead exist in substantial amounts in terrigenous environments, these sources were likely not the principal contributors. And the low HIX could support this hypothesis. Thus, we suggest that CRAM and poly phenolic compounds might be the predominant constituents in this region, consistent with the DOM composition in glacial and marine environments that are completely devoid of higher plants (Hertkorn et al., 2006; Stubbins et al., 2010). These compounds are associated with the end products of biodegradation and also exhibit recalcitrant character similar to that of humics and lignins (Hertkorn et al., 2006). In view of the complex composition and indiscernible sources of lignins/CRAM-like structures region, we focused more on the refractory nature of chemical species with elemental ratios in this region.

Undoubtedly, anthropogenic inputs (fracturing additives such as biocides, breakers, and corrosion inhibitors) also had profound influences on DOM composition, distribution, and evolution. As mentioned above, LAS, SPC, and DATS contributed largely to the corresponding class or even to the whole molecular composition. **Fig. S13** through different visualizations explicitly elucidate their appearance and patterns. What is worth being emphasized is the ambiguity in classifying thousands of organic compounds into several groups using VK diagram boundaries merely based on the two simplistic ratios (i.e., H/C and O/C), especially for water sources containing significant anthropogenic fraction, such as SGW. That is, LAS, SPC, and DATS are neither lipids

nor lignins/CRAM-like structures, but they fall in corresponding boundary regions of the VK diagram (**Fig. S13a**), respectively. Admittedly, there are various boundaries used to interpret VK diagrams from different literatures (Hertkorn et al., 2006; Kim et al., 2003; Wang et al., 2012; Xu et al., 2020), which would lead to distinct quantitative results, while similar conclusions can still be obtained here grounded in the similarities and differences between SGW and other important water types. Similarly, the ambiguity of the MLB, which is based solely on a H/C threshold value, should also be noted.

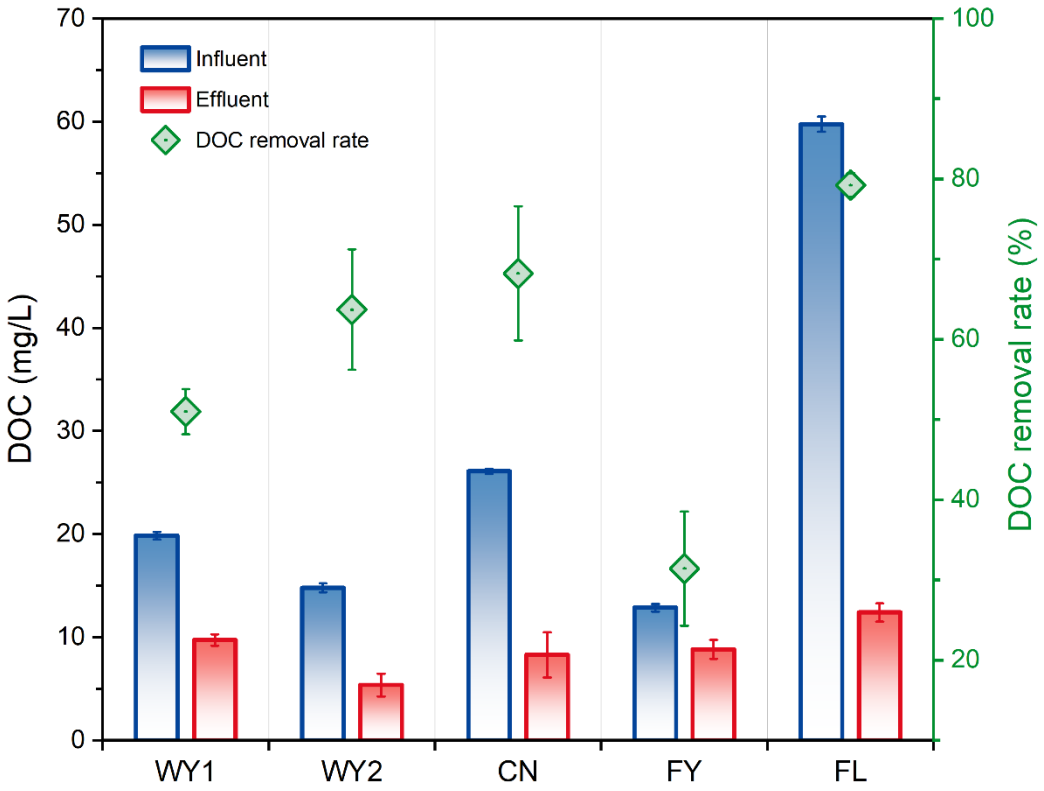
A total of 6176 molecular formulas were identified in the five SGW samples. These formulas were divided into three different groups (tCP(5): present in all of the 5 samples; pCP(2-4): present in at least 2 samples and at most 4 samples; DP(1): present in only one of the 5 samples). **Table S10** shows the distribution of formulas in the three groups. The proportion of unique molecular formulas (DP(1)) in samples WY2 and FL reached up to more than 23% (by peak number), which was much higher than that in WY1 (5.36%), CN (4.43%), and FY (10.28%). Van Krevelen diagrams for unique molecular formulas in each sample are presented in **Fig. S14a-e**. A significant proportion of the unique N-containing molecules (80.12% by peak number) was observed in WY2 (**Fig. S14b**). The unique CHO molecular formulas of sample FL were mainly located in O/C < 0.3 and 0.7 < H/C < 1.7 regions, and the unique CHOS formulas were mainly in 0.1 < O/C < 0.3 and 1.1 < H/C < 2.0 regions (**Fig. S14e**). It should be noticed that the proportion of different categories in DP(1) group calculated

by peak intensity showed the same trend as the proportion calculated by peak number for all the five samples. Moreover, 49.44%, 49.55%, 38.72%, 55.44%, and 37.93% of the molecular formulas in samples WY1, WY2, CN, FY, and FL, respectively, were classified as pCP(2-4) (**Fig. S14f**).

The 1207 common molecular formulas (tCP(5)) accounted for a considerable proportion in each sample, and the proportion calculated by intensity was much higher than that calculated by number (**Table S10**), indicating that the common formulas usually exhibited high intensities. **Fig. S15** shows that CHO (59.73%) molecules were the primary components in tCP(5). Furthermore, the inter-sample rankings analyses (Chen et al., 2021; Herzsprung et al., 2012; Zhang, L. et al., 2019) were used to evaluate the differences in DOM quality (**Fig. S16**). Sample FY exhibited ranks 1-2 (relatively high peak intensities) for lower molecular weight compounds (< 325 Da) (**Fig. S16h**), which might be caused by microbial degradation of high molecular weight organic matter into smaller molecules during the long-term storage (for 225 days) in the open-air holding ponds. Given that sample CN was collected in the relatively early stage of the well and the storage time of CN is the shortest (for 53 days), it is expected that more high molecular weight compounds could exist in CN. It is consistent with the fact that CN showed ranks 1-2 for components with higher molecular weight (400-500 Da) (**Fig. S16f**). In addition, the molecular formulas in the ranks 1-2 of sample FL were mainly located in O/C < 0.3 region (**Fig. S16i**).

3.5. Biodegradation for SGW treatment.

Camarillo *et al*. (2016) previously showed that 37% of 155 fracturing additives, which were variable in different wells and solely represented a small fraction of the DOM pool, could be effectively removed by biological treatment based on the available physical-chemical data (Camarillo et al., 2016). In the present study, numerous labile constituents of DOM in real SGW obtained from FT-ICR MS (**Section 3.4**) suggest that biodegradation would be suitable for SGW treatment and reuse. Thus, we conducted preliminary biological treatment experiments (SBR) to support this hypothesis. **Fig. 4** shows that the efficacies of DOC removal reached 50.98%, 63.69%, 68.23%, 31.42%, and 79.23% for samples WY1, WY2, CN, FY, and FL, respectively. Similar removal rates (RR) were observed in other researches treating SGW: 45.3-83.2% by biologically active filtration (Tang et al., 2022), 59.5-87.9% by membrane bioreactor (Liu, X. et al., 2022), and 52-85% by SBR (Sitterley et al., 2021).



577

578 **Fig. 4.** Removal of DOC from the five SGW samples by activated sludge SBR.

579 Note that no considerable decrease in DOC concentration was obtained in the non-

580 inoculated control experiments.

581

582 Furthermore, the relationships between DOC removal rate (RR) and some key

583 parameters that are considered likely to affect the performance of biological treatment

584 were assessed by Spearman rank correlation coefficients (**Fig. S17**). RR was positively

585 correlated with MLB_L (the sum of normalized intensities of assigned labile molecules)

586 and DOC, but not with MLB_{nL}, MLB_L(%), and MLB_{nL}(%), suggesting that higher

587 absolute content of labile components would facilitate microbial metabolism and lead

588 to more organic removal. Actually, MLB_L(%) and MLB_{nL}(%) have been generally used

589 to assess the extent of DOM lability in published researches (D'Andrilli et al., 2015; He
590 et al., 2019), while MLB_L showed greater potential to indicate the biological treatability
591 of SGW in this study.

592 Notably, since there was no attempt to optimize the acclimation process in the
593 biological treatment experiments, the DOC removal rates obtained might only represent
594 relatively low levels. Therefore, our results (31.42-79.23% of DOC removal) would
595 support the feasibility of biodegradation for SGW treatment. Given the inherently labile
596 nature of SGW, we highlight that robustly constructing a salt-tolerant microbial
597 community, disclosing key degradation pathways, optimizing reactor designs, as well
598 as combining effective pretreatment techniques (e.g., advanced oxidation, adsorption)
599 should represent future research efforts.
600

4. Conclusions

The present study provided a detailed investigation on the composition and characteristics of DOM in complex SGW from four key shale gas blocks in the Sichuan Basin, China using a suite of analysis techniques (FT-ICR MS, UHPLC-QTOF-MS, UV–Vis spectrophotometry, and fluorescence spectrophotometry). In general, SGW DOM was characterized by low average molecular weight, high saturation degree, and low aromaticity. However, the correlations between molecular-level parameters and spectral indexes observed in samples from more traditional aquatic ecosystems were absent from SGW, implying the high complexity and variability of SGW. Moreover, the identification of LAS, as well as associated biodegraded metabolites (SPC) and coproducts (DATS) suggested the distinct anthropogenic imprints and abundant microbial activities during exploitation processes. These findings also underline the need for further revealing the environmental fate of hydraulic fracturing additives and the associations between these compounds and microbial growth, with the objective of guaranteeing effective hydraulic fracturing engineering and reducing relevant environmental risks. Based on the van Krevelen diagram analysis, we found that various sources (microbially derived organics in shales and biochemical transformations) might be responsible for SGW DOM, while the limitations and ambiguity of classification boundaries in VK diagram should also be noted. Furthermore, 31.42-79.23% of DOC were effectively removed by SBR, fully supporting the inherently labile nature of SGW as indicated by FT-ICR MS and

providing the theoretical basis for full-scale biological treatment of SGW in long-term operation.

CRediT authorship contribution statement

Xuanyu Ji: Methodology, Investigation, Visualization, Writing - original draft. **Alberto Tiraferri**: Validation, Writing - review & editing. **Xiaofei Zhang**: Resources, Data curation. **Peng Liu**: Resources, Data curation, Software, Formal analysis. **Zhiwei Gan**: Resources, Investigation. **John C. Crittenden**: Methodology, Data curation, Writing - review & editing. **Jun Ma**: Methodology, Writing - review & editing. **Baicang Liu**: Conceptualization, Supervision, Writing - review & editing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (52070134, 52270075), and Sichuan University and Yibin City People's Government Strategic Cooperation Project (2020CDYB-2).

Supporting Information

Additional experimental materials, methods, and procedures for data processing; discussion on SPE and comparative analysis; general water quality of the five SGW samples; detailed distribution of organic compounds; and a number of other supporting materials are provided in Texts S1−S3, Tables S1−S10, and Figs. S1−S17.

REFERENCES

Barbot, E., Vidic, N.S., Gregory, K.B., Vidic, R.D., 2013. Spatial and Temporal Correlation of Water Quality Parameters of Produced Waters from Devonian-Age Shale following Hydraulic Fracturing. Environ. Sci. Technol. 47(6), 2562-2569. https://doi.org/10.1021/es304638h.

Battin, T.J., Kaplan, L.A., Findlay, S., Hopkinson, C.S., Marti, E., Packman, A.I., Newbold, J.D., Sabater, F., 2008. Biophysical controls on organic carbon fluxes in fluvial networks. Nat. Geosci. 1(2), 95-100. https://doi.org/10.1038/ngeo101.

Butkovskyi, A., Bruning, H., Kools, S.A.E., Rijnaarts, H.H.M., Van Wezel, A.P., 2017. Organic Pollutants in Shale Gas Flowback and Produced Waters: Identification, Potential Ecological Impact, and Implications for Treatment Strategies. Environ. Sci. Technol. 51(9), 4740-4754. https://doi.org/10.1021/acs.est.6b05640.

Camarillo, M.K., Domen, J.K., Stringfellow, W.T., 2016. Physical-chemical evaluation of hydraulic fracturing chemicals in the context of produced water treatment. J. Environ. Manage. 183, 164-174. https://doi.org/https://doi.org/10.1016/j.jenvman.2016.08.065.

Chang, H., Li, T., Liu, B., Vidic, R.D., Elimelech, M., Crittenden, J.C., 2019. Potential and implemented membrane-based technologies for the treatment and reuse of flowback and produced water from shale gas and oil plays: A review. Desalination 455, 34-57. https://doi.org/10.1016/j.desal.2019.01.001.

Chen, Q., Chen, F., Gonsior, M., Li, Y., Wang, Y., He, C., Cai, R., Xu, J., Wang, Y., Xu, D., Sun, J., Zhang, T., Shi, Q., Jiao, N., Zheng, Q., 2021. Correspondence between DOM molecules and microbial community in a subtropical coastal estuary on a spatiotemporal scale. ENVIRON INT 154, 106558. https://doi.org/https://doi.org/10.1016/j.envint.2021.106558.

Chen, W., Zhuo, X., He, C., Shi, Q., Li, Q., 2020. Molecular investigation into the transformation of dissolved organic matter in mature landfill leachate during treatment in a combined membrane bioreactor-reverse osmosis process. J. Hazard. Mater. 397, 122759. https://doi.org/https://doi.org/10.1016/j.jhazmat.2020.122759.

Cluff, M.A., Hartsock, A., MacRae, J.D., Carter, K., Mouser, P.J., 2014. Temporal Changes in Microbial Ecology and Geochemistry in Produced Water from Hydraulically Fractured Marcellus Shale Gas Wells. Environ. Sci. Technol. 48(11), 6508-6517. https://doi.org/10.1021/es501173p.

Coonrod, C.L., Yin, Y.B., Hanna, T., Atkinson, A.J., Alvarez, P.J.J., Tekavec, T.N., Reynolds, M.A., Wong, M.S., 2020. Fit-for-purpose treatment goals for produced waters in shale oil and gas fields. Water Res. 173. https://doi.org/10.1016/j.watres.2020.115467.

D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G., 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Spectrom. 29(24), 2385-2401.

https://doi.org/10.1002/rcm.7400.
D'Andrilli, J., Fischer, S.J., Rosario-Ortiz, F.L., 2020. Advancing Critical Applications of High Resolution Mass Spectrometry for DOM Assessments: Re-Engaging with Mass Spectral Principles, Limitations, and Data Analysis. Environ. Sci. Technol. 54(19), 11654-11656. https://doi.org/10.1021/acs.est.0c04557.
D'Andrilli, J., Foreman, C.M., Marshall, A.G., McKnight, D.M., 2013. Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. Org. Geochem. 65, 19-28. https://doi.org/10.1016/j.orggeochem.2013.09.013.
Di Corcia, A., Capuani, L., Casassa, F., Marcomini, A., Samperi, R., 1999. Fate of linear alkyl benzenesulfonates, coproducts, and their metabolites in sewage treatment plants and in receiving river waters. Environ. Sci. Technol. 33(22), 4119-4125. https://doi.org/10.1021/es990596u.
Di Corcia, A., Casassa, F., Crescenzi, C., Marcomini, A., Samperi, R., 1999. Investigation of the Fate of Linear Alkyl Benzenesulfonates and Coproducts in a Laboratory Biodegradation Test by Using Liquid Chromatography/Mass Spectrometry. Environ. Sci. Technol. 33(22), 4112-4118. https://doi.org/10.1021/es9905952.
Dittmar, T., Koch, B., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol Oceanogr-Meth 6, 230-235. https://doi.org/10.4319/lom.2008.6.230.
Evans, M.V., Getzinger, G., Luek, J.L., Hanson, A.J., McLaughlin, M.C., Blotevogel, J., Welch, S.A., Nicora, C.D., Purvine, S.O., Xu, C.D., Cole, D.R., Darrah, T.H., Hoyt, D.W., Metz, T.O., Ferguson, P.L., Lipton, M.S., Wilkins, M.J., Mouser, P.J., 2019. In situ transformation of ethoxylate and glycol surfactants by shale-colonizing microorganisms during hydraulic fracturing. ISME J 13(11), 2690-2700. https://doi.org/10.1038/s41396-019-0466-0.
Ferrer, I., Thurman, E.M., 2015. Chemical constituents and analytical approaches for hydraulic fracturing waters. Trends Environ. Anal. Chem. 5, 18-25. https://doi.org/https://doi.org/10.1016/j.teac.2015.01.003.
Geng, C.X., Cao, N., Xu, W., He, C., Yuan, Z.W., Liu, J.W., Shi, Q., Xu, C.M., Liu, S.T., Zhao, H.Z., 2018. Molecular Characterization of Organics Removed by a Covalently Bound Inorganic-Organic Hybrid Coagulant for Advanced Treatment of Municipal Sewage. Environ. Sci. Technol. 52(21), 12642-12648. https://doi.org/10.1021/acs.est.8b03306.
Gonsior, M., Zwartjes, M., Cooper, W.J., Song, W., Ishida, K.P., Tseng, L.Y., Jeung, M.K., Rosso, D., Hertkorn, N., Schmitt-Kopplin, P., 2011. Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. Water Res. 45(9), 2943-2953. https://doi.org/https://doi.org/10.1016/j.watres.2011.03.016.

He, D., He, C., Li, P.H., Zhang, X.W., Shi, Q., Sun, Y.G., 2019. Optical and Molecular Signatures of Dissolved Organic Matter Reflect Anthropogenic Influence in a Coastal River, Northeast China. J. Environ. Qual. 48(3), 603-613. https://doi.org/10.2134/jeq2018.09.0330.

He, P.J., Liu, W.Y., Qiu, J.J., Zhang, H., Huang, Y.L., Deng, Y.T., Shao, L.M., Lu, F., 2021. Improvement criteria for different advanced technologies towards bio-stabilized leachate based on molecular subcategories of DOM. J. Hazard. Mater. 414. https://doi.org/10.1016/j.jhazmat.2021.125463.

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr. 53(3), 955-969. https://doi.org/10.4319/lo.2008.53.3.0955.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., Hedges, J.I., 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochim. Cosmochim. Acta 70(12), 2990-3010. https://doi.org/10.1016/j.gca.2006.03.021.

Herzsprung, P., von Tümpling, W., Hertkorn, N., Harir, M., Büttner, O., Bravidor, J., Friese, K., Schmitt-Kopplin, P., 2012. Variations of DOM Quality in Inflows of a Drinking Water Reservoir: Linking of van Krevelen Diagrams with EEMF Spectra by Rank Correlation. Environ. Sci. Technol. 46(10), 5511-5518. https://doi.org/10.1021/es300345c.

Hou, D.Y., Luo, J., Al-Tabbaa, A., 2012. COMMENTARY: Shale gas can be a double-edged sword for climate change. Nat Clim Chang 2(6), 385-387. https://doi.org/10.1038/nclimate1500.

Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nat. Geosci. 8(6), 454-U452. https://doi.org/10.1038/ngeo2440.

Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. Anal. Chem. 75(20), 5336-5344. https://doi.org/10.1021/ac034415p.

Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 20(5), 926-932. https://doi.org/10.1002/rcm.2386.

Kondash, A.J., Lauer, N.E., Vengosh, A., 2018. The intensification of the water footprint of hydraulic fracturing. Sci. Adv. 4(8). https://doi.org/10.1126/sciadv.aar5982.

Krumholz, L.R., McKinley, J.P., Ulrich, F.A., Suflita, J.M., 1997. Confined subsurface microbial communities in Cretaceous rock. Nature 386(6620), 64-66. https://doi.org/10.1038/386064a0.

Lara-Martín, P.A., Gómez-Parra, A., Sanz, J.L., González-Mazo, E., 2010. Anaerobic Degradation Pathway of Linear Alkylbenzene Sulfonates (LAS) in Sulfate-Reducing Marine Sediments. Environ. Sci. Technol. 44(5), 1670-1676.

https://doi.org/10.1021/es9032887.
Li, C., Zheng, M., Cao, D., Yang, L., Wu, J., Yang, Y., Liu, G., 2021. Recognition of the molecular characterization and mechanisms of heterogeneously formed organic pollutants from metallurgical industries by FT-ICR-MS and GC/Q-TOF-MS. J. Hazard. Mater. 406, 124603. https://doi.org/10.1016/j.jhazmat.2020.124603.
Li, Y., Harir, M., Uhl, J., Kanawati, B., Lucio, M., Smirnov, K.S., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2017. How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. Water Res. 116, 316-323. https://doi.org/10.1016/j.watres.2017.03.038.
Liu, X., Tang, P., Liu, Y., Xie, W., Chen, C., Li, T., He, Q., Bao, J., Tiraferri, A., Liu, B., 2022. Efficient removal of organic compounds from shale gas wastewater by coupled ozonation and moving-bed-biofilm submerged membrane bioreactor. Bioresour. Technol. 344, 126191. https://doi.org/https://doi.org/10.1016/j.biortech.2021.126191.
Liu, Y., Wu, Q., Chen, C., Li, T., Liu, S., He, Q., Yang, P., Bai, Y., Liu, B., 2022. An efficient system of aerogel adsorbent combined with membranes for reuse of shale gas wastewater. Desalination 526, 115545. https://doi.org/https://doi.org/10.1016/j.desal.2021.115545.
Luek, J.L., Gonsior, M., 2017. Organic compounds in hydraulic fracturing fluids and wastewaters: A review. Water Res. 123, 536-548. https://doi.org/https://doi.org/10.1016/j.watres.2017.07.012.
Luek, J.L., Harir, M., Schmitt-Kopplin, P., Mouser, P.J., Gonsior, M., 2018. Temporal dynamics of halogenated organic compounds in Marcellus Shale flowback. Water Res. 136, 200-206. https://doi.org/10.1016/j.watres.2018.02.055.
Luek, J.L., Harir, M., Schmitt-Kopplin, P., Mouser, P.J., Gonsior, M., 2019. Organic sulfur fingerprint indicates continued injection fluid signature 10 months after hydraulic fracturing. ENVIRON SCI-PROC IMP 21(2), 206-213. https://doi.org/10.1039/c8em00331a.
Luek, J.L., Schmitt-Kopplin, P., Mouser, P.J., Petty, W.T., Richardson, S.D., Gonsior, M., 2017. Halogenated Organic Compounds Identified in Hydraulic Fracturing Wastewaters Using Ultrahigh Resolution Mass Spectrometry. Environ. Sci. Technol. 51(10), 5377-5385. https://doi.org/10.1021/acs.est.6b06213.
Marshall, A.G., Hendrickson, C.L., Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom. Rev. 17(1), 1-35. https://doi.org/10.1002/(sici)1098-2787(1998)17:1<1::Aid-mas1>3.0.Co;2-k.
McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnol. Oceanogr. 46(1), 38-48. https://doi.org/10.4319/lo.2001.46.1.0038.
Ni, Y., Zou, C., Cui, H., Li, J., Lauer, N.E., Harkness, J.S., Kondash, A.J., Coyte, R.M., Dwyer, G.S., Liu, D., Dong, D., Liao, F., Vengosh, A., 2018. Origin of Flowback

and Produced Waters from Sichuan Basin, China. Environ. Sci. Technol. 52(24), 14519-14527. https://doi.org/10.1021/acs.est.8b04345.

Oetjen, K., Giddings, C.G.S., McLaughlin, M., Nell, M., Blotevogel, J., Helbling, D.E., Mueller, D., Higgins, C.P., 2017. Emerging analytical methods for the characterization and quantification of organic contaminants in flowback and produced water. Trends Environ. Anal. Chem. 15, 12-23. https://doi.org/https://doi.org/10.1016/j.teac.2017.07.002.

Ohno, T., 2002. Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environ. Sci. Technol. 36(4), 742-746. https://doi.org/10.1021/es0155276.

Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. Org. Geochem. 31(12), 1765-1781. https://doi.org/https://doi.org/10.1016/S0146-6380(00)00124-8.

Qin, Y., Höglund-Isaksson, L., Byers, E., Feng, K., Wagner, F., Peng, W., Mauzerall, D.L., 2018. Air quality–carbon–water synergies and trade-offs in China's natural gas industry. Nat. Sustain. 1(9), 505-511. https://doi.org/10.1038/s41893-018-0136-7.

Riley, S.M., Ahoor, D.C., Regnery, J., Cath, T.Y., 2018. Tracking oil and gas wastewater-derived organic matter in a hybrid biofilter membrane treatment system: A multi-analytical approach. Sci. Total Environ. 613-614, 208-217. https://doi.org/10.1016/j.scitotenv.2017.09.031.

Robbins, C.A., Du, X., Bradley, T.H., Quinn, J.C., Bandhauer, T.M., Conrad, S.A., Carlson, K.H., Tong, T., 2022. Beyond treatment technology: Understanding motivations and barriers for wastewater treatment and reuse in unconventional energy production. Resour Conserv Recycl 177, 106011. https://doi.org/https://doi.org/10.1016/j.resconrec.2021.106011.

Singer, G.A., Fasching, C., Wilhelm, L., Niggemann, J., Steier, P., Dittmar, T., Battin, T.J., 2012. Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate. Nat. Geosci. 5(10), 710-714. https://doi.org/10.1038/ngeo1581.

Sirivedhin, T., Dallbauman, L., 2004. Organic matrix in produced water from the Osage-Skiatook Petroleum Environmental Research Site, Osage County, Oklahoma. Chemosphere 57(6), 463-469. https://doi.org/10.1016/j.chemosphere.2004.05.034.

Sitterley, K.A., Silverstein, J., Rosenblum, J., Linden, K.G., 2021. Aerobic biological degradation of organic matter and fracturing fluid additives in high salinity hydraulic fracturing wastewaters. Sci. Total Environ. 758, 143622. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.143622.

Sleighter, R.L., Hatcher, P.G., 2011. Fourier Transform Mass Spectrometry for the Molecular Level Characterization of Natural Organic Matter: Instrument Capabilities, Applications, and Limitations. In Fourier Transforms-Approach to Scientific Principles.

Stenson, A.C., Marshall, A.G., Cooper, W.T., 2003. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. Anal. Chem. 75(6), 1275-1284. https://doi.org/10.1021/ac026106p.

Stringfellow, W.T., Domen, J.K., Camarillo, M.K., Sandelin, W.L., Borglin, S., 2014. Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing. J. Hazard. Mater. 275, 37-54. https://doi.org/https://doi.org/10.1016/j.jhazmat.2014.04.040.

Stubbins, A., Lapierre, J.F., Berggren, M., Prairie, Y.T., Dittmar, T., del Giorgio, P.A., 2014. What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada. Environ. Sci. Technol. 48(18), 10598-10606. https://doi.org/10.1021/es502086e.

Stubbins, A., Spencer, R.G.M., Chen, H.M., Hatcher, P.G., Mopper, K., Hernes, P.J., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., Six, J., 2010. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. Limnol. Oceanogr. 55(4), 1467-1477. https://doi.org/10.4319/lo.2010.55.4.1467.

Sun, C., Zhang, Y., Alessi, D.S., Martin, J.W., 2019. Nontarget profiling of organic compounds in a temporal series of hydraulic fracturing flowback and produced waters. ENVIRON INT 131, 104944. https://doi.org/https://doi.org/10.1016/j.envint.2019.104944.

Sun, Y., Wu, M., Tong, T., Liu, P., Tang, P., Gan, Z., Yang, P., He, Q., Liu, B., 2021. Organic compounds in Weiyuan shale gas produced water: Identification, detection and rejection by ultrafiltration-reverse osmosis processes. Chem. Eng. J. 412, 128699. https://doi.org/https://doi.org/10.1016/j.cej.2021.128699.

Tang, P., Li, J., Li, T., Tian, L., Sun, Y., Xie, W., He, Q., Chang, H., Tiraferri, A., Liu, B., 2020a. Efficient integrated module of gravity driven membrane filtration, solar aeration and GAC adsorption for pretreatment of shale gas wastewater. J. Hazard. Mater., 124166. https://doi.org/https://doi.org/10.1016/j.jhazmat.2020.124166.

Tang, P., Liu, B., Zhang, Y., Chang, H., Zhou, P., Feng, M., Sharma, V.K., 2020b. Sustainable reuse of shale gas wastewater by pre-ozonation with ultrafiltration-reverse osmosis. Chem. Eng. J. 392, 123743. https://doi.org/https://doi.org/10.1016/j.cej.2019.123743.

Tang, P., Xie, W., Tian, L., Tan, B., Zhang, Y., Yang, Z., Chen, C., Zhang, W., Liu, B., 2022. Oxidation-biotreatment-membrane combined process for external reuse of shale gas wastewater. Sep. Purif. Technol. 291, 120920. https://doi.org/https://doi.org/10.1016/j.seppur.2022.120920.

Thurman, E.M., Ferrer, I., Rosenblum, J., Linden, K., Ryan, J.N., 2017. Identification of polypropylene glycols and polyethylene glycol carboxylates in flowback and produced water from hydraulic fracturing. J. Hazard. Mater. 323(Pt A), 11-17. https://doi.org/10.1016/j.jhazmat.2016.02.041.

Tong, T.Z., Carlson, K.H., Robbins, C.A., Zhang, Z.Y., Du, X.W., 2019. Membrane-

based treatment of shale oil and gas wastewater: The current state of knowledge. Front Environ Sci Eng 13(4). https://doi.org/10.1007/s11783-019-1147-y.
Vengosh, A., Jackson, R.B., Warner, N., Darrah, T.H., Kondash, A., 2014. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environ. Sci. Technol. 48(15), 8334-8348. https://doi.org/10.1021/es405118y.
Vidic, R.D., Brantley, S.L., Vandenbossche, J.M., Yoxtheimer, D., Abad, J.D., 2013. Impact of Shale Gas Development on Regional Water Quality. Science 340(6134). https://doi.org/10.1126/science.1235009.
Wang, D.D., Song, C.T., Zhang, B.L., Chen, J.W., Luo, A.L., Wang, X.S., Wu, S.D., Ye, Y.X., 2021. Deciphering dissolved organic matter from freshwater aquaculture ponds in Eastern China based on optical and molecular signatures. PROCESS SAF ENVIRON 155. https://doi.org/10.1016/j.psep.2021.09.025.
Wang, H., Lu, L., Chen, X., Bian, Y., Ren, Z.J., 2019. Geochemical and microbial characterizations of flowback and produced water in three shale oil and gas plays in the central and western United States. Water Res. 164, 114942. https://doi.org/https://doi.org/10.1016/j.watres.2019.114942.
Wang, X., Goual, L., Colberg, P.J.S., 2012. Characterization and treatment of dissolved organic matter from oilfield produced waters. J. Hazard. Mater. 217-218, 164-170. https://doi.org/https://doi.org/10.1016/j.jhazmat.2012.03.006.
Wang, X., Ji, Y., Shi, Q., Zhang, Y., He, C., Wang, Q., Guo, S., Chen, C., 2020. Characterization of wastewater effluent organic matter with different solid phase extraction sorbents. Chemosphere 257, 127235. https://doi.org/10.1016/j.chemosphere.2020.127235.
Wilson, H.F., Xenopoulos, M.A., 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nat. Geosci. 2(1), 37-41. https://doi.org/10.1038/ngeo391.
Xie, W., Tian, L., Tang, P., Cui, J., Wang, T., Zhu, Y., Bai, Y., Tiraferri, A., Crittenden, J.C., Liu, B., 2022. Shale gas wastewater characterization: Comprehensive detection, evaluation of valuable metals, and environmental risks of heavy metals and radionuclides. Water Res. 220, 118703. https://doi.org/https://doi.org/10.1016/j.watres.2022.118703.
Xu, W., Gao, Q., He, C., Shi, Q., Hou, Z.-Q., Zhao, H.-Z., 2020. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. Environ. Sci. Technol. 54(20), 12929-12937. https://doi.org/10.1021/acs.est.0c01681.
Yu, M., Weinthal, E., Patino-Echeverri, D., Deshusses, M.A., Zou, C., Ni, Y., Vengosh, A., 2016. Water Availability for Shale Gas Development in Sichuan Basin, China. Environ. Sci. Technol. 50(6), 2837-2845. https://doi.org/10.1021/acs.est.5b04669.
Zhang, B., Shan, C., Hao, Z., Liu, J., Wu, B., Pan, B., 2019. Transformation of dissolved organic matter during full-scale treatment of integrated chemical wastewater: Molecular composition correlated with spectral indexes and acute toxicity. Water

Res. 157, 472-482. https://doi.org/https://doi.org/10.1016/j.watres.2019.04.002.
Zhang, L., Peng, Y., Ge, Z., Xu, K., 2019. Fate of dissolved organic nitrogen during the Anammox process using ultra-high resolution mass spectrometry. ENVIRON INT 131, 105042. https://doi.org/https://doi.org/10.1016/j.envint.2019.105042.
Zhong, C., Zolfaghari, A., Hou, D., Goss, G.G., Lanoil, B.D., Gehman, J., Tsang, D.C.W., He, Y., Alessi, D.S., 2021. Comparison of the Hydraulic Fracturing Water Cycle in China and North America: A Critical Review. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c06119.
Zhou, S., Peng, S., Li, Z., Zhang, D., Zhu, Y., Li, X., Hong, M., Li, W., Lu, P., 2022. Characterization of microbial communities and functions in shale gas wastewaters and sludge: Implications for pretreatment. J. Hazard. Mater. 424, 127649. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.127649.
Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. Chemosphere 38(1), 45-50. https://doi.org/https://doi.org/10.1016/S0045-6535(98)00166-0.

In preparation for ***Journal of Hazardous Materials***

Date: *November 12, 2022*

Characteristics of dissolved organic matter in complex shale gas wastewater analyzed with ESI FT-ICR MS

Xuanyu Ji [a,b], Alberto Tiraferri [c], Xiaofei Zhang [d], Peng Liu [e], Zhiwei Gan [a], John C. Crittenden [f], Jun Ma [g], Baicang Liu [a,b,]*

[a] State Key Laboratory of Hydraulics and Mountain River Engineering, College of Architecture and Environment, Institute of New Energy and Low-Carbon Technology, Sichuan University, Chengdu, Sichuan 610207, PR China

[b] Yibin Institute of Industrial Technology, Sichuan University Yibin Park, Section 2, Lingang Ave., Cuiping District, Yibin, Sichuan 644000, PR China

[c] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

[*]Corresponding author. Tel.: +86-28-85995998; fax: +86-28-62138325; E-mail: bcliu@scu.edu.cn; baicangliu@gmail.com (B. Liu).

[d] State Key Laboratory of Petroleum Pollution Control, CNPC Research Institute of Safety and Environmental Technology Co., Ltd, Beijing 102206, P R China

[e] Wuxi Research Institute of Petroleum Geology, Petroleum Exploration and Production Research Institute, SINOPEC, Wuxi 214000, PR China

[f] Brook Byers Institute for Sustainable Systems, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

[g] School of Environment, Harbin Institute of Technology, Harbin 150090, PR China

Abstract: Knowledge on the composition and characteristics of dissolved organic matter (DOM) in complex shale gas wastewater (SGW) is critical to evaluate environmental risks and to determine effective management strategies. Herein, five SGW samples from four key shale gas blocks in the Sichuan Basin, China, were comprehensively characterized. Specifically, FT-ICR MS was employed to provide insights into the sources, composition, and characteristics of SGW DOM. Organic matter was characterized by low average molecular weight, high saturation degree, and low aromaticity. Notably, the absence of correlations between molecular-level parameters and spectral indexes might be attributed to the high complexity and variability of SGW. The unique distribution depicted in van Krevelen diagrams suggested various sources of DOM in SGW, such as microbially derived organics in shales and biochemical transformations. Moreover, linear alkyl benzene sulfonates, as well as associated biodegraded metabolites and coproducts, were identified in SGW, implying the distinct anthropogenic imprints and abundant microbial activities. Furthermore, high DOC removal rates (31.42-79.23%) were achieved by biological treatment, fully supporting the inherently labile nature of SGW and the feasibility of biodegradation for SGW management. Therefore, we conclude that DOM in SGW is a complex but mostly labile mixture reflecting both autochthonous and anthropogenic sources.

Keywords: Shale gas wastewater (SGW); Dissolved organic matter (DOM); FT-ICR MS; Molecular lability; Biological treatment

Environmental Implication: Global concerns on the enormous quantity of SGW have raised substantially in recent years due to the adverse environmental risks associated with this hazardous material. However, the need of efficient management of SGW is largely hindered by the poor understanding of its organic composition. This study provides a comprehensive and in-depth investigation on the composition and characteristics of DOM in SGW. Especially, the inherently labile nature of SGW is further supported by the high DOC removal rates during biological treatment. The rich information of this study would facilitate optimizing management strategies and mitigating environmental risks associated with shale gas exploitation.

Graphical Abstract:

1. Introduction

Extraction of shale gas has expanded significantly in recent years and has profoundly affected the global energy landscape (Hou et al., 2012; Vidic et al., 2013; Zhong et al., 2021). Especially, China possesses the largest shale gas reserves (25.08 trillion m^3) in the world and the shale gas industry is booming (Zhong et al., 2021). Nevertheless, concerns about the water management issues associated with shale gas extraction have concurrently grown (Qin et al., 2018; Vengosh et al., 2014). Horizontal drilling and hydraulic fracturing techniques, implemented for the economic exploitation of shale gas, are very water-intensive (Butkovskyi et al., 2017; Chang et al., 2019). Generally, more than 20000 m^3 water is needed for each well. Immediately after hydraulic fracturing, flowback water, at a flow rate up to 1000 m^3/d, returns to the surface usually within two weeks. During gas production, produced water continues to be generated (2-8 m^3/d) over the lifetime of the well. In practice, flowback water and produced water are commonly impounded together in open-air holding ponds, namely shale gas wastewater (SGW) here, for subsequent disposal, treatment, and/or reuse. A variable amount of SGW is produced for each well, ranging from 5200 to 26000 m^3 (Chang et al., 2019), and it is reported that the volume of SGW could reach 499-3585 million m^3 by 2030 (Kondash et al., 2018). SGW is a mixture of injected chemical additives, connate water, in-situ transformation products, and constituents leaching from the shale formation (Coonrod et al., 2020; Ferrer and Thurman, 2015). As a result, the composition of SGW is highly complex and variable. According to previous studies,

the concentrations of total dissolved solids (TDS) and total organic carbon (TOC) are in the ranges of 6906-31090 mg/L and 78-1975 mg/L, respectively, in SGW from the Sichuan Basin, which is the most productive shale gas region in China (Chang et al., 2019; Xie et al., 2022). Moreover, SGW exhibits high levels of radioactivity, with gross α activity of 3.71-83.4 Bq/L and gross β activity of 1.62-18.7 Bq/L (Xie et al., 2022). Therefore, the effective management of SGW is of vital importance to mitigate associated environmental risks and to ensure the sustainable development of the shale gas industry.

Advanced treatment of SGW for beneficial external reuse (such as livestock watering and irrigation) has attracted increasing attention, and has been considered as the best option for SGW management (Coonrod et al., 2020; Robbins et al., 2022). Membrane-based technologies (including reverse osmosis, forward osmosis, electrodialysis, and membrane distillation) and mechanical vapor compression have been studied for the desalination of SGW, according to the level of salinity (Chang et al., 2019; Tong et al., 2019). However, they have not been widely employed on a large scale, as these processes are technically and economically challenging. Advanced oxidation, adsorption, and biological treatment have been evaluated for the removal of organics in SGW (Liu, Y. et al., 2022; Tang et al., 2022). Whereas, remarkable negative influence of the high salinity in SGW on these methods has been reported, and more toxic by-products could be generated during the oxidation processes (Butkovskyi et al., 2017). At present, due to the critical knowledge gaps in the composition of SGW, it is

difficult to unravel the mechanisms of pollutant removal during treatment, and to establish a reasonable regulatory framework for reuse and discharge of the treated effluents. Thus, in order to design effective treatment processes and determine appropriate management strategies, it is critical to reveal the composition of SGW comprehensively.

In comparison with inorganic components whose composition and latent regulation have been extensively disclosed (Barbot et al., 2013; Ni et al., 2018), research on the complex organic constituents in SGW has been relatively limited, especially for SGW from China (Butkovskyi et al., 2017; Zhong et al., 2021). Current understanding has been heavily biased toward volatile and semi-volatile compounds, such as hydrocarbons, studied with gas chromatography paired with mass spectrometry (Luek and Gonsior, 2017), while non-volatile and more polar organic compounds remain largely uncharacterized due to the challenges caused by the intrinsically extreme complicacy and variability of SGW combined with unavailable standard methods of analysis (Oetjen et al., 2017). On the other hand, most studies have merely been devoted to targeting known chemical additives in flowback water and fracturing fluid (Oetjen et al., 2017), which is fairly insufficient as SGW also comprises the so-called produced water, rich with compounds originating from the shale formation. Hence, further investigations involving nontargeted profiling of organic compounds in SGW are needed to better understand the complex composition of SGW.

Nontargeted Fourier transform ion cyclotron resonance mass spectrometry (FT-

ICR MS) is well known as the ultimate technique for probing the composition, source, and evolution of organic matters in various ecosystems based on its unrivaled resolving power and mass accuracy (D'Andrilli et al., 2020; Li et al., 2021; Marshall et al., 1998). Particularly, FT-ICR MS coupled with electrospray ionization (ESI) has been most widely employed to identify polar DOM in aquatic environments and has extended encyclopedic knowledge of DOM much further than any former techniques, at molecular level (Chen et al., 2020; He et al., 2021; Kellerman et al., 2015; Sleighter and Hatcher, 2011). However, to the best of our knowledge, limited publications have applied this technique to characterize DOM in SGW so far (Luek et al., 2018; Luek et al., 2017; Sun et al., 2021). Two articles uncovered the possible subsurface reactions and origins of halogenated organic compounds (Luek et al., 2018; Luek et al., 2017). And Sun *et al*. (2021) concerned more about providing information on the environmental fate of three specific low molecular weight compounds (Sun et al., 2021). Their pioneering research has advanced the understanding of organic components in SGW and demonstrated the potential of FT-ICR MS for characterizing DOM in SGW, whereas, apparently, constituents they investigated were only a small fraction of the DOM pool.

Therefore, in this study five SGW samples collected in the four key shale gas blocks of the Sichuan Basin, China, were investigated. A suite of analysis techniques was combined to provide a comprehensive and in-depth understanding of the SGW DOM. Specifically, ESI FT-ICR MS was employed to reveal the possible sources,

molecular composition, and characteristics of DOM in SGW. Correlations between molecular-level parameters and spectral indexes were evaluated to reveal the complexity of this wastewater. Furthermore, biological treatment experiments were conducted to support the inherently labile nature of SGW as indicated by FT-ICR MS and the feasibility of biodegradation for SGW management. The rich information of this study is critically important to optimize management strategies and to reduce environmental risks associated with shale gas exploitation.

2. Materials and methods

2.1. Sites and sampling.

Five SGW samples analyzed in this study were collected between January 2020 and June 2021 from the four key zones for shale gas exploration in the Sichuan Basin, China: Weiyuan, Changning, Fushun-Yongchuan, and Fuling shale gas blocks (Yu et al., 2016), as shown in **Fig. 1**. In Weiyuan shale gas block, two sampling sites, located in Macaoyan village and Xinlong village, respectively, were used. SGW samples were obtained from each site and labelled as WY1 (at day 148 after hydraulic fracturing) and WY2 (at day 82 after hydraulic fracturing), respectively. In Changning block, sample labeled as CN was obtained at day 53 after hydraulic fracturing from a well in Leijiawan village. Sample FY at day 225 after hydraulic fracturing was obtained from Zhongxin village in Luxian county (Fushun-Yongchuan block), and sample FL at day 123 was from Jiaoshi village in Fuling county (Fuling block). All the hydraulic fracturing-stimulated shale gas wells were horizontally drilled into organic-rich areas of the Lower Silurian Longmaxi Formation and were in production while sampling. Samples were collected from the open-air holding ponds at each shale gas station and were then shipped to the laboratory in hermetically sealed brown high-density polyethylene plastic buckets with limited headspace. They were stored at 4 °C in the dark, and associated experiments and analyses were performed as quickly as possible. Organic characterizations were carried out immediately upon sample arrival to minimize composition variation, followed by other parameters. Moreover, field blanks of

deionized water, processed and analyzed with the same methods of the SGW samples, were used to control possible contamination during sampling and sample preparation.

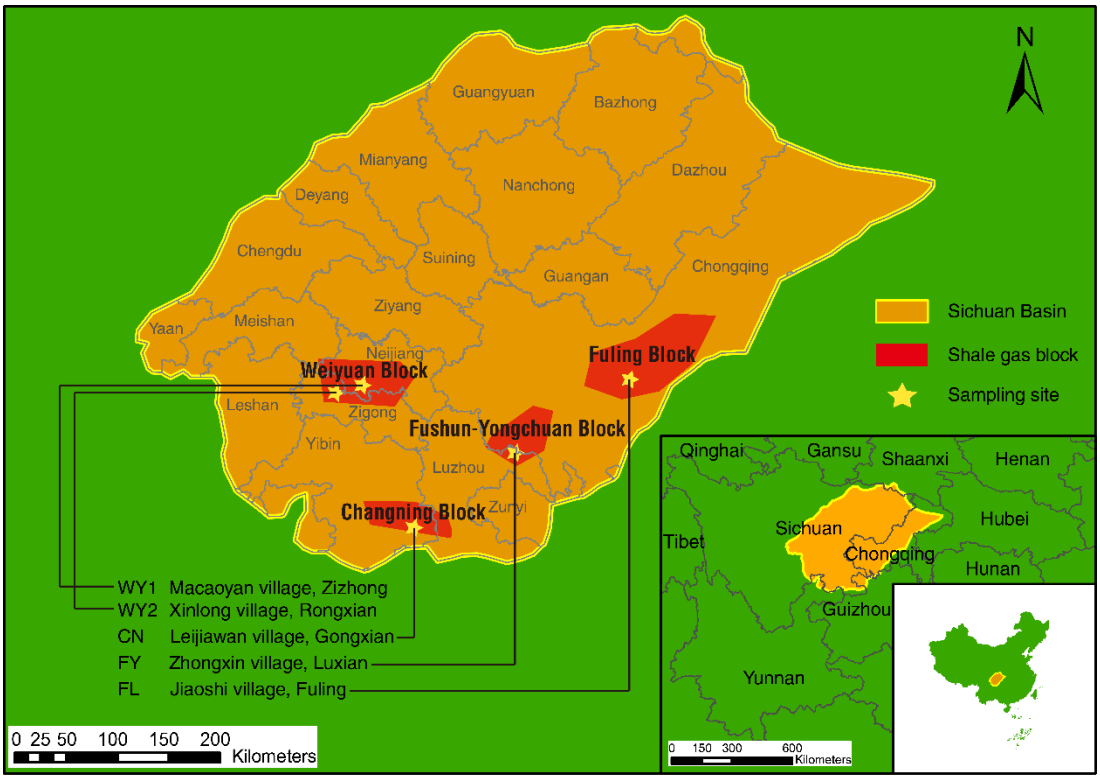


Fig. 1. Map displaying the major shale gas block locations and sampling sites within each block in the Sichuan Basin, China.

2.2. General quality parameters of SGW.

All the samples were analyzed for total dissolved solids (TDS), electrical conductivity (EC), pH, turbidity, dissolved organic carbon (DOC), total dissolved nitrogen (TDN), zeta potential, mean particle size, and inorganic ion concentrations. And we further examined the geochemical characteristics of the SGW samples by absorbance and fluorescence spectroscopy techniques. A wide set of spectral indexes,

including absorbance at 254 nm (UV_{254}), specific UV absorbance (SUVA), spectral slope (S) (Helms et al., 2008), fluorescence index (FI) (Kellerman et al., 2015; McKnight et al., 2001), biological index (BIX) (Parlanti et al., 2000; Wilson and Xenopoulos, 2009), and humification index (HIX) (Ohno, 2002; Zsolnay et al., 1999), were calculated to reflect the sources and structural characteristics of DOM in SGW. The detailed procedures for the above analyses can be found in **Text S1** of the Supporting Information **(SI)**.

2.3. Solid phase extraction and comparative study.

Previous investigations have shown that the extraction efficiency of DOM by solid phase extraction (SPE) is largely affected by SPE sorbents and DOM composition (Dittmar et al., 2008; Li et al., 2017). However, the selectivity of different SPE cartridges for isolating DOM in SGW has not yet been investigated, which might hinder an in-depth understanding of the composition and characteristics of SGW. Therefore, three widespread SPE cartridges (Bond Elut PPL, Oasis HLB, and Sep-Pak C18) were chosen in this study for isolation of SGW DOM following identical extraction procedure to obtain representative extracts. DOC recovery, UV–visible absorbance spectrometry, and FTIR spectroscopy were employed to compare the selectivity of PPL, HLB, and C18. The SPE procedure and analysis methods, as well as the corresponding results, are presented in **Text S2** and **Fig. S1**. Briefly, for the purpose of acquiring representative results, PPL was finally selected due to the consistently highest DOC recoveries and the excellent capacity to concentrate aromatic compounds and molecules

with oxygen-containing functional groups.

2.4. Analytical methods.

The molecular compositions of the PPL extracts of DOM from the five SGW samples were analyzed using a 9.4 T Bruker Apex Ultra FT-ICR mass spectrometer equipped with an Apollo II electrospray ionization (ESI) source. Based on the characterization obtained from FTIR (**Text S2**), suggesting the main occurrence of functional groups that will facilely lose a proton (such as carboxylic acids and alcohols), negative ESI mode was employed for this work (Sleighter and Hatcher, 2011). Ahead of the FT-ICR MS measurement, deuterated stearic acid ($C_{18}D_{35}H_1O_2$, 10^{-6} mol/L) was equally added to the PPL extracts diluted with methanol as an internal standard to obtain semi-quantitative results. The optimized instrumental parameters, procedures for mass calibration, data acquisition and processing can be found in **Text S3**.

The highly abundant class species indicated by FT-ICR MS were identified with UHPLC-QTOF-MS in electrospray negative ion mode. The linear alkyl benzene sulfonates (LAS) standard (C_{10}-C_{13} alkyl homologues, i.e., C_{10}-LAS—C_{13}-LAS) was obtained from Aladdin BioChem Technology Co., Ltd. (Shanghai, China) and three kinds of commercial LAS were supplied by Usolf Chemical, ChuangCheng Washing, and DongRun Chemical.

2.5. Biological treatment.

Biological treatment experiments were carried out in 300 mL glass bottles with 200 mL of liquid and 100 mL of headspace using sequencing batch reactors (SBR).

Five reactors were conducted, with one for each SGW sample. They were inoculated with activated sludge obtained from a municipal wastewater treatment plant (Chengdu, China) and aerated via aeration disks at the bottom to maintain the dissolved oxygen level at 4–6 mg/L. The reactors were operated in two-day cycles and 150 mL was replaced by new feed water at the end of each cycle. During the acclimation period, the feed water (a mixture of the municipal wastewater and SGW) TDS concentration was increased by ~3000 mg/L (by adding SGW) for every stage until it consisted of 100% SGW in each reactor. After acclimation, the reactors were operated for two additional cycles, with the influent and effluent samples taken for DOC analysis. In addition, control experiments were conducted without inoculation. The detailed steps of the biological treatment experiments are provided in **Table S1**.

3. Results and discussion

3.1. Geochemical characteristics of SGW.

The general quality parameters are summarized in **Table S2** to inform the bulk physical and chemical characteristics of SGW. It is obvious that the concentrations of the main constituents were highly variable spatially. TDS of the five SGW samples varied by a factor of 3, from 10880 to 33650 mg/L, and DOC by a factor of ~5, from 12.86 to 59.75 mg/L, representing the wide spectrum of characteristics observed for SGW globally (Chang et al., 2019; Tang et al., 2020a; Zhong et al., 2021). Especially, FI values, commonly used to reflect the sources of DOM, were all greater than 2.30, suggesting the predominant microbial sources of organic material in SGW (Kellerman et al., 2015; McKnight et al., 2001). BIX values that served as a tool to estimate biological activity were greater than 1.04, indicating the strong autogenetic characteristics of SGW DOM (Parlanti et al., 2000; Wilson and Xenopoulos, 2009). Moreover, HIX was developed to evaluate the humification extent of DOM, and the extremely low values (0.53-1.97) implied that SGW DOM was mainly derived from metabolic processes of microorganisms and was of low humification extent (Ohno, 2002; Zsolnay et al., 1999). Actually, the above analyses were mainly based on the differences in the values of the spectral indexes (FI, BIX, and HIX) between SGW and other water sources (such as freshwater and seawater), and the limitations of these indexes would be discussed in detail in **Section 3.2**.

3.2. Molecular characteristics of SGW.

The broadband mass spectra of the SGW samples showed several thousand peaks between m/z 200 and 700, and peaks were generally located below m/z 450 (**Fig. S2**). As opposed to the symmetric distribution for the mass spectra of DOM from typical water sources, such as Suwannee River Fulvic Acid (SRFA, a widely used DOM reference) (D'Andrilli et al., 2013), the SGW DOM distribution was quite irregular and dispersive. Nevertheless, the regular mass spacing patterns common to other types of water sources were also observed for SGW, including typical 14.0156 Da for CH_2 groups, 0.9953 Da (NH vs. CH_2), and 0.0364 Da (CH_4 vs. O) (Stenson et al., 2003). The CH_4 vs. O pattern is highlighted by showing the expanded spectra at m/z 275 (**Fig. S2**), and all assigned formulas at this nominal mass, sorted into different molecular series according to the substitution of CH_4 for O, are presented in **Table S3**. Most of the common formulas for the five SGW samples at this nominal mass were CHO formulas, whereas differences among the samples largely stemmed from compounds with more heteroatoms. And these trends were present along with the spectra.

The total number of assigned formulas and molecular-level parameters of the five SGW samples analyzed in this work are summarized in **Table S4**, together with those of samples from various aquatic ecosystems (including rivers, lakes, oceans, etc.) obtained from the literature (He et al., 2019; Li et al., 2017; Wang et al., 2020; Xu et al., 2020). Overall, SGW samples had apparently higher H/C ratios (1.36-1.56), and significantly lower average molecular weight (295.55-319.66), O/C ratios (0.253-

0.402), double bond equivalent (4.56-6.38) (D'Andrilli et al., 2013), and aromaticity index (0.027-0.197) (Koch and Dittmar, 2006) compared to other more traditional water samples (including freshwater, seawater, river water, and lake water). Therefore, DOM in SGW seemed to contain more aliphatic compounds with high levels of saturation and low oxidation degree. In addition, the lower N/C and S/C ratios might be associated with the narrow distribution range of the two heteroatoms. The remarkable differences observed between SGW with respect to other types of water sources, underline the importance of understanding the organic composition of SGW for its effective and safe management.

To investigate the similarity among the five SGW samples, principal component analysis (PCA) was performed based on the molecular-level parameters and spectral indexes, as shown in **Fig. S3**. The samples WY1 and WY2 from the same shale gas block were indeed similar, while FL was dissimilar to all the other samples. Furthermore, the relationships between molecular-level parameters and spectral indexes were determined by Spearman rank correlation and the color correlogram is presented in **Fig. S4** to provide a more systematic understanding of the nature of SGW DOM. In general, molecular-level parameters correlated well with each other. DBE_w, $(DBE/C)_w$, and AI_w were positively correlated with each other, while $(H/C)_w$ was negatively correlated with them, indicating the excellent power of the four parameters to predict the aromaticity and unsaturation degree of DOM. That is, DOM with lower DBE_w, $(DBE/C)_w$, and/or AI_w, and higher $(H/C)_w$ has higher degree of saturation, and

vice versa. $NOSC_w$ presented high level significant positive correlation with $(O/C)_w$. Stubbins *et al*. (2014) have suggested that fluorescent DOM (FDOM) would have the potential to track the bulk pool of DOM in freshwater (Stubbins et al., 2014). Many previous studies have also devoted to interpreting the broader DOM pool by utilizing chromophoric DOM (CDOM) and FDOM as prisms, and they have revealed the significant correlations between spectral indexes as well as between spectral indexes and molecular-level parameters: the S parameter is negatively correlated with $(m/z)_w$, and HIX is negatively correlated with BIX and positively correlated with $(O/C)_w$ (Singer et al., 2012; Wang et al., 2021; Zhang, B. et al., 2019). However, these relationships were not observed in this study, which would be attributed to the complexity of SGW. That is, the percentage and chemistry of CDOM and FDOM in SGW might be highly variable, also highlighting the limitations regarding the interpretation of DOM optical signatures in SGW. It should be noted that the molecular-level parameters are subjected to the selectivity of SPE and ESI in negative mode (D'Andrilli et al., 2020; Li et al., 2017), also probably leading to discrepancies from expected correlations. Besides, the limited number of samples should be cautioned here, and collecting more SGW samples for correlation analysis would be important in future efforts.

3.3. Molecular composition and distribution.

All SGW samples mainly contained CHO, CHNO, CHOS, and CHNOS categories, and their semi-quantitative relative abundance decreased consistently in the order CHO

(> 50%) > CHOS > CHNO > CHNOS (< 6%), except for sample CN that contained 53.75% CHOS compounds, followed by CHO, CHNO, and in which no CHNOS compounds were detected (**Fig. S5**). To further investigate the elemental composition, a total of 43, 56, 37, 44 and 28 classes were assigned for WY1, WY2, CN, FY, and FL, respectively, and the semi-quantitative relative abundance of each class for each sample is shown in **Figs. 2** and **S6-S9** in detail. The molecular composition varied significantly among the samples, and 23 classes in total including $O_{2\text{-}10}$, $N_1O_{2\text{-}8}$, $O_{3\text{-}9}S_1$ were assigned in all samples. The most abundant class species in the CHO category were $O_{2\text{-}6}$ for FL and $O_{4\text{-}8}$ for the other four samples, with much lower oxygen atom numbers than that of other traditional water types, which agreed well with the low $(O/C)_w$ values found for SGW in this study. N_nO_o always consisted of the largest number of classes, except for sample FL that had no molecular formulas containing more than one N heteroatom. Only O_oS_1 and $N_1O_oS_1$ class species were determined in CHOS and CHNOS, respectively, for all the samples.

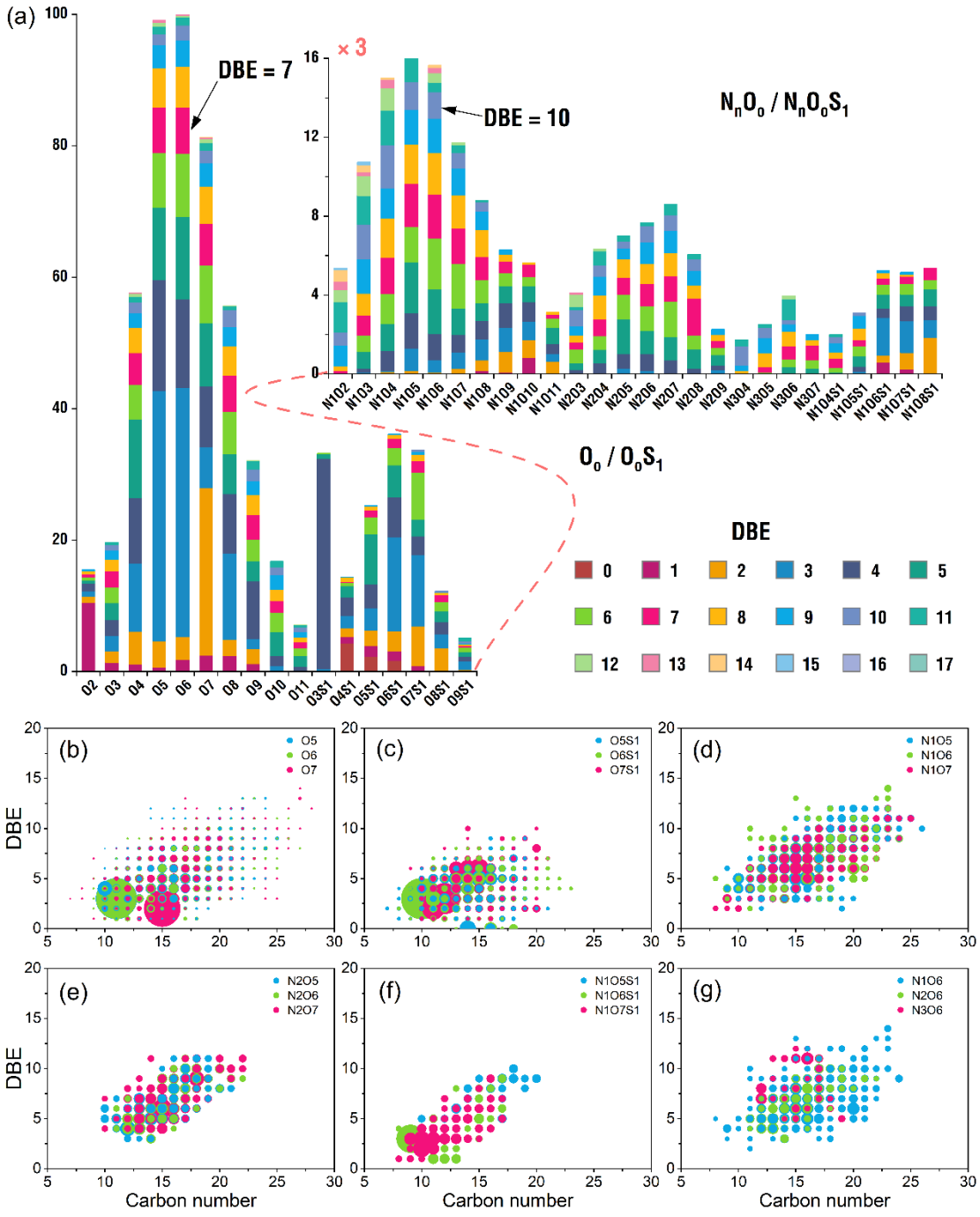


Fig. 2. Molecular composition and distribution of sample WY1. (a) Relative abundance of the assigned classes. The horizontal axis represents heteroatom classes and is split into 2 segments. The vertical scale for the N_nO_o and $N_nO_oS_s$ segment is enlarged 3 times. Columns with different colors correspond to compound with different DBE. Diagrams of DBE versus carbon numbers for O_5-O_7 (b), O_5S_1-O_7S_1 (c), N_1O_5-

340 N_1O_7 (d), N_2O_5-N_2O_7 (e), $N_1O_5S_1$-$N_1O_7S_1$ (f), and N_1O_6-N_3O_6 (g) classes. Bubble size
341 represents the relative abundance of subclasses in each diagram of DBE versus carbon
342 numbers.
343
344 The DBE_w distribution for each category is illustrated more clearly in **Table S5**.
345 CHNO compounds had the highest DBE_w (5.973-7.825), consistently for all samples.
346 Meanwhile, CHOS always had the lowest DBE_w, except for sample FY associated with
347 a slightly lower DBE_w for the CHO category (4.352). Also, the consistently higher
348 DBE_w of CHNO and CHNOS suggested that molecules with no N heteroatom were
349 much more saturated. This phenomenon might result from the more complex core
350 structures of nitrogen-containing compounds. Particularly, the fully saturated
351 compounds with DBE=0 were mainly composed of CHOS and there were no CHNO
352 compounds, see **Figs. 2** and **S6-S9**. Furthermore, according to the DBE vs. carbon
353 number data, classes with higher relative abundance extended to wider DBE and carbon
354 number ranges, and all classes showed the common tendency of increasing DBE as the
355 number of carbon atoms increased. The main O_5-O_7 classes species in sample WY1 had
356 comparatively low DBE values of 2-4 (**Fig. 2b**), indicating that hydroxyl, carbonyl,
357 and/or carboxyl might be contained in these compounds. The O_6S_1 class in WY1 had
358 DBE values of 0-9 and carbon numbers of 8-23 (**Fig. 2c**). The most abundant O_6S_1
359 species exhibited DBE=3 and the peak with corresponding neutral formula of
360 $C_{10}H_{16}O_6S_1$ had the highest relative abundance. In general, the relative abundance of

N-containing formulas was less variable (**Fig. 2d-g**), without obviously abundant subclasses (such as O_6 species with 3 DBEs and 11 carbon numbers in **Fig. 2b**). More importantly, it could be found that DBE increased with higher number of N atoms (**Fig. 2g**), suggesting more aromatic structures, and less obvious trends were also observed for most classes with increasing the number of O atoms (**Fig. 2b-f**).

Evidently, O_3S_1 class species showed an unusually high relative abundance in O_oS_1 and were characterized by the complete dominance of compounds with 4 DBEs, especially for sample CN (**Fig. S7**). For detailed examination, primary formulas common to all the samples, accounting for more than 95% of the subclass (O_3S_1 with 4 DBEs) by abundance, are listed in **Table S6**. And compounds with these assigned formulas were further confirmed as linear alkyl benzene sulfonates (LAS), the most widely used anionic surfactant, with alkyl chains in the range of 10 to 13 carbon atoms in length (C_{10}-LAS—C_{13}-LAS), based on the perfect matching of accurate mass, MS/MS spectrum, and retention time with LAS standard analyzed by UHPLC-QTOF-MS. Moreover, major compounds with formulas in subclasses of O_5S_1 with 5 DBEs and O_3S_1 with 5 DBEs were identified via MS/MS analysis as sulfophenyl carboxylic acids (SPC) and dialkyl tetralin sulfonates (DATS), respectively (**Table S6**). Note that quite similar fragmentation behaviors of the same sulfonates have been reported previously (Gonsior et al., 2011; Lara-Martín et al., 2010). Detailed descriptions and structures of the above compounds are given in **Fig. S10** and exemplary MS/MS spectra of some selected analytes are shown in **Fig. S11**.

LAS, SPC, and DATS constituted a considerable proportion of the dissolved organic sulfur (DOS) pool. It should be noted that these compounds were absent from the blanks. More importantly, the identification of LAS as fracturing additives was verified with the engineers responsible for the wells in the present study, despite the fact that the composition of hydraulic fracturing fluids has not been fully disclosed in China. Generally, the concentrations of fracturing additives in SGW decline quickly as the flowback time increases (Sun et al., 2019). Thus, it makes sense that LAS exhibited an extremely high relative abundance in sample CN (**Fig. S7**), as CN was collected in the relatively early stage of the well (at day 53 after hydraulic fracturing). Nonionic surfactants (such as polyethylene glycols and polypropylene glycols identified in electrospray positive ion mode) have been primarily reported surfactant additives that act as multifunctional chemicals (corrosion inhibitors, friction reducers or even biocides) in hydraulic fracturing in North America (Sun et al., 2019; Thurman et al., 2017). On the other hand, alcohol sulfate and alcohol ethoxysulfate were considered to be the dominant DOS classes in SGW samples from Morgantown, WV, USA (Luek et al., 2019). The results fully demonstrate the significant differences of composition and characteristics of SGW between China and North America (Zhong et al., 2021), putatively stemming from formation-specific geogenic constituents and various fracturing fluid chemistry designed as a function of well depth and geology, as well as operator discretion (Coonrod et al., 2020; Ferrer and Thurman, 2015; Stringfellow et al., 2014). Additionally, LAS would represent sensitive organic tracers of injected

fracturing fluid in such complex matrixes due to their very high ionization efficiencies in negative ESI (Luek et al., 2019).

The three kinds of commercial LAS shared similar distribution of the homologous series as measured and extracted ion chromatograms (EIC) of C_{10}-LAS—C_{13}-LAS in sample CN and in a representative commercial LAS (Usolf Chemical) are presented in **Fig. S12**. In accordance with previous studies (Di Corcia, A. et al., 1999; Di Corcia, Antonio et al., 1999), absorption to particles/soil and biodegradation processes are probably responsible for the shift of LAS distribution to homologues with shorter alkyl chain length in sample CN (**Fig. S12**). And the ubiquitous SPC, known as the most common biodegradation metabolites of LAS, indicate the abundant microbial activities in SGW during exploitation processes, simultaneously (Cluff et al., 2014; Zhou et al., 2022). However, it is unknown whether the biodegradation of LAS occurred in the deep subsurface (anaerobic), in open-air holding ponds after returning to the surface (aerobic), or both, as SPC could be generated under both aerobic and anaerobic conditions with distinct degradation pathways (Di Corcia, Antonio et al., 1999; Lara-Martín et al., 2010). It is worth noting that, although the unique reaction metabolites (4-methyl sulfophenyl dicarboxylic acids, Me-SPdC) of LAS under anaerobic conditions (Lara-Martín et al., 2010) were not identified in the five SGW samples, this observation alone was not sufficient to indicate the absence of anaerobic degradation, as we collected samples subjected to aerobic conditions for a long time in open-air holding ponds, which would also lead to the degradation of Me-SPdC (Lara-Martín et al., 2010).

Collecting and analyzing a time series of samples that newly return to the surface would help to draw convincing conclusions. In fact, it is highly undesirable when the biotransformation takes place underground as attenuation of injected additives potentially affect the well production efficiency (Evans et al., 2019; Luek et al., 2019) and stimulated microbial processes could lead to gas souring and infrastructure corrosion (Sirivedhin and Dallbauman, 2004). Therefore, biodegradation sites and associations between hydraulic fracturing additives and microbial growth need to be systematically investigated for reasonably designing fracturing fluid formulation and guaranteeing effective hydraulic fracturing engineering. Moreover, recent studies have reported the endocrine disrupting potential of LAS, SPC, and DATS (Geng et al., 2018), it is also vital to reveal the environmental fate of these compounds upon unintended release.

3.4. Van Krevelen diagram analysis.

The use of van Krevelen diagram is a prominent approach to visualizing such huge datasets and interpreting the respective chemical properties in DOM, based on the characteristic H/C and O/C ratios possessed by each major compound groups (Kim et al., 2003). In this work, detected molecules tended to be divided into seven compound groups with detailed classification boundaries discussed in **Text S3** (Hertkorn et al., 2006; Kim et al., 2003; Xu et al., 2020). Overall, the DOM in SGW distributed centrally and corresponded reasonably well with their $(H/C)_w$ and $(O/C)_w$ (**Fig. 3**). Clearly, more than 96% of the compounds with elemental ratios (semi-quantitative relative abundance)

located in lipids, aliphatic/proteins, and lignins/carboxylic rich alicyclic molecules (CRAM)-like structures, for all the samples (**Table S7**). The lignins/CRAM-like structures group was the most abundant in samples WY1 (42.74%), CN (46.19%), FY (51.67%) and FL (68.88%), while molecular signals in aliphatic/proteins region exhibited the highest percentage in WY2 (46.67%). Moreover, the excessively low proportion of unsaturated hydrocarbons, carbohydrates, tannins, and condensed aromatic structures is tabulated in **Table S7**.

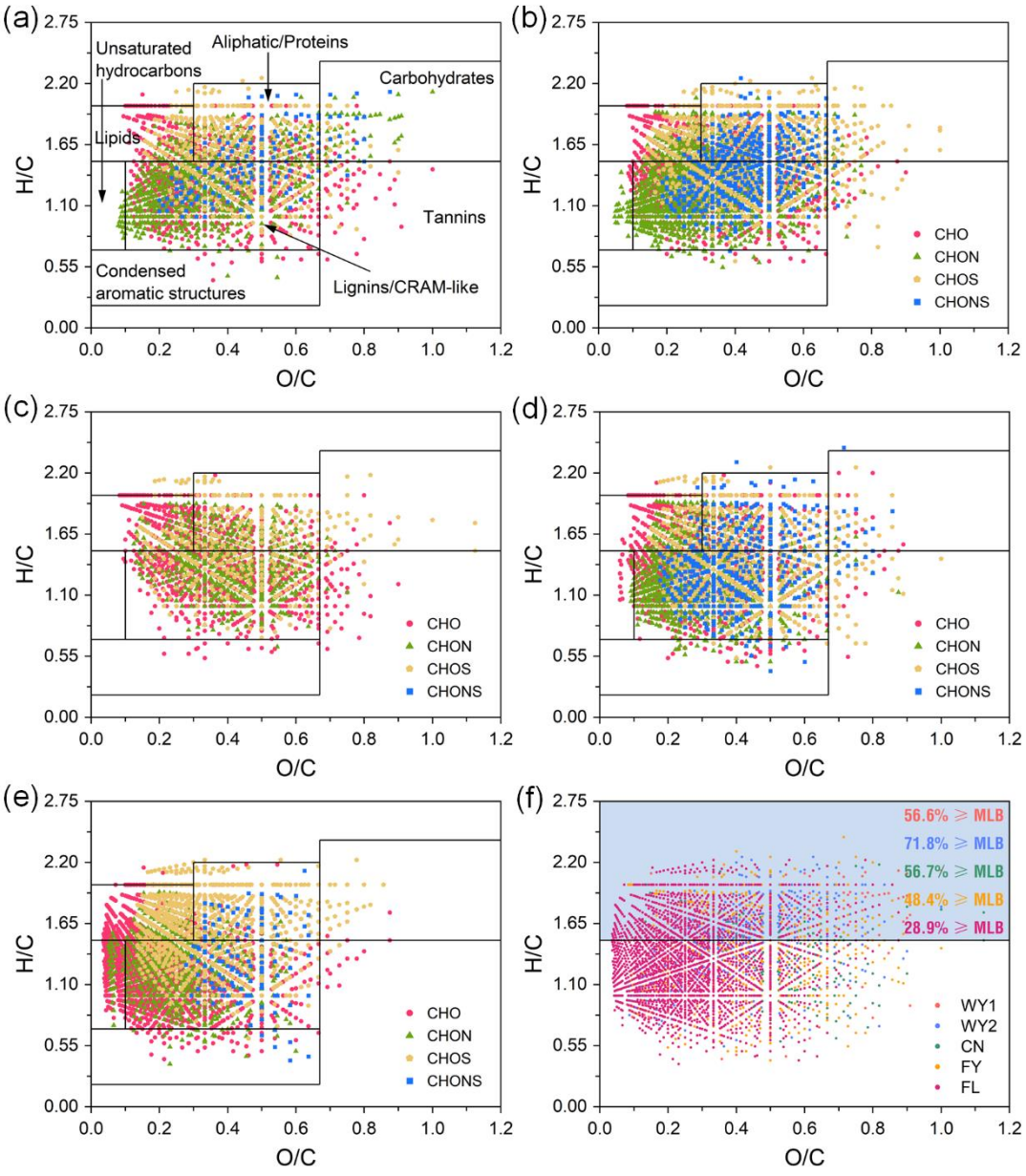


Fig. 3. Van Krevelen diagrams of sample WY1 (a), WY2 (b), CN (c), FY (d), and FL (e), constructed for each category (CHO, CHNO, CHOS, and CHNOS). The black lines depicting the boundaries corresponding to primary groups of DOM constituents are expounded in the VK diagram for WY1 (**Fig. 3a**). (f) Van Krevelen diagram for the five samples, with the proportion of intensity-weighted labile component indicated. The black line depicts the location of MLB, and the shaded area corresponds to labile region.

The chemical composition and nature of DOM in SGW would be associated with its unique sources (He et al., 2019). Especially, a considerable proportion of lipids- and aliphatic/proteins-like species, commonly used to reflect microbially influenced DOM (D'Andrilli et al., 2013), were assigned in SGW (WY1: 53.61%, WY2: 69.27%, CN: 52.36%, FY: 46.2%, FL: 27.9%). Combined with the high FI and BIX (**Section 3.1**), we suggested that DOM in SGW could partly originate from the organic material trapped within shales in the deep subsurface (Krumholz et al., 1997), a microbially dominated environment with abundant detritus and secretions of microorganisms . Although numerous research efforts have also found microbially derived protein signals in SGW (Riley et al., 2018; Tang et al., 2020b; Wang et al., 2019), it is important to note that the N-containing formulas in aliphatic/proteins region were not merely related to the proteins, given that N-containing formulas were concentrated mainly in lignins/CRAM-like region which implied various sources responsible for them.

In order to extend FT-ICR MS analysis toward further prediction of biogeochemical properties, D'Andrilli *et al*. (2015) have proposed the employment of the molecular lability boundary (MLB) at H/C=1.5 to determine the extent of DOM lability (D'Andrilli et al., 2015). Labile constituents in DOM refer to bioavailable carbon utilized in heterotrophic activity (Battin et al., 2008). Compounds above MLB at H/C≥1.5, including lipids, aliphatic/proteins, and carbohydrates in the VK diagram, exhibit a labile character (shaded region in **Fig. 3f**), whereas constituents below the MLB, H/C<1.5, correspond to generally refractory material. The normalized

intensities (MLB_L) and number (MLB_{nL}) of assigned labile molecules, as well as the relative proportion of them (MLB_L(%) and MLB_{nL}(%)) are summarized in **Table S8**, with their detailed explanations presented in **Text S3**. Interestingly, both MLB_L(%) (28.89-71.83%) and MLB_{nL}(%) (33.51-43.84%) indicated the much more labile nature of SGW compounds, compared to most other ecosystems including the glacial environments which are thought to contain a mass of labile component derived microbially (D'Andrilli et al., 2015; D'Andrilli et al., 2013). The MLB_L(%) and MLB_{nL}(%) were usually higher in CHO and CHOS compounds than in CHNO and CHNOS (**Table S8**). Furthermore, the distribution of different categories (CHO, CHNO, CHOS, and CHNOS) and groups (lipids, aliphatic/proteins, and carbohydrates) in labile region (H/C≥1.5) is illustrated more clearly in **Table S9**. As expected, CHO and CHOS compounds accounted for a considerable proportion (> 87% by peak intensity) in labile constituents. In addition, the majority (> 85% by peak intensity) of labile molecules were located in lipids and aliphatic/proteins regions, with much less contributions from molecules in carbohydrates.

The lignins/CRAM-like structures in SGW shared common features with different types of DOM from a variety of ecosystems. And this region was made up of various chemical species, including lignins, CRAM, and poly phenolic compounds. Lignin species are complicated and variable substances sourcing from higher plants and widely distributed in waters and soils (Stenson et al., 2003). The freshwater used for preparing fracturing fluid and leaching from solids during water-soil interactions might be

potential sources for this fraction of DOM in SGW. However, considering the low molecular signals in the tannins region that instead exist in substantial amounts in terrigenous environments, these sources were likely not the principal contributors. And the low HIX could support this hypothesis. Thus, we suggest that CRAM and poly phenolic compounds might be the predominant constituents in this region, consistent with the DOM composition in glacial and marine environments that are completely devoid of higher plants (Hertkorn et al., 2006; Stubbins et al., 2010). These compounds are associated with the end products of biodegradation and also exhibit recalcitrant character similar to that of humics and lignins (Hertkorn et al., 2006). In view of the complex composition and indiscernible sources of lignins/CRAM-like structures region, we focused more on the refractory nature of chemical species with elemental ratios in this region.

Undoubtedly, anthropogenic inputs (fracturing additives such as biocides, breakers, and corrosion inhibitors) also had profound influences on DOM composition, distribution, and evolution. As mentioned above, LAS, SPC, and DATS contributed largely to the corresponding class or even to the whole molecular composition. **Fig. S13** through different visualizations explicitly elucidate their appearance and patterns. What is worth being emphasized is the ambiguity in classifying thousands of organic compounds into several groups using VK diagram boundaries merely based on the two simplistic ratios (i.e., H/C and O/C), especially for water sources containing significant anthropogenic fraction, such as SGW. That is, LAS, SPC, and DATS are neither lipids

nor lignins/CRAM-like structures, but they fall in corresponding boundary regions of the VK diagram (**Fig. S13a**), respectively. Admittedly, there are various boundaries used to interpret VK diagrams from different literatures (Hertkorn et al., 2006; Kim et al., 2003; Wang et al., 2012; Xu et al., 2020), which would lead to distinct quantitative results, while similar conclusions can still be obtained here grounded in the similarities and differences between SGW and other important water types. Similarly, the ambiguity of the MLB, which is based solely on a H/C threshold value, should also be noted.

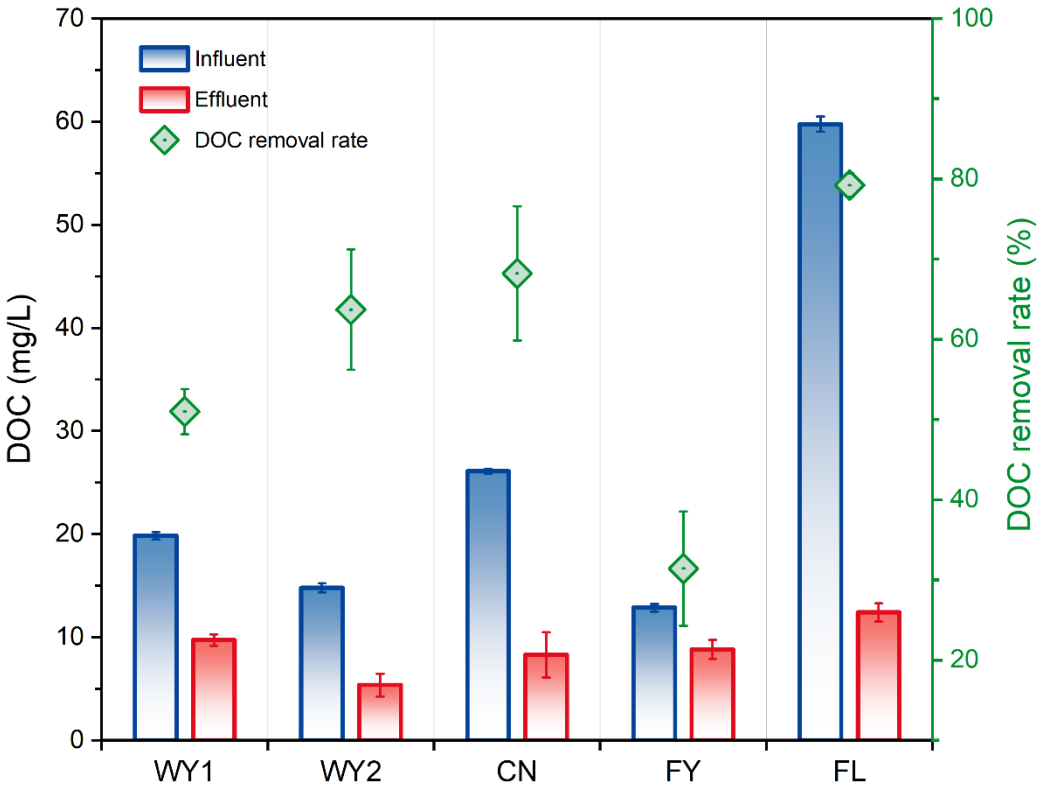
A total of 6176 molecular formulas were identified in the five SGW samples. These formulas were divided into three different groups (tCP(5): present in all of the 5 samples; pCP(2-4): present in at least 2 samples and at most 4 samples; DP(1): present in only one of the 5 samples). **Table S10** shows the distribution of formulas in the three groups. The proportion of unique molecular formulas (DP(1)) in samples WY2 and FL reached up to more than 23% (by peak number), which was much higher than that in WY1 (5.36%), CN (4.43%), and FY (10.28%). Van Krevelen diagrams for unique molecular formulas in each sample are presented in **Fig. S14a-e**. A significant proportion of the unique N-containing molecules (80.12% by peak number) was observed in WY2 (**Fig. S14b**). The unique CHO molecular formulas of sample FL were mainly located in O/C < 0.3 and 0.7 < H/C < 1.7 regions, and the unique CHOS formulas were mainly in 0.1 < O/C < 0.3 and 1.1 < H/C < 2.0 regions (**Fig. S14e**). It should be noticed that the proportion of different categories in DP(1) group calculated

by peak intensity showed the same trend as the proportion calculated by peak number for all the five samples. Moreover, 49.44%, 49.55%, 38.72%, 55.44%, and 37.93% of the molecular formulas in samples WY1, WY2, CN, FY, and FL, respectively, were classified as pCP(2-4) (**Fig. S14f**).

The 1207 common molecular formulas (tCP(5)) accounted for a considerable proportion in each sample, and the proportion calculated by intensity was much higher than that calculated by number (**Table S10**), indicating that the common formulas usually exhibited high intensities. **Fig. S15** shows that CHO (59.73%) molecules were the primary components in tCP(5). Furthermore, the inter-sample rankings analyses (Chen et al., 2021; Herzsprung et al., 2012; Zhang, L. et al., 2019) were used to evaluate the differences in DOM quality (**Fig. S16**). Sample FY exhibited ranks 1-2 (relatively high peak intensities) for lower molecular weight compounds (< 325 Da) (**Fig. S16h**), which might be caused by microbial degradation of high molecular weight organic matter into smaller molecules during the long-term storage (for 225 days) in the open-air holding ponds. Given that sample CN was collected in the relatively early stage of the well and the storage time of CN is the shortest (for 53 days), it is expected that more high molecular weight compounds could exist in CN. It is consistent with the fact that CN showed ranks 1-2 for components with higher molecular weight (400-500 Da) (**Fig. S16f**). In addition, the molecular formulas in the ranks 1-2 of sample FL were mainly located in O/C < 0.3 region (**Fig. S16i**).

3.5. Biodegradation for SGW treatment.

Camarillo *et al*. (2016) previously showed that 37% of 155 fracturing additives, which were variable in different wells and solely represented a small fraction of the DOM pool, could be effectively removed by biological treatment based on the available physical-chemical data (Camarillo et al., 2016). In the present study, numerous labile constituents of DOM in real SGW obtained from FT-ICR MS (**Section 3.4**) suggest that biodegradation would be suitable for SGW treatment and reuse. Thus, we conducted preliminary biological treatment experiments (SBR) to support this hypothesis. **Fig. 4** shows that the efficacies of DOC removal reached 50.98%, 63.69%, 68.23%, 31.42%, and 79.23% for samples WY1, WY2, CN, FY, and FL, respectively. Similar removal rates (RR) were observed in other researches treating SGW: 45.3-83.2% by biologically active filtration (Tang et al., 2022), 59.5-87.9% by membrane bioreactor (Liu, X. et al., 2022), and 52-85% by SBR (Sitterley et al., 2021).



577

578 **Fig. 4.** Removal of DOC from the five SGW samples by activated sludge SBR.

579 Note that no considerable decrease in DOC concentration was obtained in the non-

580 inoculated control experiments.

581

582 Furthermore, the relationships between DOC removal rate (RR) and some key

583 parameters that are considered likely to affect the performance of biological treatment

584 were assessed by Spearman rank correlation coefficients (**Fig. S17**). RR was positively

585 correlated with MLB_L (the sum of normalized intensities of assigned labile molecules)

586 and DOC, but not with MLB_{nL}, MLB_L(%), and MLB_{nL}(%), suggesting that higher

587 absolute content of labile components would facilitate microbial metabolism and lead

588 to more organic removal. Actually, MLB_L(%) and MLB_{nL}(%) have been generally used

to assess the extent of DOM lability in published researches (D'Andrilli et al., 2015; He et al., 2019), while MLB_L showed greater potential to indicate the biological treatability of SGW in this study.

Notably, since there was no attempt to optimize the acclimation process in the biological treatment experiments, the DOC removal rates obtained might only represent relatively low levels. Therefore, our results (31.42-79.23% of DOC removal) would support the feasibility of biodegradation for SGW treatment. Given the inherently labile nature of SGW, we highlight that robustly constructing a salt-tolerant microbial community, disclosing key degradation pathways, optimizing reactor designs, as well as combining effective pretreatment techniques (e.g., advanced oxidation, adsorption) should represent future research efforts.

4. Conclusions

The present study provided a detailed investigation on the composition and characteristics of DOM in complex SGW from four key shale gas blocks in the Sichuan Basin, China using a suite of analysis techniques (FT-ICR MS, UHPLC-QTOF-MS, UV–Vis spectrophotometry, and fluorescence spectrophotometry). In general, SGW DOM was characterized by low average molecular weight, high saturation degree, and low aromaticity. However, the correlations between molecular-level parameters and spectral indexes observed in samples from more traditional aquatic ecosystems were absent from SGW, implying the high complexity and variability of SGW. Moreover, the identification of LAS, as well as associated biodegraded metabolites (SPC) and coproducts (DATS) suggested the distinct anthropogenic imprints and abundant microbial activities during exploitation processes. These findings also underline the need for further revealing the environmental fate of hydraulic fracturing additives and the associations between these compounds and microbial growth, with the objective of guaranteeing effective hydraulic fracturing engineering and reducing relevant environmental risks. Based on the van Krevelen diagram analysis, we found that various sources (microbially derived organics in shales and biochemical transformations) might be responsible for SGW DOM, while the limitations and ambiguity of classification boundaries in VK diagram should also be noted. Furthermore, 31.42-79.23% of DOC were effectively removed by SBR, fully supporting the inherently labile nature of SGW as indicated by FT-ICR MS and

providing the theoretical basis for full-scale biological treatment of SGW in long-term operation.

CRediT authorship contribution statement

Xuanyu Ji: Methodology, Investigation, Visualization, Writing - original draft. **Alberto Tiraferri**: Validation, Writing - review & editing. **Xiaofei Zhang**: Resources, Data curation. **Peng Liu**: Resources, Data curation, Software, Formal analysis. **Zhiwei Gan**: Resources, Investigation. **John C. Crittenden**: Methodology, Data curation, Writing - review & editing. **Jun Ma**: Methodology, Writing - review & editing. **Baicang Liu**: Conceptualization, Supervision, Writing - review & editing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (52070134, 52270075), and Sichuan University and Yibin City People's Government Strategic Cooperation Project (2020CDYB-2).

Supporting Information

Additional experimental materials, methods, and procedures for data processing; discussion on SPE and comparative analysis; general water quality of the five SGW samples; detailed distribution of organic compounds; and a number of other supporting materials are provided in Texts S1−S3, Tables S1−S10, and Figs. S1−S17.

REFERENCES

Barbot, E., Vidic, N.S., Gregory, K.B., Vidic, R.D., 2013. Spatial and Temporal Correlation of Water Quality Parameters of Produced Waters from Devonian-Age Shale following Hydraulic Fracturing. Environ. Sci. Technol. 47(6), 2562-2569. https://doi.org/10.1021/es304638h.

Battin, T.J., Kaplan, L.A., Findlay, S., Hopkinson, C.S., Marti, E., Packman, A.I., Newbold, J.D., Sabater, F., 2008. Biophysical controls on organic carbon fluxes in fluvial networks. Nat. Geosci. 1(2), 95-100. https://doi.org/10.1038/ngeo101.

Butkovskyi, A., Bruning, H., Kools, S.A.E., Rijnaarts, H.H.M., Van Wezel, A.P., 2017. Organic Pollutants in Shale Gas Flowback and Produced Waters: Identification, Potential Ecological Impact, and Implications for Treatment Strategies. Environ. Sci. Technol. 51(9), 4740-4754. https://doi.org/10.1021/acs.est.6b05640.

Camarillo, M.K., Domen, J.K., Stringfellow, W.T., 2016. Physical-chemical evaluation of hydraulic fracturing chemicals in the context of produced water treatment. J. Environ. Manage. 183, 164-174. https://doi.org/https://doi.org/10.1016/j.jenvman.2016.08.065.

Chang, H., Li, T., Liu, B., Vidic, R.D., Elimelech, M., Crittenden, J.C., 2019. Potential and implemented membrane-based technologies for the treatment and reuse of flowback and produced water from shale gas and oil plays: A review. Desalination 455, 34-57. https://doi.org/10.1016/j.desal.2019.01.001.

Chen, Q., Chen, F., Gonsior, M., Li, Y., Wang, Y., He, C., Cai, R., Xu, J., Wang, Y., Xu, D., Sun, J., Zhang, T., Shi, Q., Jiao, N., Zheng, Q., 2021. Correspondence between DOM molecules and microbial community in a subtropical coastal estuary on a spatiotemporal scale. ENVIRON INT 154, 106558. https://doi.org/https://doi.org/10.1016/j.envint.2021.106558.

Chen, W., Zhuo, X., He, C., Shi, Q., Li, Q., 2020. Molecular investigation into the transformation of dissolved organic matter in mature landfill leachate during treatment in a combined membrane bioreactor-reverse osmosis process. J. Hazard. Mater. 397, 122759. https://doi.org/https://doi.org/10.1016/j.jhazmat.2020.122759.

Cluff, M.A., Hartsock, A., MacRae, J.D., Carter, K., Mouser, P.J., 2014. Temporal Changes in Microbial Ecology and Geochemistry in Produced Water from Hydraulically Fractured Marcellus Shale Gas Wells. Environ. Sci. Technol. 48(11), 6508-6517. https://doi.org/10.1021/es501173p.

Coonrod, C.L., Yin, Y.B., Hanna, T., Atkinson, A.J., Alvarez, P.J.J., Tekavec, T.N., Reynolds, M.A., Wong, M.S., 2020. Fit-for-purpose treatment goals for produced waters in shale oil and gas fields. Water Res. 173. https://doi.org/10.1016/j.watres.2020.115467.

D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G., 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Spectrom. 29(24), 2385-2401.

https://doi.org/10.1002/rcm.7400.
D'Andrilli, J., Fischer, S.J., Rosario-Ortiz, F.L., 2020. Advancing Critical Applications of High Resolution Mass Spectrometry for DOM Assessments: Re-Engaging with Mass Spectral Principles, Limitations, and Data Analysis. Environ. Sci. Technol. 54(19), 11654-11656. https://doi.org/10.1021/acs.est.0c04557.
D'Andrilli, J., Foreman, C.M., Marshall, A.G., McKnight, D.M., 2013. Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. Org. Geochem. 65, 19-28. https://doi.org/10.1016/j.orggeochem.2013.09.013.
Di Corcia, A., Capuani, L., Casassa, F., Marcomini, A., Samperi, R., 1999. Fate of linear alkyl benzenesulfonates, coproducts, and their metabolites in sewage treatment plants and in receiving river waters. Environ. Sci. Technol. 33(22), 4119-4125. https://doi.org/10.1021/es990596u.
Di Corcia, A., Casassa, F., Crescenzi, C., Marcomini, A., Samperi, R., 1999. Investigation of the Fate of Linear Alkyl Benzenesulfonates and Coproducts in a Laboratory Biodegradation Test by Using Liquid Chromatography/Mass Spectrometry. Environ. Sci. Technol. 33(22), 4112-4118. https://doi.org/10.1021/es9905952.
Dittmar, T., Koch, B., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol Oceanogr-Meth 6, 230-235. https://doi.org/10.4319/lom.2008.6.230.
Evans, M.V., Getzinger, G., Luek, J.L., Hanson, A.J., McLaughlin, M.C., Blotevogel, J., Welch, S.A., Nicora, C.D., Purvine, S.O., Xu, C.D., Cole, D.R., Darrah, T.H., Hoyt, D.W., Metz, T.O., Ferguson, P.L., Lipton, M.S., Wilkins, M.J., Mouser, P.J., 2019. In situ transformation of ethoxylate and glycol surfactants by shale-colonizing microorganisms during hydraulic fracturing. ISME J 13(11), 2690-2700. https://doi.org/10.1038/s41396-019-0466-0.
Ferrer, I., Thurman, E.M., 2015. Chemical constituents and analytical approaches for hydraulic fracturing waters. Trends Environ. Anal. Chem. 5, 18-25. https://doi.org/https://doi.org/10.1016/j.teac.2015.01.003.
Geng, C.X., Cao, N., Xu, W., He, C., Yuan, Z.W., Liu, J.W., Shi, Q., Xu, C.M., Liu, S.T., Zhao, H.Z., 2018. Molecular Characterization of Organics Removed by a Covalently Bound Inorganic-Organic Hybrid Coagulant for Advanced Treatment of Municipal Sewage. Environ. Sci. Technol. 52(21), 12642-12648. https://doi.org/10.1021/acs.est.8b03306.
Gonsior, M., Zwartjes, M., Cooper, W.J., Song, W., Ishida, K.P., Tseng, L.Y., Jeung, M.K., Rosso, D., Hertkorn, N., Schmitt-Kopplin, P., 2011. Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. Water Res. 45(9), 2943-2953. https://doi.org/https://doi.org/10.1016/j.watres.2011.03.016.

He, D., He, C., Li, P.H., Zhang, X.W., Shi, Q., Sun, Y.G., 2019. Optical and Molecular Signatures of Dissolved Organic Matter Reflect Anthropogenic Influence in a Coastal River, Northeast China. J. Environ. Qual. 48(3), 603-613. https://doi.org/10.2134/jeq2018.09.0330.

He, P.J., Liu, W.Y., Qiu, J.J., Zhang, H., Huang, Y.L., Deng, Y.T., Shao, L.M., Lu, F., 2021. Improvement criteria for different advanced technologies towards bio-stabilized leachate based on molecular subcategories of DOM. J. Hazard. Mater. 414. https://doi.org/10.1016/j.jhazmat.2021.125463.

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr. 53(3), 955-969. https://doi.org/10.4319/lo.2008.53.3.0955.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., Hedges, J.I., 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochim. Cosmochim. Acta 70(12), 2990-3010. https://doi.org/10.1016/j.gca.2006.03.021.

Herzsprung, P., von Tümpling, W., Hertkorn, N., Harir, M., Büttner, O., Bravidor, J., Friese, K., Schmitt-Kopplin, P., 2012. Variations of DOM Quality in Inflows of a Drinking Water Reservoir: Linking of van Krevelen Diagrams with EEMF Spectra by Rank Correlation. Environ. Sci. Technol. 46(10), 5511-5518. https://doi.org/10.1021/es300345c.

Hou, D.Y., Luo, J., Al-Tabbaa, A., 2012. COMMENTARY: Shale gas can be a double-edged sword for climate change. Nat Clim Chang 2(6), 385-387. https://doi.org/10.1038/nclimate1500.

Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nat. Geosci. 8(6), 454-U452. https://doi.org/10.1038/ngeo2440.

Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. Anal. Chem. 75(20), 5336-5344. https://doi.org/10.1021/ac034415p.

Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 20(5), 926-932. https://doi.org/10.1002/rcm.2386.

Kondash, A.J., Lauer, N.E., Vengosh, A., 2018. The intensification of the water footprint of hydraulic fracturing. Sci. Adv. 4(8). https://doi.org/10.1126/sciadv.aar5982.

Krumholz, L.R., McKinley, J.P., Ulrich, F.A., Suflita, J.M., 1997. Confined subsurface microbial communities in Cretaceous rock. Nature 386(6620), 64-66. https://doi.org/10.1038/386064a0.

Lara-Martín, P.A., Gómez-Parra, A., Sanz, J.L., González-Mazo, E., 2010. Anaerobic Degradation Pathway of Linear Alkylbenzene Sulfonates (LAS) in Sulfate-Reducing Marine Sediments. Environ. Sci. Technol. 44(5), 1670-1676.

https://doi.org/10.1021/es9032887.
Li, C., Zheng, M., Cao, D., Yang, L., Wu, J., Yang, Y., Liu, G., 2021. Recognition of the molecular characterization and mechanisms of heterogeneously formed organic pollutants from metallurgical industries by FT-ICR-MS and GC/Q-TOF-MS. J. Hazard. Mater. 406, 124603. https://doi.org/10.1016/j.jhazmat.2020.124603.
Li, Y., Harir, M., Uhl, J., Kanawati, B., Lucio, M., Smirnov, K.S., Koch, B.P., Schmitt-Kopplin, P., Hertkorn, N., 2017. How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. Water Res. 116, 316-323. https://doi.org/10.1016/j.watres.2017.03.038.
Liu, X., Tang, P., Liu, Y., Xie, W., Chen, C., Li, T., He, Q., Bao, J., Tiraferri, A., Liu, B., 2022. Efficient removal of organic compounds from shale gas wastewater by coupled ozonation and moving-bed-biofilm submerged membrane bioreactor. Bioresour. Technol. 344, 126191. https://doi.org/https://doi.org/10.1016/j.biortech.2021.126191.
Liu, Y., Wu, Q., Chen, C., Li, T., Liu, S., He, Q., Yang, P., Bai, Y., Liu, B., 2022. An efficient system of aerogel adsorbent combined with membranes for reuse of shale gas wastewater. Desalination 526, 115545. https://doi.org/https://doi.org/10.1016/j.desal.2021.115545.
Luek, J.L., Gonsior, M., 2017. Organic compounds in hydraulic fracturing fluids and wastewaters: A review. Water Res. 123, 536-548. https://doi.org/https://doi.org/10.1016/j.watres.2017.07.012.
Luek, J.L., Harir, M., Schmitt-Kopplin, P., Mouser, P.J., Gonsior, M., 2018. Temporal dynamics of halogenated organic compounds in Marcellus Shale flowback. Water Res. 136, 200-206. https://doi.org/10.1016/j.watres.2018.02.055.
Luek, J.L., Harir, M., Schmitt-Kopplin, P., Mouser, P.J., Gonsior, M., 2019. Organic sulfur fingerprint indicates continued injection fluid signature 10 months after hydraulic fracturing. ENVIRON SCI-PROC IMP 21(2), 206-213. https://doi.org/10.1039/c8em00331a.
Luek, J.L., Schmitt-Kopplin, P., Mouser, P.J., Petty, W.T., Richardson, S.D., Gonsior, M., 2017. Halogenated Organic Compounds Identified in Hydraulic Fracturing Wastewaters Using Ultrahigh Resolution Mass Spectrometry. Environ. Sci. Technol. 51(10), 5377-5385. https://doi.org/10.1021/acs.est.6b06213.
Marshall, A.G., Hendrickson, C.L., Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom. Rev. 17(1), 1-35. https://doi.org/10.1002/(sici)1098-2787(1998)17:1<1::Aid-mas1>3.0.Co;2-k.
McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnol. Oceanogr. 46(1), 38-48. https://doi.org/10.4319/lo.2001.46.1.0038.
Ni, Y., Zou, C., Cui, H., Li, J., Lauer, N.E., Harkness, J.S., Kondash, A.J., Coyte, R.M., Dwyer, G.S., Liu, D., Dong, D., Liao, F., Vengosh, A., 2018. Origin of Flowback

and Produced Waters from Sichuan Basin, China. Environ. Sci. Technol. 52(24), 14519-14527. https://doi.org/10.1021/acs.est.8b04345.

Oetjen, K., Giddings, C.G.S., McLaughlin, M., Nell, M., Blotevogel, J., Helbling, D.E., Mueller, D., Higgins, C.P., 2017. Emerging analytical methods for the characterization and quantification of organic contaminants in flowback and produced water. Trends Environ. Anal. Chem. 15, 12-23. https://doi.org/https://doi.org/10.1016/j.teac.2017.07.002.

Ohno, T., 2002. Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environ. Sci. Technol. 36(4), 742-746. https://doi.org/10.1021/es0155276.

Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. Org. Geochem. 31(12), 1765-1781. https://doi.org/https://doi.org/10.1016/S0146-6380(00)00124-8.

Qin, Y., Höglund-Isaksson, L., Byers, E., Feng, K., Wagner, F., Peng, W., Mauzerall, D.L., 2018. Air quality–carbon–water synergies and trade-offs in China's natural gas industry. Nat. Sustain. 1(9), 505-511. https://doi.org/10.1038/s41893-018-0136-7.

Riley, S.M., Ahoor, D.C., Regnery, J., Cath, T.Y., 2018. Tracking oil and gas wastewater-derived organic matter in a hybrid biofilter membrane treatment system: A multi-analytical approach. Sci. Total Environ. 613-614, 208-217. https://doi.org/10.1016/j.scitotenv.2017.09.031.

Robbins, C.A., Du, X., Bradley, T.H., Quinn, J.C., Bandhauer, T.M., Conrad, S.A., Carlson, K.H., Tong, T., 2022. Beyond treatment technology: Understanding motivations and barriers for wastewater treatment and reuse in unconventional energy production. Resour Conserv Recycl 177, 106011. https://doi.org/https://doi.org/10.1016/j.resconrec.2021.106011.

Singer, G.A., Fasching, C., Wilhelm, L., Niggemann, J., Steier, P., Dittmar, T., Battin, T.J., 2012. Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate. Nat. Geosci. 5(10), 710-714. https://doi.org/10.1038/ngeo1581.

Sirivedhin, T., Dallbauman, L., 2004. Organic matrix in produced water from the Osage-Skiatook Petroleum Environmental Research Site, Osage County, Oklahoma. Chemosphere 57(6), 463-469. https://doi.org/10.1016/j.chemosphere.2004.05.034.

Sitterley, K.A., Silverstein, J., Rosenblum, J., Linden, K.G., 2021. Aerobic biological degradation of organic matter and fracturing fluid additives in high salinity hydraulic fracturing wastewaters. Sci. Total Environ. 758, 143622. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.143622.

Sleighter, R.L., Hatcher, P.G., 2011. Fourier Transform Mass Spectrometry for the Molecular Level Characterization of Natural Organic Matter: Instrument Capabilities, Applications, and Limitations. In Fourier Transforms-Approach to Scientific Principles.

Stenson, A.C., Marshall, A.G., Cooper, W.T., 2003. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. Anal. Chem. 75(6), 1275-1284. https://doi.org/10.1021/ac026106p.

Stringfellow, W.T., Domen, J.K., Camarillo, M.K., Sandelin, W.L., Borglin, S., 2014. Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing. J. Hazard. Mater. 275, 37-54. https://doi.org/https://doi.org/10.1016/j.jhazmat.2014.04.040.

Stubbins, A., Lapierre, J.F., Berggren, M., Prairie, Y.T., Dittmar, T., del Giorgio, P.A., 2014. What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada. Environ. Sci. Technol. 48(18), 10598-10606. https://doi.org/10.1021/es502086e.

Stubbins, A., Spencer, R.G.M., Chen, H.M., Hatcher, P.G., Mopper, K., Hernes, P.J., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., Six, J., 2010. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. Limnol. Oceanogr. 55(4), 1467-1477. https://doi.org/10.4319/lo.2010.55.4.1467.

Sun, C., Zhang, Y., Alessi, D.S., Martin, J.W., 2019. Nontarget profiling of organic compounds in a temporal series of hydraulic fracturing flowback and produced waters. ENVIRON INT 131, 104944. https://doi.org/https://doi.org/10.1016/j.envint.2019.104944.

Sun, Y., Wu, M., Tong, T., Liu, P., Tang, P., Gan, Z., Yang, P., He, Q., Liu, B., 2021. Organic compounds in Weiyuan shale gas produced water: Identification, detection and rejection by ultrafiltration-reverse osmosis processes. Chem. Eng. J. 412, 128699. https://doi.org/https://doi.org/10.1016/j.cej.2021.128699.

Tang, P., Li, J., Li, T., Tian, L., Sun, Y., Xie, W., He, Q., Chang, H., Tiraferri, A., Liu, B., 2020a. Efficient integrated module of gravity driven membrane filtration, solar aeration and GAC adsorption for pretreatment of shale gas wastewater. J. Hazard. Mater., 124166. https://doi.org/https://doi.org/10.1016/j.jhazmat.2020.124166.

Tang, P., Liu, B., Zhang, Y., Chang, H., Zhou, P., Feng, M., Sharma, V.K., 2020b. Sustainable reuse of shale gas wastewater by pre-ozonation with ultrafiltration-reverse osmosis. Chem. Eng. J. 392, 123743. https://doi.org/https://doi.org/10.1016/j.cej.2019.123743.

Tang, P., Xie, W., Tian, L., Tan, B., Zhang, Y., Yang, Z., Chen, C., Zhang, W., Liu, B., 2022. Oxidation-biotreatment-membrane combined process for external reuse of shale gas wastewater. Sep. Purif. Technol. 291, 120920. https://doi.org/https://doi.org/10.1016/j.seppur.2022.120920.

Thurman, E.M., Ferrer, I., Rosenblum, J., Linden, K., Ryan, J.N., 2017. Identification of polypropylene glycols and polyethylene glycol carboxylates in flowback and produced water from hydraulic fracturing. J. Hazard. Mater. 323(Pt A), 11-17. https://doi.org/10.1016/j.jhazmat.2016.02.041.

Tong, T.Z., Carlson, K.H., Robbins, C.A., Zhang, Z.Y., Du, X.W., 2019. Membrane-

based treatment of shale oil and gas wastewater: The current state of knowledge. Front Environ Sci Eng 13(4). https://doi.org/10.1007/s11783-019-1147-y.
Vengosh, A., Jackson, R.B., Warner, N., Darrah, T.H., Kondash, A., 2014. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environ. Sci. Technol. 48(15), 8334-8348. https://doi.org/10.1021/es405118y.
Vidic, R.D., Brantley, S.L., Vandenbossche, J.M., Yoxtheimer, D., Abad, J.D., 2013. Impact of Shale Gas Development on Regional Water Quality. Science 340(6134). https://doi.org/10.1126/science.1235009.
Wang, D.D., Song, C.T., Zhang, B.L., Chen, J.W., Luo, A.L., Wang, X.S., Wu, S.D., Ye, Y.X., 2021. Deciphering dissolved organic matter from freshwater aquaculture ponds in Eastern China based on optical and molecular signatures. PROCESS SAF ENVIRON 155. https://doi.org/10.1016/j.psep.2021.09.025.
Wang, H., Lu, L., Chen, X., Bian, Y., Ren, Z.J., 2019. Geochemical and microbial characterizations of flowback and produced water in three shale oil and gas plays in the central and western United States. Water Res. 164, 114942. https://doi.org/https://doi.org/10.1016/j.watres.2019.114942.
Wang, X., Goual, L., Colberg, P.J.S., 2012. Characterization and treatment of dissolved organic matter from oilfield produced waters. J. Hazard. Mater. 217-218, 164-170. https://doi.org/https://doi.org/10.1016/j.jhazmat.2012.03.006.
Wang, X., Ji, Y., Shi, Q., Zhang, Y., He, C., Wang, Q., Guo, S., Chen, C., 2020. Characterization of wastewater effluent organic matter with different solid phase extraction sorbents. Chemosphere 257, 127235. https://doi.org/10.1016/j.chemosphere.2020.127235.
Wilson, H.F., Xenopoulos, M.A., 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nat. Geosci. 2(1), 37-41. https://doi.org/10.1038/ngeo391.
Xie, W., Tian, L., Tang, P., Cui, J., Wang, T., Zhu, Y., Bai, Y., Tiraferri, A., Crittenden, J.C., Liu, B., 2022. Shale gas wastewater characterization: Comprehensive detection, evaluation of valuable metals, and environmental risks of heavy metals and radionuclides. Water Res. 220, 118703. https://doi.org/https://doi.org/10.1016/j.watres.2022.118703.
Xu, W., Gao, Q., He, C., Shi, Q., Hou, Z.-Q., Zhao, H.-Z., 2020. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. Environ. Sci. Technol. 54(20), 12929-12937. https://doi.org/10.1021/acs.est.0c01681.
Yu, M., Weinthal, E., Patino-Echeverri, D., Deshusses, M.A., Zou, C., Ni, Y., Vengosh, A., 2016. Water Availability for Shale Gas Development in Sichuan Basin, China. Environ. Sci. Technol. 50(6), 2837-2845. https://doi.org/10.1021/acs.est.5b04669.
Zhang, B., Shan, C., Hao, Z., Liu, J., Wu, B., Pan, B., 2019. Transformation of dissolved organic matter during full-scale treatment of integrated chemical wastewater: Molecular composition correlated with spectral indexes and acute toxicity. Water

Res. 157, 472-482. https://doi.org/https://doi.org/10.1016/j.watres.2019.04.002.
Zhang, L., Peng, Y., Ge, Z., Xu, K., 2019. Fate of dissolved organic nitrogen during the Anammox process using ultra-high resolution mass spectrometry. ENVIRON INT 131, 105042. https://doi.org/https://doi.org/10.1016/j.envint.2019.105042.
Zhong, C., Zolfaghari, A., Hou, D., Goss, G.G., Lanoil, B.D., Gehman, J., Tsang, D.C.W., He, Y., Alessi, D.S., 2021. Comparison of the Hydraulic Fracturing Water Cycle in China and North America: A Critical Review. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c06119.
Zhou, S., Peng, S., Li, Z., Zhang, D., Zhu, Y., Li, X., Hong, M., Li, W., Lu, P., 2022. Characterization of microbial communities and functions in shale gas wastewaters and sludge: Implications for pretreatment. J. Hazard. Mater. 424, 127649. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.127649.
Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. Chemosphere 38(1), 45-50. https://doi.org/https://doi.org/10.1016/S0045-6535(98)00166-0.

Figure 1

Click here to access/download;Figure;Figure 1.tif

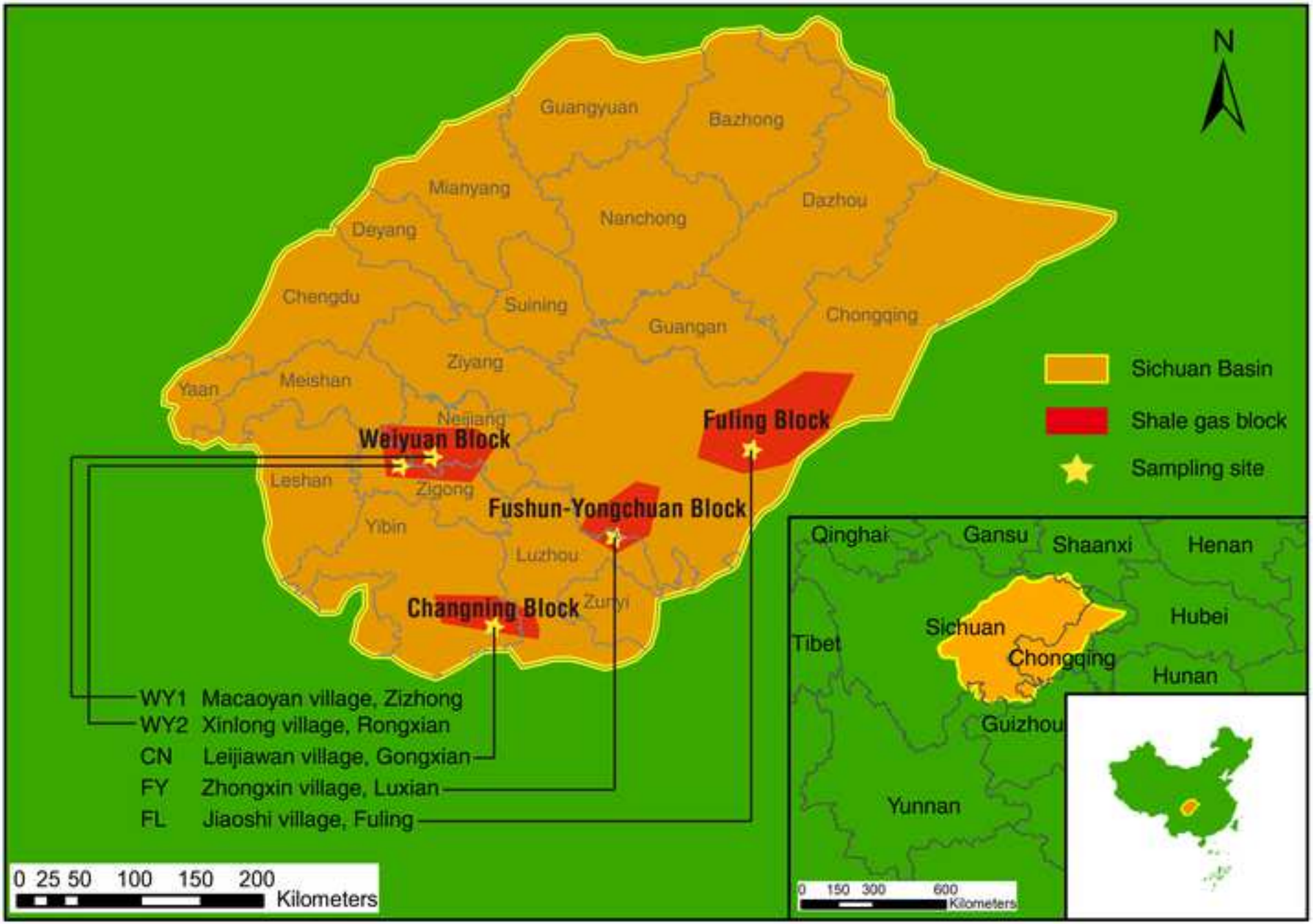


Figure 2

Click here to access/download;Figure;Figure 2.tif

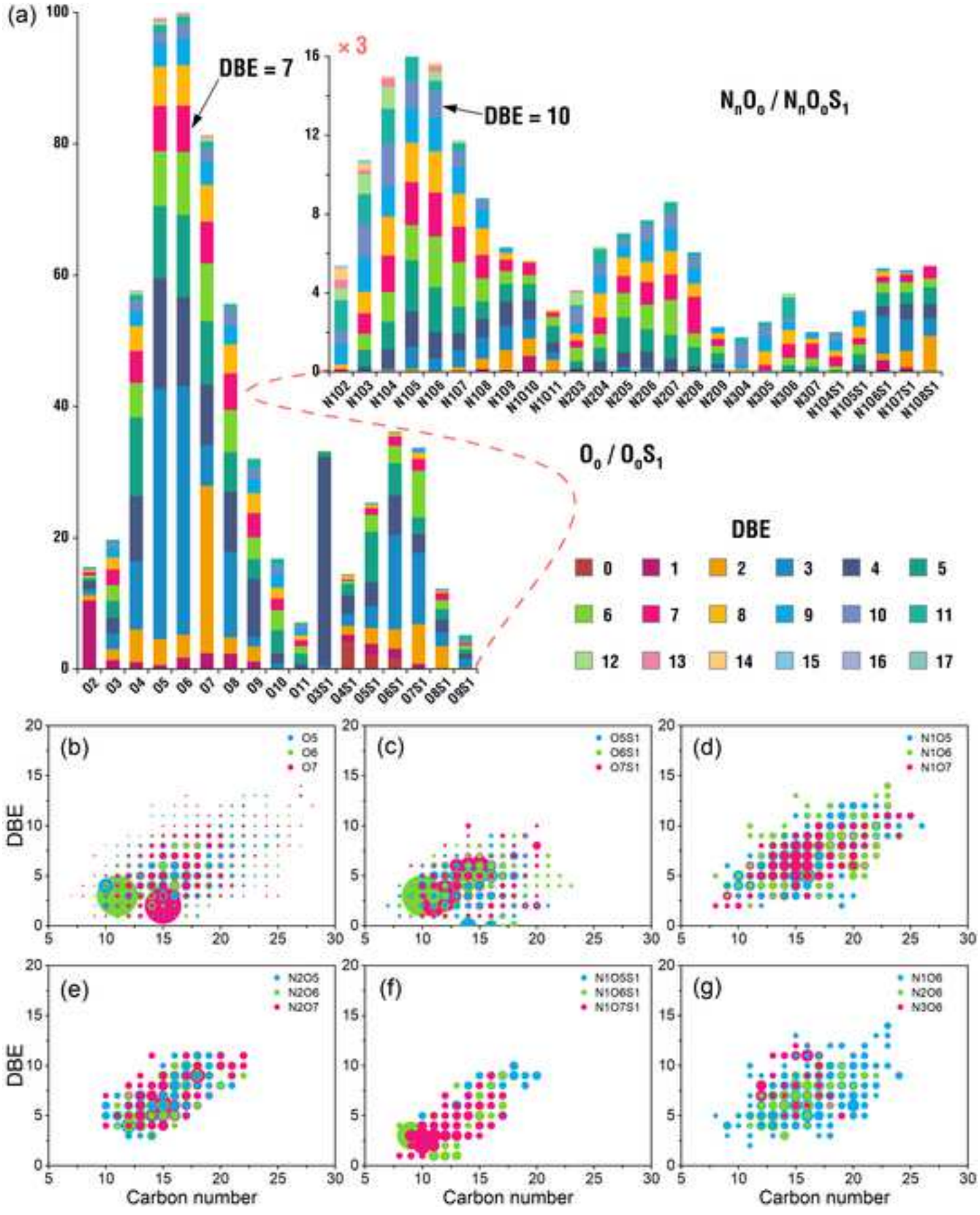


Figure 3

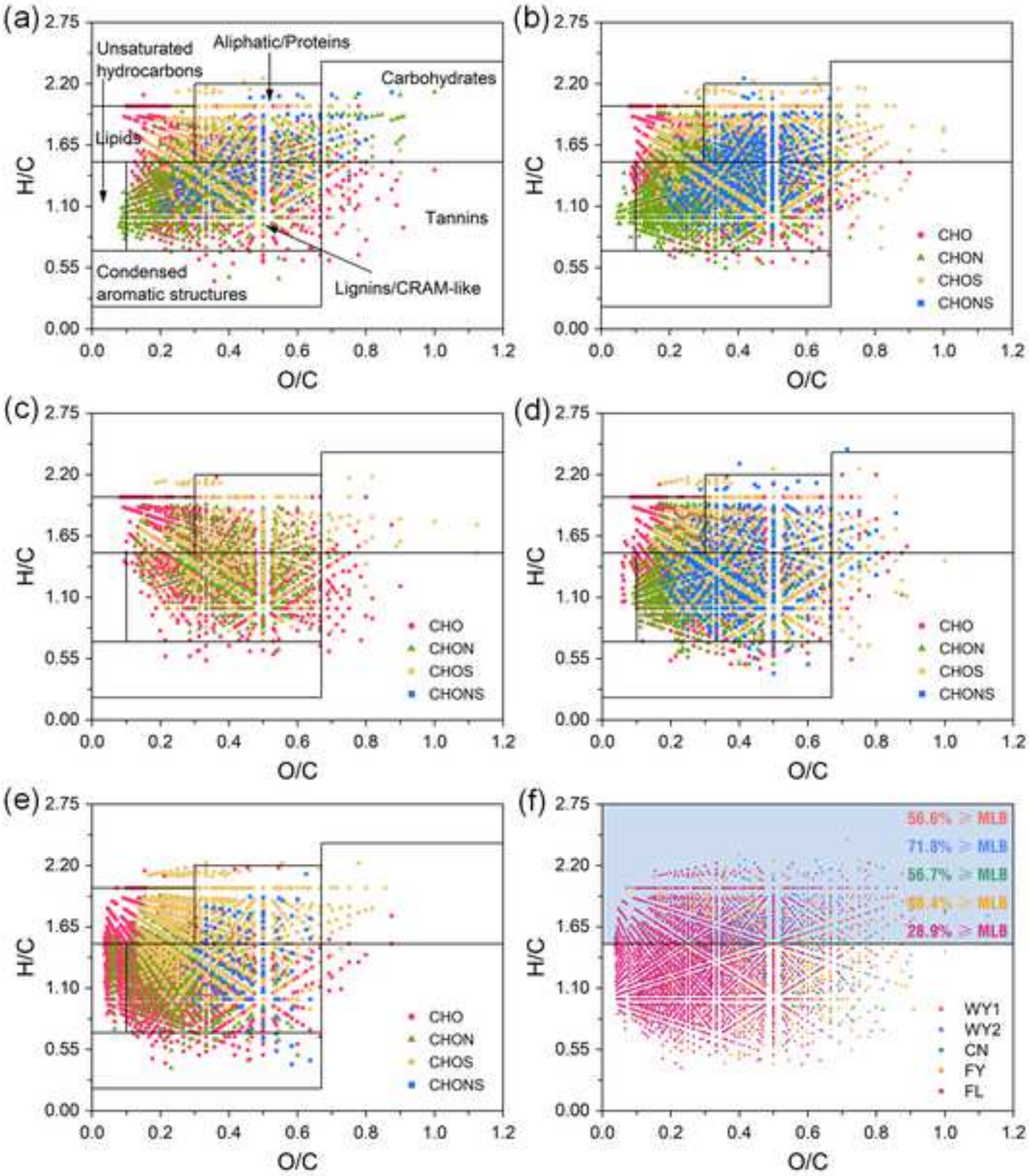


Figure 4

Click here to access/download;Figure;Figure 4.tif